

Available online at www.sciencedirect.com**ScienceDirect**

Transportation Research Procedia 27 (2017) 11–18

**Transportation
Research
Procedia**

www.elsevier.com/locate/procedia

20th EURO Working Group on Transportation Meeting, EWGT 2017, 4-6 September 2017,
Budapest, Hungary

Adarules: Learning rules for real-time road-traffic prediction

Rafael Mena-Yedra ^{a,b}*, Ricard Gavaldà ^a, Jordi Casas ^b,

^aUniversitat Politècnica de Catalunya, Campus Nord, C/ Jordi Girona, 1-3, 08034 Barcelona, Spain

^bTSS-Transport Simulation Systems, Ronda de la Universitat, 22b, 08007 Barcelona, Spain

Abstract

Traffic management is being more important than ever, especially in overcrowded big cities with over-pollution problems and with new unprecedented mobility changes. In this scenario, road-traffic prediction plays a key role within Intelligent Transportation Systems, allowing traffic managers to be able to anticipate and take the proper decisions. This paper aims to analyze the situation in a commercial real-time prediction system with its current problems and limitations. We analyze issues related to the use of spatiotemporal information to reconstruct the traffic state. The analysis unveils the trade-off between simple parsimonious models and more complex models. Finally, we propose an enriched machine learning framework, *Adarules*, for the traffic state prediction in real-time facing the problem as continuously incoming data streams with all the commonly occurring problems in such volatile scenario, namely changes in the network infrastructure and demand, new detection stations or failure ones, among others. The framework is also able to infer automatically the most relevant features to our end-task, including the relationships within the road network, which we call as “structure learning”. Although the intention with the proposed framework is to evolve and grow with new incoming big data, however there is no limitation in starting to use it without any prior knowledge as it can start learning the structure and parameters automatically from data.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 20th EURO Working Group on Transportation Meeting.

Keywords: real-time traffic prediction; autolearning; adaptation to change

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .

E-mail address: rafael.mena@cs.upc.edu ; rafael.mena@aimsun.com

1. Introduction

1.1. Motivation

Currently, the growth of global population, urbanization and changes in people density and distribution along with their mobility needs have caused traffic congestion to become a major problem globally. Concerning the strategies adopted in traffic management, the tendency has been moving from focusing uniquely on classical measures, such as increasing infrastructure or restricting the use of private car, to more intelligent solutions. These solutions include the provision of information (Traveler Information Services) to both agents, the driver and the traffic manager. The former can make use of this predictive information to perform suitable travel decisions before the departure time (pre-trip information) and during the journey (on-trip information), while the traffic control agent in a Traffic Management system can take the appropriate decisions for accommodating traffic flow in an efficient, effective and safe manner whose assortment of techniques and strategies is known as Active Traffic Management (ATM). In the end, the use of this kind of information has a beneficial impact on the network performance in terms of throughput, congestion length and average network speeds.

1.2. Context

This research work has been developed in the context of Aimsun. The Aimsun transport modelling software (Casas et al., 2010) was originally the focus of a multi-year research programme (Ferrer and Barceló, 1993) at the Universitat Politècnica de Catalunya (UPC), and currently it is in its 8th commercial major version with ongoing development at TSS-Transport Simulation Systems. Aimsun is one of the most advanced and complete transport modelling software used by consultants, universities and city halls around the world. More recently, it includes a decision support system for real-time traffic management known as Aimsun Online, which is used by traffic control centers to make real-time decisions about the management of a road network, since it can be used to dynamically forecast future traffic conditions based on the current state of the network and to evaluate incident response or traffic management strategies. However, there has been some shortcomings concerning the data-driven analytical subsystem for the short-term traffic prediction. Because it has relied to be built in an offline batch setting with just historical data for the learning procedure, it has presented problems related to a lack of adaptation to change, and a heavy reliance on the data analyst or traffic engineer for deciding when to renew the models set with new gathered data for maintenance purposes. In addition, it is especially important that an enough representative historical data set which covers different scenarios, namely seasonality, incidents, among others, must be available for the models learning, otherwise they will not respond accurately to such situations during real-time operation.

