FIATS-M 2011

Boris Shishkov (Ed.)

Future Internet Applications for Traffic Surveillance and Management

Proceedings of the
1st International Workshop on
Future Internet Applications for Traffic Surveillance and Management
FIATS-M 2011

Sofia - Bulgaria, October, 2011

SciTePress

Organized by

CREST

In cooperation with

CTIT
Centre for Telematics and Information Technology

Boris Shishkov (Ed.)

Future Internet Applications for Traffic Surveillance and Management

Proceedings of the
1st International Workshop on
Future Internet Applications for Traffic Surveillance and Management
FIATS-M 2011

Sofia, Bulgaria, October 2011

SciTePress
Portugal

Volume Editor

Boris Shishkov (Bulgaria)

1st International Workshop on
Future Internet Applications for Traffic Surveillance and Management
Sofia, Bulgaria, October 2011
Boris Shishkov (Ed.)

Printed in Portugal

ISBN: 978-989-8425-87-4
Depósito Legal: 338082/11

Foreword

This volume contains the proceedings of the First International Workshop on Future Internet Applications for Traffic Surveillance and Management (FIATS-M 2011), held on October 27-28 in Sofia, Bulgaria.

The FIATS-M workshop is a scientific event of IICREST, the *Interdisciplinary Institute for Collaboration and Research on Enterprise Systems and Technology*. FIATS-M provides a platform to researchers and practitioners, from academia and industry, to discuss challenges, solutions, ideas, and experiences that relate to the broad field of road traffic management. Each year, a special theme is chosen within this broad field, in order to make presentations and discussions more focused. The theme of FIATS-M 2011 is: **Service-Oriented and Context-Aware Road Traffic Surveillance and Management Systems**.

It is agreed that road traffic is directly related to sustainability and resource optimization. Hence, it is considered worthwhile proposing innovative solutions that concern road traffic. Moreover, improvements in this direction would have also other positive effects, such as increased safety of people. Even though this is widely recognized, we still miss exhaustive (IT-driven) technological solutions to the major road traffic – related problems. There are examples of technological achievements, such as highly dense electromagnetic induction loop sensor networks, car sensing systems, automated and remotely operated traffic lights, and advanced surveillance facilities. Nonetheless, these have never been adequately integrated in an IT system that essentially monitors and supports the road traffic or at least such a system has not been successfully deployed. Moreover, the personalized needs and emotions of drivers have not been factored into the design of these systems. Research on ad hoc mobile networks is actively pursued, and the information infrastructures are available to create them. However, information management systems are not keeping pace with the potential offered up by this network technology. For example, besides on-car sensors simply sensing that the host car is slowing down, cars may also inform each other that they are slowing down. Information about a traffic situation can be backwards propagated to let others anticipate dangerous situations. Traffic information can be exchanged, i.e., at the micro level, where a vehicle can inform others in the vicinity about their intentions/actions, to help prevent accidents. Information exchange can also happen at a macro level, e.g. cars, through their navigation systems, may exchange their destination information. Traffic management systems may use this information to optimize traffic management by, for example, adapting flows through changing the timing of traffic lights.

Hence, the contribution of FIATS-M 2011 concerns the goal of relating all these issues (mentioned above) to the challenge of designing and developing powerful and deployable service-oriented and context-aware road traffic surveillance and management systems able to support individual mobility and network wide operations. By 'service-oriented' we mean operating through services whose underlying technological complexity remains 'hidden' from users. By 'context-aware' we mean adapting their behavior depending on the surrounding context.

Following the FIATS-M'11 Call for Papers and received submissions, 10 papers were selected for a 30-minutes oral presentation during the workshop and for publication in these proceedings. The selected papers are a good illustration of different topics relevant to the above-mentioned challenge, that are currently under research: some papers are more oriented towards business modeling as well as (technology-independent) models for road traffic services and technologies while other papers are directed more to traffic flow modeling, pattern matching, and network monitoring/management; there are also some papers touching upon traffic data acquisition and information quality while others address user modeling.

Taking this opportunity, we would like to express our gratitude to all who have contributed to FIATS-M 2011, including the authors (who have provided the main content for these proceedings) as well as the organizers (who have had crucial role in building up the FIATS-M Community). We would also like to thank Vitor Pedrosa for the excellent work and support in preparing the proceedings, and also for his great collaboration in general. Last but not least, we tremendously appreciate the willingness of SciTePress to publish the proceedings.

We wish all presenters and attendees an inspiring workshop and a nice stay in the beautiful city of Sofia.

October 2011

Boris Shishkov

Workshop Chair

Boris Shishkov
IICREST, Bulgaria

Supporting Organizations

CTIT - Centre for Telematics and Information Technology
(The Netherlands)

INSTICC - Institute for Systems and Technologies of Information, Control and Communication (Portugal)

QlikTech Netherlands B.V.
(The Netherlands)

Table of Contents

Papers

Papers

Technology-independent Functionality Models for Road Traffic Systems

Boris Shishkov

IICREST
53 Ivan Susanin Str., 1618 Sofia, Bulgaria
b.b.shishkov@iicrest.eu

Abstract. Improving road traffic management, by addressing concerns including safety, pollution, congestion, and travel time is important. Adequately specifying technology-independent functionality models for road traffic surveillance and management systems is the focus of this work, as well as the adequate specification of corresponding supportive information systems. Technology-independent functionality models are not only needed for better understanding the problems and discussing them of full value with both developers and users but also for establishing appropriate traceability that would allow updating the technology accordingly based on desired updates in the (real-life) business processes and regulations. The contribution of our work is thus in the following directions: (i) Methodological support that concerns the technology-independent functionality modeling for road traffic systems and (related to them) supportive information systems; (ii) Conceptualization of incorporating service-orientation and context-awareness in the functionality of specified systems.

1 Introduction

We observe overpopulation as well as increasing standards of living in many urban areas around the world and hence mobility (of people and goods) is becoming crucially important. In such areas, mobility is realized mainly by means of road transport and rail transport. Logically, road traffic is mostly contributing to transportation-related problems because of its being more complex (as flow) and less predictable (with very many human beings involved), compared to rail traffic. Further, road transport provides higher mobility of people and goods (compared to rail transport) with its door-to-door service [1]. We thus focus on road traffic as an important means of transport for people and goods. In particular, we are interested in finding solutions to the biggest concerns (in our view) related to road traffic, namely: safety, pollution, congestion, and travel time.

The development of traffic surveillance and management systems that go beyond the current state of the art and which can be made service-oriented and context-aware and which are capable of supporting both individual mobility and optimal network wide management is claimed to be crucial in this regard [24]. Adequately specifying technology-independent functionality models for such road traffic systems is the focus of this work as well as the adequate specification of corresponding supportive information systems. Technology-independent functionality models are not only needed for better understanding the problems and discussing them of full value with both developers and users but also for establishing appropriate traceability that would allow updating the technology accordingly based on desired updates in the (real-life) business processes and regulations [5]. Nevertheless, most current efforts in this direction have their underlying experimental applications which are specific to the type of sensors and other technological facilities they depend on; for example, there may be reference to image processing algorithms and related image feature extraction techniques that aim at capturing traffic dynamics along a road. Anyway, although such examples require specific technologies in their implementation, an abstraction could be made that generalizes the contemporary trends in traffic management approaches. Further, the traffic related parameters that need to be measured remain the same, i.e. traffic volume in cars per time unit, car speed, road lane occupancy, length of vehicle per lane, gap between successive cars etc, as they are essential in a following stage to feed the traffic prediction models that traffic management agencies apply in their daily routine operations [1]. The contribution of our work is thus in the following directions: (i) Methodological support that concerns the technology-independent functionality modeling for road traffic systems and (related to them) supportive information systems; (ii) Conceptualization of incorporating service-orientation and context-awareness in the functionality of specified systems.

The outline of the remaining of the current paper is as follows: Section 2 outlines our conceptual views concerning not only the stakeholders and processes related to road traffic but also the role of road traffic systems and (related to them) supportive information systems. Road traffic systems and their supportive information systems are further addressed in more detail in Section 3, where the specification of a road traffic system as a business system and the design of a supportive information system are discussed. Relating this to service-oriented and context-aware solutions is then addressed in Section 4. Further, Section 5 contains partial exemplification of the solution directions proposed in the previous sections. Section 6 contains a brief analysis of related work. Finally, Section 7 contains the conclusions.

2 Conceptual Views

Presenting our conceptual views concerning not only the stakeholders and processes related to road traffic but also the role of road traffic systems and (related to them) supportive information systems, is considered important because any adequate IT support should be consistent with stakeholders' activities and their facilitations [16].

In the first sub-section of the current section, our discussion addresses stakeholders and processes; in Sub-Section 2.2, we discuss the roles of road traffic systems and (related to them) supportive information systems.

2.1 Stakeholders and Processes

As already mentioned we will present our views on the main stakeholders and processes relevant to road traffic.

Considering roads, in general, one would logically observe different entities and processes, as shown on Figure 1. The most typical entity types are vehicles (for whom roads are built), human beings (whose pedestrian activities (as opposed to their activities as drivers of vehicles) are 'interrupted' by road traffic), and devices (supposed to help controlling road traffic); the most typical process types are driving (a vehicle), road crossing (by a human being), monitoring (by a human being and/or a device), and sending control commands (by a human being and/or a device). Further, some entity (e) /process (p) types can fall into several different categories:

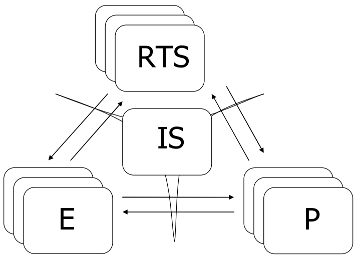

Fig. 1. Conceptual view on Entities (E), Processes (P), Road Traffic Systems (RTS) and (related to them) supportive Information Systems (IS).

(e-i) With regard to vehicles, they can be categorized based on their general purpose (cars, trucks, buses, and so on), based on their ‘institutionalization’ (private/municipal vehicle, police vehicle, ambulance, customs vehicle, and so on), and so on;

(e-ii) With regard to human beings, they can be categorized based on gender (male/female), based on age (child/grown-up), based on their function (pedestrian/driver/policeman/…), and so on;

(e-iii) With regard to devices, they can be categorized based on their general purpose (signal devices, such as traffic lights, monitoring devices, such as sensors, and so on) or based on their placement (fixed devices, for example attached to traffic lights, mobile devices, for example attached to cars, and so on), and so on;

(p-iv) With regard to driving, it can be categorized based on urgency (normal/emergency driving), based on lawfulness (lawful/illegal driving), and so on;

(p-v) With regard to road crossing, it can be categorized based on authority (pedestrian/policeman road crossing), based on lawfulness (lawful/illegal road crossing), and so on;

(p-vi) With regard to monitoring, it can be categorized based on the beneficiary (police/statistics/…), continuity (constant/upon-violation), and so on;

(p-vii) With regard to sending control commands, it can be categorized based on the location type (on-the-road/crossroad controlling), based on the signal type (traffic lights, speed instruction, and so on), and so on.

This is illustrated in Figure 2:

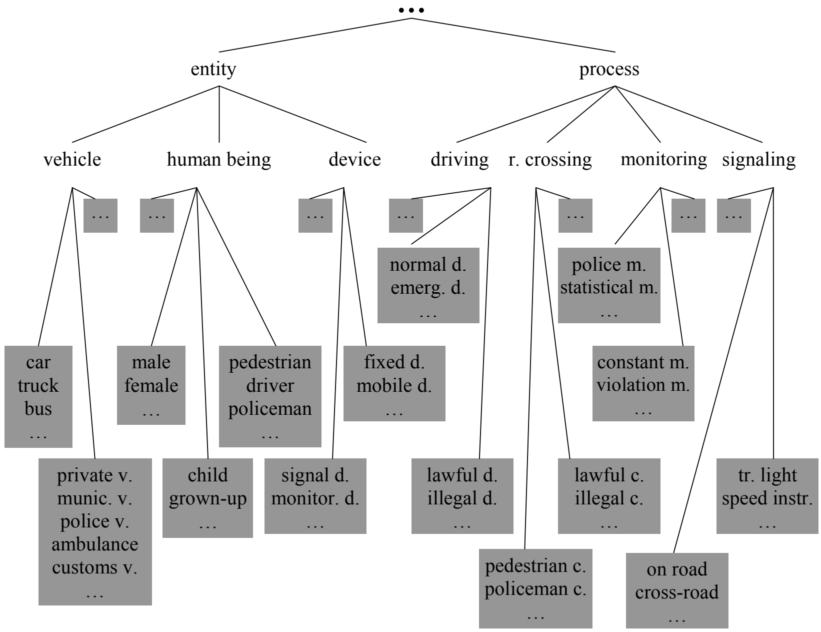

Fig. 2. View on main types and categories of entities and processes in road traffic.

Discussing so far just entities and processes, we have not discussed the roles of road traffic systems and (related to them) supportive information systems. We will do this in Sub-Section 2.2.

2.2 Road Traffic Systems and Information Systems

Road Traffic Systems and (related to them) supportive Information Systems, as exhibited in Figure 1, have great importance in managing road traffic, because such systems should guarantee that all involved stakeholders follow the regulations and are appropriately instructed [1].

Road Traffic Systems (or RTSs, for short) are referred to (in this paper) as all human beings, facilities and regulations realizing surveillance and management (including synchronization) with regard to stakeholders and processes that concern road traffic. These are typically the traffic lights, the instruction boards, the surveillance devices, and so on, as well as the corresponding regulations and supportive personnel.

Information Systems (or ISs, for short) are also needed, for supporting RTS and stakeholders with all that relates to information storage and information processing, and this mainly concerns sensor data – all raw data collected through sensors, needs to be processed and stored properly.

Hence, an RTS is responsible mainly for control and synchronization, and an IS – mainly for data storage and processing, as illustrated in Fig. 3:

Finally, the overall picture, as presented in Fig. 1, suggests that: there is link not only between the stakeholders (entities) involved in road traffic and corresponding

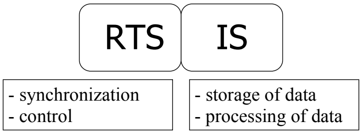

Fig. 3. Main responsibilities of Road Traffic Systems (RTS) and (related to them) supportive Information Systems (IS).

processes (for example, between a driver and a security check the driver is to pass through) but also between entities / processes and RTS (this is obvious since it is expected that RTSs establish synchronization and control); further, this all comes 'through' an IS involvement since all that takes place in managing road traffic is about (underlying) data flows.

It is thus concluded that both RTS and IS have great importance in road traffic management. Further, a technology-independent view on such crucial systems is essential, as already mentioned in the Introduction. Hence, we will especially address system design in the following section, from a technology-independent perspective, relating this particularly to the design of RTS and IS – for road traffic management.

3 System Design Views

We consider an RTS (Road Traffic System) and an IS (Information System) as a business system and an information system, respectively, following the terminology in [5]:

- A *system* is characterized by composition (it is composed of some entities), environment (there should be a delimitation between what is inside the system and what remains outside it), and structure (there are relations concerning the composition and the environment).

- A *business system* is composed of human beings collaborating among each other through actions which are driven by the goal of delivering products to entities belonging to the environment of the system; further, it may be that there are technical devices involved but it is the responsible human being 'behind' who 'counts', not the device itself.

- An *information system* is composed of human beings (often facilitated by ICT applications as well as technical and/or technological facilities) collaborating among each other driven by the goal of supporting informationally a corresponding business system.

Hence, we will discuss firstly the specification of an RTS and then information systems design.

3.1 Specifying an RTS

With regard to specifying a business system, in general, and an RTS in particular, we would follow principles of the Language-Action Perspective (LAP) and Organizational Semiotics [3] because of their proven strengths concerning technology-independent modeling of business systems [5]. Due to the limited scope of this paper, we will only briefly discuss some important issues; interested readers can find more about LAP in [6,23] and about Organizational Semiotics in [4].

The **first challenge** is specifying the RTS functionality in terms of actor roles and interactions that concern these roles as well as drawing the delimitation between what is inside the system and what remains outside it.

With regard to this challenge, we start by addressing *actor roles*. Here it is to be mentioned that in considering an action, we need to take into account the human responsibility behind rather than considering just the human/technical entity (partially) realizing the action [5]. Take for example a surveillance camera; taking it as a technical device on its own would bring little value because it is of crucial importance also what is the purpose behind - it is one thing if the camera is used to monitor traffic for statistical purposes and it is another thing if the camera is used to 'catch' violations. Take the latter example. In this case, whatever the technical facilitation, the responsible human being is of essential importance - the human being behind who is responsible for the inspection, including also the decision to impose sanction, for example. Thus, a typical actor role here would be VIOLATION INSPECTOR; fulfilling this role would include monitoring, violation identification, sanction determination, and so on. Each of these 'atomic' activities point to the responsibility of the violation inspector but it is possible that some of them are realized by a technical device (for example, the monitoring); still, this would be in the end responsibility of the violation inspector (a machine obviously cannot be kept responsible).

As *interactions* are concerned, they are considered always as occurring between two actors (we speak of an *actor* when we have an actor role fulfilled by a concrete entity) and if the setting is more complex, it is assumed that decomposition to sets of interactions involving exactly two actors is possible [6]. Further, we apply the *request-promise-state-accept* interaction pattern that is inspired by the Language-Action Perspective: according to this pattern, one of the actors would execute something (this execution would lead to the *production fact* characterizing the interaction), in response to a request by the other actor; once the production fact has occurred, this has to be announced to the requesting actor (state) and accepted by this actor (accept). This interaction pattern allows for modeling real-life business processes in a realistic way, as studied in [23]. For the sake of brevity, we will not discuss this further in the current paper, leaving interested readers to find more in the mentioned references.

Finally, it is of crucial importance to adequately draw *delimitation* between what (which actor roles) is inside the business system (in our case – the RTS) and what remains in its environment. We do not consider any particular guidelines for this, it just has to be done when identifying actor roles and interactions; then it could be that we identify: (i) interactions concerning only actor roles that are inside the business system; (ii) interactions concerning only actor roles that are in the environment of the business system; (iii) interactions 'crossing over' – one actor role is inside the business system and the other actor role is in the environment of the business system.

The **second challenge** is establishing the regulations underlying the interactions within an RTS and with regard to this, we consider as background Organizational Semiotics, in general, and NAN - the Norm Analysis Method [4], in particular.

Norms, which include formal and informal rules and regulations, define the dynamic conditions of the pattern of behavior existing in a community and govern how its members (agents) behave, think, make judgments and perceive the world.

Norms are developed through practical experiences of agents in a community, and in turn have functions of directing, coordinating and controlling their actions within the community. When modeling agents and their actions, which may reveal the repertoire of available behaviors of agents, norms will supply rationale for actions. Norms will also provide guidance for members to determine whether certain patterns of behavior are legal or acceptable within a given context. An individual member in the community, having learned the norms, will be able to use the knowledge to guide his or her actions, though he or she may decide to take either a norm-conforming or a norm-breaking action. When the norms of an organization are learned, it will be possible for one to expect and predict behavior and to collaborate with others in performing coordinated actions. Once the norms are understood, captured and represented in, for example, the form of deontic logic, it will serve as a basis for programming intelligent agents to perform many regular activities.

The long established classification of norms distinguishes between perceptual, evaluative, cognitive and behavioral norms; each governing human behavior from different aspects. However, in business process modeling, most rules and regulations fall into the category of behavioral norms. These norms prescribe what people must, may, and must not do, which are equivalent to three deontic operators of “obligation”, “permission”, and “prohibition”. Hence, the following format is considered suitable for specification of behavioral norms:

whenever <condition>
if <state>
then <agent>
is <deontic operator>
to <action>

We give also an example for a norm governing interest charges concerning a fine to be paid by a driver:

whenever an amount of outstanding fine
if more than 25 days after posting
then the driver
is obliged
to pay the interest

Thus, semiotic norms can be useful in elaborating actor roles and/or interactions. Nevertheless, even though identified when actor roles / interactions are considered, each norm needs to be analyzed further and positioned in a corresponding hierarchy of norms where higher level (‘constitutional’) norms govern lower level norms [3]. For the sake of brevity, we will not go in more detail regarding this.

In summary, business systems, such as RTS, are to be specified methodologically and applying the request-promise-state-accept LAP pattern can be helpful if actor roles and corresponding interactions are properly identified. Considering underlying regulations is as well necessary and can be usefully done through semiotic norms (in particular – behavioral norms). This all concerns the specification of a technology-independent model that may be further elaborated in terms of technical platforms and/or facilities/devices which may be (partially) fulfilling some actor roles; however, considering this is left beyond the scope of this paper, as already mentioned.

3.2 IS Design

Systems, such as RTS, are business systems in a sense that they are neither essentially automating something (even though there may be technical devices within them) nor they are there to informationally support particular processes. It is only that some tasks are performed by human beings and other tasks may be preformed by technical devices. Still, there are neither underlying IT architectures nor IT system requirements. Thus, with regard to RTS we only need to have a business process model, we need to identify actor roles and interactions, and also address the underlying regulations. This all holds also for the IS design but there are other issues as well to be taken into account when considering an information system. What is most important is IS's supportive role to what already exists, and this can be (as mentioned above) achieved through: (i) automation of processes or (ii) informational support with regard to processes. What we are addressing in the current paper, and also following what is outlined in Figure 1, is in line with (ii); hence, the role of the IS (as considered in the paper) is to support processes involving different stakeholders and RTS.

Hence, we are not reflecting 'real-life' level entities in IS entities but we are to specify new entities and new processes (for the IS). We firstly need to establish what is the desired support to be delivered (by the IS) and then go to the identification and specification of requirements.

Since requirements relate directly to the IS specification [2], they are to describe properly how the system-to-be should behave; application domain information is also needed as well as information on constraints on the system's operation. Considering all this, it could be distinguished between requirements that concern the IS fitting in its environment (*domain-imposed requirements*) and requirements that concern the desired (technical) characteristics of the system (*user-defined requirements*). For more information on requirements identification and specification, interested readers are referred to [5].

The requirements model is to be the basis for identifying actor roles and interactions, and also projecting important underlying regulations – as discussed in the previous sub-section.

We would need then the right-granularity functionality pieces, for furthering the information system design, and these can in principle correspond to any of the following: (i) a particular actor role can be reflected in a piece of functionality; (ii) an actor role can be reflected in several functionality pieces; (iii) several actor roles together can be reflected in a piece of functionality. This choice would depend on suitabilities with regard to the applied technology platform(s).

Such functionality pieces can be supported either by system components or by external IT services [18].

IT service solutions are more interesting for us not only because service-orientation becomes increasingly popular [9] but also because relying on (global) IT services is characterizing most current road traffic systems, as it is well known.

For this reason, we will continue the discussion in the following section, from a service perspective, suggesting possible benefits from applying distributed web services and (related to them) context-aware solutions. Then, in Section 5 we will provide partial exemplification of what has been proposed in the paper.

4 Towards Distributed and Adaptable IT Service Solutions for Traffic Management

As already mentioned, this section discusses service-orientation and adaptability (achieved through context awareness), on top of the system specification visions presented in the previous section.

4.1 Distributed Web Services

Most of the latest software technology innovations are centered around the service concept [8] and service-orientation in turn demands a heavy distribution [8,7]. We thus consider these together in the current sub-section, starting with a discussion on the service concept.

From an abstract point of view, a service represents a piece of well-defined functionality that is available at some network endpoint and is accessible via various transport protocols and specialization formats [21]. The functionalities provided by services cover a vast spectrum reaching from low level features like offering storage capabilities, over simple application functions like changing a customer address, to complex business processes like hiring a new employee [18].

The ability to create new ICT applications from existing services, independently on who provides these services, where they are provided, and how they are implemented, would mean usefully utilizing the service perspective in application development [16]. Such kind of application development is innovative not only because the application is not constructed from the scratch (actually, this is true also for component-based application development) but also because the development itself is fully centered around the desired end functionality to be consumed by users (this leads to service compositions and hence developers would no longer possess full control over all software components that play roles in delivering the application functionality). Hence, the application development task (as considered in general) might split into: (i) development of small software modules delivering generic adjustable services to whoever might be interested in using them, and (ii) composition of complex functionalities, by using available generic services. This all inspires new middleware developments also [17].

According to such latest middleware visions, on the basis of the specification of a needed functionality, the middleware would determine a composite service that would deliver the required functionality. This would be done either automatically or through a developer's intervention. This would obviously concern not only the application development but also its performance because services would require in most cases processing power of back-end server systems – hence both application creation and application execution would rely on support from the 'Cloud', especially when web services (services created and executed through the Web) are considered.

Furthermore, in order to be of actual use, such services would demand enabling technology standards and some recent views of Papazoglou [8] appear to be actual in this respect. Transportation protocols are to be mentioned firstly because logically, web services' relying on a transportation protocol is crucial. Although not tied to any specific transportation protocol, web services build on ubiquitous Internet connectivity and infrastructure to ensure nearly universal reach and support. Hence, their mostly relying on HTTP (the connection protocol that is used by web services and browsers) and XML (a widely accepted format for all exchanging data and its corresponding semantics) looks logical. Having this as foundation, we have to briefly discuss three core web service standards, namely SOAP, WSDL, and UDDI: (i) SOAP (Simple Object Access Protocol) is a simple XML-based messaging protocol on which web services rely in exchanging among themselves information. SOAP implements a request/response model for communication between interacting web services. (ii) WSDL (Web Service Description Language) is a language that specifies the interface of a web service, providing to the requestors a description of the service in this way. (iii) UDDI (Universal Description, Discovery, and Integration) represents a public directory that not only provides the publication of online services but also facilitates their eventual discovery. And finally, as part of the web services composition, we need to introduce some orchestration defining their control flows [18], such as sequential, parallel, conditional, and so on, and to also determine complex processes that would usually span many parties. BPEL4WS (Business Process Execution Language for Web Services) can usefully support such composition activities [22]. Finally, as the collaboration among many parties (through their web services) is concerned, a common observable behavior (choreography) would often need to be defined. CDL4WS (Choreography Description Language for Web Services) can usefully support such collaboration descriptions.

4.2 Adaptable Context-aware Solutions

The utilization of a generic service for a specific user-related situation logically relates to acquiring knowledge on the context of the user and also exploiting this knowledge to provide the best possible service, which is labeled as context-awareness by Shishkov & Van Sinderen [19]. We hence claim that taking the end-user context into account is important in adequately delivering a service. Examples of end-user context are the location of the user, the user's activity, the availability of the user, and so on. We do assume that the end-user is in different contexts over time, and as a consequence (s)he has changing preferences or needs with respect to services. Relating this to application creation and execution: the application is informed (for example through sensors) of the context (or of context changes), where the sensing is done as unobtrusively (and invisibly) for the end-user as possible. Sensors sample the user's environment and produce (primitive) context information, which is an approximation of the actual context, suitable for computer interpretation and processing. Higher level context information may be derived through inference and aggregation (using input from multiple sensors) before it is presented to applications which in turn can decide on the current context of the end-user and the corresponding service(s) that must be offered [20].

As according to previous studies, there are three challenges related to achieving adequate context aware support, namely: capturing, delivery, and response capacity [16]: (i) it appears to be non trivial establishing (through sensors) what change in the end user's situation has occurred – raw sensor data can easily be misinterpreted and also establishing adequate Quality-of-Context levels remains a challenge; (ii) even though deterministic service delivery approaches (when it is clear still at design time what behavior corresponds to what end user situation) should work OK, situations would occur which could not have been predicted at design time and thus require non-deterministic solutions – with respect to this, we doubt how much reliable such solutions would be, taking into account observed failures in many artificial-intelligence –related projects. We thus need better link to data analysis and more powerful run-time solution facilities. (iii) When delivering non-deterministic solutions, it may happen that such services are required that need unavailable resources; this would hence make it more difficult delivering timely the proper service(s) to the end-user.

5 Illustrating Example

As mentioned at the end of Section 3, we provide in the current section partial exemplification with regard to what has been proposed in this paper as solution directions. It needs to be emphasized however that the exemplification is just partial, due to the limited scope of the current paper, and the example is deliberately a 'toy example', adapted from another case study, since the purpose is just to briefly illustrate some important modeling steps discussed in the previous sections.

We take a monitoring-related example, putting things such that road traffic is monitored for 'catching' violations of traffic rules. There is a mediating system, called 'The Monitoring Mediator' (or 'Mediator', for short), this system is part of the Information System (IS), and is supporting stakeholders and the road traffic system, by informationally checking and advising on whether or not a violation has occurred. To make the example simpler, we abstract from the rest of the IS and also from the road traffic system and the stakeholders. We abstractly introduce the 'Customer' as the one who receives the advice that is delivered by the Mediator.

Thus, as it can be seen from Fig. 4, the Customer requests and the Mediator delivers an advice – if we put things as simple as that, we may have just these two actor roles, namely Customer and Mediator. We need however to 'peek' inside the Mediator, in order to be able to specify its functionality, and as seen from the figure, we go to another level of granularity where we can model actor roles which are internal with regard to the Mediator. Considering requirements and other issues (not mentioned for brevity), we arrive at a model where the following 4 actor roles are presented (these are 4 major roles and the model does not claim exhaustiveness): (i) Advisor (A) – delivering the advise to the Customer (C) (on behalf of the Mediator) on whether or not a violation has occurred; (ii) Match-maker (MM) – realizing match between what has been captured (through sensors) and analyzed (through reasoning), on one hand,

and what are the violations specifications (as it is according to the law), on the other hand; (iii) Request processor (R) – taking raw sensor data and realizing interpretation through reasoning techniques; (iv) Data searcher (D) – making a list of all relevant violation specifications as according to the law. Hence, the Mediator can deliver the advice (through A) to C, only after MM has delivered to A, which in turn needs to be preceded by the deliveries of R and D.

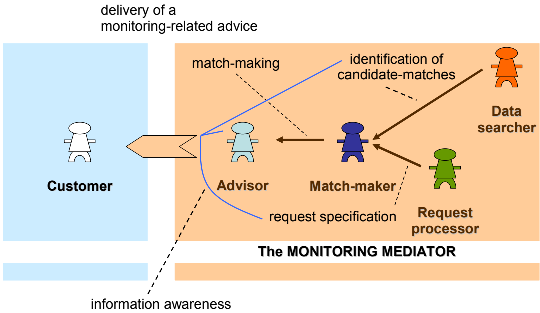

Fig. 4. General business process model for the monitoring mediation case.

We then go to the Entity model, as exhibited in Figure 5, where we have the dashed line delimiting what is inside the system from the rest, and we have there all actor roles discussed so far. Hence, the actor roles represent the entities in this model and there are some lines connecting two entities – this reveals the structure (how the entities are related to each other). These relations (representing the interactions) follow the 'deliveries' already discussed: A delivers to C, MM delivers to A, R delivers to MM, and D delivers to MM (the grey boxes indicate the one who is delivering).

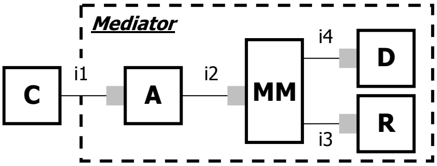

Fig. 5. Entity model for the monitoring mediation case.

The Entity model is then straightforwardly reflected in an Interaction model (Fig. 6) where the emphasis is on the interactions which are modeled using the ISDL technique [23]. As it can be seen, workflow details, such as sequence, synchronization, and so on, are reflected: as it can be seen: i3 and i4 should both be completed (synchronization) before i2 is completed, and i1 would be completed after i2 has been completed (sequence). On the top right corner of the figure is the higher-granularity-level model (as already discussed) where we look at all as one interaction (a delivery between the Mediator and the Customer).

Finally, we elaborate each interaction, by applying the request-promise-state-accept LAP pattern, as it is shown on Figure 7. The 4 grey ovals correspond to the 4 interactions' production facts and these are elaborated with corresponding coordination acts (such as request, promise, state, and accept); thus, the first 'line' from top to

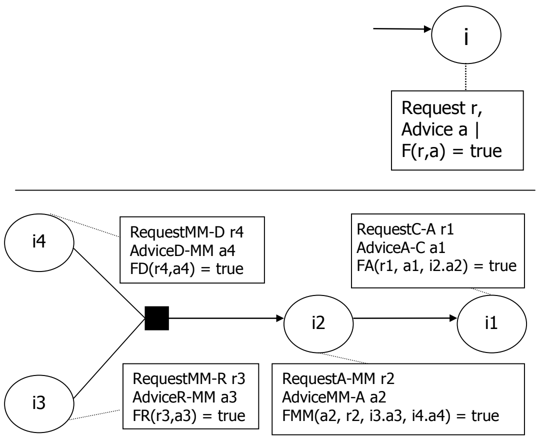

Fig. 6. Interaction model for the monitoring mediation case.

bottom is to be read as follows: Start, request by C, promise by A, production fact, state by A, accept by C, and so on. There are decision points – allowing not only for promise but also for decline (d), and these in turn allow for failure scenarios as well. We are not going to explain the model in more detail – interested readers can find more information on LAP in the references already mentioned.

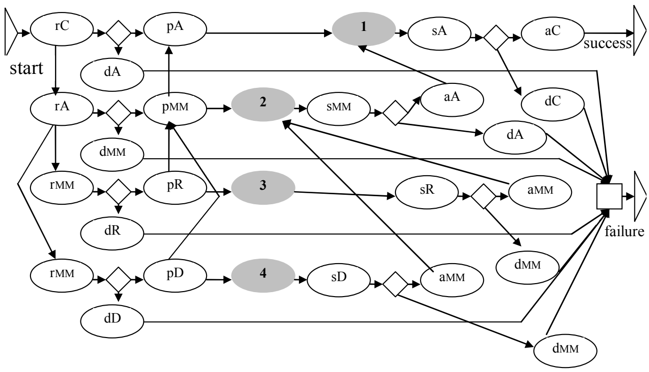

Fig. 7. Service-level model for the monitoring mediation case.

This model is considered as a 'service-level' model since on its basis one could usefully derive services.

Deriving service specifications, based on such a model, is nevertheless not trivial and requires resolving issues, such as tight coupling (possibly through introducing an orchestrating entity) and context-awareness, for instance, which would mean introducing new entities and interactions. Still, this way of progressing from very general real-life information to a well-structured and theory-rooted model is claimed to be beneficial. Such a model can be elaborated further, by adding semiotic norms.

However, for the sake of brevity, we are not going in further details with regard to this case. The purpose of this section (as also mentioned before) was to just give an impression on how part of what has been proposed can be realized through particular modeling techniques.

6 Related Work

Discussing road traffic, one could point to numerous relevant research activities, R&D projects, and initiatives. We do not claim exhaustiveness however in our brief analysis of related work, for two reasons: (i) the work reported in this paper is part of the FIATS-M initiative [1] and we find it most important to appropriately align to the works of other FIATS-M participants; (ii) for the sake of brevity, we cannot go in much detail in the current section, and we thus focus only on analyzing related work limited to the FIATS-M initiative, in general, and the work presented at FIATS-M'11, in particular.

