Foot Traffic Prediction for Large-Scale Events Based on Pattern-Aware Neural Regression

Takafumi Okukubo[1,2], Yoshiaki Bando[2], Masaki Onishi[2], Hiroyasu Ando[1]

[1] Graduate School of Sciece and Technology, Degree Programs in Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba-shi, Ibaraki, Japan

[2] National Institute of Advanced Industrial Science and Technology (AIST), 2-3-26 Aomi, Koto-ku, Tokyo, Japan

s1920553@s.tsukuba.ac.jp, y.bando@aist.go.jp, onishi@ni.aist.go.jp, ando@sk.tsukuba.ac.jp

Abstract

This paper presents a statistical method that can efficiently predict foot traffic from a limited amount of training data. The traffic prediction is essential for forecasting and controlling excessive congestion after a large-scale event. To handle a small amount of training data due to the nature of a large-scale event, we propose a pattern-aware neural regression method that interprets traffic data as a sum of latent behavior patterns weighted by their activations. The neural regression model predicts the weights of the patterns from the event information. The method trains the regression model with a variational expectation-maximization algorithm to prevent overfitting. With real traffic data at a soccer stadium, the experiments demonstrate that the proposed method outperforms the conventional direct regression methods. We also show that our Bayesian approach performs more stable prediction than an existing maximum likelihood approach.

Introduction

Predicting traffic congestion is an essential function for urban transportation to keep its safety and comfortability (Zhang et al. 2011; Alsrehin, Klaib, and Magableh 2019). A typical traffic congestion of vehicles and people often occurs after a large-scale event, such as sports and entertainment events. A large number of visitors try to leave the event venue at the same time, and the resulting traffic jam often causes serious accidents around the event site if no precautionary measures are take against such a disaster. Various methods have been investigated for forecasting and controlling excessive traffic congestion (Jiang et al. 2019; Liao et al. 2018; Ni, He, and Gao 2016).

Since the existing prediction methods rely on the training data of past traffic histories (Nihan and Holmesland 1980; Xu et al. 2017; Kang, Lv, and Chen 2017), their performance deteriorates with a limited amount of the training data. Although the traffic prediction for large-scale events is an important task to avoid serious accidents, it is often difficult to prepare a large amount of traffic data enough to train a prediction method. A soccer stadium, for example, holds a few major games, where a large number of spectators visit in a year. The traffic data of other similar events, in addition, would not improve the performance because the traffic situation varies depending on their surrounding environments.

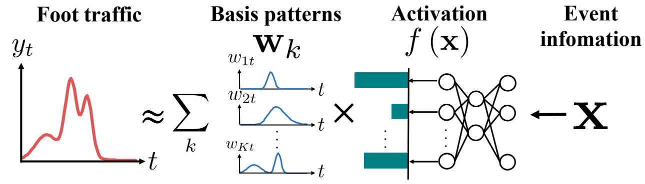


Figure 1: Overview of the proposed foot traffic prediction.

A promising approach to handle a limited amount of training data is to assume a small set of behavior patterns of people. A group of people, for example, leaves an event site immediately after the event is over while another group stays longer to bask in the afterglow of the event. The observed traffic could be considered as a sum of such latent behavior patterns weighted by the number of people corresponding to each group. By predicting the weights of the basis behavior patterns instead of directly predicting the raw time series, the traffic prediction will be much improved because we can significantly reduce the number of predicted values.

In this paper, we present a pattern-aware neural regression method for traffic prediction that can make great use of a limited amount of training data (Figure 1). We predict a time series of foot traffic at an exiting gate of a large-scale event by taking as input the event information (e.g., starting time, weather, the number of spectators). The behavior patterns are handled by non-negative matrix factorization (NMF) (Cemgil 2008), which interprets the traffic time series as a sum of non-negative basis behaviors weighted by their activations. More specifically, our method trains a neural regression model to predict the activations while jointly finding the basis behaviors for the training data. This framework is formulated as a Bayesian generative model for a traffic series to avoid the overfitting of the model parameters.

The main contribution of this study is to introduce the low-rank approximation by NMF into traffic prediction for handling a small amount of training data. We formulate a probabilistic model suitable for the traffic prediction task, whereas several studies have utilized NMF for predicting a time series (Zhu and Yan 2018; Mei et al. 2018). We evaluate the proposed method by real foot traffic data at an exiting gate of a soccer stadium. The experimental results show that the proposed method outperforms the conventional methods.

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Related Work

This section introduces existing traffic prediction methods and non-negative matrix factorization methods.