2. Literature survey: Short term traffic forecasting

Because road traffic is the visible result of the complex interplay between traffic demand (the amount of travelers making a trip at a particular place and time) and traffic supply (network infrastructure), when modeling it is usual to find that the input–output data relationship is noisy and that the relationships between these variables are multivariate and (highly) nonlinear (van Lint and van Hinsbergen, 2012), additionally the process is usually high-dimensional, non-stationary and tackled in real-time.

In the literature, there are two main approaches adopted for road traffic prediction: model-driven and data-driven. Model-driven approaches try to reproduce the road network behavior through simulation, and depending on the level of detail and the underlying traffic flow theory in which are based they could be distinguished among microscopic, mesoscopic, macroscopic and hybrid variants (Barceló, 2010; Treiber and Kesting, 2013). One requirement of such model-driven approaches to obtain accurate predictions is to have a detailed knowledge about the network topology. Model-driven approaches are within the parametric category as their number of parameters are fixed in advance as well as the model structure, however their main drawback is that a detailed knowledge about each road network being modeled is required to obtain accurate predictions, but also it is especially important the inherent fact that road network infrastructures are continuously changing and such changes must be reflected into the model to maintain the accuracy.

On the other hand, the data-driven approach aims to reproduce the input–output mapping but usually neglecting the underlying data generation process (i.e. well-founded mathematical models that are based on macroscopic and microscopic theories of traffic flow) and disregarding, in general, the network topology. Despite this, integrating the network spatio-temporal information within the short-term traffic prediction task is of ultimate importance (Ermagun and Levinson, 2016). This branch has taken advantage from the fact that over time different measuring devices have been deployed within road networks to measure and verify road traffic conditions. For additional review references, (Ermagun and Levinson, 2016; van Lint and van Hinsbergen, 2012; Vlahogianni et al., 2014),

3. Methodology

As stated in recent reviewed literature, nowadays the trend for short-term traffic forecasting is data-driven empirical approaches, given the growing data availability. This creates the necessity to handle both structured and non-structured data as well as to take advantage from contextual information and data coming from multiple sources and technologies. Additionally, the short-term traffic forecasting task is inherently a real-time task that must deal itself with the common challenges found in this field, namely high-dimensionality and non-linearity, noisy data from the measurement devices, missing data from faulty or disabled ones, volatility, and adaptation to change in the traffic demand and the traffic supply characteristics. For these reasons, it is widely accepted that a non-parametric approach is usually required to manage the growing complexities as new data is collected.

To this end, we propose a framework named Adarules, which is built from different machine learning and data analysis components that ultimately make up a predictive system robust to outliers, irrelevant features and missing data, able to grow its complexity with new data and adaptive to changes (concept drift), it is also able to scale along with the network size. The framework is inspired by the works of (Gama, 2010) applied to data streaming scenarios (Almeida et al., 2013), but tailored to the requirements for this application.

3.1. Automatic knowledge discovery through rule identification

Adarules works in a supervised manner, meaning that for each desired prediction target, it is going to discover or unveil a set of rules to gain knowledge about the supervised task, having past observations with their correct prediction. Then, each rule R contained within each ruleset \mathcal{R} is composed of an antecedent A and a consequent C with the logical form: $A \Rightarrow C$. The rule antecedent can be composed of several literals L , where a literal L is a single condition over a specific attribute with a specific split-point v ; with the form $(x_j > v)$, $(x_j \leq v)$ if it is numerical, or $(x_j = v)$ if it is categorical. The antecedent is interpreted as a conjunction. In this way, a rule R is said to *trigger*, or to cover, an example x_i if all its literals (the antecedent) are evaluated to *True* on the example.