The work presented by Brahmananda Sapkota [9] focuses on context-aware global IT services that interact among each other in delivering support to drivers and other stakeholders involved in road traffic. This work can be complementary with regard to our work since the focus there is mainly on utilizing services while in our work we address their specification rooted in methodological information system development. The work presented by Bjorn Kijl [10] focuses on assessing viability of the entire road traffic system both from a business sustainability perspective and an innovation perspective. This could link to our work since such an input may be relevant with regard to the technology-independent road traffic models that need constant analysis and updating, and this needs to be in turn reflected in the technology evolution. The work presented by Thomas Jackson [11] focuses on predicting situations on the road based on establishing a pattern whose occurrence has preceded the situation. We can relate to this in our work, by incorporating such 'mechanisms' at design time; then the run time support would adequately fit in the overall system functionality. The work presented by Apostolos Kotsialos [12] focuses on Internet based network control and this work's relation to our work is two-fold: (i) the network management and control are relevant to the functionalities of both the road traffic system and its supportive information system (as considered in our work); (ii) any Internet-based support should be realized through web services (addressed in our work). The work presented by Tom Thomas [13] focuses on monitoring and analyzing road traffic, for using this in possible re-designs. This again relates to our technology-independent models that may incorporate such 'mechanisms' at design time, for the sake of better run time support in the end. The work presented by Maria Virvou [14] focuses on user monitoring and modeling, in general, and on considering the emotional states of drivers, in particular; this is done in order to trigger such kind of support to a driver that would stimulate his or her good emotional state while on the road. If this would be applied in the context of FIATS-M, then again, it needs to be incorporates in the system specification, still at design time, and this relates the work on user monitoring and modeling to our work. Finally, the work presented by Jos Vrancken [15] focuses on network management and flows optimization for road traffic, considering not only the traffic inside the network under consideration but also on its border – different criteria may be applied in prioritizing the desired effects of traffic management.

Again, this is essential for supporting road traffic and needs to be incorporated in the systems functionalities at design time; this is how this work relates to our work.

7 Conclusions

In this paper, we have discussed the design of business/information systems and related to this specification of (context-aware) IT services, putting all this in the perspective of road traffic management, providing as basis conceptual views on road traffic. We have emphasized on the importance of properly specifying business/information systems in a technology-independent way, as a guarantee for the adequacy of the technology system to be developed on top of that, and we have provided methodological guidelines accordingly. Finally, we have partially exemplified our views.

Hence, the contribution of this paper is two-fold: (i) our having established and justified the role of technology-independent system modeling, especially with regard to the technical and technological facilitation of road traffic. (ii) the proposed solution directions and provided partial methodological guidelines (and exemplification) regarding technology-independent business/information system modeling for road traffic.

As further research, we plan projecting this work to a lower level, establishing adequate relations between technology-independent models (as focused in the current work) and technology-specific solutions.

References

1. FIATS-M (Home), http://www.fiats-m.org
2. Software Engineering Institute (Home), http://www.sei.cmu.edu
3. Shishkov, B., Dietz, J. L. G., Liu, K.: Bridging the Language-Action Perspective and Organizational Semiotics in SDBC. In ICEIS'06, 8th Int. Conf. on Enterpr. Inf. Systems (2006)
4. Liu, K.: Semiotics in Information Systems Engineering. Cambridge University Press, Cambridge (2000)
5. Shishkov, B.: Software Specification Based on Re-usable Business Components. Delft, The Netherlands: Sieca Repro (2005)
6. Dietz, J.: Enterprise Ontology, Theory and Methodology. Springer-Verlag, Berlin Heidelberg (2006)
7. Object Management Group (OMG): MDA – Guide, V1.0.1, omg/03-06-01 (2003)
8. Papazoglou, M. P.: Web Services: Principles and Technology. Pearson Pr. Hall (2008)
9. Sapkota, B. and Van Sinderen, M. J.: An Adaptive Service Platform for Traffic Management and Surveillance. In: FIATS-M'11, 1st International Workshop on Future Internet Applications for Traffic Surveillance and Management. SciTePress (2011)
10. Kijl, B., Nieuwenhuis, L. J. M., Meertens, L. O., Iacob, M. E.: Assessing the Viability of Service Innovations – a Structured Business Modeling Approach. In: FIATS-M'11, 1st International Workshop on Future Internet Applications for Traffic Surveillance and Management. SciTePress (2011)

11. Hodge, V.J., Jackson, T., Austin, J.: Intelligent Decision Support using Pattern Matching. In: FIATS-M'11, 1st International Workshop on Future Internet Applications for Traffic Surveillance and Management. SciTePress (2011)
12. Kotsialos, A.: Perspectives of Internet Based Road Network Traffic Flow Modelling and Control. In: FIATS-M'11, 1st International Workshop on Future Internet Applications for Traffic Surveillance and Management. SciTePress (2011)
13. Thomas, T. and Veensta, S.: Network Monitoring and Personalized Traffic Control: A Starting Point Based on Experiences from the Municipality of Enschede in The Netherlands. In: FIATS-M'11, 1st International Workshop on Future Internet Applications for Traffic Surveillance and Management. SciTePress (2011)
14. Virvou, M.: User Modelling and Emotion Recognition of Drivers through a Multi-Modal GPS Interface. In: FIATS-M'11, 1st International Workshop on Future Internet Applications for Traffic Surveillance and Management. SciTePress (2011)
15. Vrancken, J.: The Future of Road Traffic Management, Moving from Local to Network-wide, From Reactive to Proactive. In: FIATS-M'11, 1st International Workshop on Future Internet Applications for Traffic Surveillance and Management. SciTePress (2011)
16. Shishkov. B. and Van Sinderen, M. J.: Service-Oriented Coordination Platform for Technology-Enhanced Learning. In: I-WEST'09, 3rd International Workshop on Enterprise Systems and Technology. INSTICC Press (2009)
17. Leymann, F.: Combining Web Services and the Grid: Towards Adaptive Enterprise Applications. CAiSE Workshops (2005)
18. Van Sinderen, M. J.: From Service-Oriented Architecture to Service-Oriented Enterprise. In: I-WEST'09, 3rd International Workshop on Enterprise Systems and Technology (2009)
19. Shishkov. B. and Van Sinderen, M. J.: On the Design of Context-Aware Applications. In: IWEST'08, 2nd Int. Workshop on Enterprise Systems and Technology. INSTICC Press (2008)
20. Alonso, G., Casati, F., Kuno, H., Machiraju, V.: Web Services, Concepts, Architectures and Applications. Springer-Verlag, Berlin-Heidelberg (2004)
21. Bosworth, A.: Developing Web Services. In: 17th Int. Conference on Data Engineering (2001)
22. Pasley, J.: How BPEL and SOA are Changing Web Services Development. Internet Computing, IEEE (2005)
23. Shishkov, B., Van Sinderen, M.J., Quartel, D.A.C.: SOA-Driven Business-Software Alignment. In: ICEBE'06, 2nd Int. Conference on e-Business Engineering. IEEE Computer Society (2006)
24. Vrancken, J. L. M., Schuppen, J. H. van, Soares, M. S., Ottenhof, F.: A Hierarchical Network Model for Road Traffic Control. In: Proceedings of the 2009 IEEE Int. Conference on Networking, Sensing and Control, Okayama, Japan (2009)

An Adaptive Service Platform for Traffic Management and Surveillance

Brahmananda Sapkota and Marten van Sinderen

Department of Electrical Engineering, Mathematics and Computer Science
University of Twente
The Netherlands
{b.sapkota, m.j.vansinderen}@utwente.nl

Abstract. The increasing number of road vehicles has given rise to increasingly adverse consequences in the society. Some of the major concerns that arise due to such an increase in road vehicles are: safety of the people using the road, cost and efficiency of the traffic management and the environmental footprint in terms of, e.g., air quality, acidification, climate change and noise pollution. Moreover, the increased road traffic, if not managed properly leads to severe congestion resulting into increased delay for people and goods. We argue that we can alleviate these problems when using an ICT-based service platform that supports localized monitoring and management of traffic and environmental information collected from various information sources such as sensors, surveillance camera, weather station, etc. Such information should be made available through services in order to increase reusability, loose coupling and management of different information and their analysis. In this paper, we discuss different functionalities that can be or even should be supported by the service platform and provide an architecture, following the service-oriented architecture principles, of such a platform.

1 Introduction

The mobility of people as well as goods is one of the main reasons for having road infrastructures. Motorized road vehicles have increased the efficiency of mobility enabling people and goods to travel to longer distances in shorter time. The number of road vehicles is increasing with the increase of populations as well as their socio-economic interests. This has caused the saturation of road infrastructures and has ultimately affected our daily lives. This increase in the number of vehicles has in fact given rise to increasingly adverse consequences not only in the lives of individuals but in society as a whole. Some of the major emerging concerns are: safety of the people using the road, cost and efficiency of the traffic management and the environmental footprint in terms of, e.g., air quality, acidification, climate change and noise pollution [1, 2]. The increased road traffic, if not managed properly leads to severe congestion resulting into increasing delay for people and goods.

Several attempts have been made for developing smart infrastructures and systems to efficiently manage the road traffic [3–7]. Collection and processing of road users data is performed using information and communication technologies. These systems and infrastructures are developed independently and are embedded in different vendor-specific systems. This makes it difficult for most of the current road users to take full advantage of these technologies due to the lack of a suitable service platform with shared conventions and standards. The existing solutions heavily rely on centralized servers and do not consider the information available at the surroundings of the the users (e.g., information available at the nearby user) [8]. Moreover, these solutions are designed to assist managing the road infrastructure and therefore the drivers are unable to utlise the instantly available information in their surroundings. For example, it is more appropriate to ask the vehicles ahead for the road condition instead of asking such context-dependent information from a central system.

To resolve the above mentioned issues, we need a context-aware service platform where interacting services coexist. Such a platform should provide support for: 1) uniform access to data and functionality offered by individual system providers and 2) seamless communication with vehicles as well as other services systems. In order to support this, the service platform should deal with localized monitoring and management of traffic and environmental information collected from various information sources (e.g., sensors, surveillance camera, weather condition, etc.). These information sources can be made available as information providing services. The services approach is required to increase reusability, loose coupling and management of different information and their analysis, which is otherwise inefficient because the services platform has to deal with an enormous amount of data. The services approach also allows for the provisioning of the services on the cloud which is useful in achieving performance requirements such as scalability and efficiency.

Since the service platform has to deal with real-time information several challenges exist. We need to define a service description that allows to specify context information. A service composition framework capable of utilising real-time (context) information is required to provide useful information to the road users. This composition framework should be guided by service intelligence to exploit knowledge acquired from analyzing observed behavior of the road users. It is also required to deal with service management (e.g. lifecycle, versioning) because the services need to be composed on a per user basis and are usually short-lived. Generic functions available in vehicles (e.g., cars) like speed information, engine parameters, gas usage etc. need to be provided using standardized interfaces. The differences between existing vendor-specific technology platforms and applications have to be bridged by an open service platform for the automotive industry. Automotive telematics services have to be identified, satisfying the needs of the business domain and using the technical capabilities of the network and sensor domain. Such services, possibly run on the open service platform, may also interact with information and computation services in the public domain, e.g., in the cloud, in order to reach its full potential.

In this paper, we focus on the problem of providing a services platform to support communication and coordination between road users, road infrastructure services and third party services. Through such communication and coordination, it is possible to increase road safety, to decrease environmental pollution and to increase the level of comfort of the drivers at the road. To define this service platform we follow the principles of service oriented architecture [9] and context-aware computing [10, 11]. The service oriented architecture (SOA) allows for the integration of varieties of services and technologies whereas context-aware computing allows for real-time adaptation system behaviour according to the changes in the users environment.

The rest of the paper is structured as follows: Section 2 describes an application scenario to motivate the requirements of the work presented in this paper. Section 3 highlights technical challenges that need to be considered while providing a service platform in the traffic domain. Section 4 presents an initial architecture of the proposed service platform based on SOA principles. Section 5 discusses some related works. Section 6 points some of the issues that needs to be further considered and Section 7 concludes this paper with some suggestions for future work.

2 Application Scenario

Bob lives in the outskirt of Enschede with his wife and two children. He is scheduled to have a project meeting in Sofia at 11:00 PM on Friday. He is occupied the entire day because of the kick-off meeting of his recently acquired project on Thursday. Because Bob is mostly busy with his work (delivering lectures, attending meetings, and doing research) during the weekdays, he spends his weekend with his family as much as possible. When his children know about his forthcoming trip to Sofia on Friday, they were sad that they will not see him during the weekend. So he promises his children that he will return to take them to the world-famous zoological garden in Emmen at the weekend.

He decides to travel Friday morning to Schiphol where he will take an early flight to Sofia. Since taking a train would not leave him enough time to check in, he takes his car, which is equipped with Intelligent Route Planning (IRP) agent, radio and Global Positioning System (GPS) devices.

He books the flight accordingly and downloads his e-ticket to his smartphone. When the e-ticket is downloaded, his smart phone recognizes it and wirelessly communicates with an IRP agent installed on his car. This agent communicates with the GPS device installed on the car and determines the required travel time to reach to the Schiphol airport. The IRP agent, knows that Bob normally wants to arrive at the airport 30 minutes before the normal time as suggested by the airlines and thus calculates the time Bob needs to start his journey. The IRP agent communicates this information to Bob's smart phone. Bob's mart phone then uses this information and sets his alarm accordingly.

When he follows the route shown on his GPS system, he suddenly encounters that the road is blocked because construction works. He then ignores the advice from the GPS system and drives on a different road than suggested by the GPS system. The GPS system apparently does not know about this situation and road that Bob is driving because it is a newly constructed road, it keeps advising Bob to take a U-turn if possible. Bob keeps ignoring the advice and keeps driving using his own instinct and sense of direction. After a while, the GPS system recognizes the stretch of road that Bob is driving and recalculates the route for Bob. The road that Bob was driving based on his own sense of direction turns out to be a faster section of the road in early morning travel time. The IRP agent on his car records this newly discovered route and updates the map and broadcasts the plan to the passerby cars.

While on his way, the IRP agent installed on a car coming from the opposite direction communicates information of long jam of cars 10 KM ahead because of a recent accident to the IRP agent installed on Bob's car. The IRP agent then communicates this information to the GPS system to re-calculate the route.

When he is driving on the re-calculated route, the IRP agent communicates with the Road-Side Infrastructure (RSI) and finds out that the traffic near the next junction where Bob has to turn right is congested (the RSI can determine such a situation by using information from loop detectors). The IRP agent informs Bob to change the lane well in advance. The IRP agent also predicts, based on the current weather conditions, total number of current road users and their average speed, that the joining road ahead of the next junction could have black ice. The IRP agent then informs Bob to drive at safe speed to avoid a possible slippery road condition.

When Bob drives some 100 KM, The IRP agent receives information from the RSI that there is a poor visibility 20 KM ahead of the road and schedules the light control system to brighten their light calculating the time required to reach that spot. When Bob passes the poor visibility area, the IRP agent identifies that the visibility is OK and resets the high to their original intensity through the light control system.

When at parking lot at the airport, Bob's car recognises that his friend Dave is also at the airport, and sends him an invitation for a coffee if he has time. Dave replies with a call and they meet at a nearby coffee shop. After having a chat with his friend, Bob goes to check-in his flight and leaves for Sofia.

After his meeting in Sofia, Bob returns to the Netherlands. When he lands at the Schiphol airport, he turns his smart phone on. His smart phone then wirelessly communicates with the IRP Agent at his car. The agent then communicates with the GPS system and calculates the time required to reach his home and informs his wife Alice about his arrival time. Bob then continues his journey towards his home following the route displayed on his GPS system.

After driving 45KM, the road RSI communicates to the radio device installed on his car that the road further ahead is busy (which is expected because it is a Friday night). The RPI agent receives this information through the radio device installed on Bob's car and communicates with the GPS system to recalculate the new route and new time required to reach Bob's home. It appears that Bob will arrive home 30 minutes later than previously expected, the IRP agent then informs Alice that Bob will be late by 30 minutes because of busy traffic.

The new road that Bob is driving now is relatively empty ahead of him, however there are few cars behind him. When he approaches Enschede, the IRP agent communicates with the RSI and finds that an ambulance is coming on the joining road at the junction ahead and Bob will not be able to cross it safely. The IRP agent then informs Bob to slow down because the traffic light at the junction is going to turn red because of the high priority vehicle on the other road. When he starts decelerating, the IRP agent communicates with the IRP agent on the car behind Bob (which was out of the range of RSI communication) and informs that Bob is decelerating. The IRP agent on the car behind Bob then informs his driver Tim to start decelerating to avoid possible environmental pollution (noise, air) and a possible collision because the car in front is decelerating for some reason.

When the ambulance crosses the junction, RSI broadcasts the message that it is going to turn the traffic light to green because there are no other vehicles on the joining road. The IRP agent informs Bob to smoothly accelerate and move forward. Finally, when Bob arrives at home, Alice is waiting for him with a hot cup of coffee, he starts talking with Alice while drinking his coffee.

2.1 Functional Requirements

Based on the application scenario described in Section 2, we define a set of functional requirements that should be provided by a service platform. Below we describe some of these functionalities and the information required to provide these functionalities.

Lance Changing. The lane changing is one of the complex tasks, which needs to be decided by the driver based on the current situation at road. There might be various reasons for decising on changing the lane while driving on the road. Drivers need to assess the situation, find the appropriate gap, ensure there is no bliend-spot situation and make the decision instantly. The reasons for changing the lane could either be mandatory or discretionary. It is mandatory to change the lane when the current lane is ending or the vehicle is facing the hazardious situation or the driver intends to take a turn at upcoming junction and therefore the vehicles needs to be on the different lane. It is discretionary to change the lane if the preceding vehicle is driving slower than normal speed. In such situations drivers can be assisted to change their lane safely based on route calculation (e.g., turn left at next junction) or current traffic situation (e.g., overtake to move faster). While assisting the drivers, bliend-spot situation must be assesed based on run-time information possibly collected through inbuilt sensors to prevent possible crash with a nearby vehicle.

Curve Speed. Driving at the curved can be difficult especially when the weather condition and visibility are poor. Informing the driver as far ahead as possible about the existence of a curve and the weather condition will allows drivers in negotiating curves at appropriate speeds. This will help in reducing the possible risk of sliding and unexpected crashes. This can be supported possibly by combining weather information, current speed of the vehicle, its GPS position and the digital map of the street.

Collision Avoidance. A vehicle on the road is vulnerable to collision not only because of the carelessness of its driver but also because of the actions of the drivers on other vehicles. It can also happen because of the occurrence of the unexpected and emergency events at the road. For example, icy road surface can cause the collision between cars. If emergency situation arise, a preceding car (not only the one at immidiate front but could also be the one at ten cars ahead) could suddenly break, leaving insufficient time to react safely. In such situation, drivers can be assisted based on the speed of the preceding cars which can possibly be collected by backward propagation of preceeding cars. Drivers can be informed to slow down if the preceding cars suddenly decelerate because of some unexpected reasons.

Pollution Control. Increasing number of vehicles on the road is one of the contributors to air pollution in urban areas. In fact, the driving/speeding pattern of a vehicle can be used to determine how much it contributes to the environmental pollution. The drivers can be assisted to maintain their speed at optimal level on the road thereby assisting in minimising their pollution contribution. Since the speed of a vehicle is dependent on the speed of other vehicles, weather condition, condition of the road and other situations at the road (e.g., maintenance, accident) such information could possibly be backward propagated for avoiding abrupt deceleration/acceleration of succeding vehicles.

Hazard Situation Avoidance. Drivers on the road often encounter hazardious situations. This could be because of bad weather condition, poort visibility, construction works at the road or sudden emergency etc. In such situation, drivers should be adviced in good time such that such situation can be avoided. Drivers can be advised, for example, to prepare to slow down because the road ahead is slippery or road ahead has poor visibility. In order to provide such advices, information could be gathered from different sources such as weather station, road infrastructure services or oncoming vehicles, etc.

Rerouting. When the road ahead is blocked because of some unexpected situation, it might be beneficial to advice drivers to take different route to their destination. The alternative route can be calculated based on information collected from vehicles coming from opposite direction, road infrastructure services or information from central server (providing traffic information of the local areas). Vehicles coming from the opposite direction can provide information based on their observation of the road, e.g., re-route because the road ahead is blocked due to recent accident, which may not be available through other sources.

The list of functionalities discussed above is not exclusive and these functionalities are also not orthogonal to each other. However, the list gives the impression about what type of functionalities are useful to maintain safety of the drivers, environmental sustainability as well as the management of the road traffic.

3 Challenges

In order to fulfill the requirements of traffic management and surveillance, the service platform should support integration of and interoperation between different services. The services are heterogenous in nature, i.e., they could be context dependent, short lived, localized and situation aware. Similarly, the interaction between these services could also be ad hoc and use heterogeneous communication protocols. In addition, these services has to deal with near real-time information. There are several technical challenges that need to be tackled to achieve integration and interoperation between these services in this domain.

Since the services are typically context dependent, exhibit dynamic behaviour and heterogeneous, the existing tools and technologies cannot be used as it is. We need a suitable service description model, which allows to describe not only the functional charecteristics of the services but also the context and situation at which these services can be used. This further requires a suitable composition model, which allows to compose such services based on (near) real-time information. The real-time information can be complemented with the knowledge extracted from the past behaviour of the road users. In order to support this, a suitable service intelligence model is required such that the extracted knowledge can be utilised in delivering the useful services to the users.

Drivers on the road are busy controlling the vehicle with their hands and legs occupied. This requires for a suitable iterface to the systesm such that the drivers can still concentrate on driving while being assisted. Similarly, the provisioning of services taking into account the interests of the users and the technical capabilities of the communication infrastructures and the sensors is another challenge which needs a serious attention. Moreover, providing consistency and the performance guarantee of such system is far from trivial.

4 Solutions

We describe a service platform to support communication between vehicles as well as between vehichles and road infrastructure through the concept of services orientation. The concept of service orientation is used to integrate various types of systems and services. It is also used for supporting interoperation between these services and systems. Figure 1, shows the high level interaction between these systems and services. Furthermore, the service orientation allows us to deploy services in the cloud to achieve performance requirements such as scalability and efficiency.

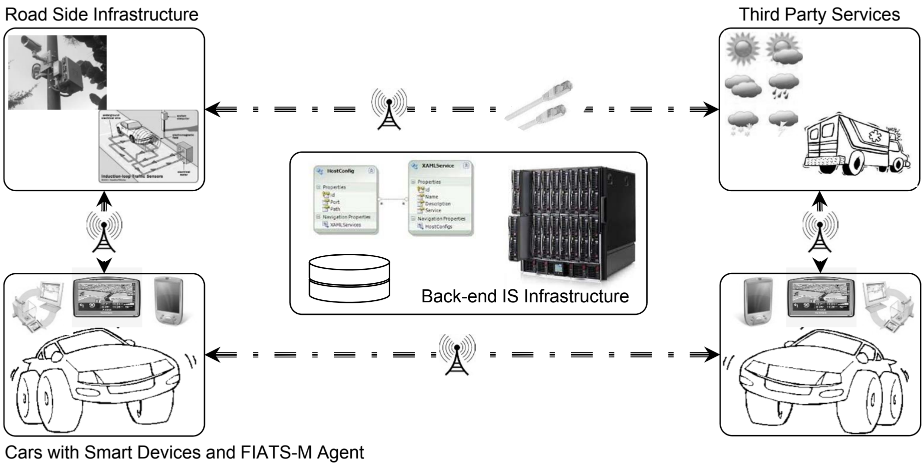

Fig. 1. Communication between Cars, Infrastructure and Services.

As seen in the figure, the service platform needs to provide supprot for different communication protocols as well as service descriptions. To supprot this requirement, the service platform provides a standard communication interface which bridges the protocol heterigeneity through the use of adapter. The heterogeneity between service descriptions can be handled by defining an intermediate description language which can allows us to define mappings without knowing the targed description language. The back-end IS infrastructure is used to process the collected information and to derive useful information or the composition of services for the user. The services can either be registered to the back-end IS infrastructure or be discovered on demand. The vehicles in the range, can communicate with themselves without requiring to go through the back-end IS infrastructure.

5 Related Works

There is significant ongoing work in the area of road traffic management and surveillance through smart technologies. An agent based approach is used in [12], aiming at providing a semantic middleware for context-aware smart road dnvironments. The context information is extracted using data mining techniquest over the collection of data obtained from various sensors.

A Peer-to-Peer based infrastructure supporting communication between vehicles and with the road-side infrastructure is defined in [10]. This work is focused on sharing context-aware road and safety information between different road users. Though their work supports such a communication, it does not spacify how up-to-date are the shared information.

In [13], an approach for supporting vehicle to vehicle communication is proposed. It employes message broadcasting as a means to support communication between vehicles. Using some intelligent communication pattern, a vehicle can notify about certain events to the nearby vehicles.

Using the color and edge information, a technique to detect traffic light is proposed in [14]. The proposed technique is defined based on image processing and statistical techniques. Detection of traffic light helps in replanning the route or suggesting drivers to stop safely at right time.

The existing systems focus on traffic management and surveillance taking into account the interests of the infrastructure providers and not the interests of the road users. Moreover, these systems do not fully utilise the collective intelligence and spatio-temporal correlations of the vehicles and their movement pattern.

6 Discussion

In the traffic-domain, a huge amount of information about road user is collected and processed to manage the road traffic in an efficient manner. The sensors, loop detectors and cameras are the commonly used tools and techniques for collecting these information. The information thus collected are sparse, dealing with these information to make certain decision is not as easy as it should be. Therefore, it is necessary to identify what information is needed and how these information can be gathered to provide usefuls services to the users.

The speed of a vehicle depends on the speed of the preceding vehicle. The preceding vehicle may be intentionally driving slowly. It is not clear how to deal with such a situation. If the preceding vehicle is the main source of congestion and hence the cause of other problems (e.g., air pollution), the vehicle behind it is left with no option that just contributing to the problem caused by the preceding vehicle. This problem may possibly require some business intelligence to panalize the intentionally troubling vehicles possibly through communication between vehicles and/or vehicle to infrastructure.

When such communication takes place, the information need to be shared between the communicating parties. It is therefore necessary to identify how privacy sensitive is the information being shared. If the information is privacy sensitive, it has to be dealt with accordingly to avoid any unauthorised manipulation of the information.

7 Conclusions

This paper discusses how a service platform can support communication between vehicles and between vehicles and road infrastructure services as well as third party services. The service platform is based on the concept of service oriented architecture and is aimed at supporting integration of and interoperation between different systems, services and information in road traffic domain. This further allows for providing various user-centric services which help at maximizing user safety, minimizing congestion and environmental pollution, and optimizing the use of road with maximum efficiency. In the current work, we identified different challenges which need to be technically tackled. In our future work, we aim at finding the solution to these challenges and extending the architecture accordingly.

Acknowledgements. This material is based upon works jointly supported by the IOP GenCom U-Care project (http://ucare.ewi.utwente.nl) sponsored by the Dutch Ministry of Economic Affairs under contract IGC0816 and by the DySCoTec project sponsored by the Centre for Telematics and Information Technology (CTIT), University of Twente, The Netherlands.

References

1. Figueiredo, L., Jesus, I., Machado, J. A. T., Ferreira, J. R., de Carvalho, J.L.M.: Towards the Development of Intelligent Transportation Systems. In: Proc. of the 2001 IEEE Intelligent Transportation Systems. (2001) 1206–1211
2. Mahmod, M. K. M.: Using Co-operative Vehicle-Infrastructure Systems to Reduce Traffic Emissions and Improve Air Quality at Signalized Urban Intersections. PhD thesis, Faculty of Engineering Technology, University of Twente (2011)
3. Anda, J., LeBrun, J., Ghosal, D., Chuah, C. N., Zhang, M.: VGrid: Vehicular AdHoc Networking and Computing Grid for Intelligent Traffic Control. In: Proc. of the Vehicular Technology Conference. (2005) 2905–2909
4. Dikaiakos, M. D., Florides, A., Nadeem, T., Iftode, L.: Location-aware Services over Vehicular Ad-Hoc Networks using Car-to-Car Communication. IEEE Journal on Selected Areas in Communications 25 (2007) 1590–1602
5. Hawas, Y. E., Napeñas, M. J. B., Hamdouch, Y.: . Journal of Intelligent Transportation Systems 13 (2009) 199–217
6. Zalila-Wenkstern, R., Steel, T., Leask, G.: A Self-organizing Architecture for Traffic Management. In Weyns, D., Malek, S., de Lemos, R., Andersson, J., eds.: Self-Organizing Architectures. Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2010) 230–250
7. Gascueña, J. M., Fernández-Caballero, A.: On the use of Agent Technology in Intelligent, multisensory and Distributed Surveillance. The Knowledge Engineering Review 26 (2011) 191–208

8. Paridel, K., Yasar, A. U. H., Vanrompay, Y., Preuveneers, D., Berbers, Y.: Teamwork on the Road: Efficient Collaboration in VANETs with Context-based Grouping. In: Proc. of the 2nd International Conference on Ambient Systems, Networks and Technologies. (2011) 4857
9. Erl, T.: Service-Oriented Architecture Concepts, Technology, and Design. Prentice Hall Professional Technical Reference (2005)
10. Santa, J., Gómez-Skarmeta, A. F.: Sharing Context-Aware Road and Safety Information. IEEE Pervasive Computing 8 (2009) 58–65
11. Schilit, B. N., Adams, N., Want, R.: Context-aware Computing Applications. In: Proc. of the Workshop on Mobile Computing Systems and Applications. (1994) 85–90
12. Terziyan, V., Kaykova, O., Zhovtobryukh, D.: UbiRoad: Semantic Middleware for Context-Aware Smart Road Environments. In: Proc. of the Fifth International Conference on Internet and Web Applications and Services. (2010) 295–302
13. Biswas, S., Tatchikou, R., Dion, F.: Vehicle-to-Vehicle Wireless Communication Protocols for Enhancing Highway Traffic Safety. IEEE Communications Magazine 44 (2006) 74–82
14. Omachi, M., Omachi, S.: Traffic Light Detection with Color and Edge Information. In: Proc. of the 2nd IEEE International Conference on Computer Science and Information Technology. (2009) 284–287

Assessing the Viability of Service Innovations: A Structured Business Modeling Approach[1]

Björn Kijl, Lambert J. M. Nieuwenhuis, L. O. Meertens and M. E. Iacob

University of Twente, School of Management & Governance,
PO Box 217, 7500 AE Enschede, The Netherlands
{b.kijl, l.j.m.nieuwenhuis, l.o.meertens, m.e.iacob}@utwente.nl
http://www.smg.utwente.nl

Abstract. Currently, business modeling seems to be an art instead of a science, as no scientific method for business modeling exists. This causes many service innovation projects to end after the pilot stage, unable to fulfill their apparent promises. We propose a structured method to create "as-is" business models in a structured manner. The method consists of the following steps: identify the involved value network roles, recognize relations among these roles, specify their main activities and develop a quantitative model using realistic estimates. The resulting quantitative business model is suitable for analysis of the current situation. This is the basis for further predictions, like business cases, scenarios and alternative business model designs. We offer two extra steps to develop and analyse these alternatives. Using our method may increase the viability of service innovation projects by helping to improve the underlying service innovation business model design.

1 Introduction

The use of information and communication technology (ICT) is increasingly proliferating in transportation. It is applied to support main functions like car management and navigation. Other applications include navigation systems that guide drivers to their destination taking traffic information into account and ICT systems that support the entertainment of travelers with games, music, video and connectivity. Next to this, traffic management systems are emerging that capitalize on the possibilities of ad hoc car networks (e.g. by using sensor technology) in order to optimize traffic management by for example adapting flows through changing the timing of traffic lights. Also sensor and actuator networks may be used for increasing safety and lowering the prevalence of car accidents.

[1] This chapter is primarily based on L.O. Meertens, M.E. Iacob and L.J.M. Nieuwenhuis, Developing the business modeling method, Proceedings of the First International Symposium on Business Modeling and Software Design 2011 (BMSD 2011), pp. 88-95 and Kijl, B., Nieuwenhuis, L.J.M., Deploying e-health service innovations – an early stage business model engineering and regulatory validation approach, International Journal of Healthcare Technology and Management, Vol. 12, No. 1, 2011.

The functionalities as mentioned above – car sensors, engine management, traffic information systems, entertainment and telecommunications – have mostly been developed independently of each other. It is expected that in the near future, these technological systems will become integrated and less car / brand independent. In such a context, these technologies may provide an architectural platform for many new service innovations.

At the service platform domain we have to define and develop an environment where interacting services can coexist. These new services are typically context aware. Next to more technological challenges like service intelligence and service management, assessing the commercial viability of these service innovations will also be a critical aspect.

2 Early Stage Service Viability Assessment

For assessing the viability of future traffic management systems and related service innovations in early project development stages, we propose to analyse the underlying business models. A business model essentially describes how value can be created (for users) and captured (e.g. in the form of profits) with a specific product or service.

In order to perform such an early stage analysis, we first need to identify all stakeholders involved. One can think of individuals participating in the traffic as well as organizations responsible for infrastructural services. However, a wide range of other service providing organizations are also expected to play an important role, like governments (probably interested in road pricing strategies and traffic management issues), the energy sector (for fueling or taking care of charging electronic car batteries), the health care sector (monitoring the wellness of drivers and their passengers), telecom operators (connectivity), entertainment and media companies (video, music, …) and of course also car manufacturers (by building excellent cars with competitive services they may create the basis for open service platforms).

Typical business related research questions relevant for organizations involved in offering new service innovations are:

- How will the value network that is needed for creating these new service innovations exactly look like – which organizations need to work together and in what way?
- Which revenue models can be used (relevant for private as well as private parties)?
- How can we optimally balance costs and revenues among the organizations participating in the value network?

In earlier research projects (see http://www.myotel.eu and http://www.utwente.nl/ewi/ucare/), we developed a structured business model validation approach aimed at answering these questions and tested it on specific e-health service innovations.

After introducing the concept of business models in Section 3 and 4, we will describe our business modeling approach by describing the six steps it consists of in Section 5. In Section 6, we will apply this approach to an illustrative case study.

3 Business Modeling Background

A business model is critical for any company. Its importance has been recognized over the past few years by several authors that have created different business model frameworks aimed at identifying the main elements of a business model (for example, Osterwalder [9]; for an overview, see Pateli and Gialis [10], Al-Debei and Avison [1] and Vermolen [14]). However, the state in which this field finds itself is one of pre-scientific chaos [7]: there are several competing schools of thought and progress is limited because of a lack of cumulative progress. Because of this, there are no clear and unique semantics related to business models. The concept of a business model is associated with many different definitions [14]. The elements of such a business model differ significantly from one approach to another. Furthermore, to the best of our knowledge, there are no methodological approaches in the literature for the design and specification of business models [14]. This lack of cohesion in the field clearly diminishes the added value of business models for companies and makes business modeling an art, rather than a science. This state of affairs motivated us to propose such a method, which enables the development of business models in a structured manner. Thus the contribution of this chapter is three-fold:

- A proposal for a business model development method;
- A definition of the concept of business models and the identification of its core elements, captured by the deliverables of the method steps;
- An illustration of the method by means of an illustrative case study from the healthcare domain.

4 Theoretical Background

A simple analysis of the two words “business model” already gives an idea of what a business model is about. On the one hand, there is “business”: the way a company does business or creates value. On the other hand, there is “model”: a conceptualization of something – in this case, of how a company does business.

We extend this common and simplistic interpretation of a business model as “the way a company earns money” into a broader and more general definition of the concept: “a simplified representation that accounts for the known and inferred properties of the business or industry as a whole, which may be used to study its characteristics further, for example, to support calculations, predictions and business transformations.”

The last part of the definition above, namely the indication of the possible uses of a business model is of particular importance in the context of this chapter. The method we propose not only facilitates the development of such a design artefact – a business model – but also takes a business engineering perspective. Thus, its application will result in essentially two (or more) business models: one that reflects the “as-is” situation of the business and one or more alternative “to-be” business models that represents possible modifications of the business as result of, for example, adoption of innovative technologies or more efficient business processes.

To the best of our knowledge, such a method does not exist yet (Vermolen 2010). In the remainder of this section, we position our work in the contexts of design science and method engineering, to which it is related.

4.1 Design Science

A business modeling method can be seen as a design-science artefact. It is the process of creating a product: a business model. We use the nine guidelines of Hevner et al. [4] to frame how we use the methodology engineering approach from Kumar & Welke [8] to create our method.

The first guideline advises to design as an artefact. Design-science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation. As said, we produce a method.

