Predictive Population Behavior Analysis from Multiple Contexts with Multilinear Poisson Regression

Masamichi Shimosaka
Tokyo Institute of Technology
Tokyo, Japan
simosaka@miubiq.cs.titech.ac.jp

Takeshi Tsukiji
The University of Tokyo
Tokyo, Japan
tsukiji@miubiq.cs.titech.ac.jp

Hideyuki Wada
The University of Tokyo
Tokyo, Japan
wada@miubiq.cs.titech.ac.jp

Kota Tsubouchi
Yahoo Japan Corporation
Tokyo, Japan
ktsubouc@yahoo-corp.jp

ABSTRACT

Predicting behaviors of a population from location-oriented log data from smartphones, i.e., urban population dynamics, has become more common in mobile and pervasive computing. A bilinear representation approach has been proposed to improve the prediction accuracy of urban population dynamics by adding contexts such as geographical information and day of the week. However, this approach has a strong limitation in that additional contexts can not be directly utilized in this representation with a unified manner. To resolve this issue, we propose a new predictive model for urban population dynamics based on *multilinear* Poisson regression so as to handle multiple contexts in a systematic manner. The model is parameterized using a tensor and can be optimized by using an efficient convex optimization with a sequence of matrix parameter optimizations. An empirical evaluation with large-scale smartphone location data showed that our model outperforms conventional approaches.

CCS CONCEPTS

• **Information systems** → **Information systems applications**; *Spatial-temporal systems*; Geographic information systems;

KEYWORDS

Poisson regression, GPS logs, Tensor-based model

1 INTRODUCTION

The spread of mobile phones and the accumulation of massive logs of location data have motivated the research of urban computing for analyzing crowd behavior. For example, urban analyses such as of population density [3], population dynamics patterns [1] and population predictions [7] have been conducted. Fan et al. [1] showed that time series behaviors of an active population, which we call the dynamics patterns, are closely related to demographics (on, e.g., residential areas, sightseeing spots, and traffic nodes). Analyses of the dynamics patterns could, for instance, be used to recommend a suitable location for opening a new shop.

Thanks to the properties of location data for long-term crowd data spanning a range of days, times of the day, and locations, researchers often use tensor or matrix factorization to identify dynamics patterns in the dataset [1]; most of these studies are inspired by the success of recommendation systems that use factorization approaches [5]. Although factorization approaches have gained attention in the community, they still face a problem called as the cold start issue [2] when rating new items for new users. In terms of population dynamics prediction, new users correspond to a new region, while new items correspond to a day in the near future.

In contrast to those trends using factorization techniques, Shimosaka et al. [7] proposed a predictive approach to avoid the cold start issue. In their approach, the issue of predicting population dynamics in the near future can be resolved by parametrizing the context affecting the population dynamics. In terms of parametrizing contexts affecting population dynamics, a matrix is used instead of a vector as a parameter of a predictive model, and a representation differentiating the peak variances is enriched with respect to multiple factors, e.g., weather conditions and the day of the week.

However, the parameterization issue has not been resolved in a systematic manner for use to be able to consider other external factors in addition to weather conditions or day of the week. To the best of our knowledge, there is no systematic feature design for urban dynamics prediction when multiple contexts are used for richer representation. This means the main challenge with predictive models is how to utilize additional external factors such as city attributes and etc., that have been used in other research. In addition, we have to consider how to resolve the overfitting issue that arises as the number of parameters increases.

The goal of our research is to construct a new predictive model based on a tensor parameter, which is a multilinear prediction, and to verify the superiority of our proposed model over conventional approaches.

Our contributions of this paper can be summarized as follows: 1) we propose a novel predictive model for urban population dynamics based on multilinear Poisson regression for exploiting multiple contexts such as the weather, day of the week, holidays, and land attributes by leveraging tensor parameterization; 2) we provide an efficient convex optimization scheme for inducing a low-rank parameter tensor and reducing the risk of overfitting that arises from the increase in the number of parameters; 3) we conducted experiments using both long-term and short-term population scale smartphone location datasets and showed the superiority of our model against previous and state-of-the-art models.

2 POPULATION PREDICTION WITH MULTILINEAR POISSON REGRESSION

2.1 Problem Setting

Here, we present the model of the time series of a population per day in each target region. One day is divided into a number of time segments, and we define the total number of smartphone location data in a certain time segment as the active population in the target area of the time segment. We model urban dynamics on the assumption that the active population, which is a non-negative integer, of each time segment follows a Poisson distribution.

