DISCLAIMER

This deliverable has been submitted but has not been approved by the EC yet

"Enhanced data management techniques for real-time logistics planning and scheduling"

Deliverable D3.1: State of the art in artificial intelligence techniques focussed on prediction

Dissemination level:

☒ Public ☐ Confidential, only for members of the consortium (including the Commission Services)

Version number: 1.0

Submission deadline: 31/03/2019

www.logistar-project.eu

DOCUMENT INFORMATION

Authors

Name	Organisation
Bastien Pietropaoli	UCC
Nic Wilson	UCC
Ken Brown	UCC

Reviewers

Name	Organisation
Enrique Onieva	DEUSTO
Naia Merino	DEUSTO
Leire Serrano	DEUSTO

Document control

Version	Date	Comment
0.1	25/03/2019	First version
1.0	09/04/2019	Comments included. Final version.

Document approval

Version	Date	Partners
1.0	09/04/2019	All partners

BASIC PROJECT INFORMATION

Horizon 2020 programme

H2020 - Mobility for Growth- 5-2-2017. Innovative ICT solutions for future logistics operations

Grant Agreement No. 769142

TABLE OF CONTENTS

LIST OF FIGURES

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Executive Summary

This document provides a review of the literature on artificial intelligence (AI) with a particular focus on predictions for logistics networks. General principles and methods used in AI are first presented, followed by a review of the literature on intelligent transport systems (ITS) for roads, airways, waterways, and railways. Insights on current and future research are then provided with a focus on the difficulties specific to LOGISTAR.

1. Introduction

Artificial intelligence (AI) is becoming more and more important in all types of industries, and major IT companies are publishing powerful open-source frameworks to ease the development of AI-based applications (e.g. TensorFlow from Google or CUDA-X AI from Nvidia). These frameworks provide many statistical and machine learning models that can be applied to different types of data, for classification, regression (e.g. to estimate the value of something), prediction, clustering and other inference tasks.

One of the main objectives of LOGISTAR is to improve transport operations in logistics. Following the discussions conducted in WP1 and the needs expressed by the various partners in the consortium, we have identified a few key prediction types that WP3 will be focusing on:

- Travel time: *How long will it take to go from A to B?* Accurate estimates of travel time can be extremely important for planning, as they enable tighter schedules and reduced delays, hence improving the quality of the delivery service. This can be seen as a spatiotemporal series problem.
- Turnaround time: *How long will it take to go through client X? How long will it take to switch from a truck to a train?* Better predictions of turnaround time will enable tighter schedules of deliveries. This can be seen as a regression problem and a time series since the turnaround times might be time-dependent (i.e. seasonality).
- Delays and delay propagation: *Will this delivery be late? Will this affect other deliveries?* Linked to travel time predictions, predictions of delays and the propagation of delays will enable proactive solutions instead of reactions after the fact. This can be seen as a decision-making problem based on travel time and turnaround time predictions.
- Delivery risk assessment: *Will this delivery fail and if so, why?* Delayed deliveries which arrive outside of contracted time windows may be rejected. . Certain paths may be more prone to damaging the products, which also causes rejected deliveries. Assessing the risk associated with every delivery will help better plan the logistics operations. This can be seen as a classification problem (i.e. success / failure with an associated level of confidence) but also a time series problem since certain periods of the year might more prone to failures (e.g. clients might be busier around Christmas and more prone to reject a late delivery).
- Orders and deliveries: *Who will order what and when?* Order and delivery predictions will also help planning by providing a longer horizon of probable orders and deliveries to be scheduled. This is a typical case of a time series problem.

The aim of this document is to provide a quick overview of the State-of-the-Art in AI, mainly focussed on predictions for logistics. Since AI (encompassing both statistical and machine learning methods) is a subject too vast to be fully covered, we quickly present in section 2 the main principles of classical methods found in the literature. This covers all the required tools for the predictions types presented above.

Section 3 and 4 will then focus mostly on Intelligent Transport Systems (ITS) for various types of routes as synchro-modality is one of the use cases of LOGISTAR. However, since the literature seems to focus a lot more on terrestrial route problems than any other means of transportation, they are covered in most detail in this report. Since the study of traffic goes back to the 1950s (Lighthill & Whitham 1955a; Lighthill & Whitham 1955b), we try to provide a representative sample of the various approaches used in the domain, more particularly in the last few years, by highlighting the recent trends found in the literature. Also, it is important to note that most of the techniques reviewed here, while applied to specific problems, could easily be ported to be used for the different types of predictions WP3 will focus on.

We then provide insights on current and future research in the domain in section 5. The main difficulties and the most promising approaches are discussed. Finally, we summarise the main findings in section 6.

2. AI and statistics for predictions

Artificial intelligence is a large sub-domain of computer science with many aspects (Russell and Norvig, 2009). Covering the entirety of AI is clearly out of the scope of this document. In this section, we quickly present some of the main aspects of AI. The next section focuses on the application of the principles presented here to problems linked to ITS.

2.1. Statistical methods

The simplest approach for time series prediction consists of averaging historical data. It can be performed at different levels of granularity to provide better accuracy. In the example of traffic prediction, multiple averages can be considered to differentiate weekdays from weekends and peak hours from free-flow hours (e.g. at night). While limited in terms of accuracy, averages tend to be more accurate for longer-term predictions (Xie and Choi, 2017).

Another classic approach in the prediction of time series is using the Box-Jenkins method, often called autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) (Box and Jenkins, 1970; Cryer and Chan, 2008). Those models are particularly good at dealing with stationarity (e.g. stochastic processes that stay stable over time) and seasonality (i.e. changes at regular time intervals).

Various models have been derived from ARMA and ARIMA to take into account more variables. For instance, vector autoregressive (VAR) models are capable of handling multiple varying values at once. Those models can also include explanatory variables to create models usually called ARIMAX or VARMAX (Hyndman and Asthanopoulos, 2018). These statistical methods are widely used in econometrics (Stock and Watson, 2003).

The average of historical data and ARIMA models are often used as a baseline in the comparison of the accuracy of prediction algorithms.

2.2. Models and simulations

Model-driven prediction methods use expert knowledge of a domain and physics-like equations to model the behaviour of a system, as opposed to the data-driven methods described in section 2.1. Oh et al. (2015, 2017) discuss how model-driven approaches have been applied to traffic and travel times. Traffic modelling has been done by physics equations to model the behaviour of queues of vehicles (Lighthill & Whitham 1955a; Lighthill & Whitham 1955b), cellular automata (Dilip et al. 2018), or large commercial simulators such as INRIX.

Model-based methods and simulators will not be considered for LOGISTAR since they require a deep understanding of the underlying mechanisms in action in a system and are computationally costly as they require fine granularity that seems incompatible with the large scale of LOGISTAR, and hence are not considered further in this document.

2.3. Machine learning (ML)

Machine learning is a subdomain of AI that has gained a lot of traction in recent years (Burkov, 2019; James *et al.*, 2017; Goodfellow *et al.*, 2016). It consists of "the programming of a digital computer to behave in a way which, if done by human beings or animals, would be described as involving the

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

process of learning" (Samuel, 1959). ML includes a plethora of methods, and covering the entirety of the subject is out of the scope of this document. We present here the general principles of a few methods that are commonly found in the literature on machine learning and predictions for logistics.

2.3.1. k-nearest neighbours

The k-nearest neighbours (kNN) algorithm is a simple yet efficient technique used for clustering, classification, and regression (Altman, 1992). It usually consists of first defining a distance measure between points in a dataset and then averaging a value based on the k-nearest neighbours, in the dataset, given the defined distance measure. Figure 1 illustrates a classification example of a 1-NN (a single nearest neighbour) algorithm.