The second guideline tackles relevance. The objective of design-science research is to develop technology-based solutions to important and relevant business problems. Viable business models lie at the heart of business problems. However, our solution is not yet technology-based. Partial automation of the method is left for future research.

The utility, quality, and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods. We demonstrate the business modeling method using a case study. We leave more rigorous evaluation for further research.

Research contribution is the topic of the fourth guideline. Effective design-science research must provide clear and verifiable contributions in the areas of the design artefact, design foundations and/or design methodologies. We provide a new artefact to use and study for the academic world. The methodology may be extended, improved and specialized.

Guideline five expresses the scientific rigour: Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artefact. We aim to be rigorous through using the methodology engineering approach. Existing, proven methods are used as foundation and methods where applicable. Evaluation was handled in the third guideline.

The sixth guideline positions design as a search process. The search for an effective artefact requires utilizing available means to reach desired ends while satisfying laws in the problem environment. Whenever possible, we use available methods for each of the steps. Following the methodology engineering approach helps us to satisfy the laws for creating a new methodology.

The final guideline instructs us to communicate our research. Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences. This chapter is one of the outlets where we present our research.

4.2 Methodology Engineering

Methodologies serve as a guarantor to achieve a specific outcome. In our case, this outcome is a consistent and better-informed business model. We aim to understand (and improve) how business models are created. With this understanding, one can explain the way business models help solve problems. We provide a baseline methodology only, with a limited amount of concepts. Later, we can extend, improve and tailor the methodology to specific situations or specific business model frameworks.

The business modeling method has both aspects from the methodology engineering viewpoint: representational and procedural [8]. The representational aspect explains what artefacts a business modeler looks at. The artefacts are the input and deliverables of the steps in the method. The procedural aspect shows how these are created and used. This includes the activities in each step, tools or techniques and the sequence of steps.

5 Defining the Business Modeling Method

We defined six individual business modeling steps. We will concisely describe each of these steps in this section, by using the following elements:

- Inputs of the step;
- Activities to perform during the step;
- Possible techniques to use for these activities, and
- Deliverables resulting from the step.

Each step in the proposed method requires specific methods, techniques or tools that are suitable for realizing the deliverables. We will mention examples of those. However, others may also be useful and applicable, and it is not our aim to be exhaustive in this respect. **Table 1.** shows an overview of our method.

Table 1. Business modeling method.

Step	***Inputs***	***Technique or Tools***	***Deliverables***
Roles	Documentation, domain literature, interviews, experience, previous research	Pouloudi and Whitley [12]	Role list
Relations	Role list, stakeholder map, value exchanges	e3-value	Role-relation matrix
Activities	Role-relation matrix, role list, business process specifications	BPM methods, languages and tools	List of activities
Quantification	Process specifications, accounting systems and annual reports	Activity based costing	Total cost of the business "as-is"
Alternatives	As-is business model, ideas for innovations and changes	Business model method (step 1-4), brainstorming	One or more alternative business models
Analysis	Alternative business models	Sensitivity analysis, technology assessment, interpolation, best/worst case scenarios	Business case for each alternative

5.1 Create As-Is Model

As mentioned in the previous section, our business model development method takes a business engineering perspective. Thus, the first four steps of our method focus on creating a business model that reflects the current state of the business. Therefore, steps one through four create a so-called as-is model.

Step 1: Identify Roles

Identifying the relevant parties or roles involved in a business model should be done as systematically as possible. The business modeler must carry out a stakeholder analysis to identify all roles. The inputs to this step include for example documentation, domain literature, interviews, experience and previous research. The output is a list of roles.

For an example of a stakeholder analysis method, we refer to Pouloudi & Whitley [12]. They provide an interpretive research method for stakeholder analysis aimed at inter-organizational systems. The method consists of the following steps:

1. Identify obvious groups of stakeholders;
2. Contact representatives from these groups;
3. (In-depth) interview them;
4. Revise stakeholder map;
5. Repeat steps two to four, until...

Pouloudi and Whitley do not list the fifth step, but mention that stakeholder analysis is a cumulative and iterative approach. This may cause the number of stakeholders to grow exponentially, and the question remains when to stop. Lack of resources may be a reason to stop the iterative process at some point. Closure would be good, but seems hard to achieve when the model is more complex. Probably, the modeler has to make an arbitrary decision. Nevertheless, one should choose stop criteria (a quantifiable measure of the stakeholder's relevance for the respective business model and a threshold for the measure) before starting the process [11].

For step four, "Revising the stakeholder map", the information on stakeholders as gathered from interviews can be complemented with information found in the literature. The business modeler then refines the list of stakeholders by focusing, aggregating and categorizing.

Step 2: Recognize Relations

The second step of our method aims to discover the relations between roles. The nature of these relations may vary substantially, but it always involves some interaction between two roles and some exchange of value. Much of the work and results from the previous step can be reused as input for this step. In theory, all roles could have relations with all other roles. However, in practice, most roles only have relations with a limited number of other roles. Usually, these relations are captured in a stakeholder map, which often follows a hub-and-spoke pattern, as the focus is on one of the roles. This pattern may be inherent to the approach used, for example if the scope is defined as a maximum distance from the focal or nodal role.

To specify all relations, we suggest the use of a role-relation matrix with all roles on both axes as technique. Within this matrix, the cells point out all possible relations among the roles. Each of the cells could hold one or more relations between two roles. Assuming that roles have a limited number of relations, the role-relation matrix will be partially empty. However, one can question for each empty cell whether a relation is missing or not.

Cells above and below the diagonal can represent the directional character of relations. Usually, relations have a providing and consuming part (above and below the diagonal) but sometimes constructions that are more complex occur, such as loops including multiple roles.

The output of this step is a set of relations.

Step 3: Specify Activities

For a first qualitative specification of the business model, the next step is to determine the main activities. Relations alone are not sufficient: the qualitative model consists of these main business activities (business processes) too. These activities originate from the relations identified in the previous step. Each of the relations in the role-relation matrix consists of at least one interaction between two roles, requiring activities by both roles. Besides work and results from the previous steps, existing process descriptions can be valuable input. Techniques from business process management may be used.

The output from these first three steps is a first qualitative business model, including roles, relations and activities. It reveals what must happen for the business to function properly.

Step 4: Quantify Model

Quantifying the business model helps us to see what is happening in more detail and compare innovations to the current situation. To turn the qualitative model into a quantitative model, numbers are needed. The numbers are costs and volumes of activities (how often they occur). Together, these numbers form a complete view of the costs captured by the business model.

Several sources for costs and volumes are possible, such as accessing accounting systems or (annual) reports.

5.2 Develop To-Be Model

The as-is model, created in previous steps, is suitable for analysis of the current state only. However, from the as-is model, it is possible to derive alternatives. Such alternatives can be created to assess how reorganisations, innovations or other changes influence the business. These are the to-be models.

Step 5: Design Alternatives

From here on, we aim to capture a future state of the business in a business model. To make predictions, the model may need further instantiations. Each instantiation is an alternative development that may happen (to-be). Using techniques such as brainstorming and generating scenarios, business modelers create alternative, qualitative, future business models. These alternatives are used to make predictions. Usually, such alternatives are built around a(n) (technical) innovation. This may include allocating specific roles to various stakeholders. A base alternative, which only continues an existing trend without interventions, may help comparing the innovations. Next to the business model, ideas for innovations serve as input. The resulting alternative business models show future (to-be) possibilities.

Step 6: Analyse Alternatives
The final step for a business modeler is to analyse the alternative business models. Besides the qualitative business models, several sources of input are possible to quantify the alternatives. Applicable techniques include sensitivity analysis, technology assessment, interpolation and using best/worst case scenarios. Each alternative can be tested against several scenarios, in which factors change that are not controllable. We can use the models to predict the impact. This step and the previous one can be repeated several times to achieve the best results. The final output is a business case (including expected loss or profit) for each alternative.
In the next section, we will describe by way of illustration the main results of the application of our business model engineering approach to an e-health service innovation project called Myotel.

6 The Myotel Case

The business modeling case consists of a so-called myofeedback based teletreatment service, aimed at patients with 1) chronic neck shoulder pain or 2) whiplashes – directly in the R&D deployment phase of the service innovation project.

The myofeedback teletreatment system monitors muscle relaxation during daily activities via sensors and actuators implemented in a wearable garment which is connected to a PDA. The system provides continuous feedback when there is too little muscle relaxation. The monitoring data is sent wirelessly – e.g. via a GPRS, UMTS or HSDPA connection – to a back end system which can be accessed by health care professionals. These health care professionals can use the system for optimizing treatment, working more efficiently by saving on face-to-face contact hours with their patients and giving them more personalized feedback as well (see Fig. 1 for a high level architectural overview of the system).

Earlier clinical evaluation studies already proved the clinical effectiveness of the specific treatment [3]. Clinical trials proved a work ability increase during treatment for both the whiplash patients (of whom 68% were employed) and chronic neck shoulder pain patients (of whom 100% were employed) that got treated with the teletreatment service. The work ability concept shows how well workers are able to perform their work [5, 6] – it can be seen as an indicator for employee productivity increase or decrease.

Below, a concise description is given of the results of each of the business modeling steps as described above with a special focus on the “to-be” business models.

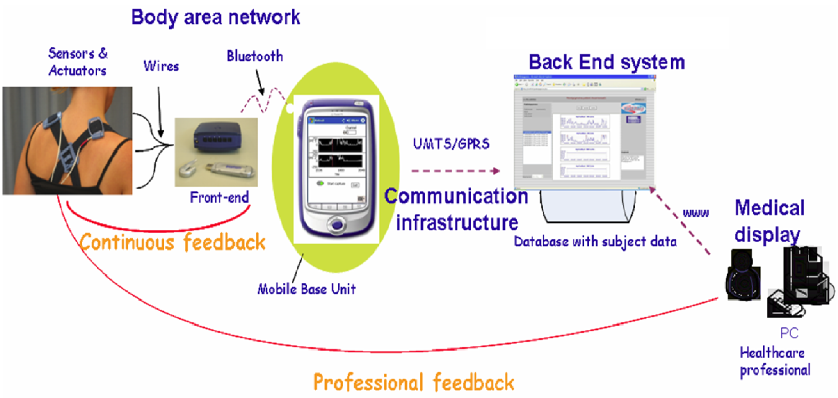

Fig. 1. The teletreatment service.

Step 1: Identify Roles

We organized a half-day business model design workshop for twelve experts within the field of myofeedback and teletreatments from four European countries in which the service could be offered – The Netherlands, Belgium, Sweden and Germany.

Based on the expert workshop, different value network roles and potential actors were identified. A value network role is performed by a specific value network actor who may perform the actual activities in the value network (cf. a specific actor playing a certain role in a movie). All value network roles and possible value network actors as identified can be found in Table 2.

Table 2. Value network roles, actors and activities identified by the experts.

Roles	***Actor***
End user / patient	Therapist patient
Network provider	Telecom operator
Hardware provider (e.g. for communication devices and sensors)	Hardware company
Telerehabilitation (myofeedback) service provider	(Spin-out) company / independent organization
Health care professional	Therapist organization
Software developer	Company
Software platform provider	Company
Insurance company	Company
Employer	(Non) commercial organization
Medical research & development organization	A (group of) medical institution(s) that support(s) the commercial exploitation of the myofeedback service

Step 2: Recognize Relations

By developing a matrix with all value network roles on both the x and y axis, we identified all possible relationships between the roles. This analysis formed the basis for Step 3, defining the activities of the value network roles.

Step 3: Specify Activities

Based on the previous two steps, the experts identified the activities of each value network role that had to be modeled in the quantitative model. Here we especially focused on the most critical, so-called first tier roles as identified by the experts: the service provider, the healthcare provider, the employer and the patient (because of resources limitations and time constraints, we decided not to model the other, less critical roles). According to the experts, these roles would be affected most by the introduction of the new teletreatment service. For these roles, the experts also had to identify the activities that are expected to lead to the most substantial cost and benefit changes. These activities are italicized in Table 3 and modelled in the quantitative model in the next step.

Table 3. Main value network roles, actors and activities identified by the experts.

Role	***Activities***
Myofeedback service provider	*Manage telerehabilitation service (overhead)* *Develop telerehabilitation market (marketing)* *Acquire telerehabilitation customers* *Build back office* *Manage back office* *Build device service* *Manage devices needed for treatment* *Train myofeedback service delivery personnel* Deliver myofeedback service Request reimbursement myofeedback treatment Receive payment for myofeedback service
Healthcare provider	Develop telerehabilitation treatment *Train personnel telerehabilitation treatment* *Diagnose patient* *Consult patient with traditional treatment* *Consult patient with telerehabilitation treatment* Request reimbursement treatment Receive payment for treatment
Employer	*Employ traditionally treated employee* *Employ telerehabilitation treated employee*
Patient	*Undergo traditional treatment* *Undergo telerehabilitation treatment*

Step 4: Quantify Model

Based on the previous steps, we started to develop a quantitative model. We used a traditional activity based costing approach to determine costs of the activities. For each activity we determined the number of times N the activity is carried out and the cost price P per activity. With respect to cost price, we distinguished between investments and yearly costs. Investments are onetime costs e.g. for training and education (needed when new employees get involved) as well as investments for equipment (needed when more devices are needed or old devices are worn out). Examples of yearly costs are costs for personnel and housing. The volume and cost tables are part of a spread sheet that simulates the provisioning of the telerehabilitation service in year *i* from 2008 to 2018. Multiplying N_i and P_i gives the overall costs for year *i* for each activity. The values for N_i are based on an S shaped technology adoption curve [13]. The values for P_i are based on today's market prices that develop over time, i.e., technology prices decline (deflating prices), whereas, e.g., salary costs for professionals increase (inflating prices). This enables Net Present Value calculations over the ten years period as well.

For calculating the costs and benefits of the myofeedback teletreatment service, a so-called variables cockpit was developed. This cockpit gives an overview of the most important variables that influence the costs and volumes as mentioned before and based on which the actual benefit and cost calculations can be made. Important variables in this context are e.g. the hourly cost price of a health care professional, the expected productivity increase resulting from treatment, the number of therapy appointments per treatment and the myofeedback device costs. All figures for the cockpit variables were determined based on results from literature research, medical research results, surveys filled in by and related workshops held with the experts as mentioned before.

Based on these variables, the volumes and costs for each of the activities identified were calculated over the period 2008 – 2018. Because the model was designed in the form of a spread sheet, the model is very flexible: the effects of changing one or more variables can be calculated directly. Based on a different set of cockpit variables for different countries, the model can automatically estimate the expected volumes and costs on a country-by-country level over the period 2008 – 2018. In this way, the model formed a useful tool for evaluating the costs and benefits related to a new service innovation like the telerehabilitation service in direct interaction with field experts.

Step 5: Design Alternatives

Based on the analysis, two alternative business models were developed – one in which the insurance company of the end user or patient would form the main source of revenue (see Fig. 2) and one in which the employer (or its occupational healthcare / disability insurance organization) formed the main source of revenue (see Fig. 3).

Step 6: Analyse Alternatives

The quantitative model as described in Step 4 revealed three critical insights (The related figures as mentioned below show the calculations for the Dutch market. For the other three countries, similar conclusions can be made):

1. The new myofeedback treatment is more expensive compared with traditional treatment – mostly because of IT related investments and operational costs (see Fig. 4).
2. Although the myofeedback treatment is expected to be more efficient compared with the traditional treatment method (less treatment hours are needed), the IT investments are likely to exceed the related labour cost savings on a health care professional level (see Fig. 5).

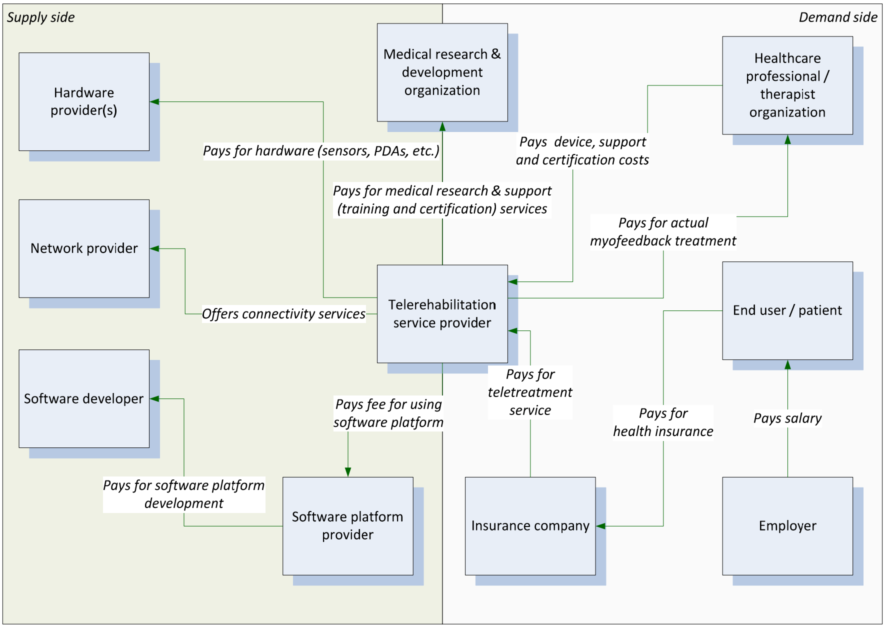

Fig. 2. The first business model design alternative – the end user / patient as revenue source.

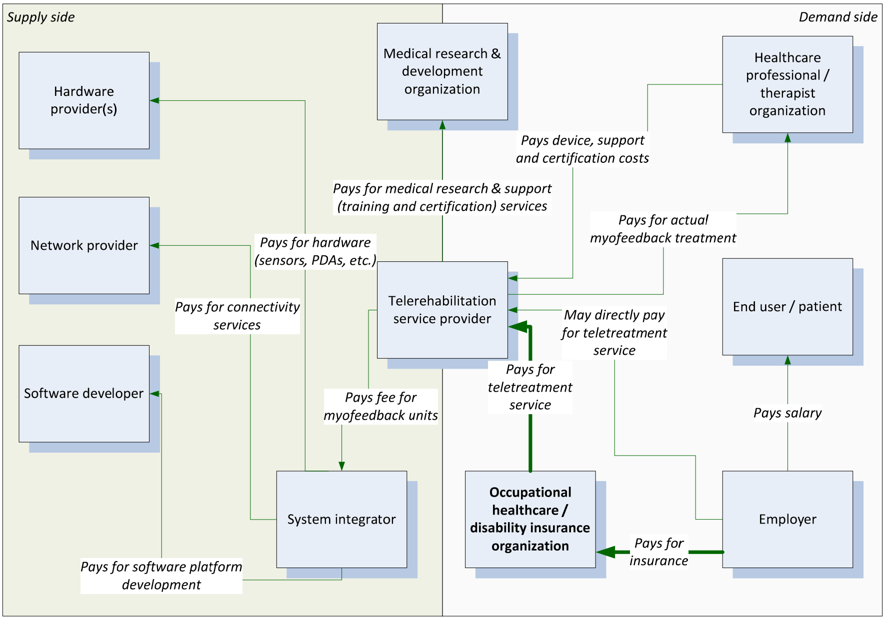

Fig. 3. The second business model design alternative – the employer of the end user / patient as revenue source.

3. However, the expected absence reduction and productivity increase of working myofeedback patients does compensate the investments needed on an employer level (see Fig. 6).

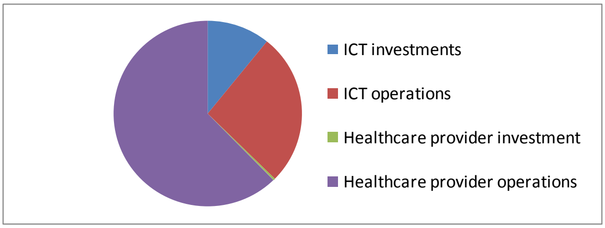

Fig. 4. The cost benefit model showed that the new myofeedback treatment is more expensive than traditional treatment – mainly because of ICT related investments and operational costs; the total costs are expected to increase with about € 100 (from ~€400 to ~€ 500) per patient with chronic neck shoulder problems.

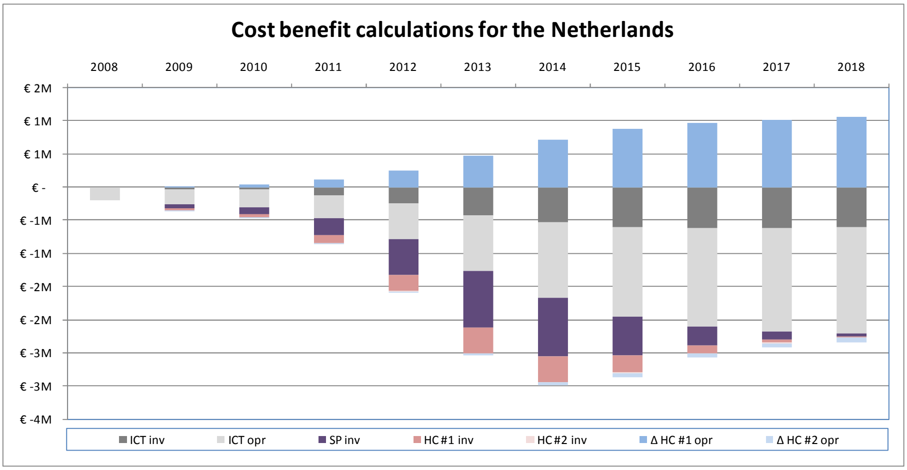

Fig. 5. A graphical overview of the cost benefit analysis in the Netherlands; showing respectively expected ICT investments and ICT operational costs, service provider and health care professional investments and the expected operational benefits/costs for health care professionals.

Based on these three critical insights, we expected the second business model design (where the employer of the end user / patient formed the main source of revenue source; see also Fig. 3) to have a higher viability than the first business model design (where the insurance organization of the patient formed the main source of revenue; see Fig. 2).

The service innovation is expected to deliver a more efficient, less labor-intensive treatment, but these savings do not outweigh the related technological investments and costs (see Fig. 5) – as a result, we don't expect healthcare insurance organizations to be very interested in paying for a new, effective but also more expensive treatment. However, our analysis showed that the service may lead to an absence reduction and productivity increase that is substantially outweighing the related technology costs

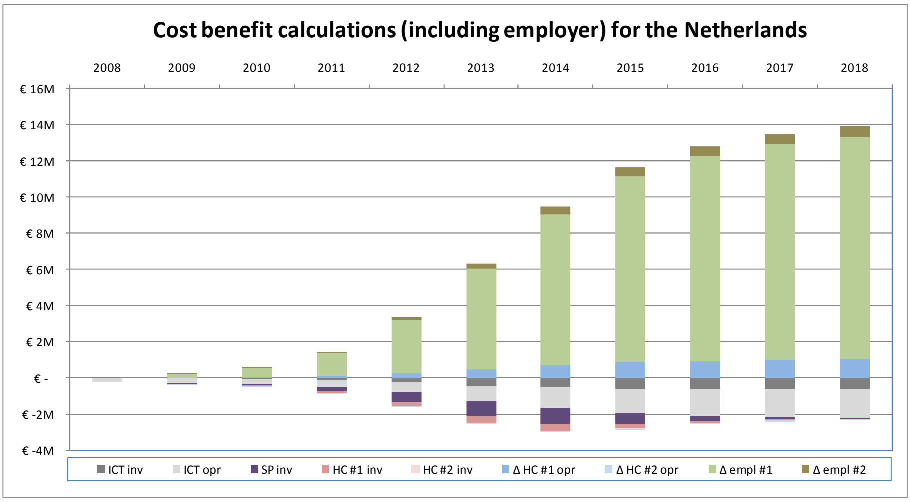

Fig. 6. A graphical overview of the cost benefit analysis in the Netherlands. Compared with Fig. 5, now also the maximum employer benefits related to absence reduction and productivity improvement are shown.

and investments (see Fig. 6). Therefore, employers may be more interested in paying for this new service innovation than healthcare insurance organizations.

7 Conclusions

Three contributions are made in this paper. First, we created a business model development method. Second, we defined the concept of a business model and identified its core elements, captured by the deliverables of the method steps. Finally, we demonstrated the method by means of an illustrative case study from the healthcare domain.

The business modeling method provides a way to create business models. Innovators can apply the steps to systematically create business cases for their ideas. This may help them to show the viability and get service innovations implemented. With our business modeling method, we provided a new design science artefact to use and study for the academic world. As business modeling has several goals, conducting only the first few steps of this method may be sufficient in some cases. For example, if one's goal is to achieve insight in the current state only, the last two steps are not useful.

The business modeling method could be extended further. Situational method engineering seems suitable for this [2]. For example, for information system development, it is interesting to analyse the possible integration of and linkages between enterprise architectures and business model designs. Also, a domain analysis step could be added before the role identification step (each domain requires different improvements). The steps in the method could be further specified and detailed as well. One way of doing this is to tailor the techniques discussed for each of the method steps. In the future, new tools and techniques may also support partial automation of the steps.

In general, we expect that by using our method for early stage business modeling the quality of the service innovation business models may be substantially improved. By early stage interaction with the more technology oriented activities within a service innovation project, we think our approach can play a critical role in maximizing the added value of the project outcome.

References

1. Al-Debei, M. M., Avison, D.: Developing a unified framework of the business model concept. European Journal of Information Systems 19(3), 359-376 (2010). doi:10.1057/ejis.2010.21
2. Henderson-Sellers, B., Ralyté, J.: Situational Method Engineering: State-of-the-Art Review. Journal of Universal Computer Science 16(3), 424-478 (2010)
3. Hermens, H. J., Vollenbroek-Hutten, M. M. R.: Towards remote monitoring and remotely supervised training. J Electromyogr Kines 18(6), 908-919 (2008). doi:DOI 10.1016/j.jelekin.2008.10.004
4. Hevner, A. R., March, S. T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quart 28(1), 75-105 (2004)
5. Ilmarinen, J.: The Work Ability Index (WAI). Occup Med (Lond) 57(2), 160- (2007). doi:10.1093/occmed/kqm008
6. Ilmarinen, J., Tuomi, K., Klockars, M.: Changes in the work ability of active employees over an 11-year period. Scandinavian journal of work, environment & health 23, 49 (1997)
7. Kuhn, T.S.: The Structure of Scientific Revolutions. University of Chicago Press (1970)
8. Kumar, K., Welke, R.J.: Methodology Engineering: a proposal for situation-specific methodology construction. In: Challenges and Strategies for Research in Systems Development. pp. 257–269. Wiley, Chichester (1992)
9. Osterwalder, A.: The Business Model Ontology – a proposition in a design science approach. PhD Thesis, Universite de Lausanne (2004)
10. Pateli, A. G., Giaglis, G. M.: A research framework for analysing eBusiness models. European Journal of Information Systems 13(4), 302-314 (2004). doi:10.1057/palgrave.ejis.3000513
11. Pouloudi, A.: Stakeholder Analysis for Interorganisational Information Systems in Healthcare. PhD thesis, London School of Economics and Political Science (1998)
12. Pouloudi, A., Whitley, E. A.: Stakeholder identification in inter-organizational systems: gaining insights for drug use management systems. European Journal of Information Systems 6, 1–14 (1997)
13. Rogers, E.: Diffusion of innovations. Free press (1995)
14. Vermolen, R.: Reflecting on IS Business Model Research: Current Gaps and Future Directions. In: Proceedings of the 13th Twente Student Conference on IT. University of Twente, Enschede, Netherlands (2010)

Intelligent Decision Support using Pattern Matching

Victoria J. Hodge, Tom Jackson and Jim Austin

Department of Computer Science, University of York, U.K.
{vicky, tom.jackson, austin}@cs.york.ac.uk

Abstract. The aim of our work is to develop Intelligent Decision Support (IDS) tools and techniques to convert traffic data into intelligence to assist network managers, operators and to aid the travelling public. The IDS system detects traffic problems, identifies the likely cause and recommends suitable interventions which are most likely to mitigate congestion of that traffic problem. In this paper, we propose to extend the existing tools to include dynamic hierarchical and distributed processing; algorithm optimisation using natural computation techniques; and, using a meta-learner to short-circuit the optimisation by learning the best settings for specific data set characteristics and using these settings to initialise the GA.

1 Introduction

Intelligent Decision Support (IDS) systems are an important computerised tool in many problem domains. The aim of our work is to provide an IDS tool to assist traffic network operators to optimally manage traffic. Within the FREEFLOW transport project (http://www.freeflowuk.net/), we have developed a pattern matching tool for Intelligent Decision Support [1, 2, 3, 4, 5]. The tool is able to:

- detect traffic problems,
- identify likely causes,
- recommend suitable control interventions,
- predict future traffic flows.

The pattern matching tool implements the k-nearest neighbour data mining algorithm [6, 7]. It uses the Advanced Uncertain Reasoning Architecture (AURA) for pattern matching [8, 9] to find the nearest neighbours. AURA is based on binary associative neural networks and can store large amounts of data and allows fast searches [8, 9, 10]. The AURA pattern matching tool employs a unified framework for both intervention recommendation and traffic state prediction. The only variation lies in how the matches are post-processed. It can fuse data from various traffic sources such as sensors embedded in the road, journey times, queue lengths, weather, time of day, etc. It can identify the cause of traffic problems by matching and cross-referencing historical data and use the results to make recommendations to traffic network managers [1, 3, 4, 5]. The AURA pattern matching tool can also predict traffic variable values to plug gaps in the data [2]. It can predict values to overcome a sensor failure or to look ahead and anticipate congestion problems, for example. We call the AURA pattern matching tool that implements the k-nearest neighbour algorithm *AURA k-NN*. A brief description of how we implement AURA k-NN for both intervention recommendation and traffic state prediction is given next. For a more detailed description see [1, 2, 3, 4, 5, 8, 9].

2 AURA Pattern Matcher

The foundation of AURA is the Correlation Matrix Memory (CMM). A CMM is a binary matrix used to store and retrieve binary vectors. All matrix elements are initialised to zero as in equation (1).

$$CMM_0 = 0 \tag{1}$$

AURA k-nearest neighbour (AURA k-NN) is an implementation of the k-nearest neighbour (k-NN) [6, 7] classification or prediction method [1, 2, 3, 4, 5, 9, 11] using AURA. K-NN is a widely used data-mining algorithm that uses similar procedures for clustering [12], outlier detection [13, 14], classification and prediction by examining the distances to the nearest neighbours. In the AURA implementation of k-NN, each column of the CMM is a tuple or record (set of traffic variable observations) and each row indexes a variable value. For traffic data, each column represents a date/time record of traffic variable observations. The variable observations can encompass sensor readings, weather data, journey times, queue lengths etc. In one column, the set of all rows represents all values for all variables for that date/time record.

Training. In AURA in general [15] and in the AURA k-NN implementation, the CMM is trained by learning associations between a binary input vector I_n which indexes matrix rows and a binary output vector O_n which indexes matrix columns as shown by equation (2) and Fig. 1.

$$CMM_k = CMM_{k-1} \oplus I_k \bullet O_k^T \tag{2}$$

The data are quantised and binned prior to training allowing numeric data to map onto the binary input vectors required for CMM training. Each row of the CMM represents a bin (range of variable values). Thus, an input vector representing the quantised variable values for a record is associated with an output vector uniquely identifying that record for all N records in the data set.

Recall. Recall involves identifying the best matching patterns stored in the CMM (the best matching columns). The recall procedure varies across AURA implementations [15]. In AURA k-NN recall, the query vector Q is a weighted (positive integer) vector that activates rows in the CMM with integer-valued scores. To emulate Euclidean distance, AURA k-NN applies parabolic kernels to the query vector [1, 2, 3, 4, 5, 8]. There is one kernel per variable and each kernel is centred on the value of that variable in the query. The kernel scores represent distance. The score is at a maximum at the query variable value and decrease with bin distance from the query value. To retrieve the k nearest neighbours, AURA effectively calculates the dot product of the integer-valued query vector Q and the CMM as in equation (3).

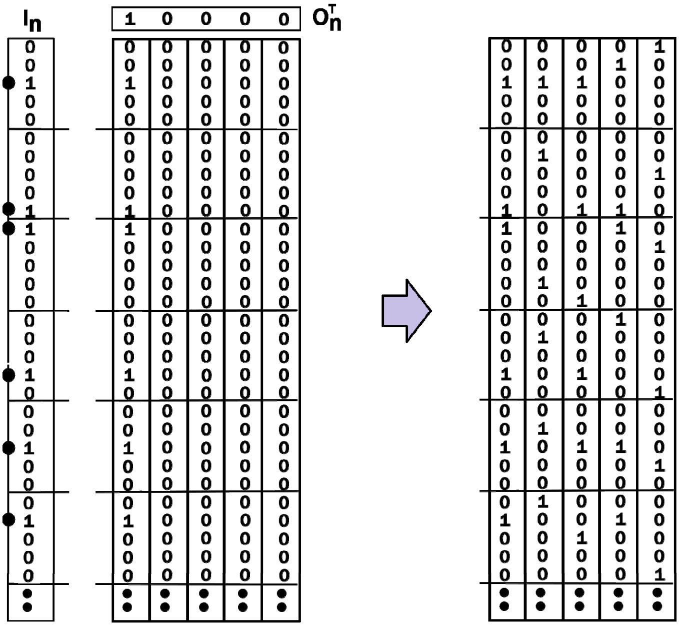

Fig. 1. Showing the CMM storing the first association $I_0 \times O_0$ on the left and a CMM trained with five associations on the right.

$$S^T = \sum CMM \bullet Q \tag{3}$$

The columns of the matrix are summed according to the value on the rows indexed by the integer-valued query vector Q producing a positive integer-valued output vector S (the summed output vector). This output vector S is thresholded by selecting the k columns in S with the highest scores. Thresholding produces a binary thresholded vector (T) with a bit set in each of the k highest scoring columns. Thus, after thresholding, T effectively lists the top k matching columns from the CMM thus identifying the top k matching records. These k nearest neighbours can then be used to provide a traffic intervention recommendation [1, 3, 4, 5] or to produce a traffic state prediction [2].

Results. The AURA k-NN has been applied to real-time intelligent transport monitoring [1, 2, 3, 4, 5]. For both recommendation and prediction, the AURA k-NN identifies similar historical traffic patterns: time periods in the past when the traffic state, as represented by a set of traffic sensor readings, is as close to the current state as possible. The only variation between recommendation and prediction lies in processing these matches. For recommendation, the set of traffic control interventions implemented during these similar historical time periods are cross-referenced from the historical data and recommended to the traffic operator for implementation, [1, 3, 4, 5]. For prediction, AURA k-NN extrapolates and produces a prediction of the future traffic value by averaging the variable value across the set of matches [2]. Results include tests of the AURA k-NN for recommendation performed using data from Hyde Park Corner (HPC) in London, UK [3]. There are 32 traffic sensors in the

HPC area with five serious or severe congestion events during the data recording period. We compared the state (congested/uncongested) of the top 5 matches selected by AURA k-NN against the recorded (actual) state for each sensor for each congestion event giving 800 sensor comparisons. AURA k-NN produced only 43 errors from the 800 sensor comparisons.

2.1 Hierarchical

We propose extending this AURA k-NN framework to a hierarchical distributed approach for intervention recommendation across large-scale networks.

A hierarchical processing approach allows the geographical location under consideration to be varied according to the spread of the traffic problem. The location could vary from a single junction up to a large city area as appropriate. We will fuse traffic and other relevant data from multiple locations, detect the spread of congestion at regular intervals and search for similar historical incidents with similar spreads of congestion, i.e. at a similar level in the hierarchy. Thus, the location and spread can vary dynamically at each time interval as the congestion spreads or contracts. AURA permits partial matching so historical incidents that are most similar with respect to both the incident features and the geographical spread will be found. The process is illustrated conceptually in Fig. 2.

