Reduction

2011-2014

Deliverable 2.1

Report with Data Flow Analysis and System Architecture

30-12-2012

Project acronym: Reduction
Project full title: Reducing Environmental Footprint based on Multi-Modal Fleet management Systems for Eco-Routing and Driver Behaviour Adaptation

Work Package: 2

Document title: **Report with data flow analysis and system architecture**
Version: 1.1
Official delivery date: 2012.09.01
Actual publication date: 2012.12.30
Type of document: Report
Nature: Public

Authors: Yubin Wang(TRI), Marcel Valé(TRI), Josif Grabocka(UHI), Umer Khan(UHI), Lars Schmidt-Thieme(UHI)

Approved by: ____________________________

Version	**Date**	**Sections Affected**
1.0	2012.09.01	Initial version
1.1	2012.12.04	Review comments processed
1.2	____________	Updated to reflect changes in the implementation
1.3	____________	Updated to reflect changes in the implementation

Executive Summary

In this deliverable, the functionalities from WP2 are summarized based on the novel algorithms for creating predictive analytics models. The main functionalities are Eco-Routing, Eco-Driving and Distributed Data Mining. Then the requirements which are used for system architecture design are collected from the REDUCTION partners. Based on the requirements, the fundamental and optional elements of the predictive analytics models are analyzed and the system architecture is proposed. A system can be designed and implemented based on the architecture. Next, a scenario is described to explain how the system interactive with users, such as on-line feedback to drivers, justify heuristic approach to eco-driving and data security issues. In the end, an alternative approach by using estimation models is proposed to estimate fuel consumption and CO2 emission in case of CANBus data is not possible to get such that fuel consumption and CO2 emission are no possible to measure directly.

Contents

1 Introduction **7**
- 1.1 Project Overview 7
- 1.2 Work Package Objectives and Tasks 8
- 1.3 Objective of this Deliverable 8

2 Related Work and Reduction **9**
- 2.1 ITS Tasks and the Ecological Impact 9
- 2.2 Eco-Routing 10
- 2.3 Eco-Driving 11
- 2.4 Distributed Data Mining 13
  - 2.4.1 Test Instance Propagation and Consensus based Data Mining 13
  - 2.4.2 Model Propagation with-in Local Neighborhood 14

3 Framework and Methodology **16**
- 3.1 Requirements specification 16
- 3.2 Software architecture 16
- 3.3 Data flow analysis 18
  - 3.3.1 Eco-driving 18
  - 3.3.2 Safe driving 19
  - 3.3.3 Personalized Eco-routing 19

4 Scenario **19**

5 Risk Assessment **21**

6 Conclusion **22**

List of Figures

1 System Archiecture 17
2 Objects relations in DSS. 18
3 Personal Eco-Drive Assistance (Courtesy: Co Eco-Driving: Pilot Evaluation of Driving Behavior Changes among U.S. Drivers by Boriboonsomsin et al. University of California). 20

1 Introduction

The analysis [32] shows that a bad driving behaviour decreases the road capacity and increases GHG emissions about 20% to 30%. In [32], the driving behavior in big cities of China is compared with the driving behavior in two Dutch cities. The long and irregular time headways and sudden lane changing are observed in Chinese cities, it is due to the fact that Chinese drivers adapt themselves to the local conditions and behave differently from Dutch drivers, giving a less efficient traffic system and high fuel consumption. Furthermore, traffic becomes more congested due to the limitation of the road capacity. Therefore, a better driving behavior is expected to result in more safety, less queues, less congestion and a larger reserve capacity, which would reduce GHG emissions.

1.1 Project Overview

Reduction of CO2 emissions is the great challenge of the transport sector nowadays. Despite progress in vehicle manufacturing and fuel technology, additional innovative technologies are needed to address this challenge. According to the International Association of Public Transport, a significant fraction of CO2 emissions in EU cities is resulting from public transport and other mass transport means, which are commonly organized into multi-modal transport fleets, because their vehicles have, on average, nearly substantial mileage and fuel consumption. The REDUCTION project focuses on advanced ICT solutions for managing multi-modal fleets and reducing their environmental footprint. REDUCTION collects historic and real-time data about driving behavior, routing information, and emissions measurements, that are processed by advanced predictive analytics to enable fleets enhancing their current services as follows: 1) Optimizing driving behavior: supporting effective decision making for the enhancement of drivers education and the formation of effective policies about optimal traffic operations (speeding, braking, etc.), based on the analytical results of the data that associate driving-behavior patterns with CO2 emissions. 2) Eco-routing: suggesting environmental-friendly routes and allowing multi-modal fleets to reduce their overall mileage automatically. 3) Support for multi-modality: offering a transparent way to support multiple transportation modes and enabling co-modality.

REDUCTION follows an interdisciplinary approach and brings together expertise from several communities. Its innovative, decentralized architecture allows scalability to large fleets by combining both V2V and V2I approaches. Its planned commercial exploitation, based on its proposed cutting-edge technology, aims at providing a major breakthrough in the fast growing market of services for ”green” fleets in EU and worldwide, and present substantial impact to the challenging environmental goals of EU.

1.2 Work Package Objectives and Tasks

The main objective of work package 2 is to develop novel algorithms for creating predictive analytics models that will operate in the decentralized environment of REDUCTION. The proposed prediction models in the work package will enable the generation of knowledge for supporting driver-behavior adaptation in order to educate drivers about ways of energy-efficient driving both off-line and on-line. The reason why both off-line and on-line are interesting for optimal driving behaviour is that drivers can be trained off-line and advices can be given to them on-line directly.