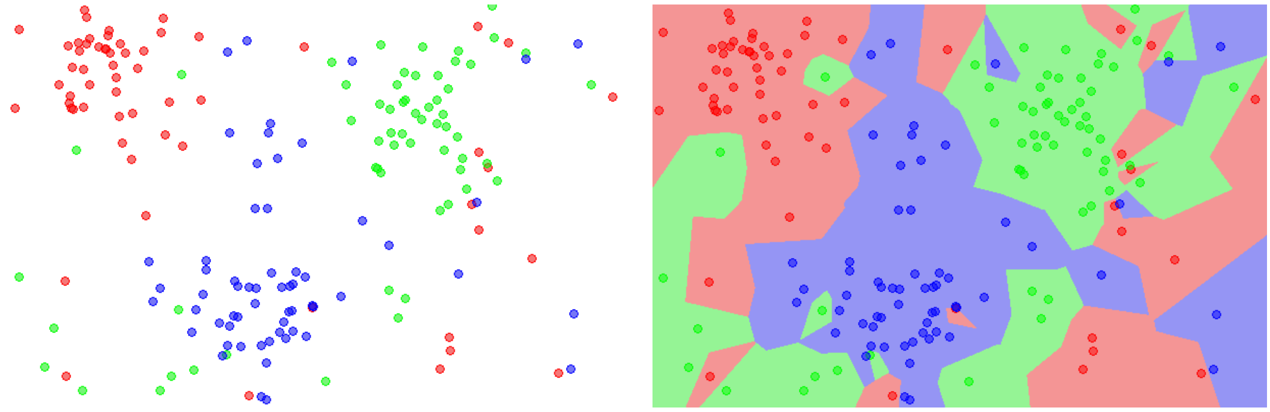

Figure 1 A dataset of points of 3 different classes (on the left) and the membership map based on 1-NN (on the right) showing to which class the rest of the space would be associated.
Source: [Wikipedia]

In the case of ITS, it is often used as a regression method to predict traffic or travel time with a spatiotemporal distance measure taking into account the physical proximity of road segments and the closest related times in the historical data (Cai *et al.*, 2016). This usually consists of recent traffic or travel time measures if available, but also measures taken at similar times such as the same time of the day, for the same day of the week. It is a common method used in traffic and travel time predictions (Cheng *et al.*, 2018; Hou *et al.*, 2013; Li *et al.*, 2012; Lim and Lee, 2011; Klunder, 2007).

2.3.2. Artificial neural networks

One of the most popular and probably the most emblematic method used in ML is the use of artificial neural networks (ANN) as a general framework. It usually consists of nodes (or neurons) connected to inputs and outputs. Multiple layers of nodes may be chained to create what is called deep neural networks (Bengio, 2015). In its simplest form as illustrated in Figure 2, often called a perceptron (Rosenblatt, 1958), inputs are provided to a single hidden layer of neurons. Each neuron is usually composed of an activation function that will be "learned" based on a training set. Each neuron is thus creating unique features.

A vast literature exists on the subject of ANNs with numerous variations of them. In recent years, recurrent neural networks (RNN) (Hochreiter *et al.,* 2001) have gained a lot of traction with a plethora of successful applications, most notably in natural language processing (NLP) (Hinton *et al*, 2012; Fernandez *et al.*, 2007). This new interest in neural networks has come from new models such as

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Long Short-Term Memory (LSTM) networks (Hochreiter and Schmidhuber, 1997) and a large community of popularisers.

RNNs and more particularly Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) (see Figure 3) have been applied to predict all sorts of sequential data (The Unreasonable Effectiveness of Recurrent Neural Networks) and are particularly interesting in the case of LOGISTAR since many of our predictions concern time series that are sequential by definition. Their remarkable efficiency makes them an excellent potential candidate for our prediction systems.

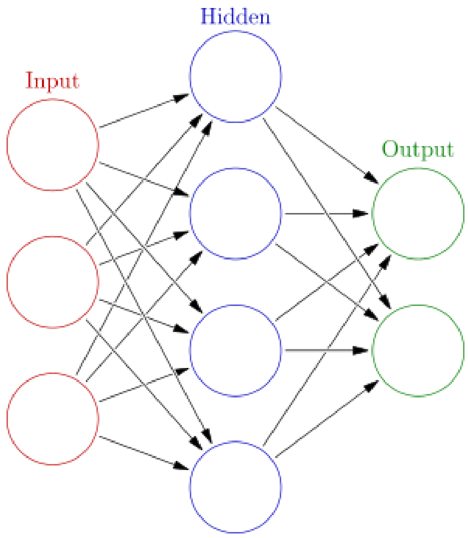

Figure 2 Artificial neural network with a single hidden layer (also called perceptron). Source: Wikipedia

However, even though neural networks seem to perform well in all sorts of tasks (e.g. classification or regression), in particular in the framework of deep learning, there are a few criticisms that cannot be ignored.

First, the learned models can be quite obscure and difficult for a human to interpret. These black boxes may lead to situations in which if a mistake was made, it is extremely difficult to interpret why that happened (The U.S. Military Wants Its Autonomous Machines to Explain Themselves). Recent progress has been made towards interpretable neural networks (Chen *et al.* 2018, Li *et al*. 2018).

Secondly, the larger the network the more powerful the results, but the larger the training set required. This can lead to situations in which enormous training sets are required even for simple tasks which leads to performance and scalability issues (Edwards, 2015).

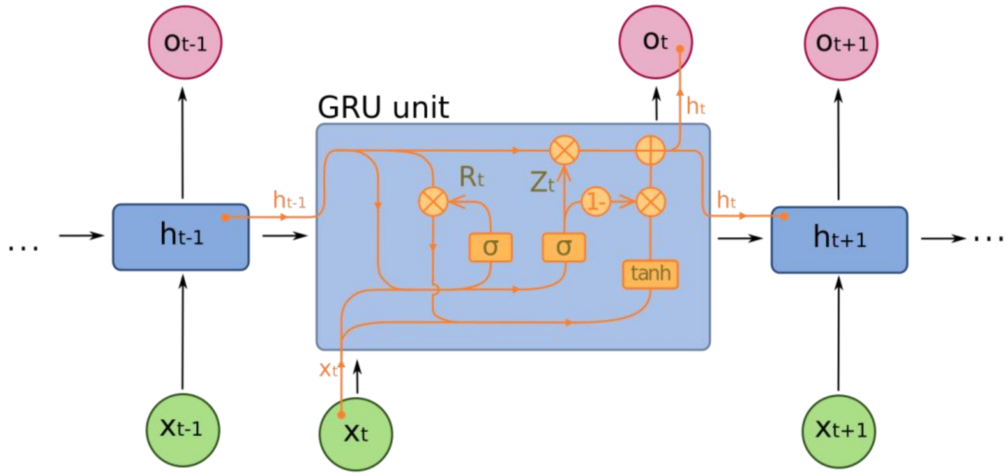

Figure 3 Unraveled RNN using GRU units. Source: [Wikipedia]

2.3.3. Trees and random forests (RF)

Decision trees and regression trees are simple decision-support tools that can be built automatically to generate decision models (Utgoff, 1989; Quinlan, 1983). As illustrated in Figure 4, they consist of a multi-step decision based on values associated with a point in a dataset. They are well-known for being used in loan default predictions (Heryati *et al.*, 2019, Khemakhem and Boujelbene, 2018).

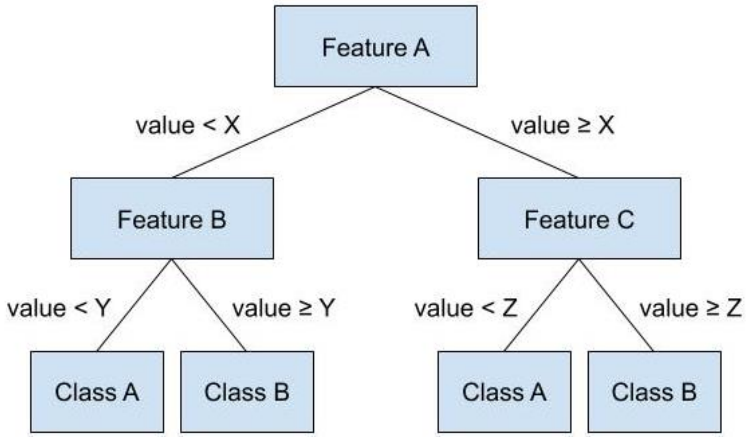

Figure 4 Example of a decision tree based on 3 features to classify something into 2 possible classes.