Traffic Prediction

Traffic prediction has been studied by using various frameworks such as autoregressive models (Nihan and Holmesland 1980), Kalman filters (Xu et al. 2017), and recurrent networks (Kang, Lv, and Chen 2017). The amount of traffic can be predicted better by using additional information, such as weather (Abrishami, Kumar, and Nienaber 2017) and the traffic data observed at other places (Duan et al. 2019; Lin et al. 2019). Such a technique has also been utilized for predicting the traffic after a large-scale event by using tracking data of Global Positioning System (GPS) (Jiang et al. 2019), search queries of a web mapping service (Liao et al. 2018), and event-related information from social media (Ni, He, and Gao 2016). Since these methods predict a traffic time series from long-term traffic history as training data, they suffer from the overfitting due to lack of the training data.

Non-negative Matrix Factorization

A popular approach to efficiently reduce the number of parameters is to use non-negative matrix factorization (NMF) (Xu, Liu, and Gong 2003), which approximates a non-negative vector with a small number of template (basis) vectors. NMF has been utilized for audio source separation (Févotte, Bertin, and Durrieu 2009), document recognition (Xu, Liu, and Gong 2003), and social media analysis (Pei, Chakraborty, and Sycara 2015). NMF has also been combined with multivariate regression models to preventing overfitting (Zhu and Yan 2018; Mei et al. 2018).

The most of NMF methods are based on a minimum mean square error (MMSE) criterion, which corresponds to the maximum likelihood criterion with the Gaussian likelihood (Xu, Liu, and Gong 2003). Since the observation of NMF is non-negative, the Gaussian likelihood that can take negative values is inconsistent in a sense of the probabilistic generative model. It has been reported that the performance of NMF can be improved by appropriate selection of the likelihood function. Poisson and exponential likelihoods are, for example, known to be suitable for audio source separation (Févotte, Bertin, and Durrieu 2009). It is also reported that the performance can be improved by estimating the latent vectors in a Bayesian manner (Cemgil 2008).

Proposed Method

This section describes our foot traffic prediction that introduces the low-rank assumption of NMF for efficient training.

Problem Specification

The problem setting for this study is specified as follows:

Inputs: a feature vector $\mathbf{x}^* \in \mathbb{R}^D$ of the target event
Outputs: the expected value of foot traffic series $\hat{\mathbf{y}}^* \in \mathbb{R}_+^T$
Training data: N pairs of $\mathbf{x}^{(n)} \in \mathbb{R}^D$ and $\mathbf{y}^{(n)} \in \mathbb{N}$ that are collected in advance,

where T represents the number of time frames. The feature vector represents the characteristics of the predicted event.

Model Formulation

We formulate the foot traffic $y_t \in \mathbb{N}$ at time t by using a Poisson regression model as follows:

$$y_t \mid f, \mathbf{x} \sim \text{Poisson}\left(f_t\left(\mathbf{x}\right)\right), \tag{1}$$

where $f_t : \mathbb{R}^D \rightarrow \mathbb{R}_+$ is a nonlinear mapping that predicts the expected value of y_t. This direct regression, however, often fails to predict the time series due to the large number of parameters for all the nonlinear mappings f_t ($t = 1, \ldots, T$). To solve this problem, we introduce K basis vectors $\mathbf{w}_k \in \mathbb{R}_+^T$ each of which represent a template traffic pattern of visitors as follows:

$$y_t \mid f, \mathbf{x}, \mathbf{W} \sim \text{Poisson}\left(\sum_{k=1}^{K} w_{kt} f_k\left(\mathbf{x}\right)\right) \tag{2}$$

where $f_k : \mathbb{R}^D \rightarrow \mathbb{R}_+$ is a nonlinear mapping that predicts the weight for the k-th traffic pattern. Here, we formulate the traffic as the sum of the basis vectors weighted by f_k instead of the direct regression by f_t. Setting a small number of bases $K \ll T$, this formulation significantly reduces the number of the parameters for the nonlinear mapping. In addition, to prevent $\mathbf{W}$ from the overfitting, we put the following gamma prior on $\mathbf{W}$, which is a conjugate prior of the Poisson likelihood:

$$w_{kt} \sim \text{Gamma}\left(a_0, b_0\right), \tag{3}$$

where a_0 and b_0 are hyperparameters to control the sparseness and scale of w_{kt}, respectively.

