

Urban Computing: a challenging problem for Semantic Technologies

Emanuele Della Valle^{1,2}, Irene Celino¹, Daniele Dell'Aglio¹, Kono Kim³, Zhisheng Huang⁴, Volker Tresp⁵, Werner Hauptmann⁵, Yi Huang⁵, and Ralph Grothmann⁵

¹ CEFRIEL – Politecnico of Milano, Via Fucini 2, 20133 Milano, Italy
`{name.surname}@cefriel.it`

² Dip. di Elettronica e Informazione, Politecnico di Milano, Milano, Italy
`emanuele.dellavalle@polimi.it`

³ Saltlux Inc., Seoul, Korea `kono@saltlux.com`

⁴ Computer Science Department, Vrije Universiteit Amsterdam, De Boelelaan 1081, Amsterdam, The Netherlands `huang@cs.vu.nl`

⁵ Corporate Technology, Siemens AG, Information and Communications, Munich, Germany `{name.surname}@siemens.com`

Abstract. In this paper we present the Urban Computing challenge and in particular we exemplify it in the context of traffic management. From our previous experiences in the field we draw requirements in terms of capacity to cope with heterogeneity in representation, semantics and defaults; with scale; with time-dependency of data; and with noisy, uncertain and inconsistent Data. Existing reasoning techniques fail to fulfill all those requirements once and for all.

1 Introduction

Our cities face many challenges well expressed by the following questions posed to the international community by the Urban Land Institute (<http://www.uli.org/>):

- How can we redevelop existing neighborhoods and business districts to improve the quality of life?
- How can we create more choices in housing, accommodating diverse lifestyles and all income levels?
- How can we reduce traffic congestion yet stay connected?
- How can we include citizens in planning their communities rather than limiting input to only those affected by the next project?
- How can we fund schools, bridges, roads, and clean water while meeting short-term costs of increased security?

The Information and Communication Technology (ICT) for sure cannot provide an answer to those question on its own, but it is one of the most important enabling factors. A signal that ICT for Urban area is growing was the publication in 2007 of a special issue of IEEE Pervasive Computing dedicated to Urban Computing [1–4] – the integration of computing, sensing, and actuation technologies into everyday urban settings and lifestyles.

Urban settings range from our own cars, while we drive them in town, to public spaces such as streets and squares including semipublic ones like cafés and tourist attractions. Urban lifestyles are even broader and include people living, working, visiting and having fun in those settings. Not surprisingly, people constantly enter and leave urban spaces, occupying them with highly variable densities and even changing their usage patterns between day and night [3].

Some years ago, due to the lack of data, solving Urban Computing problems with ICT looked like a Sci-Fi idea. Nowadays, a large amount of the required information can be made available on the Internet at almost no cost: computerized systems contain maps with the commercial activities and meeting places (e.g., Google Earth), events scheduled in the city and their locations, positions and speed information of public transportation vehicles and of mobile phone users [3], parking availabilities in specific parking areas, and so on.

However, current ICT technologies are not up to the challenge of solving Urban Computing problems: this requires the combination of a huge amount of static knowledge about the city (i.e., urbanistic, social and cultural knowledge) with an even larger set of dynamic data (originating in real time from heterogeneous and noisy data sources) and reasoning above the resulting time-varying knowledge. A new generation of reasoners is clearly needed.

For this reason we are running the LarKC project, which is aiming at a configurable platform for infinitely scalable semantic web reasoning [5, 6], to address one of the Urban Land Institute open questions: “how can we reduce traffic congestion and yet stay connected?”. We select such question because we have been working in this area for years and we can derive from our previous experiences challenging requirements not only for the LarKC project, but also for the entire community working on scalable, tolerant and dynamic reasoning.

In the rest of the paper, we present a Story Board to make our vision more explicit (Section 2). We identify traffic management (Section 3) as a special case of Urban Computing and we present state-of-the-art in this field. Then, we describe our past implementation experiences (Section 4), from which we derive requirements (Section 5) for LarKC project and the entire community. In Section 6, we provide a short description of the solutions we are working on, claiming that they are only partial and that new reasoning techniques are needed to fulfill all the requirements of Urban Computing once and for all. Finally, in Section 7 we give a broader vision of the potential impact of solving the Urban Computing challenge.

2 A Story Board: Getting to Milano

This use case shows the added value of creating a Urban Computing System (UCS) that collects a broad set of information around traffic, integrates it and uses it to support a citizen that has to go to Milano from a city nearby.

