12th ITS European Congress, Strasbourg, France, 19-22 June 2017

Paper ID #

Application of Artificial Intelligence Techniques to Traffic Prediction and Route Planning: the vision of TIMON project

Eneko Osaba[1,2*], Pedro Lopez-Garcia[1,2], Enrique Onieva[1,2], Antonio D. Masegosa [1,2,3], Leire Serrano[1,2] , Hugo Landaluce[1,2]

1. Facultad Ingeniería, Universidad de Deusto, Avda. Universidades, 24, 48007, Bilbao.

2. DeustoTech-Fundación Deusto, Avda. Universidades, 24, 48007, Bilbao.

3. IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain.

* e.osaba@deusto.es

Abstract

TIMON is an European research project under the Horizon 2020 programme. The main objective of this project is to provide real-time services through a web based platform and a mobile APP for drivers, Vulnerable Road Users (VRUs) and businesses. These services will contribute to increasing drivers and VRUs assistance and safety. To provide these services, one of the core technologies developed inside TIMON will be the design and development of Artificial Intelligence (AI) techniques for traffic prediction and route planning. The DeustoTech-Mobility research group is in charge of this part of the project. The objective of this technical paper is to describe the approach followed in TIMON to develop traffic congestion prediction and route planning services based on AI techniques and the progress done so far. Additionally, the deployment and the result obtained in the first test done is also detailed in this study.

Keywords:

Artificial Intelligence, Traffic Congestion, Route Planning.

1. Introduction

TIMON is a research and innovation project supported by the European Union's Horizon 2020 programme[1]. The principal goals of this project are to increase the sustainability, efficiency, flexibility and safety of road transport systems by taking advantage of cooperative communication and by processing open data related to traffic. To this end, a mobile application and a cooperative open web based platform will be developed with the purpose of delivering services and information of drivers, businesses and Vulnerable Road Users (VRU) in real time.

TIMON project counts with a consortium formed by 11 organisations (Universities, RTOs, SME and big industry) from 8 different European countries (Spain, Italy, Belgium, the United Kingdom, Germany, Hungary, the Netherlands, and Slovenia), and an external panel of experts, composed by vehicles suppliers, road transport experts, City Councils etc. In line with this, DeustoTech Mobility is

---

1 http://www.timon-project.eu/

the coordinator of the project and the partner in charge of developing one of the core technologies of the project related to the application of Artificial Intelligence (AI) methods for route planning and traffic forecasting. Briefly explained, DeustoTech-Mobility[2] is a research group inside the Deusto Institute of Technology, placed at the Faculty of Engineering of the University of Deusto. This research group promotes the application of new technologies to address present and future transport and mobility needs of society and industry.

The general scheme of the approach followed in TIMON to develop advanced AI techniques to improve the state-of-the-art on traffic congestion prediction and multimodal route optimization is described below:

- Pre-processing and processing of data gathered from different sources, such as traffic sensors and weather data from Helmond and Ljubljana given by the partners of the project, in order to use it for both prediction and routing systems.
- Identifying behaviour patterns and models in road transport and mobility basing on fuzzy-evolutionary techniques.
- Perform quality predictions of the traffic state, and reliable evaluations of the current traffic state, using fuzzy-evolutionary methods.
- Delivery of optimized multimodal routes using evolutionary techniques.

The objective of this paper is to describe more in detail this approach and to review the progress done so far. In this sense, in the following section (Section 2) the main work regarding data gathering and preprocessing is detailed. In the following section (Section 3) the techniques that will be applied for the traffic prediction and route planning are described. After that, the first testing results are detailed in Section 4. This paper ends with the main conclusions and future work in Section 5.

2. Data gathering and pre-processing

Data used in this project is collected from different sources. Given by the different partners of TIMON project, the creation of the datasets used during the first 18 months has been made. This data contains, on the one hand, the extracted data by JP-LPT[3] and ISKRA[4] from the inductive loop sensors installed in the city of Ljubljana. In overall, the treated data includes 513 files of information collected over three different months (November-2015 to January-2016). On the other hand, a total of 150 files of information collected by traffic cameras from Helmond over five months (from May 27th 2015 to November 18th 2015), forms the rest of the datasets. Besides the traffic data, the weather data was also provided and it contains the corresponding information for Ljubljana and Helmond in the same periods

[2] http://mobility.deustotech.eu/

[3] http://www.lpt.si/

[4] http://www.iskra.si/

described above. These open data is provided by Sensible Code Company[5]. Additionally, a month of data from the I5 highway in California has been used in order to perform additional tests and the good behaviour of the techniques that will be used in future experimentations [7].