The consequent (of a rule) is composed of an adaptive output using the multiple rule predictors that the rule may hold. The individual outputs are built at prediction time from the examples gathered in the scope of that rule, then the adaptive output is generated from that population of individual outputs (also could be called experts, following an expert advice schema) weighted by their respective online errors. In addition to the prediction point estimate, an uncertainty interval is given based on the error seen which approximates the real one as the uncertainty associated with covariates is neglected. Finally, each rule R has an associated data structure \mathcal{L} which contains updated statistics from the observed streams (attributes, targets and errors) for those observations gathered by the rule.

3.2. Framework modules

3.2.1. Variable selector

The variable selector is the unit aimed to handle the prior information usually put as expert knowledge. In our case, it just set a normalized attractiveness value to each feature separated in different categories (e.g. count, occupancy, speed, time, weather...). Thus, they have an associated probability (that could be updated with new gathered evidence) that is used to select features stochastically as will be explained later in the rule expansion process and online learning. Therefore, features associated with detectors in the road network have a normalized attractiveness ($1/d$, where d is the orthodromic distance which could be easily replaced by using travel times coming from a transport network model) based on their distance to the point to be predicted, while discrete attributes

(e.g. time, weekday, weather...) have a uniform probability in the scope of its own category to reduce the computation time stochastically. Anyway, all this kind of prior knowledge can be adjusted manually beforehand, or set a function to adjust these probabilities in runtime.

3.2.2. Anomaly detection

Detection of outliers or anomalous examples is very important in on-line learning because of the potential impact in the performance of the learner. For this reason, incoming samples are analyzed to detect anomalous samples and to avoid its learning. Considering the probability $P(X_j = x_{ij} | \mathcal{L})$ of observing a certain value x_{ij} in a rule R given the observed statistics in \mathcal{L} . We calculate this probability using the Cantelli's inequality (Bhattacharyya, 1987), which is a generalization of Chebyshev's inequality in the case of a single tail.

3.2.3. Change detection

Change detection, also known as *concept drift detection* in the machine learning community (Gama et al., 2014), is a critical component for modelling non-stationary processes as it is our case. For this purpose, each rule has associated a change detector which monitors their error. The idea is that, after a rule has been expanded and, thus, two new rules are created: their individual rule predictors are trained in their respective 'batch' mode with their corresponding gathered observations. So, starting from there, the residual mean error should be located at zero and it is started to be monitored for changes. When a change is detected, a signal is sent to the concept drift handler and the rule is removed from the ruleset. The current implemented approach for detecting a change is based on the Page-Hinkley (PH) test (Page, 1954), although other approaches are being considered (Bifet and Gavaldà, 2009, 2007).

3.2.4. Handling missing data

The framework gathers online statistics for each attribute in the context of each rule (which corresponds to a specific road conditions). So, in the long term, with enough sample size each rule has a good view of their data distribution for each recognized road condition. Thus, for each missing attribute, the framework reconstructs a normal distribution with the gathered mean and dispersion, but limiting the probability density at zero at the current minimum and maximum values in order to avoid extrapolation in the covariates.

3.2.5. Winsorizing

When extreme values (outliers) are received, i.e. those whose probability is extremely low in the scope of a specific rule, it is often better to filter them, or else replace them using the handler for missing data described above. Again, assuming a Gaussian distribution, this means considering outliers those values beyond or above approximately 3 standard deviations from the mean.

3.3. Expanding a rule

Rules could be viewed as high-level features discovered in the road network with the aim of reducing the uncertainty around the prediction target using a specific goodness of fit function. For this purpose, existing rules have a chance to run a rule expansion evaluation process. If the evaluation process is favorable, the current rule disappears and it is specialized into two new rules with their respective observations and statistics. The parameter N_{min} dictates the minimum amount of observations which must be seen, separately on each rule scope, to proceed with an evaluation. This threshold N_{min} is preset to an initial value N_{min0} , that is later dynamically adjusted (never increasing) based on the dispersion of the rule error in a logarithmic scale. More specifically:

$$N_{min} = N_{min0} \frac{1}{\ln(e + \sigma_{error})}$$

This dynamic adjustment aims at relaxing the tradeoff between prompt but expensive checks and slow but inefficient checks, because rules which are doing it well with a narrow error do not need to be specialized so often.

