Online Predictive Model for Taxi Services

Luís Moreira-Matias[1,2], João Gama[2,3], Michel Ferreira[4,5], João Mendes-Moreira[1,2], and Luís Damas[5]

[1] Departamento de Engenharia Informática, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, s/n 4200-465 Porto – Portugal
[2] LIAAD-INESC TEC. Rua de Ceuta, 118, 6º, 4050-190 Porto – Portugal
[3] Faculdade de Economia, Universidade do Porto Rua Dr. Roberto Frias, s/n 4200-465 Porto – Portugal
[4] Instituto de Telecomunicações, Departamento de Ciência de Computadores, Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal
[5] Geolink, Pct. Adelino Amaro Costa, 772 4º Dto, 4050-012 Porto – Portugal
{luis.matias,jmoreira}@fe.up.pt, jgama@fep.up.pt, michel@dcc.fc.up.pt, luis@geolink.pt

Abstract. In recent years, both companies and researchers have been exploring intelligent data analysis to increase the profitability of the taxi industry. Intelligent systems for online taxi dispatching and time saving route finding have been built to do so. In this paper, we propose a novel methodology to produce online predictions regarding the spatial distribution of passenger demand throughout taxi stand networks. We have done so by assembling two well-known time series short-term forecast models: the time-varying Poisson models and ARIMA models. Our tests were performed using data gathered over a period of 6 months and collected from 63 taxi stands within the city of Porto, Portugal. Our results demonstrate that this model is a true major contribution to the driver mobility intelligence: 78% of the 253745 demanded taxi services were correctly forecasted in a 30 minutes horizon.

Keywords: ARIMA, Time-Varying Poisson Model, Taxi Services, Time Series, Data Streams.

1 Introduction

In the last decade, real-time vehicle location systems have attracted the attention of both companies and researchers for a new kind of rich spatio-temporal information. Taxi networks have largely been affected by this phenomenon, as such networks produce multiple data streams that can be explored using intelligent data analysis. Online systems for efficient taxi dispatching [1] and time-saving route finding [2] (among others) have already been developed to improve taxi service reliability.

The taxi driver mobility intelligence is a crucial feature to maximize both profit and reliability. There are few works which focus on this topic and use real GPS data. The work presented in [3] uses spatial-based clustering applied to the GPS historical data of a taxi network running in a large urban zone. Their goal consisted on

J. Hollmén, F. Klawonn, and A. Tucker (Eds.): IDA 2012, LNCS 7619, pp. 230–240, 2012.

discovering passenger demand patterns over a period of time, by building a departure-destination-time cubic matrix. Recently, an innovative study has been presented in [4] to validate the triplet Time-Location-Strategy as the key feature to build a good passenger finding strategy. Here the L1-Norm-SVM was used as a feature selection tool to discover both efficient and inefficient passenger finding strategies based on a large-scale GPS dataset, from a taxi network running in a large city in China. An empirical study on the impact of the selected features was conducted and its conclusions were validated by the feature selection tool.

Both works represent recent studies about the selection of the best route (over a period of time) to maximize the number of passengers picked-up by each driver. However, there are three important issues that are not convincingly handled by these authors: (1) conditions which are not common to every urban area are assumed (e.g.: drivers are able to choose their routes freely and without any regulation); (2) their insights consist of offline patterns, which do not fulfill the ubiquitous potential of the data to provide decision support information in real-time; (3) the rising cost of fuel is clearly disallowing the economic viability of online or offline cruising strategies to get passengers.

In our work, we focus on the choice problem concerning which is the best taxi stand to go to after a passenger drop-off at a given location and time. One of the most important factors for this is the passenger demand at each taxi stand over a time span. In this paper, we present a streaming model to predict the number of services of a taxi network over a space span (taxi stand), for a short-time horizon of P-minutes. Based on both historical and real time GPS location and service data (passenger drop-off and pick-up) transmitted by the telematics installed in each vehicle, time series histograms are built for each stand containing the number of services with an aggregation of P-minutes. The predictive model was developed by adapting well-known time series forecasting techniques, such as the time varying Poisson model [5] and ARIMA (Autoregressive Integrated Moving Average) [6] to our problem. Our goal is to predict at an instant t how many services will be demanded during a period $[t, t+P]$ at each existent taxi stand, reusing the real service count on $[t, t+P]$ which has been extracted from the data to do the same for the instant $t+P$ and so on (i.e. the framework runs continuously in a stream). To the best of our knowledge, such approach has no parallel in the literature.