Decision trees and regression trees gained traction again in recent years with the development of boosted trees (Freund and Schapire, 1996) and random forests (Breiman, 2001). This consists of creating thousands of shallow decision trees based on various samples of features and training points and then merging together the result of all of these trees into a single prediction.

Obviously, ML is much broader than what we have presented in this section. A plethora of other methods exists such as linear regressions (Seal, 1967), support vector machines (Cortes and Vapnik, 1995), or even ad hoc time series predictions based on unsupervised clustering (Leverger *et al.*, 2018).

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

3. Predictions for terrestrial routes

Prediction for terrestrial routes is a vast subject with literally hundreds of papers published on the subject since the 1950s. The interested reader can refer to the numerous surveys of the domain for more complete reviews of the State-of-the-Art (Oh *et al.* 2017; Oh *et al.* 2015; Barros *et al.* 2015; Mori *et al.* 2014; Bolshinsky *et al.*, 2012; Khrisnan and Polak, 2008; Lin *et al.*, 2005, Taylor *et al.*, 2005; Vlahogianni *et al.* 2004).

As explained above, in this section, we focus on data-driven approaches. The intention is to cover the most interesting and/or promising approaches found in the recent literature.

3.1. Traffic flow / conditions

In the literature, traffic flow and traffic conditions are two terms that refer to the fluidity of the road network or specific road segments. They encompass a set of measurable values: number of vehicles per minute, average vehicle speed, road occupation, average distance between vehicles, etc.. The most common measures used are the number of vehicles per minute and the average vehicle speed. It is important to notice that the average vehicle speed is usually not equivalent to the travel time since it usually does not take into account interruptions such as stop signs, or signalled intersections, nor does it take into account the turning delay at intersections.

Du *et al.* (2018), perform a 15-minute ahead prediction of the number of vehicles going through all the UK motorways. To do so, they use a mix of a Convolutional Neural Network (CNN) and an RNN based on GRU. They apply this mixed neural network to various time series provided in their dataset (vehicle flow, vehicle speed, etc.) and merge the results with a custom method. They achieve extremely high accuracy overall, particularly because their method is capable of correctly adapting to rush hour peaks. They compare their approach to more regular approaches such as ARIMA, Support Vector Regression (SVR), and various configurations of RNNs, LSTMs and GRUs, showing that their method outperforms all of the others.

Q. Zhang *et al.* (2017) use Extreme Learning Machines (ELM) to predict daily vehicle counts on California highways using the PeMS dataset. They compare their method with linear regressions. The interesting part of their work is that they include historical data but also air pollution data, and search engine data to retrieve information about special events in San Francisco. ELMs are a type of neural networks designed to learn faster than classic neural networks. As shown in their work, ELM unsurprisingly outperforms linear regressions, even when the size of the training set is drastically reduced. The accuracy of ELM is also less affected by the reduction of the training set size.

Yuan & Tu (2017) use a classic artificial neural network (ANN) (see Figure 2) but exploit mutual information to select the best features. By also using the PeMS dataset, they show that their method for feature selection improves the prediction results, notably a lot more than Principal Component Analysis (PCA), which is a common method in machine learning used to perform feature selection.

Xie & Choi (2017) predict the average speed on road segments in Hong Kong. They compare the results provided by a simple historical average with an ARIMA model and a Periodical Moving Average (PMA). While ARIMA performs well for short-term predictions, its accuracy decreases with higher prediction horizons while PMA seems to have a more constant accuracy. Leveraging this, they use a neural network to combine the result of the two approaches by adding information about the current time and the horizon. They then complete their approach with a Bayesian Network (BN) to better handle non-recurrent events such as traffic accidents. Their approach offers slightly better predictions than the approaches they compared to.

Soua *et al.* (2016) exploit historical traffic flow data as well as weather data and localised Twitter data to feed a deep belief network using Restricted Boltzmann Machines (RBM). RBMs are again a special type of neural network. They then use the Dempster-Shafer theory (Dempster, 1969; Shafer, 1976) to combine the results of multiple RBMs.They also test their approach with data from the PeMS dataset. Their approach outperforms a classic ARIMA approach as well as a classic neural network.

From the work cited above, we can extract that neural networks and their variations are a major trend in the modelling of traffic flow. They seem to consistently outperform more classic statistical approaches such as ARIMA, which is often used as a baseline for comparison. However, a combination of multiple algorithms is often used, showing that ensemble methods might be the way to improve the prediction performance, even when combining weak predictors (e.g. as in (Xie & Choi 2017)).

3.2. Travel time

While traffic flow can be extremely useful for route planning, travel time seems to get more attention in the literature. In this subsection, we try to provide a quick overview of the various ways travel times are estimated and predicted.

3.2.1. Travel time estimation

In order to improve travel time predictions, one might need to better estimate the travel time from incomplete and stochastic data provided by sensors placed on roads. Multiple papers can be found in the domain of travel time estimation providing probability distributions of travel times:

- Jabari *et al.* (2018) propose an estimation method from sparse time travel measurements (e.g. coming from probe vehicles) based on a mixture distribution using Mittag-Leffler functions instead of the classic Gaussian distributions. They achieve better goodness of fit while providing compression by drastically reducing the number of distributions used in the mixture. With their method, they manage to provide precise memory-efficient time-varying travel time distributions.
- In (Shi, B. Y. Chen, *et al.* 2017), an estimation of travel time is achieved based on two types of road sensors: point detectors (e.g. loop detectors) and interval detectors (e.g. automated vehicle identification systems). To achieve this, they merge the two travel time distribution estimates using Dempster-Shafer theory and then impute travel times for neighbouring links. Their method outperforms the estimates computed from only one of the two types of sensors.
- Another interesting way to compute the travel time is to combine travel time for road segments and the delay incurred by turning from one road segment to another. That is what authors in (Shi, B. Chen *et al.* 2017) do using sparse probe vehicle data from overlapping paths.

Simple predictions based on stochastic estimates of travel time can already provide richer travel time predictions than the mean of the historical travel times (also called historical means). A historical probability distribution can already provide a better image of what the travel time might look like for a road segment.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

3.2.2. Segment-based approaches

Segment-based (also called link-based) approaches aim at predicting travel time for each road segment separately. The methods used for predictions for road segments are numerous:

- A probabilistic principal component analysis (PPCA) combined with a simple smoothing based on incomplete data provided by probe vehicles is used in (Jenelius and Koutsopoulos, 2018). Their approach is interesting since it enables predicting for an entire network and their experiment on the city centre of Shenzen, China shows promising results.
- A CNN to capture spatiotemporal features and an LSTM to predict time series, both mixed together with a regression layer is used by Z. Zhang *et al.* (2017). Tested on probe vehicle data for an urban expressway in Beijing, China, their approach demonstrates better performance than if only a temporal approach was used, showing the importance of taking the spatiotemporal aspect of traffic into account.
- Wang *et al.* (2017) use a Gaussian Process Regression (GPR) to predict the travel time of a highway section in Delft, the Netherlands using sensors placed on the road providing information about the current travel time and the traffic flow. The simplicity of their method and the good results are promising but their method has been tested only for highways.
- Using probe data from taxis in Stockholm, Sweden, Rodriguez-Deniz *et al.* (2017) also used a GPR to predict travel time but achieved only slightly better than historical means.
- Liu *et al. (*2017) apply various deep learning approaches to data from the PeMS dataset. Their experiments show the importance of hyperparameter tuning when it comes to deep neural networks. Also, they observed that LSTM combined with a deep neural network (DNN) performed best compared to other DNNs.

While segment-based approaches work well for highways composed of long single segments, they suffer a significant decrease in terms of accuracy when it comes to shorter road segments. Obviously, this is an issue for urban environments composed mainly of short segments. Moreover, the lack of consideration for turning delays, and the numerous events that can influence the travel time in an urban environment (e.g. interactions with pedestrians, cyclists, and other vehicles) render the prediction at the segment level extremely difficult (Zheng *et al.* 2017).