Data pre-processing is a very important step for obtaining the final datasets that can be considered correct and useful for further data mining algorithms. In case of TIMON, these datasets will feed the prediction system. The fact of not applying pre-processing methods can produce misleading results. For example, missing values, inconsistent and superfluous data, huge sizes of examples and/or features are problems that can appear in real-world datasets, and could lead to errors during the runtime of the technique, or the bad behaviour of the algorithm. Figure 1 shows a first approach to the division of data pre-processing techniques.

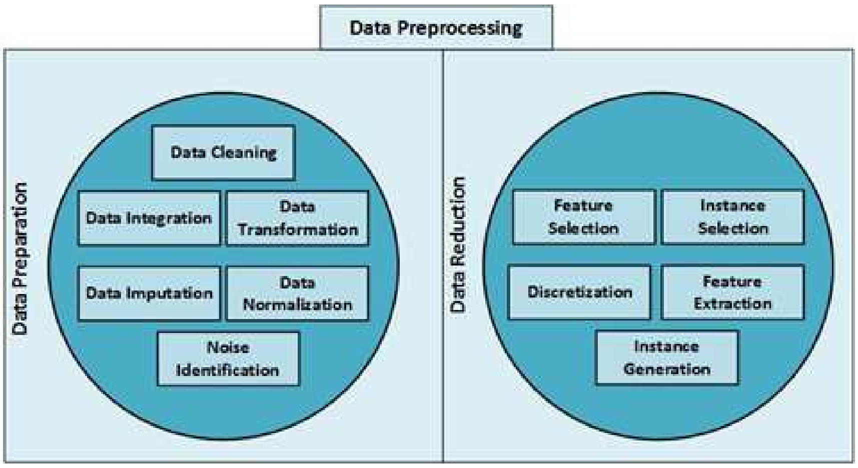

Figure 1 - Data Pre-processing techniques

In the first stage of the project, the creation of the datasets has been made with the data provided by the different partners of TIMON. The RAW data contains, on the one hand, the extracted data from the inductive loop sensors installed in the city of Ljubljana (Slovenia), collected over three different months (November-2015 to January-2016). On the other hand, information collected by traffic cameras from the city of Helmond (The Netherlands) over five months (May 27th 2015 to November 11th 2015), forms the rest of the datasets. Besides the traffic data, the weather data has been also provided, and it contains the corresponding information for Ljubljana and Helmond in the same periods described above. A Data Pre-processing Platform has been developed in JAVA in order to automatize the following pre-processing steps: cleaning, discretization, transformation and normalization. In overall, a total of 92 datasets have been created for Helmond, one per camera, line, and horizon of time. On the other hand, for Ljubljana, 530 datasets have been created, one per sensor and horizon of time.

Besides all the mentioned pre-processing steps, it is important to apply attribute and instance selection techniques to datasets in order to reduce its size without affecting the precision of the models. To

[5] http://sensiblecode.io/

perform this task, the WEKA software[6] has been used. Among the attribute selection techniques, those based on rankings or greedy stepwise have been applied to the datasets. Some examples are the methods OneR [1], Relief [2] or Principal Components [3].

After the applications of those algorithms, some conclusions have been drawn. Firstly, regarding features with contextual information, the hour of the day, the day of the week, the summary of the weather, the precipitation type and the temperature were the ones that provided most valuable information. Secondly, as for features with traffic information, the number of vehicles, the average and the minimum speed in the current and previous time intervals, were the most important. Those features will be kept in the final version of the datasets.

3. Fuzzy Systems and Evolutionary Algorithms for traffic forecasting and route planning

This section includes a brief summary of the work carried out for the routing and prediction systems. First, the application of artificial intelligence techniques, as Evolutionary Fuzzy Rule-Based Systems to a first approach of the problem of forecast traffic congestion is mentioned. After that, the presentation of the approach to the route planning problem based on Evolutionary Algorithms as well as different solutions to solve it is presented.