Our model was applied to data collected from a large-sized taxi network which contains a total of 63 taxi stands and has 441 vehicles running in the city of Porto, Portugal. In this city, the existing regulations force taxi drivers to pick a route to one of the existing stands after a passenger drop-off. Our study just uses the services obtained directly at the stands or which were automatically dispatched to the parked vehicles as input/output. This was done because the passenger demand at each taxi stand is the main feature to aid the taxi drivers' decision (since it represents 76% of the total number of services).

Our experiments use the real-time data arriving in a stream to produce predictions about the expected number of services demanded in each stand. Firstly, 20 weeks' worth of data was acquired to build our historic. Secondly, the demand for the next 8 weeks was predicted, by updating the historic time series over time. The results obtained were promising: our model predicted more than 78% of the services that actually emerged using an average computational time (i.e. average time per prediction) of 99.77 seconds.

The remainder of the paper is structured as follows. Section 2 formally describes our model. Section 3 describes how the dataset used was acquired and preprocessed, as well as some statistics about it. Section 4 outlines the testing of the methodology in a concrete scenario. Firstly, we introduce the evaluation metrics of our model, along with the experimental setup. We then present the obtained results. Section 5 concludes the paper and describes future work we intend to carry out.

2 The Model

Consider $\boldsymbol{S} = \{\boldsymbol{s_1}, \boldsymbol{s_2}, \ldots, \boldsymbol{s_N}\}$ to be the set of $\boldsymbol{N}$ taxi stands of interest and $\boldsymbol{D} = \{\boldsymbol{d_1}, \boldsymbol{d_2}, \ldots, \boldsymbol{d_j}\}$ to be a set of $\boldsymbol{j}$ possible passenger destinations. Our problem consists of choosing the best taxi stand at the instant $\boldsymbol{t}$ according to our forecast about passenger demand distribution over the time stands for the period $[\boldsymbol{t}, \boldsymbol{t+P}]$, as is illustrated in Fig. 1.

Consider $\boldsymbol{X_k} = \{\boldsymbol{X_{k,0}}, \boldsymbol{X_{k,1}}, \ldots, \boldsymbol{X_{k,t}}\}$ **to** be a time series for the number of demanded services at a taxi stand $\boldsymbol{k}$. Our goal is to build a model to determine the set of service count $\boldsymbol{X_{k,t+1}}$ for the instant $\boldsymbol{t+1}$ and for all taxi stands $\boldsymbol{k \epsilon \{1, N\}}$**.** To do so, we propose three distinct short-term prediction models and a well-known data stream ensemble framework to use them all. We formally describe those models along this section.

2.1 Time Varying Poisson Model

The following section presents a model firstly proposed in [5]. The demand of taxi services exhibit, like other transportation means [7], a periodicity in time on a daily basis that reflects the patterns of the underlying human activity, making the data appear non-homogeneous [5]. Fig. 2 illustrates a one month taxi service analysis extracted from our dataset that illustrates this periodicity (the dataset is described in detail in Section 3).

Consider the probability to have *n* taxi assignments in a determined time period – ***P(n)*** - following a *Poisson distribution*. We can define it using the following equation

$$\mathrm{P}(\mathrm{n};\lambda) = \frac{\mathrm{e}^{-\lambda}\lambda^{\mathrm{n}}}{\mathrm{n!}}, \tag{1}$$

where λ represents the rate (averaged number of the demand on taxi services) in a fixed time interval. However, in this specific problem, the rate λ is not constant but time-variant. As a result, we adapt it as a function of time, i.e. $\lambda(t)$, transforming the Poisson distribution into a nonhomogeneous one. Consider $\lambda(t)$ to be defined as follows

$$\lambda(\mathrm{t}) = \lambda_0 \delta_{\mathrm{d(t)}} \eta_{\mathrm{d(t),h(t)}}, \tag{2}$$

where $d(t)$ represents the weekday {1=Sunday, 2=Monday,…}; $h(t)$ the period in which time *t* falls (e.g. the time 00:31 is contained in period 2 if we consider 30-minute periods).

