



HAL
open science

A Multi-objective Data Mining Approach for Road Traffic Prediction

Ilias Kalamaras, Anastasios Drosou, Konstantinos Votis, Dionysios Kehagias,
Dimitrios Tzovaras

► **To cite this version:**

Ilias Kalamaras, Anastasios Drosou, Konstantinos Votis, Dionysios Kehagias, Dimitrios Tzovaras. A Multi-objective Data Mining Approach for Road Traffic Prediction. 14th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), May 2018, Rhodes, Greece. pp.425-436, 10.1007/978-3-319-92007-8_36 . hal-01821080

HAL Id: hal-01821080

<https://inria.hal.science/hal-01821080v1>

Submitted on 22 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A multi-objective data mining approach for road traffic prediction

Ilias Kalamaras, Anastasios Drosou, Konstantinos Votis, Dionysios Kehagias and
Dimitrios Tzovaras

Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki,
Greece

{kalamar, drosou, kvotis, diok, tzovaras}@iti.gr

Abstract. Road traffic prediction for the efficient traffic control has lately been in the focus of the research community, as it can solve significant urban issues, such as city evacuation plans, increased concentration of CO₂ emissions and delays caused by extended traffic jams. The current paper proposes a novel approach for multi-variate data mining from past traffic data (i.e. average speed values per road), so as to dynamically detect all significant correlations between the road network components (i.e. the segments of the roads) by mapping the latter onto a low dimensional embedding. Multiple traffic-related features (e.g. speed correlation, spatial proximity, phase difference, etc.) are utilized in a multi-objective optimization framework, producing all Pareto-optimal embeddings, each one corresponding to a different trade-off between the objectives. The operator is provided with the option to interactively select among these Pareto-optimal solutions, so as to explore the most descriptive sets of road influences. The proposed method has been evaluated on real traffic data, while the evaluation of the forecasting performance of the multi-objective approach exhibited accuracy improvement with respect single-objective approaches.

Keywords: Data Mining, Multi-objective Optimization, Pareto-optimal Solutions, Clustering, Graph-based Representation.

1 Introduction

Traffic conditions on road networks have grown to affect several aspects such as the amount of productivity and the lifestyle of the citizens in big urban centers. The optimization of transport, in terms of both individuals and/or fleets, has a serious socio-economic and environmental impact. The study towards the development of intelligent transportation systems (ITS) is constantly gaining attention by the research community. Although common traffic detectors could help alleviate certain traffic problems, the fact that traffic is a dynamic, constantly altering variable, poses new challenges in its confrontation. Consequently, the problem of forecasting traffic within short time intervals ahead of time has arisen as a crucial task, the satisfactory analysis of which is believed to result in more efficient and dynamic routing solutions.

Traffic may be interpreted in various ways through traffic descriptors, such as travel time and instantaneous vehicle speed, observed using GPS sensors. Traffic prediction techniques use this information, taking also into account the fact that the traffic in one region can affect the traffic in another. However, most techniques utilize a single notion of influence between roads, e.g. the correlation between their traffic time series. However, such influences are usually multi-dimensional, depending on more than one parameters, e.g. phase similarity of the corresponding time series, geographical proximity of the roads, etc.

In this paper, an approach for traffic prediction is proposed, which takes into account multiple types of influence among roads. The proposed method considers multiple notions of dissimilarity between roads, based on correlation, phase, geographical distance, etc. The dissimilarity measures are used in a multi-objective Multidimensional Scaling framework, for computing the influence between roads and using it for prediction. The multi-objective nature of the framework allows the selection of trade-offs among the dissimilarity measures, allowing data exploration by operators.

2 Related Work

Existing approaches used for **traffic prediction** are divided into parametric and non-parametric methods. *Parametric methods* are based on specific pre-determined models that are trained in order to deduce their parameters. Common parametric methods include the Auto-Regressive Integrated Moving Average (ARIMA) model [1], and its variations, such as the Auto-Regressive Moving Average (ARMA) model [2]. In contrast with the univariate analysis of the ARIMA model, its multivariate counterpart, the Space-Time ARIMA (STARIMA), first introduced in [3], takes into account several time series that are related to each other, introducing new parameters to account for spatial and temporal lags. A quite different line of parametric methods is the widely used Kalman Filters [4][5], which are based on updating a state variable upon receiving each new measurement.