1.3 Objective of this Deliverable

In this deliverable, the requirements (which will also be used in the system integration in work package 4) are collected; the software architecture (that will be developed in T2.2) is defined; the fundamental and optional elements of the predictive analytics models (that will be implemented in work package 5) are analysed; a scenario is described to explain on-line feedback to drivers, justify heuristic approach to eco-driving and data security issues; the fallback plan of use-case failures or data unavailability is discussed.

2 Related Work and Reduction

2.1 ITS Tasks and the Ecological Impact

In deliverable D2.2 [17], we are presenting a list of problem categories related to ITS, including the description of the scope and short briefings of the task, while providing references for interested readers. Once the list is presented, our aim is to identify a subset of relevant problems with thee highest possible fuel consumption impact. The list of ITS tasks' categories, reflecting our intensive related work research within this deliverable, is as follows:

- **Routing (Path Finding):** Finding the optimal or quasi-optimal path among two destinations in order to optimize an overall quality metric. Most applications address total distance as the quality metric to be minimized [50], even though paths that optimize time have also been addressed [11, 45].
- **Trajectory Planning:** Reasonably large researcher have elaborated the task of predicting the movement trajectory of a vehicle. Efficient predictions bring benefits, among which traffic congestions avoidance and road safety increase [16].
- **Eco-Driving:**

  Eco-driving is a generalist term used to describe the adaptation of driving behaviors, through providing informatory or cautionary feedback to the driver. The feedback typically represents the result of methodological evaluation of the driving behavior data [30].
- **Collision Avoidance**

  Car accidents represent a major concern for our societies, therefore automatic avoidance of the vehicle collisions has gathered important focus within the ITS community.

  Lane Change: Predicting the intent of a driver to change lane during highway drivings, is a current hot topic for the ITS community. The ultimate goal relies in increasing safety of life and avoiding collisions [37, 40].
- **Travel Time Estimation**

  Estimating the travel time helps people plan their activities better and increase the productivity of the time usage. Aside the obvious practical aspect, accurate travel time prediction can be combined with routing and path finding, yielding time-optimized path calculation [31, 29].
- **Formal Vehicle and Traffic Modeling**

  A special dedication was delivered to formal modeling of vehicle and traffic. Formal models of vehicles typically model the motion dynamics via a set of mathematical equations, by using a set of parameters [55]. Larger scale modeling refer to the mathematical formalization of the traffic flow.

2.2 Eco-Routing

The effect of route choice in reducing emissions has been addressed by several authors. In [46] authors have discussed main routing algorithms and presented the Coolest path algorithm which enables multi-criteria personalization based on travel distance, travel time, points of interest, and path simplicity. The coolest path approach allows users to set the level of importance for each of these criteria. Moreover, a prototype was built in order to validate the proposed ideas. [7] presented a novel personalized route planning framework that considers user movement behaviours. The proposed framework comprises familiar road network construction and route planning. In the first component, familiar road segments are mined from a driver's historical trajectory dataset, and a familiar road network is constructed. For the second component, an efficient route planning algorithm is proposed to generate the top-k familiar routes given a start point and a destination point.

[20] presents a list of the most relevant studies carried out in the field of route choice optimization, considering energy and air quality. Furthermore, it indicates the methodology for emissions and fuel use estimation (m micro-scale models; MMacro; MeMeso, ffield measurement) and whether route characteristics (RC) (type of road, incidents, traffic signals, neighborhood, safety) are considered, yes (y) or not (n). These studies can be divided into two groups. In [51, 38, 13], the authors developed mathematical formulations to assign traffic on a virtual network considering air quality. In a recent study, mathematical programs integrating emission objective into system wide travel time minimization were developed. It was found that in an idealized system optimum (SO) scenario, CO emissions reduction and travel time minimization can occur when drivers chose longer routes with low speed profiles instead of all users selecting the route with the shortest travel time, i.e., user equilibrium. Another study developed a theoretical emissions-optimized (EO) traffic assignment model. First the authors focused on standard user-equilibrium (UE) and system-optimal (SO) assignment methods, and then they derived two new functions to characterize link emissions and vehicle emissions. It was found that under EO conditions the percentage of traffic assigned to freeways is very low since emissions rates are extremely high at freeway free-flow speeds. Moreover, the EO assignment is most effective when the network is under low to moderately congested conditions [39]. Other authors employed different solutions such as Genetic Algorithms [12] to solve complicated optimization problems with non-linear terms. Overall, reductions in emissions can be achieved, if travelers choose eco-friendly routes. However, in the extreme case of everyone choosing an eco-friendly route, a shift to all-or-nothing assignment from UE assignment may occur.

In the second group [20, 15, 36, 52] field experiments were conducted and a wide range of models were applied to evaluate the impact of route selection in terms of emissions and energy use, over several case-studies. The research of Rakha, Frey and Barth should be highlighted for different reasons. Ahn and Rakha [3] studied two alternative routes using different type of emission models which allowed concluding that macroscopic emission estimation tools can yield incorrect conclusions. In 2011, Rakha et al. also presented a framework for modeling eco-routing strategies [43]. Barth et al. developed and patented an environmentally-friendly navigation system [42]. Frey et al. [15] carried out experiments using a portable emission measurement system (PEMS) under real world driving cycles. Then, in order to standardize the comparisons of emission rates for different vehicles and routes, the authors employed the Vehicle

Specific Power (VSP) approach to characterize fuel use and emissions [15]. VSP variable can explain a significant portion of variability in fuel consumption and emissions [9]. The majority of studies have concluded that route choice has a significant impact on emissions and energy use. However, few studies have addressed the effect of congested periods on emissions [56]. The distribution of vehicle speeds and accelerations in traffic vary by type of road facility and amount of traffic volume, generating large discrepancies in emission levels [41]. Possibly, this fact has contributed to some inconsistency on literature about this issue. On the one hand, some studies pointed out that time minimization paths often also minimize energy use and emissions [15, 42]. On the other hand, research demonstrated that frequently the faster alternatives are not the best from an environmental perspective [21, 3]. What has arisen from literature review is that it is not possible to generalize conclusions, considering the limited study areas, and thus more research is needed to evaluate a wider range of driving patterns conditions, namely at different periods of the day. A more extensive analysis including different scales, and different traffic volumes, as performed here, may better reflect the reality and improve the knowledge to develop further traffic management strategies.