1. Carlo lives in Varese 60 km from Milano. Tomorrow, he has to go to the Lombardy Region premises in Milano where he arranged a meeting at 11.00.
2. He opens the “plan a journey” service of UCS and fills in the required data
 - FROM: via Luigi Sacco, 1 Varese after 8.00 tomorrow

- TO: via Taramelli, 20 Milano before 11.00 tomorrow
 - USING: any means of transportation
3. UCS provides Carlo with three alternatives, two using public transportation (i.e., A and B) and one by using his private car (i.e., C)
 - (a) Using railroad “LeNord” and Metro M3; leaving home at 8.30 and arriving between 10.15 and 10.30.
 - (b) Using railroad “Ferrovie dello Stato” (alternative to “LeNord”) and Metro M3; leaving home at 8.20 and arriving between 10.05 and 10.20.
 - (c) Using private car; leaving home after 9.30 (when the commuters traffic on motorway A8 is almost over) and arriving between 10.10 and 10.40.
 4. Carlo is tempted by option C, he could sleep a little longer, but while traveling by train he could complete the presentation for the meeting, so he chooses the option A and he uses the ticket-less option to buy the train ticket.
 5. Before exiting UCS, Carlo asks to be alerted if the option he chose will no longer be the best one (e.g., due to problems to the railroads).
 6. The day after at 7.14 UCS learns from the railroad information system of railroad “LeNord” that a technical problem is causing an average delay of 45 minutes to all “leNord” trains from Varese to Milano.
 7. UCS estimates that an accident of such kind will not be solved before 11.00, therefore it checks if any planned journey is at risk. It finds Carlo’s journey.
 8. UCS checks if the other options it proposed to Carlo are still valid. Apparently they are, so UCS sends an SMS to Carlo informing him that a accident is causing 45 minutes delay for all trains on railroad “LeNord” and he can either use the railroad “Ferrovie dello Stato” (option B) or take his private car, in this case Carlo can convert his train ticket into a daily parking ticket for one of the parking lots of the sub-urban metro stations in Milano.
 9. Carlo receives the SMS, he enters UCS and checks the two alternatives. He can take option B, but he knows that when a problem of this kind happens on “LeNord”, all commuters take the railroad “Ferrovie dello Stato” and he will never be able to find a seat. On the other hand, UCS (taking into account weather data through the route and real time traffic congestion status on top of historical traffic congestion statistics) predicts that being a rainy day the traffic on A8 will be slower and he should leave around 9.00.
 10. Carlo decides to take his car, in this way he has all the time to complete the presentation before leaving. He leaves home around 9.00 and instructs his GPS Navigator to interact with UCS traffic service and to find the cheapest gas station along the road. While driving Carlo receives the instructions for the gas station and refuels the car.
 11. At a certain moment his GPS Navigator receives alert from UCS: Milano North-West area is hit by heavy showers and the traffic is getting slower. Instead of going to the planned North-West parking lot, the GPS suggests to go to one in the South-West; the metro from there will only take 10 minutes more than from the planned parking lot, but the estimated time to the planned parking lot is 25 minutes more than the planned one.
 12. Carlo considers the option and decides to follow it.
 13. Carlo parks the car and taking the metro arrives to his appointment on time.

3 Traffic Management

Traffic demand has been growing steadily for decades and this growth is foreseen to continue in the future. For many years, the primary way of dealing with this increasing demand has been the increase of the roadway network capacity, by building new roads or adding new lanes to existing ones. However, financial and ecological considerations are posing increasingly severe constraints on this process. Hence, there is a need for additional intelligent approaches designed to meet the demand while more efficiently utilizing the existing infrastructure.

Public authorities have taken steps in this direction through the installation of traffic management systems intended to equalize traffic demand both temporally (by spreading out trips in time) and spatially (by redistributing demand). The domain of traffic management solutions has recently experienced a significant demand of advanced information technology. Control centers for traffic management are connected online to different devices (such as detectors on roads and traffic lights) making it feasible for operators to supervise the state of the road network by consulting data bases with recent information from detectors and to modify the state of control devices. The use of such traffic monitoring and management facilities requires sophisticated tools for traffic modeling, estimation, prognosis and decision support for online operators to help them in dealing with the complexity and diversity of information sensor data and control devices.

Traffic System Infrastructure Setup. Today, in a typical information infrastructure for real-time traffic control, as it can be found in different cities, usually the following basic components can be discriminated. There are sensors (e.g. loop detectors, cameras, traffic eyes, radar detectors) on major roads recording several traffic magnitudes such as vehicle speed (km/h), traffic flow (vehicle/h) and occupancy or traffic density, i.e., the percentage of time the sensor is occupied by a vehicle (vehicles/km). The distance between successive sensors on a freeway is typically in the order of about 500 meters. The information is periodically transmitted to a control center, which also receives information about the current state of control devices. Such control devices include traffic signals at intersections, traffic signals at sideways entry-ramps, variable message signs that can display different messages to drivers (e.g., warning about existing congestions, accidents or alternative path recommendations), radio advisory systems to broadcast messages to drivers, and reversible lanes (i.e., freeway lanes whose direction can be selected according to the current and expected traffic demand). In the control center, operators interpret the sensor data and detect the presence of problems and their possible causes. Problems are congested areas at certain locations caused by lack of capacity due to accidents, excess of demand like rush hours, etc. In addition, operators determine control actions to solve or reduce the severity of existing problems (e.g. extended green-phase of a traffic signal, switch of variable message signs).