3.2.3. Path-based approaches

To overcome the issues of the segment-based approaches, many researchers work directly with paths. Those paths are usually obtained either via planned routes or via probe vehicle data (GPS tracks). Path-based (also sometimes called route-based) approaches offer the advantage of reduced variance compared to segment-based approaches.

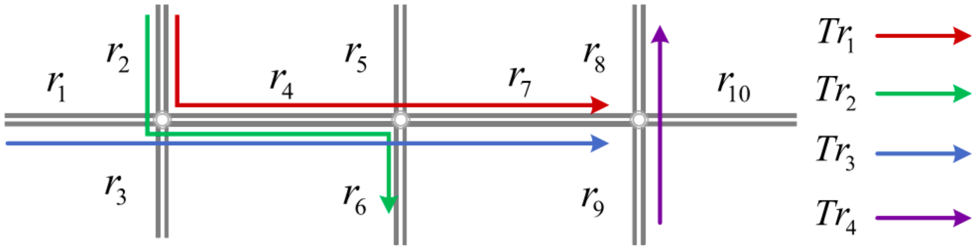

Figure 5 - Different trajectories (paths) may overlap and provide information about specific road segments. Source: (Tang et al. *2018)*

As for segment-based approaches, there exist numerous ways to perform path-based travel time predictions:

- Waury *et al.* (2018) use a custom method to determine histograms of travel times of road segments based on historical path data provided by GPS tracks of vehicles in Aalborg, Denmark. They test the impact of multiple parameters on the quality of the predictions: driver-specific data, weather data, sample size, temporal filtering, congestion, and road segment type. They observe large differences from one type of road segments to another, confirming other works on that matter. Surprisingly, they find that the use of weather data only slightly improves the accuracy of their predictions.
- Utilising tensors combined with a probabilistic approach and a regularisation to impute missing data, Tang *et al. (*2018) provide travel time predictions from sparse probe vehicle data provided by taxis equipped with GPS in Beijing, China. While outperforming more basic approaches, the time and space complexity of their method might make it undesirable for a live system.
- Xu *et al.* (2017) predict the travel time between three important regions of a road network (in Shanghai, China) using a classic neural network trained on historical travel time and weather data. They use GPS tracks of taxis to learn the travel times for various routes. They can then predict travel times of planned routes given the route and the weather forecast. Their approach is interesting because it could easily be applied to delivery routes in LOGISTAR without having to process the entire European road network.
- Wen *et al.* (2017) study the travel time between locations in a road network, more specifically, they study what appears to be a logistics network. By having overlapping routes, they manage to infer estimates for sub-routes. They then use a probabilistic approach based on historical data to provide travel time distribution estimates for routes.

Path-based approaches seem more suitable to logistic networks since they can easily provide estimates and predictions for frequently used routes, or by helping to select the best route given historical data when planning future routes. This type of approaches will hence be seriously considered in LOGISTAR.

3.2.4. Neighbours-based approaches

Neighbours-based approaches use the neighbouring links of a road segment to infer its traversal time. The most well-known and most commonly used method for this is the k-nearest neighbours (kNN). Here are some of the works found in the literature that exploit neighbours to predict travel times:

- Cheng *et al.* (2018) build on top of kNN by adding a dynamic selection of k and a dynamic time window depending on spatiotemporal considerations such as the presence of traffic jam. This enables their kNN algorithm to make sense of contextual data. They test their approach for short-term predictions on data from PeMS and data from highways in Beijing, China. Their approach outperforms a classic application of kNN, the historical average, and also an Elman neural network (which is a type of RNN). They also investigated the impact of the measure distance used on the performance of their kNN algorithm.
- Vu *et al.* (2017) do not use kNN but instead use a classic neural network that learns travel time patterns of a road segment from its neighbouring links. Improving the learning by filtering outliers using a Gaussian mixture model, their model outperforms a classic statistical approach and a linear regression. The shortcoming of this work comes from the fact that travel time on the road segment used for predictions is still necessary for the training phase. Such data might be sparse due to limited coverage by probe vehicles for instance.
- Enhancing the classic kNN with their own spatiotemporal distance measure, Cai *et al.* (2016) improve even further the already good performance of kNN compared to a historical average, a Support Vector Machine (SVM), and an Elman-NN. They validate their approach on data collected via probe vehicles in Beijing China.

3.2.5. Hybrid approaches

Hybrid approaches combine two or more of the aforementioned approaches to provide travel time predictions. Usually, the idea is to overcome the shortcomings of any single method by combining the results of multiple ones.

As already mentioned, (Shi, B. Chen, *et al.* 2017; Shi, B. Y. Chen, *et al.* 2017) use a mixed approach to estimate the travel time of road segments. However, mixed approaches are fairly rare in the prediction literature. Lim & Lee (2011), in one of the few papers using a mixed approach for prediction, use both point and interval detectors and combine local and neighbour measurements using a kNN approach. Their method is thus mixing the three main branches described above. Testing their approach on data from sensors placed along a highway in Yangjae IC, South Korea, they obtain good results but do not compare them to any other approach.

On the subject of travel time prediction, we can notice that most of the literature focuses on short-term prediction (5-60 minutes). Obviously, the longer the prediction horizon, the lower the accuracy, whichever method is used. The only method found in the literature that seems to perform acceptably well for long-term predictions (at least a day ahead) is using a historical average of the measured travel times (Klunder, 2007). Methods using historical probabilistic distributions of travel times might also be a possibility for long-term predictions.

3.3. Travel time variability

Travel time variability aims at understanding how the traffic varies over time and how it depends on various conditions, for instance under various weather conditions. While this is not directly linked to predictions, it still provides insights into what are the main factors influencing traffic. This data analysis/mining approach and its results could be used to create hybrid models leveraging the best of model-driven methods and data-driven methods. In this section, we present some interesting work found in the literature.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Chen *et al.* (2018) study the travel time variability of different types of roads in Beijing. To do so, they gathered data using probe vehicles (i.e. taxis equipped with a GPS) roaming in the city over a week. They covered 200 road segments of four different types: urban expressways, auxiliary roads of urban expressways, major roads, and secondary roads. Their work is particularly interesting since it highlights that there are real differences in terms of patterns between urban highways and smaller urban roads. Notably, they observe that small roads do not necessarily show peak patterns during the weekend.

Zheng *et al.* (2017) study the travel time variability of segments roads in Changsha City, China using measures provided by an Automated Number Plate Recognition (ANPR) system. Notably, they argue that variability occurs even when the traffic conditions are known, due to a plethora of factors such as interactions with pedestrians, cyclists, other vehicles, etc.. Their results demonstrate the stochastic nature of travel time. Finally, they show that while lognormal distributions provide good results to model the travel time distribution, the Johnson curve offers an even better way of modelling the travel time distributions.

Tran-The *et al.* (2017) explore the correlation between major weather events and traffic conditions in the region of Kansai, Japan. Using DBSCAN clustering, they manage to discover co-occurring spatiotemporal events linking degraded traffic conditions to torrential rains. While not directly linked to travel time, their method enables the detection and potentially the prediction of traffic disasters due to natural causes. This also highlights the extreme importance of weather when predicting traffic conditions and travel time.

Figure 6 Traffic jam centres observed in Beijing, China. Source: (Yinan et al. 2017)

Yinan *et al.* (2017) develop a method to detect traffic jams in urban environments and model their propagation (Figure 6). They experimentally test their technique with data from Beijing, China. They were able to demonstrate that certain jam centres dominate the creation of a traffic jam and that it propagates as a wave whose amplitude decreases with distance from the traffic jam centre. This supports methods performing estimations and predictions of traffic conditions of road segments based on neighbouring links. Their study illustrates how segment roads can greatly influence one another.

4. Predictions for other routes

The literature on routes other than terrestrial is limited. In this section, we quickly review some recent papers developing predictions for railways, waterways, and airways.