2.3 Eco-Driving

There has been significant progress in realizing fuel economy transportation systems through the development of hybrid cars, advanced engine control systems, solar- or electric-power-driven cars, road infrastructure, and many other aspects to reduce the impact on the ecological balance of the earth. The following sections are based on the survey of Kamal et al. [24].

Eco-driving is a way of maneuvering a vehicle with a human driver that is intended to minimize fuel consumption while coping with varying and uncertain road traffic by trading off the most efficient driving point of the vehicle whenever necessary. The positive impact of influencing the driver behaviour to a more economic driving style could already be demonstrated in several studies. For example, Changxu Wu et al. [54] showed that especially in urban driving conditions an improved accelerating and decelerating behavior through the use of an in-car application can reduce fuel consumption by 12 31%. Some vehicle manufacturers already have solutions in place delivering trip information post-driving, aiming to improve driving behaviour [54]. FIAT, for example, could demonstrate that overall fuel savings of up to 16% could be realised in the short term [1]. A recent experiment conducted on urban roadways showed that the reduction in fuel consumption can be as high as 25% [49]. Generally, fuel economy is maximized when accelerating and braking events are minimized [14, 2]. Therefore, a fuel-efficient or ecological strategy is to anticipate what is happening ahead, drive with acceleration and braking as little as possible, cruise at the optimal velocity, and slowly decelerate at stops. A driver can easily maneuver a car on a road tackling uncertain, discontinuous, and complex traffic situations using his perception of the environment. For optimal eco-driving, the same driver needs to be more anticipative on road-traffic situations with perfect knowledge of the engine dynamics of his car, which is hardly attainable by a human driver. Therefore, the driver can be technologically assisted to ecologically drive his car.

Recently, various assistance schemes for eco-driving have emerged. Speculative features of ecological driving are available in the form of driving tips [14, 2]. Some recently manufactured cars have an ecological indicator that shows a green ECO mark to a driver when it consumes little or no fuel. A driver would find his driving as ecological only when he maintains a steady velocity at a reasonable level or brakes the car. Nissan launched an off-board eco-driving support service for some users in which, after the driving record is sent to a telemetric data center for offline analysis, advice is sent to the driver to improve his driving style the next time. Based on past performance, they have proposed an on-board assist system to motivate the driver to ecologically drive the car by showing his comparative driving efficiency, his position in fuel composition ranking, etc... [22]. Recent work in determining ecological strategy uses an optimal control approach in which only the model of the engine in terms of velocity, gear ratio, and load is considered [44]. A more realistic approach to assisting a driver uses information of traffic signal, jams, road gradient, and distance between cars, and the advice is given in a very rough form such as keep driving or reduce pressure on pedal, depending on the motivation of the driver [48]. However, existing approaches to eco-driving assistance are very superficial; they do not provide concrete information such as the level of velocity or acceleration required for long-term fuel-efficient driving by analyzing current vehicleroadtraffic situation and its trend. Recently, an ecological driver-assistance system (EDAS) that can guide a driver for fuel-efficient driving on a flat urban road has been proposed [23]. Using current roadtraffic information, the EDAS anticipates the future states of vehicles using their dynamic models, and based on the anticipated future states and the fuel consumption model of the engine, it calculates the optimum vehicle control input required for ecological driving. This unique approach is found to be very promising on a flat urban road. Countries such as Japan have many hilly areas, where many roads have many sections with updown slopes. Fuel consumption significantly varies on such a road with updown slopes [34]. Some earlier approaches relating to the optimal driving on roads with varying grades mainly used dynamic programming to solve a known drive cycle [19, 27], which are not suitably applicable with immediately perceived road grade information. This paper extends the scope of the EDAS to roads with updown slopes that are neither crowded nor have too many traffic signals, which is a scenario typically found in non-urban areas. Using information about road gradients and anticipation of the vehicles future state on the road, it generates control inputs that are required for optimally fuel-efficient driving. The road gradient angle is computed using road altitude data obtained from the digital road map and introduced in the optimization process in a simple way. Once the optimal control input becomes available, a suitable human interface may convey it to the driver in the case of an assistance system. For simplicity in this paper, however, any interface component is omitted, and the control input generated by the system is directly fed to the host vehicle, focusing mainly on the vehicle control aspects on hilly roads.

In [24], authors presented the concept of eco-driving system on hilly roads. They implemented an efficient way of using data from the digital road map. Optimum control inputs can only be computed through anticipated rigorous reasoning using information concerning road terrain, model of the vehicle dynamics, and fuel consumption characteristics. The developed system is comprised of a nonlinear model predictive control method with a fast optimization algorithm is implemented to derive the vehicle control inputs based on road gradient conditions obtained from digital road maps. The fuel consumption model of a typical vehicle is formulated using engine efficiency characteristics and used in the objective function to ensure fuel economy driving.

2.4 Distributed Data Mining

In this section, at first we mention about state-of-the-art in distributed data mining in peer-to-peer (P2P) and wireless ad-hoc networks (WANETS) in general. Later we present related work about application of data mining in VANETS.