Traffic management systems must be reactive to the different states of traffic flow in the controlled network. In the early systems, the approach was based on a

library of signal plans applied online in different predefined situations according to some time-based criteria or to the traffic data collected by roadside sensors. However, this precalculated-plan approach usually lacked the conceptual granularity required by the system to be adaptive enough to the variety of situations, in time and space, which may occur in the network. Later, more adaptive systems were introduced, where an intelligence for understanding traffic situations in real time was designed and integrated with a model for decision making.

In the last years, a considerable amount of work has concentrated on the fields of traffic modeling and estimation and the analysis and forecast of traffic conditions. One of the main tasks in traffic management systems is to model the current traffic condition for the entire road network. Traffic flow is normally only measured at certain points along the roadway. Employing appropriate models this data is used to estimate traffic conditions for the major part of the network. Here, different approaches are applied: methods for statistical evaluations and visualization of stationary and mobile traffic data, methods for evaluating and projecting traffic correlations from current and historical traffic flows and propagation methods for calculating the current traffic condition on the basis of origin-destination matrices as well as on statistical analysis of traffic data surveyed online. In general, the propagation method is based on the assumption that the traffic volume measured at a cross section is a superposition of different traffic flows. They branch out before and after the cross section within the network. From the assignment calculation the operator knows the different traffic flows which amount to the measured value. As a result, it is possible to allocate the percentage of each traffic flow to the routes within the network. The propagation method allows the user to dynamically visualize congestion impacts which are plausibly running upstream over several time slices.

Recent developments not only consider stationary traffic data provided by standard detectors but also allow to integrate so-called floating car data, as an increasing number of operators of advanced traffic management systems also use mobile traffic data. With this mobile data the level of detail of analysis and forecast methods can be considerably improved and information about areas not covered by roadside detectors can be provided.

Evolution of Traffic Management Solutions. Despite the increasing sophistication of the traffic management and control infrastructures run by public authorities, such collective systems for traffic management suffer from several limitations. One aspect is that they are unable to provide continuous, up-to-the-minute information to drivers. Another aspect is that it is impossible to restrict an advice or control action to a targeted subgroup of drivers, for example, those with a particular destination area. The options in collective route guidance are essentially “all or nothing”. On the other hand, private service operators will in the future be providing an increasing level of traffic information services targeted to the user’s need for the fastest or shortest route. Modern navigation systems are designed to take delays due to incidents or congestions dynamically into account, provided that these delays have been previously reported

and transmitted by some means to the device, e.g., by broadcast media such as TMC (Traffic Message Channel) in Europe. The reaction of present route guidance systems to delays and incidents is a short-term and/or small-scale strategy. Comprehensive optimization of dynamic routing strategies is not provided. In addition, current systems are unable to include the influence of public traffic management strategies, such as traffic signal coordination. Therefore, the integration of commercial route guidance recommendations with public policy and collective interests as well as advancing the further development of vehicle navigation are primary issues for the evolution of traffic management solutions. Particularly, new approaches and algorithms have to be provided for dynamically incorporating diverse available sources of traffic data and information as well as public traffic management and control strategies and priorities into integrated comprehensive systems. These kinds of integrated route guidance systems are also referred to as third-generation or traffic-responsive navigation solutions. They are to enable services targeted to the drivers' needs such as recommendation of routes with coordinated traffic signals in urban areas or congestion-free alternative routes on motorways.

Possible Approaches for Future Real-time Traffic Management. An analysis of the current and predicted traffic state in the entire road network and the identification of reserve capacities comprise the basis for advanced city traffic management and navigation solutions. Mobile and stationary sensors collect the appropriate traffic data and transmit it to a central unit. Similarly to the weather forecast, the different and heterogeneous information sources are combined to obtain an estimation of the traffic state during a period ranging from minutes to hours or even longer. Thus, a comprehensive knowledge base can be built up to support optimal individual route guidance.