Training

Given N training pairs of $\mathbf{x}^{(n)}$ and $\mathbf{y}^{(n)}$, we find the regression model f_k that maximizes the log-marginal likelihood $\log p(\mathbf{Y} \mid \mathbf{X}, f)$. Because this log-marginal likelihood is intractable, we approximately optimize the likelihood by introducing the following variational posterior distribution q:

$$p(\mathbf{W} \mid \mathbf{X}, f) \approx q(\mathbf{W}) \triangleq \prod_k q(\mathbf{w}_k). \tag{4}$$

The training maximizes a lower-bound of the log-marginal likelihood, namely the evidence lower bound (ELBO):

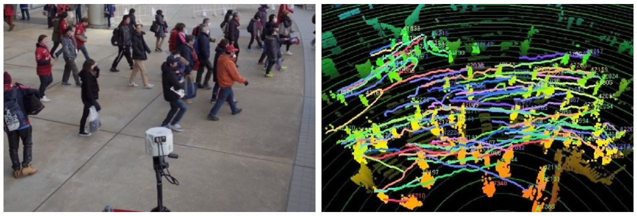
$$\mathcal{L} = \mathbb{E}_q[\log p\left(\mathbf{Y} \mid \mathbf{X}, f, \mathbf{W}\right)] - \mathcal{D}_{\mathrm{KL}}[q\left(\mathbf{W}\right) | p\left(\mathbf{W}\right)]. \tag{5}$$

The whole training procedure is formulated as a variational expectation-maximization (EM) algorithm (Bishop 2006) that iteratively and alternately updates the nonlinear mapping f_k and the variational posterior q. In the E step, we update the variational distribution $q\left(w_{tk}\right)$ to maximize the ELBO as in the Bayesian NMF (Cemgil 2008):

$$q\left(w_{tk}\right) \leftarrow \text{Gamma}\left(a_{tk}, b_{tk}\right), \tag{6}$$

$$a_{tk} = a_0 + \sum_{n=1}^{N} y_t^{(n)} \psi_{tk}^{(n)}, \;\; b_{tk} = b_0 + \sum_{n=1}^{N} f_k\left(\mathbf{x}^{(n)}\right), \tag{7}$$

where $\psi_{tk}^{(n)} \propto \exp\left(\mathbb{E}_q\left[\log w_{tk}\right]\right) f_k\left(\mathbf{x}^{(n)}\right)$ is an auxiliary variable satisfying $\sum_{k=1}^{K} \psi_{tk}^{(n)} = 1$. In the M step, on the other hand, we update f_k to maximize the ELBO of Eq. (5) by using a stochastic gradient descent method. In this paper, we used the Adam optimizer for updating f_k.



(a) Exiting people and LiDAR (b) Visualized tracking results

Figure 2: Traffic data collection.

Table 1: Description of event information.

Feature	Value range	Description
First half win	{0, 1}	Win or not at the 1st half
Kickoff time	[14, 19]	The start time of the game
Spectator count	[13502, 34312]	Sum of Spectators
Half-time temp.	[9.8, 28.3]	Temperature (°C) at the end of the 1st half

Prediction

The predicted value $\hat{\mathbf{y}}^*$ for a new feature vector $\mathbf{x}^*$ is calculated as an expected value of the predictive distribution $p\left(\mathbf{y}^* \mid \mathbf{x}^*, f, \mathbf{Y}, \mathbf{X}\right)$ as follows:

$$\hat{y}_t^* = \frac{1}{L} \sum_{l=1}^{L} \sum_{k=1}^{K} w_{kt}^{(l)} f_k\left(\mathbf{x}^*\right) \tag{8}$$

where $w_{kt}^{(l)} \sim q(w_{kt})$ is a sample of the variational distribution obtained at the training stage.

Experimental Evaluation

This section describes the experimental results with foot traffic data recorded at an actual soccer stadium.

Dataset

The proposed method was evaluated with real foot traffic data recorded in the Kashima Soccer Stadium, Japan. We recorded the traffic time series of exiting visitors for 15 games between December 2018 and December 2019. The traffic data collection was conducted at a gate of the stadium by using a human tracking system based on a light detection and ranging (LiDAR) sensor (Onishi 2015), which can observe human movement in a range between 0.5 m and 22 m from the sensor at 10 fps (Figure 2). The foot traffic was recorded for 120 minutes from 90 minutes after the beginning of each game, which corresponds to approximately 30 minutes before the end of the second half. The recorded data were split into time frames with the window length of one minute to count the number of people leaving the stadium in each frame depending on their moving directions. To predict the foot-traffic time series, we used the event information summarized in Table 1. These features were determined from our empirical experience at the stadium. They were standardized by their means and standard deviations. For continuous variables in them, we also used the second, third, and fourth powers of them as the feature. In summary, we used a $D = 13$ dimensional vector for predicting.