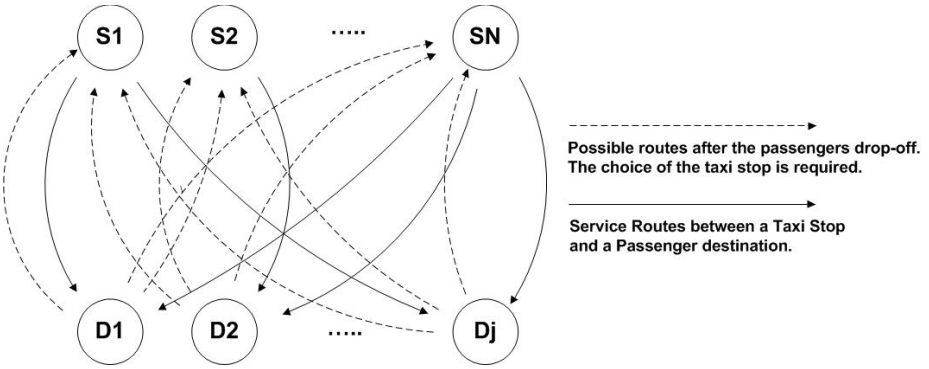

Fig. 1. A schema to illustrate our problem

It requires the validity of both equations

$$\sum_{i=1}^{7} \delta_i = 7, \tag{3}$$

$$\sum_{i=1}^{D} \eta_{d,i} = D \quad \forall d, \tag{4}$$

where *D* is the number of time intervals in a day. To ease the interpretation of these equations, we can define the remaining symbols as follows:

- λ_0 is the average (i.e. expected) rate of the Poisson process over a full week;
- δ_i is the relative change for the day ***i*** (Saturdays have lower day rates than Tuesdays);
- $\eta_{j,i}$ is the relative change for the period ***i*** on the day ***j*** (the peak hours);
- $\lambda(t)$ is a discrete function representing the expected demand of taxi services distribution over a period of time for a taxi stand of interest ***k.***

2.2 Weighted Time Varying Poisson Model

The model previously presented can be viewed as a time-dependent average. However, it is not guaranteed that every taxi stand will have a highly regular passenger demands: in fact, the demand at many stands can often be *seasonal*.

To face this specific issue, we propose a weighted average model based on the one already presented: our goal is to increase the relevance of the demand pattern observed in the previous week, by comparing it with the patterns observed several weeks ago (e.g. what happened on the previous Tuesday is more relevant than what happened two or three Tuesdays ago). The weight set ω is calculated using a well-known time series approach to these kind of problems: the Exponential Smoothing [8]. We can define ω as follows

$$\omega = \alpha * \{1, (1-\alpha), (1-\alpha)^2, \dots, (1-\alpha)^{\gamma-1}\}, \tag{5}$$

where γ is the number of historical periods considered in the initial average, α is the smoothing factor (i.e. a user-defined parameter) and $0 < \alpha < 1$.

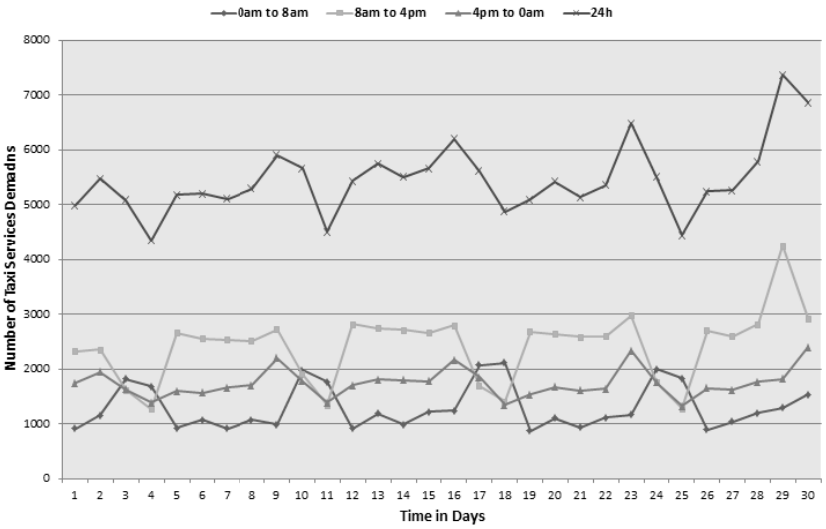

Fig. 2. One month data analysis (total and per driver shift)

2.3 Autoregressive Integrated Moving Average Model

The two previous models assume the existence of a regular (seasonal or not) periodicity in the taxi service passenger demand (i.e. the demand at one taxi stand on a regular Tuesday during a certain time period will be highly similar to the demand verified during the same time period on other Tuesdays).