In order to obtain a high-resolution picture of the current traffic state as well as of weather conditions and other environmental factors, current research activities are directed towards both utilizing existing stationary detection facilities, such as loop detectors, and advanced vehicle-based data source known as XFCD (eXtended Floating Car Data). Floating cars act as mobile sensors and can collect a range of information including speed and position data. During a trip, XFCD-vehicles perform an analysis using position, speed, and other data that gives important information on the local traffic state as well as the traffic context and surroundings (e.g., dynamic control systems, rain sensor, driver assistance systems, braking activity). If there is relevant information available it is transmitted anonymously to a traffic center and fused with other data sources. The advantage of this distributed data source is that measurements of traffic occurrences are possible in principle within the entire road network without the requirement of expensive stationary infrastructure.

The quality of dynamic route guidance crucially depends on the quality of the available dynamic traffic state reconstruction in the road network. The higher the quality of the reconstruction, the more reliable the traffic prediction. As mentioned above, one important contribution to this quality could be provided

by comprehensive measurement of XFCD. In addition, further concepts envision the integration of the origins, destinations and route information as planned by the individual onboard navigation systems of vehicles, which are provided to the traffic management center. On the other hand, innovative technologies are required in order to process and integrate the resulting collection of distributed information bits within a complex, diverse information environment. Here, a major task is the provision of appropriate solutions for the integration and fusion of heterogeneous information sources, where each source of information can have distinct characteristics with respect to availability, precision, reliability, resolution and representation (see Section 5).

4 Implementation experiences

u-City – Intelligent Transportation System. u-City is a new intelligent real-time city project currently in progress in South Korea⁶. The goal of the project is to build an ubiquitous computing based environment. Roads, cars and buildings that physically exist, plus all things that take up electronic space, cell phones, PDA, DMB devices are modeled in a formal manner and are interlinked. As a result a new 3D city space is created. u-City offers an infrastructure through which any information can be accessed anytime, from anywhere, using any device without any obstruction in a seamless connection.

One of u-City objectives is to provide a transportation system that allows an increased efficiency in the whole city as well as a decrease of the city operation fees. In particular we noticed that many users of GPS navigators complain about the inability of such systems to anticipate road conditions. At the same time, Intelligent Transportation Systems (ITS) are able to accumulate data from the sensors located on city roads and analyze them in real time, but it is currently not possible to embed an ITS in a GPS navigator.

So we decided to realize a u-City service that interlinks ITS with GPS navigators and exploits semantic technology to provide a new type of transportation service that increases the satisfaction of the urban inhabitants (see Figure 1). Such service is able to change itinerary, to recommend different road itineraries, and to send information to ITS for further automatic actions.

We developed rules in F-logic for ITS, by expressing car, road and traffic conditions as ontology assertions and axioms. Our ontology is similar, in terms of capability of supporting the computing of the best trip, to the one [7] developed by the European FP6 Project REVERSE (IST-506779), but it is less focused on general terms and it has more sophisticated features allowing real-time traffic management. To exemplify how we use rules, in the following listing we show the one we used for calculating expected traffic resolution time when a traffic accident occurred.

⁶ See for example http://www.udongtan.or.kr/english/cyber/cyb_01_7.aspx

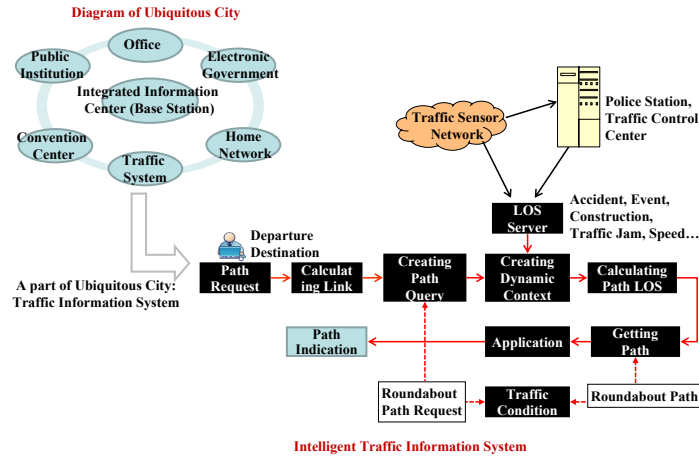


Fig. 1. Overall system flow of ITS

IF takeIn(TrafficAccident, Link) AND hasCarSituation(TrafficAccident, CarSituation(fire)) AND hasRelatedTrafficAccidentStat(TrafficAccident, TrafficAccidentStat) AND hasCarSituation(TrafficAccidentStat, CarSituation(fire)) AND solvedTime(TrafficAccidentStat, X) AND necessaryTime(TrafficAccidentStat, Y) AND add(X,Y,Z)
THEN ofEvent(TrafficAccidentSolvedTime, TrafficAccident) AND relatedLink(TrafficAccidentSolvedTime, Link) AND hasEventTime(TrafficAccidentSolvedTime, TrafficEventTime) AND hasRelatedTrafficAccidentStat(TrafficAccidentSolvedTime, TrafficAccidentStat) AND solvedTime(TrafficAccidentSolvedTime, Z)

The u-City navigation system takes the departure and destination points and, in real time, suggests to the driver alternative routes based on the real situation. The system proved to be effective in various traffic conditions and it expanded the possibilities of a traditional navigation service.