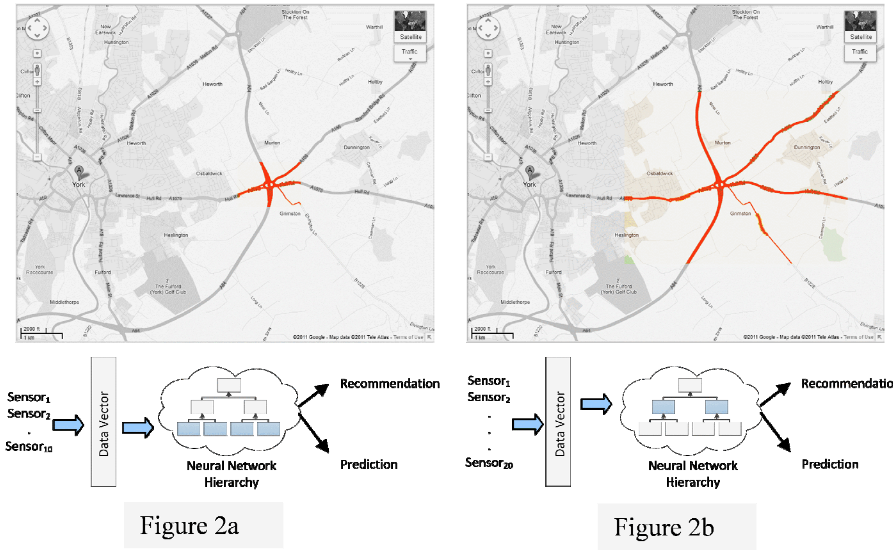

Fig. 2. In figure 2a, congestion is detected over a small area (10 sensors) and the neural network is used for recommendation and prediction against the data vector for the 10 sensors. In figure 2b, the congestion has spread to cover 20 sensors so the data vector for the 10 sensors expands to cover 20 sensors.

We can achieve a hierarchical representation using CMM row masking. Masking varies the number of variables to query. Thus, for a CMM storing data for forty sensors, to query sensors 1-10 we only examine the CMM rows that relate to those ten sensors and "mask off" all other rows, i.e., we do not activate those other rows. To query sensors 1-40, we examine all CMM rows (four times as many rows). This allows a simple, efficient "hierarchical" representation with no data repetition. Masking means that the number of sensors to query can be varied dynamically according to the geographical spread of congestion on the road network. Only sensors within the region of congestion need to be queried. The partial match capability of AURA allows the AURA k-NN to find the best match with respect to both the variables (sensor readings, weather etc.) and the geographical spread of the congestion.

The levels in the hierarchy represent different sized geographical locations and store data for different numbers of traffic sensors. The levels may represent just a few sensors at the bottom of the hierarchy but the levels at the top of the hierarchy may represent large geographical areas covering, potentially, thousands of sensors. Storing and processing data for large numbers of sensors requires careful consideration with respect to memory usage and processing speed. Storing data for more sensors increases the size of the CMM in the AURA k-NN which increases memory usage and slows processing [8]. However, the hierarchical processing approach described above lends itself to distributed processing. This could be processing the traffic data at the same geographical location across multiple compute nodes (parallel search) or even processing the traffic data at multiple geographical locations and assimilating the results (distributed processing). When the number of sensors is large but the sensor data is stored in one location then parallel searching is preferable, for example, sensor data for all traffic sensors of one local government authority area stored at one site. Where the number of traffic sensors is large and the data is physically spread across geographical locations such as traffic sensors covering multiple local authority areas where each authority stores data for its own sensors then distributed searching across the multiple sites is preferable.

Parallel. In [16], we demonstrated a parallel search implementation of AURA. This entails "striping" the data across several parallel CMM subsections. The CMM is effectively subdivided vertically across the output vector as shown in Fig. 3. In the traffic data, the number of variables m (sensor readings, journey times, weather data etc.) is much less than the number of records N (date/time records). Hence, $m << N$. Therefore, we subdivide the data along the number of records N (column stripes) as shown in Fig. 3.

Splitting the data across multiple CMMs using the date/time dimension (columns) means that the CMM can store data for all sensors, journey times, and weather data as separate rows within a single stripe. Each date/time is a separate record all contained within a single stripe.

Each separate CMM stripe (each separate date/time section) outputs a thresholded vector containing the top k records from that CMM stripe and their respective scores. All top k matches from all CMM stripes are aggregated and the top k matches overall can then be identified by finding the k matches with the overall highest scores.

Note: if the number of variables is large (for a large city area there may be a high number of sensors and other data) then it is possible to subdivide the data across multiple CMMs. The CMM is divided by the data/time of the records (column stripes) and then the column stripes are subdivided by the input variables (row stripes).

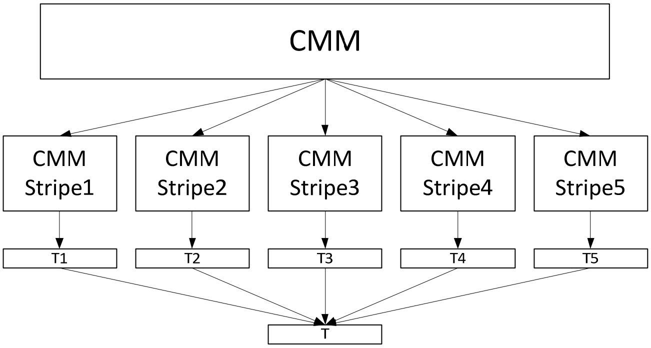

Fig. 3. The CMM covers a large time period so, to speed processing, the CMM is subdivided (striped) across a number of CMM stripes. Each CMM stripe produces a thresholded output vector T_n containing the top k matches (and their respective scores) for that stripe. All T_n are aggregated to form a single thresholded output vector T that lists the top k matches overall.

Dividing the CMM using the variables (row stripes) makes assimilating the results more complex. Each row stripe produces a summed output vector containing column subtotals for those variables within the stripe. The column subtotals need to be assimilated from all row stripes that hold data for that column. Thus, we sum these column subtotals to produce a column stripe vector C holding the overall sum for each column in that stripe and find the k top scoring columns as previously. Row striping involves assimilating integer vectors of length c where c is the number of columns for the column subdivision (column stripe).

Distributed. A distributed approach for AURA search has been proven in the context of condition monitoring for civil aerospace [17]. We wish to extend the principles developed in that domain and apply to them to the challenges of city or region wide monitoring of traffic patterns. There are two central challenges; maintaining a distributed data archive such that sensing and traffic data does not have to be moved to a central repository and secondly, orchestrating the search process across the distributed data. The distributed data issue can be addressed with existing software solutions, such as Hadoop and Storage Request Broker [18, 19].

To address the requirements for orchestrated search we have developed a grid solution that relies on a middleware stack for farming search queries across the distributed data resources. This component is termed the Pattern Match Controller (PMC). It provides a mechanism by which a front-end grid/web service client with an IDS system can submit queries to all of the known data resources in a parallel, asynchronous manner, and to manage the processing and analysis of the data at the remote repositories. Enforcing the pattern matching process to take place at the remote data repositories removes the costly requirement to shift large volumes of data during the search. The PMC builds on the Storage Request Broker (SRB) service to permit the virtualization of data repositories and data assets. This combination of technologies provides a scalable, high-volume, solution for pattern matching in complex signal processing and diagnostic domains. These methods are able to operate on Terabytes of data and hence are potentially scalable to wide regional traffic monitoring. We wish to extend this architecture to incorporate the requirements of

IDS for transport management, particularly the addressing the issues of impact analysis on changing traffic conditions in one area of a city, and investigating the consequences that may extend to other regions. We also wish to explore how the system can integrate traffic information from diverse sources, for example Government as well as regional traffic authorities, weather stations, smart sat-navigation systems and satellite information. There are many challenges with optimising the search process in highly distributed systems of this nature, and these issues are addressed in the following section.

3 Optimisation

One particular research focus is the area of pattern match optimisation. As with most machine-learning algorithms, AURA k-NN has parameters that need to be optimised to ensure highest accuracy. Optimisation is a combinatorial problem and traffic is dynamic over time so the AURA system needs re-optimising periodically to model the new data and keep the AURA memory up-to-date.

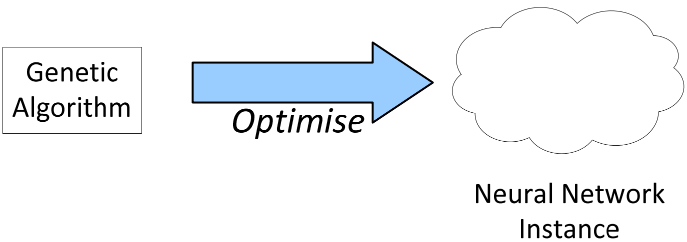

Fig. 4. The neural network parameter settings are optimised using a genetic algorithm.

We propose optimising a single node AURA system using, for example, genetic algorithms (GAs) or particle swarm optimisation (PSO) which have been used widely in the literature for this problem. In the following, we discuss using a genetic algorithm for optimisation as an example as shown in Fig. 4.

3.1 Optimisation using a GA

Genetic algorithms [20, 21] are inspired by Darwin's theory of evolution. In a GA, a chromosome represents a solution for the problem and chromosomes are selected according to their fitness; chromosomes with higher fitness stand a better chance of being selected. New chromosomes are produced using the genetic operators: cross-over and mutation. Generating the population of chromosomes involves the following stages.

1. Produce a random population of n chromosomes
2. Compute the fitness (score) for each chromosome, c, using a fitness function $f(c)$.
3. Repeat until the stopping criterion is met and the population is complete:
   - Select chromosomes from a population according to their fitness

- Apply cross-over to the selected chromosomes to create new chromosomes
- Apply mutate to the new chromosomes at each locus (position in chromosome)
- Compute the fitness of each new chromosome
- Update the population by replacing old chromosomes with the new chromosomes

The fundamental task in GAs is mapping the function to be optimised on to the chromosome. We propose using a binary valued chromosome where each gene encodes a parameter value of the AURA k-NN. This requires $log_2(r)$ bits per component where r represents the range of the values to optimise over and the $\log_2$ value is rounded up to the nearest integer. For example, to optimise the k value for k-NN across a range of values $5 \le k \le 50$ has a range of 45 then $log_2(45) \rightarrow 6$. Hence, optimising k requires a gene of 6 bits and the k values are encoded into their binary equivalents using 6 bits. This is illustrated in Fig. 5.

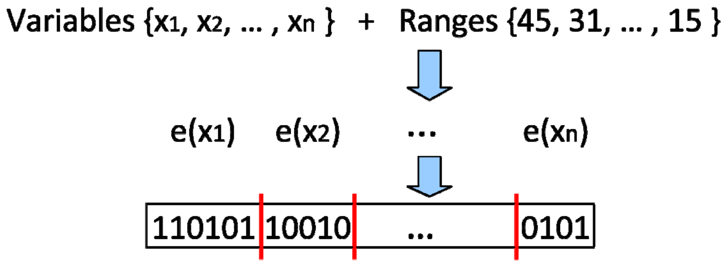

Fig. 5. Diagram demonstrating how the variables (parameter settings) are encoded in a GA chromosome. The range of value determines the number of bits in the gene. Here, variable x_1 has range 45 and uses 6 bits; variable x_2 has range 31 and uses 5 bits.

Each cycle of the GA generates a new chromosome that contains the parameter values to use for the AURA k-NN on that iteration. Running AURA k-NN with those parameter values produces a fitness score. This fitness score is either the recall accuracy for classification or the prediction accuracy for variable prediction. A high score reflects a set of parameters that work well. The chromosome is associated with that score and forms part of the chromosome population. Low scoring chromosomes "die" and are removed from the population. This process is repeated until the stopping criterion is met.

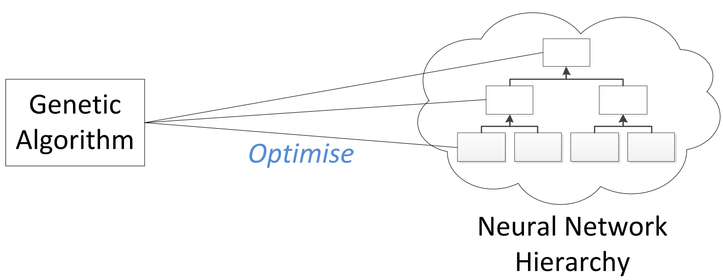

Fig. 6. The parameter settings for each separate CMM in the hierarchy are optimised using a genetic algorithm.

We then propose expanding this optimisation to the hierarchical representation so an identical GA optimisation process would run against each node in the hierarchy as shown in Fig. 6. In the hierarchy, each AURA k-NN is processing a different portion of the overall data set, from small regional areas to large geographical areas. These different data portions are likely to require different algorithm and data settings. Hence, a GA needs to be run against each AURA k-NN within the hierarchy to optimise its algorithm and data parameters.

3.2 AURA Meta-learner

GAs are easily parallelisable and do not get stuck in local minima. However, they are computationally expensive. The optimisation process needs to be rerun periodically to accommodate new data into an existing historical database or, if the road layout changes, then the appropriate database needs to be remodelled etc. Hence, we finally propose a k-NN-based meta-learner as shown in Fig. 7.

Authors have considered using a k-NN based meta-learner for selecting datasets that are similar to the current data set. Learning algorithms that performed best on the most similar previous dataset can then be selected for use for the current data set [22]. Our approach will use an instance of the AURA k-NN to store the results of the optimisations run previously and learn the best data and parameter settings for the AURA k-NN that were used to process previous data sets. These best settings may then be used to bootstrap future optimisations and short-circuit the optimisation process which is a combinatorial problem and, hence, computationally intensive. The AURA k-NN distance function to calculate data set similarity will be based on the features of the dataset such as the data types (integer-valued, real-valued or symbolic), the data ranges or the data set size (number of variables or number of records). Matching the features of the new data set against the features of the stored data sets using AURA k-NN will elicit the most similar data sets (neighbours). The AURA k-NN can then cross-reference the parameter settings used to process the most similar dataset(s) (nearest neighbours), that is, the best settings to use to initialise the GA for optimising the current dataset. These AURA k-NN parameter settings can then be used to initialise the GA, i.e., produce an initial set of chromosomes and, hopefully, reduce the search space of the GA through focused initialisation of the GA chromosomes.

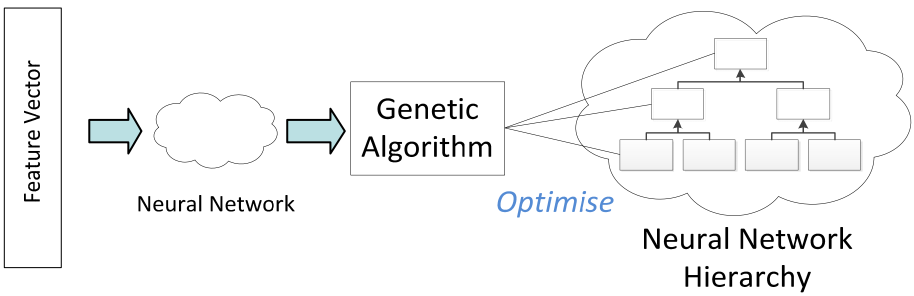

Fig. 7. A neural network can function as a meta-learner to retrieve the best initialisation settings for the GA using characteristics (features) of the data set.

4 Conclusions

The paper describes an Intelligent Decision Support tool for pattern matching aimed at data processing, optimisation, recommendation and prediction. The tool is based on a binary neural implementation of the k-nearest neighbour algorithm, AURA k-NN. AURA k-NN is fast and scalable. It can vary the region of interest dynamically, process data in parallel by subdividing processing using the time dimension and process data across a number of sites using distributed processing. We propose using a genetic algorithm (or similar) to optimise the algorithm and data settings for the pattern matcher. Additionally, the pattern matcher itself can be used to store initialisation settings for the genetic algorithm thus short-circuiting the optimisation process of the genetic algorithm which is computationally intensive. The pattern matcher stores characteristics of the data sets as feature vectors and matches the characteristics of the new data set against the stored data sets to find the most similar stored data set. The optimisation settings that were used for this stored data set can then be used to initialise the genetic algorithm for optimising the new data set.

References

1. Hodge, V., Krishnan, R., Austin, J. and Polak, J. (2010). A computationally efficient method for online identification of traffic incidents and network equipment failures. Presented at, *Transport Science and Technology Congress: TRANSTEC 2010,* Delhi, Apr. 4-7, 2010.
2. Hodge, V., Krishnan, R., Jackson, T., Austin, J. and Polak, J. (2011). Short-Term Traffic Prediction Using a Binary Neural Network. Presented at, *43rd Annual UTSG Conference,* Open University, Milton Keynes, UK, Jan. 5-7, 2011.
3. Krishnan, R., Hodge, V., Austin, J. and Polak, J. (2010a). A Computationally Efficient Method for Online Identification of Traffic Control Intervention Measures. Presented at, *42nd Annual UTSG Conference,* University of Plymouth, UK: Jan. 5-7, 2010.
4. Krishnan, R., Hodge, V., Austin, J., Polak, J. and Lee, T-C. (2010b). On Identifying Spatial Traffic Patterns using Advanced Pattern Matching Techniques. In, *Proceedings of Transportation Research Board (TRB) 89th Annual Meeting, Washington, D.C., Jan. 10-14, 2010.* (DVD-ROM: Compendium of Papers).
5. Krishnan, R., Hodge, V., Austin, J., Polak, J., Jackson, T., Smith, M. and Lee, T-C. (2010c). Decision Support for Traffic Management. In, *Proceedings of 17th ITS World Congress: (CD-ROM), Busan: Korea, Oct. 25-29, 2010.*
6. Cover, T. M. and Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory,* 13(1):21–27.
7. Fix, E. and Hodges, J. L. (1951). Discriminatory analysis, nonparametric discrimination: Consistency properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas.
8. Hodge, V. and Austin, J. (2005). A Binary Neural k-Nearest Neighbour Technique. *Knowledge and Information Systems*, 8(3): pp. 276-292, Springer-Verlag London Ltd, 2005.
9. Hodge, V., Lees, K. and Austin, J. (2004). A High Performance k-NN Approach Using Binary Neural Networks. *Neural Networks*, 17(3): pp. 441-458, Elsevier Science, 2004.
10. Hodge, V. and Austin, J. (2001). An Evaluation of Standard Retrieval Algorithms and a Binary Neural Approach. *Neural Networks*, 14(3): pp. 287-303, Elsevier Science.

11. Weeks, M., Hodge, V., O'Keefe, S., Austin, J. and Lees, K. (2003). Improved AURA k-Nearest Neighbour Approach. In, *Proceedings of IWANN-2003*, Mahon, Spain. June 3-6, 2003. *Lecture Notes in Computer Science (LNCS) 2687*, Springer Verlag, Berlin.
12. Bubeck, S. and von Luxburg, U. (2009). Nearest Neighbor Clustering: A Baseline Method for Consistent Clustering with Arbitrary Objective Functions. Journal of Machine Learning Research, 10(10): 657-698.
13. Hodge, V. and Austin, J. (2004). A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 22: pp. 85-126, Kluwer Academic Publishers.
14. Hodge, V. (2011). *Outlier and Anomaly Detection: A Survey of Outlier and Anomaly Detection Methods*. Lambert Academic Publishing (LAP), ISBN-13: 978-3846548226
15. AURA -- Advanced Computer Architectures Group web page (accessed 15/12/11) http://www.cs.york.ac.uk/arch/neural-networks/technologies/aura
16. Weeks, M., Hodge, V. and Austin, J. (2002). A Hardware Accelerated Novel IR System. In, *Proceedings of the 10th Euromicro Workshop (PDP-2002), Las Palmas de Gran Canaria, Canary Islands, Jan. 9–11, 2002.* IEEE Computer Society, CA.
17. Austin, J., Davis, R., Fletcher, M., Jackson, T., Jessop, M., Liang, B. and Pasley, A. (2005). DAME: Searching Large Data Sets within a Grid-Enabled Engineering Application. *Proceedings IEEE - Special Issue on Grid Computing,* 93(3): 496-509, ISBN 0018-9219
18. Shvachko, K., Hairong K., Radia, S. and Chansler, R. (2010). The Hadoop Distributed File Store System. In, *Proceedings of 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST*), 28 June 2010.
19. SDSC Storage Request Broker [Online] (accessed 15/12/11). Available: http://www.sdsc.edu/srb/index.php/Main_Page
20. Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning* Addison-Wesley Pub. Co. ISBN: 0201157675
21. Holland, J.H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*, University of Michigan Press.
22. Brazdil, P. B., Soares, C. and Da Costa, J.P. (2003). Ranking Learning Algorithms: Using IBL and Meta-Learning on Accuracy and Time Results. *Machine Learning,* 50(3): 251-277.

Perspectives of Internet based Road Network Traffic Flow Modelling and Control

Apostolos Kotsialos

School of Engineering and Computing Sciences, Durham University
South Road, Durham, DH1 3LE, U.K.
apostolos.kotsialos@durham.ac.uk

Abstract. This paper is concerned with possible future internet applications within the framework of traffic flow modelling and control. Two modelling related uses are described based on the assumption that vehicles in the near future will be equipped with smart on-board devices that can communicate with each other through reliable *ad-hoc* wireless networks. This capability opens new ways of thinking about traffic flow and requires the explicit consideration of the drivers' behaviour when more information is available to them pertaining to downstream traffic conditions. Finally, a web-based application for supporting ramp metering is discussed as well.

1 Introduction

The ubiquitous traffic light is probably the most important device related to road traffic control. It was introduced at the beginning of the 20th century for improving the safety of road junctions, by giving right of way to passing vehicles. A century later and with the advent of transport telematics, the traffic light and its operation is still a hot topic of research. The scope of course is much wider now, but the simple fact remains that a device introduced a hundred years ago is still the effector of our most sophisticated traffic control strategies. Why has this simple device endured a century of use? The reason is that it is a simple and effective means of communication, highly tuned to human cognition norms.

The advent of the internet and wireless communication networks has brought forward a new communications medium which interfaces both with transport units/infrastructure and drivers. Personalised devices are increasingly becoming connected and interactive getting access in diverse databases and having significant computation power for data processing and decision making support. This personalised device, in the form of GPS, smartphones, tablets etc, can now become the new traffic control sensor and effector. This technology has the potential to become the new "traffic light" and in that sense the new Variable Speed Limit (VSL) sign and Variable Message Sign (VMS) as well. The added feature is the personalised information and the custom demand for traffic control decisions on the vehicle level. With the exception of special types of vehicles (ambulances, buses etc) it is difficult for the traffic light at a junction to be aware of each of the queued vehicles needs, as at most there can be some OD information estimate available. Personalised devices can communicate this information effectively pre-trip and en-route. This gives rise to a whole new range of traffic control strategies which go beyond the minimisation of an aggregate control objective, and go to the custom made individual and collective traffic management decisions. Important questions regarding the level of aggregation and emergent traffic flow behaviour need to be posed and answered. Hence, for traffic control purposes, the internet can be viewed as an additional extra system layer added to the classical Infrastructure-Vehicle-Driver (IVD) transport system. The internet and wireless communication technologies facilitate the addition of Information in that triplet, i.e. the IVD-Information (IVDI) transport system.

This calls for a revision of traffic flow models both at the micro and the macro (hence also at the meso) level and that should be fed to traffic surveillance and control applications. Traffic flow models should include information in a more fundamental and structured way as one of the constituent elements of traffic. Especially Model Predictive Control (MPC) for integrated control measures in mixed corridors, has to gain a lot in terms of efficiency from this revision. Thus, the internet calls for a new generation of traffic controllers over wide mixed corridors where information is a system component, rather than an external input.

As connectivity becomes an essential and widely adopted feature of our societies, similar to water and electricity network access, internet access for drivers, vehicles and on-board devices will become the norm. We can envisage a distributed virtual traffic control centre spread over wide geographic areas using robust, highly reliable and redundant communications networks, and mobile computing power to support traffic management in terms of surveillance and control. This can be achieved by adopting hierarchical, decentralised and peer-to-peer architectures using networks of sensors and also notions of networked control over limited bandwidth communication channels. Hopefully, this effort will result to the introduction of new notions, ideas, concepts and technologies that will last for the good part of this century.

This paper is structured as follows. Section 2 provides a brief overview of some ideas about ways of using the internet for traffic modelling and control purposes. Section 3 describes a possible application affecting microscopic car following dynamics; macroscopic modelling is discussed in section 4. Section 5 discusses the use of web-based applications affecting driver departure in support of ramp metering operations. Section 6 concludes this paper.

2 Internet based Applications for Traffic Management

The internet is a diverse network of clients and servers communicating using different protocols. Its most important feature is the networked structure supporting its robustness. Within the framework of traffic management systems, the most obvious use of internet for surveillance and control purposes is exploitation of the information dissemination potential of cyber space. In this case, the general model is that of a driver receiving information and responding to it. The informed drivers make their choices based on this, something which may be viewed as a limiting factor to the application of control measures. Indeed, the prevailing paradigm with respect to traffic control applications is that of a TCC being responsible for the operation of compulsory control measures in an effort to regulate some crucial traffic variable(s). Responsibility and liability for these operations are placed on the traffic authority. Drivers are required to comply with the control measures' operational decisions, e.g. in case of ramp metering they have to wait at the on-ramp until they get a green light to enter into the mainstream; in case of compulsory variable speed limits, they must observe the speed limit communicated to them. The scope of internet based applications in this framework is limited as it is defined by the needs of the information/communication infrastructure and software development.

However, there are other possibilities worth exploring by changing the way traffic control measures operate. High connectivity and real-time information may be used for designing the traffic control operations. This can be done by means of timely delivery of information allowing or motivating drivers to behave in such ways as to optimise the traffic flow process. This requires accepting the fact that drivers are intelligent agents able to operate efficiently in changing environmental conditions. Hence, a new type of control approach can be envisaged, that assigns more emphasis on driver behaviour rather than just TCC issued compulsory suggestions or actions. Thus, information as disseminated by internet based applications can be considered as a structural element of modelling and control design of road network traffic.

The key feature in this approach is the support of emergence, i.e. the process by which a desired global road network-wide macro-state emerges from the application of local interactions in the micro-state. This is a potential of information intensive applications that needs to be tapped. By informing drivers with the appropriate pieces of information individual decision making and driver behaviour can be steered towards establishing a desired global state. This scenario supports the "informed drivers" who are responsible for their own decision; the traffic authority is responsible for facilitating the efficient and reliable collection and dissemination of information, and the maintenance of this infrastructure, rather than the control measures' decision making.

Different applications can be envisaged based on this approach. Three such scenarios are discussed here. The first one is based on classical microscopic car-following dynamics. The second, is a comment on the analysis of macroscopic dynamics based on non-local information. Finally, a third one is concerned with the possibility of using the internet for managing the demand in conjunction with ramp metering.

3 Changing the Car-following Dynamics

In this section a possible application of *ad-hoc* wireless networks [1] is considered for changing the behaviour of a traffic stream towards more efficient use of the road capacity.

Let us consider a homogeneous traffic stream moving along a single-lane motorway as shown in Figure 1, [2], [3], [4]. Modelling of this stream on the microscopic level is usually done by considering the car-following dynamics theory developed in the 50s and 60s. One of the most common general family of model are those developed by General Motors (GM) researchers and are based on the stimuli-response driver behaviour description. In these models, drivers respond to external stimuli and the response is a function of the drivers' sensitivity and the stimuli's strength, i.e.

$$\text{response} = f\left(\text{driver sensitivity, stimulus strength}\right). \tag{1}$$

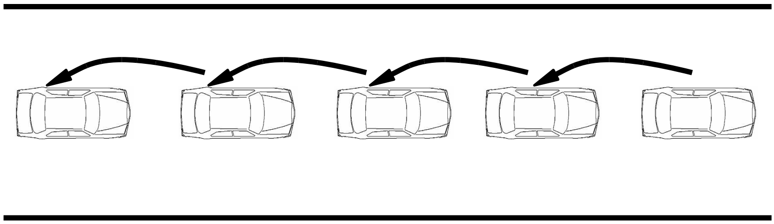

Fig. 1. Car-following with interactions on consecutive vehicles.

In the GM models, the response is always expressed as the vehicle's acceleration or deceleration. The general car-following model set can be viewed in Figure 2. The leading vehicle has index n and the follower $n+1$. Vehicle $n+1$ responds, i.e. changes its acceleration profile, to changes in the state of the leader n. L_n is the vehicles n's length and $x_n(t)$ is its position at time t. The gap between the follower and the leader is g_{n+1} and their distance at time t is considered to be $x_n(t) - x_{n+1}(t)$. Then, the general (ℓ, m) car-following model reads

$$\ddot{x}_{n+1}(t+\Delta t) = \frac{\alpha \dot{x}_{n+1}(t+\Delta t)^m}{[x_n(t) - x_{n+1}(t)]^\ell} [\dot{x}_n(t) - \dot{x}_{n+1}(t)] \tag{2}$$

where α is the driver $n+1$ sensitivity to changes in the relative speed between vehicles n and $n+1$, $\dot{x}_n(t) - \dot{x}_{n+1}(t)$ at time t and their distance. Driver $n+1$ reacts with a delay Δt and this is influenced by the speed at the time of reaction $t + \Delta t$. The pair (ℓ, m) are parameters that allow consideration of a whole family of models, but for our purposes here the $(\ell = 1, m = 0)$, i.e.

$$\ddot{x}_{n+1}(t+\Delta t) = \frac{\alpha [\dot{x}_n(t) - \dot{x}_{n+1}(t)]}{[x_n(t) - x_{n+1}(t)]} \tag{3}$$

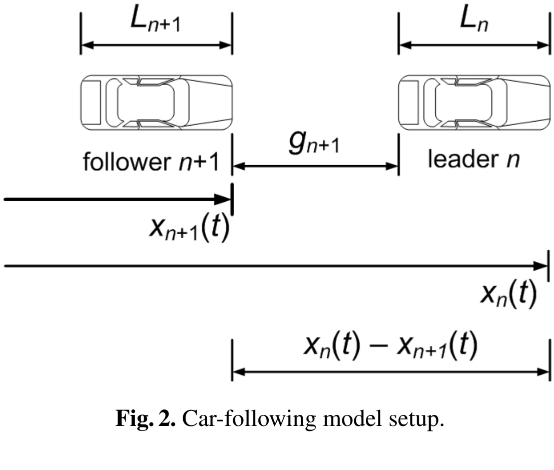

Fig. 2. Car-following model setup.

These car-following models have proved very useful and they are in constant improvement since they are at the core of most microscopic simulators. The pattern of eqn. (1) is preserved, and essentially it models the interaction between two consecutive vehicles in a traffic stream, Figure 1. The drivers' visual sensor and perception of speed allows them to react to the next downstream vehicle motion. Now, let us assume that the same kind of information is given to the follower, through an on-board device, *regarding the vehicle with minimum speed within range R from vehicle* $n+1$. Hence, it is assumed that each vehicle in the traffic stream has an on-board device with a range R that discerns the vehicle f at detectable distance $x_f(t) - x_{n+1}(t)$ further downstream and has the smallest speed of all vehicles downstream $n+1$ within distance R. Hence, driver $n+1$ receives two stimuli of the same nature, but from different sources:

- a stimulus from the vehicle n immediately downstream through visual perception and
- a stimulus from the on-board device for the downstream vehicle f with the minimum speed within distance R.

Obviously, the design of such an on-board device is not a trivial task and requires the study of drivers' behavioral and cognitive features, but this is not within the scope of this discussion. It is sufficient to assume that this information is transmitted via a robust communication network in the form shown in Figure 3, where each vehicle sends its location and speed to those preceding it. Furthermore, let us assume that the drivers' reaction to both stimuli above follows the same model of eqn. (1). Then the acceleration (drivers' response) is given by

$$\ddot{x}_{n+1}(t + \varDelta t) = \min \left\{ \frac{\alpha \left[\dot{x}_n(t) - \dot{x}_{n+1}(t)\right]}{\left[x_n(t) - x_{n+1}(t)\right]}, \frac{\alpha_f \left[\dot{x}_f(t) - \dot{x}_{n+1}(t)\right]}{\left[x_f(t) - x_{n+1}(t)\right]} \right\} \tag{4}$$

where a_f the drivers' sensitivity to the stimulus coming from the on-board device.

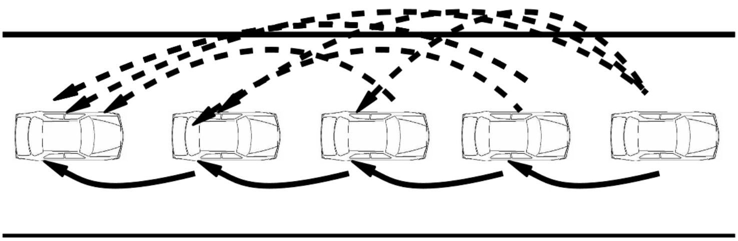

Fig. 3. Car-following with vehicle interactions with on-board devices.

Again, eqn. (4) models an idealised situation, since the traffic flow process is much more complex. However, it does provide an initial insight into how to use information for improving the use of road capacity without the direct implementation of a traffic control measure.

In order to see that, let us conduct a simulation experiment using models (3) and (4) for a 10 km road stretch with vehicles with the following parameters:

- vehicle maximum speed possible $v_{\max} = 120$ km/h;
- vehicle maximum acceleration $\gamma_{\max} = 12,960.0$ km/h^2;
- vehicle maximum deceleration $\gamma_{\min} = 19,440.0$ km/h^2;

– stimulus sensitivities $\alpha = \alpha_f = 29.412$ km/h.

Now, assume that the lead vehicle of the traffic stream, i.e. the first vehicle that enters the road, moves at maximum up to a point where it decelarates (with maximum constant deceleration) for some reason over a period of time. After that, it moves with constant speed for a time period after which, it begins to accelerate (with maximum constant acceleration) until it reaches its maximum speed. Afterwards it moves with constant maximum speed, having the road empty in front of it. The result of this pattern of motion of the stream lead vehicle for a length of 10 km can be seen in the time-distance diagram in Figure 4. A number of shockwaves are created and high density areas can be discerned. This is a typical pattern formed predicted by car-following theory.

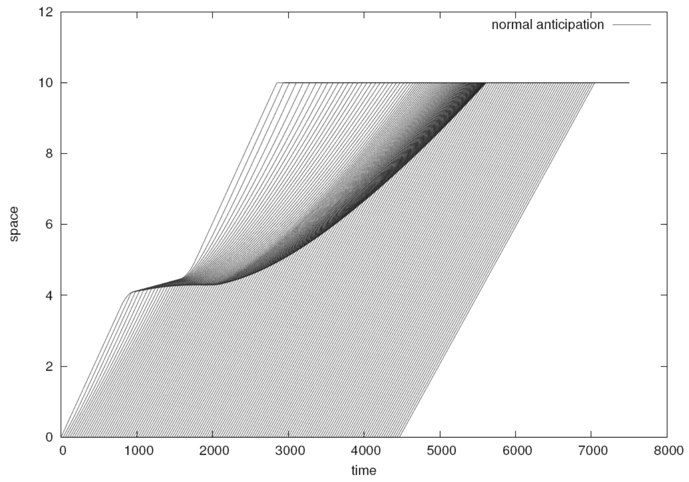

Fig. 4. Time-distance diagram for normal car-following.

Let us assume now the scenario of Figure 3 and the application of eqn. (4) for a range $R = 16$ meters. The results are shown in the time-distance diagram of Figure 5. The earlier warning produced by the on-board devices results to better utilisation of the road facility, since more vehicles are served per unit of time, where service in this framework is the arrival of a vehicle at point 10 km downstream the road. Figure 6 depicts the time-space diagram of model (4) when the range is set to 100 meters. In this case, the efficiency of the resulting traffic stream is reduced, but the high concentration areas, observed in Figures 4 and 5, are dispersed, which means an improvement to safety. Figure 7 depicts the exponentially smoothed outflows at the end of the road. It can be shown that for any traffic system, the minimisation of the total time spent by vehicles in it, travelling and queueing, is equivalent to the maximisation of the time weighted system outflows, [5]. Hence, the traffic stream's efficiency of the three different scenarios is shown in Figure 7. The most efficient is the one where the peak in outflows happens earlier, which clearly is the scenario with anticipation $R = 16$ meters. The most inefficient is the scenario with $R = 100$ meters, but as mentioned it is the safest.