3.1. Evolutionary Fuzzy Rule-Based Systems for Traffic Congestion

Traffic congestions and travel times can now be more accurately predicted thanks to advanced traffic data collection techniques, combined with modelling and prediction methods. Classical modelling techniques such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees (DT), Regression or Kalman Filters (KF), among others, have been widely used in literature for modelling the traffic flow and, consequently, to predict future states of the road [8]. These techniques are well supported by theoretical results and their effectiveness has been widely tested.

In contrast, they use to lack in the representation of the information and returning models are hardly interpretable by final user, or by traffic managers. In general, any kind of modelling usually comes with two contradictory requirements: interpretability (the capability to express the real system behaviour in a comprehensible way) and accuracy (the capability to faithfully represent the real system).

Due to the multi-objective nature of the problem of constructing both accurate and interpretable systems has not been solved yet. The progress in this line is the use of multi-objective evolutionary algorithms in the optimization of the models. In addition, thanks to the capabilities of fuzzy rule-based systems to deal with uncertainty, and to structure the information in a set of interpretable IF-THEN rules, the predictions systems developed in TIMON will provide to an equilibrium between the complexity of the model and its precision to obtain a higher level of interpretability and accuracy.

The datasets created and pre-processed in the previous stage (described in the above section) are those which will be used to train these forecasting methods in order to determinate the traffic in a point of

[6] http://www.cs.waikato.ac.nz/ml/weka/

interest with different horizons of time. Then, one of the principal activities to carry out is to define a fuzzy based model of classifier/regressor, and its codification as a structure suitable for being optimized by an evolutionary algorithm in order to forecast traffic congestion. The multi-objective implementation of the technique must be oriented to reduce the complexity of the model without affecting its precision. For that reason, measures about the simplicity of fuzzy models have to be considered. To accomplish this task, an ensemble approach has been used due to their good results in many different application fields. Concretely, the focus has been done on boosting methods in general and AdaBoost in particular. AdaBoost is a boosting algorithm, which repeatedly invokes a learning algorithm to successively generate a committee of simple, low-quality classifiers. One of its many variants is the Fuzzy AdaBoost [5]. The objective of Fuzzy AdaBoost is to obtain fuzzy weak classifiers to boost. Then, the overall classifier outperforms any of the fuzzy classifiers. At this stage of the project, the work performed has been focused in the modification of the Fuzzy AdaBoost to adapt it to the main challenges that traffic congestion prediction poses: imbalance datasets and ordinal multiclass classification. The base algorithm code is offered by KEEL software and the modifications are made in JAVA platform. Other methods, as FARCHD [6], are also being studied due to its good performance in the same problems.

Another of the main activities carried out in this task is the implementation of a system that makes use of the models offered by the above describe algorithms to provide traffic congestion prediction over a complete street network. To this end, the OpenTraffic (OT) platform[7] has been used. This framework has also been used to manage the information related with the current state of the traffic.

One of the main advantages of using OT is that it is integrated with Open Street Maps (OSM). In this sense, the platform provides a complete structure to deal with a whole street network and the information related to each segment of this network. This platform has been adapted to use the models provided by the improved Fuzzy AdaBoost algorithm and the information provided by the TIMON Cloud for the segments of a street network. In line with this, it is interesting to mention that the original version of OT does not contemplate the integration of traffic forecasting. For this reason, some modifications have been done to the original version of OT, to add this new functionality.

In the first stage of the process, OT obtains from the TIMON Cloud all the data coming from the different traffic sensors and open data sources. All this information is efficiently pre-processed using the pre-processing engine developed in this work package. Once the traffic and contextual data are obtained and treated, it is used to feed the above described improved Fuzzy AdaBoost algorithm, in order to predict the traffic congestion for four different time horizons: 15, 30, 45 and 60 minutes. Furthermore, and using also the coming traffic data, OT calculates the current traffic state for all the street segments with available information. All this process is executed every 15 minutes, and all the resulting information is assigned to every street segment with traffic data associated. All these predictions, as well as the information related to the current state of the traffic is sent to the TIMON

[7] http://opentraffic.io/

Cloud for its use. The general scheme of this process is represented in Figure 2.