Let S be the number of time segments in a day and $h = [h_1, h_2, \cdots, h_S] \in \mathbb{R}^S$ be a one day active population transition, where h_s indicates the active population of the s-th time segment. Following the previous research [7], we assume that each h_s follows a Poisson distribution, and write the likelihood of h_s as $p(h_s) = \text{Pois}(h_s|\lambda_s) = \lambda_s^{h_s} \exp(-\lambda_s)/\Gamma(h_s + 1)$, where λ_s is the parameter of the Poisson distribution. In this setting, our task is to determine λ_s as an estimate of the true active population h_s by using many factors such as time, day, and weather.

2.2 Definition of Variables

Since our work is inspired by the success of the predictive model with bilinear regression [7], we will borrow the basic variable definition from [7]. We use the time factor and external factors other than time factor affecting urban dynamics. The time factor $t \in \mathbb{R}^S$ is a feature value with the spread of the Gaussian distribution, the mean of which is the corresponding time segment s. Let t_k be the k ($k = 1, 2, \cdots, S$)th element of the time factor t of the time segment s, and following [7], we express t as a Gaussian distribution with mean parameter s and variance parameter σ: $t = \{t_k | t_k = \mathcal{N}(k|s, \sigma), k = 1, \cdots S\}$. When t is a feature in the time segment s, $\lambda(t)$ is the predicted active population, and h_s is the actual active population. The external factors d include ones such as the weather, day of the week, and land attributes.

With the bilinear representation, when the weight parameter matrix is set to $W \in \mathbb{R}^{D \times S}$, where $D = \dim(d)$, $\lambda(t, d) = \exp(d^\top W t)$. In the state-of-the-art bilinear model [7], W is calculated as the product of $U \in \mathbb{R}^{D \times K}$ and $V \in \mathbb{R}^{K \times S}$, where K is an arbitrary value that satisfies $DS >> (D + S)K$, and trained by the alternating

least squares method. Let me note that their bilinear model falls into local optima due to its non-convex optimization scheme, and cannot provide systematic means of describing the feature when multiple contexts are plugged into feature d, which is the limitation of bilinear representation.

2.3 Formulating the proposed model

For the sake of simplicity, we describe a case where the weight parameter is an Nth-order tensor. The multilinear Poisson regression model uses the time factor $t \in \mathbb{R}^{r_1}$ and the external factors $d_n \in \mathbb{R}^{r_n}$ ($n = 2, 3, \cdots, N$). This notation in which n starts from 2 enables us to regard t as d_1, one of the factors of the multilinear model, and makes the following explanation easier. Here, r_1 denotes the dimension of the time factor t (note that $r_1 = S$), and r_n indicates the dimension of the n-th external factor d_n. When external factors are divided into $d_2, d_3, \cdots, d_N$ and when the weight parameter tensor is set to $\mathcal{W} \in \mathbb{R}^{r_1 \times r_2 \cdots \times r_N}$, the Poisson distribution parameter λ is given by the following equation: $\lambda(t, d_2, d_3, \cdots, d_N) = \exp(\mathcal{W} \times_1 t \times_2 d_2 \times_3 d_3 \cdots \times_N d_N)$.

Note that $\mathcal{W} \times_n d_n$ is the mode-n multiplication of a tensor $\mathcal{W} \in \mathbb{R}^{r_1 \times r_2 \cdots \times r_N}$ by a vector $d_n \in \mathbb{R}^{r_n}$ and its product has dimension $r_1 \times \cdots r_{n-1} \times r_{n+1} \times \cdots \times r_N$. Using element-wise format we have

$$\{\mathcal{W} \times_n d_n\}_{i_1, \cdots, i_{n-1}, i_{n+1}, \cdots, i_N} = \sum_{i_n=1}^{r_n} \{\mathcal{W}\}_{i_1, \cdots, i_n, \cdots, i_N} \{d_n\}_{i_n}$$

(1)

where curly braces with subscripts denote the element of a tensor / vector enclosed by the braces, and i. denotes the index of a tensor.