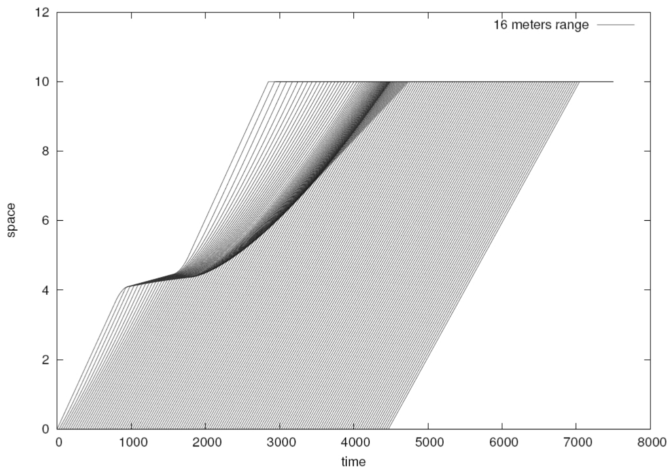

Fig. 5. Time-distance diagram for on-board based car-following with $R = 16$ m.

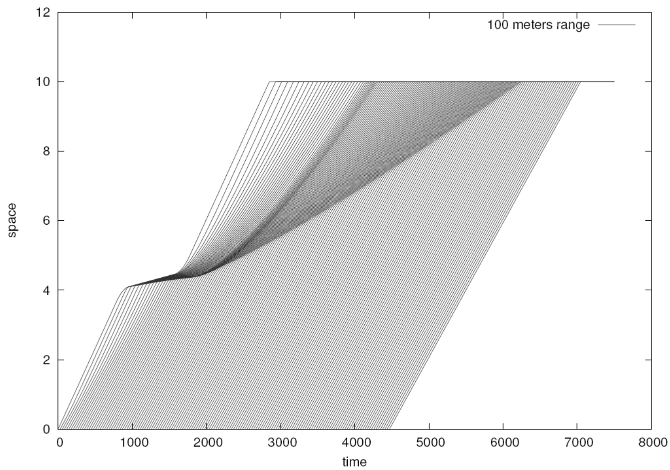

Fig. 6. Time-distance diagram for on-board based car-following with $R = 100$ m.

The main issue to be noted here is the capability provided by the on-board devices for achieving a desirable traffic state, in terms of efficiency or safety, without the direct mediation of the TCC, e.g. by providing compulsory speed advice to drivers. It is the drivers themselves that do the decision making and based on that process the desired traffic state emerges. The traffic authority need only provide the networking support applications that will allow the setup of a robust, reliable and fast communication network.

This kind of analysis has been based here on a lot of idealised assumptions, but this line of reasoning can be extended to more realistic situations, where a lot of the parameters characterising the system are stochastic variables rather than deterministic. More detailed results will be reported elsewhere.

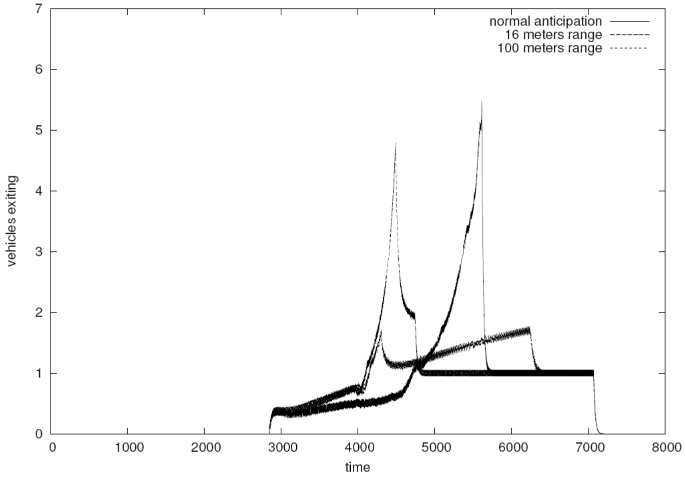

Fig. 7. Smoothed system outflows for the three scenarios.

4 The Impact of On-board Devices on Macroscopic Models

Macroscopic traffic flow models describe traffic in terms of vehicular density, average speed and volume (flow). The vehicle conservation equation that lies in the core of every macroscopic model reads [6], [7]

$$\frac{\partial \rho(x,t)}{\partial t} + \frac{\partial}{\partial x}\left[\rho(x,t)v(x,t)\right] = 0. \tag{5}$$

where $\rho(x,t)$ and $v(x,t)$ are the vehicular density and average speed at point x at time t, respectively. When the traffic flow is in equilibrium at point x at time t, the relationship between speed and flow is given by the fundamental diagram $V\left[\rho(x,t)\right]$. Second order models use an empirical speed equation in conjunction with the conservation equation (5). One of the possible forms of this equation reads [8], [9]

$$v(x,t+\tau) = V\left[\rho(x+\Delta x,t)\right] \tag{6}$$

Equation (6) states that the average driver will need some time τ to react to the stimulus provided by traffic density downstream at distance Δx. A similar assumption was made for the microscopic models, but here the behaviour is averaged. A Taylor expansion argument on both sides of (6) yields the following dynamic speed equation [9]

$$\begin{aligned} & v(x,t) + \tau\frac{\partial v(x,t)}{\partial t} + \tau v\frac{\partial v(x,t)}{\partial x} = \\ & V\left[\rho(x,t)\right] - \frac{\nu}{\rho(x,t)}\frac{\partial \rho(x,t)}{\partial x} \Rightarrow \\ & \frac{\partial v(x,t)}{\partial t} = -\frac{\partial v(x,t)}{\partial x} \end{aligned}$$

$$+\frac{1}{\tau}\left[V\left[\rho(x,t)\right]-v(x,t)\right]$$
$$-\frac{\nu}{\tau}\frac{1}{\rho(x,t)}\frac{\partial\rho(x,t)}{\partial x} \tag{7}$$

where due to microscopic considerations

$$\Delta x=\frac{0.5}{\rho} \tag{8}$$

has been used and ν is a model parameter given by

$$\nu=-0.5\frac{\partial V}{\partial\rho}>0 \tag{9}$$

which is constant when a linear fundamental diagram is assumed.

The key parameter that is of interest here is the determination of Δx. In the original approach, eqn. (8) gives sufficient small values to Δx for the Taylor expansion theorem to be valid. However, with the use of on-board devices in a scenario as that shown in Figure 3, the distance Δx in eqn. (6) can take much larger values. Hence, the Taylor expansion argument is not straightforward anymore. New or revised models need to be developed that will consider the impact of information from on-board devices to the macroscopic description of traffic.

5 Using the Internet for Improving Ramp Metering Control

Ramp metering is one of the most effective control measures applied in motorway networks. Figure 8 depicts its basic principle of operation. An on-ramp o is used by demand originating from the residential areas adjacent to the motorway $d_o(k)$ (veh/h) during period k. Vehicles are queued into the on-ramp forming a queue of length $w_o(k)$ (number of vehicles). A traffic light installed at entrance of the on-ramp to the motorway mainstream regulates the inflow $q_o(k)$ (veh/hour) from the queue into the traffic stream of the first segment of link μ.

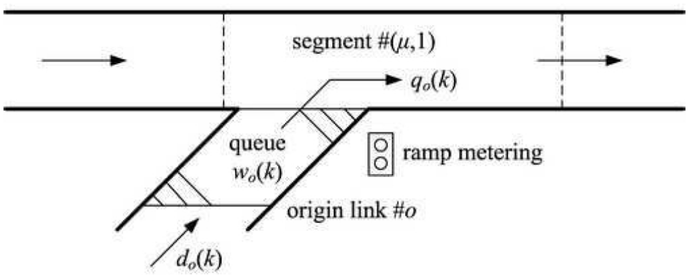

Fig. 8. Ramp metering control measure setup.

One of the most efficient approaches to coordinated ramp metering is based on discrete-time nonlinear optimal control. The general discrete time nonlinear optimal control problem reads [10]

Minimise

$$J = \vartheta\,[K] + \sum_{k=0}^{K-1} \varphi\,[\mathbf{x}(k), \mathbf{u}(k), \mathbf{d}(k)] \qquad (10)$$

subject to

$$\mathbf{x}(k+1) = \mathbf{f}\,[\mathbf{x}(k), \mathbf{u}(k), \mathbf{d}(k)]\,,\ \mathbf{x}(0) = \mathbf{x}_0 \qquad (11)$$

$$u_{i,\min} \leq u_i(k) \leq u_{i,\max}\ \forall i = 1, \ldots, m \qquad (12)$$

where K is the time horizon, k the discrete time index, $\mathbf{x}$ the system's state vector, $\mathbf{u}$ the bounded control vector and $\mathbf{d}$ the disturbance vector, i.e. the uncontrolled inputs to the process. ϑ and φ are smooth cost functionals. $\mathbf{f}$ is the controlled process's model.

In the case of ramp metering of a motorway network, a discrete time nonlinear macroscopic model of the whole network is developed based on a discretisation scheme of the macroscopic equations, e.g. eqns. (5) and (7). The objective function selected represents an appropriate cost criterion; the most commonly used is the TTS, which can easily be expressed in terms of the model's macroscopic variables [10]. The state vector consists of the density and average speed of the segments into which the motorway has been divided into and the queue lengths at the origins (on-ramps and motorway entrances). The control vector consists of the ramp metering inflow rates allowed and regulated by the traffic light at every on-ramp. Finally, the disturbance vector consists of the demand originating from the adjacent residential area and the outflows of the vehicles to their destination off-ramps.

This approach to ramp metering strategy design has been shown to be very effective and tends to exploit the capacity of the on-ramps in an optimal way. Detailed investigations have revealed that one of the most important parameters for the efficient use of ramp metering is the storage capacity of the on-ramps, i.e. the maximum number of vehicles that can be stored into the on-ramps at any given point in time [10]. This is a crucial parameter for the surrounding surface road network as well, since small on-ramps tend to spill vehicles into residential areas and therefore degrade environmental conditions in sensitive areas. Hence, high demands $\mathbf{d}$ will result to higher queues. But if a maximum queue constraint is imposed on the ramp metering strategy, then efficiency will have to be sacrificed to the benefit of reduced interference of motorway with urban street traffic (for reasons of equity as well).

The ramp metering strategy itself cannot do anything to change this condition, since $\mathbf{d}$ is a disturbace, i.e. an uncontrolled input to the traffic flow process. It is exactly here that internet applications can be used to support the operation of coordinated ramp metering. It is relatively straightforward for a traffic authority to set up a web-based service informing in *real time* drivers who want to use the motorway network and access it via a specific on-ramp (or a set of possible on-ramps) about the travel conditions and expected travel time as they are waiting home to depart. This information can be highly customised to the individual needs. Such a service would result to drivers changing their departure time from their homes, e.g. to the morning commute.

In terms of ramp metering control, a service like this exploits the storage capacity of the residential areas. Instead of storing vehicles just in the limited space of the on-ramps, an on-line application providing a motive to drivers to wait and depart later, stores the demand further upstream the on-ramps without degrading environmental conditions in residential areas. That increase in storage capacity will increase the efficiency of ramp metering strategies. Furthermore, real time demand distribution could take place by providing advice regarding which on-ramp should be used for the trip. Obviously, this requires a highly reliable service and real-time information given to users and the development of the appropriate interfaces, but this is something feasible with existing technology without major problems.

6 Conclusions

In this paper a few ideas have been put forward regarding the future use of internet for traffic flow modelling and control. Two modelling related uses have been described based on the assumption that vehicles in the near future will be equipped with smart on-board devices that can communicate with each other through reliable networks setting up their own internet. This capability opens new ways of thinking about traffic flow and requires the explicit consideration of the drivers' behaviour when more information is available to them pertaining to downstream traffic conditions.

A discussion was provided under idealised assumptions regarding the microscopic dynamics of a traffic stream and how the information provided by such on-board devices can be used to achieve the same effect as if there was a traffic control strategy explicitly used. The second example was concerned with possible changes or revisions in the macroscopic modelling of traffic, were again the impact of information on driver behaviour needs to be considered, as it becomes an important feature. Finally, the use of a classical web-based application has been described, that can work in conjunction with and in support of ramp metering operations. By providing information about travel times and traffic conditions and even suggestions, the time of departure can be influenced so that in effect the residential areas are used as storage areas for vehicles, similarly to the use of the on-ramp storage capacity from ramp metering strategies.

These ideas, briefly described here, can be further elaborated as they offer some good research directions to be followed and pose significant challenges.

References

1. Sharif, B. S., Blythe, P. T., Almajnooni, S. M., Tsimenidis, C. C.: Inter-vehicle mobile ad hoc network for road transport systems. IET Intelligent Transport Systems. 1 (2007) 47–56
2. Herman, R., Montroll, E. W., Potts, R. B., Rothery, R. W.: Traffic dynamics: Analysis of stability in car following. Operations Research, 7(1) (1959) 86–106
3. Herman, R., Potts, R. B.: Single-lane traffic theory and experiment. In: Theory of Traffic Flow. Proceedings of Symposium Held at Research Laboratories General Motors Corporation, Warren, Michigan, U.S.A., Elsevier (1961)
4. A. D. May: Traffic Flow Fundamentals Prentice Hall (1990)
5. Papageorgiou, M.: Application of Automatic Control Concepts to Traffic Flow Modelling and Control. Springer, New York (1983)
6. Lighthill, M. J., Whitham, G. B.: On kinematic waves II: a traffic flow theory on long crowded roads. Proc. of the Royal Society of London Series A, 229 (1955) 317–345

7. Richards, P. I.: Shock waves on the highway. Operations Research, 4 (1956) 42–51
8. Payne, H. J.: Models of freeway traffic and control. Simulation Council Proc., 1 (1971) 51–61
9. Papageorgiou, M., Blossville, J. M., Hadj-Salem, H.: Modelling of traffic flow on the Boulevard Périphérique in Paris. Transportation Research B, 23(1) (1989) 29–47
10. Papamichail, I., Kotsialos, A., Margonis, I., Papageorgiou, M.: Coordinated ramp metering for freeway networks. A model-predictive hierarchical control approach. Transportation Research: Part C 18(3) (2010) 311–331

Network Monitoring and Personalized Traffic Control: A Starting Point based on Experiences from the Municipality of Enschede in the Netherlands

Tom Thomas and Sander Veenstra

Center of Transport Studies, University of Twente, P.o box 217
7500 AE Enschede, The Netherlands
t.thomas@utwente.nl, s.a.veenstra@utwente.nl

Abstract. An increasing number of cities have severe traffic problems. We identify three main challenges for managing these problems. The first one is to achieve a proper amount of monitoring. Secondly, predictions of the effects of network wide management measures require knowledge of the underlying travel behaviour. Finally, measures should be in line with needs and expectations of travellers to be effective. In this paper we focus on these challenges. We use loop detectors near traffic lights in the Dutch city of Enschede to monitor the traffic situation in its network. We developed a method to estimate delays from these measurements. We also use a simple forecasting algorithm to predict flows and travel times for different time horizons. Regarding travel behaviour, we used a license plate survey to study route choice. We discuss how the results from these studies may be used to improve urban traffic management.

1 Introduction

Congestion has increased significantly in the last few decades. This has led to serious problems, especially in urban areas. Travel times for travellers have increased and the livability in residential areas has declined due to pollution and issues concerning safety. The efficient use of existing infrastructure is one of the strategies to reduce these problems. This requires sufficient monitoring of the traffic status of (parts of) the network. Only recently systematic data collection by roadside systems like loop detectors and cameras has increased significantly in urban areas. As a result, several traffic management centers have been applied for urban networks, e.g. [1], [2], [3].

In large cities in developing countries and Eastern Europe, however, still only few roadside sensors are available to collect the necessary traffic data. At the same time, an increasing number of travelers use smart devices which enable them to trace their position by e.g. GMS or GPS. With these devices traffic information, such as travel time or amount of congestion, can be provided to users in real-time. Increasingly, this is done by private companies. However, the underlying data they use, is often not freely available. It is therefore likely that in the near future, traffic managers will still need to use some roadside systems combined with limited amount of floating car data. One of the challenges will be to generate sufficient traffic data in cities which have a limited amount of sensors.

With sensors throughout the network, network wide control scenarios can be developed for managing the network. For unsaturated flows, local vehicle actuated controls can minimize delays, and green waves, e.g. [4], at a main arterial road can reduce the number of stop and go moments, but in large cities with many congested roads these measures are often not sufficient. However, optimization of network wide strategies is a complex task, especially when real-time data are used. Simplification of the network therefore appears to be inevitable. In that regard, a hierarchical network architecture, in which controls are clustered in a tree structure, appears to be a promising approach, e.g. [5].

Although a lot of progress has been made in adopting network wide control strategies, their possibilities would be greatly enhanced when historical data and underlying traffic patterns, i.e. origin destination flows and route choice behavior, would be included. From historical traffic data, (short term) predictions of the traffic status can be made, for example by pattern matching, e.g. [6]. These predictions enable controllers to anticipate on the (near) future. Moreover, information about OD patterns and route choice can be used to simulate the outcomes of many different control scenarios in advance, e.g. [7]. These simulations may cover whole periods, like complete rush hours. In this way, timely measures to control the amount of traffic on access roads, may for example prevent congestion in the city center at a later time. In addition, dynamic assignment models may be used to improve forecasts of route choice fractions under different circumstances, e.g. [8].

Another challenge is that individual travel advice is sometimes not aligned with the objective of a traffic manager. For instance, a navigation tool may advise a traveler to drive through the center, while the objective of road authorities may be to reduce through traffic in the city center. It is therefore important to align the objectives of travelers and traffic managers. Through intelligent communication devices, road authorities may communicate their control strategies to individual travelers. Not only will this help travelers to anticipate certain measures, but if travelers understand the reasons behind certain control strategies, they might also follow advice that enhance these strategies. Such a win-win situation will only be successful if road authorities are informed about the needs and expectations of travelers. Hence, some knowledge of underlying travel behavior appears to be crucial for providing tailor made traffic control to different types of users.

In this paper, we will touch some of these issues. In section 2, we present some results of a route choice study based on a license plate study in the city of Enschede. The results could be used as input for specific control strategies in Enschede. In section 3, we describe a simple method to estimate travel times from detection loops at signalized intersections in Enschede. Together with volumes, these travel times describe the traffic status on the network. In section 4, we describe how we can use historic data to predict volumes and travel times, which enable traffic managers to anticipate on certain bottlenecks. In section 5, we conclude with a short discussion.

2 Route Choice from a License Plate Survey

Route choice plays an important role in predicting traffic flows. For given origin (O) and destination (D) pairs, route choice behavior determines how trips are distributed over the network. Hence, description of route choice behavior is essential in estimating traffic loads. Initially, travelers were assumed to choose the shortest travel time route, e.g. [9]. In general, both travel advice and traffic control are often still based on this assumption. However, many other attributes are found to be important, such as for example directness and number of intersections and turns, e.g. [10], [11].

The influence of these many different attributes on route choice can be evaluated by all kinds of observations. A rather indirect way to calibrate a route choice model is with the help of aggregate, instantaneous data like traffic volumes and travel times. It is not trivial, however, how individual preferences match with aggregate states of the transport system and many combinations of attribute values in the model might lead to link volumes that are reasonably comparable to observed volumes.

Most authors therefore prefer to gather observations from which route choice can be derived in a more direct way. Revealed preference techniques are probably most suited for this task, because they measure the actual choices of participants. Like stated preference, many of these revealed preference studies have been carried out by questionnaires, e.g. [12], [13]. Questionnaires enable the researcher to study individual preferences in detail, but the small samples are often not representative for the whole population. Routes can also be observed by floating car data, e.g. by GPS tracking, e.g. [14], [15], [16]. GPS tracking is in some sense complementary to questionnaires. Although fewer individuals are usually in the sample, partly due to privacy restrictions, individuals can be followed over longer time periods which enables researchers to describe dynamic aspects of route choice behavior. In addition, observed routes are often represented by a unique path in the network. However, like in questionnaires, samples are quite specified, i.e., aimed at a small group of individuals and hence a small set of arbitrary OD pairs.

Roadside systems can also be used to estimate route choice behavior. We used a license plate survey in the municipality of Enschede to study route choice. Although license plate surveys have been used for this purpose before, e.g. [17], this remains quite rare. License plate surveys are however complementary to questionnaires and GPS data. A license plate survey does not provide the explanatory power from a questionnaire or the details from GPS data. However, it provides a complete dataset with which average route choice behavior of many users and different types of OD pairs can be estimated.

2.1 License Plate Survey

The city of Enschede has about 130.000 inhabitants. Although the city is rather small and compact in an international context, it can be considered a large city (13th in the Netherlands) in the Dutch context. The monitoring stations were positioned in concentric circle cordons, covering all main roads. We distinguished main roads from residential streets. A main road has a speed limit of 50 km/h or larger. Residential streets have a speed limit of 30 km/h. Car license plates and their time passages were registered during the off peak and evening rush hour on a Tuesday, and on a Saturday afternoon. These time slots represent periods with different traffic situations, e.g., on Saturdays a significant fraction of traffic consists of shoppers visiting the city center. Within each period the traffic situation is quite stable. The registrations were carried out by human observers.

Figure 1 illustrates the cordons around the central part of the city. The stations with the labels STK are on the ring road, and the stations with labels CK and CTK are on roads through the city center. The network in the figure consists of all main roads.

Trips were defined as a sequence of measurements of the same license plate. We used certain criterion to split sequences in separate trips when it was quite likely a driver actually had multiple destinations. However, due to significant variations in travel times, e.g. due to the unpredictability of traffic control phases encountered during a trip, it is quite difficult to distinguish multiple trips when the time at an intermediate destination is short. Actually, this problem also plays an important role in GPS data from which modelers want to identify different destinations and trip purposes. In some cases, it will remain difficult to distinguish between a stop at a destination and a traffic related stop. Although we do not know the exact number of these cases, they probably only constitute at most a few percent of all trips. For general trends, this issue is therefore not very relevant. However, it may be important for modeling route choice probability distributions. Probabilities for long routes may then be overestimated, because some of them actually constitute multiple trips.

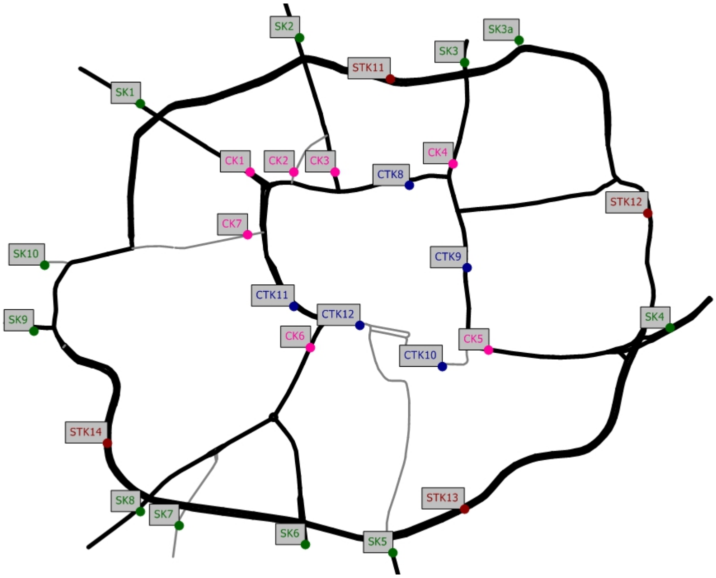

Fig. 1. Network of main roads and monitoring stations (labels) for the central part of Enschede.

Because travel times between consecutive stations are quite variable, we used the average travel time rather than the individual travel times. For each pair of consecutive stations, we estimated the average travel time and standard deviation of all observed trips containing that pair, irrespective of origin, destination or route. Due to

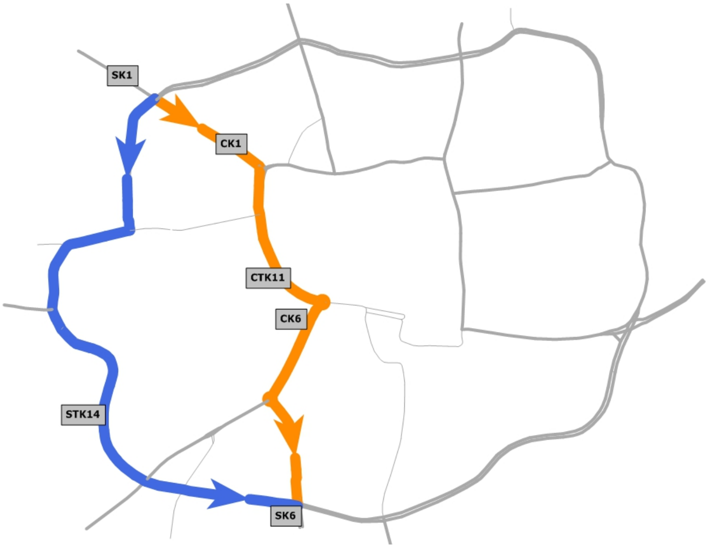

Fig. 2. Observed routes for sequences between SK1 and SK6.

the large number of observations, average travel times are quite accurate. By adding the average travel times between all consecutive stations in a sequence, the travel time of the corresponding sequence was obtained.

For some individual trips, deviating travel times are the result of something atypical going on, for example, a driver who makes a stop at a petrol station or just an error in the registration process. These so-called outliers were identified when the travel time between two stations deviated more than 3 times the standard deviation from the average travel time. As a result about 10% of trips were removed as outlier. However, as suggested in the previous paragraph, some "bona fide" trips were probably also removed, while short activities in between two trips, like collecting or delivering someone or something, may not have been filtered out.

Figure 2 shows an example of the trips that were observed between station SK1 and SK6. The blue route is along the ring road, and the orange route through the city center.

2.2 Route Set

Route assignment models use either implicit or explicit path generation. In explicit path generation, the generation of the choice set and the assignment of choices (by choice probabilities) are clearly distinct processes. Several authors (see for a review, [18]) have argued that both should be done independently, because choice probabilities may be affected by choice set size and composition.

However, even when paths are generated explicitly in advance, the alternatives will always be biased in some way, because the modeler determines which "logical" paths, i.e. those likely to be chosen, will be selected. For instance, large detours may be perceived as illogical, but it is not certain whether they may not occur in some cases. It is the purpose of route choice studies to reveal how likely both logical and illogical routes are chosen in practice. This bias of using “a priori knowledge” may even be enhanced when requirements are set to include used (observed) paths in the route set, while equally valid unobserved paths may be left out. In such a bias against “unseen” data, the occurrences of observed events are over estimated.

The aforementioned issues are in particular relevant for our survey. We do not observe paths. Instead, our so called observed routes only consist of sequences of monitoring stations. In between these locations, multiple paths are still possible. Fortunately, this problem can be dealt with, because the following is true for this survey. First, most drivers take main roads. The use of residential streets is quite rare. Secondly, between consecutive stations, there is often one unique path of main roads. If there are multiple paths, these can be considered as overlapping, which means they are actually considered as the same by the driver. Thirdly, different sequences are considered to be non-overlapping, because they contain different key intersections, which make the corresponding routes quite distinct. Hence, each observed route (sequence of links) can be represented by one unique underlying path.

Since link sequences represent unique paths, our route set only consists of link sequences. We do not need to know the paths, and therefore do not need a network of roads. This will simplify the route set generation enormously, but also has the advantage that observed routes can mapped one-on-one to routes in the route set. Another advantage is that we almost do not need to make assumptions about the plausibility of certain routes in advance. Thus, almost all possible routes are selected, including unobserved routes. This means that we are careful not to introduce biases against “unseen” or “illogical” routes.

We used a simple route set method in which new routes, i.e. sequences, were created by appending stations to the last station of the previous sequence. Each sequence only contained pairs of consecutive stations which also appeared in observed sequences. Also, (sub) circular routes were not allowed, and a time limit relative to the shortest time route was used to stop the creation of very long alternatives. Yet, with these restrictions still about 80 alternative routes per OD pair were found, which is much more than in other route set generation methods.

2.3 Main Results

We found that most drivers use the shortest time route. However, 25% of the trips did not use the shortest time route, but a detour route. This is quite a significant fraction, but it is smaller than typically found in the literature, e.g. [13], [14], [15]. In those studies, typically fewer than 50% of the drivers take the shortest time route. The difference may be related to the relatively “low resolution” of our data, which does not allow us to distinguish different, but resembling, paths. On the other hand, the samples sizes of participants and corresponding OD pairs were relatively small (in order of magnitude of 100) in the aforementioned studies. Moreover, in most studies the samples only contained university staff members and / or commuters. Even if we would consider the samples in these studies to be representative, they are far from complete. It should be noted, however, that while our sample may be complete for the city of Enschede, it only contains city trips and no highway trips. It is therefore not unlikely that the sample differences may be responsible for differences in results.

For a subsample of OD pairs, which cut the city center, we found that even more drivers, i.e. 88%, took the shortest time route when the ring road was faster than the route through the center. However, we found that only 14% took the shortest time route when the ring road was slower than the route through the center. Travelers thus preferred the route along the ring road even if it was not the fastest route. For this particular network constellation, road hierarchy appears to be crucial, and just as important as travel time. We also conclude that it is only useful to compare route choice fractions over shortest time routes when the context is the same.

Whereas route frequencies are quite sensitive to the characteristics of OD pairs, overall detour times are much more robust. We found that the average detour time is about 8% of the average travel time over the shortest time route. This results is quite comparable with the literature, e.g. [14], [16]. This can be explained by the fact that for OD pairs with only long alternative routes, a relatively small percentage of trips over these alternatives will be offset by larger detour times for these alternatives.

3 Travel Time Estimates for Signalized Intersections

The municipality of Enschede has the objective to reduce the number of vehicle kilometers by 5%. Consequently, the accessibility will be improved. For this purpose, travelers may be given incentives to change their behavior. This can for example already be done by confronting them with their "bad" travel behavior in comparison with that of other travelers. Another way is to inform travelers about the traffic status, and provide them with alternative modes or routes, which may yield more sustainable trips.

In the latter case, travel times and volumes in the network should be estimated. Travel times can be observed by floating car data, e.g. by GPS tracking. Roadside systems can also be used to estimate travel times. For example, Bluetooth data may provide quite reliable travel time estimates, e.g. [19]. We used data from inductive loops near all important signalized intersections in Enschede to estimate travel times or delays. These data were obtained from January 2010 till June 2011. Although travel times can only be estimated in an indirect way by inductive loops, they provide a complete dataset of all drivers who passed these intersections. They therefore also provide volumes, which are typically not observed from GPS data alone. The advantage of volume measurements is that they enable an assessment on why travel times deviate. In some cases, travel times increase, for example due to bad weather, while volumes remain constant or even decrease. In other cases, for example at the start of a normal rush hour, travel times increase due to increasing road volumes. If an explicit distinction can be made between capacity and demand related travel time deviations, this may yield more reliable short term travel time predictions.

3.1 Loop Data

A signalized intersection consists of 4 to 12 signal groups, indicated by the traffic light symbols in figure 3. The figure shows that a signal group consists of 1 to 5 inductive loops. The loops associated to a signal group all have particular functions. The loop closest to the stop line (stop line loop) is mostly used to detect vehicles at the stop line and to estimate the volume of vehicle passing through. The second loop (long loop, if present) is generally situated 10 to 15 meters upwards from the stop line and is used to detect the first hints of a queue. The other loops (distant loops) are used to detect approaching vehicles. These loops can also be used to count the inflow of vehicles at a particular arm of the intersection. Most inductive loops are connected to one signal group, but distant loops can sometimes be associated with more than one signal group.

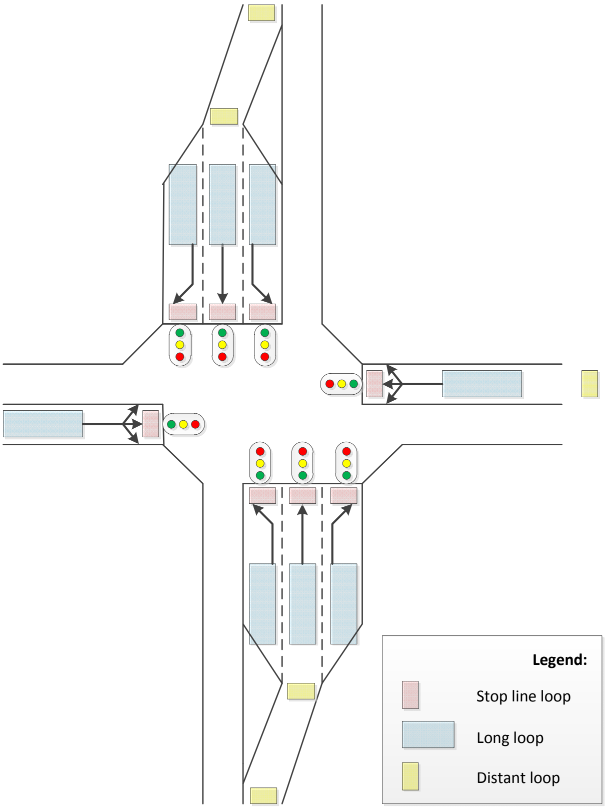

Fig. 3. Example of an intersection and its inductive loops.

The data were aggregated in 5-minute intervals. For different types of loops the data were further aggregated to the signal group level. Although we plan to use all the data, especially for estimating delays of oversaturated flows, so far we only used information from stop line loops. Flow volumes were estimated by adding vehicle counts from stop line loops belonging to the same signal group, and signal group occupancy rates were estimated as the weighted (by vehicle count) average over the corresponding loops. Also, the total red time and number of times the signal group switched from red to green were recorded. Finally, we also kept record of the occupancy time during red as fraction of the total time (5 minutes).

Unreliable data were flagged. This was done as follows. If the volume was 0 or the vehicle count of at least one loop exceeded a maximum limit, the observation was flagged (flag = 0) as unreliable. The limit was set such that the time headway between two vehicles during green time should not be smaller than 1.5 seconds. Smaller values are unrealistic and indicate rapid and artificial variation in the loop's induction current. A volume of 0 does not necessarily have to point to a malfunctioning of the loops. In fact, in the middle of the night it is possible that no cars are passing during a five minute period. However, it is no problem to flag these "bona fide" measurements, because they are not relevant to the traffic manager, and cannot be used to estimate the travel time anyway.

We mapped the signal groups to a network of roads. This enabled us to provide route travel times by combining (free flow) travel times on road segments with the estimated travel times or delays at the signalized intersections. For intersections without measurements, we estimated the travel time by using a macroscopic traffic model.

3.2 Delay Estimation

Because delays cannot be measured directly from loop data, queuing theory is often used to estimate delays, e.g. [20]. If we assume a homogenous arrival rate of vehicles near saturation (green time is just sufficient to let all queuing vehicles through), the average delay per vehicle is half the red time of one cycle. For a random arrival rate, the delay increases somewhat, but the largest delays occur when cars arrive in clusters, i.e. the intensity during arrival is the same as the discharge intensity. In other words, all drivers will wait the same amount of time for a red traffic light. Near saturation, this would yield an average delay of the red time of one cycle.

In figure 4, we show the case in which arrival intensities are the same as discharge intensities for an arbitrary average 5 minute flow volume. The figure shows that all cars have the same waiting time. Hence, the average delay is the delay of the first car in the queue. This delay can be calculated by multiplying the occupancy rate during 'red' and the red time per cycle. Both quantities can be derived from the recorded observations.