In *non-parametric methods*, the model is not known a-priori. Non-parametric methods can be categorized to memory-based ones, which retain historical samples in order to perform prediction, and model-based ones, which only need the extracted model, discarding historical data upon the training phase. The most typical example of a memory-based model is the k -Nearest Neighbor (kNN) method. Although simple in nature, kNN seems to produce satisfactory results [6], suggesting that performance lies mainly in the proper representation of the dataset features rather than the blind application of a robust algorithm. Model-based methods construct a model using training data. Typical examples include Random Forests (RFs) [7], Artificial Neural Networks (ANNs), such as Multi-Layered Perceptrons (MLPs) [8][9], and Support Vector Machines (SVMs) [10][11].

Multi-objective optimization deals with problems having many conflicting objectives [12][13]. Such problems arise frequently, especially in engineering and economics, for instance maximizing speed while minimizing fuel consumption or maximizing profit while minimizing cost. A solution that is optimal for one objective is subopti-

mal for the other. One way to deal with this conflict is by scalarizing the objective functions, i.e. combining them in a single objective, which is then minimized with traditional optimization methods. Scalarization methods include using a weighted sum of the objectives [12], the ε -constraint method, where only one of the objectives is minimized, with the others used as constraints [12], or achievement function-based methods, that measure the distance of a solution from a reference one [14]. Scalarizing the objectives involves setting preferences for the multiple objectives, (e.g. the weights of a sum, or the reference point for achievement functions), which is not a trivial task. Another class of multi-objective optimization methods produce a set of solutions [12] instead of a single one, namely the *Pareto front*, containing different trade-offs among the objectives. Such an approach can often discover solutions that scalarization approaches cannot, e.g. solutions lying in concave parts of the Pareto front. The Pareto front is commonly calculated using genetic algorithms, since the fact that they maintain a population of solutions, instead of a single one, suits the goal of calculating multiple solutions [13][15]. For the determination of the fitness function of the genetic process, different approaches have been followed [15], including weighted sums of the objectives with varying weights [16], alternating among the objectives [15] and using dominance relations [17].

Most of the existing methods for traffic prediction usually deal with one traffic-related characteristic (i.e. modality), e.g. speed velocity at a given time instant, travel time required to traverse a road, etc. However, traffic prediction is a multivariate problem and such approaches pose some information loss, since they restrict the cognitive understanding of the traffic to one dimension. In this respect, the current paper proposes a framework that is able to merge multiple variables on a common space via the utilization of Multi-Dimensional Scaling and multi-objective optimization, overcoming thus the so-called “curse of dimensionality”, while a set of Pareto-optimal cases is provided to the traffic operator as possible choices for traffic prediction.

3 Motivation and contribution

Existing research has shown that the choice of the data mining approach to use for traffic prediction can affect the capability traffic modeling. The work of [18] shows that just considering the flow of a road at previous times without considering the flow of neighboring roads discards information that is essential for producing more representative traffic models. Extending this further, the utilization of only a single notion of influence between roads, e.g. graph neighborhoods, may discard other types of influence that may lead to more accurate models. Influences among roads usually depend on more than one parameters, such as correlation between their traffic flow, flow phase similarity, geographical proximity, etc.

When multiple types of information are available, multi-modal processing techniques have proven useful in combining all available information, in order to produce outcomes that are more accurate than using each type of information separately. Combining multiple modalities has been especially useful in the field of multimedia analysis [19]. Multi-objective optimization has shown promising results in this field,

for managing multiple modalities [20], as a generalization of other multimodal techniques, able to discover more solutions and to present the operator with a set of limited optimal trade-offs. Adapting this approach for combining multiple traffic-related characteristics seems promising for improving traffic prediction and allowing parameter exploration by the operator.

The contribution of this paper is the combination of multiple traffic-related features, using multi-objective optimization techniques, in order to measure the influence of one road to other roads of the network, and thus define different notions of “neighborhood”, based on Multi-Dimensional Scaling, to be used in STARIMA traffic models. The proposed framework seems to produce promising results, and is adaptive to more traffic features and further notions of road similarity, providing space for further experimentation and research. From an application point of view, the operators are presented with a set of trade-offs among the multiple traffic features, allowing them to select the most important ones for prediction. Such interactivity facilitates the operator of an ITS in exploring the available data and making decisions.