However, the demand can present distinct periodicities for different stands. The ubiquitous characteristics of this network force us to rapidly decide if and how the model is changing. The Autoregressive Integrated Moving Average Model (ARIMA) [9] is a well-known methodology to both model and forecast univariate time series data such as traffic flow data [10], electricity pricing [11] and other short term prediction problems like our own. The ARIMA main advantage when compared to other algorithms is its versatility to represent very different types of time series: the autoregressive (AR) ones, the moving average ones (MA) and a combination of those two (ARMA). A brief presentation of one of the simplest ARIMA models (for non-seasonal stationary time series) is enunciated below following the existing description in [12] (however, our framework can also detect both seasonal and non-stationary ones). For a more detailed discussion, the reader should consult a comprehensive time series forecasting text such as Chapters 4 and 5 in [13].

In an autoregressive integrated moving average model, the future value of a variable is assumed to be a linear function of several past observations and random errors. The underlying process that generates the time series (taxi service over time for a given stand $\boldsymbol{k}$) can be formulate as

$$\begin{aligned} R_{k,t} = \theta_0 + \phi_1 X_{k,t-1} + \phi_2 X_{k,t-2} + \cdots + \phi_p X_{k,t-p} \\ + \varepsilon_{k,t} - \theta_1 X_{k,t-1} - \theta_2 X_{k,t-2} - \cdots - \theta_q X_{k,t-q} \end{aligned} \quad (6)$$

where $R_{k,t}$ and $\varepsilon_{k,t}$ are the actual value and random error at time period t, respectively; $\phi_l (l = 1,2, \dots, p)$ and $\theta_m (m = 0,1,2, \dots, q)$ are the model parameters/weights while p and q are positive integers often referred to as the order of the model. Both order and weights can be inferred from the historical time series using both the autocorrelation and partial autocorrelation functions as introduced by Box and Jenkins in [9]. They are useful to detect if the signal is periodic and, most important, which are the frequencies of these periodicities.

2.4 Sliding Window Ensemble Framework

In the last decade, regression and classification tasks on stream data have attracted the community attention due to its special characteristics: the traditional algorithms had computational efforts which were not compatible with the time available to make a decision. Fast and adaptive ensembles of these were also built. One of the most popular models is the weighted ensemble [14]. The model proposed below is also based on this one.

Consider $M = \{M_1, M_2, \dots, M_z\}$ to be a set of z models of interest to model a given time series and $M_t = \{M_{1t}, M_{2t}, \dots, M_{nt}\}$ to be the set of forecasted values to the next period on the interval t by those models. The ensemble forecast E_t is obtained as

$$E_t = \sum_{i=1}^{z} \frac{M_{it}}{\beta}, \qquad \beta = \sum_{i=1}^{z} \rho_{iH} \tag{7}$$

where ρ_{iH} is the forecasting error obtained for the model M_1 in the periods contained within the time window/period $[t - H, t]$ (H is a user-defined parameter to define the window size). As the information is arriving in a continuous manner for the next periods $\{t, t+1, t+2, \dots\}$ the window will also *slide* to determine how the models are performing in the *last H periods*. To calculate such error, we used a well-known time series forecasting error metric: the Symmetric Mean Percentage Error (*sMAPE*) [15].

