Layered Ensemble Model for Short-Term Traffic Flow Forecasting with Outlier Detection

A. Abdullatif, F. Masulli, S. Rovetta

***Abstract*—Real time traffic forecasting is a necessary requirement for traffic management in order to be able to evaluate the effects of different available strategies or policies. This paper focuses on short-term traffic flow forecasting by taking into consideration both spatial (road links) and temporal (lag or past traffic flow values) information. We propose a Layered Ensemble Model (LEM) which combines Artificial Neural Networks and Graded Possibilistic Clustering obtaining an accurate forecast of the traffic flow rates with outlier detection. Experimentation has been carried out on two different data sets. The former was obtained from real UK motorway and the later was obtained from simulated traffic flow on a street network in Genoa (Italy). The proposed LEM model for short-term traffic forecasting provides promising results and given the ability for outlier detection, accuracy, robustness of the proposed approach, it can be fruitful integrated in traffic flow management systems.**

I. Introduction

Traffic flow forecasting is one of the fundamental components in Intelligent Transportation Systems (ITS). There are mainly three approaches to solve problem of traffic forecasting in urban and network scale areas: Parametric models ([1], [2], [3]); Machine Learning and Computational Intelligence methods (non-parametric models) ([4], [5], [6]); Hybrid approaches which combine the features and capabilities of parametric and non-parametric models as in ([7], [8], [9]). Since the eighties, short term traffic forecasting has been part of most intelligent transportation systems. Many traffic forecasting approaches focus on the problem of freeway traffic forecasting in which the state of the road traffic is almost stable. In contrast, traffic forecasting in urban and network-scale areas is more complex because of the rapid change of traffic behavior and limited availability of sensors that can cover the whole network.

Many researches have developed approaches bases on non-parametric models, such as multilayer perceptron with a learning rule based on a Kalman filter [10], wavelet-based neural network [11], fuzzy-neural model [12], ARIMA model [13], graphical-lasso neural network [14], multi-task neural network [15], multi-task ensemble neural network [16], k-nearest neighbor non-parametric regression [17]. A good survey on the current literature is given in In [18].

In this paper we present a robust model for short-term traffic forecasting with online outlier detection combining clustering and Neural Network (NN).

II. Methodology

The proposed method consists of an ensemble of Neural Networks (NN), each specialized on a region of the data space. The first layer provides this local specialization, as well as providing a measure of pattern outlierness that is used to estimate the expected quality of the forecast. In the following we outline the methods used for clustering and forecasting.

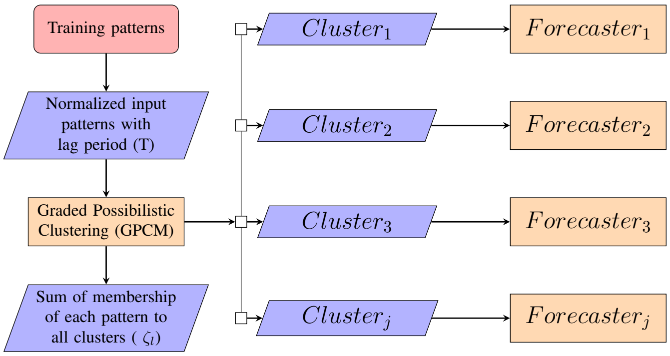


Fig. 1. Flow chart of training stage in the proposed model

A. LEM model

The Layered Ensemble Model (LEM) consists of 2 layers as shown in Figure 1. In the first layer we employ a clustering process called Graded Possibilistic c Means (GPCM) [19] that is able to adapt to the changes in the traffic flow, by implementing a continuous learning that exploits the input patterns as they arrive. So the first layer has two tasks: 1) Group patterns into K groups where pattern in the same group have similar behavior. 2) Measure outlierness degree of each pattern. The measure of outlierness will be used to assess and improve the accuracy of the second layer (forecaster). In the second layer we have an ensemble method based on base learners (in our case multilayer perceptron neural networks); each of them is trained on traffic flow patterns belonging to a different cluster.