We deviated from standard assumptions in queuing theory, because urban traffic flows are quite clustered. This may in particular be true when controls are vehicle actuated, such is the case in Enschede. Both in quiet traffic conditions and conditions near saturation, these estimations may therefore be quite accurate. For intermediate conditions the delay is probably somewhat over estimated. These simple estimates could therefore be improved by including a flow dependent factor between 0.5 and 1 for under saturated flows.

3.3 Accuracy of Delay Estimations

To test the accuracy of our delay estimates, we compared our estimates with other data sources. For the city of Enschede, a database with average speed profiles based

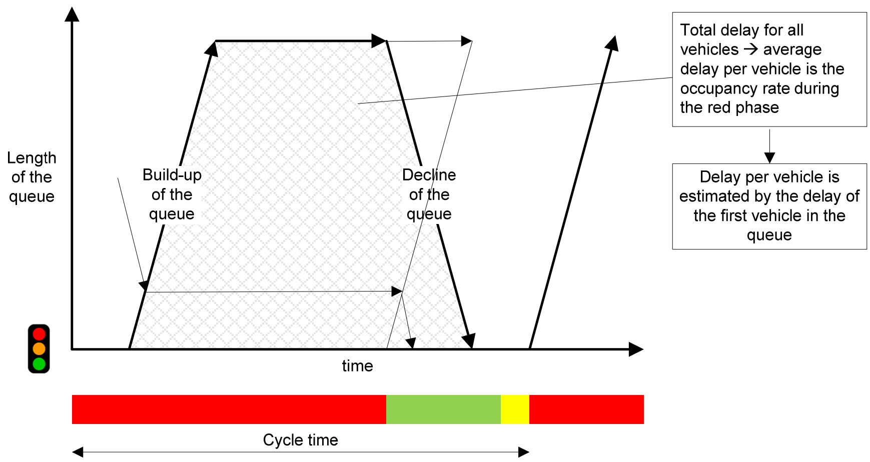

Fig. 4. Illustration of queuing when cars in under saturated flows arrive in clusters.

on floating car data (called ViaStat) is available. From this database average travel times on road sections can be extracted. We converted our delay estimates per signal group into travel times by assuming free flow travel times on the corresponding road segments. We chose the urban ring road for comparison. For both directions on the urban ring road we estimated the travel time on an average Thursday morning peak hour (between 8 and 9 AM).

In table 1, we show the comparison. The table suggests that our results are quite valid. However, we should be careful interpreting these results. The comparison consists of aggregates over many signal groups. Systematic errors in individual delays could still be present. In particular, over estimation of delays in under saturated conditions, could have been offset by an underestimation of delays in potential saturated conditions. So far we did not consider saturated traffic conditions. In a saturated condition the delays will increase dramatically when vehicles have to wait several cycle times. For the estimations of these travel times, we might use data from previous 5 minute intervals to describe the growth of the queue, and apply correction factor larger than 1 to capture the increasing delays.

Table 1. Travel times based on ViaStat and our estimates.

	Based on ViaSTAT	Our estimate
Ring road 'right'	14 min 11 sec	13 min 35 sec
Ring road 'left'	13 min 50 sec	14 min 11 sec

4 Urban Traffic Flow Predictions

One of the objectives is to generate accurate information on flow volumes and travel times. Users need to be able to plan their journey based on this information. Thus not only real-time information should be generated, but also predictions for the (near) future. Different approaches exist for predicting volumes and travel times. Extrapolation models (both spatial and temporal) are often used for short term predictions, e.g. [21], [22]. Extrapolations can give accurate predictions, but only for prediction horizons smaller than 15-20 minutes. For longer prediction horizons, volume measurements can also be matched to historical patterns. For these predictions neural networks, e.g. [23], [24], or clustering methods, e.g. [25], [26], are applied.

4.1 Base-line Prediction

Autoregressive models (e.g. ARIMA models) are common in time-series forecasts. Their forecasts are based on linear combinations of measurements from previous time-intervals. Travel demand variations are often non-linear. Several authors, e.g. [21], [22], [26], have therefore indicated that they prefer to use the average historical profile of a whole day for predictions of a future day. In this case, non-linear features may already be captured by the historical profile. Based on these results, our first prediction for day *d*, link *l* and time interval *t* is equal to the historical mean of the group to which day *d* belongs.

$$q_{dlt}^{base} = \overline{q}_{dlt} = \frac{\sum_{d' \in D} q_{d'lt}^{obs}}{n_D} \tag{1a}$$

$$D_{dlt}^{base} = \overline{D}_{dlt} = \frac{\sum_{d' \in D} D_{d'lt}^{obs}}{n_D} \tag{1b}$$

In equation (1a), we call q_{dlt}^{base} the base-line prediction, and q_{dlt}^{obs} is the measured volume on day *d*, link *l* and time interval *t*. Equation (1b) provides the base-line prediction for delay *D*. Figure 5 illustrates base-line predictions for 5 minute intervals of a particular traffic control during Mondays, Thursdays and Saturdays. The upper panel shows significant day-to-day variation. The characteristic morning rush hour peak is absent during Saturdays, while Thursdays show extra traffic in the evening due to extra shopping hours. The lower panel shows the average delay for these weekdays. The figure nicely illustrates that during busy periods, travel times also increase.

4.2 Short Term Predictions

When large events take place traffic flows may be influenced by the visitors of these events. At certain locations this will lead to a significant increase of traffic just before the event has started and after the event has finished. These events should be taken into account, implicitly by pattern recognition or explicitly by using historical data of events and knowledge about the occurrence of a new event. However, base-line predictions sometimes also do not follow the measurements in a regular situation. Apparently, traffic counts show systematic variations in time, which cannot be described by the regular day-to-day variation alone. Such variations can have different causes,

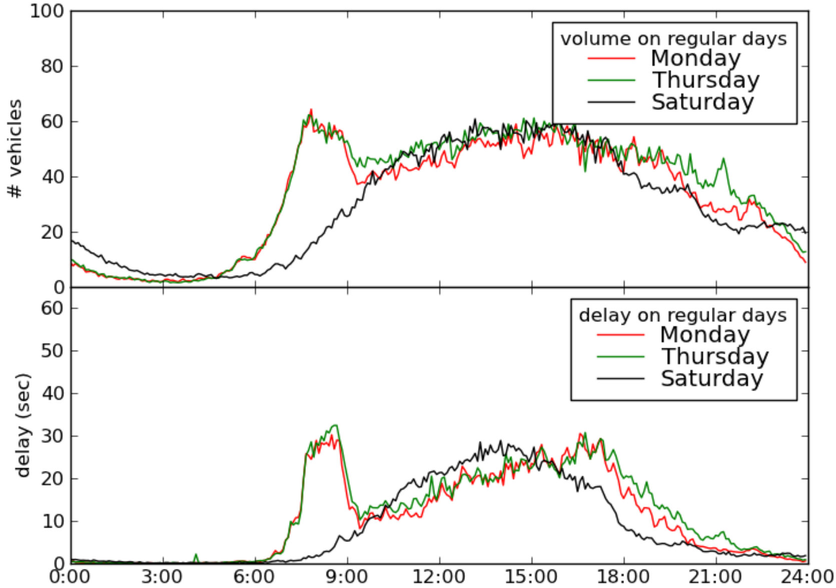

Fig. 5. Volumes (upper panel) and delays (lower panel) for a specific intersection and per 5 minute interval averaged over all Mondays, Thursdays and Saturdays.

like for example seasonal effects, changing weather conditions or road works.

From a visual inspection of time-series, we suspect that a large amount of the variation between successive time-intervals is random. This variation is called noise. The amount of noise is an important quantity. If the amount of noise increases, systematic variations can be detected less easily. It also gives an under limit for the predictive power, because noise cannot be predicted. Noise can have different causes. It can be caused by the random arrival process of cars. This process results in different headways between following cars, which is an important source of variation on highways. In urban areas traffic flows are interrupted by traffic signals. However, if green times of these signals are unknown, they may even contribute to the noise. In practice, all variations which have short time-scales and which do not follow a recurrent pattern can be considered as noise.

A measurement of quantity *x* on day *d*, at link *l* and in time interval *t* can thus be described in the following terms:

$$x_{dlt}^{obs} = x_{dlt}^{pred} + \varepsilon_{dlt} + \nu_{dlt} \tag{2}$$

with x_{dlt}^{pred} the prediction (e.g. the base-line prediction), ε_{dlt} the systematic variation between measurement and prediction, i.e. the prediction error and ν_{dlt} the noise on day *d*, at link *l* and in time interval *t*.

The objective is to develop a prediction scheme that minimizes the systematic variation or prediction error. There are two extreme approaches to reach this objective. First, the external processes that lead to systematic variations can be studied in detail, so that the relation between the two can be modeled (e.g. the relation between weather and travel demand). The advantage of this approach is that it provides insight in the variation of travel demand. The disadvantage is that it is complicated and requires many reliable data sources, which are often not available. Another approach is a black-box approach. In this approach correlations in historical data are found by certain mathematical techniques (e.g. neural networks, pattern matching) and these correlations are used in the prediction scheme.

We applied an intermediate approach. Our method is based on the following assumption. The single most important temporal correlation in the systematic variation is that between *successive* epochs (which can have different time-scales), i.e. there is a positive correlation between the systematic variation ε_{dlt} and $\varepsilon_{dlt+\Delta t}$. For example, due to seasonal effects, we assume that if there is more traffic than average on a particular day, than the probability is high that the next day will also show more traffic. The improvement of the base-line prediction is quite simple in this case. The relative systematic variation c (with $\varepsilon = cq^{pred}$) results from the ratio between the observation and base-line prediction. Suppose that this ratio is 1.10, i.e. c is estimated to be 10%. Dependent on the strength of the correlation between the relative systematic variation of successive time intervals, the updated prediction then lies between 1.00 (in case of no correlation) and 1.10 (in case the correlation coefficient is 1) times the base-line prediction.

Hence, the short term predictions up to 24h ahead can be described as:

$$q^{pred}_{dlt+\Delta t} = \overline{q}_{dlt+\Delta t} * \left(\frac{q_{dlt}}{\overline{q}_{dlt}} \right)^{\beta_q} \tag{3a}$$

$$D^{pred}_{dlt+\Delta t} = \overline{D}_{dlt+\Delta t} * \left(\frac{D_{dlt}}{\overline{D}_{dlt}} \right)^{\beta_D} \tag{3b}$$

In which the base-line prediction at $t + \Delta t$ is updated by the ratio between real-time observations and base-line predictions at t. We used hourly volumes to estimate this ratio. In doing so, we reduced the influence of noise, while still capturing most of the changes in the systematic variation. As mentioned, the update factor β depends on the strength of the correlation between the relative systematic variation at t and $t + \Delta t$. In Figure 6, we show the relative systematic variation of successive days for volumes (upper panel) and delays (lower panel) for a particular traffic control. We did this for each hour of the day between 6.00 and 22.00h. The figure shows positive correlations (with correlation coefficient 0.79 for volumes and 0.68 for delays). The correlation is less strong for the delay, because of the larger amount of noise in the delay estimations. These correlations yield βs of around 0.7.

Besides seasonal influences, which can be used to make predictions on a longer timeframe (i.e. one to two days ahead), real-time data may be used to improve predictions for shorter time horizon. We estimated the correlation between systematic variation in successive hours, and found these to be comparable to that of successive days.

These are still preliminary results. Due to the high noise level, in especially delay estimates, systematic variations are not easily resolved. As a result, at the moment short term predictions are only slightly better than base-line predictions. However, short term predictions can be improved when the noise in the observations are filtered [27]. Without a noise filter, the quality of the prediction quickly reaches an upper limit set by the amount of noise in the observations.

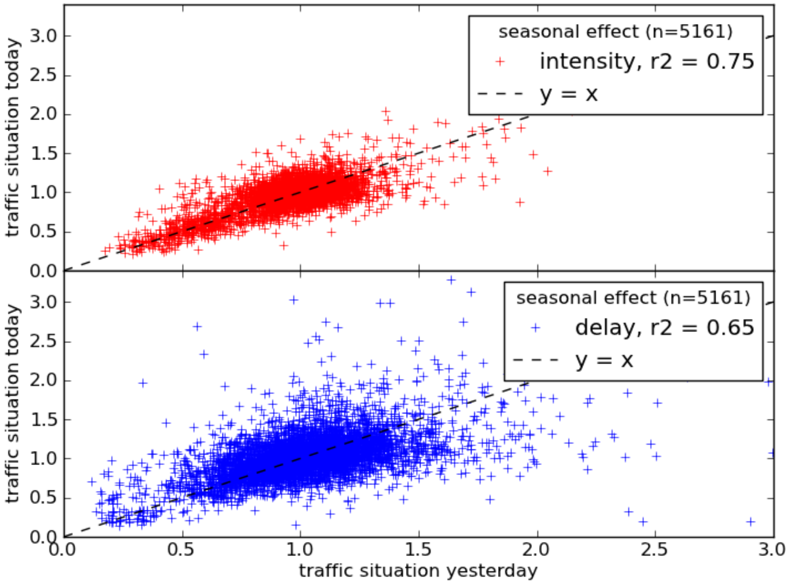

Fig. 6. Correlation between relative systematic variations of successive days for volumes (upper panel) and delays (lower panel).

5 Discussion

Traffic management has developed significantly in the last few decades. However, especially for urban areas, future developments are required.

The first development concerns monitoring. Network wide measures can only be effective if the traffic situation is monitored throughout the whole network. This poses a challenge. Roadside sensors do not always cover the whole network. For example, in sections 3 and 4 we used data from detection loops to monitor the traffic situation in the Dutch city of Enschede. Because some (important) intersections are not equipped with sensors, our estimates and forecasts are incomplete. Ideally, roadside measurements could be supplemented by floating car data, e.g. by GPS, from individual travellers. GPS data alone are probably not sufficient either, because traffic is quite dispersed in urban networks, which implies GPS data from a high fraction of motorists are needed. A fusion between the different data types might be the solution, and could be one of the challenges in a new project.

Floating car data may also be used to improve travel time estimates throughout the network. We used a quite simple algorithm to estimate delays at signalized intersections using occupation rates and red times. However, in particular for saturated condition, delays are probably underestimated by this simple method. We therefore need to improve the travel time algorithm. In addition, car floating data can be useful for validation purposes.

Monitoring in itself is not sufficient. Predictions about the traffic situation enable controllers to anticipate on the (near) future, such that timely measures to control the amount of traffic in one part of the network may prevent bottlenecks in other parts. In section 4, we used a simple prediction algorithm to forecast volumes and travel times in the near future. The predictions can however be improved. First, noise filters could reduce the noise in the observations and as a result yield better predictions. Secondly, so far we only considered single time series. However, volumes and delays may be spatially correlated. We might improve the predictions by including spatial correlations between sensors, but this will only be effective when most adjacent intersection are equipped with sensors. Unfortunately, this is not always the case. Finally, we focus on slow changing systematic variations in traffic. However, for managers, sudden changes due to events or incidents are often more relevant. The related traffic situations may in those cases be better predicted by pattern matching algorithms. It may be possible to develop a hybrid method, which combines pattern matching and forecasts of slow changing variations.

The second development concerns personalized traffic control. Predictions are not only useful for traffic control managers, but also for road users. If they are informed about future traffic flows and control strategies, they may make choices that enhance those strategies. This will only be possible if personal needs and expectations of road users are taken into account by traffic control strategies.

For this purpose, we need to explore travel behavior. We used a license plate survey to study route choice in the municipality of Enschede. We found that most drivers use the shortest time route, but that 25% of the trips did not use the shortest time route. Moreover, we found that most travellers took the ring road instead of the route through the city centre, even when the ring road was slower. This result suggest that travellers do not need to take the fastest route, but that they may prefer larger roads (higher up in the hierarchy). Many traffic managers would like to see drivers to take the ring road instead of small roads through the city center. These results suggest that strategies to increase the use of ring roads are probably acceptable to users and may also be quite successful.

References

1. Hasberg, P., Serwill, D.: Stadtinfoköln – a global mobility information system for the Cologne area, 7th World Congress on Intelligent Transport Systems, Turin, Italy (2000)
2. Kellerman, A., Schmid, A.: Mobinet: Intermodal traffic management in Munich –control centre development, 7th World Congress on Intelligent Transport Systems, Turin, Italy (2000)
3. Leitsch, B.: A Public-privat partnership for mobility – Traffic management Center Berlin, 9th World Congress on Intelligent Transport Systems, Chicago (2002)
4. Nagatani, T.: Vehicular traffic through a sequence of green-wave lights, Physica A: Statistical Mechanics and its Applications, Vol. 380, (2007) 503-511
5. Vrancken, J., Van Schuppen, J.H., Dos Santos Soares, M., Ottenhof, F.: A HierarchicalNetwork Model for Road Traffic Control, Proceedings of the 2009 IEEE International Conference on Networking, Sensing and Control, Okyama, Japan, (2009) 340-344
6. Hodge, V.J., Krishnan, R., Jackson T., Austin, J., Polak, J.: Short Term Traffic Prediction Using a Binary Neural Network, 43rd Annual UTSG Conference, Open University, Milton Keynes, UK (2011)

7. Wismans, L.J.J, Van Berkum, E.C., Bliemer, M.C.J.: Comparison of Evolutionary Multi Objective Algorithms for the Dynamic Network Design Problem. ICNSC – IEEE conference, Delft (2011)
8. Mitsakis, E., Salanova, J.M., Giannopoulos, G.: Combined dynamic traffic assignment and urban traffic control models, Procedia - Social and Behavioral Sciences, Vol. 20, (2011) 427 – 436
9. Wardrop, J. G.: Some theoretical aspects of road traffic research, Proceedings, Institute of Civil Engineers, PART II- 1, (1952) 325-378.
10. Bar-Gera, H., Mirchandani, P., Fan, W.: Evaluating the assumption of independent turning probabilities, Transportation Research part B, Vol. 40, (2006) 903 - 916
11. Chen, T.Y., Chang, H.L., Tzeng, G.H.: Using a weight-assesing model to identify route choice criteria and information effects, Transportation Research Part A, Vol. 35, (2001) 197 – 224
12. Mahmassani, H.S., Jou, R-C.: Transferring insights into commuter behavior dynamics from laboratory experiments to field trials. Transportation Research Vol. 34A(4), (2000) 243-260
13. Prato, C. G., Bekhor, S.: Applying Branch-and-Bound Technique to route choice set generation. Transportation Research Record, Vol. 1985, (2006) 19-82
14. Jan, O., Horowitz, A. J., Peng, Z. R.: Using global positioning system data to understand variations in path choice. Transportation Research Record, Vol. 1725, (2000) 37 – 44.
15. Zhu, S., Levinson, D.: Do people use the shortest path? An empirical test of Wardrop's first principle, 4th International Symposium on Transportation Network Reliability, July Minneapolis, USA (2010)
16. Papinski, D., Scott, D. M.: A GIS Toolkit for route choice analysis, Journal of Transport Geography, Vol. 19, (2009) 434 – 442
17. Hamerslag R.: Investigation into factors affecting the route choice in "Rijnstreek-West" with the aid of a disaggregate logit model. Transportation, Vol. 10, (1981) 373 – 391.
18. Bovy, P. H. L.: On modeling route choice sets in transportation networks: a synthesis, Transport Reviews, Vol. 29 (1), (2009) 43 – 68
19. Tarnoff, P. J., Wasson, J. S, Young, S. E., Ganig, N., Bullock, D. M., Sturdevant, J. R.: The Continuing Evolution of Travel Time Data Information Collection and Processing, Transportation Research Board Annual Meeting, Paper ID: 09-2030 (2009)
20. Mak, W. K., Viti, F., Hoogendoorn, S. P., Hegyi, A.: Online travel time estimation in urban areas using the occupancy of long loop detectors, 12th IFAC symposium on transportation systems, Redondo Beach (2009)
21. Wild, D.: Short-term forecasting based on a transformation and classification of traffic volume time series, International Journal of Forecasting, Vol. 13, (1997) 63 – 72
22. Van Grol, R., Inaudi, D., Kroes, E.: On-line traffic condition forecasting using on-line measurements and a historical database, 7th World Congress on Intelligent Transport Systems, Turin, Italy (2000)
23. Dia, H.: An object-oriented neural network approach to short-term traffic forecasting," European Journal of Operational Research, Vol. 131, (2001) 253 – 261
24. Yin, H., Wong, S. C., Xu, J., Wong, C. K.: Urban traffic flow prediction using a fuzzy-neural approach, Transportation Research Part C, Vol. 10, (2002) 85 – 98
25. Chung E.: Classification of traffic pattern, 10th World Congress on Intelligent Transport Systems, Madrid (2003)
26. Weijermars, W. A. M,: Analysis of urban traffic patterns using clustering, Ph.D. Thesis, University of Twente, Enschede, The Netherlands (2007)
27. Thomas, T., Weijermars, W. A. M, Van Berkum E.C.: Predictions of Urban Volumes in Single Time Series, IEEE Transactions on Intelligent Transportation Systems, Vol. 11 (1), (2010) 71 – 80

User Modelling and Emotion Recognition of Drivers through a Multi-modal GPS Interface

Maria Virvou

Department of Informatics, University of Piraeus,
80 Karaoli & Dimitriou St., Piraeus 18534
mvirvou@unipi.gr

Abstract. Drivers play an important role on road traffic. Traffic frequently has a big impact on drivers' emotions and drivers' emotions have a big impact on traffic. Traffic congestions may be the cause of human drivers' frustration, loss of their patience and control, leading to aggressiveness and so on. On the other hand, drivers' aggressiveness may cause dangerous driving, car accidents and drivers' fighting. This results in an endless loop of traffic problems that is propagating along many drivers. However, even excessive enjoyment may also lead to dangerous driving, since people may underestimate the road dangers and drive carelessly. Thus, it is very important to aim at keeping drivers calm, happy and alert when they drive. In view of this, it would be extremely useful to extend the functionalities of existing GPSs to include user modelling and emotion recognition abilities so that they may provide spontaneous assistance that would be dynamically generated based on the results of the user modelling and emotion recognition module. The action of GPSs would be to provide automatic recommendation to drivers that would be compatible with their own preferences concerning alternative routes and make them feel happier and calmer.

Keywords. User modelling, Traffic, Emotion recognition, Affective computing, Multi-modal GPS.

1 Introduction

With more and more people in the world and in the workforce, roads are becoming increasingly crowded; when we're all frustrated with traffic, sometimes people make mistakes or pull impolite driving maneuvers, which can lead to anger from other frustrated drivers; this often results in road rage, which can pose a significant threat to health and safety for everyone on the road [1].

Counseling psychologist Jerry Deffenbacher and his colleagues [2] point out that: "Those high-anger drivers are a source of alarm. Even typically calm, reasonable people can sometimes turn into warriors behind the wheel; when provoked, they yell obscenities, wildly gesture, honk and swerve in and out of traffic, and may endanger their lives and others."

In the official site of the city of Santa Rosa at the Section of Traffic [3] there is a question: Which is an example of aggressive driving?

According to the same site, among other things the above question includes the following examples:
Speeding up to make it through a yellow light.
Switching a lane without signaling first.
Going over the speed limit in a school zone or neighborhood.
Approaching so fast that the driver of another car that is stopped, feels threatened.
Tailgating a car to pressure the driver to go faster or move over.
Tailgating a car to punish the driver for something.
Driving with an alcohol level above the legal limit.
Drive while drowsy enough to have droopy eyes.
Making an obscene gesture at another road user.

Moreover, the Official U.S. Government site for distracted driving [4] warns people, that: “Distracted driving is unsafe, irresponsible and in a split second, its consequences can be devastating.” On the other hand, research and experience demonstrate that happy drivers are better drivers [5].

In view of the above, it seems that human emotions play a very important role to traffic management and it is to the benefit of traffic to put research energy on recognizing automatically human emotions of drivers and building systems that would react accordingly. In view of this, it would be extremely useful to extend the functionalities of existing GPSs to include user modelling and emotion recognition abilities so that they may provide spontaneous assistance that would be dynamically generated based on the results of the user modelling and emotion recognition module. The action of GPSs would be to provide automatic personalized recommendation to drivers that would be compatible with their own preferences concerning alternative routes and make them feel happier and calmer.

The main body of this paper is organized as follows: In Section 2, related work on research of ourselves and others is surveyed and discussed. In Section 3, the aims of the proposed research is presented. In Section 4, the proposed solution is presented. Finally in Section 5 the conclusions of this paper are drawn and also connections to proposals of other participants are highlighted.

2 Related Work

Affective Computing is a recent area of Computer Science that studies human emotions:

- Emotion Recognition by the computer
- Emotion Generation from the computer

Until recently, human emotions were not considered at all by the designers of user interfaces. However, research that flourished during the past decade has been based on the important argument that human feelings play an important role on human decision making and affect all areas of human computer interaction. There has been a lot of research on automatically recognizing human feelings and generating emotions from the computing. This kind of research is labeled affective computing. So far, there has been significant progress in this field. Nevertheless, there is still a lot of basic research needed and thus affective computing remains a hot research topic.

Another area that has been investigated by many researchers during the past decade is that of user modeling and generation of personalized recommendations to computer users. Recommender systems constitute an area of research that attracts researchers from a wide area of computer science and applications varying from e-commerce to electronic libraries.

2.1 Previous Work on Affective Computing and GPS Recommender Systems in our Lab

In our own research lab we have made extensive research in the areas of affective computing and recommender systems [e.g. 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16].

This research has resulted in four Ph.D.s that were completed successfully, two monographs and many papers in the area.

One recent relevant basic research project that we had was entitled:

"Technologies of affective human-computer interaction and application in mobile learning" and it was developed in the Software Engineering Lab of the Department of Informatics at the University of Piraeus. Our Industrial Partner was Sony Ericsson. The main research problem that was investigated was recognition and generation of emotions through multi-modal hand-held devices. In the context of this project, two Ph.Ds were supervised and completed:

1. Ph.D. student: Efthimios Alepis (Supervisor M.Virvou)
2. Ph.D. student: : Ioanna-Ourania Stathopoulo (Supervisor G. Tsihrintzis).

The research topics that we dealt with were:

- Emotion recognition through microphone and keyboard
- Emotion generation in animated agents
- Incorporation of animated agents in mobile learning
- Visual Analysis of facial expressions
- Construction of our own database of facial expressions

2.2 Previous Work of Affective Computing and Traffic

Two powerful partners — a well-funded consumer-facing company and a top research university, such as MIT – joined forces to produce inventive solutions to real-world problems. Audi of America wanted to be involved in conversations about America's urban future and provide cars that fit into tomorrow's tech-dominated cities. The carmaker wanted to encourage people to admire and buy Audi cars by giving them an online tool with information about their roadways. The MIT lab Research Laboratory of Electronics studied exactly the kind of data that the Audi idea needed. The result is the Audi Road Frustration Index (Fig. 2), an entertaining web-

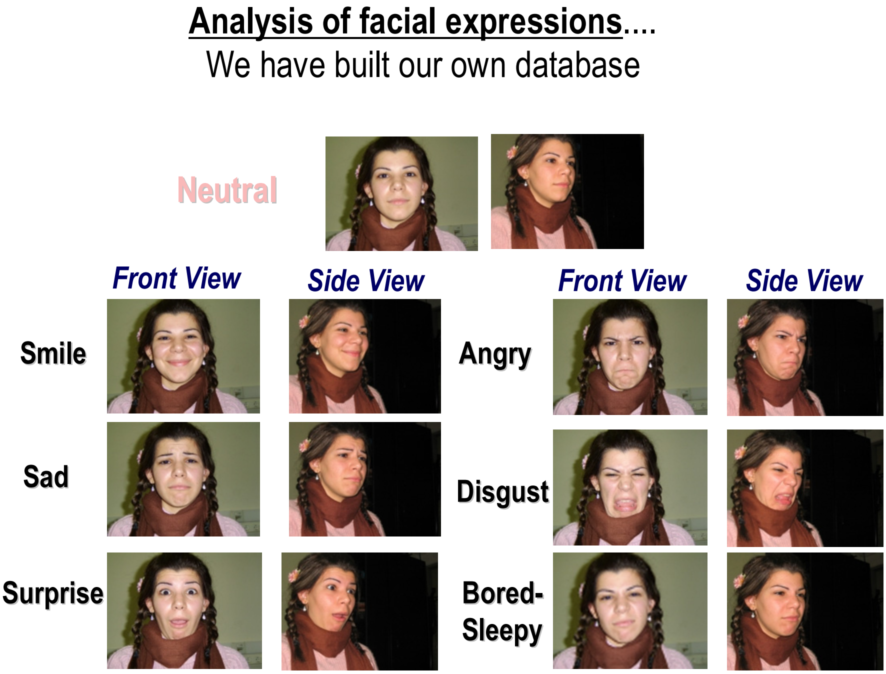

Fig. 1. Facial expressions denoting human emotions.

Fig.2. Audi's Frustration Index.

site that launched in beta in mid-September. It tells users at any given hour how the roadways and drivers' moods in their city rank compared to others nationwide. For instance, Sacramento, Calif., is often as miserable as New York City [17].

Thus, Audi's latest ad campaign for the new 2012 A6 claims to want to make the road "a more intelligent place," starting with asking drivers to pledge to be on their best behavior while behind the wheel. At the top of the German automaker's list of sins is driving while drinking a latte, leading us to perceive the effort as only half serious [18].

Another joint effort on researching the influence of emotions on drivers arose from a leading University, Stanford, and a leading car company, Toyota [19]. This effort resulted in a study that examines whether characteristics of a car voice can influence driver performance and affect.

- In a 2 (driver emotion: happy or upset) x 2 (car voice emotion: energetic vs. subdued) experimental study, participants (N=40) had emotion induced through watching one of two sets of 5-minute video clips. Participants then spent 20 minutes in a driving simulator where a voice in the car spoke 36 questions (e.g., "How do you think that the car is performing?") and comments ("My favorite part of this drive is the lighthouse.") in either an energetic or subdued voice.
- Participants were invited to interact with the car voice. When user emotion matched car voice emotion (happy/energetic and upset/subdued), drivers had fewer accidents, attended more to the road (actual and perceived), and spoke more to the car.
- To assess drivers' engagement with the voice, participants were invited to speak to the Virtual Passenger.

Other recent research efforts include Using Paralinguistic Cues in Speech to Recognise Emotions in Older Car Drivers by Christian Jones and Ing-Marie Jonsson [20]. Finally, there has been research on "Analysis of Real-World Driver's Frustration" [21], to name some very recent research projects in the area.

2.3 Conclusions from the Related Work

It seems that there is interest from World Leading Universities, such as MIT and Stanford and leading car manufacturers such as AUDI and TOYOTA respectively, to produce new affective systems for the drivers.

However, there are not yet many such research attempts. This means that there is a lot of scope in this particular research topic that seems to be gaining research interest. In this respect, our proposed approach is very innovative to the field.

3 Aims of the Research Proposed

The aims of this work package are the following:

1. Recognition of basic emotions of drivers based on visual-facial and audio-lingual recognition and contextual information.
    a. Visual facial recognition through a camera
    b. Audio lingual recognition through a microphone
    c. Recording of contextual information that contributes to change of drivers' feelings.
2. Monitoring and recording drivers' preferences with respect to traffic and making inferences leading to recommendations.

Technically, the above aims are going to be pursued using the following:

- Neural network-based and support vector machine-based classifiers for the visual facial recognition,
- User stereotypes and multi-criteria decision making theories for the audio-lingual recognition and the contributing contextual information,
- User stereotypes, machine learning algorithms and user monitoring for the acquisition of user models of drivers with respect to their needs, preferences and knowledge level of routes.
- Multi-criteria decision making for the selection of appropriate recommendations.
- Advanced multi-level recommender systems, which produce recommendations by combining a specific driver's preferences with preferences of 'similar' drivers.

4 Proposed Solution

We propose to build a user modeling module that will take into account

- Individual features of drivers such as route preferences, age, car type.
- Emotions of drivers in particular situations
- Traffic information

The driver would be monitored by a camera into the car so that image analysis of his/her face may take place. The driver will also have a microphone to interact with the system. The habits and behaviour of drivers will be analysed and recorded in a long term user model over the web.

In return, the driver will receive

1. Personalised recommendations about routes
2. Personalised advice on handling emotions of drivers in particular situations

The proposed solution will include the following:

- A navigation system which will provide location-based services with a personalized way, taking into account the preferences and the interests of each user.
- Location-Based Services are provided via Web Services
- A personalization mechanism

The term “location-based services” (LBS) is a rather recent concept that integrates geographic location with the general notion of services.
The five categories in Fig. 3 characterize what may be thought of as standard location-based services.

Fig.3. Standard location-based services.

One of the most basic characteristics of the LBS, is their potential of personalization as they know which user they are serving, under what circumstances and for what reason.

A system architecture that can be used is illustrated in Fig. 4. This architecture illustrates how user modeling can be incorporated in order to record drivers’ preferences and the use an inference engine to produce hypotheses on future preferences of users on other similar situations. Moreover, it shows how information from LBS can also be used. This information can be processed and passed to a user interface device, such as a GPS in a car. Moreover, Figure 5 illustrates three input devices, such as camera, microphone and keyboard that can be used to process information about a driver in terms of his/her emotional state. This kind of processing will be incorporated in the user modeling component as illustrated in Fig. 4.

5 Conclusions and Connection with other Research

We propose to extend the functionalities of existing GPSs to include user modelling and emotion recognition abilities so that they may provide spontaneous assistance that would be dynamically generated based on the results of the user modelling and emotion recognition module. In return, the action of GPSs would be to provide automatic recommendation to drivers that would be compatible with their own preferences concerning alternative routes and make them feel happier and calmer.

Our contribution could use information on traffic and routes that could be developed by other partners of the project such as Thomas Jackson and Tom Thomas. Relevant goals to Tom Thomas individual preferences of drivers concerning favourite routes. Also we see ourselves in the user interface (human computer interaction) as mentioned by Brahmananda Sapkota in his talk about functionalities to drivers. Finally, we can provide individualised information to the “informed driver” of Apostolos Kotsialos.

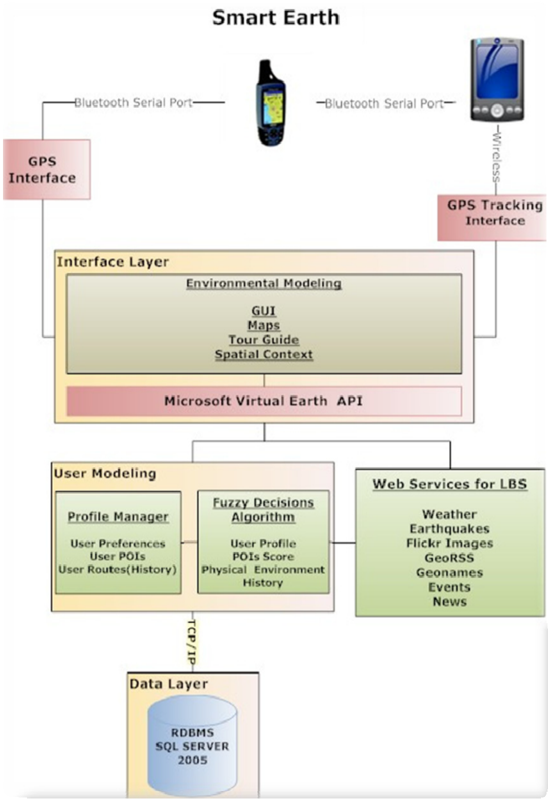

Fig. 4. A system architecture.