Our conclusions are twofold; on the one hand we are convinced that our ontology based system is an extremely effective navigation method; on the other hand we face significant technological challenges in terms of scalability and reasoning performance.

Siemens Previous Experiences in the Field. The department of learning systems at Siemens Corporate Technology (<https://www.ct.siemens.com/>) has a lot of research and project experience in the field of traffic modeling and forecasting. In the last 15 years they have been using machine learning techniques, e.g., fuzzy clustering, neural networks and reinforcement-learning to model, predict and optimize traffic flows. The following papers represent only a brief summary of our work.

In 1995 Hellendorn and Baudrex1 combined fuzzy methods and feedforward neural networks to control and to forecast traffic [8]. They built a fuzzy system for traffic flow control and incident recognition that has been in use for some time. Furthermore, the system was used in forecasting whether a particular parking garage is full or not.

An approach to traffic modeling is presented by Wagner et al. in 1996 [9]. The authors derived a second-order traffic flow model from microscopic equations. The model incorporated different driver characteristics on the microscopic level called the desired velocity. The authors explored dynamical quantities for the mean and variance of the desired velocity, and the covariance between actual and desired velocity. Through these quantities an alternative explanation for the onset of traffic clusters can be given, i.e., a spatial variation of the variance of the desired velocity can cause the formation of a traffic jam.

Lenz et al. extended the microscopic car-following model by Bando et al. by incorporating multi-vehicle interactions [10]. The authors showed that the reaction to more than one vehicle ahead leads to a stabilization of the dynamical behavior, i.e., the stable region increases. The fundamental macroscopic properties of traffic, namely free flow and congested flow, were still described. Due to the multi-anticipative driving behaviour driving in narrow platoons is forced such that a third fundamental property of traffic flow, the so-called synchronized flow is modelled as well. Related work on traffic modelling can be found in [11]. Here, anticipative schemes are used to switching between speed limits based on the density of the downstream segment.

Stutz and Runkler classified and predicted road traffic by using an application specific fuzzy clustering approach [12]. The authors used fuzzy methods for traffic data analysis. The results of the data analysis were classification and prediction systems. The work was focused on fuzzy clustering methods. The known clustering models were extended to: constrained prototypes, the use of a mix of different prototypes for one data set, partial supervision of the clustering, and the estimation of the number of clusters by cluster merging. A successful application example was given for the classification of traffic jam on a German motor highway.

Within the public funded research project LEONET⁷ Appl and Sollacher used adaptive learning systems for intelligent traffic control. In the paper [13] the authors described the application of modern reinforcement-learning approaches for the automatic control of groups of traffic lights. As an extension to this work Sollacher and Klein applied feedforward neural networks in combination with information of origin-destination traffic flows to forecast upcoming traffic volumes in short-term (up to one week in 15 min. time buckets). The feedforward neural network incorporated a so-called bottleneck coordinate transformation to cluster the daily traffic variation curves.

The bottleneck enables us to identify the non-linear principal components of the traffic variation curves. In order to predict the future development of the traffic volume the neural network concentrates on the principal components. The dimensionality of the forecasting problem is therefore dramatically reduced. The principal components are forecasted by taking into account external influences. The reconstruction of the traffic variation curve for every time step of the future is assured by the reuse of the bottleneck network [14]. A long-term forecast model

⁷ <http://www.inb.uni-luebeck.de/research/leonet>

has also been developed on the basis of this clustering approach of daily traffic variation curves and the estimation of origins and destinations.

5 Requirements

In this section, we will investigate requirements of Urban Computing. As argued before, we are particularly interested in the reasoning requirements for LarKC, but we believe such requirements are interesting for the entire community working on the complex relationship of the Internet with space, places, people and content.

Coping with Heterogeneity. Dealing with heterogeneous data has been a challenge for a long time in many areas in computer science and engineering, which include database systems, multimedia applications, network systems, and artificial intelligence. Here, we would like to propose a comprehensive notion of heterogeneity processing for semantic technologies. We distinguish the following different levels of heterogeneity: Representational Heterogeneity, Semantic Heterogeneity, and Default Heterogeneity.

Representational Heterogeneity means semantic data are represented by using different specification languages. Systems supporting Representational Heterogeneity would allow semantic data specified by multiple semantic languages, rather than using a single metadata or ontology language, like OWL or RDF/RDFS. However, note that different representation of semantic data does not necessarily mean that they have different semantics. The problem of merging and aligning ontologies is a structural problem of knowledge engineering and it is always considered when developing an application of semantic technologies.