3 Data Acquisition and Preprocessing

In this work, we studied the taxi driver mobility intelligence of one company operating in the city of Porto, Portugal. This city is the center of a medium size urban area (1.3 million inhabitants) where the passenger demand is lower than the number of running vacant taxis, provoking a huge competition between both companies and drivers. The existing regulations force the drivers to not run *randomly* in search of passengers but to choose a specific taxi stand out of the 63 existing ones in the city – see Fig. 3 to observe its spatial distribution - to go park and wait for the next service immediately after the last passenger drop-off. There are three main ways to pick-up a passenger: (I) a passenger goes to a taxi stand and picks-up a taxi – the regulations also force the passengers to pick-up the first taxi in the line (First In, First Out); (II) a passenger calls the taxi network central and demands a taxi for a specific location/time –parked taxis have priority over running vacant ones in the central taxi dispatch system; (III) a passenger picks a vacant taxi while it is going to a taxi stand, on any street.

Fig. 3. Taxi Stand spatial distribution over the city of Porto, Portugal

In this section, we describe the studied company and the data acquisition process and the preprocessing applied to it.

3.1 Data Acquisition

The data was continuously acquired using the telematics installed in each one of the 441 running vehicles (GPS and wireless communication) of the studied company. These vehicles usually run in one out of three 8h shifts: midnight to 8am, 8am-4pm and 4pm to midnight. Each data chunk arrives with the following six attributes: (1)

TYPE –relative to the type of event reported and has four possible values: busy – the driver picked-up a passenger; assign – the dispatch central assigned a service previously demanded; free – the driver dropped-off a passenger and park – the driver parked at a taxi stand. The attribute (2) STOP is an integer with the ID of the related taxi stand. The attribute (3) TIMESTAMP is the date/time in seconds of the event and the attribute (4) TAXI is the driver code; the attributes (5) and (6) refer to the LATITUDE and the LONGITUDE corresponding to the acquired GPS position.

This data was acquired over a non-stop period of 28 weeks. Our study just uses as input/output the services obtained directly at the stands or those automatically dispatched to the parked vehicles (more details in the section below). This was done as the passenger demand at each taxi stand is the main feature to aid the taxi drivers' decision.

3.2 Preprocessing and Data Analysis

As preprocessing, a time series of taxi demand services aggregated for a period of P-minutes was continuously built. There are three types of accounted events: (1) the *busy* directly set at a taxi stand; (2) the *assign* directly set to a taxi parked in a taxi stand and (3) the *busy* set while a vacant taxi is cruising. Type 1 events are accounted directly as type 2 events. However, for each type 2 event, the system receives a *busy* event a few minutes later – as soon as the driver effectively picked-up the passenger – this is ignored by our system. Type 3 events are ignored unless they occur in a radius of W meters from a taxi stand (where W is a user defined parameter). If it does, it is considered as being a type 1 event related with the nearest taxi stand according the

defined criteria. This was done because many regulations disallow to picking-up passengers in a pre-defined radius of a stop (in Porto a 50m radius is defined).

Table 1 details the number of taxi services demanded from any taxi stand per daily shift and day type. Table 2 has information about all services (independently of the pick-up place) per taxi and duration. The *service* column in Table 2 represents the number of services picked-up by taxi drivers, while the second one is related to the time distance of each service done.

Additionally, it could be stated that the central service assignment is 24% of the total service (*versus* the 76% of the one demanded directly in the street) while 77% of the service is demanded directly to taxis parked at a taxi stand (and 23% is picked-up while they are cruising). The average waiting time (to pick-up passengers) of a taxi parked at a taxi stand is 42 minutes while the average time distance for a service is only 11 minutes and 12 seconds.

The data presented in the tables highlight this, despite the regularity exhibited in the service (especially on the weekends), there are huge mobility intelligence discrepancies between the drivers (i.e. a large variance in both number and time distance of the services).

4 Experimental Results

In this section, we firstly describe the experimental setup developed to test our model on the available data. Secondly, we present and discuss the results achieved.

4.1 Experimental Setup

Our model produces an online forecast for the demand of taxi services at all taxi stands at each period of P-minutes. The scripts used were developed using the R statistical software [16].

The data was continuously acquired and processed through messages sent over a socket. The pre-defined functions used and the values set for the models parameters are detailed along this section.

An aggregation period of 30 minutes was set (i.e. a new forecast is produced each 30 minutes; P=30) and a radius of 100 meters (W = 100 > 50 defined by the existing regulations). This aggregation was set according the average waiting time at a taxi stand, i.e. the forecast horizon lower than 42 minutes.