B. Data preprocessing

For any forecaster model there are some issues that affect its performance:

- **Lag period:** As shown in Figure 1 the first issue is the proper selection of the lag period which is crucial because it affects the correct representation of the traffic flow in time. If the lag period is chosen to be smaller than needed, then we will not be able to distinguish between the time-lag vectors in the vector space (redundancy) [20]; hence, the prediction process will be practically impossible because data doesn't carry enough valuable information. If the lag period is chosen to be larger than needed, the vectors in the vector space will be almost uncorrelated (irrelevance) [21]. In [22] they proposed an

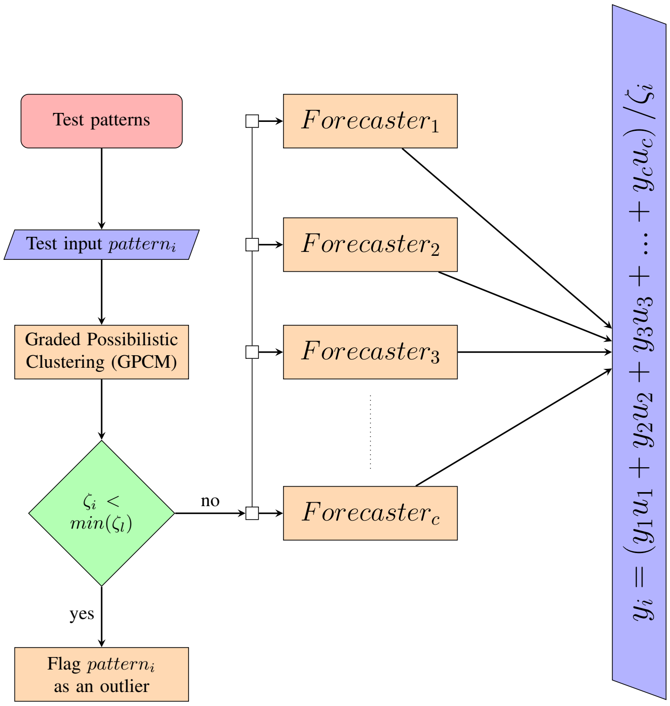


Fig. 2. Flow chart of the test stage in the proposed model

approach to tackle this problem by using the time-lag mutual information as a tool to determine a reasonable value of the lag period.

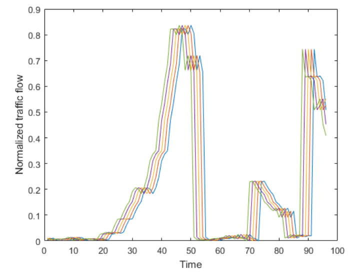
- **Historical period:** This refer to the number of observation patterns that will be used to train the forecasters.
- **Traffic flow patterns and outliers:** Learning patterns with a different behavior together in a Large data set tends to reduce the models performance. As a solution to this problem, we employ a clustering process called GPCM that is able to group patterns into K groups, where patterns in the same group have similar behavior as shown in Figure 3, and measures the degree of outlierness of each pattern. This measures will be used to improve the forecaster accuracy.

C. The Graded Possibilistic c Means Model

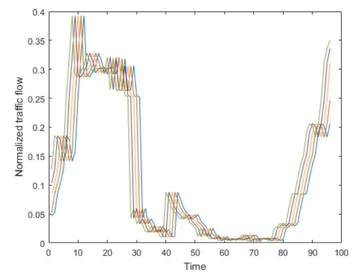
In central clustering data objects are points or vectors in data space, and c clusters are represented by means of their "central" points or centroids $\mathbf{y}_j$. The Graded Possibilistic model is a *soft* central clustering method, implying that cluster membership can be partial. This is usually represented by means of cluster indicators (or membership functions) which are real-valued rather than integer.

In many cases methods are derived as the iterative optimization of a constrained objective function [23], usually the mean squared distortion:

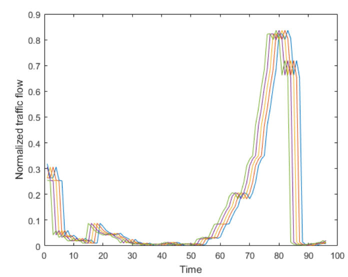
$$D = \frac{1}{n} \sum_{l=1}^{n} \sum_{j=1}^{c} u_{lj} ||\mathbf{x}_l - \mathbf{y}_i||^2 \tag{1}$$



(a)



(b)



(c)

Fig. 3. First five traffic flow patterns (UK motorway Site 30012533 (AL2989A) 2009) in (a) cluster 1; (b) cluster 2; (c) cluster 3.