Urban Computing-related data can come from different and independent data sources, which can be developed with traditional technologies and modeling methods (e.g., relational DBMS) or expressed with “semantic” formats and languages (e.g., RDF/S, OWL, WSML); for example, geographic data are usually expressed in some geographic standard⁸, events details are published on the Web in a variety of forms, traffic data are stored in databases; etc. The integration and reuse of those data, therefore, need a process of conversion/translation for the data to become useful together.

Reasoning Heterogeneity means the systems allow for multiple paradigms of reasoners. For instance, many applications of Urban Computing may need different reasoners for temporal reasoning, spatial reasoning, and causal reasoning. However, it does not necessarily mean that we have to develop a single but powerful reasoner which can cover all of those reasoning tasks. A system which supports Reasoning Heterogeneity would find a way to allow multiple single-paradigm-based reasoners to achieve the result of Reasoning Heterogeneity.

Some data related to Urban Computing need precise and consistent inference; e.g., knowing if two roads are connected for a given kind of vehicle; telling that

⁸ http://en.wikipedia.org/wiki/Geographic_Data_Files

at a given junction all vehicles, but public transportation ones, must go straight; checking if private cars are allowed to enter a specific urban area. Other data need approximate reasoning or imperfect estimations; e.g., calculating the probability of a traffic jam given the current traffic conditions and the past history. Therefore, the requirement is for different kinds of techniques and reasoners to deal with those kinds of data; moreover, another requirement is for a system which dynamically selects and runs a specific reasoner on the basis of the available data and the desired processing tasks.

By **Default Heterogeneity**, we mean that systems support for various specification defaults of semantic data. Well-known specification defaults of semantic data are closed world assumption, open world assumption, unique name assumption and non-unique name assumption. In the Semantic Web community, it is widely accepted that semantic data for the Web should take the open world assumption and the non-unique name assumption, as taken by the popular ontology language OWL.

However, as we have observed in many applications of Urban Computing, we should not commit to any single specification default. Take the example of traffic and transportation ontologies: although in many cases we can take the open world assumption and non-unique name assumption, because of our limited knowledge and information about the data, sometimes it is much convenient to take a *local* closed world assumption. For example, for a time table of a bus station, it is well reasonable to assume that the information about the bus schedule in the time table is locally complete, in the sense that if you cannot find any information about a bus which is scheduled at specific time, it would mean that there are no bus scheduled for that time. The same scenario is also applied to a city map: if there is no information which states a road connects two streets directly on the map, that would mean that there is no road which connects those two streets directly.

The examples above show that the semantic systems of Urban Computing should support multiple specification defaults. It should allow users or knowledge engineers feel free to state any data with any reasoning assumption. Some part of semantic data may be based on the open world assumption, and some part may be based on the closed world assumption.

Coping with Scale. The advent of Pervasive Computing and Web 2.0 technologies led to a constantly growing amount of data about urban environments, like information coming from multiple sensors (traffic detectors, public transportation, pollution monitors, etc.) as well as from citizens' observation (black points, commercial activities' ratings, events organization, etc.). The result, however, is that the amount of data available to be used and integrated is not manageable by state-of-the-art technologies and tools and a severe focus on scalability issues must be taken into account. For example, intelligent methods for data sampling or selection should be adopted before employing traditional reasoning techniques, e.g. to select traffic data to employ in predictions.

Although we encounter large scale data which are not manageable, it does not necessarily mean that we have to deal with all of the data simultaneously. Usually, there are only very limited amounts of data which are relevant for a single query/processing for a specific application. For example, when Carlo is driving to Taramelli, Milano, only part of the Milano map data are relevant. Furthermore, it is impossible and unnecessary to store large scale data at the same memory level, we only need a portion to be easily accessible for the processing. The idea of the data scheduling is to move relevant data from a memory level to another memory level in advance to make them of easier access. For example, when Carlo during the journey gets close to his final destination, the local parking information may become active by a prediction of the causal relation between driving and parking. We consider this idea of the data scheduling as a partial solution for the scalability, which will be discussed in the next section.

Coping with time-dependency. Knowledge and data can change over the time. For instance, in Urban Computing names of streets, landmarks, kind of events, etc. change very slowly, whereas the number of cars that go through a traffic detector in five minutes changes very quickly. This means that the system must have the notion of "observation period", defined as the period when the system is subject to querying.