4 Proposed approach

The proposed approach proceeds as follows. Initially, traffic-related features (vehicle speed, etc.) are extracted for each road in order to compute pairwise distance matrices, using multiple notions of road distance. The distance matrices are used to construct objective functions, whose minimization leads to an embedding of the roads as points on the 2D plane. Multidimensional Scaling (MDS) is used to formulate these objectives, while multi-objective optimization techniques are used to compute a set of Pareto-optimal placement solutions. This way, a set of alternative placements of the roads are provided, that can be used to define neighborhoods of influence, for use in STARIMA [3] traffic prediction. The placement solutions can be interactively selected by the operators, with reference to the indicated current state of the selected road segment, while traffic prediction is visually annotated in real time on the map.

4.1 Problem formulation

A set $O = \{R_1, R_2, \dots, R_N\}$ of N road segments is considered. A road segment is the part of a road between two subsequent intersections. If there are two opposite lanes in the same road between two intersections, two separate road segments are considered, one for each lane. Without loss of generality, the road segments are hereby considered to be straight lines.

For a particular day, each road segment R_i , $i = 1 \dots N$, is modeled as a set of certain attributes:

$$R_i = \{\mathbf{s}_i, \mathbf{e}_i, \mathbf{v}_i\} \in P, \quad (1)$$

where $\mathbf{s}_i \in \mathbf{R}^2$ is the starting point of the segment, in map coordinates (latitude and longitude), $\mathbf{e}_i \in \mathbf{R}^2$ is similarly the ending point of the segment and $\mathbf{v}_i \in \mathbf{R}^M$ is a vector whose j -th element, $j = 1 \dots M$, is the average speed of the vehicles traveling

on the road segment at the j -th time interval, where the day has been split into M time intervals. The road segments R_i are considered to belong in the space P of all sets of these attributes.

The problem addressed hereby is, given the road segments and the vehicle speeds for a specific day, to discover which road segments are related to a selected road segment, in order to predict which segments will be influenced by a change in the vehicles' speed, e.g. denoting congestion, at the selected segment, on the same day another week.

4.2 Road segment distance measures

The proposed approach is based on defining notions of distance among the road segments, which encode different types of influence among them. The following distance measures are used, covering various spatiotemporal characteristics, although others can be incorporated as needed. The distance measures are used to construct distance matrices among the data, containing the distances among each pair of road segments.

Cross Correlation similarities. This metric stands for the cross correlation value between the sequences of speeds of two road segments. Given two road segments R_i and R_j , $i, j \in 1 \dots N$, with their corresponding speed vectors \mathbf{v}_i and \mathbf{v}_j , the *correlation distance* d_{cor} is defined as follows:

$$d_{\text{cor}}(R_i, R_j) = \max_{k=-M \dots M} \left| \frac{E[(v_{i,t} - \mu_i)(v_{j,t+k} - \mu_j)]}{\sigma_i \sigma_j} \right|, \quad (2)$$

whereby $v_{i,t}$ is the t -th element of vector \mathbf{v}_i , μ_i is the mean value of vector \mathbf{v}_i , σ_i is the standard deviation of \mathbf{v}_i and $E[\cdot]$ stands for the mean value of the enclosed values, for all values of t . The correlation value derives from the absolute value of the Pearson Product-Moment Correlation Coefficient (PPMCC), with the modification of considering a phase parameter k , in order to take all the possible alignments of the two time series, as in [21], since one of them may be delayed with respect to another due to time needed for traffic to pass from one segment to another.

Phase similarities. The correlation value above considers the maximum value of the correlation coefficient for all possible alignments of the two time series. The *phase similarity* d_{phase} between two time series is hereby defined as the amount of sliding needed in order to achieve this maximum coefficient value. It is therefore defined as:

$$d_{\text{phase}}(R_i, R_j) = \left| \arg \max_{k=-M \dots M} \left| \frac{E[(v_{i,t} - \mu_i)(v_{j,t+k} - \mu_j)]}{\sigma_i \sigma_j} \right| \right| \quad (3)$$

Only the absolute value of the delay k is used, in order for the distance measure to be symmetric. However, the sign of the delay, i.e. whether the second segment precedes or follows the first, can be used in applications, in order to demonstrate which road segments will be influenced in the future.