Data was acquired over a period of 28 weeks. 20 weeks' worth of data was initially stored just to train the model. Then the number of services for each stand was continuously forecasted every 30 minutes over a total of 8 weeks. This framework continuously runs on a single computer using just one core (i.e. without any parallelization).

The ARIMA model (p,d,q values and seasonality) was firstly set (and updated each 24h exactly at 3am) by learning/detecting the underlying model running on the historical time series curve for each considered taxi stand. This was done by using an automatic time series function in the [forecast] R package [17] - `auto.arima` - with the default parameters. The weights/parameters are specifically fit for each period using the `arima` function from the built-in R package [stats].

Table 1. Taxi Services Volume (per daytype/Shift)

	Total Services Occurred	Averaged Service Demand per Shift		
		0am to 8am	8am to 4pm	4pm to 0am
Workdays	781398	1263	2227	1719
Weekends	289364	1048	2085	1534
Total	1070762	**2311**	**4312**	**3253**

Table 2. Taxi Services Volume (per Taxi/Duration)

	Services	Time Running Busy (minutes)
Máx.	5174	50605
Min.	38	384
Mean Total	1811	23424
Std. Dev. Total	816	9996

The time-varying Poisson averaged models (both weighted and non-weighted) were updated every 24 hours. A sliding window of 4 hours (H=8) was considered in the ensemble. The error of each model was measured using the metric also proposed to weight each model in the ensemble – sMAPE. Distinct results for two distinct values (0.4 and 0.5) of the parameter alpha (α in the weighted average) are presented below.

4.2 Results

The main results are presented in Table 3. A percentage is presented for each model corresponding to a similarity measure between the real service time series and the predicted one (i.e.: like a distance between the two time series calculated using the *sMAPE*). The results are firstly presented per shift and then globally. The values presented below are calculated through an average weight of the error obtained through each one of the time series (i.e. the forecast error of the demand for taxi services at each one of the 63 taxi stands). Each error rate was weighted according to the number of services demanded at the corresponding taxi stand along the whole test period.

Table 3. Models Success Rate per Shift

MODEL	PERIODS							
	alpha (α) = 0.4				*alpha (α) = 0.5*			
	00->08	08->16	16->00	24h	00->08	08->16	16->00	24h
Poisson Mean	79,03%	76,25%	76,79%	77,34%	79,03%	76,25%	76,79%	77,34%
W. Poisson Mean	78,15%	74,51%	75,25%	75,53%	77,05%	73,20%	73,99%	74,29%
ARIMA	79,01%	73,37%	76,79%	75,72%	79,01%	73,37%	76,79%	75,72%
Ensemble	**81,06%**	**76,87%**	**78,64%**	**78,36%**	**81,01%**	**76,73%**	**78,58%**	**78,26%**
Nr. Of Events	54.276	114.344	85.125	253.745	54.276	114.344	85.125	253.745

Each model returns a time series which has a similarity measure above 74% (the Weighted Poisson Mean and the Ensemble are only affected by the changes of the *alpha* (α) parameter) for the real one. The sliding window ensemble is always the best model for every shift and period considered, with a similarity measure greater than 78% (197921 of the 253745 total taxi services were correctly forecasted in both time and space using an aggregation of 30-minutes). Our framework also had a satisfactory performance in another important aspect: the computational time. It was implemented using an iterative process (i.e. a loop) forecasting the next value for each existing time series (for a total of 63). Such process took, in average, 99.77 seconds of processing time. The ARIMA model update (it is done every 24 hours) on average lasted 48.12 seconds.

5 Conclusions and Future Work

Sensor networks are providing a growing number of challenges to researchers along with an explosive number of opportunities for companies. In the last decade, the taxi industry has already been exploring intelligent data analysis on a daily basis (efficient taxi dispatching, time saving route finding, among others). This paper presents a novel application work where well-known time series forecasting models are adapted to predict the spatial distribution of the passenger demand for taxi services in a short term horizon (30 minutes periods). To do so, two distinct models have been assembled - the Time-Varying Poisson Model [5] and the ARIMA [9] one – using a Sliding Window Weighted Ensemble model [14] (i.e. it uses the error of each model through a pre-determined time window as their weight in the mean calculation). These models were chosen due to their well-known efficiency in short-term point forecast problems like our own. At our best knowledge, such an online predictive model for the passenger demand is a true novelty in this specific industry.