Approaches to distributed data mining have focused on developing some primitive operations as well as more complicated data analysis/mining algorithms. Researchers have developed several different approaches for computing primitive operations (average, sum, max, random sampling) on P2P networks [10]. For example, Kempe et al. [26] investigate gossip based randomized algorithms. They prove that the error will go to zero in probability if the algorithm runs uninterrupted. Jelasity and Eiben [28] develop the newscast model as part of the DREAM project1. They rely on empirical accuracy results rather than guaranteed correctness. Both of the above approaches used an epidemic model of computation. Bawa et al. [6] have developed an approach in which similar primitives is evaluated to within an error margin. A main goal of these works is to lay a foundation for applications and more sophisticated data analysis/ mining algorithms (efficient complex algorithms can, in principle, be developed from the application of efficient primitives). The common feature of all the approaches mentioned so far is that they all require resources that scale directly with the size of the system. This feature distinguishes these from local algorithms. Such algorithms [5] computed their result using information from just a handful of nearby neighbours. Even still, it is possible to make definite claims regarding correctness. The resources required by these algorithms are independent of the size of the system in many cases. The obvious benefit is their superb scalability, which make them a good fit for networks spanning millions of peers. They are also very good at adjusting to failure and changes in the input locally, so far as the output need not change. However, a disadvantage of local algorithms is the limited class of functions to which they can apply.

In general, all the research work in the field of distributed data mining can be categorized by two kinds of approaches, as follows:

2.4.1 Test Instance Propagation and Consensus based Data Mining

In these approaches, a test instance in propagated with-in the network or just network neighborhood, to query intended peers or directly connected neighbors, to ask about their decision for this particular instance. The answering peer replies with the predicted class or value. Such approaches performs good in the sense that the communication cost for building the model is far less than the approaches which require a bundle of data to be sent in network. But on the other hand, these approaches suffer with larger response times and communication delays, when the classification task is frequently occurring.

These approaches include - Majority voting over an ensemble of classifiers [33, 53, 35, 8] - Incremental and reactive model building approaches by Continuously monitoring every change in data. Then they maintain a consensus about data statistics [10].

2.4.2 Model Propagation with-in Local Neighborhood

In these approaches, each peer builds a local classifier and propagates this model to only directly connected neighbors. Therefore, it also receives learned models of its neighbors, and merge them with its own local knowledge. An obvious hypothesis is that local knowledge of each peer will be enhances and so is the global performance of this network neighborhood. These approaches are communication costly while building the neighborhood's global classifier. But once the model updation is complete, the classification task can be performed fast and frequently. The actual challenge in these approaches is to ensure least communication cost while keeping the accuracy as close to as centralized learning. Another goal is, instead of performing some compression before transmitting models on network, how we can respect the representation of each local training instance at a peer. These goals really make such approaches interesting and challenging.

These approaches include, the use of reduced or compressed form of model to be propagated, where the target is achieving accuracy close to centralized approaches [4].

In terms of distributed data mining in VANETS, the most significant related work is by Kargupta et al. [25]. They have developed MineFleet, distributed vehicle performance data mining system designed for commercial fleets. MineFleet analyses high throughput data streams on-board the vehicle, generates the analytics, sends those to the remote server over the wide-area wireless networks and offers them to the fleet managers using stand-alone and web-based user-interface. The main unique characteristics of the MineFleet system that distinguish it from traditional data mining systems are as follows:

- Distributed mining of the multiple mobile data sources with little centralization of the data.
- On-board data stream management and mining using embedded computing devices.
- Designed to pay careful attention to the following important resource constraints:

  -Minimize data communication over the wide area wireless network.

  -Minimize on-board data storage and the footprint of the data stream mining software.
- Process high throughput data streams using resource constrained embedded computing environments.
- Respect privacy constraints of the data, whenever necessary.

It builds on the existing work on vehicle telematics. Existing vehicle telematics systems collect vehicle performance data and offer them to the fleet managers or vehicle owners. OnStar [1] for General Motors vehicles and Sync2 [2] from Ford are examples of such telematics systems. There are some major differences between the MineFleet and traditional telematics systems. Some of them are listed below:

[1] `www.onstar.com`

[2] `www.fordvehicles.com/technology/sync/`

Advanced data analytics: MineFleet is powered by advanced distributed data mining and statistical analysis algorithms. Most telematics systems are designed for in-car infotainment and security application based on relatively simple data management operations.

On-board data mining: MineFleet offers dramatic reduction of wireless communication by performing data analysis on-board the vehicle. Unlike most conventional telematics systems, MineFleet sends the results of the on-board analysis to the server over the wireless network, not the raw data.

Not a GPS-based tracking/navigation system: Unlike most conventional telematics devices MineFleet is primarily focused on vehicle performance data analysis not tracking and navigation.

3 Framework and Methodology

3.1 Requirements specification

A good requirements specification is critial to develope a successful system, therefore we have collected all possible requirements together with the REDUCTION partners for WP2. The requirements are based on our experiences to fullfil the functionalities within WP2. More detailed requirements from the actual users will be collected later when the users start to use the production which will be developed within the REDUCTION project. Imaging the scenario which is described in the later section, a driver will use a device in their car. All the functionalites such as Eco-Routing, Eco-Driving and Distributed Data Mining from this work package will be integrated in the device. Based on the functionalities, the following requirements are listed:

1. GPS data can be collected and stored in database while
2. GPS data have to be mapped to a digital map.
3. CO2 emission data from CANBus can be collected and stored in database.
4. A off-line advisory feedback is provided to the driver in response to his driving records.
5. An on-line advisory feedback is provided to the driver in response to his driving records, for example a warning message is displayed to the driver in case non Eco-friendly driving pattern was detected.
6. An on-line feedback is given to drivers to prevent events possible in the immediate road neighborhood in order to achieve safe and time efficient driving, for example a warning message is displayed to the driver in case an event is detected ahead.
7. Personalized Eco-routing is provided to deliver personalized route paths that would be eco-friendly by the drivers personal standards.