4.1. Predictions for railways

For railways, the main interest seems to be in predicting train delays. Nilsson and Henning (2018) compare a neural network and boosted trees (using AdaBoost) applied to data from the trains around Stockholm, Sweden and to weather data. Cerreto *et al.* (2018) use a common data analysis method, k-means clustering, to identify train delay patterns in a single high traffic railway north of Copenhagen, Denmark. An application to the Italian railway network can be found in (Oneto *et al.* 2016) Extreme Learning Machines (ELM), Random Forests (RF), and Kernel Regularised Least Squares (KRLS), in which they input historical train travel times and weather forecasts. Random forests, an ensemble method based on randomly generated decision trees, outperforms the other two tested methods.

4.2. Predictions for waterways

In waterways, the research seems to be interested in two main types of predictions: route predictions, and traffic predictions.

Route predictions consist in being able to determine the path of the various vessels roaming on a waterbody. Nguyen *et al.* (2018) propose an interesting approach similar to natural language processing using LSTMs to predict the next movements of a vessel based on its current path with an application to vessels in the Mediterranean Sea. A route discovery application using historical data of vessels roaming the Baltic Sea is proposed in (Fernandez Arguedas *et al.* 2018). They exploit DBSCAN (a clustering algorithm) to filter outliers and then cluster the main roads to reduce the dimensionality of their dataset. By doing so, they manage to provide with great compression and accuracy the main routes in the Baltic Sea (as illustrated in Figure 7).

Maritime traffic is predicted in (Xiao *et al.* 2017). To achieve this, they first use DBSCAN to reduce the dimensionality of their dataset, then use a kernel density estimate to predict the motion behaviour of different types of vessels. They demonstrate the validity of their approach with an application to Singapore waters.

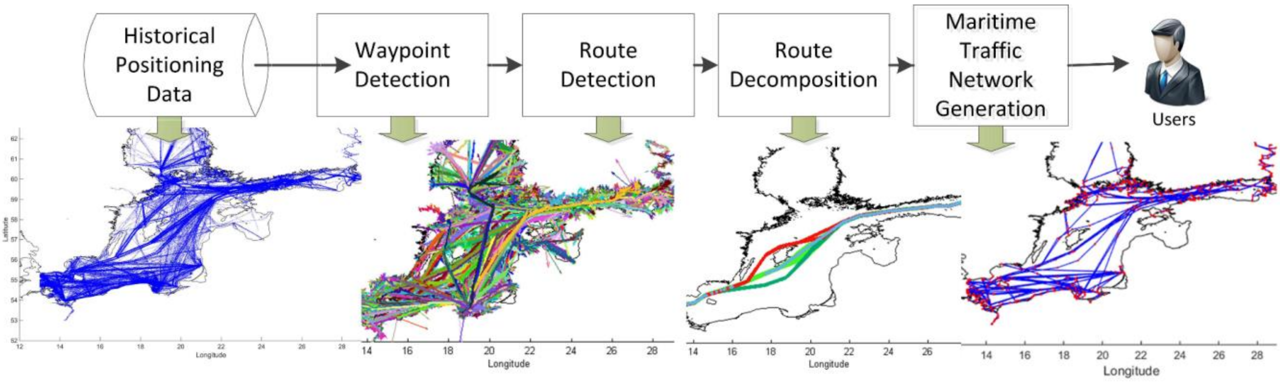

Figure 7 Maritime traffic discovery. Source: (Fernandez Arguedas et al. 2018)

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

4.3. Predictions for airways

As for railways, the main interest in predictions for airways concerns delays. Manna *et al.* (2017) use a gradient boosted decision tree on data provided by the U.S. Department of Transportation to predict flight departure and arrival delays. Their method achieves good accuracy on the five busiest U.S. airports. For more information on the prediction of delays in flights, the interested reader can refer to the review of the state-of-the-art on the subject by Sternberg *et al.* (2017).

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

5. Insights on current and future research

As presented in this document, AI and more particularly machine learning is becoming more and more popular in research and applications. A lot of work has been done on predictions of traffic and travel since the 1980s. In this section, we present some of the shortcomings of the current research and the future difficulties WP3 will have to overcome.

5.1. Freight-specific problems

The literature reviewed in this document rarely refers to goods transportation and freight. However, it is worth noting that any approach presented would be applicable to the problems we are facing in LOGISTAR. One possible opening for research would be to consider freight specific problems and data such as the impossibility for trucks or vessels to go through certain routes, the load of the trucks and vessels, etc. As shown in (Waury *et al.* 2018), recurrent information about the driver can also help increase the accuracy of the predictions. While impossible to apply for general traffic, this information could be available in a logistics network in which truck drivers are known.

5.2. Evaluation and comparison

Evaluating and comparing the quality of predictions made by various methods can be tedious and complicated. First of all, implementing multiple methods for the sake of comparison can be time-consuming. Second, the comparison generally requires data on top of training sets (hence usually reducing the size of the training sets). Finally, selecting evaluation measures can be complicated since each one usually reflects a specific aspect of the performance of a model. For instance, the Root-Mean-Square-Error (RMSE) gives a lot more weight to outliers compared to a classic mean that will reflect more the general goodness of fit without any accent put to error peaks. Hence, the performance measure to use to select the best model might be application specific.

In the papers reviewed in this document, we can see multiple issues with the evaluations:

- **Different papers have a tendency to use different datasets.** There does not seem to be a reference dataset used for comparison. Even PeMS, publicly available and often used, is rarely considered in its entirety and only a section of the covered roads is usually selected. This renders the comparison of different approaches extremely difficult, sometimes even leading to discrepancy since some researchers might rank the same methods differently (Mori *et al.* 2014).
- **Different papers might use different measures** or measures that are incomparable from one application to another (e.g. RMSE). Some of the measures found in the literature include the RMSE, the Normalised Root-Mean-Square Deviation (NRMSD), the Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE). All those measures are measurements of the quality of a single-value prediction and do not provide any way to measure the quality of a travel time distribution or interval for instance. (Shi, B. Chen, *et al.* 2017; Shi, B. Y. Chen, *et al.* 2017) proposed two new measures working for travel time prediction intervals: the probability outside the predicted time interval (POPI) and the probability outside of the observed time interval (POOI).
- **The computational resources (CPU and memory) are almost never considered**. This is problematic, especially when comparing methods. If a new approach offers predictions just slightly better at the cost of a lot more required CPU and memory, then its usefulness could be debated. In LOGISTAR, the resource efficiency of our system will be an important component of its evaluation and this will have to be seriously considered when comparing different methods. In

particular, offline predictions for long-term planning (i.e. multiple days ahead) could easily use prediction methods running over a few minutes to an hour. However, live predictions (as they could be used for live replanning for instance) would require a compromise between speed and accuracy much more oriented towards speed to reduce the reaction time of our applications.

- **Various time horizons for predictions might require various predictors.** The literature mostly focuses on short-term predictions (<1h). However, long-term and very long-term predictions (multiple days ahead) will be required for the scheduling of deliveries. The literature seems to support the idea that different predictors have different levels of reliability for different time horizons. Hence, selecting the best predictor might depend on the actual task to perform. Two main approaches can be used for long-term forecasts: iterative predictors and direct predictors (Shi and Yeung, 2018). Iterative predictors just use a single-step predictor iteratively (i.e. using the same predictor on its own predictions multiple times in a row) while direct predictors provide a direct prediction for a specific time horizon, requiring a different predictor for each time horizon. A mixed approach can also be used by combining direct predictors in the spirit of binary coding (Shi and Yeung, 2018). The two approaches with various predictors will need to be investigated thoroughly for LOGISTAR.

All of this opens the possibility to contribute to the construction of standards for the domain of ITS, on the used datasets as well as on the evaluation measures used. A standard method already exists to statistically compare classification algorithms (Demšar, 2006) and can probably be extended to regression algorithms. Evaluation and comparison of our approaches with the existing ones will have to be considered seriously within LOGISTAR.

5.3. Integrating more sources of data

Many of the papers presented in this document integrate exogenous data sources into their systems (e.g. weather forecasts, social network signals, or special event calendars) but this is only a recent trend and still seems to represent a minority of the work found in the literature. In most cases, such integration improves the quality of the predictions. However, once again, the resource cost against the gain in accuracy tradeoff will have to be considered since some of the external sources might be more accessible than others. Typically, weather forecasts are widely available with even sometimes public APIs for live querying. However, social network data might be more difficult to extract and to be made meaningful. Still, this offers interesting possible paths of research to investigate for LOGISTAR.