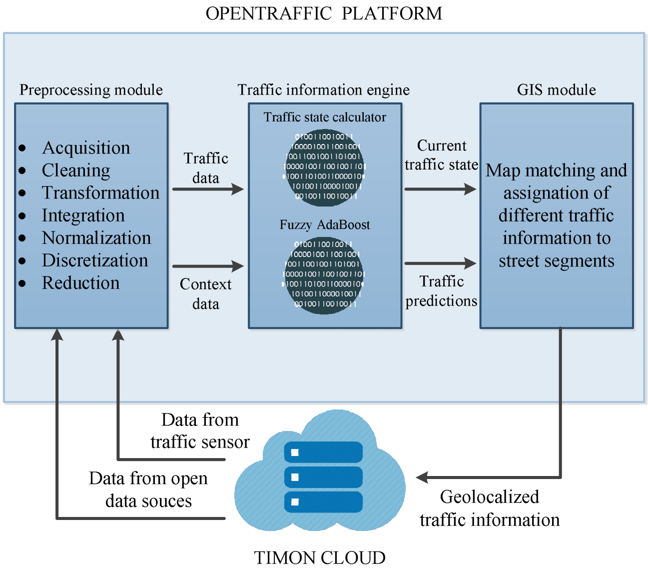

Figure 2 - OT platform process for the traffic predictions and current traffic state data distribution

As pointed out, all the information is updated every 15 minutes, which means that the data must be stored in periods of 15 minutes. In any case, the original version of OT is designed to work with periods of 60 minutes. For this reason, an adaptation process has been made in order to work with these shorter intervals. This adaptation is one additional modification that has been made in order to fill the requirements of TIMON.

Finally, it should be briefly highlighted that every 24 hours OT updates the historical traffic data associated to every segment in which traffic information is available. This information will be used by the routing system, as explained in the next section.

3.2. Evolutionary Fuzzy Rule-Based Systems for Traffic Congestion

The main objective of the routing services of TIMON is to generate optimized routes based on the preferences of the users, and taking into account the traffic predictions obtained so far. Specifically, this task involves mainly the integration of the predictions of the state of the road in the route calculator, and the implementation of the route optimizers, which considers the status of the road at the time the vehicle/user is expected to depart or arrive.

First of all, it is convenient to highlight that the OpenTripPlanner (OTP) platform has been used[8] as a route calculator engine. One of the main activities performed in this task is the efficient communication between OT and OTP in order to incorporate the traffic congestion predictions as well as the current state of the traffic into the route calculation. For this communication, both platforms use the Google's Protocol Buffer[9]. It should be highlighted the basic versions of OTP and OT count with a preliminary

[8] http://www.opentripplanner.org/

[9] https://developers.google.com/protocol-buffers/

structure for working with one, and only one, exchange format. This exchange format takes only into account the historical data. For TIMON, a second exchange format has been generated, in order to incorporate the current traffic state and the congestion prediction data when the routes are calculated.

In this way, each time the routing engine calculates a route, it takes into account the state of the traffic of every street segment. The main objective is to design efficient routes which are dependent on the historical, current and future state of the traffic. In this sense, there are three different considerations, depending on the moment in which the route crosses the different street segments:

- **For street segments crossed in a time horizon between the current moment and the next 15 minutes**: The current state of the traffic will be taken into account.
- **For segments transited in a time horizon between 15 and 60 minutes (from the current moment)**: The proper traffic prediction will be considered, taking into account the four time horizons (15, 30, 45 and 60 minutes).
- **For street segments crossed in a time horizon greater than 60 minutes (from the current moment)**: The historical traffic data will be used.

At this stage of the project, DeustoTech-Mobility has started with the development of the routing algorithm, which has been already designed. This algorithm will consist on an Evolutionary Algorithm to improve the performance of current route calculator of the OTP. In this regard, current route calculators are based on fixed user preferences according to some standard or heuristic values or by the inputs that the user has given. Some common examples of this preferences are the maximum distance the user is willing to walk, the departure or the arrival time, how much a user prefer only one transfer with respect to two, etc. However, many users are flexible with respect to the fixed values provided if a variation leads to a better route. For instance, a user may prefer departing 5 minutes before the time established if in that case he/she can arrive 15 minutes earlier because of the frequency of the buses.