After selected combinations of features and split-points have been scored, the success of the rule expansion evaluation process is determined by using the ratio of the two best scores and a predetermined confidence-level on

the split must be guaranteed so that it can be expanded, by means of the Hoeffding bound (Hoeffding, 1963), as used in (Almeida et al., 2013; Gama, 2010).

3.3.1. Reducing the search

When it is time to run the expansion evaluation, it is needed to decide which attributes and split points are going to be measured. Perhaps the intuitive idea is to evaluate all the features, but in such high-dimensional problem this can lead to time-consumption problems and overfitting if, for instance, detectors that are very far away are selected as antecedents. Therefore, the candidates to be evaluated are selected probabilistically using the variable selector.

Continuous attributes considered include the traffic count, occupancy and speed from the whole road network. Discrete attributes considered include the time of the day, weekday and weather.

3.3.2. Scoring

The goodness of fit used to evaluate the different combinations of features and split-points is based on entropy minimization (or information gain). From an information theory perspective, entropy $H(X)$ measure the randomness of the information in the random variable X . The entropy is maximized if the distribution is vague (i.e. uniform with equal probability in the whole space), this is the situation of maximum uncertainty as it is most difficult to predict the outcome. When there is less uncertainty, this is when the outcome is peaked around certain location values, then this reduced uncertainty is quantified in a lower entropy. At the extreme case, when there is no uncertainty because we are sure about the outcome the entropy is zero (MacKay, 2003). The goodness of fit function also considers the missing data ratios of the feature candidates, penalizing those whose missing data ratio is higher (untrustworthy).

3.4. Rule prediction strategies (experts)

3.4.1. Weighted mean

The prediction is simply the weighted mean of the true target of the past examples covered by the rule.

3.4.2. Penalized linear regression

A linear regression model is built using the examples covered by the rule. Although short-term traffic prediction is a highly non-linear problem, this way we use the rules discover the nonlinearities and combine a population of lower-level, specialized linear models.

In addition to the sum of squared error loss, a penalty term (L1 norm) has been included in the minimization problem looking for shrinkage and sparse solutions, this approach is also known as LASSO (Hastie et al., 2015; Tibshirani, 1996). Concerning the learning procedure, coordinate-wise gradient descent has been used to obtain the parameter estimates because it applies well to our case where $n \ll p$, and it has been successfully applied to this problem in high-dimensionality settings (Friedman et al., 2010), and demonstrated to be efficient in large problems (Nesterov, 2012). When training in batch mode, a regularization path is obtained efficiently using warm-up starts for different penalty λ values, which means that multiple solutions exist ranging from the least penalized (ordinary least squares solution) to most penalized and sparse solution. In online learning, the coordinate descent is also applicable using mini-batches of data, a small learning rate and the soft-thresholding technique (Shalev-Shwartz and Tewari, 2011)..

3.4.3. Adaptive strategy

Finally, an adaptive strategy is proposed which combines the forecasters population derived from the above two strategies within a rule into a single point-estimate value using for it the on-line estimation of the mean absolute error (MAE). This on-line estimation T of the error is updated following a fading factor strategy which controls the importance of the oldest and newest examples. Finally, the contribution from each forecaster to the final point-estimate prediction is determined inversely proportional to their current online estimation of the error. A similar strategy is followed when multiple rules cover a single observation and an adaptive response is given in addition to the individual rule responses. Finally, in addition to the prediction point estimate, an uncertainty interval is given based on the error seen.

3.5. Regularization

The previous procedures, namely the error-based timing for rule expansion evaluation process, the Hoeffding bound, the drift detection, the penalty to untrustworthy features, the minimum number of observations for splitting, and the shrinkage and sparsity achieved in the penalized linear regressions, are used altogether as a way of regularization to avoid overfitting.