Also, one of the major current trends in AI is in data fusion (Faouzi *et al.* 2016; Lim & Lee 2011). Integrating more data sources, and more methods to then fuse them has been shown to provide better predictions in some of the papers presented in this document (Soua *et al.* 2016; Waury *et al.* 2018; Xu *et al.* 2017).

5.4. Ensemble methods and data fusion

The combination of the results of multiple methods can be seen as an ensemble method. Combination methods can be numerous and can consist of a simple mean or rely on more complex approaches such as Dempster-Shafer theory. The benefits of ensemble methods have been particularly well demonstrated with Random Forests (RF) (Fernández-Delgado *et al.*, 2014) and have even been formalised (Schapire *et al.* 1998). Ensemble methods should be considered in LOGISTAR to provide good predictions using a combination of poor but extremely computationally cheap predictors.

Each predictor may provide predictions of a time series in the form of a probability distribution. For instance, travel from A to B could be given as a discrete histogram as illustrated in Figure 8. As demonstrated in recent work (Shi *et al.,* 2017), Dempster-Shafer theory can be used to merge into a single probability distribution multiple ones provided by different prediction algorithms.

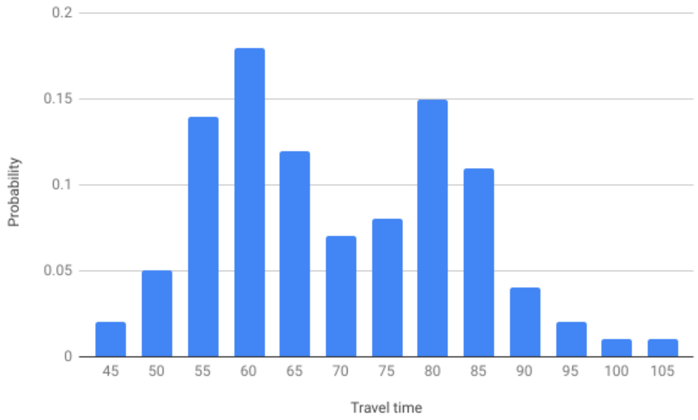

Figure 8 Example of a probability distribution of travel time with a granularity of 5min.

Belief functions on continuous values in the form of intervals are currently under investigation for the fusion of multiple predictors. This would enable the fusion of probability distributions provided as histograms, even with irregular, misaligned and conflicting histograms.

5.5. Scalability

One extremely important aspect of this literature review is the lack of very large-scale applications. While certain prototypes work on extremely large and complex cities such as Beijing, China (Jiang *et al.,* 2017), none of the papers found in the literature has an application integrating a network larger than a city. LOGISTAR will look at a transport network at the scale of a country or even the continent.

This will lead to major scalability issues when it comes to prediction and the amount of data that will need to be processed. First, computationally, the larger the network, the more resources will be necessary to provide network-wide predictions. Secondly, this means that integrating exogenous data will require a notion of locality. For instance, the weather might drastically differ in various parts of the network considered. Also, routes crossing multiple countries over multiple days will have to be considered in the predictions, which means various conditions (traffic, weather, events) will have to be integrated at a large scale.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

6. Conclusion

In this document, we have presented a quick overview of the State-of-the-Art in methods used in AI with a focus on predictions. Section 2 provided a summary of popular methods used in AI and in particular in the domain of machine learning. Section 3 provided a focus on predictions for intelligent transport systems, in particular for terrestrial routes. Section 4 provided a quick overview of predictions applied to waterways, airways, and railways. In section 5, we have provided insights on the limitations of current research and offered ideas for future research, in particular within LOGISTAR.

As explained, the large scale of LOGISTAR which will potentially cover multiple types of routes, continent-wide, will impose scalability constraints that will need to be tackled carefully when applying State-of-the-Art methods for predictions. The time horizon we will be predicting for, a few days ahead, is also rarely treated in the literature and promises to lead to interesting new research.

While scalability and temporality issues are fundamental, building a standard for both the research community and the industry could also be highly beneficial to the domain. The creation of large, highly detailed, and open datasets could bring more support to LOGISTAR by engaging with the existing research community. The datasets could consist of multiple types of routes (terrestrial, airways, railways, waterways), multiple types of information such as travel time, turnaround time, the quantities of products transported, weather information (current measurements and forecasts), special events (sports, festivals, etc.), and events detected within WP3 (incidents, accidents, unpredictable events).

Open datasets could also be integrated with automatic evaluation tools to help the community compare their various approaches. An automated tool could provide all sorts of evaluation metrics (e.g. accuracy or resources consumed) based on the type of prediction (single-valued vs probability distribution), the time horizon (short-term to very long-term), and the physical granularity (road segment vs paths). By integrating test tools to the dataset, any researcher could showcase their particular tradeoff in terms of the different metrics, enabling finding the best method depending on the application or service one wants to build.

List of abbreviations and acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
ANPR	Automated Number Plate Recognition
API	Application Programming Interface
ARIMA	AutoRegressive Integrated Moving Average
ARIMAX	AutoRegressive Integrated Moving Average with eXplainable variables
ARMA	AutoRegressive Moving Average
BN	Bayesian Network
CNN	Convoluted Neural Network
DNN	Deep Neural Network
ELM	Extreme Learning Machine
GPR	Gaussian Process Regression
GPS	Global Positioning System
GRU	Gated Recurrent Unit
IT	Information Technology
ITS	Intelligent Transport System
kNN	k-Nearest Neighbours
KRLS	Kernel Regularised Least Squares
LSTM	Long Short-Term Memory
ML	Machine Learning
NLP	Natural Language Processing
PCA	Principal Component Analysis
PMA	Periodical Moving Average
PPCA	Probabilistic Principal Component Analysis
RBM	Recurrent Boltzmann Machine
RF	Random Forest
RNN	Recurrent Neural Network
SVM	Support Vector Machine
SVR	Support Vector Regression
VAR	Vector AutoRegressive
VARMAX	Vector AutoRegressive Moving Average with eXplainable variables

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

References

Altman, N. S., 1992. *An introduction to kernel and nearest-neighbor nonparametric regression.* The American Statistician. 46 (3): 175–185. Available at: https://doi.org/10.1080%2F00031305.1992.10475879

Barros, J., Araujo, M. & Rossetti, R.J.F., 2015. *Short-term real-time traffic prediction methods: A survey*. In 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). Available at: http://dx.doi.org/10.1109/mtits.2015.7223248.

Bengio, Y., LeCun, Y., Hinton, G., 2015. *Deep Learning.* Nature. 521 (7553): 436–444. Available at: https://doi.org/10.1038%2Fnature14539

Bolshinsky, E. and Freidman, R., 2012. *Traffic Flow Forecast Survey*, Technical Report

Box, G., Jenkins, G., 1970. *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

Breiman, L. 2001. *Random Forests.* Machine Learning (2001) 45: 5. Available at: https://doi.org/10.1023/A:1010933404324

Burkov, A., 2019. *The Hundred-Page Machine Learning Book*, ISBN-13: 978-1999579500

Cai, P., Wang, Y., Lu, G., Chen, P., Ding, C., Sun, J., 2016. *A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting.* Transportation Research Part C: Emerging Technologies, 62, pp.21–34.

Cerreto, F., Nielsen, B.F., Nielsen, O.A., Harrod, S.S., 2018. *Application of Data Clustering to Railway Delay Pattern Recognition.* Journal of Advanced Transportation, 2018, pp.1–18.

Chen, C., Li, O., Tao, C., Barnett, A.J., Su, J., Rudin, C., 2018. *This Looks Like That: Deep Learning for Interpretable Image Recognition.* arXiv: 1806.10574

Chen, P., Tong, R., Lu, G., Wang, Y., 2018. *Exploring Travel Time Distribution and Variability Patterns Using Probe Vehicle Data: Case Study in Beijing.* Journal of Advanced Transportation, 2018, pp.1–13.

Cheng, S., Lu, F., 2018. *Short-term traffic forecasting: An adaptive ST-KNN model that considers spatial heterogeneity.* Computers, environment and urban systems, 71, pp.186–198.