References

1. Elizabeth Scott, M.S. (2010), Article on Road Rage: How to Manage Road Rage to Stay Healthier and SaferAbout.com Guide, Updated November 22, 2010.
2. Deffenbacher, J.L., Filetti, L.B., Lynch, R.S., Dahlen, E.R. & Oetting E.R. (2002): Cognitive-behavioral treatment of high anger drivers. Behaviour Research and Therapy, Volume 40, Issue 8, August 2002, Pages 895-910
3. http://ci.santarosa.ca.us/departments/publicworks/traffic/streetsmarts/drivers/Pages/AggressiveDrivers
4. Official U.S. Government site for distracted driving http://www.distraction.gov/
5. Groeger, J.A. (2000): Understanding driving: Applying cognitive psychology to a complex everyday task. Philadelphia, PA: Psychology Press.
6. Alepis, E. & Virvou, M. (2012): "Object Oriented User Interfaces For Personalized Learning" Book Series: Intelligent Systems Reference Library, SPRINGER, to appear.

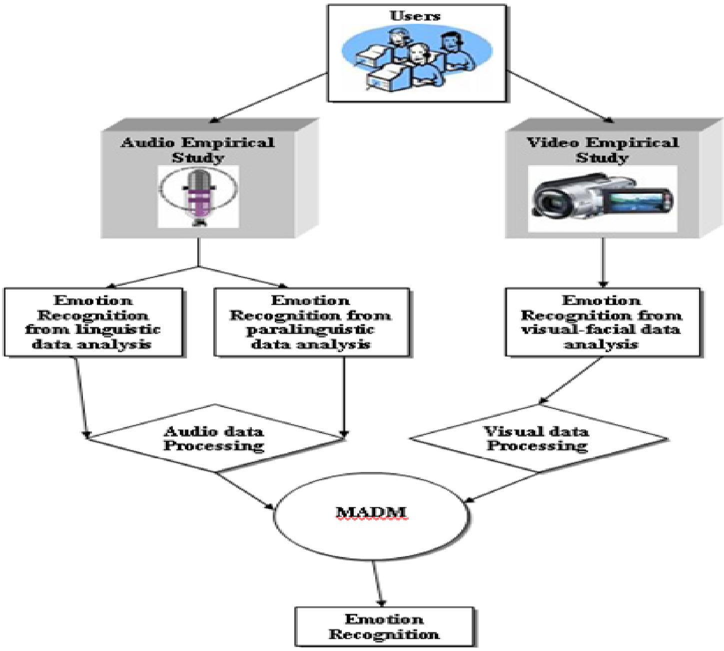

Fig. 5. Affective recognition through 2 modalities of interaction.

7. Alepis, E. & Virvou, M. (2011): Automatic generation of emotions in tutoring agents for affective e-learning in medical education. Expert Systems with Applications 38(8): 9840-9847 (2011)
8. Alepis, E., Stathopoulou, I.O., Virvou, M., Tsihrintzis, G.A. & Kabassi, K.(2010): Audio-lingual and Visual-facial Emotion Recognition: Towards a Bi-modal Interaction System. IEEE International Conference on Tools in Artificial Intelligence ICTAI (2) 2010: 274-281
9. Alepis, E., Virvou, M., & Kabassi, K. (2011): Combining Two Decision Making Theories for Affective Learning in Programming Courses. CSEDU (1) 2011: 103-109
10. Chalvantzis, C. & Virvou, M. (2008): Fuzzy Logic Decisions and Web Services for a Personalized Geographical Information System. New Directions in Intelligent Interactive Multimedia 2008: 439-450
11. Katsionis, G. , Virvou, M. (2008): Personalised e-learning through an educational virtual reality game using Web services. Multimedia Tools Appl. 39(1): 47-71 (2008)
12. Stathopoulou, I.O. & Tsihrintzis, G.A “Visual Affect Recognition” , Book Series: Frontiers in Artificial Intelligence and Applications, Volume 214, 2010
13. Stathopoulou, I.O., Alepis, E., Tsihrintzis, G.A & Virvou, M (2010): On assisting a visual-facial affect recognition system with keyboard-stroke pattern information. Knowledge - Based Systems 23(4): 350-356 (2010)
14. Stathopoulou, I.O., Alepis, E., Tsihrintzis, G.A. & Virvou, M. (2009) On Assisting a Visual-Facial Affect Recognition System with Keyboard-Stroke Pattern Information. SGAI Conf. 2009: 451-463
15. Virvou, M., Tsihrintzis, G.A., Alepis, E. & Stathopoulou, I.O (2008).: Designing a multi-modal affective knowledge-based user interface: combining empirical studies. Joint Conference on Knowledge Based Software Engineering JCKBSE 2008: 250-259

16. Virvou, M. & Katsionis, G. 2008. “On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE” COMPUTERS & EDUCATION, Vol. 50 (1) (Jan. 2008), pp. 154–178, Elsevier Science.
17. Voight, J. (2011) Salt Lake City Roads Worse Than New York's? Road Frustration Index Says Yes. CNBC.com, http://www.cnbc.com/id/45150075/Salt_Lake_City_Roads_Worse_Than_New_York_s_Road_Frustration_Index_Says_Yes
18. Autoblog: http://www.autoblog.com/2011/10/07/audis-road-frustration-index-plugs-new-a6/
19. Nass, C., Jonsson, I.M., Harris, H., Reaves, B., Endo, J., Brave, S. & Takayama, L. (2005) Improving automotive safety by pairing driver emotion and car voice emotion, CHI '05, April 2005, Publisher: ACM
20. Jones, C. & Jonsson I.E. (2008) Using Paralinguistic Cues in Speech to Recognise Emotions in Older Car Drivers. AFFECT AND EMOTION IN HCI, LNCS 2008, Vol. 4868
21. Malta, L.; Miyajima, C.; Kitaoka, N.; Takeda, K. (2011). Analysis of Real-World Driver's Frustration” IEEE Transactions on Intelligent Transportation Systems, Volume: 12 Issue:1, pp: 109 - 118

The Future of Road Traffic Management: From Local to Network-wide, from Reactive to Proactive

Jos Vrancken

Delft University of Technology, Delft, The Netherlands
j.l.m.vrancken@tudelft.nl
http://tbm.tudelft.nl/index.php?id=30785

Abstract. This paper describes the current status and future of Road Traffic Management and the key challenges to achieve this future. It proposes a way forward that covers both the traffic management theory development, the development of supporting systems and the necessary standardization effort needed to make large scale deployment of Network Management feasible.

1 Introduction

Road traffic is the most important means of transport, both for people and goods, but also the most problematic one. There are no prospects that this will change fundamentally in the coming decades. On the contrary, with increasing standards of living in many countries in the world, car ownership will continue to grow in an already, at many places, oversaturated network. The heavier the use of a roads network, the more reason there is to manage traffic, even though, in dense conditions, it is also hardest to do this effectively. We can observe each day that just self-organization, although it certainly plays a important and increasing role, is not enough to prevent congestion. Traffic management is necessary for an efficient use of the network near its capacity. The majority of traffic management measures, especially the automated measures such as traffic signals and ramp metering, are stil local measures with a scope of at most several hundred meters. Local measures have obvious limitations: they have a tendency to solve a congestion problem within their own scope by shifting it to a neighbouring area. Moreover, without looking at the network context, it is hard to look into the near future, because the future of one place in a network is, to a large extent, determined by the current status of traffic elsewhere in the network. Proactive traffic management, trying to prevent problems rather than solving them, is hard when looking only locally. So, network management (NM), which is the activity of managing traffic in larger areas, is urgently needed, even more so as networks become more and more overloaded. But denser traffic is also traffic in which events at one point have much wider consequences, which means that a network with denser traffic is a more complex system and harder to manage. The more NM is needed, the harder it is. In order to understand the true nature of the problem of developing NM, one must observe that the problem contains a number of, mutually dependent, components. It's not only about traffic management, but also about system development, both control systems and decision support systems for operators, about the deployment of NM-systems in practice and about measuring the effectiveness of NM. In the following section we will describe the current status of NM, other relevant developments in the transport domain, the main reasons why development of NM is inevitably a slow process, and finally we will propose an approach to move forward in this field. The key issues that will be addressed are:

– a framework to describe the problem of control in a network
– the problem of partitioning a network in order to make it manageable
– the problem of assigning priorities to network parts in order to support control strategies
– the necessary standardization of interoperability among traffic control systems, in order to make deployment of NM systems practically feasible

2 Current Status of Network Management

Network management is growing slowly but steadily. The subject is best developed in the area of coordination of traffic lights control systems in urban areas. Research in this area dates back to the seventies and nowadays, many products are available that offer various degrees of coordination over more or less extended, urban areas ([10]). It is well developed, but it is restricted to typical urban traffic and only one kind of control measure. Integrated Corridor Management ([7]) considers virtually all kinds of control measures available but it considers only a linear kind of network, a so-called corridor, the main traffic artery in an area. An example of a system that considers mixed networks and various types of control measures, together with advanced support for managing control scenarios, is the SCM system (Scenario Coordination Module, [14], [15]) in the Amsterdam area. This system was installed in September 2010 and is now in the middle of the process of evaluation and full-scale deployment. In [5], an interesting systems-oriented approach is described that aims at integrating current traffic control systems into a network control system in Sweden. Thus we can conclude that network management does take off, but progress is slow. Mature, full-scale network management systems, covering all networks in an area and coordinating all local measures, are not yet in operation to the best of our knowledge, let apart that it has been proven that these systems have a positive effect on network performance.

3 Other Developments in Road Transport

Several other developments in Road Transport are relevant for traffic management. We mention only a few:

3.1 Cooperative Systems

Cooperative systems are about the increase in data processing and communications on the road, with connections between vehicles and between vehicles and the roadside. This is highly relevant to traffic management in several ways. First of all, this development increases the process of self-organization among vehicles. Especially at the micro level, cars can improve traffic properties by means such as adaptive or connected cruise control. In addition, through the communications facilities, drivers are much better informed about traffic conditions on their intended route. Second, the vehicles become an important source of traffic data. Suppliers of navigation equipment, such as TomTom in The Netherlands, are already using this kind of data from their connected devices, for the purposes of providing their customers with real-time traffic information. To some extent, they are influencing traffic in a way that might be called traffic management (or even network management) but a private party that depends on its customers, will necessarily manage traffic for the user optimum and not the system optimum that governmental traffic management authorities focus on. It is therefore not to be expected that this ”private” traffic management will eventually replace the traffic management authorities. Cooperative systems offer interesting capabilities to supply the necessary sensors and actuators for NM, so instead of making traffic management unnecessary, they greatly extend the facilities for this activity.

3.2 Cross-sectoral Connections

A second important development might be summarized as the increasing level of connection between the road transport system and other systems in society, both within transport and in other sectors. Examples of this are:

Multimodal Traffic. This development is in essence about connecting road transport with the public transport system (both on the road and other forms, such as trains, streetcars, etc.). This is about finding optimal trips, in terms of travel time and cost, by combining different modes of transport (private car, bus, train, shared car, ...);

The Energy Sector. Road transport is highly energy consuming and, by its very nature, strongly dependent on liquid fuels. Much research goes into the introduction of electric vehicles, and the required charging infrastructure, but convincing breakthroughs are still to come. The extra logistics needed for electric driving, the different driving properties, and the stronger emphasis on driving limited distances, multimodal trips and car sharing will certainly have an impact on the characteristics of traffic and will change traffic patterns.

Car Sharing. Car sharing is based on the fact that most cars are idle for 95% of the time, which is of course tremendously inefficient in terms of resources that went into building so many cars, the cost of ownership and the parking place occupied by so many vehicles. Car sharing is an obvious solution to this but it requires extensive information processing in order to discover nearby shared cars that are available for use, and for the administrative handling of trips such that costs are correctly attributed to the users of a shared car. Car sharing will thoroughly change the relationship between a car and a driver, as the car is no longer his or hers. It is no longer considered an extension of private space. It looks more like public transport. This in turn will pave the way towards more outside influence on driving behavior and route choice, which offers interesting leads for traffic management. Spreading traffic over the network, for instance, is a lot easier if one can influence the route choice behavior of a fair percentage of vehicles.

4 What is so Hard about Developing Network Management?

However obvious the need for NM may be, its development is slow. The blanket reason for this is the high complexity of traffic in a network. A single crossing with traffic signals can already be very complex, let apart a large network containing urban, motorway and other types of roads. A number of different factors contribute to NM's complexity, and the list below is in no way exhaustive.

Chicken-egg Dependencies. There are several chicken-egg dependencies in the problem of NM development that greatly hamper quick progress in this area. One of them is the mutual dependency between equipment and network management theory. In order to develop NM theory, we need real-life examples. Such examples require adequate equipment installed in a given area. The roadside equipment that is available was not designed for the purposes of NM. But the investment in adequate equipment is high and should be guided by insight into NM theory. A comparable dependency can be found in the education of operators. NM will strongly depend on human operators in traffic management centers, simply because humans have many capabilities that are very hard or impossible to automate. These operators have to be trained. But as long as we don't know for sure how to do NM, it remains hard to train operators. Without trained operators, we don't get the real-life examples needed for theory development. It can safely be concluded that traffic operators play a crucial role in the development of NM, a role that is still often underestimated.

Measuring Effects of NM is Hard. For those cases in which we do have some degree of NM installed in real life, it turns out that it is hard to measure its effects on network performance. Such effects are inevitably statistical effects, averages over longer periods. They apply to many different traffic states (a specific approach to NM is likely to be more effective in one state than in others), which is a second reason why we need longer periods of measurement, in order to see some differentiation in the data for different traffic states. But networks change during longer periods, for instance due to road works. And also, traffic demand patterns change during longer periods. Both are reasons that make the results of one measurement difficult to compare or to combine with other measurements at different periods in time. More on this problem can be found in [11].

Legacy Systems. Systems for NM will inevitably consist for a large part of existing, legacy control systems. As mentioned before, these systems were not designed to function within an NM system, but in addition to this, it is often hard to connect with these systems. In practice, most of the effort of installing NM systems is spent on connecting to legacy systems, as experience with the SCM system mentioned above has shown. Often these systems are owned by different manufacturers, which may be competitors of the NM system manufacturer. The willingness to cooperate is therefore not always obvious.

Sensors and Actuators. However difficult it may be to connect with roadside equipment, in many cases one will have to cope with the absence of such equipment. The activity of NM is greatly complicated by the fact that it has to be mapped onto the existing equipment. It is always difficult to install extra equipment for purposes that are essentially experimental, not only because of the cost involved, but also because of the necessary permits, the interruption it may cause to traffic, privacy objections, etc. From the point of view of NM, the existing equipment is usually scattered more or less randomly over the area under consideration.

Simulation of networks is Hard. Traffic management research heavily relies on modeling and simulating traffic. The models and simulators currently available are not very scalable. They have difficulty to cope with larger areas, already for performance reasons. Moreover, models and simulators are often purpose specific, or at least have to be configured in a purpose specific way. In order to study effects of network management realistically, calibration data is needed, but the lack of real-life NM systems makes that such data is virtually non-existent. To make things worse, traffic itself has a strongly chaotic nature, due to the fact that it is strongly influenced by its environment, which contains many chaotic, unpredictable effects, so much so that the ”regular” situation is a notion in traffic management which is often used but equally often contested. Predictions based on simulation are most needed in dense traffic, but as mentioned above, dense traffic in a network happens to be the most complex, hardest case to manage or to simulate. Dense traffic in a large area is particularly sensitive to the chaotic influences from the environment. This chaotic behavior of traffic applies not only at very short time scales, but also on larger time scales, in which, for instance, demand patterns change for reasons outside of the transport domain.

5 The Approach to Developing Network Management

Just like the original chicken-egg problem, the problem of NM development will have to be solved in a very gradual, evolutionary fashion. Essential steps and corresponding milestones in the approach are:

Network Partitioning We need a way to partition a network into parts, such that the control problem for the network as a whole, can abstract from the control problem inside the parts. For purposes of scalability, this partitioning will have to be done recursively. In some approaches, a network is split up into a number of main arteries, each artery containing several junctions. Such split-ups are not area-oriented but artery-oriented. In our opinion, this approach is not a good idea. With such a split-up, it is obvious that one can do a further split-up of an artery into pieces, such that each piece contains at most one junction. If two arteries cross each other and share a junction, this would lead to two systems controlling one network element (the junction), which would have to be resolved by giving one of the two arteries priority over the other, which would then effectively be split up. In this way, one obtains an area-oriented split-up in which control is easier.

Control Framework. We need a control framework to organize the many units of control present in an NM system, in order to guarantee that control scopes do not overlap, and that each control unit is governed by only one parent system. A proposal for this has been made in [13]. More research is needed on the control of distributed systems.

Priorities. It is obvious that some parts of a network are more important than others, in the sense that congestion in important road segments has more impact than in less important parts (f.i.: a belt road is more important than a residential area). This plays a crucial role in network management which necessarily must be focused on keeping fluid important parts at the expense of less important parts. Therefore, an effective priority assignment to network parts is needed, which reflects the importance that network parts have and which can be used in control strategies.

System Architecture. NM will have to be supported by automated systems. Although still many details of effective network management are not yet clear, such systems have to be built, lest there will be no progress in NM theory development. Theory, systems, and operator experience will have to evolve together. Systems will have to cooperate, which can be facilitated by a systems architecture which, at the architectural level, explains how the traffic management theory can be expressed in monitoring and control systems and decision support systems for human operators.

Human Operators. Human operators will play a key role. Because of the chaotic nature of traffic, the activity of traffic management is more like stock trading than like controlling an automated production process with few unpredictable disturbances. The capabilities of experienced human operators in assessing traffic patterns and in responding to unpredicted situations with no precedents, are not so easy to automate. In the course of time, some of these capabilities will be automated, but on the other hand, the capabilities of operators will also be strengthened by better visualization of traffic states and better decision support. NM will remain a matter of humans + machines in the foreseeable future, thus operators should have strong involvement in the development process of NM.

Interoperability Standard. In order to make real-life deployment of NM systems feasible from the point of view of cost and effort, and also from the point of view of market reform, an open interoperability standard would be most welcome. Due to the many connections an NM system needs with both the local control systems it coordinates and with the neighboring NM-systems, such a standard should support both horizontal and vertical communication.

6 Key Issues in Network Management

In this section we will go further into a number of the key challenges mentioned in the approach above.

6.1 Abstraction and Network Partitioning

Abstraction and Network Partitioning both aim at reducing the complexity of NM. There is an obvious candidate for abstraction, which is also a basis for network partitioning. This candidate is the property that, when considering a subnetwork S, only the boundary, which consists of a integer number of entries and exits of S, is of relevance to the rest of the network. All internal behavior can be ignored, only what happens on the boundary counts. This abstraction principle is also reflected in the so-called network fundamental diagram ([2], [4], [1]) which relates total flow through the boundary to the number of cars inside the subnetwork. Although mathematically attractive, such a strong reduction to a single number maybe somewhat too strong. It holds only if some degree of homogeneity can be assumed for the network, which in a rectangular streets and avenues network may be a sensible assumption, but for European cities is not very realistic. There are other, more informative quantities, such as OD-matrices for flow and for travel time, that are still pure boundary quantities but contain far more information, and therefore make fewer assumptions about the subnetwork involved. This abstraction principle implies a way of partitioning a network: namely such that boundaries between parts are always cross sections ('points') of carriageways in one driving direction. This partitioning can be applied recursively, resulting in a tree-structure split-up of the network, similar to (and partially coinciding with) the political partitioning of the surface of the earth [8] into countries, provinces, districts, etc. The big challenge however is how exactly to do this for the purpose of effective NM. A number of heuristic rules can be given, such as the high cohesion, low coupling principle and the rule that boundaries should be positioned in quiet areas. In the examples that we know of, the split up has been done by hand, but this is a tedious and error-prone job, and it makes the split up not very flexible. Some people suggest that the partitioning should be traffic state dependent. So for special events or sudden accidents, it should be possible to adapt the partitioning on the spot. For these reasons, it would be most welcome if effective partitioning of networks can be automated. There is some research on this subject [9], [6], [3], [8], which among others points out that many network algorithms tend to be NP-hard, and that the network structure has a big influence on its performance, but a convincing breakthrough in this area is still to be found.

6.2 Priorities

However obvious it may seem that some parts of a network are more important than others, it is still largely unknown how to assign priorities to network parts, and how to use these priorities in control strategies. In addition, priorities are also assigned to kinds of traffic, such as public transport. Both types of priorities will play an indispensable role in future NM, but still a lot of research is needed on the topic of finding effective and implementable algorithms to assign priorities to network parts.

6.3 Interoperability Standardization

An interoperability standard is needed to make deployment of sizeable NM-systems feasible. Here again, there is a chicken-egg dependency. Such a standard should be able to express what one control system would like to request from another one. As long as NM is still not well developed, this remains only very partially known. A recent initiative for an interoperability standard for traffic control systems, focusing on NM systems and connections between NM and local control systems, is described in [12]. This approach harnesses the extensibility of XML to solve the chicken-egg dependency in this challenge.

7 Conclusions

Network Management for road traffic is badly needed in order to cope with our increasingly congested road networks, but its development is slow and its current level of deployment in practice is still limited. The slow development is caused by the inherent high complexity of the problem and the chicken-egg dependencies between the various components of this activity, such as between traffic management theory and the control systems needed for NM. Important components in the approach that we propose in this paper are: real-life pilots, algorithms for network partitioning and priority assignment, architecture and standardization for the systems development needed to support the NM activity, and strong involvement of operators in the development process.

Acknowledgements. This work was supported in part by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. INFSOICT-223844, project C4C, aka Con4Coord.

References

1. Cassidy, M. J., Jang, K., Daganzo, C. F.: Macroscopic Fundamental Diagrams for Freeway Networks: Theory and Observation. TRB 90th Annual Meeting Compendium of Papers (2011), Washington, DC 20001 USA.
2. Daganzo, C. F., Geroliminis, N.: An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transportation Research Part B: Methodological, 42, (2008) 771–781
3. Etemadnia, H., Abdelghany, K. F.: A network partitioning methodology for distributed traffic management applications. TRB 90th Annual Meeting Compendium of Papers (2011), Washington, DC 20001 USA.
4. Hoogendoorn, S., Hoogendoorn-Lanser, S;, Kooten, J. van, Polderdijk, S.: Integrated Network Management: Towards an Operational Control Method. TRB 90th Annual Meeting Compendium of Papers (2011), Washington, DC 20001 USA.
5. Julner, T., Olsson, K.: Deployment of a National Traffic Management System. Proceedings of the 18th ITS World Congress, 2011, Orlando, Florida, USA.
6. Li, Y., Yang, J., Guo, X., Abbas, M. M.: Urban Traffic Signal Control Network Partitioning Using Self-Organizing Maps. TRB 90th Annual Meeting Compendium of Papers (2011), Washington, DC 20001 USA.
7. Marcuson, J. K.: The integration of ICM and ATM. Proceedings of the 18th ITS World Congress, 2011, Orlando, Florida, USA.
8. Parthasarathi, P., Hochmair, H., Levinson, D. M.: Network Structure and Spatial Separation. TRB 90th Annual Meeting Compendium of Papers (2011), Washington, DC 20001 USA.
9. Parthasarathi, P., Levinson, D. M.: Network Structure and Metropolitan Mobility. TRB 90th Annual Meeting Compendium of Papers (2011), Washington, DC 20001 USA.
10. Pooran, F., Martinez, J. C. R.: True Adaptive Control Algorithms A Comparison Of Alternatives. Proceedings of the 18th ITS World Congress, 2011, Orlando, Florida, USA.
11. Vrancken, J. L. M., Ottenhof, F., Boel, R. K., Schuppen, J. H. van, Tassiulas, L., Valé, M.: WP4 Road Network Control: Design of the Measurement System. Deliverable D-WP4-3 of the C4C project, http://www.c4c-project.eu/index.php/filedb/showFile/143

12. Vrancken, J. L. M., Ottenhof, F., Valé, M., Lagerweij, R., Goossens, P.: DVM-Exchange, an interface standard for Traffic Management Systems. Proceedings of the 18th ITS World Congress, 2011, Orlando, Florida, USA.
13. Vrancken, J. L. M., Schuppen, J. H. van, Soares, M. S., Ottenhof, F.: A Hierarchical Network Model for Road Traffic Control. Proceedings of the 2009 IEEE International Conference on Networking, Sensing and Control, Okayama, Japan (2009), 340–344.
14. Yubin Wang, Ottenhof, F., Vrancken, J. L. M.: Integrating Bottom-up Traffic Control into the Scenario Coordination Module. Proceedings of the ITS Europe Congress, 2011, Lyon, France.
15. Yubin Wang, Vrancken, J. L. M., Ottenhof, F., Valé, M.: Next Generation Traffic Control in the Netherlands. Proceedings of the 18th ITS World Congress, 2011, Orlando, Florida, USA.

Applications of Video Surveillance Systems for Traffic Data Acquisition

Vasil Lakov

University of Architecture, Civil Engineering and Geodesy
1046 Sofia, Bulgaria
supervas220@gmail.com
http://www.uacg.bg/

Abstract. The process of urbanization raises city population and extends city boundaries, which leads to heavier road traffic, environmental pollution and bigger financial losses. Due to spatial structure and influence of many factors on transport conditions, the transport system requires complex management methods. To be widely adopted, traffic management projects have to have the lowest possible price and highest effect on transportation. To accomplish these requirements, city authorities have to know what and where the halting points are. Using existing CCTV cameras for traffic purposes can empower authorities to easily obtain city-wide traffic data. This paper analyzes the benefits and problem areas of CCTV systems when used for traffic surveillance and present different ways for improving reliability and accuracy of these systems. Examined areas are camera settings, optical filters, mounting place, digital image processing techniques for image enhancing and object extraction and objects filtering based on their properties.

1 Introduction

The rising of the world population and the process of urbanization are some of the reasons for differentiation of densely populated city areas. The growth of cities in both horizontal and vertical direction, strengthening of economic links among them and constant searching for profit require more effective ways for people and goods transportation. Road transport arteries connecting residential with industrial and commercial city areas, and highways between cities, gather extremely important role in functioning and growth of every modern country.

Mass usage of automobile transport is constantly increasing since its biggest advantage is door-to-door service. To be able to guarantee normal functioning of the global economical system, road transport systems must be:

- efficient (providing alternative routes, enough capacity, suitable for all vehicles),
- effective (with minimal expense of time and fuel),
- reliable (with minimal disturbance from planned and random events),
- ecological (with minimal environmental pollution),
- safe (with minimal risk for traffic participants).

As reasons for traffic jams, and the ways of reducing them are complex. The ways for improving transport quality may be grouped in the following categories:

- restricting or stimulation of using one or another type of vehicles and traffic reorganization,
- city areas and roads planning consistent with current and future transport needs,
- renewal of old one and constructing of new road infrastructure,
- enhancing of traffic flows surveillance and control.

2 The Role of Traffic Sensors

Intelligent Transportation Systems (ITS) are in use worldwide nowadays and their aim is to improve safety, performance, environmental impact and to provide sustainable development of road transport systems by modern transport technologies.

For identifying the bottlenecks and their locations, as also for revealing and evaluating the possible corrective actions, it is mandatory to have information about the road infrastructure and the traffic flows. The realization of a specific traffic project can influence different aspects of transport conditions and evaluating the project consequences in advance is an important step in its planning. By creating computer models of city or national road networks and by using real stress data acquired by traffic sensors optimal decisions are more easily taken in various situations during traffic management and planning in response to current demand and conditions [1,2].

Traffic flow management relies on up-to-date data for different traffic parameters. In addition to individual vehicle parameters sensors compute generalized indicators describing automobile flow as a whole. This data is used for a wide range of tasks such as:

- evaluation of road load,
- simulation of computer transport models,
- adaptive control of traffic lights,
- identifying of peek hours,
- identifying locations with frequent traffic congestions,
- evaluation of sound and atmospheric pollution,
- determination of the predominant types of vehicles,
- route determination of major traffic flows.

Examples of measured vehicle parameters are:

- used lane,
- direction,
- speed,
- type of vehicle,
- weight of the vehicle,
- length of the vehicle,
- time of passage,
- registration number,

On the basis of these parameters following indicators can be computed:

- average flow speed,
- distance between consecutive vehicles,
- loading degree of the road section,
- air and noise pollution,
- existing or impending congestion,
- traffic violations.

For the realization of the most suitable project meeting expected improvement of transportation conditions it is necessary to clarify beforehand the following questions:
- what is the transport problem,
- what information is needed for solving the problem,
- what are the specifics of the road section and the traffic.

Having defined the above points a project may proceed with the choice of technology and specific road sensors meeting the requirements.

3 Traffic Sensors

The two main groups of traffic sensors are:
- in-vehicle sensors,
- road sensors.

The first group of sensors represents devices such as mobile phones, GPS receivers and RFID transponders, mounted in each vehicle and monitoring only this vehicle. In this case, many devices are used for ensuring proper data and covering larger area of the road network.

The second group of sensors is installed at a specific point of the network for measuring local traffic parameters only. One road sensor can monitor multiple lanes with the same or different directions. This group includes video surveillance systems for acquiring and analyzing traffic data. These are passive sensors covering long road section, providing simultaneously video signal from the place, which is the major difference compared to other types of road sensors. For extracting useful traffic data from the video signal different algorithms for video image processing are being applied [2].

4 Video Road Sensors

Video road sensors are based on video surveillance systems and techniques from machine (computer) vision for analyzing of video frames in the process of acquiring data for the vehicles visible in the frames. The video camera frames can be processed either centralized in the control centre or locally by each video camera.

In the first case, computers are located in the data centre and are processing video streams from remote cameras, as one computer or controller can handle multiple video streams. In the second case, video cameras are equipped with integrated computer modules for digital image processing. These video cameras transmit only traffic data to the control centre, but it is also possible to stream live video. The choice of the video system type depends on various factors, whereas it is possible to establish a hybrid system including already installed analogue cameras as well as new smart digital cameras. Streaming of the video signal from the cameras can be permanently or when certain conditions are being registered such as abnormal parameter values, a traffic incident or a violation [3].

Video road sensors implement two types of algorithms for measuring vehicle parameters. The first one measures parameters only once in a set point of the video frame, normally when vehicle crosses a virtual line, thus simulating physical road sensors. The other type measure constantly vehicle parameters, as long as vehicles are visible. This process includes tracking each vehicle on each consecutive video frame.

In order to be appropriate for road traffic monitoring, video road sensors must fulfil following requirements:

- automatic detection of all vehicles,
- detection of both moving and stopped vehicles,
- operability under various weather conditions,
- operability in real time.

To be considered a universal tool in transportation projects video road sensors must also provide:

- high accuracy of the measured parameters,
- simplicity of installation and maintenance,
- high fault tolerance,
- ability to exchange information with other systems,
- ability to visualize and analyze collected data.

5 Advantages of Video Sensors

In the big cities, there are already established systems for video surveillance of various public places, including important road arteries for the purpose of security. The signal from these video cameras can be used simultaneously for the purpose of traffic monitoring and safety. In this case the whole infrastructure is ready and necessary investments are substantially lower. What remains to be done is the installation of computer equipment in data centres for processing video signals from cameras. This significantly reduces the time for building the system.

With the ability to cover a large number of lanes and a long road section by one video sensor different road situations can be detected e.g.:

- speeding,
- stopped or parked vehicles,
- driving in forbidden lane,
- driving in wrong direction,
- crossing a continuous line,
- red light crossing,
- smoke in tunnels,

– debris on the road.

This allows obtaining extensive information about the traffic by installing a small number of devices. Installation of new cameras can be done on the side of the road, as not to hinder normal traffic.

Adding new features to the processing of the video signal can be done in stages, starting with vehicle counting and completing the system with their classification and violations and congestion monitoring. The modular structure of the software system will further accelerate the introduction of the system in operation and will allow the creation of optimized solutions for every specific project. The general use of video cameras eliminates the need to install other cameras for obtaining video images used by supervisory authorities to assess the traffic situation.

Thanks to the digital nature of the video sensor functioning and modern communications, it is possible to analyze video signals from anywhere in the world. To reduce needed bandwidth of the communication system, it is advisable to send first only a picture and after the situation has been assessed by operator to begin video broadcast in real time.

6 Disadvantages of Video Sensors

Difficulties in the operation of video sensors are a consequence of the difficulty in extracting the necessary data from the video footage in degraded frame quality, caused by the following factors:

– low light and bad weather conditions,
– low quality of video frames,
– scene complexity.

6.1 Low Light and Bad Weather Conditions

Insufficient light on the road is a major problem for video sensors. Separating cars from the background becomes unreliable, which leads to missed cars. Many algorithms for detecting cars do not work as well at night.

Bad weather such as heavy rain, snow and fog reduces the road lighting and image contrast and may introduce noise in the frame. Visibility is also reduced and the recognition of distant vehicles becomes difficult. The algorithms which use specific points or object edges for identifying and tracking of vehicles will not be able to reliably detect vehicles due to reduced contrast. Reduced number of the contours in the image as a result of decreased contrast is used by some systems for smoke detection in tunnels.

6.2 Low Quality of Video Frames

The quality of video frames is influenced by the following video camera elements:

– camera lens,

– image sensor type,
– automatic video functions.

Optical distortions are defects caused by a simple single lens or other imperfect optical system due to some phenomena and geometrical characteristics of light passing through each lens. Each lens suffers, to some extend, from various optical defects. As a result of these effects, the image created by the optical system is poorly focused, distorted or with changed colours. Distortions are of several types and are removed by a combination of different lenses. Completely removal of all distortions is impossible, and only these distortions that hinder the specific task are removed. This is done through complex combinations of lenses of different shapes, made of glass with different refraction ratios and combined in a way so that to remove all undesired image faults. Compensating the various distortions is necessary for the proper measurement of vehicle speed, trajectory determination and classification [4].

A smaller photosensitive element creates a stronger noise in the image and if the frame resolution is also low, the details in the image become indistinguishable, which worsens the overall accuracy of the extracted data.

CCD and CMOS sensors represent two different technologies for capturing digital images. Each has its strengths and weaknesses depending on the application and neither of the two is clearly superior to the other. For the purpose of traffic monitoring the frame speed is the one of the standard TV and the time for digitalization of the frame is not so important. But it is important to minimize the noise, keep the colour uniformity and light sensitivity at high levels because traffic cameras work 24/7. Characteristics of both technologies are shown in table 1.

Table 1. Characteristics of image sensors.

Indicator	CCD	CMOS
Chip output	Voltage (analogue)	Bits (digital)
Number of amplifiers	One	For each pixel
Noise level	Low	Medium
Complexity	Low	High
Dynamic range	High	Medium
Colour uniformity	High	Low to medium
Conversion speed	Medium to high	Higher
Power consumption	High	Low
Heating	High	Low
Pixels density	High	Medium

CCTV cameras have a set of functions for automatic determination of parameters for frame capturing depending on ambient conditions. For road sections monitoring, changes in the scene are small and slow, and the most important factor is the total light during the day. Some functions are not needed and removing or disabling them facilitates the processing and reduces the cost of the camera. These unwanted functions are BLC (Back-Light Compensation), AWB (Automatic White Balance), AF (Automatic Focus) and AI (Automatic Iris). In the absence of light sources behind the road BLC should be blocked. The colour change with AWB can reduce the reliability of the algorithms using colour information for detecting vehicles. AF and AI functions apply mechanical techniques which reduce the reliability of the camera and increase the complexity and cost. After mounting the camera the focus remains unchanged and the AI function is replaced by optical filters, AGC and AES.

6.3 Scene Complexity

When the video camera is mounted close to the ground or directed almost perpendicular to the road, it is possible for nearby vehicles to get overlapped and even hidden one by another. This leads to wrong counts or incorrect parameter values of these cars. Merging of close cars is possible also during periods when the sun is close to the horizon. Then cast shadows are longer and may fall on adjacent cars [5].

The presence of moving objects near the roadway or of ones often passing through it can also lead to the adoption of these objects by the processing algorithms as real cars. In this case, false alarms of traffic violations or other dangerous situations are more likely to be generated by the system. Such unwanted objects can be:

- wild animals,
- pedestrians,
- moving trees,
- buildings.