3.2 Software architecture

The software architecture that will be developed in T2.2 is described in this section. The architecture is displayed in Figure 1. The system architecture is based on the Trinité's system. However, the system architecture can be used by any other system with a distributed middleware. It consists of four components which are explained below.

DSS datapool Dynamic Subscribe System (DSS) datapool [47] is a real-time publis-subscribe distributed middleware. It provides a level of abstraction, by hiding the complexity of a variety of platforms (Different servers with different operational systems), networks and low-level process communication. Application developers may concentrate on the current requirements of the software to be developed, and use lower-level services provided by the middleware when necessary.

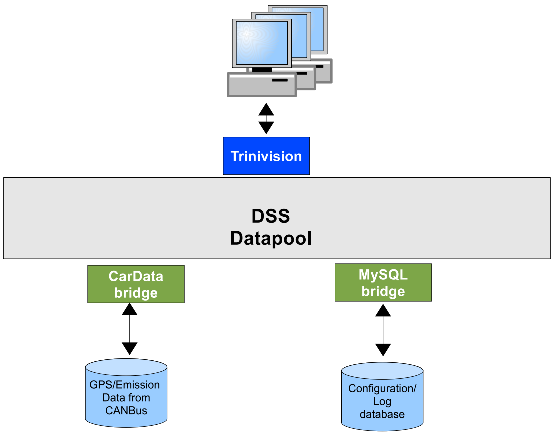

Figure 1: System Archiecture

Trinivision A standard DSS user interface which is presenting DSS visual objects. It can be used in both personal computers and smart phones via web application.

Database bridge The interface between the DSS Datapool and database. DSS uses it to read/write data from/to database. MySql database bridge is used within the Trinite enviroment.

CarData bridge The interface between the DSS Datapool and In-car systems. DSS uses it to read all kind of vehicle related data, such as GPS and CO2 emission data.

Besides the middleware functionality the DSS datapool also addresses the business logic. All above objects are implemented in the datapool. Figure 2 zoomed in into a small part of the DSS datapool. The balloons represent the objects, the arrows repersent the data flows between each object. Each bridge only connect one external hardware, so there are multiple CarData bridge to connect to many In-car systems and only one MySql bridge to connect one database. The objects within the DSS datapool are explained below.

CarObjectItem CarObjectItem is responsible for all the communications and providing data of the corresponding CarObject to the Desktop (User interface).

CarObject Carobject represents a vehicle. It gets real-time data of vehicle (including personal cars, trucks, busses, minibuses and other vehicle types) related data and stores the data in database. The prediction model in T2.2 will be applied in this object.

In general, the real-time data of vehicle related data is collected via CarData Bridge and stored both in the database and CarObject. The CarObject inlcuding the final result of the prediction model can be presented on user interface by using CarObjectItem.

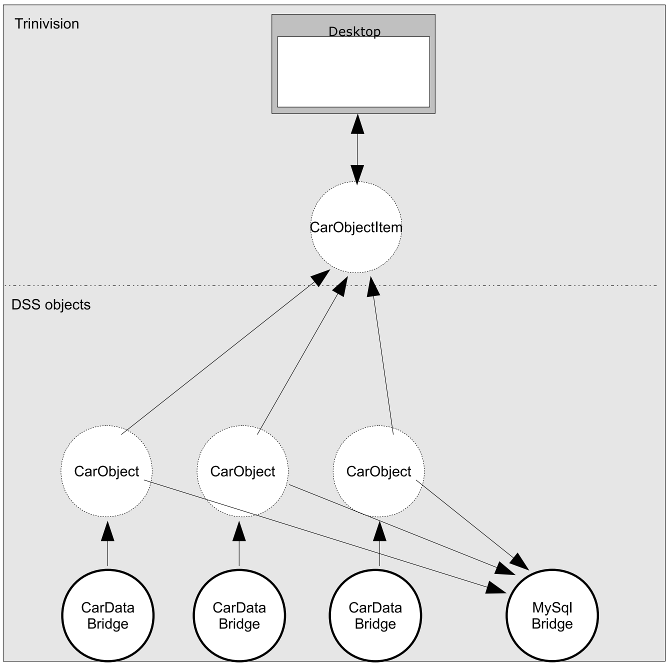

Figure 2: Objects relations in DSS.

3.3 Data flow analysis

In this section, the fundamental and optional elements of the predictive analytics models are analysed. Data flow of three main functionalities as follows:

3.3.1 Eco-driving

The data analysis of Eco-driving can be divided into two cases: the off-line case and the on-line case, since there are two options to deliver the feedback to the vehicle. The categorization is done based on the relative delay of the response. On-line feedback reflects the instantaneous messaging of the feedback to the vehicle, reflecting his behavior in short and recent data measurements. In typical applications the vehicle driver is able to adopt his behavior almost immediately. On the other side, the off-line analysis refers to providing feedback as a result of an analysis of a longer period of behavior data, typically days, weeks or months. In the end of the period, off-line analysis helps to detect trends, anomalies or regularities along the whole duration of analysis. The ultimate feedback is given periodically, so the driver can adopt the behavior in longer future terms. More details of two cases can be seen from [17].

The off-line case The vehicle related data including GPS and CO2 emission data are stored in database. The data could be accessed from a web interface. Then the data constituting the driving pattern over long-term distances will be collected and analyzed. Such off-line processing aims at detecting and identifying certain driving patterns of a driver in comparison to other drivers' behaviors under similar driving conditions. In the end of the analysis classification and cluster feedback will be provided to the drivers and the fleet managers. Then drivers can be educated to adapt their driving behavior off-line.