Cortes, C., Vapnik, V. N., 1995. *Support-vector networks.* Machine Learning. 20 (3): pp. 273–297.

Cryer, J.D., Chan, K.-S., 2008. *Time Series Analysis With Applications in R (2nd ed.)*, Springer Texts in Statistics

Dempster A.P., 1969. *Upper and Lower Probability Inferences for Families of Hypotheses with Monotone Density Ratios*. Ann Math Stat. 1969;40: 953–969.

Demšar, J., 2006. *Statistical Comparisons of Classifiers over Multiple Data Sets*, In The Journal of Machine Learning Research, Vol. 7, pp 1-30

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Dilip, D.M., Lin, D.C., Jabari, S.E., 2018. *Learning Traffic Flow Dynamics using Random Fields*, arXiv:1806.08764

Du, S., Li, T., Gong, X., Yu, Z., Horng, S.-J., 2018. *A Hybrid Method for Traffic Flow Forecasting Using Multimodal Deep Learning*, arXiv:1803.02099

Edwards, C., 2015. *Growing pains for deep learning*. Communications of the ACM. 58 (7): 14–16. Available at: https://doi.org/10.1145%2F2771283

Faouzi, N.-E.E., El Faouzi, N.-E. & Klein, L.A., 2016. *Data Fusion for ITS: Techniques and Research Needs.* Transportation Research Procedia, 15, pp.495–512.

Fernandez, S., Graves, A.,Schmidhuber, J., 2007. *An application of recurrent neural networks to discriminative keyword spotting*, Proceedings of ICANN (2), pp. 220–229. Available at: https://mediatum.ub.tum.de/doc/1289941/file.pdf

Fernandez Arguedas, V., Pallota, G., Vespe, M., 2018. *Maritime Traffic Networks: From Historical Positioning Data to Unsupervised Maritime Traffic Monitoring.* IEEE Transactions on Intelligent Transportation Systems, 19(3), pp.722–732.

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D.; 2014. *Do we need hundreds of classifiers to solve real-world classification problems?*, In The Journal of Machine Learning Research, Vol. 15, pp 3133-3181

Freund, Y., Schapire, R.E., 1996. *Experiments with a new boosting algorithm.* In Proceedings of the Thirteenth International Conference on Machine Learning, pp. 148-156

Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning.* MIT Press. Available at: http://www.deeplearningbook.org

Heryati Syed Nor, S., Ismail, S., Wah Yap, B., 2019. *Personal bankruptcy prediction using decision tree model*. Journal of Economics, Finance and Administrative Science, Available at: https://doi.org/10.1108/JEFAS-08-2018-0076

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., Kingsbury, B., 2012. Deep Neural Networks for Acoustic Modeling in Speech Recognition --- The shared views of four research groups. IEEE Signal Processing Magazine. 29 (6): 82–97. Available at: https://doi.org/10.1109%2Fmsp.2012.2205597

Hochreiter, S., Schmidhuber, J, 1997. Long Short-Term Memory. Neural Computation. 9 (8): 1735–1780. Available at: https://doi.org/10.1162%2Fneco.1997.9.8.1735

Hochreiter, S., Bengio, Y., Frasconi, P., Schmidhuber, J., 2001. *Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.* In Kolen, John F.; Kremer, Stefan C. A Field Guide to Dynamical Recurrent Networks. John Wiley & Sons. ISBN 978-0-7803-5369-5.

Hou, X., Wang, Y., Hu, S., 2013. *Short-term Traffic Flow Forecasting based on Two-tier K-nearest Neighbor Algorithm*, Procedia - Social and Behavioral Sciences, Vol. 96, pp. 2529-2536, Available at: https://doi.org/10.1016/j.sbspro.2013.08.283.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Hyndman, R.J., & Athanasopoulos, G., 2018. *Forecasting: principles and practice (2nd ed.)*, OTexts: Melbourne, Australia. Available at: https://otexts.com/fpp2/.

Jabari, S. E., Freris, N., M., Dilip, D. M., 2018. *Sparse Travel Time Estimation from Streaming Data*, arXiv:1804.08130

James, G., Witten, D., Hastie, T., Tibshirani, R., 2017. *An Introduction to Statistical Learning With Applications in R (8th printing)*. Springer Texts in Statistics. Available at: http://www-bcf.usc.edu/~gareth/ISL/index.html

Jenelius, E., Koutsopoulos, H. N., 2018. *Urban Network Travel Time Prediction Based on a Probabilistic Principal Component Analysis Model of Probe Data*, in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 2, pp. 436-445, Feb. 2018. doi: 10.1109/TITS.2017.2703652

Jiang, Y., Kang, R., Li, D., Guo, S., Havlin, S., 2017. *Spatio-temporal propagation of traffic jams in urban traffic networks*, arXiv:1705.08269

Khemakhem, S., Boujelbene, Y., 2018. *Predicting credit risk on the basis of financial and non-financial variables and data mining*, Review of Accounting and Finance, Vol. 17 Issue: 3, pp. 316-340, Available at: https://doi.org/10.1108/RAF-07-2017-0143

Klunder, G., A, 2007. *Long-term travel time prediction algorithm using historical data*, In 14th World Congress on Intelligent Transport Systems, ITS 2007; Beijing; China; October 2007, 1191-1198

Krishnan, R., Polak, J., 2008. *Short-term travel time prediction: An overview of methods and recurring themes*, In Transportation Planning and Implementation Methodologies for Developing Countries Conference (TPMDC 2008)

Leverger, C., Lemaire, V., Malinowski, S., Guyet, T., Rozé, L., 2018. *Day-ahead time series forecasting: application to capacity planning*. arXiv:1811.02215

Li, O., Liu, H., Chen, C., Rudin, C., 2018. *Deep Learning for Case-Based Reasoning through Prototypes: A Neural Network that Explains Its Predictions.* The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)

Li, S., Shen, Z., Xiong, G., 2012. *A k-nearest neighbor locally weighted regression method for short-term traffic flow forecasting*, 2012 15th International IEEE Conference on Intelligent Transportation Systems, Anchorage, AK, pp. 1596-1601. Available at: https://ieeexplore.ieee.org/document/6338648

Lighthill, M.J., Whitham, G.B., 1955a. *On Kinematic Waves. I. Flood Movement in Long Rivers.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 229(1178), pp.281–316.

Lighthill, M.J., Whitham, G.B., 1955b. *On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 229(1178), pp.317–345.

Lim, S., Lee, C., 2011. *Data fusion algorithm improves travel time predictions.* IET Intelligent Transport Systems, 5(4), pp.302–309.

Lin, H.-E., Zito, R., Taylor, M. A. P, 2005. *A review of travel-time prediction in transport and logistics,* In Proceedings of the Eastern Asia Society for Transportation Studies, Vol. 5, pp 1433-1448

Liu, Y., Wang, Y., Yang, X., Zhang, L., 2017. *Short-term travel time prediction by deep learning: A comparison of different LSTM-DNN models.* In 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). Available at: http://dx.doi.org/10.1109/itsc.2017.8317886.

Manna, S., Biswas, S., Kundu, R., Rakshit, S., Gupta, P., Barman, S., 2017. *A statistical approach to predict flight delay using gradient boosted decision tree*. In 2017 International Conference on Computational Intelligence in Data Science (ICCIDS). Available at: http://dx.doi.org/10.1109/iccids.2017.8272656.

Mori, U., 2014. *A review of travel time estimation and forecasting for Advanced Traveller Information Systems*. Transportmetrica A: Transport Science, 11(2), pp.119–157.

Nguyen, D.-D., Le Van, C. Ali, M.I., 2018. *Vessel Destination and Arrival Time Prediction with Sequence-to-Sequence Models over Spatial Grid.* In Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems - DEBS ’18. Available at: http://dx.doi.org/10.1145/3210284.3220507.