Geographical proximity. The *geographical distance* d_{geo} between two road segments is defined as the Euclidean distance between their midpoints:

$$d_{\text{geo}}(R_i, R_j) = \|\mathbf{m}_i - \mathbf{m}_j\|, \quad (4)$$

where $\mathbf{m}_i = \frac{1}{2}(\mathbf{s}_i + \mathbf{e}_i)$ and $\|\cdot\|$ denotes the Euclidean norm.

Dynamic Time Warping difference. Provided the strong dependence on temporal relations of the fluctuations in the recorded velocities during the day, another metric that has been utilized for the estimation of the intra-distances between the roads is the Dynamic Time Warping (DTW) algorithm [22] that sufficiently manages to capture the spatiotemporal characteristics of these signals. The DTW algorithm has been widely used in a series of matching problems, varying from speech processing [22] to biometric recognition applications. Its main advantages are its simple implementation and its satisfactory performance given the required processing time.

4.3 Multi-objective multidimensional scaling

The hereby proposed approach is, using the distance measures described in Section 4.2, to embed the road segments into a low dimensional space, where nearest neighbors can be calculated, given a selected segment. Formally, the goal is to find an embedding $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N$, $\mathbf{p}_i \in \mathbb{R}^L$, for the N road segments R_1, R_2, \dots, R_N , so that the distances among the points \mathbf{p}_i in the low dimensional space \mathbb{R}^L correspond to all the notions of distance between road segments which are defined in Section 4.2.

Multidimensional Scaling (MDS) [23] is a common technique used to find an embedding of a set of points, when only the distances among them are known. Let the known target distances be δ_{ij} , $i, j \in 1 \dots N$. The desired embedding $\mathcal{P} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N) \in \mathcal{P}$ of the objects is one in which the target distances among points are best preserved in the embedding. Let $J: \mathcal{P} \rightarrow \mathbb{R}_{\geq 0}$ be a cost function (objective function), evaluating the capability of an embedding $p \in \mathcal{P}$ for preserving the target distances among the data. $\mathbb{R}_{\geq 0}$ is the set of non-negative real numbers. A commonly used objective function for MDS is the following:

$$J(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N) = \sum_{i,j}^N (\|\mathbf{p}_i - \mathbf{p}_j\| - \delta_{ij})^2 \quad (5)$$

Since there are multiple notions of distance between two data points (Section 4.2), multiple objective functions for MDS can be defined, one for each distance measure used. This is handled as a *multi-objective optimization* problem [12], using genetic algorithms [17], resulting in a Pareto front of optimal trade-offs among the objectives. The genetic algorithms proceed by examining different placements of the N points on the 2D plane, so that there are $2N$ variables. Crossover strategies can combine two placements by randomly keeping, for each object, the corresponding point from either parent as the child's point. Mutation operates by adding random noise on the child's position. The initial population size can be chosen as multiple of the solutions that are kept to represent the final Pareto front, which can be in the order of 10 - 20, so that

they are not overwhelming to the operator. The focus of this paper is more on the assessment of how the combination of multiple traffic characteristics can benefit prediction than on finding the optimal crossover/mutation functions and population size. Further work can use different optimization techniques, in order to achieve more representative Pareto fronts faster.

An example Pareto front for a problem of two objectives is illustrated in Fig. 1(a). The Pareto diagram depicts the Pareto optimal solutions as points, having the values of the objective functions as coordinates. The gray-shaded area represents the set of all feasible solutions, while the bold border in the lower left of the feasible area is the Pareto front. Point P_2 dominates P_1 , as well as the whole hatched area, since both objectives have smaller values at P_2 . On the other hand, points P_2 and P_3 are incomparable, since decreasing one objective leads to increasing the other one.