The objective proposed by DeustoTech-Mobility is to design an Evolutionary Algorithm that exploits this tolerance in the users's preferences in order to find small variations (in a range previously established by the users) that lead to better routes in terms of travel time, costs or emissions. In this sense and after having done a depth analysis of the structure of OTP, some of the most convenient user preferences that will be optimized by the Evolutionary Algorithm have been identified. In Table 1 a brief list of the modifiable user's preferences that have been identified so far is listed.

Table 1 - Main classes modified and generated for the efficient communication between OTP and OT

Attribute	**Description**
MaxWalkDistance	The maximum distance (in meters) the user is willing to walk
MaxTransferWalkDistance	The maximum distance (in meters) the user is willing to walk for transfer legs
MaxPreTransitDistance	The maximum time (in seconds) of pre-transit travel
DepartureDateTime	The epoch date/time that the trip should depart

MaxSlope	The maximum slope of streets
MaxTransfers	The maximum number of transport transfers to accept

4. Deployment and results obtained in the first pilot test

The first testing phase of the project was conducted the past September (2016), where some initial tests were carried out which contribute to the future consecution of the objectives planned for the project. In this sense, the system tested had two main functionalities. On one hand, predict future traffic congestions at specified time horizons, and on the other hand, provide traffic density and average speed information in real-time. Both functionalities were tested in the city of Helmond (The Netherlands), at one way of the A270 highway section. At this highway section, some traffic cameras owned by one of the project's partners (TASS International[10]) are installed. The map with the specific location of the road stretch and the cameras is shown in Figure 3.

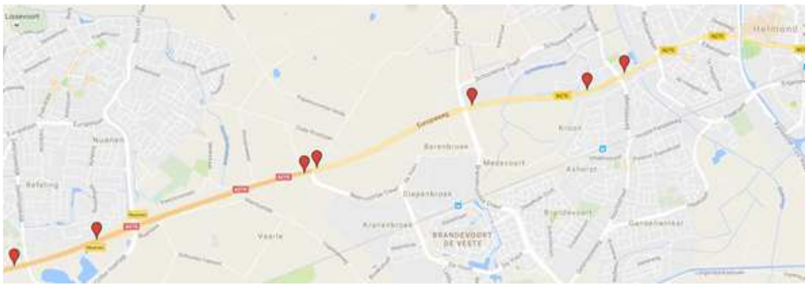

Figure 2 - Position of the used traffic cameras (from left to right: 43, 33, 54, 9, 49, 57, and 48)

The prediction system estimated four levels of congestion (free, slight, moderate and severe) at each line of the seven segments in which the cameras divide the mentioned highway section. To this end, it made use of the real-time information provided by the traffic cameras whose position can be seen in the map above. All these sensors were placed in roads of different nature: four of them were placed in straight road segments, while the remaining ones were located in road bifurcations. In this sense, the level of congestion was calculated using the current average speed shown by the vehicles in the area. Summarizing, the main items tested in this first testing phase can be divided in four different bullets:

- **Traffic congestion prediction in straight road segments**. The traffic congestion prediction system was tested on straight segments where the vehicle's speed is usually high and the probability of congestion low or inexistent.
- **Traffic congestion prediction in road bifurcations**. The prediction system was also tested on road bifurcations or crossings with traffic lights where the vehicle's' speed is usually low and the probability of congestion higher.
- **Real-Time Traffic Information**. The real-time traffic information system is in charge of calculating and offering diverse information about the current state of the traffic, such as the

[10] https://www.tassinternational.com/

density, and the average speed. This information is provided for those road segments for which traffic data are available.

- **Scalability of the system**. In order to test the performance of the system in a more demanding scenario, different experiments replicating the data provided by the existing cameras in their respective lines were executed. In this sense, tests simulating up to 240 traffic sensors were performed.