The on-line case The real-time vehicle related data are also available in the CarObject of the DSS datapool. The driving behavior will be described via collected signals and sensor inputs regarding vehicle CANBus messages, such as speed, accelerator, gear changes, brakes and fuel consumption. Our prediction model will use the recent messages of a specified time window and build an advisory method which outputs certain feedback warning messages to the driver in case non Eco-friendly driving pattern was detected. In-car devices such as a touch-pad device can be used to display the warning messages.

3.3.2 Safe driving

In addition to optimal driving behavior by means of the off-line and on-line feedback, safe and time efficient driving is also an important aspect of driving behavior. Therefore, a distributed data mining model is designed to give an on-line feedback to drivers that can predict events possible in the immediate road neighborhood. A local learning algorithm on each vehicle will receive position, velocity, acceleration, heading and yaw rate measurements from all connected vehicles in its immediate road neighborhood in order to predict future event in this road segment. The local algorithm will use this data to learn a model periodically, and each vehicle will also share its decision/prediction about a certain event.

3.3.3 Personalized Eco-routing

Futhermore, personalized Eco-routing consists of a set of methods which compute an estimated itinerary between requested endpoints, such that an overall fuel-efficient objective function is optimized [17]. Each of the driver's driving history (including travelled road segments and used time to pass across the segments) is stored in database. This will be done automatically by recording GPS data that the vehicle transmits real time to the system. So the GPS data will be sent frequently, online from vehicle to the central system. Next in order to check for a route from A to B, the driver will send the query to central system. The system will compute a path using any path-finding algorithms on graph, where the predicted time of the travel will be estimated as described in [17], using his profile.

4 Scenario

In this section, we describe a scenario to explain on-line feedback to drivers; justify heuristic approach to eco-driving; data security issues.

The eco-driving device can be imagined which consists of four components: 1) Personal Eco-Drive Assistance Screen 2) On-board Diagnostics module and reader to read CAN messages every 2 seconds 3) a GPS chip that is programmed to log the position (i.e. latitude and longitude) and speed of the vehicle and finally 4) a GPRS modem that allows data to be transmitted wirelessly to a central server periodically. The data from the CAN bus and the GPS chip are synchronized before forwarding them to server.

The server hosts the modules that improve Eco-driving analysis by using time-series methodologies. These methods are based on the observation that the series of driving related actions/parameters history/records composes a set of time series. For instance, actions/parameters can be: velocity of car at a specific time, acceleration per time, throttle per time or fuel consumption per time. The message exhibits time series continuity, therefore time series methodologies can be employed in order to classify the time series, detect anomalies or regularities in the records. These methodologies are described in section 3.2 of deliverable 2.2.

The learning algorithm at server side application, classifies driver behavior as poor or good w.r.t fuel economy, and the application attaches an advice message (either to regulate [poor driving] or strengthen [good driving] driving behavior). This message is sent back to vehicle via same wireless medium. Once received at vehicle side, Personal Eco-Drive Assistance Screen (See Figure 3) will display this message in a suitable way e.g. the screen can display real-time fuel economy and CO2 emission in a color scheme from red (poor) to green (good), and also driving advice can be displayed.

Figure 3: Personal Eco-Drive Assistance (Courtesy: Co Eco-Driving: Pilot Evaluation of Driving Behavior Changes among U.S. Drivers by Boriboonsomsin et al. University of California).

5 Risk Assessment

The problem which we are facing now is that the CANBus data is very difficult to get. But I have an alternative approach. In case we cannot directly get the measurment data from CANBus data, Fuel consumption model and CO2 emission model which have been investigated in WP4 will be used to estimate fuel consumption and CO2 emission. The methods for estimating fuel consumption and GHG emission have been developed for several years. In the context of road transportation, Fuel consumption and GHG emission models are classified into macroscopic and microscopic models in [18]. The models are a good alternative source to get the CANBus data based on GPS data.

6 Conclusion

In this deliverable, the functionalites from WP2 are summeried based on the novel algorithms for creating predictive analytics models. The main functionalities are categorized as follows: Eco-Routing, Eco-Driving and Distributed Data Mining. Then the requirements of the system have been collected from the REDUCTION partners. Based on the requirements, the fundamental and optional elements of the predictive analytics models are analysed and the system architecture is proposed. A system can be designed and implemented based on the architecture. Next, a scenario is described to explain how the system interative with users, such as on-line feedback to drivers, justify heuristic approach to eco-driving and data security issues. In the end, in the risk assessment section an alternative approach by using estimation models is proposed to estimate fuel consumption and CO2 emission in case of fuel consumption and CO2 emission are no possible to measure directly due to the fact that CANBus data is not possible to get such.

References

[1] Eco-driving facts. http://www.fiat.co.uk/ecodrive/en/#ecodrive_en/thefacts, 2011.

[2] Team Minus 6%. 10 items of eco-driving performance (in japanese). [online]. Available http://www.team-6.jp/.

[3] Kyoungho Ahn and Hesham Rakha. The effects of route choice decisions on vehicle energy consumption and emissions. *Transportation Research Part D: Transport and Environment*, 13(3):151 – 167, 2008.

[4] Hock Ang, Vivekanand Gopalkrishnan, Steven Hoi, and Wee Ng. Cascade rsvm in peer-to-peer networks. In Walter Daelemans, Bart Goethals, and Katharina Morik, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 5211 of *Lecture Notes in Computer Science*, pages 55–70. Springer Berlin / Heidelberg, 2008.

[5] B. Awerbuch, A. Bar-Noy, N. Linial, and D. Peleg. Compact distributed data structures for adaptive routing. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, STOC '89, pages 479–489, New York, NY, USA, 1989. ACM.