Nilsson, R., Henning, K., 2018. *Predictions of train delays using machine learning*, Technical report, Examensarbete inom Datateknik, KTH, Sverige

Oh, S., Byon, Y.-J., Jang, K., Yeo, H., 2015. *Short-term Travel-time Prediction on Highway: A Review of the Data-driven Approach*. Transport Reviews, 35(1), pp.4–32.

Oh, S., Byon, Y.-J., Jang, K., Yeo, H., 2017. *Short-term travel-time prediction on highway: A review on model-based approach*. KSCE Journal of Civil Engineering, 22(1), pp.298–310.

Oneto, L. Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., Anguita, D., 2016. *Advanced Analytics for Train Delay Prediction Systems by Including Exogenous Weather Data*. In 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA). Available at: http://dx.doi.org/10.1109/dsaa.2016.57.

Quinlan, R., 1983. *Learning efficient classification procedures*, Machine Learning: an artificial intelligence approach, Michalski, Carbonell & Mitchell (eds.), Morgan Kaufmann,, pp. 463-482. Available at: https://doi.org/10.1007%2F978-3-662-12405-5_15

Rodriguez-Deniz, H., Jenelius, E. & Villani, M., 2017. *Urban network travel time prediction via online multi-output Gaussian process regression.* In 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). Available at: http://dx.doi.org/10.1109/itsc.2017.8317796.

Rosenblatt, F., 1958. *The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain.* Psychological Review. 65 (6): 386–408. Available at: https://doi.org/10.1037%2Fh0042519

Russell, S. J., Norvig, P., 2009. *Artificial Intelligence: A Modern Approach (3rd ed.)*. Upper Saddle River, New Jersey: Prentice Hall. ISBN 978-0-13-604259-4.

Samuel, A. L., 1959. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development. 3 (3): 210–229. Available at: https://doi.org/10.1147%2Frd.33.0210

Shafer G., 1976. *A Mathematical Theory of Evidence.* Princeton University Press.

Schapire, R.E., Freund, Y., Barlett, P., Sun Lee, W., 1998. *Boosting the margin: a new explanation for the effectiveness of voting methods.* Annals of statistics, 26(5), pp.1651–1686.

Seal, H.L, 1967. *Studies in the History of Probability and Statistics. XV The historical development of the Gauss linear model.* Biometrika, Volume 54, Issue 1-2, June 1967, pp. 1–24, Available at: https://doi.org/10.1093/biomet/54.1-2.1

Shi, C., Chen, B.Y., Lam, W.H.K., Li, Q., 2017. *Heterogeneous Data Fusion Method to Estimate Travel Time Distributions in Congested Road Networks.* Sensors, 17(12). Available at: http://dx.doi.org/10.3390/s17122822.

Shi, C., Chen, B. & Li, Q., 2017. *Estimation of Travel Time Distributions in Urban Road Networks Using Low-Frequency Floating Car Data.* ISPRS International Journal of Geo-Information, 6(8), p.253.

Shi, X., Yeung, D.-Y., 2018. *Machine Learning for Spatiotemporal Sequence Forecasting: A Survey*. arXiv:1808.06865

Soua, R., Koesdwiady, A. & Karray, F., 2016. *Big-data-generated traffic flow prediction using deep learning and Dempster-Shafer theory.* In 2016 International Joint Conference on Neural Networks (IJCNN). Available at: http://dx.doi.org/10.1109/ijcnn.2016.7727607.

Sternberg, A., Soares, J., Carvalho, D., Ogasawara, E., 2017. *A Review on Flight Delay Prediction*, arXiv:1703.06118

Stock, J.H., Watson, M.W., 2003. *Introduction to Econometrics.* Addison Wesley.

Tang, K., Chen, S. & Liu, Z., 2018. *Citywide Spatial-Temporal Travel Time Estimation Using Big and Sparse Trajectories.* IEEE Transactions on Intelligent Transportation Systems, pp.1–12.

Taylor, M. A. P., Zito, R., Lin, H.-E. 2005. *A review of travel-time prediction in transport and logistics*, EASTS

Tran-The, H., Hung Tran-The & Zettsu, K., 2017. *Discovering co-occurrence patterns of heterogeneous events from unevenly-distributed spatiotemporal data.* In 2017 IEEE International Conference on Big Data (Big Data). Available at: http://dx.doi.org/10.1109/bigdata.2017.8258023.

Utgoff, P. E. 1989. *Incremental induction of decision trees.* Machine learning, 4(2), pp. 161-186. Available at: https://doi.org/10.1023%2FA%3A1022699900025

Vlahogianni, E.I., Golias, J.C. & Karlaftis, M.G., 2004. *Short-term traffic forecasting: Overview of objectives and methods.* Transport Reviews, 24(5), pp.533–557.

Vu, L.H., Passow, B.N., Paluszczyszyn, D., Deka, L., Goodyer, E., 2017. *Neighbouring link travel time inference method using artificial neural network.* In 2017 IEEE Symposium Series on Computational Intelligence (SSCI). Available at: http://dx.doi.org/10.1109/ssci.2017.8285221.

Wang, D., Wu, Y. & Xiao, Z., 2017. *A Gaussian process regression method for urban road travel time prediction.* In 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). Available at: http://dx.doi.org/10.1109/fskd.2017.8393394.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Waury, R., Jensen, C.S. & Torp, K., 2018. *Adaptive Travel-Time Estimation: A Case for Custom Predicate Selection.* In 2018 19th IEEE International Conference on Mobile Data Management (MDM). Available at: http://dx.doi.org/10.1109/mdm.2018.00026.

Wen, R., Yan, W. & Zhang, A.N., 2017. *Adaptive spatio-temporal mining for route planning and travel time estimation.* In 2017 IEEE International Conference on Big Data (Big Data). Available at: http://dx.doi.org/10.1109/bigdata.2017.8258311.

Xiao, Z., Ponnambalam, L., Fu, X., Zhang, W., 2017. *Maritime Traffic Probabilistic Forecasting Based on Vessels' Waterway Patterns and Motion Behaviors.* IEEE Transactions on Intelligent Transportation Systems, 18(11), pp.3122–3134.

Xie, J. & Choi, Y.-K., 2017. *Hybrid traffic prediction scheme for intelligent transportation systems based on historical and real-time data*. International Journal of Distributed Sensor Networks, 13(11), p.155014771774500.

Xu, T., Li, X. & Claramunt, C., 2017. *Trip-oriented travel time prediction (TOTTP) with historical vehicle trajectories*. Frontiers of earth science, 12(2), pp.253–263.

Yuan, Z. & Tu, C., 2017. *Short-term traffic flow forecasting based on feature selection with mutual information*. Available at: http://dx.doi.org/10.1063/1.4982544.

Zhang, Q., Jian, D., Xu, R., Dai, W., Liu, Y., 2017. *Integrating heterogeneous data sources for traffic flow prediction through extreme learning machine.* In 2017 IEEE International Conference on Big Data (Big Data). Available at: http://dx.doi.org/10.1109/bigdata.2017.8258443.

Zhang, Z., Chen, P., Wang, Y., Yu, G., 2017. *A hybrid deep learning approach for urban expressway travel time prediction considering spatial-temporal features.* In 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). Available at: http://dx.doi.org/10.1109/itsc.2017.8317889.

Zheng, F., Liu, X., van Zuylen, H., Li, J., Lu, C., 2017. *Travel Time Reliability for Urban Networks: Modelling and Empirics.* Journal of Advanced Transportation, pp.1–13.

Web links

Andrej Karpathy on RNNS: http://karpathy.github.io/2015/05/21/rnn-effectiveness/

CUDA-X AI: https://www.nvidia.com/en-us/technologies/cuda-x/

INRIX: http://inrix.com/

PeMS dataset: http://www.dot.ca.gov/trafficops/mpr/source.html

TensorFlow: https://www.tensorflow.org/

US Military and AI: https://www.technologyreview.com/s/603795/the-us-military-wants-its-autonomous-machines-to-explain-themselves/

Wikipedia - Artificial Neural Network: https://en.wikipedia.org/wiki/Artificial_neural_network

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Wikipedia - k-Nearest Neighbors Algorithm: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

Wikipedia - Recurrent Neural Networks: https://en.wikipedia.org/wiki/Recurrent_neural_network