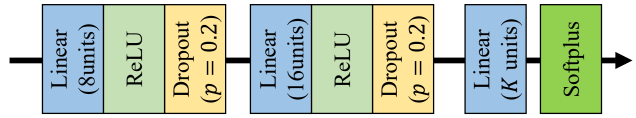


Figure 3: Configuration of neural regression model.

Table 2: Comparison of prediction performance.

Method	Likelihood	K	RMSE	MAPE
Direct regression	Gaussian	N/A	34.60	0.572
	Poisson	N/A	36.92	0.585
NMF-based regression	Gaussian	8	30.14	0.511
	Poisson	7	27.39	0.454
BNMF-based regression	**Poisson**	**11**	**26.77**	**0.440**

Experimental conditions

Since we had only traffic data for 15 games, we conducted a leave-one-out cross validation by using one game for evaluation and the other $N = 14$ games for training in each validation. As shown in Figure 3, we used a three-layer neural network as the regression model f_k. The training was conducted by using an Adam optimizer with a learning rate of 1.0×10^{-2}. The hyperparameters a_0 and b_0 were set to 1.0 and $1.0 \times K$, respectively. The iteration of the EM algorithm was conducted for 500 times. The network weights were updated 100 times for every update of the basis vectors. These parameters were determined experimentally by using the cross validation.

The proposed method was evaluated with two baselines. One was a direct method that predicts a traffic time series directly from the feature vector. The other was an NMF-based method similar to the proposed framework, in which the basis vectors were updated by the maximum likelihood manner instead of our Bayesian manner. For both the methods, we evaluated the loss functions based on the Gaussian and Poisson likelihood functions. We used the same network configuration for the baseline methods as our method. The prediction performances were evaluated by root mean square errors (RMSEs) and mean absolute percentage errors (MAPEs). MAPEs were calculated as the mean absolute errors divided by the mean value of the whole foot traffic.

Experimental Results

Table 2 summarizes the prediction performance of each method with K having the best RMSE. Firstly, NMF-based regression improved the prediction performance in both RMSE and MAPE more than the direct regressions. The performance in the NMF-based regression was also improved by replacing the Gaussian likelihood with the Poisson likelihood. These results demonstrated that our proposal of the pattern-aware regression and generative model approach with the Poisson likelihood was effective for traffic prediction under the small amount (14 samples) of training data. In addition, our Bayesian extension (denoted by BNMF-based regression), which estimates the posterior distribution of basis behavior vectors, further improved the performance in both the RMSE and MAPE. This result shows

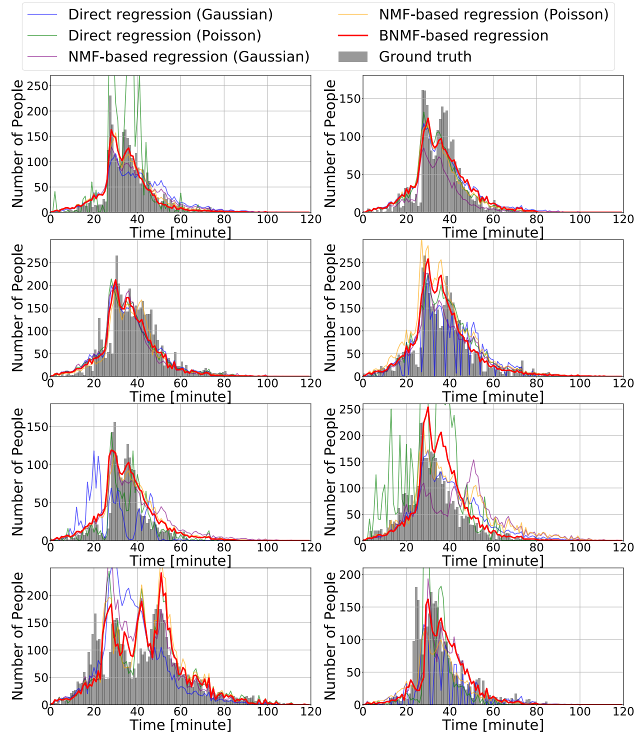


Figure 4: Excerpts of predicted foot traffic.

that our proposal of the Bayesian inference was effective to prevent the model parameters from overfitting. Figure 4 shows excerpts of the predicted foot traffics. We can see that the direct regression methods (blue and green lines) often drastically failed to predict the traffic. In contrast, the proposed BNMF-based regression (red lines) stably predicted the foot traffic.