Moreover the system, within a given observation period, must consider the following four different types of knowledge and data:

- *Invariable knowledge*:
 - it includes obvious terminological knowledge (as an address made up by a street name, a civic number, a city name and a ZIP code) and
 - less obvious nomological knowledge that describes how the world is expected to be (e.g., given traffic lights are switched off or certain streets are closed during the night) or to evolve (e.g., traffic jams appear more often when it rains or when important sport events take place).
- *Invariable data* do not change in the observation period, e.g. the names and lengths of the roads.
- *Periodically changing data* change according to a temporal law that can be
 - Pure periodic law, e.g. every night at 10pm Milano overpasses close.
 - Probabilistic law, e.g. traffic jams appear in the west side of Milano due to bad weather or when San Siro stadium hosts a soccer match.
- *Event driven changing data* are updated as a consequence of some external event. They can be further characterized by the mean time between changes:
 - Fast, e.g. the intensity of traffic for each street in a city;
 - Medium, e.g. roads closed for accidents or congestion due to traffic;
 - Slow, e.g. roads closed for scheduled works.

Coping with Noisy, Uncertain and Inconsistent Data. We distinguish the following different types of data:

- Noisy Data: part of data are useless or semantically meaningless.

- Inconsistent Data: parts of data are in logical contradiction with each another, or are semantically impossible.
- Uncertain data: the semantics of data are partial, incomplete, or they are conceptually arranged into a range with multiple possibilities.

Traffic data are a very good example of such data. Different sensors observing the same road area give apparently inconsistent information. For example, a traffic camera may say that the road is empty whereas an inductive loop traffic detector may tell 100 vehicles went over it. The two information may be coherent if one considers that a traffic camera transmits an image per second with a delay of 15-30 seconds, whereas a traffic detector tells the number of vehicles that went over it in 5 minutes and the information may arrive 5-10 minutes later.

Moreover, a single data coming from a sensor in a given moment may have no certain meaning. For example, consider an inductive loop traffic detector, it tells you 0 cars went over it, what does it mean? Is the road empty? Is the traffic completely stuck? Did somebody park the car above the sensor? Is the sensor broken? Combining multiple information from multiple sensors in a given time window can be the only reasonable way to reduce the uncertainty.

6 Partial Solutions

Within the LarKC, we are envisioning a set of partial solutions to address the challenges of Urban Computing including: Traffic Prediction using recurrent neural networks, Data Scheduling to address scalability and Stream Reasoning to address time-dependency.

Predicting Traffic Using Recurrent Neural Networks. Given that a forecast model should focus on the underlying dynamics of the traffic flow and external influences on the traffic volume should be incorporated in the model, we intend to use time-delay recurrent neural networks for the traffic predictions [15]. With this approach we presume that the traffic volume is the outcome of an open dynamical system which combines an autonomous development with external influences (e.g. calendar effects, special events etc.). Recurrent neural networks offer a new way to model (nonlinear, high dimensional) open dynamical systems based on time series data. Our recurrent neural networks are formulated as state space models in discrete time to identify the traffic dynamics and the impact of the external influences. In state space formulation a recurrent neural network is described by a hidden state-transition- and an output-equation. The temporal equations are transformed into a spatial neural network architecture using shared weights (so-called unfolding in time). Prior knowledge about the application (e.g. topology of the traffic network or the temporal structure of the traffic flows) can be easily incorporated in the neural network architecture. For instance, an error correction mechanism can be used to consider the impact of unplanned construction sites, traffic accidents or holdups. This is also the key for robust forecasting [16].

Data Scheduling. The idea of data scheduling takes inspiration from memory management techniques developed and adopted in computer systems and

software engineering (e.g., garbage collection, memory caching and direct memory access). Large scale data are organized at different memory levels based on their relevance and on the context of applications: working data, which should be accessed by systems immediately without any over-heading cost; neighboring data, which can be accessed by the system with a moderate cost; and remote data, which can be accessed by the system with a significant amount of cost. The research problem is finding automatic ways to move data from higher access cost memory into lower access cost memory and vice versa. Such memory shift should take place in parallel with reasoning.

Stream Reasoning. Periodically changing data and event driven changing data are best represented as data streams. Processing of data streams has been largely investigated in the last decade [17] and specialized systems have been developed. While reasoners are year after year scaling up in the classical, time invariant domain of ontological knowledge, reasoning upon rapidly changing information has been neglected or forgotten. By coupling reasoners with powerful, reactive, throughput-efficient stream management systems, we introduce the concept of Stream Reasoning [18]. We expect future realization of such a concept to have a strong impact on Urban Computing because it enables reasoning in real time, at a throughput and with a reactivity not obtained in previous works.

7 Vision and Conclusions

In this paper we focus on presenting the Urban Computing challenge and in particular some requirements for future mobility management systems. We also presented some novel multi-disciplinary ideas about ways to address the Urban Computing challenge by partially satisfying one or more requirements. More solutions and, in particular, broader ones should be explored. As a matter of fact, if we were able to cope with requirements present in Section 5 we would be able to solve a broad range of Urban Computing problems.