We tested it using the data received from a company network in the city of Porto, Portugal. The test ran continuously for the duration of 28 weeks over the existing 63 taxi stands.

The results are promising: the model correctly predicted 78% of the 253745 received services. Our iterative implementation took, an average, 99.77 seconds to produce a prediction (1.58 seconds for each taxi stand) for the next period. These numbers demonstrate that this model can be a major contribution to this industry, improving and aiding the drivers mobility intelligence in real time.

In the near future, we will improve this model in two ways: 1) using parallel computation to reduce the current processing time - the time series produced by each stand are independent from the remaining ones; 2) modeling the weather changes as a factor to change our predictions (as it is addressed by using Hidden Markov Models to handle the occurrence of bursty events in [5]).

It will also be used as a feature of a recommendation system (to be developed) which will produce smart live recommendations to taxi drivers about which taxi stand they should head to after a drop-off. This decision support framework will also address other features like the distance or live traffic conditions, among others.

Acknowledgements. This work is part-funded by the ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for

competitiveness), by the Portuguese Funds through the FCT (Portuguese Foundation for Science and Technology) within project FCOMP - 01-0124-FEDER-022701.

References

1. Glashenko, A., Ivaschenko, A., Rzevski, G., Skobelev, P.: Multiagent real time scheduling system for taxi companies. In: Proceedings of 8th Int. Conf. on Autonomous Agents and Multi-Agent Systems, Budapest, Hungary (2009)
2. Lee, J., Park, G.-L., Kim, H., Yang, Y.-K., Kim, P., Kim, S.-W.: A Telematics Service System Based on the Linux Cluster. In: Shi, Y., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) ICCS 2007, Part IV. LNCS, vol. 4490, pp. 660–667. Springer, Heidelberg (2007)
3. Junghoon, L., Inhye, S., Park, G.: Analysis of the Passenger Pick-Up Pattern for Taxi Location Recommendation. In: Conference on Networked Computing and Advanced Information Management, vol. 1, pp. 199–204 (2008)
4. Bin, L., Daqing, Z., Lin, S., Chao, C., Shijian, L., Guande, Q., Qiang, Y.: Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. In: 2011 IEEE International Conference on Pervasive Computing and Communications Workshops, pp. 63–68 (2011)
5. Ihler, A., Hutchins, J., Smyth, P.: Adaptive Event Detection with Time-Varying Poisson Processes. In: 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, pp. 207-216 (2006)
6. Box, G., Jenkins, G., Reinsel, G.: Time series analysis. Holden-day, San Francisco (1976)
7. Matias, L., Gama, J., Mendes-Moreira, J., Sousa, J.F.: Validation of both number and coverage of bus Schedules using AVL data. In: Proceedings of IEEE Conference on Intelligent Transportation Systems, Funchal, Portugal, pp. 131–136 (2010)
8. Holt, C.: Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting 20, 5–10 (2004)
9. Box, G., Pierce, D.: Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. Journal of the American Statistical Association 65, 1509–1526 (1970)
10. Williams, B., Hoel, L.: Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. Journal of Transportation Engineering 129, 664–672 (2003)
11. Contreras, J., Espinola, R., Nogales, F.J., Conejo, A.J.: ARIMA models to predict next-day electricity prices. IEEE Transactions on Power Systems 18, 1014–1020 (2003)
12. Zhang, G.P.: Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159–175 (2003)
13. Cryer, J., Chan, K.: Time Series Analysis with Applications in R. Springer, USA (2008)
14. Wang, H., Fan, W., Yu, P.S., Han, J.: Mining concept-drifting data streams using ensemble classifiers. In: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 226–235. ACM, Washington, DC (2003)
15. Makridakis, S., Hibon, M.: The M3-Competition: results, conclusions and implications. International Journal of Forecasting 16, 451–476 (2000)
16. R Development Core Team, R: A Language and Environment for Statistical Computing, Vienna, Austria (2005)
17. Yeasmin, K., Hyndman, R.: Automatic Time Series Forecasting: The *forecast* Package for R. Journal of Statistical Software 27 (2008)