Centroids are obtained by imposing $\nabla D = 0$:

$$\mathbf{y}_j = \frac{\sum_{l=1}^{n} u_{lj} x_l}{\sum_{l=1}^{n} u_{lj}} \ . \tag{2}$$

Usually constraints are placed on the sum $\zeta_l = \sum_{j=1}^{c} u_{lj}$ of all memberships for any given observation x_l. The value ζ_l can be interpreted as the total membership mass of observation x_l. We now survey from this perspective two related soft clustering method.

The Maximum Entropy (ME) approach [24] imposes $\zeta_l = 1$, so we are in the "probabilistic" case, where memberships are formally equivalent to probabilities.

The objective J_{ME} includes an entropic penalty with weight β, and $\nabla J_{\mathrm{ME}} = 0$ yields

$$u_{lj} = \frac{e^{-d_{lj}/\beta}}{\zeta_l} \ . \tag{3}$$

On the other end of the spectrum, the Possibilistic c-Means in its second formulation (PCM-II) [25] does not impose any constraint on ζ_l, so memberships are not formally equivalent to probabilities; they represent degrees of typicality.

The objective $J_{\mathrm{PCM-II}}$ includes an individual parameter β_j for each cluster, and $\nabla J_{\mathrm{PCM-II}} = 0$ yields

$$u_{lj} = e^{-d_{lj}/\beta_j} \ . \tag{4}$$

Note that eq. (4) is a special case of eq. (3) with ζ_l replaced by 1 $\forall l$, whereas eq. (3) is a special case of eq. (4) with $\beta_j = \beta \ \forall j$. In other words, both can be generalized to a unique formulation. This was done in [19] as follows:

$$u_{lj} = \frac{v_{lj}}{Z_l}, \tag{5}$$

where the *free membership* $v_{lj} = e^{-d_{lj}/\beta_j}$ is normalized with some term Z_l, a function of (but not necessarily equal to) ζ_l. This allows us to add a continuum of other, intermediate cases to the two extreme models just described, respectively characterized by $Z_l = \zeta_j = \sum_{j=1}^{c} v_{lj}$ (probabilistic) and $Z_l = 1$ (possibilistic). Here we use the following formulation:

$$Z_l = \zeta_l^{\alpha} = \left(\sum_{j=1}^{c} v_{lj} \right)^{\alpha} , \quad \alpha \in [0, 1] \subset \mathbb{R} \tag{6}$$

The parameter α controls the "possibility level", from a totally probabilistic ($\alpha = 1$) to a totally possibilistic ($\alpha = 0$) model, with all intermediate cases for $0 < \alpha < 1$.

D. Ensemble forecast model

As shown in Figure 1, for each cluster a forecaster with architecture shown in Table I is trained and ζ_l is obtained which is a vector of size size m where m is number of traffic patterns in the training data set.

After training stage, we start online test stage as shown in Figure 2 where patterns comes as a streams and for each upcoming **test pattern**$_i$ ζ_i is computed and compared to a threshold. In the proposed model we use min(ζ_l) as a threshold where $\zeta_i < min(\zeta_l)$ means that the test pattern is an extreme outlier and will be dropped. The drop rate of the test pattens depend on the value of α which controls the sensitivity of the model to outliers. A high value if α means less sensitivity to outlier and a lower drop rate.

The final output of the LEM is computed as a weighted sum of the individual base learner forecasts, as follows:

$$y_i = \sum_{j=1}^{c} y_j u_j / \zeta_i \tag{7}$$

In eq. (7) we see that the output of each forecaster is weighted by u_j, which is the membership degree of each pattern to each cluster so that u_j will have high value for the most suitable forecaster(s) and low to the others.

III. Experiments and Results

A. Data sets

UK road network: Multiple data sets are obtained from different road links in the United Kingdom (UK) [26]. This data series provides traffic flow information for 15-minute periods since 2009 on most of road links in UK. The data sets are obtined from different loop sensors but we selected the loop sensor id AL2989A (TMU Site 30012533) to obtain traffic flow as done in [27].

TABLE I
LEM model parameters

	UK data set	Genoa dataset
Lag period	1day (96 observations)	25 min (5 observations from 3 road links)
Observation period	15 min	5 min
Historical period for training	9 months	6 hours
Test period (forecasting period)	3 months	3 hours
Validation	10-fold cross validation	10-fold cross validation
Forecaster	Artificial Neural Network	Artificial Neural Network
Number of layers	3 (input,hidden,output)	3
Neural Network architecture	95-10-1	4-10-1
Number of clusters	5	5
α	$0 < \alpha < 1$	$0 < \alpha < 1$
β	.1	.01

Genoa Data set: The data were obtained as follows. An urban area of the city of Genoa, in the north-west of Italy, was mapped with the aid of Open Street Map data. Traffic parameters were obtained from actual observations and several days of traffic were estimated by using the SUMO open source traffic simulator [28]. The simulation yielded observations at time intervals of five minutes. in this data set we obtain traffic flow not only from main link but also from the adjacent links to forecast the future.