4. Experiments

The data used in this paper comes from the Caltrans Performance Measurement System (PeMS) maintained by the California Department of Transportation. More specifically, the current work has focused the attention into the Caltrans District 11. It spans the entire California-Mexico Border from the Pacific coast to Arizona and reaches north from the international border to Orange and Riverside counties. The district has roughly 1,000 centerline miles of urban and rural freeways, monitored with over 1,500 detection stations. Collected data for experiments spans for three years ranging from 2013/01 to 2015/12 with an initial 5-min resolution with has been aggregated to 15-min for three reasons: (a) mitigating the inherent noise in road network measuring devices, (b) reducing the running time for the experiments in the current research work without compromising the validity of the results, and (c) convenience for commercial purposes from TSS-Transport Simulation Systems and its product Aimsun Online. Finally, because the final intention is that the resulting product from this research work is to be integrated within Aimsun Online we have put our attention into predicting traffic volume.

For the experiments, a subset of detection stations from the road network has been chosen in order to focus on a small set of locations which are far apart, but exhibiting high distributional changes.

In addition, the goal of our experiments is to compare the performance of our real-time learning framework versus a more classical approach of training a model in batch model with past data and putting it into production. The batch approach is based on the same linear regression penalized with L1 (LASSO), the best λ penalty value is selected using cross-validation on the training dataset, and the input information is composed from the traffic volumes existing in the whole road network at the same time. Since the short-term traffic prediction task is non-linear and we are trying to model it with a linear model, we cover up this non-linear behavior by discretizing the task according to the data resolution (15 minutes) and training a separate set of coefficients for each time point. Another approach has been compared, which is based on a blind adaptation; this is retraining the whole models set every 1 week or 1 month, and using the last 1 month or 6 months for the training dataset.

Finally, a 60-min forecasting horizon has been chosen to evaluate the different approaches because it is a challenging and interesting horizon for commercial purposes, although a great part of the literature has been found to focus on forecasting horizon shorter than 15-minutes.

The performance metric considered to compare results is the mean absolute percentage error (MAPE) because it gives an intuitive measure of the performance independent of the unit and scale, and a time interval of one month has been used to aggregate the MAPE values over.

The results of the experiments are shown in Fig 1. As can be seen, the MAPE for the adaptive approach is initially high in all the stations because the framework starts with no knowledge and then it starts to learn and adapt its parameters as in real-time. On other hand, the approaches based on batch training start with a low error because it is just the data which has been used for their training (the first 6 or 12 months from the 3 years). Obviously, it is not a fair comparison, but the aim is just to show how Adarules lower its error as more data is seen and more knowledge is acquired. It can be seen also that the batch approach trained with more data (1 year) has a lower error than the batch approach trained with less data (6 months) but the difference is slight. Another interesting point is observing how the adaptive approach deals better when sudden changes happen, while the performance for the batch approaches deteriorates and does not seem to recover. It can be seen also that the performance using the adaptive approach is improving until the end of the experiments. Finally, when it is compared to the blind adaptation approaches; it can be seen that the accuracy performance is similar on the long term, however a crucial difference is the Adarules autonomy to decide the training times avoiding unnecessary training costs every 1 week or 1 month. Besides collateral benefits from Adarules explained in previous section, another crucial difference not noticeable in the picture because of the results aggregation, is that Adarules is giving responses even with missing data.