[6] Mayank Bawa, Aristides Gionis, Hector Garcia-Molina, and Rajeev Motwani. The price of validity in dynamic networks. *Journal of Computer and System Sciences*, 73(3):245 – 264, 2007. Special Issue: Database Theory 2004.

[7] Kai-Ping Chang, Ling-Yin Wei, Mi-Yeh Yeh, and Wen-Chih Peng. Discovering personalized routes from trajectories. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, LBSN '11, pages 33–40, New York, NY, USA, 2011. ACM.

[8] Nitesh V. Chawla, Lawrence O. Hall, Kevin W. Bowyer, and W. Philip Kegelmeyer. Learning ensembles from bites: A scalable and accurate approach. *J. Mach. Learn. Res.*, 5:421–451, December 2004.

[9] Margarida C. Coelho, H. Christopher Frey, Nagui M. Rouphail, Haibo Zhai, and Luc Pelkmans. Assessing methods for comparing emissions from gasoline and diesel light-duty vehicles based on microscale measurements. *Transportation Research Part D: Transport and Environment*, 14(2):91 – 99, 2009.

[10] Souptik Datta, Kanishka Bhaduri, Chris Giannella, Ran Wolff, and Hillol Kargupta. Distributed Data Mining in Peer-to-Peer Networks. *IEEE Internet Computing*, 10(4):18–26, 2006.

[11] Bolin Ding, Jeffrey Xu Yu, and Lu Qin. Finding time-dependent shortest paths over large graphs. In Alfons Kemper, Patrick Valduriez, Noureddine Mouaddib, Jens Teubner, Mokrane Bouzeghoub, Volker Markl, Laurent Amsaleg, and Ioana Manolescu, editors, *EDBT*, volume 261 of *ACM International Conference Proceeding Series*, pages 205–216. ACM, 2008.

[12] Erin Molly Ferguson. Minimizing vehicle emissions through transportation road network design incorporating demand uncertainty, 2010.

[13] M. Figliozzi. missions minimization vehicle routing problem, 2010.

[14] FORD-WERKE. Schneller schalten, weiter kommen, cologne ford eco-driving, 2003.

[15] Rouphail NM Frey HC, Zhang K. Fuel use and emissions comparisons for alternative routes, time of day, road grade, and vehicles based on in-use measurements. `http://www.ncbi.nlm.nih.gov/pubmed/18504985`, 2008.

[16] Sebastien Glaser, Benoit Vanholme, Sad Mammar, Dominique Gruyer, and Lydie Nouveliere. Maneuver-based trajectory planning for highly autonomous vehicles on real road with traffic and driver interaction. *IEEE Transactions on Intelligent Transportation Systems*, 11(3):589–606, 2010.

[17] Josif Grabocka, Umer Khan, and Lars Schmidt-Thieme. Deliverable d2.2. 2012.

[18] Chenjuan Guo(AU), Bin Yang(AU), Christian S.Jensen(AU), and Kristian Torp (AAU). Deliverable d3.2. 2012.

[19] Erik Hellström, Jan Åslund, and Lars Nielsen. Design of a well-behaved algorithm for on-board look-ahead control. In *IFAC World Congress*, Seoul, Korea, 2008.

[20] Asad Khattak Nagui Rauphail Jorge Bandeira, Dario O. Carvalho and Margariada C. Coelho. A comparative empirical analysis of eco- friendly routes during peak and off-peak hours. `http://transportes-tema.web.ua.pt/publications/83.pdf`, 2011.

[21] Nagui Rauphail Jorge Bandeira, Asad Khattak and Margariada C. Coelho. Generating emission information for route selection, 2011.

[22] M. Sugimoto K. Satou, R. Shitamatsu and E. Kamata. Development of an on-board eco-driving support system (in japanese). *Nissan Tech. Rev., no. 65(2009-9)*, 12:68–71, 2009.

[23] M.A.S. Kamal, M. Mukai, J. Murata, and T. Kawabe. Ecological driver assistance system using model-based anticipation of vehicle–road–traffic information. *IET Intelligent Transport Systems*, 4(4):244–251, 2010.

[24] Murata Kamal, Mukai and Kawabe. Ecological vehicle control on roads with up-down slopes. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, 12:783–794, 2011.

[25] Hillol Kargupta, Kakali Sarkar, and Michael Gilligan. Minefleet: an overview of a widely adopted distributed vehicle performance data mining system. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 37–46, New York, NY, USA, 2010. ACM.

[26] David Kempe, Alin Dobra, and Johannes Gehrke. Gossip-based computation of aggregate information. pages 482–491. IEEE Computer Society, 2003.

[27] F. Kirschbaum, M. Back, and M. Hart. Determination of the fuel-optimal trajectory along a known route. In *Proc. IFAC 15th Triennial World Congress*, Barcelona, Spain, 2002.

[28] Wojtek Kowalczyk, Mark Jelasity, and A. E. Eiben. Towards data mining in large and fully distributed peer-to-peer overlay networks. In *In Proceedings of BNAIC03*, pages 203–210, 2003.

[29] Neal Lathia, Jon Froehlich, and Licia Capra. Mining public transport usage for personalised intelligent transport systems. In *Proceedings of the 2010 IEEE International Conference on Data Mining*, ICDM '10, pages 887–892, Washington, DC, USA, 2010. IEEE Computer Society.

[30] Heewon Lee, Woohun Lee, and Youn-Kyung Lim. The effect of eco-driving system towards sustainable driving behavior. In Elizabeth D. Mynatt, Don Schoner, Geraldine Fitzpatrick, Scott E. Hudson, W. Keith Edwards, and Tom Rodden, editors, *CHI Extended Abstracts*, pages 4255–4260. ACM, 2010.

[31] Wei-Hsun Lee, Shian-Shyong Tseng, and Sheng-Han Tsai. A knowledge based real-time travel time prediction system for urban network. *Expert Syst. Appl.*, 36(3):4239–4247, April 2009.