The selection of a Pareto-optimal embedding to use for detecting roads with similar characteristics and making predictions is performed by a human user. The solutions of the Pareto set are presented to the user in the form of a Pareto diagram, such as the one in Fig. 1(a). By selecting among different trade-offs, the operator can focus on different aspects of traffic flow and be assisted in predicting future states. After the selection of a Pareto-optimal embedding, the road segments are represented by points in the 2D space (see Fig. 1(b)). Nearby points represent roads that are similar with respect to the distance measures used and the Pareto trade-off that has been selected. When an operator selects a road segment on the map, wishing to view the segments which are influenced by it, the nearest roads segments to the selected one in the embedding are found and presented on the map. The amount of closeness of a road segment can be depicted graphically on a map by e.g. controlling its color or opacity.

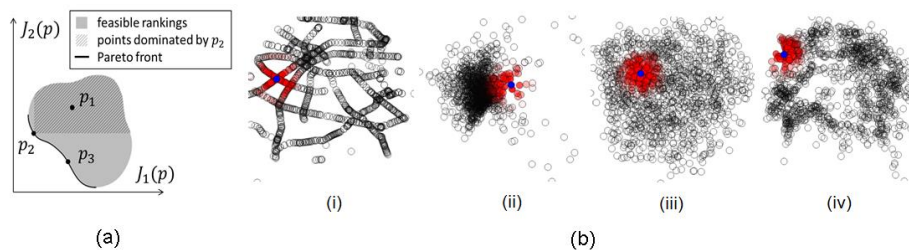


Fig. 1. (a) Example Pareto diagram. (b) Low-dimensional embeddings solely based on the (i) geographical proximity (ii) temporal proximity (iii) cross-correlation (iv) DTW-based modality.

5 Evaluation of the proposed approach

5.1 Dataset Description

The proposed method has been tested on the so-called *Berlin dataset*, which was recorded from 18/03/2012 to 31/03/2012, using the open TomTom API [24]. The data contain real vehicle speed measurements from Berlin, collected from several road

points using GPS locators. Each measurement contains the instantaneous speed at a specific road point, measured in millisecond intervals, and is accompanied by the map coordinates of the road location. However, due to the absence of speed values in many time moments, a preprocessing of the original data was performed, as described in [21]. The preprocessing procedure, for a day of measurements is as follows.

First, the measurement points were grouped in road segments, i.e. parts of a road between two subsequent intersections, resulting in about 7000 road segments, each associated with the speeds obtained for the road points inside it. Since each road segment may contain a different number of speed measurements, the raw measurements were grouped in five-minute intervals, considering the harmonic mean of the speed values at each interval. The size of the resulting speed vector assigned to a road segment was at most 288, the number of five-minute intervals in a day. This significantly reduced the size of the dataset and compensated for many missing values. However, there were still many intervals within a day with no measurements, so we kept a time period of 15 hours that had a small overall number of missing values throughout the whole dataset. This truncated the speed vectors to 180 values. The remaining missing values were filled using cubic spline interpolation. Finally, in order to limit the experiments to busy, and thus interesting, roads, the road segments were further filtered by keeping only those within a radius of approximately 5km from the city center. The final dataset consists of about 1300 road segments for each day of the week, each associated with a 180-dimensional speed vector (of average speed value 33.5 km/hour).

5.2 Experimental Results

The proposed method has been employed for the task of predicting the speed of a road segment at a future time interval. For evaluation purposes, it has been compared to the performance of methods which use a single objective to determine which road segments are influenced by the traffic of a selected segment. The model used for prediction is a simplified form of the STARIMA model, such as the ones used in [21]. For a selected road segment, the model is described by the following equation:

$$z_{t+1} = \phi_0 z_t + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \psi_0 u_t + \psi_1 u_{t-1} \quad (6)$$