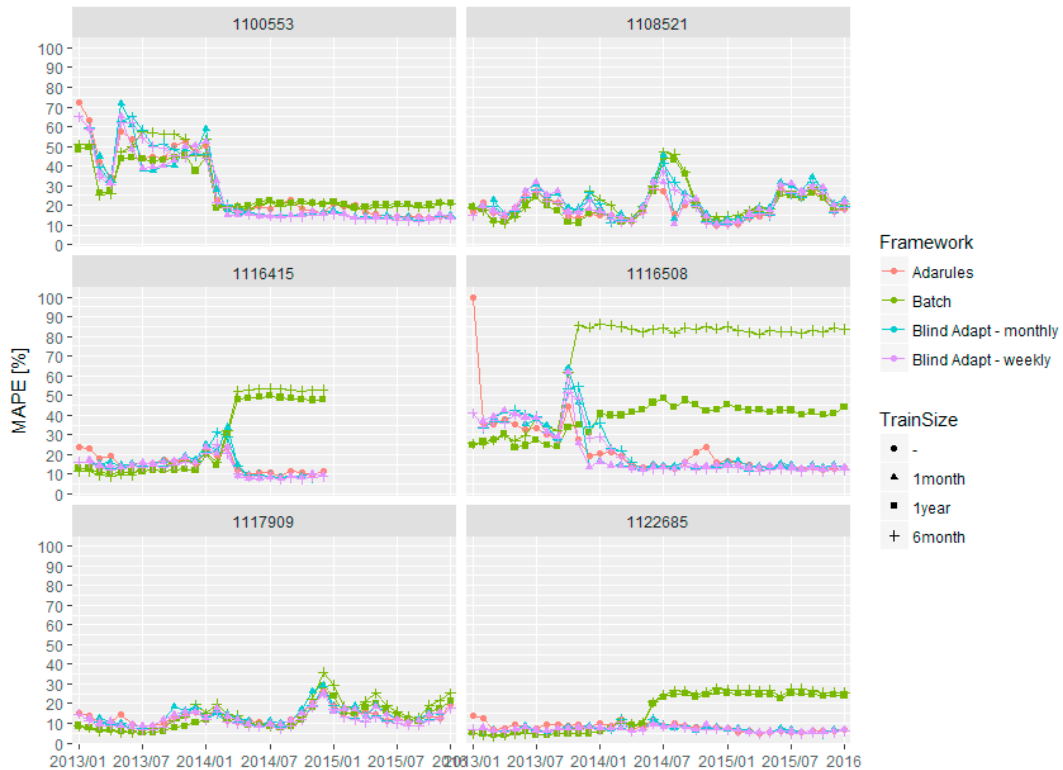


Fig 1. Experiments results for six detection stations and 60-minute forecasting horizon, showing the monthly aggregated MAPE for the three approaches and its evolution over time during the three years.

5. Conclusions

The experimental results have confirmed the expectations about the proposed Adarules, which are a good tolerance and fast adaption to change, specially sudden changes and long term changes associated with seasonality or traffic demand growth. Second, that as it sees more data, the framework learns and maintains its predictive accuracy. Third, it is valuable that the framework unveils the inherent dependencies in the road network (which can be seen as high-level features in the machine learning argot) and, also importantly, these can be easily interpreted and evaluated by traffic managers. The use of contextual information (e.g. date, time and weather) and the measurement of its impact is especially attractive for traffic managers. Another interesting advantage of Adarules is that it can give responses for the predictive point even if it is temporally malfunctioning if there is enough knowledge acquired and thus reconstructing its typical behaviour. Finally, Adarules is an autonomous framework and can manage the tradeoff for deciding the proper training times to adapt quickly to changes while keeping a good prediction accuracy.

Finally, an important remark about road network traffic prediction is that prediction accuracy is very important, but it cannot be the only criterion when choosing the appropriate modelling methodology (Kirby et al., 1997). Given that this task is a non-stationary stochastic process tackled in real-time, other matters concerning the adaptability to changing behaviors and traffic demand changes, transferability to new locations with scarce data or information about the traffic supply characteristics, causality and interpretability about the process, and cost in time and effort for model development and, more importantly, maintenance must be considered. Some of these challenges are pointed out in (Vlahogianni et al., 2014), which can be summarized in the following points that we believe that the proposed predictive framework Adarules in the current research work aims to deal with them: (1) responsive forecasting schemes for non-recurrent conditions, (2) developing prediction systems with increased algorithmic complexity, (3) attempting to understand data coming from novel technologies and fuse multi-source traffic data to improve predictions, (4) the applicability of artificial intelligence (AI) methodologies to the short-term traffic prediction problem.