B. Experiments parameters and Results

The results in Figure 4 show on the top the scatter plot of the forecasted traffic flow in UK, and on the bottom the forecasted traffic flow in Genoa both with zero drop rate. In Figure 5 we show the effect of α on the accuracy (Mean square error) of the LEM model. The selected range of alpha values are $.93 \leq \alpha \leq 1$ and the drop rate is computed as follow:

$$\text{drop rate} = 1 - \left(\frac{\text{Number of output patterns}}{\text{Number of target patterns}} \right) \tag{8}$$

IV. Conclusion

The experiments show that the proposed method offers the ability to deal with typical patterns, tuning the forecaster in the most appropriate manner, while detecting atypical patterns.

In future work, the method will be expanded by using the outlierness information to estimate the presence of concept drift in the data.

References

[1] A. J. Fowe and Y. Chan, "A microstate spatial-inference model for network-traffic estimation," vol. 36, pp. 245–260, Jan 2013.

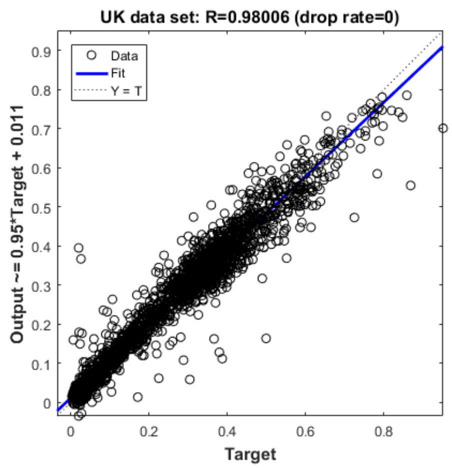
[2] M. Treiber and A. Kesting, "Validation of traffic flow models with respect to the spatiotemporal evolution of congested traffic patterns," vol. 21, pp. 31–41, Jan 2012.

[3] B. Ghosh, B. Basu, and M. O'Mahony, "Multivariate short-term traffic flow forecasting using time-series analysis," vol. 10, pp. 246–254, Jan 2009.

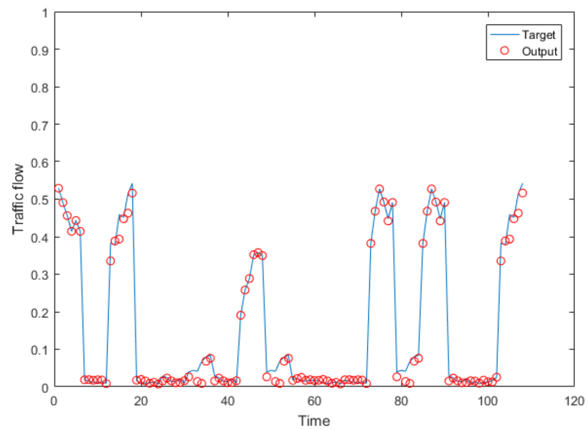
[4] F. Zheng and H. Van Zuylen, "Urban link travel time estimation based on sparse probe vehicle data," *Transportation Research Part C: Emerging Technologies*, vol. 31, pp. 145–157, Jan 2013.

[5] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," vol. 19, pp. 387–399, Jan 2011.

[6] F. Jin and S. Sun, "Neural network multitask learning for traffic flow forecasting," *IEEE*, pp. 1897–1901, Jun 2008.



(a)



(b)

Fig. 4. (a) Scatter plot between the target and the output (UK data set) (b) Forecasted output and the target (Genoa data set).