Hereby, z_{t+1} is the predicted value of the speed of the road segment at time interval $t + 1$. This value is calculated as a weighted combination of the speed values of the same road at previous time intervals, as well as of roads that are influenced by the selected road segment. The values z_t , z_{t-1} and z_{t-2} in Eq. (6) are the speeds of the road segments at the current time interval, t , and at the two previous time intervals, $t - 1$ and $t - 2$, respectively. The parameters ϕ_0 , ϕ_1 and ϕ_2 are used as weights to denote the importance of each previous speed value for the determination of the predicted one. The last two factors of Eq. (6) are related to the speeds of the road segments which are most influenced by the selected road segment. The parameter u_t is defined as:

$$u_t = \frac{1}{k} \sum_{R=\{s,e,v\} \in \mathcal{N}_k} v_t \quad (7)$$

The set \mathcal{N}_k is the set of the k segments which are most influenced by the selected segment. The value of u_t is thus the mean value of the speeds at time interval t of the road segments which belong to the set \mathcal{N}_k . The determination of the most influenced road segments is hereby performed by selecting, in the 2-dimensional embedding, the k closest points to the point corresponding to the selected road segment. The most influenced road segments are thus determined by considering all notions of distance (Section 4.2) in a multi-objective manner, instead of using a single influence measure, as e.g. in the lag-based STARIMA model of [21].

In order to demonstrate the effectiveness of using multiple objectives, a comparison has been performed between two Pareto-optimal solutions. The first corresponds to a combination of the d_{cor} and d_{tim} distances measures, i.e. it is similar to the lag-based STARIMA method of [21], which uses correlation to find the roads that are most influenced, but also exploits time lags. The second Pareto point corresponds to a combination of the d_{cor} , the d_{tim} and the d_{geo} measures, i.e. it also includes geographical information. For each solution, the model of Eq. (6) is trained with data from a specific day of the week. The unknown model parameters ϕ_0 , ϕ_1 , ϕ_2 , ψ_0 and ψ_1 are learned using a least-square error estimate. In order to test the model, predictions are made for the same day at another week. Given the current speed for a specific road segment, along with the previous speeds of the same segment and of the most influenced ones, the speed at the next time interval is calculated using Eq. (6).

Fig. 2(a) illustrates the Root Mean Square Error (RMSE) between the predicted and the real values, for different time intervals. For each interval, the RMSE for the prediction of the next interval is depicted. The RMSE is defined as the square root of the mean of the squared error between the predicted and the real values, where the mean is taken over all road segments. The combination of the three measures results in generally smaller prediction errors than the combination of just two of them, especially at the beginning of the day by an improvement of 0.0301 in the RMSE. Fig. 2(b) depicts similar results for another pair of sub-cases. Hereby, the first Pareto solution considers only the geographical distances among the roads, while the second solution considers a combination of the d_{DTW} , the d_{tim} and the d_{geo} measures of Section 4.2. Again the combination of three objectives outperforms the use of only one. The average improvements in the RMSE between the two curves is 0.0717.

The capabilities of the proposed approach have also been exhibited via a visualization application. Let us consider that, after the low-dimensional embedding has been performed and one of the Pareto-optimal solutions has been selected, an operator selects a specific road segment (blue line in Fig. 3(a) - Fig. 3(d)). The point in the 2D embedding that corresponds to the selected road segment is found and the road segments corresponding to nearby points in the embedding are colored on the map. The operator can thus have an overview of which roads are influenced by the traffic conditions of the selected road segment.

In Fig. 3(a) - Fig. 3(d), examples of different segment selections are depicted. When the operator selects a different segment, different roads are colored, denoting those which are mostly influenced by the selected segment. The color of the road segments is an indication of the direction of traffic. The faint red color corresponds to segments whose speed series precede the speed series of the selected segment, in

terms of correlation, while segments with bright red color have speed series which follow the ones of the selected segment. As an example, in Fig. 3(d), the colors of the roads indicate that the general direction of traffic is towards the center of the city.