[32] Jie Li, Henk J. Van Zuylen, Yusen Chen, and Ruihua Lu. Comparison of driver behavior and saturation flow between china and the netherlands. *International Conference on Transportation Engineering 2009, Proceedings of the Second International Conference on Transportation Engineering*, 2009.

[33] Ping Luo, Hui Xiong, Kevin Lü, and Zhongzhi Shi. Distributed classification in peer-to-peer networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '07, pages 968–976, New York, NY, USA, 2007. ACM.

[34] J. Murata M. A. S. Kamal, M. Mukai and T. Kawabe. Ecological vehicle control based on road shapes prediction. *Proceedings of SICE 10th Conference on Control Systems, Paper 164-3-3*, 2010.

[35] Sabine McConnell and David B. Skillicorn. Building predictors from vertically distributed data. In *Proceedings of the 2004 conference of the Centre for Advanced Studies on Collaborative research*, CASCON '04, pages 150–162. IBM Press, 2004.

[36] C.F. Minett. Eco-routing: Comparing the fuel consumption of different routes between an origin and destination using field test speed profiles and synthetic speed profiles, 2011.

[37] Brendan Morris, Anup Doshi, and Mohan M. Trivedi. Lane change intent prediction for driver assistance: On-road design and evaluation. In *Intelligent Vehicles Symposium*, pages 895–901. IEEE, 2011.

[38] Anna Nagurney, Padma Ramanujam, and Kanwalroop Kathy Dhanda. A multimodal traffic network equilibrium model with emission pollution permits: compliance vs noncompliance. *Transportation Research Part D: Transport and Environment*, 3(5):349 – 374, 1998.

[39] Anna Nagurney, Padma Ramanujam, and Kanwalroop Kathy Dhanda. A multimodal traffic network equilibrium model with emission pollution permits: compliance vs noncompliance. *Transportation Research Part D: Transport and Environment*, 3(5):349 – 374, 1998.

[40] Yoshihiro Nishiwaki, Chiyomi Miyajima, Norihide Kitaoka, Ryuta Terashima, Toshihiro Wakita, and Kazuya Takeda. Generating lane-change trajectories of individual drivers. In *IEEE International Conference on Vehicular Electronics and Safety*, pages 271–275. IEEE, 2008.

[41] US Department of Transportation. The congestion mitigation and air quality improvement program: Assessing 10 years of experience, 2002.

[42] Ahn Rakha and Moran. Environmentally-friendly navigation. 2011.

[43] Ahn Rakha and Moran. Integration framework for modeling eco-routing strategies: Logic and preliminary results, 2011.

[44] Y. Saboohi and H. Farzaneh. Model for developing an eco-driving strategy of a passenger vehicle based on the least fuel consumption. *Applied Energy*, 86(10):1925–1932, 2009.

[45] Elvis Rodrigues da Silva, Cláudio de Souza Baptista, Luciana Cavalcante de Menezes, and Anselmo Cardoso de Paiva. Personalized path finding in road networks. In *Proceedings of the 2008 Fourth International Conference on Networked Computing and Advanced Information Management - Volume 02*, NCM '08, pages 586–591, Washington, DC, USA, 2008. IEEE Computer Society.

[46] Elvis Rodrigues da Silva, Cláudio de Souza Baptista, Luciana Cavalcante de Menezes, and Anselmo Cardoso de Paiva. Personalized path finding in road networks. In *Proceedings of the 2008 Fourth International Conference on Networked Computing and Advanced Information Management - Volume 02*, NCM '08, pages 586–591, Washington, DC, USA, 2008. IEEE Computer Society.

[47] M. S. Soares, Jos L. M. Vrancken, and Yubin Wang. Application of a publish-subscribe middleware for road traffic measurements visualization. pages 329–333, March 2009.

[48] Daisuke Yamaguchi Yoichi Sato Takashi Ichihara, Shiro Kumano and Yoshihiro Suda. Driver assistance system for eco-driving. *ITS World Congress 2009*, 2009.

[49] M. Taniguchi. Eco-driving and fuel economy of passenger cars. *In Proceedings of Annual Meeting IEE Japan*, pages p. S21 (5–8), 2008.

[50] Yuan Tian, Ken C. K. Lee, and Wang-Chien Lee. Monitoring minimum cost paths on road networks. In Ouri Wolfson, Divyakant Agrawal, and Chang-Tien Lu, editors, *GIS*, pages 217–226. ACM, 2009.

[51] Theodore Tsekeris and Stefan Vo. Design and evaluation of road pricing: state-of-the-art and methodological advances. *NETNOMICS*, 10:5–52, 2009. 10.1007/s11066-008-9024-z.

[52] Gwo Hshiung Tzeng. Multiobjective decision making for traffic assignment, 1993.

[53] Ran Wolff and Assaf Schuster. Association rule mining in peer-to-peer systems. *Data Mining, IEEE International Conference on*, 0:363, 2003.

[54] Changxu Wu, Guozhen Zhao, and Bo Ou. A fuel economy optimization system with applications in vehicles with human drivers and autonomous vehicles. *Transportation Research Part D: Transport and Environment*, 16(7):515 – 524, 2011.

[55] Wenda Xu, Wen ZhaYao, Huijing Zhao, and Hongbin Zha. A vehicle model for micro-traffic simulation in dynamic urban scenarios. In *ICRA*, pages 2267–2274. IEEE, 2011.

[56] Kai Zhang, Stuart Batterman, and Franois Dion. Vehicle emissions in congestion: Comparison of work zone, rush hour and free-flow conditions. *Atmospheric Environment*, 45(11):1929 – 1939, 2011.