Trees located adjacent to the road may hide part of it or be source of movement in the frame (moving cars) in the presence of wind. Light sources at night can blind the camera, illuminate excessively part of the road or be recognized as car headlights. Glare from road surface (especially when wet) or vehicles surfaces in some cases can be the cause for the impossibility to determine the vehicle type, the fusion of cars and road and the loss of colour information. Even if these unwanted objects are out of the analyzed area of the frame, they can change various parameters used in frame processing, such as image thresholding and objects filtering values.

During observation of roundabouts and squares with divers streams of cars and pedestrians, where bus stops are present together with traffic lights, recognition and filtering of the various objects and determination of their behaviour is very difficult. If the road is curved and the system checks for violations as prohibited overtaking or driving in prohibited lane, the vehicle trajectory will be an arc and not straight line, which complicates its analysis.

7 Hardware Methods for Improvements

In creating a universal tool for traffic monitoring for all tasks and conditions, with regard to the accuracy and reliability of the final results, it is appropriate to overcome the problems at each stage. The optimization process has to start with choosing the camera and its location, to continue with setting camera parameters and to close with video footage processing and filtrating algorithms and parameters.

7.1 Mounting Place and Video Camera Parameters

The cameras are mounted on a certain height above or on the side off the road. They have to be pointed down towards the road, and all observed lanes have to leave the frame at the top and bottom, not sideways. The sky must not enter the frame in order to reduce ultraviolet light, brightness changes, camera dazzle by headlights, etc. Surrounding objects and light sources, which may hinder processing algorithms, must also be minimized.

Stationary cameras without zoom capability and progressive frame scanning are used. Interlaced scan cameras require further signal processing before information from the signal can be extracted.

To avoid having to adjust the iris, lenses with shorter focal length have to be used to keep the depth of field long enough during the night, when the iris opens to allow more light. A wide-angle lens provides a larger viewing area, but it introduces also heavier geometric distortions.

When expected that the substructure will experience vibrations, it is necessary to use a camera with image stabilization. It is sufficient to compensate the linear displacements of the image sensor, because there is usually no rotation. Stabilization can be achieved in two ways – mechanical, by movable lens and digital, by image processing.

Useful video functions are AGC (Automatic Gain Control) and AES (Automatic Electronic Shutter). AGC function adjusts the gain of analogue signal from image sensor at different levels of illumination. This function is useful during the evenings and night. AES function is responsible for the shutter speed, which affects the brightness of the frame. For many algorithms is important that the difference between consecutive frames is minimal. To achieve this, the two functions should work, so that the parameters of the frame remain constant, even when the amount of sunlight changes. Rapid change in brightness of the image is observed when a large vehicle passes near the camera. Such a vehicle occupies big part of the frame and the large surface reflects more light toward the camera.

7.2 Additional Optical Filters

The overall quality of the video frames can be improved by adding optical filters in front of the camera lens. Filters for contrast enhancement are suitable for places with frequent fog or haze. Polarizing filters reduce the glare from water, wet and metal surfaces such as the coverage of most vehicles. Ultraviolet and infrared light filters improve colour reproduction and durability of the image sensor in the video camera.

7.3 Processing Units

The performance of microprocessors is growing steadily and today it is sufficient for the purpose of digital image processing in real time and in particular for video traffic monitoring. There are different technologies for the construction of computer modules for general and specialized applications, such as:

– digital signal processors (DSP),
– system on a chip (SoC),
– field-programmable gate arrays (FPGA),
– application-specific integrated circuit (ASIC).

Some of the solutions use hard-coded algorithms that can not be changed once the chips are manufactured or once software is written to the chip. Since all functions of video traffic sensors are realized by a software program it is desirable to be able to change and enhance them after initial system deployment. In this sense it is better to use standard solutions such as DSP processors, enabling easy software update.

When using a dual-core DSP processor, each core can process an individual frame, so the cores will have twice the time to finish the processing. Another option for optimization is to separate the series of processing operations between the two cores. So the volume of tasks for each core will be smaller and will require less time.

8 Software Methods for Improvements

Having selected the optimal location and camera, here comes the optimization of the filtering and processing algorithms for video signal analysis. Filtering is a basic method for removing erroneous values in the data, which increases the accuracy of the final result.

8.1 Camera and Traffic Parameters

In visual identification of vehicles, the ability to set different system parameters is an advisable step. The configuration of the observed road section is static so it is possible to preset parameters describing the road, camera characteristics and normal ranges of traffic parameters. Some systems rely on the relationship between the size of the frame in pixels and the actual size of the observed road. It is even possible by setting the actual coordinates of several points of the frame to determine three-dimensional location and type of vehicles, without using stereoscopic camera [6]. Such configuration parameters can be:

– road length,
– number and lane boundaries,
– driving direction for each lane,
– real distance between virtual sensors,
– mounting height of the video camera,
– viewing angle or focal length of the lens,
– central point of the intersection,
– zones not subject to analysis.

8.2 Processing and Filtering Parameters

Noise filtering should be done in two stages. The first stage covers video frame processing before measuring vehicle parameters and the second involves the evaluation of the estimated parameters.

The first stage could include reducing colour noise, contrast adjusting, colour normalization, application of smoothing filters, frame thresholding or difference between two consecutive frames and more. Reducing noise in the image is useful in cases where optical flow is calculated. Normalization of frames is aimed at eliminating unevenness in brightness. The application of morphological operation "erosion" on the binary image removes small objects occurring due to noise, rain, movement of trees, birds, animals, etc. In order to eliminate false objects or those which are not important, an area of interest must be set and only in this area objects are analyzed. This is useful when is not possible to avoid areas close to the road with common movements like sidewalks and parking spaces.

It is possible that even after frame filtering and separation of potential vehicles, the extracted data contains objects that are not cars or with incorrect parameters due to some reason. Such objects are moving with unreal speed, suddenly changing direction, got too large or too small. Characteristics of the vehicles slightly change between successive frames and the big difference have to be considered as not normal and corresponding object filtered out. Example of a jump-like change of the parameters of a vehicle is where two or more vehicles overlap visually or by their shadows. Additional filtering parameters can be the valid ranges for vehicle size. For example, motorcyclists and cyclists occupy significantly less space than cars, and trucks and buses more space. This space can be measured during on-site trial tests and later setup in the software. Only after this stage the system can proceed by reporting the real count of vehicles, calculation of additional traffic indicators and checking for violations.

8.3 Vehicle Detecting Algorithms

In the theory of digital image processing, there are numerous operations. Different analyzing algorithms apply different sets of operations for a given task and therefore their efficiency and accuracy is varying and depends on many factors. The main task of these algorithms is the detection of moving and stopped vehicles on the road. Other task may include speed and vehicle type recognition. Development of fast and reliable algorithm is a process requiring time and deep knowledge. Different weather and road conditions may require the use of several methods to achieve accurate data in a wide range of situations.

At present there are numerous methods for detecting objects in video footage. They can be classified into 3 groups:

- methods using prior information about objects,
- methods based on movement or change in video footage,
- methods based on wave analysis.

Preliminary information used to identify an object as a vehicle can be symmetry, colour, contours of the vehicle, the presence of round objects (wheels), 3D models and headlights. Symmetry in the horizontal and vertical directions can be used to distinguish vehicles from other objects. Due to the typical colour of asphalt and its uniformity, opening the car can be based on colour differences. The presence of a dark shadow is a sure sign of the presence of a car. This method is limited to places with no side objects near the road.

On the basis of uniform colour and the lack of contours in the image and during vehicle passage contours appear, which can be used to classify vehicle by type by creating its wire 3D model. Change in a certain way of intensity of the image, also could be signal for the presence of a vehicle.

Classification of vehicles according to the distance between the front and rear axle (wheels) requires the camera to be positioned laterally and preferably perpendicular to the road. Another method used during the night time is the detection of a pair of punctuated bright objects representing vehicle’s headlights.

All methods discussed so far use spatial features to separate vehicles from the background. Another used method is the calculation of optical flow. The vector field of the moving object is called optical flow. Here several frames are used, i.e. the temporal characteristics of the objects. The calculation of optical flow is a heavy computational task that is only applicable to moving objects. This prevents the method from applying on congested roads and intersections, because stopped vehicles can’t be detected.

Wave transformation is a new method for image processing. Movement of vehicles is described in full 3D spatial-temporal model covering several video frames.

8.4 Computation Difficulty and CPU Power

Another area influencing the final results and functionality of the system is the computational difficulty of the algorithms. Since it is important to obtain data in real time, it is necessary to pay attention to the following factors:

- volume of the video signal,
- the efficiency of the processing algorithms,
- the performance of microprocessors.

Volume of the input signal W is defined as a product of width and height of the frame in pixels multiplied by the number of frames per second and is measured in pixels/s:

$$W = F_W F_H F_S . \qquad (1)$$

Where W - volume, F_W - width of frames, F_H - height of frames, F_S - frames per second. The bigger is the frame and higher the frame rate is, the larger the volume of the input signal per second is, which requires a faster processor.

The implementation of additional filtration of the frames or more precise algorithms requires the calculations to be optimized. Common approach for speeding up calculations is to use integer arithmetic, since this type of operations are faster compared to floating point operations. Another optimization can be reducing the volume of the input signal via a pre-resizing of the frames by a given ratio smaller than 1.

Frame rate of 10 fps is sufficient to obtain accurate data, but typically the standard TV speed of 25 or 30 fps is used to obtain smooth video.

9 Conclusions

The video sensors for traffic monitoring are well-suited for a wide range of transport tasks because in addition to measuring individual vehicle parameters, they can also detect traffic violations, queues of cars and unwanted situations such as traffic jams. They provide coverage of a large number of lanes on a long road section, which makes them attractive. The lack of necessity for stopping the traffic during installation and maintenance activities, as well as the possibility for using already installed video cameras, makes this solution even more valuable for transport authorities. The opportunity to expand their software functions remotely from the control centre, without changing the cameras, makes from video traffic sensors a promising technology with long life.

The guidelines for future development in the sphere of video traffic monitoring can be in the direction of improving the techniques and algorithms for vehicle detection in the cases of merging several nearby vehicles and working during night time. There are video sensors using stereoscopic cameras addressing the problem with merging of adjacent vehicles. These video sensors determine exact spatial location of vehicles on the road and their type.

References

1. Mimbela, E., Klein, L.: A Summary of Vehicle Detection and Surveillance Technologies used in Intelligent Transportation Systems, New Mexico State University (2007)
2. Lakov, V.: Application of video surveillance in traffic management, Utilities, ISSN: 1312-3017, Vol. 91 (2011)
3. Coifmana, B., Beymerb, D., McLauchlanb, P., Malikb, J.: [2] A real-time computer vision system for vehicle tracking and traffic surveillance, PERGAMON Transportation Research Part C 6 (1998)
4. Markishki, P.: The Magic of the Light, http://magic-light.photopia-bg.net/Photo1b.htm (2010)
5. Bramberger, M., Pflugfelder, R. P., Maier, A., Rinner, B.: A Smart Camera for Traffic Surveillance, Institute for Technical Informatics, Graz University of Technology
6. Han, D., Leotta, M., Cooper, D., Mundy, J.: Vehicle Class Recognition from Video based on 3D Curve Probes, Brown University of Rhode Island

Monitoring Information Quality: With Applications for Traffic Management Systems

Markus Helfert[1], Fakir Hossain[1] and Owen Foley[2]

[1]School of Computing, Dublin City University, Glasnevin, Ireland
{Markus.Helfert, Fakir.Hossain}@computing.dcu.ie
[2]Galway Mayo Institute of Technology, Dublin Road, Galway, Ireland
Owen.Foley@gmit.ie

Abstract. Information quality (IQ) plays a critical role in all management systems; however for traffic management systems IQ is often of limited consideration. The general approach to the study of IQ has offered numerous management approaches, IQ frameworks and list of IQ criteria. As the volume of data increases, IQ problems become pervasive. An example is decisions within traffic surveillance and management system, with its large amount of real time data and short decision times. Recognizing limitation in the applicability of current IQ frameworks with our work we aim to provide a practical-orient approach, and propose a process centric IQ monitoring framework that can be useful for traffic surveillance and management systems. Key to our work is the perspective of information systems as information manufacturing system (IMS). Objective of our information quality monitoring framework is to develop a comprehensive monitoring system that complements traffic management and surveillance systems.

1 Introduction

Ensuring the quality of information has become an increasingly important factor in decision support and management systems [7,11,33,34]. However for traffic management systems IQ is often of limited consideration. Thus it is not surprising that traffic surveillance and management systems need to address the issues of improving information quality in order to be successful. Recognizing the importance of Information Quality (IQ), practitioners and researchers have considered for many years ways to improve its quality. Scientists have worked on mathematical and statistical models to introduce constrain based mechanism to prevent data quality problems. Management of the process of data generation and the management of Information Manufacturing Systems (IMS) have also attracted many researchers. With the increasing importance of IQ, much research in recent years has been focused on IQ assessment. Researchers have developed many frameworks, criteria lists and approaches for assessing and measuring IQ. The frameworks most widely used have been recently documented and adopted by the International Standards Organizations (ISO) [16].

IQ has been often defined as a measure for ‘fitness for use’ of information [32]. The discussion follows the general quality literature by viewing quality as the capability to ‘meet or exceed users’ requirements.’ Common examples of IQ dimensions are accuracy, completeness, consistency, timeliness, interpretability, and availability. Over the last decade, many studies have confirmed that IQ is a multi-dimensional concept [e.g. 3,15,28,31,32] and its evaluation should consider different aspects. The literature provides numerous definitions and taxonomies of IQ dimensions analyzing the problem in different contexts. Also, literature provides us with numerous case studies, investigating IQ in practice.

However, the practical application of most of the proposed approaches is still very limited and continuous improvement activities for data quality are rarely integrated. Furthermore although some studies examine effects of IQ [e.g. 6], applications of IQ Management to traffic management and surveillance systems are rare. Therefore the IQ problem continues to exist. In addition as the volume of data and the complexity of traffic management system increases [9], IQ problems become pervasive. Most frameworks are aiming to adopt IQ criteria and to develop suitable and domain specific measurements. Currently this process requires intensive domain expertise, as the adoption of the frameworks is limited. Furthermore, Knight and Burn [21] point out that despite the sizeable body of literature available relatively few researchers have tackled quantifying some of the conceptual definitions. We also observed that despite the inherent subjective nature of IQ, most researchers focus on providing a general applicable IQ framework without considerations of its adoption in different environments. Neither process mapping nor data modeling provides sufficient provision to define the required quality that data/information must conform to. Furthermore, on-going monitoring of the conformance of the information production process is not possible without developing a cost and time prohibitive data monitoring system. The problem of data and information quality increases as the volume of data and the time requirements increase.

Recognizing the limitations of current approaches and aiming to provide a practical-orient approach, we propose a process centric data quality approach that can be useful for traffic surveillance and management systems. In contrast to other approaches we do not aim to develop a domain specific IQ approach. In our work we aim to develop a more general approach that can then be used in several contexts, include traffic management and surveillance systems. It also incorporates a technique to specific suitable data product qualities and assesses its conformance. Objective of data quality monitoring framework is to develop a comprehensive monitoring system that for instance can be used to complement the traffic management systems in form of an information manufacturing system. To test the approach on key develop step is to design and implement a process independent monitoring system that will continuously monitor data in traffic management systems to ensure various aspects of data and information quality.

In this paper we present the overall framework, consider the benefit of a process centric framework for on-going data quality monitoring and discuss its application to traffic surveillance and management systems. Our results show that the context dimension is crucial in IQ assessment and that our framework helps to form context-aware IQ assessments that indeed can be applied to other contexts. The paper is structured as follows. In Section 2 we reflect our work with related research and outline limitations of current approaches. In Section 3 we propose a context-aware IQ framework which is then used to outline or data quality monitor framework in Section 4. Section 5 concludes the article and presents indications for further research in form of implementing the framework for traffic management and surveillance systems.

2 Related Work

IQ has been investigated for many years and numerous frameworks and criteria lists have been proposed. Although claims are made to provide generic criteria lists [32], on closer examination most research has been focused on investigating IQ within a specific context [e.g.1,4,10,13,20,24]; however traffic surveillance and management systems are little or not considered [12]. Analyzing some popular IQ frameworks [18,26,32] we can observe a large number of dimensions and criteria associated with IQ. One of the most popular and referenced frameworks was proposed by Wang and Strong [32], and since then has been applied to many contexts and research. A critical element of any IQ assessment is to assign specific values for each IQ criteria through objective, repeatable and reliable measures. Over the years a variety of IQ assessment methodologies have been proposed. [2,8,15,19,22,27,28,30] provide examples of typical methodologies that can be compared by various criteria [14]. On the one hand, IQ is often measured with subjective perceptions from information users. On the other hand, research has developed objective IQ measures on the basis of quality criteria (mostly for intrinsic IQ characteristics such as accuracy, completeness and correctness). But as of today no widely accepted IQ framework with generic, generally applicable measurements is available. This makes the application of IQ concepts to traffic management and surveillance systems challenging. Furthermore, most frameworks do not provide any guidelines to apply the framework to various contexts. Most frameworks only provide very limited assistance for analyzing causes of insufficient IQ and often do not provide any plan for solving identified problems. Furthermore most frameworks are limited in considering specific requirements.

The limited work on IQ in traffic management and surveillance systems together with the challenges to apply foremost IQ approaches to new domains, underpins the requirement for our current work to design a more general applicable information quality management approach.

3 A Context-oriented IQ Framework

In order to develop our general applicable IQ framework, we base our concept on two traditional and well established concepts. In order to structure characteristics of information, we follow the theory of semiotics. In addition, in order to provide different quality views, we follow general quality literature and structure quality along “quality of conformance” and “quality of design”. Semiotic is a relatively widely established discipline, which has recently received increasing attention. Indeed, since the publication of Stamper [29] semiotic has revealed its relevance to information systems (IS) in many research. Stamper extended the traditional three layers of semiotics (syntactics, semantics and pragmatics) with additional aspects (physical, empirical and social aspects) forming the “semiotic ladder” that consists of the views on signs from the perspective of physics, empirics, syntactics, semantic, pragmatics, and the social world [25].

Furthermore, numerous discussions related to quality indicate that defining quality is at least as challenging as the term information itself [12,17]. This approach comprises of two aspects of quality:

(1) Quality represents certain product characteristics, which meet customer needs and thereby provide customer satisfaction.

(2) The absence from deficiencies that result in customer dissatisfaction [17].

In general, the first aspect refers to *quality of design* whereas the second aspect refers to *quality of conformance* [12]. Quality of design addresses the aspect of information requirements and information product design. "How good are the requirements met by the information product design?"

The conformance of the final information product with the product design is addressed by quality of conformance. Quality of conformance takes the divergence of design with the final product into consideration. Because low quality of design and low quality of conformance have different causes and therefore different solutions, it is fundamental to consider both aspects. High quality of design does not mean high quality of conformance and vice versa. Increasing quality of design tends to result in higher costs, whereas increasing in quality of conformance tends to results in lower costs. In addition, higher conformance means fewer complains and therefore increased customer satisfaction. In this article we limit our discussion of this view on IQ and refer to [12], in which an application to Data Warehouse Systems is illustrated.

Having established an IQ framework for relevant IQ dimensions (Table 1), it needs to be applied to a particular context such as traffic management and surveillance systems. In order to evaluate the application context, the application of an IQ framework requires an analysis of the IS environment (e.g. traffic management) prior to the measurement of IQ dimensions.

Table 1. Information quality dimensions based on Semiotic and Quality aspects [12].

Semiotic Level	Quality Aspects		Measurement Approach
	Quality of Design	Quality of Conformance	
Pragmatic	Relevance, completeness	Timeliness, actuality, efficiency	Information process, application
Semantic	Precise data definitions, easy to understand and objective data definitions.	Interpretability, accuracy (free-of error), consistent data values, complete data values, , believability, reliability	Comparison with real world and experience
Syntax	Consistent and adequate syntax	Syntactical correctness, consistent representation, security, accessibility	Syntactical standards and agreements

The dimensions for an environment are many and varied. For example for traffic and surveillance environment timeliness, completeness and accuracy might be of high priority. They are also highly dynamic, time-depended and different user groups will priorities different dimensions. Knight and Burn [21] indicate that the choice and implementation of quality related algorithms for Internet searching is very much dependent on the characteristics of the World Wide Web. In addition, the emergences of new information system architectures and service oriented architecture (SOA) have underpinned the importance of the environment and context to the fore. The ability for organizations to distinguish between the impact of the environment and the traditional view of IQ dimensions is vital. The employment of traditional IQ frameworks does not allow for this. This observation led us to the development of an context-oriented IQ Framework that includes the context dimension and its relation to information quality measurements.

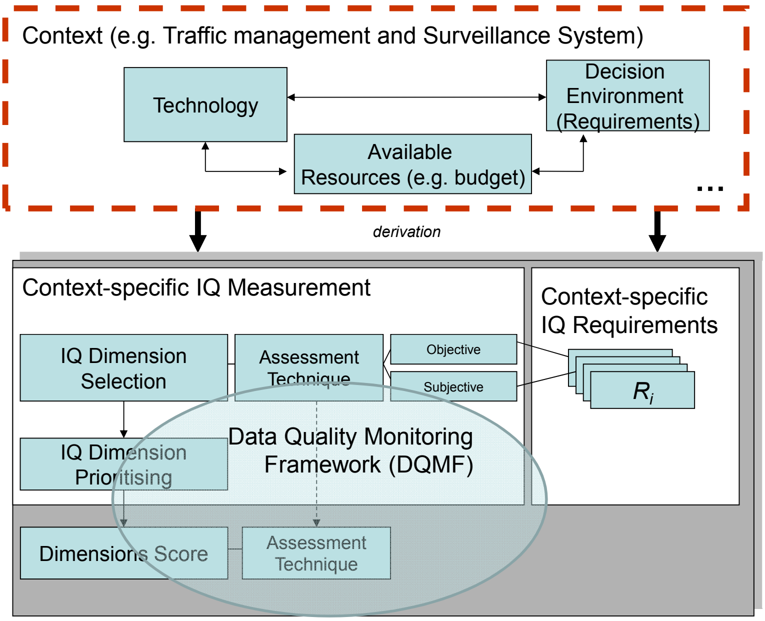

Fig. 1. Information quality framework *(grey area indicates context dimensions derived e.g. from traffic management system).*

Depending on the particular context, we are able to select and prioritize different IQ dimensions. For example, by applying Leung's [23] metric for importance, urgency and cost we can select the dimensions that are most suitable for traffic management and surveillance systems. In general, this will vary from environment to environment and within the same environment different user groups may select different IQ dimensions that most reflect the particular view of quality or the software services they require. These dimensions may change over time for reasons such as user skill level or change in software service.

The application of the framework should be on a regular basis in order to maintain the currency of the dimensions. This dynamic allows for adaptation of IQ dimensions to the constantly changing environment that more and more IS will find themselves. In our future research we aim to apply this framework to traffic management and surveillance systems, in order to identify the most important IQ dimensions. For example following dimensions could be identified (see Table 2).

Table 2. Information quality dimensions for traffice mangement and surveillance system (example).

Semiotic Level	Quality Aspects	
	Quality of Design	Quality of Conformance
Pragmatic	Completeness	Timeliness
Semantic	Easy to understand	Interpretability, believability, reliability
Syntax	Consistent syntax	Syntactical correctness, consistent representation, accessibility

4 Introducing the Data Quality Monitoring Framework (DQMF)

In order to measure the IQ dimensions, the above framework will be extended by a data quality monitoring framework which can be implemented as a comprehensive mentoring system. The key is to develop a process independent monitoring system that will continuously monitor data to ensure various aspects of IQ. In an example scenario, if traffic data could be continuously monitored to ensure notifications are sent timely, a problem could be detected much earlier and rectified with no impact to client/decision maker.

Buildings form the discussion above, we specify a data quality block in form of rules that incorporates data quality constrains and allows us to monitor certain aspects of ensuring quality. In the context of our example, if a text message fails to be sent from the traffic monitoring system, client might be not aware of a new situation. Instead of including the quality block into the system, an IMS independent quality conformance monitor would generate far better results in terms of performance. However, developing a parallel system to monitor data can also be time and cost prohibitive. Our aim in modeling quality block is also to develop data quality rules in such a way so that it can be feed to an independent data quality monitor. The framework consists of three core components. Data Quality Monitor (DQM), Data Product Markup Language (DPML) and Information Quality Markup Language (IQML).

4.1 Data Quality Monitor (DQM)

The data quality monitor is an application that accepts data product quality rules as its input and continuously monitors data product to ensure that it meets the agreed quality as defined. When designing the quality block, usually Business Process Modeling Notation (BPMN) can be supplemented by metadata about each information manufacturing block. Objective of the monitor is not to intervene in the process, but merely to monitor the data products to see if the data meets the quality requirement of the product relevant to the stage of its production. If the product fails to meet the requirement, it will report the inconsistency in accordance with agreed protocol to facilitate immediate intervention for corrective measures.

4.2 Data Product Markup Language (DPML)

A key element of our framework is DPML. In order to be effective quality controller, Information System models must describe sufficiently and accurately static, dynamic and organizational aspect of IMS. In a traditional manufacturing assembly line, as a product reaches various stages of its development, it can be inspected to ensure that it has met the requirement to be achieved at the relevant stage of the production. This is possible because a product in traditional sense will be predefined to achieve certain quality criteria that will be developed as part of designing the product. For our framework to work, we treat data as a product of information manufacturing system. At the design phase, we must then define the quality criteria that a data must meet at various stages of its production. In order to achieve this objective, we developed a Data Product Markup Language as an IP Unified Modeling Language (UML) based data product definition language. By using UML we can build on previous work to create visualized mapping of the data processes [5]. Furthermore, UML/BPMN is widely accepted is that it can be exported to code directly by cutting down on development time. This was further developed by IP MAP which extended a systematic method of representing the process involved in manufacturing of IP. Flow of data at various stages is also visualized by IP MAP. However, it lacks the ability to bridge various process and information product. There is also a need to, as described in the next section, to export the quality rules for automated execution. Hence we also base DPML on BPMN. We extend this model to model an integrated approach to define data quality requirements and business process together.

4.3 Information Quality Markup Language (IQML)

Once we are able to model data product using DPML, as described above, we need to translate it into an executable that can be processed by automated software. Otherwise, for each system a separate monitoring tool have to be developed. This is likely to make it cost and time prohibitive. This is why there needs to be ability to convert this DPML into and XML based rules that can be accepted by the monitoring tool. Information Quality Markup Language (IQML) is an XML based data product definition language. The purpose and nature of IQML is identical to that of DPML. Difference is that while DPML is UML based, IQML is XML based. IQML is either auto generated from DPML or generated independent of it. It is merely a means to facilitate data product definitions to be consumed by the Data Quality Monitor.

5 Conclusions and Further Research

As discussed above, IQ research provided numerous frameworks, criteria and methodologies to guide enterprise in the assessment, analysis, and improvement of

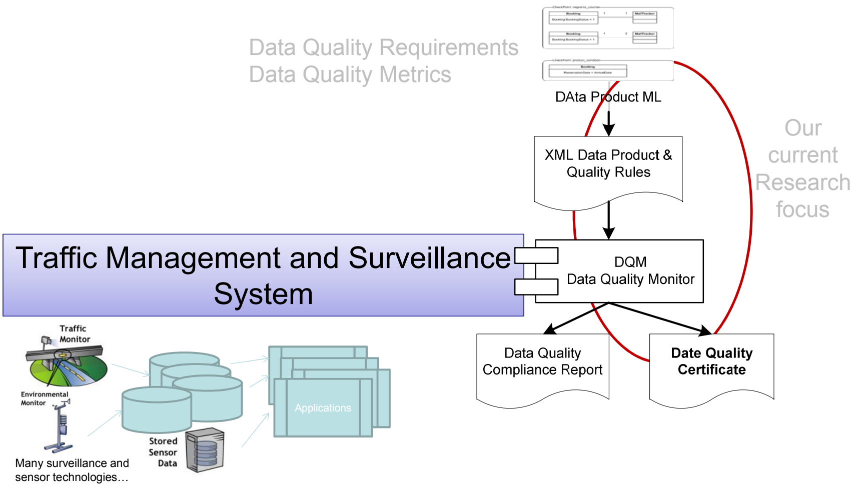

Fig. 2. Data Quality Monitor and Research Focus.

IQ. However, focusing on the critical issues related to the assessment phase, the literature does not provide an exhaustive set of metrics or guidelines that organizations can apply. Indeed examining literature on Traffic Management and Surveillance Systems, IQ in general and IQ assessments in particular are underrepresented. Our current research aims to apply common IQ approaches to various contexts, including traffic management. As illustrated in Figure 2, we aim to build a Data Quality Monitor. Most enterprises are developing their own approaches to address IQ issues although several algorithms have been developed for a subset of dimensions, such as accuracy, completeness, consistency, and timeliness. In fact the practical relevancy and generalization of some frameworks can be argued. Most common approaches used to obtain an IQ assessment is to consider domain specific measures associated with the different quality dimensions. Our research and discussion above shows the importance of context in measuring IQ. Indeed, the variations of IQ frameworks for different application scenarios indicate the significance of context in assessing IQ. Considering this observation we proposed a context-aware IQ framework. A critical element of our framework is the recognition of pragmatics (in the sense of semiotics) within our framework and the differentiation of quality of conformance and quality of design. This was used to propose a process centric Data Quality Monitoring Framework (DQMF), which can be useful for traffic surveillance and management systems incorporating data product quality and conformance. Objective of data quality monitoring framework is to develop a comprehensive mentoring system that can complement traditional traffic surveillance and management systems. The framework consists of three core components. Data Quality Monitor (DQM), Data Product Markup Language (DPML) and Information Quality Markup Language (IQML).

In future work we will implement our monitoring framework and develop a software tool that considers the context of IQ as outlined in our framework. This can be applied in the context of traffic surveillance and management systems. This will be considered for the empirical validation of the proposed context-aware IQ framework.

Furthermore, future work will also focus on the definition of an algorithm to obtain an aggregate quality measure able to assess the organizations' IQ level.

References

1. Alexander, J. E., and Tate, M. A. (1999) Web Wisdom: How to Evaluate and Create Information Quality on the Web, Lawrence Erlbaum, Mahwah, NJ.
2. Amicis, F. D. and Batini, C. (2004), A methodology for data quality assessment on financial data, Studies in Communication Sciences, 4(2), pp. 115-137.
3. Ballou D. P. and Pazer H. L.(1995) Designing Information Systems to Optimize the Accuracy-Timeliness Tradeoff, Information Systems Research, 6,1, 51-72.
4. Ballou, D. P. and Tayi, G. K. (1999). Enhancing data quality in data warehouse environments. Communications of the ACM. 42, 1, 73-78.
5. Ballou, D. P., Wang R., Pazer, H., Tayi, G. K. (1998) Modelling information manufacturing systems to determine information product quality, in Management Science, April 1998, Vol. 44, Issue 4, pp. 462-484.
6. Chen, Peter Shen-Te; Karthik K. Srinivasan, Hani S. Mahmassani: Effect of Information Quality on Compliance Behavior of Commuters Under Real-Time Traffic Information, Journal Transportation Research Record : Journal of the Transportation Research Board, 1676 (1999), pp. 53-60.
7. Chengalur-Smith, I. N., Ballou, D. P. and Pazer, H. L. (1999), The impact of data quality information on decision making: an exploratory analysis. IEEE Transactions on Knowledge and Data Engineering, 11(6), pp. 853-864.
8. Dedeke, A. (2000) A conceptual framework for developing quality measures for information systems. Proceedings of 5th International Conference on Information Quality, MIT, USA.
9. Dunkel, Jürgen , Alberto Fernández, Rubén Ortiz, Sascha Ossowski, Event-driven architecture for decision support in traffic management systems, Expert Systems with Applications, 38, 6, June 2011, Pages 6530-6539
10. English, L.: Improving Data Warehouse and Business Information Quality, New York: Wiley & Sons, 1999.
11. Fisher C.W. and Kingma B.R. (2001), Criticality of data quality as exemplified in two disasters, Information & Management, 39, 2, 109-116.
12. Helfert, M. (2001), Managing and measuring data quality in data warehousing, Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics.
13. Helfert, M. and Heinrich B. (2003), Analyzing data quality investments in CRM: a model-based approach, 8th International Conference on Information Quality, MIT, USA.
14. Helfert, M., Foley, O., Ge, M., Cappiello, C. (2009) Limitations of weighted sum measures for Information Quality 15thAmericas Conference on Information Systems, San Francisco, California August 6th-9th 2009
15. Huang K. T., Lee Y. W., Wang R.Y.(1999), Quality Information and Knowledge Management, Prentice Hall.
16. ISO/IEC 25012 (2008) Software engineering -Software product Quality Requirements and Evaluation (SQuaRE) -Data quality model, International Organization for Standardization.
17. Juran, J. M. (1998) How to think about Quality, in Juran, J. M., Godfrey A. B. (ed.): Juran's quality handbook, 5th ed., New York: McGraw-Hill, 1998, pp. 2.1-2.18.
18. Kahn B. K. & Strong D. M. (1998) Product and Service Performance Model for Information Quality: An Update, in Proceedings of the 1998 International Conference on Information Quality, MIT, USA.

19. Kahn, B. & Strong, D. M. (2002) Information Quality Benchmarks: Product and Service. Communications of the ACM, 45, 4,184-192.
20. Katerattanakul, P. & Siau, K. (1999) Measuring information quality of web sites: Development of instrument, in Proceedings of the 20th international conference on Information Systems. Charlotte, North Carolina.
21. Knight, S. & Burn, J. (2005) Developing a Framework for Assessing Information Quality on the World Wide Web. Informing Science, 8, 159-172.
22. Lee Y., Strong D., Kahn B., and Wang R. Y. (2002), AIMQ: A Methodology for Information Quality Assessment, Information & Management, 40, 2, 133-146.
23. Leung, H. K. N. (2001) Quality Metrics for Intranet Applications. Information & Management, 38, 137-152.
24. Li, S. and Lin, B. (2006), Accessing information sharing and information quality in supply chain management, Decision Support Systems, 42(3), pp.1641-1656.
25. Liu, K. (2000) Semiotics in information systems engineering. Cambridge, England: Cambridge University Press.
26. Miller, H. (1996) The multiple dimensions of information quality, in Information Systems Management, Vol. 13, Issue 2, Spring 1996, pp. 79-82.
27. Pipino, L. Lee, Y. W. & Wang, R. Y. (2002) Data Quality Assessment. Communications of the ACM, 45, 4, 211-218.
28. Redman, T.(1996) Data Quality For The Information Age, Publisher: Artech House.
29. Stamper, R. K. (1973) Information in Business and Administrative Systems. New York: John Wiley and Sons (1973)
30. Stvilia, B., Gasser, L., Twidale M. B. and Smith L. C. (2007). A Framework for Information Quality Assessment, Journal of the American Society for Information Science and Technology. 58(12), 1720-1733
31. Wand Y., and Wang R. Y. (1996), Anchoring data quality dimensions in ontological foundations, Communications of the ACM, 39,11, pp.86-95.
32. Wang, R. Y. and Strong, D. M. (1996) Beyond accuracy: What Data Quality Means to Data Consumers. Journal of Management Information System, 12,4, pp. 5-34.
33. Xu, H. and Koronios, A. (2004), Understanding information quality in E-business, Journal of Computer Information Systems, 45(2), pp. 73-82.
34. Xu, H., Horn, J., Brown, N. and Nord G. D. (2002), Data quality issues in implementing an ERP, Industrial Management & Data Systems, 102(1/2), pp. 47-59.

Author Index

Proceedings of the 1st International Workshop on
Future Internet Applications for Traffic Surveillance and Management - FIATS-M 2011
ISBN: 978-989-8425-87-4