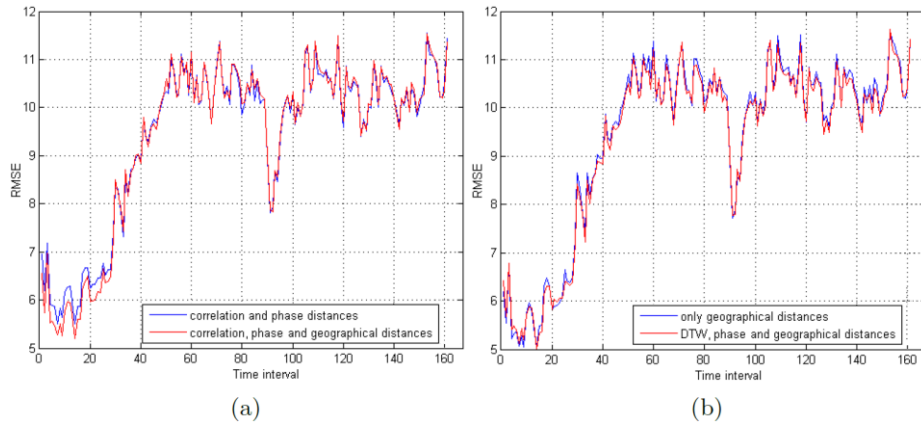


Fig. 2. Comparison of RMSE for predictions performed using: (a) a combination of d_{cor} and d_{tim} (blue curve) and a combination of d_{cor} , d_{tim} and d_{geo} (red curve), (b) only d_{geo} (blue curve) and a combination of d_{DTW} , d_{tim} and d_{geo} (red curve).

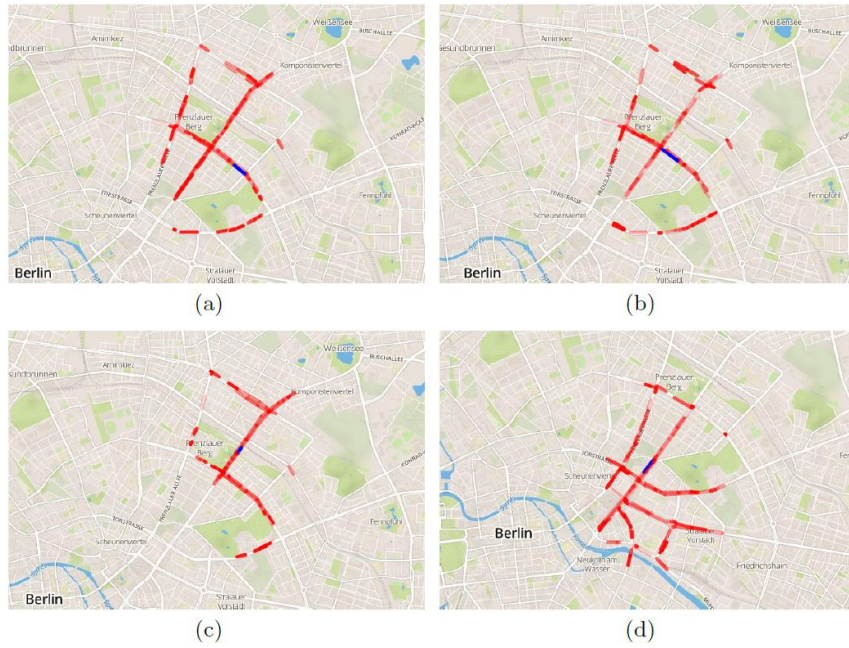


Fig. 3. Examples of the neighboring roads, in terms of “merged”, low dimensionality distance. The brightest the red colored lines, the biggest is their phase difference in the future. (d) It can be noted that the traffic congestion in the next minutes will be moving towards the center of the city.

6 Conclusions and next steps

In this study, the problem of forecasting traffic speeds was addressed, using a multi-objective framework that supports the combination of several traffic-related modalities via Multidimensional Scaling. The proposed approach facilitates increased flexibility and efficient human interaction, while exhibiting improved traffic prediction results when the broadly utilized STARIMA algorithm is applied. Furthermore, the proposed framework can form the basis for useful interactive applications, via suggesting only the (Pareto-)optimal solutions to the operator.

The potential for further improvement in the prediction accuracy of the proposed approach should be highlighted, since it supports the integration of an unlimited amount of additional traffic-related modalities. Apart from speed measurements, traffic volume measurements can also be exploited as additional modalities in the future. Traffic volume adds additional information, which, combined with traffic speed can lead to more accurate traffic prediction. The utilized distance measures can be extended by using geodesic distance measures, which exploit the graph-like structure of the road network, and may be more informative than actual geographic proximity. The graph-based structure can also be used to improve the distance measures already used, by adjusting the distance matrices, based on their geodesic proximity of the graph. Such improvements are the objectives of future extensions of the current work, with the goal of achieving more accurate traffic prediction.