Furthermore, the main results obtained in this first testing phase can be also divided in four different bullets, each one for each item tested:

- **Results on traffic congestion prediction in straight segments**. Using the sMAPE [4] as a quality measure, the main conclusions obtained for the cameras located in straight road segments are the following. All the results obtained for these tests are below 9%. These results are good enough to conclude that the tested system is reliable. As evidence of this reliability is that the increase of the time horizon does not necessarily imply a lower accuracy of the system. Additionally, there are lots of cases in which the obtained value is lower than 2.5%. In overall, it can be said that the predictions made by the system present a good accuracy.
- **Results on traffic congestion prediction in bifurcations**. Using the same measure, the results obtained were better than the ones presented for the straight roads. This is due to the fact that in bifurcations there is a higher probability of congestions and therefore more examples to "learn" when they occur. In this case, the results obtained are lower than 2% in most of the cases, showing a performance higher that the 5% only in 6 specific cases.
- **Results of the Real Time Traffic System**. The results of this system were also successful since it fulfills the time constraints required for this service, updating the state of the traffic (number of vehicles and average velocity) every 15 minutes.
- **Results of study about the scalability of the system**. The main conclusion obtained is that as the number of traffic sensors grows, the time required for both prediction and a real time traffic system does not grow in a linear way. Typically, this growth tends to resemble an exponential increase, as happens in this case. For these tests, the number of the used sensors was increases up to 240 sensors, showing promising results for both systems.

5. Conclusions and further work

TIMON is a research and innovation project under the Horizon 2020 programme. The main objective of this project is to provide real-time services through a web based platform and a mobile APP for drivers, Vulnerable Road Users and businesses. These services will contribute to increasing drivers and VRUs assistance and safety. To provide these services, one of the core technologies developed inside the project will be the design and implementation of Artificial Intelligence methods for traffic prediction and route planning. The partner in charge of this part of the project is DeustoTech-Mobility research group. In this technical paper, a description of the approach followed in TIMON to develop traffic congestion prediction and route planning services based on AI techniques and the progress done so far is provided. Along with this description, this work also details the deployment and the result obtained in the first test done. The work package detailed in this work ends in the upcoming March 2017, so, the future work is to complete the development of the AI methods and the development of the corresponding deliverables. After the closing of the package, the effort of DeustoTech-Mobility will be focused on the integration of the algorithms in the whole system, and in the writing of some research papers with the obtained results.

Acknowledgements

TIMON project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 636220. Open data shared by SENSIBLE CODE. Close data in Slovenia provided by JP-LPT and ISKRA. Testing scenario deployed by TASS International. The rest of the partners participating in TIMON are: CORTE, FRAUNHOFER ESK, GEOX KFT, CTTC, INTECS SPA, XLAB.

References

1. A. MacKinnon, B. Kramer. (1981). *One-parameter scaling of localization length and conductance in disordered systems*. Physical Review Letters, 47 (21), pp. 1546-1549.

2. K. Kira and L. Rendell, (1992). *A practical approach to feature selection.* In Proceedings of the ninth international workshop on Machine learning, pp. 249-256.

3. L. Eklundh, A. Singh. (1993). *A comparative analysis of standardised and unstandardised principal components analysis in remote sensing*. International Journal of Remote Sensing, 14 (7), pp. 1359-1370

4. D. Sikora, M. Štěpnička, L. Vavříčková. (2011) *Combining techniques in time series forecasting*. Mendel, pp. 419-426.

5. M. J. del Jesús, F. Hoffmann, L. Junco, L. Sánchez. (2004) *Induction of Fuzzy-Rule-Based Classifiers with Evolutionary Boosting Algorithms*. IEEE Trans. on Fuzzy Systems 12, 296-308.

6. J. Alcalá-Fdez, R. Alcalá, F. Herrera. (2011) *A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning*. IEEE Transactions on Fuzzy Systems, 19 (5), art. no. 5756477, pp. 857-872.

7. E. Onieva, P. Lopez-Garcia, A.D. Masegosa, E. Osaba, A. Perallos. (2016) *A Comparative Study on the Performance of Evolutionary Fuzzy and Crisp Rule Based Classification Methods in Congestion Prediction*. Transportation Research Procedia, Volume 14, Pages 4458-4467

8. P. Lopez-Garcia, E. Onieva, E. Osaba, A.D. Masegosa, A. Perallos (2016). *A hybrid method for short-term traffic congestion forecasting using genetic algorithms and cross entropy*. IEEE Transactions on Intelligent Transportation Systems, 17(2), 557-569.