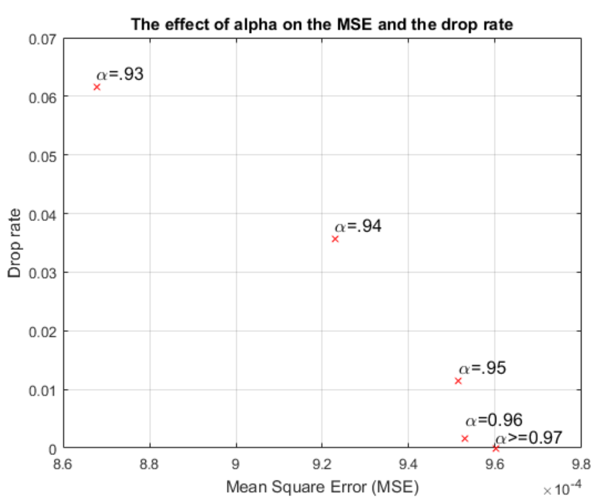


Fig. 5. LEM model accuracy w.r.t α in UK for 3 months.

[7] A. Stathopoulos and L. Dimitriou, "Fuzzy modeling approach for combined forecasting of urban traffic flow," *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, p. 521, 2008.

[8] Y. Zhang, "Hourly traffic forecasts using interacting multiple model (imm) predictor," *IEEE Signal Processing Letters*, vol. 18, pp. 607–610, Oct 2011.

[9] E. Vlahogianni, M. Karlaftis, and D. Tselentis, "Improving short-term traffic forecasts: to combine models or not to combine?" *IET Intelligent Transport Systems*, Jul 2014.

[10] P. Vythoulkas, "Alternative approaches to short term traffic forecasting for use in driver information systems," in *International Symposium on the Theory of Traffic Flow and Transportation (12th : 1993 : Berkeley, Calif.). Transportation and traffic theory*, 1992.

[11] J. Gao, Z. Leng, Y. Qin, Z. Ma, and X. Liu, "Short-term traffic flow forecasting model based on wavelet neural network," in *2013 25th Chinese Control and Decision Conference (CCDC)*, May 2013, pp. 5081–5084.

[12] H. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transportation Research Part C: Emerging Technologies*, vol. 10, pp. 85–98, Feb 2002.

[13] M. Hamed, H. AI-Masaeid, and Z. Bani Said, "Short-term prediction of traffic volume in urban arterials," *Journal of Transportation Engineering*, vol. 121, pp. 249–254, 1995.

[14] S. Sun, R. Huang, and Y. Gao, "Network-scale traffic modeling and forecasting with graphical lasso," *Advances in Neural Networks – ISNN 2011*, vol. 6676, pp. 151–158, 2011.

[15] Y. Gao and S. Sun, "Multi-link traffic flow forecasting using neural networks," in *2010 Sixth International Conference on Natural Computation*, vol. 1, Aug 2010, pp. 398–401.

[16] S. Sun, "Traffic flow forecasting based on multitask ensemble learning," in *Proceedings of the First ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, ser. GEC '09. New York, NY, USA: ACM, 2009, pp. 961–964.

[17] B. Yoon and H. Chang, "Potentialities of data-driven nonparametric regression in urban signalized traffic flow forecasting," *Journal of Transportation Engineering*, vol. 140, p. 04014027, 2014.

[18] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, p. 3, Jan 2014.

[19] F. Masulli and S. Rovetta, "Soft transition from probabilistic to possibilistic fuzzy clustering," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 4, pp. 516–527, August 2006.

[20] M. Casdagli, S. Eubank, J. Farmer, and J. Gibson, "State space reconstruction in the presence of noise," *Physica D: Nonlinear Phenomena*, vol. 51, no. 1, pp. 52 – 98, 1991.

[21] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. Cambridge University Press, 2003, cambridge Books Online.

[22] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A*, vol. 33, pp. 1134–1140, Feb 1986.

[23] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers, 1981.

[24] K. Rose, E. Gurewitz, and G. Fox, "A deterministic annealing approach to clustering," *Pattern Recognition Letters*, vol. 11, pp. 589–594, 1990.

[25] R. Krishnapuram and J. M. Keller, "The possibilistic C-Means algorithm: insights and recommendations," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 3, pp. 385–393, August 1996.

[26] "Uk agency network journey time and traffic flow data," tris.highwaysengland.co.uk/detail/trafficflowdata.

[27] A. Wibisono, W. Jatmiko, H. A. Wisesa, B. Hardjono, and P. Mursanto, "Traffic big data prediction and visualization using fast incremental model trees-drift detection (fimt-dd)," *Knowledge-Based Systems*, vol. 93, pp. 33 – 46, 2016.

[28] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO - Simulation of Urban MObility," *International Journal On Advances in Systems and Measurements*, vol. 5, no. 3&4, pp. 128–138, December 2012.