City Planning. Urban Computing applications can extract statistics and synthetic descriptions of citizens' movements, habits and opinions in order to position new housing complex, office buildings, shops, parking lots, green areas and to optimize public and private transportation routes and timetables. The City Planning can also lower pollution and enhance energy savings.

Tourism and Culture. Urban Computing applications analyze tourists' movements and enhance the appeal of current places of interest and create targeted promotional campaigns to increase tourism.

Public Safety. Urban Computing applications can perform continuous statistical analysis of people movements to find abnormal behavior and correlate them with the ones coming from law enforcement and public protection forces to enhance city safety.

Acknowledgments

This research has been partially supported by the LarKC EU-funded project (FP7-215535). For more information visit <http://www.larkc.eu>.

References

1. Kindberg, T., Chalmers, M., Paulos, E.: Guest editors' introduction: Urban computing. *IEEE Pervasive Computing* **6**(3) (2007) 18–20
2. Arikawa, M., Konomi, S., Ohnishi, K.: Navitime: Supporting pedestrian navigation in the real world. *IEEE Pervasive Computing* **6**(3) (2007) 21–29
3. Reades, J., Calabrese, F., Sevtsuk, A., Ratti, C.: Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing* **6**(3) (2007) 30–38
4. Bassoli, A., Brewer, J., Martin, K., Dourish, P., Mainwaring, S.: Underground aesthetics: Rethinking urban computing. *IEEE Pervasive Computing* **6**(3) (2007) 39–45
5. Fensel, D., van Harmelen, F., Andersson, B., Brennan, P., Cunningham, H., Della Valle, E., Fischer, F., Huang, Z., Kiryakov, A., il Lee, T.K., School, L., Tresp, V., Wesner, S., Witbrock, M., Zhong, N.: Towards larkc: a platform for web-scale reasoning, *IEEE International Conference on Semantic Computing* (2008)
6. Fensel, D., van Harmelen, F.: Unifying reasoning and search to web scale. *IEEE Internet Computing* **11**(2) (2007)
7. : Ontology of Transportation Networks. Technical report, REWERSE Project (2005)
8. Hellendoorn, H., Baudrexel, R.: Fuzzy neural traffic control and forecasting. In: *Fuzzy Systems, 1995. International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium., Proceedings of 1995 IEEE International Conference on.* Volume 4. (1995) 2187–2194
9. Hoffmann, C., Sollacher, R., Wagenhuber, J., Schürmann, B.: Second-order continuum traffic flow model. In: *Phys. Rev. E* **54**. (1996) 5073 – 5085
10. Lenz, H., Wagner, C.K., Sollacher, R.: Multi-anticipative car-following model. In: *Eur. Phys. J. B* **7**. (1998) 331–335
11. Lenz, H., Sollacher, R., Lang, M.: Standing waves and the influence of speed limits. In: *Proceedings of the European Control Conference 2001, Porto, Portugal* (2001)
12. Stutz, C., Runkler, T.: Classification and prediction of road traffic using application-specific fuzzy clustering. In: *IEEE Transactions on Fuzzy Systems.* Volume 10. (2002) 297–308
13. Appl, M., Sollacher, R.: Intelligent traffic control with adaptive, learning systems. In: *Automatisierungstechnik.* Volume 49. (2001) 512
14. Zimmermann, H.G., Neuneier, R., Grothmann, R.: Modeling of the german yield curve by error correction neural networks. In Leung, K., Chan, L.W., Meng, H., eds.: *Proceedings Intelligent Data Engineering and Automated Learning 2000 (IDEAL 2000), Hong Kong* (2000) 262–267
15. Zimmermann, H.G., Neuneier, R.: Modeling dynamical systems by recurrent neural networks. In Ebecken, N., Brebbia, C., eds.: *Data Mining II, WIT Press* (2000) 557–566
16. Zimmermann, H.G., Neuneier, R., Grothmann, R.: Modeling of dynamical systems by error correction neural networks. In Soofi, A., Cao, L., eds.: *Modeling and Forecasting Financial Data, Techniques of Nonlinear Dynamics, Kluwer Academic Publishers* (2002)
17. Garofalakis, M., Gehrke, J., Rastogi, R.: *Data Stream Management: Processing High-Speed Data Streams (Data-Centric Systems and Applications).* Springer-Verlag New York, Inc., Secaucus, NJ, USA (2007)
18. Della Valle, E., Ceri, S., Barbieri, D.F., Braga, D., Campi, A.: A first step towards stream reasoning. In: *Proceedings of the Future Internet Symposium.* (2008)