Conclusion

In this paper, we presented a framework to efficiently predict foot traffic congestion from a small amount of data. Our method predicts the weights of the basis traffic patterns from the event feature vectors. We formulated a probabilistic generative model of traffic data based on NMF and trained a neural regression model in a variational EM algorithm to prevent the overfitting of the model parameters. The generative model is formulated with a Poisson likelihood, which is a natural representation of count data. The experimental results with real foot traffic data from large-scale events showed that our method outperformed the conventional direct regression method. The proposed method also outperformed the NMF-based regression methods trained by maximum likelihood estimation with a Gaussian likelihood, which is used in the conventional methods. Our future work is to enable online (real-time) model updating to deal with unexpected or difficult-to-predict situations.

Acknowledgement

This work was partially supported by JSPS KAKENHI Grant Number JP19H04943.

References

Abrishami, S.; Kumar, P.; and Nienaber, W. 2017. Smart stores: A scalable foot traffic collection and prediction system. In *ICDM*, 107–121.

Alsrehin, N. O.; Klaib, A. F.; and Magableh, A. 2019. Intelligent transportation and control systems using data mining and machine learning techniques: A comprehensive study. *IEEE Access* 7: 49830–49857.

Bishop, C. M. 2006. *Pattern recognition and machine learning*.

Cemgil, A. T. 2008. Bayesian inference for nonnegative matrix factorisation models. *Computational intelligence and neuroscience* 2009.

Duan, P.; Mao, G.; Kang, J.; and Huang, B. 2019. Estimation of Link Travel Time Distribution With Limited Traffic Detectors. *IEEE Trans. on ITS* .

Févotte, C.; Bertin, N.; and Durrieu, J.-L. 2009. Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *Neural computation* 21(3): 793–830.

Jiang, R.; Song, X.; Huang, D.; Song, X.; Xia, T.; Cai, Z.; Wang, Z.; Kim, K.-S.; and Shibasaki, R. 2019. Deepurbanevent: A system for predicting citywide crowd dynamics at big events. In *KDD*, 2114–2122.

Kang, D.; Lv, Y.; and Chen, Y.-y. 2017. Short-term traffic flow prediction with LSTM recurrent neural network. In *IEEE ITSC*, 1–6.

Liao, B.; Zhang, J.; Wu, C.; McIlwraith, D.; Chen, T.; Yang, S.; Guo, Y.; and Wu, F. 2018. Deep sequence learning with auxiliary information for traffic prediction. In *KDD*, 537–546.

Lin, Z.; Feng, J.; Lu, Z.; Li, Y.; and Jin, D. 2019. Deepstn+: Context-aware spatial-temporal neural network for crowd flow prediction in metropolis. In *AAAI*, volume 33, 1020–1027.

Mei, J.; De Castro, Y.; Goude, Y.; Azais, J.-M.; and Hébrail, G. 2018. Nonnegative matrix factorization with side information for time series recovery and prediction. *IEEE Trans. on KDE* 31(3): 493–506.

Ni, M.; He, Q.; and Gao, J. 2016. Forecasting the subway passenger flow under event occurrences with social media. *IEEE Trans. on ITS* 18(6): 1623–1632.

Nihan, N. L.; and Holmesland, K. O. 1980. Use of the Box and Jenkins time series technique in traffic forecasting. *Transportation* 9(2): 125–143.

Onishi, M. 2015. Analysis and Visualization of Large-Scale Pedestrian Flow in Normal and Disaster Situations. *ITE Trans. on MTA* 3(3): 170–183.

Pei, Y.; Chakraborty, N.; and Sycara, K. 2015. Nonnegative matrix tri-factorization with graph regularization for community detection in social networks. In *IJCAI*.

Xu, D.-w.; Wang, Y.-d.; Jia, L.-m.; Qin, Y.; and Dong, H.-h. 2017. Real-time road traffic state prediction based on ARIMA and Kalman filter. *Frontiers of Information Technology & Electronic Engineering* 18(2): 287–302.

Xu, W.; Liu, X.; and Gong, Y. 2003. Document clustering based on non-negative matrix factorization. In *SIGIR*, 267–273.

Zhang, J.; Wang, F.-Y.; Wang, K.; Lin, W.-H.; Xu, X.; and Chen, C. 2011. Data-driven intelligent transportation systems: A survey. *IEEE Trans. on ITS* 12(4): 1624–1639.

Zhu, W.; and Yan, Y. 2018. Joint linear regression and nonnegative matrix factorization based on self-organized graph for image clustering and classification. *IEEE Access* 6: 38820–38834.