7 Acknowledgments

This work was supported by the EU-funded projects RESOLUTE (H2020-DRS-2014, grant agreement no. 653460) and My-TRAC (H2020-S2RJU-OC-2017, grant agreement no. 777640).

References

1. G. E. P. Box and G. Jenkins: Time Series Analysis, Forecasting and Control. Holden-Day, Incorporated, 1990.
2. P. J. Brockwell and R. A. Davis: Introduction to Time Series and Forecasting. Springer, 2nd ed., Mar. 2002.
3. P. E. Pfeifer and S. J. Deutsch: A three-stage iterative procedure for space-time modeling, *Technometrics*, vol. 22, no. 1, pp. 35-47, 1980.
4. R. E. Kalman: A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35-45, 1960.

5. I. Okutani and Y. J. Stephanedes: Dynamic prediction of traffic volume through Kalman filtering theory, *Transportation Research Part B: Methodological*, vol. 18, no. 1, pp. 1 { 11, 1984.
6. W. Shen, Y. Kamarianakis, L. Wynter, J. He, Q. He, R. Lawrence, and G. Swirszcz: Traffic velocity prediction using GPS data: IEEE ICDM contest task 3 report, in *Data Mining Workshops (ICDMW)*, 2010 IEEE International Conference on, pp. 1369-1371, 2010.
7. L. Breiman: Random forests, *Mach. Learn.*, vol. 45, pp. 5-32, Oct. 2001.
8. E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias: Optimized and metaoptimized neural networks for short-term traffic flow prediction: A genetic approach, *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 211-234, 2005.
9. S. Innamaa: Short-term prediction of travel time using neural networks on an interurban highway, *Transportation*, vol. 32, no. 6, pp. 649-669, 2005.
10. V. N. Vapnik: *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.
11. C.-H. Wu, J.-M. Ho, and D. T. Lee: Travel-time prediction with support vector regression, *Trans. Intell. Transport. Sys.*, vol. 5, pp. 276-281, Dec. 2004.
12. M. Ehrgott: *Multicriteria optimization*, vol. 491. Springer Science & Business Media, 2005.
13. C. A. C. Coello, G. B. Lamont, D. A. Van Veldhuizen, et al: *Evolutionary algorithms for solving multi-objective problems*, vol. 5. Springer, 2007.
14. A. P. Wierzbicki: On the completeness and constructiveness of parametric characterizations to vector optimization problems, *Operations-Research-Spektrum*, vol. 8, no. 2, pp. 73-87, 1986.
15. E. Zitzler, M. Laumanns, and S. Bleuler: A tutorial on evolutionary multiobjective optimization, in *Metaheuristics for multiobjective optimisation*, pp. 3-37, Springer, 2004.
16. P. Hajela and C.-Y. Lin: Genetic search strategies in multicriterion optimal design, *Structural optimization*, vol. 4, no. 2, pp. 99-107, 1992.
17. E. Zitzler, M. Laumanns, and L. Thiele: SPEA2: Improving the strength Pareto evolutionary algorithm, *TIK-report*, vol. 103, 2001.
18. Y. Kamarianakis and P. Prastacos: Space-time modeling of traffic flow, *Comput. Geosci.*, vol. 31, pp. 119-133, Mar. 2005.
19. P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli: Multimodal fusion for multimedia analysis: a survey, *Multimedia systems*, vol. 16, no. 6, pp. 345-379, 2010.
20. I. Kalamaras, A. Drosou, and D. Tzovaras: Multi-objective optimization for multimodal visualization, *IEEE Transactions on Multimedia*, vol. 16, no. 5, pp. 1460-1472, 2014.
21. T. Diamantopoulos, D. Kehagias, F. G. König, and D. Tzovaras: Investigating the effect of global metrics in travel time forecasting, in *Proceedings of 16th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2013.
22. H. Sakoe and S. Chiba: Dynamic programming algorithm optimization for spoken word recognition, in *Readings in speech recognition*, pp. 159-165, Elsevier, 1990.
23. M. A. Cox and T. F. Cox: Multidimensional scaling, in *Handbook of data visualization*, pp. 315-347, Springer, 2008.
24. TomTom open maps API, <https://developer.tomtom.com/online-traffic>, last accessed 2018/03/20