Acknowledgements

The authors would like to thank California Department of Transportation for providing the traffic data within the Caltrans Performance Measurement System (PeMS). Work by R. Mena-Yedra is funded by the PhD grant: Pla de Doctorats Industrials (DI-2014) by AGAUR. Work by R. Gavalda is partially funded by AGAUR project 2014 SGR-890 (MACDA) and by MINECO project TIN2014-57226-P (APCOM).”

References

- Almeida, E., Ferreira, C., Gama, J., 2013. Adaptive Model Rules from Data Streams, in: Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases - Volume 8188, ECML PKDD 2013. Springer-Verlag New York, Inc., New York, NY, USA, pp. 480–492.
- Barceló, J., 2010. Fundamentals of traffic simulation, International series in operations research & management science. Springer.
- Bhattacharyya, B., 1987. One sided Chebyshev inequality when the first four moments are known. *Commun. Stat.-Theory Methods* 16, 2789–2791.
- Bifet, A., Gavalda, R., 2009. Adaptive Learning from Evolving Data Streams, in: Proceedings of the 8th International Symposium on Intelligent Data Analysis: Advances in Intelligent Data Analysis VIII, IDA '09. Springer-Verlag, Berlin, Heidelberg, pp. 249–260. doi:10.1007/978-3-642-03915-7_22
- Bifet, A., Gavalda, R., 2007. Learning from Time-Changing Data with Adaptive Windowing, in: Proceedings of the 2007 SIAM International Conference on Data Mining, Proceedings. Society for Industrial and Applied Mathematics, pp. 443–448.
- Casas, J., Ferrer, J.L., Garcia, D., Perarnau, J., Torday, A., 2010. Traffic simulation with Aimsun, in: Fundamentals of Traffic Simulation. Springer, pp. 173–232.
- Ermagun, A., Levinson, D.M., 2016. Spatiotemporal Traffic Forecasting: Review and Proposed Directions.
- Ferrer, J.L., Barceló, J., 1993. AIMSUN2: advanced interactive microscopic simulator for urban and non-urban networks. Research report. Departamento de Estadística e Investigación Operativa, Facultad de Informática, Universidad Politécnica de Cataluña.
- Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1.
- Gama, J., 2010. Knowledge discovery from data streams. CRC Press.
- Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., Bouchachia, A., 2014. A Survey on Concept Drift Adaptation. *ACM Comput Surv* 46, 44:1–44:37. doi:10.1145/2523813
- Hastie, T., Tibshirani, R., Wainwright, M., 2015. Statistical Learning with Sparsity: The Lasso and Generalizations. Chapman and Hall/CRC.
- Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* 58, 13–30.
- Kirby, H.R., Watson, S.M., Dougherty, M.S., 1997. Should we use neural networks or statistical models for short-term motorway traffic forecasting? *Int. J. Forecast.* 13, 43–50.
- MacKay, D.J., 2003. Information theory, inference and learning algorithms. Cambridge university press.
- Nesterov, Y., 2012. Efficiency of Coordinate Descent Methods on Huge-Scale Optimization Problems. *SIAM J. Optim.* 22, 341–362. doi:10.1137/100802001
- Page, E., 1954. Continuous inspection schemes. *Biometrika* 41, 100–115.
- Shalev-Shwartz, S., Tewari, A., 2011. Stochastic methods for l_1 -regularized loss minimization. *J. Mach. Learn. Res.* 12, 1865–1892.
- Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* 58, 267–288.
- Treiber, M., Kesting, A., 2013. Traffic flow dynamics. Traffic Flow Dyn. Data Models Simul. Springer-Verl. Berl. Heidelb.
- van Lint, H., van Hinsbergen, C., 2012. Short-Term Traffic and Travel Time Prediction Models. *Transp. Res. E-Circ.*
- Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2014. Short-term traffic forecasting: Where we are and where we're going. *Transp. Res. Part C Emerg. Technol., Special Issue on Short-term Traffic Flow Forecasting* 43, Part 1, 3–19. doi:10.1016/j.trc.2014.01.005