In the multilinear Poisson regression model, the logarithm of the likelihood $L(\mathcal{W})$ is written as

$$\begin{aligned}
\ln L(\mathcal{W}) &= \sum_s \sum_m \ln \text{Pois}(y_{m,s}|\lambda_{m,s}) \\
&= \sum_s \sum_m \{y_{m,s}(\mathcal{W} \times_1 t_{m,s} \times_2 d_{2m,s} \cdots \times_N d_{Nm,s}) \\
&\quad - \exp(\mathcal{W} \times_1 t_{m,s} \times_2 d_{2m,s} \cdots \times_N d_{Nm,s}) \\
&\quad - \ln(\Gamma(y_{m,s} + 1))\},
\end{aligned}$$

where m is the index of days of the dataset and s is the index of the time segment in a day. When tensor parameters $\mathcal{W} \in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_N}$ are of the Nth-order (r_i is the dimension of d_i), the matrix $W_{(n)} \in \mathbb{R}^{r_n \times \frac{\prod_{i=1}^N r_i}{r_n}}$ is an expansion of $\mathcal{W}$ in the form of a matrix for each dimension and is treated as a matrix expanded in mode n. We consider the expanded matrix in a plurality of modes because the tensors are different from the matrices.

2.4 Parameter optimization

Because of the increase in the number of the parameters in our model, the overfitting issue must be dealt with in the optimization process. Although an explicit reduction in the number of parameters via Tucker decomposition or CP decomposition can resolve this issue [9], explicit rank reduction has the disadvantage of making it easy for the optimization to fall into a local minimum. Instead, convex relaxation is more useful for making the tensor parameters low-rank because it guarantees convergence to a global solution.

Inspired by the success of [6], we use trace norm regularization as a convex relaxation for regularizing tensor parameters instead of using explicit rank reduction. Note that the trace norm of a matrix is a sum of the singular values of the matrix [4]. Analogous to the case of matrices, the trace norm in tensors is a sum of the trace norms of fiber matrices. The properties of the trace norm of matrices promote sparsity in terms of singular values and lead to low-rank matrices. Thus, it helps to reduce the number of parameters. Using trace norm regularization enables the global optima to be reached and provides stable results in terms of prediction accuracy.

When it comes to the trace norm regularization term, we should selectively avoid making the time direction (order) low-rank because time-specific information such as the peak time is important in urban dynamics. In our model, we employ a modified version of the trace norm regularization term $\Omega(\mathcal{W})$, defined as $\Omega(\mathcal{W}) = \|\mathcal{W}\|_{\mathrm{tr}} = \sum_{n=2}^{N} \|W_{(n)}\|_{\mathrm{tr}}$. Finally, we consider the following optimization problem: $\hat{\mathcal{W}} = \mathrm{argmin}_{\mathcal{W}}\{-\gamma \ln L(\mathcal{W}) + \Omega(\mathcal{W})\}$, where γ is the coefficient determined through empirical evaluation. To promote sparsity in terms of singular values and get a low-rank matrix, the objective function should be optimized at each of the modes of $\mathcal{W}$. Therefore, we employ the alternating direction multiplier method (ADMM) [8], in which each sub-problem (i.e. optimization in each mode) is treated as a matrix trace norm minimization.

3 EXPERIMENTAL RESULTS

To evaluate the performance of our method with active population data from inside cities, we conducted two experiments on urban dynamics prediction with large-scale smartphone location data. In the first experiment, we created a single multilinear Poisson regressor with land attribution data for several cities to show that interaction across the factors must be considered in order to predict urban dynamics precisely. In the second experiment, we created multiple multilinear models where a single model corresponded to a region of interest, to show that our method is better when we consider multiple contexts. The results are discussed in Section 4.

3.1 Dataset

We used location information obtained from a disaster prevention application provided by Yahoo! JAPAN[1] as input data. The application records the GPS data of individuals who have agreed to provide their location data. The original dataset can be described by a set of 4-tuples: anonymized user ID, timestamp, longitude, and latitude. Essentially, GPS logs are generated when the smartphone user starts to move. The application has more than 1 million registered users (sample rate approximately 1% throughout Japan) from GPS logs that are continuously collected every day. Information about the user was not included in any of the experimental data.

We used location data that had been acquired during one year, from July 1, 2013 to June 30, 2014, from users in the Kanto area of Japan, and the number amounted to about 15 million per day. We selected 300 points in the Kanto region as points of interest (POI). Each POI corresponded to a 900-meter square area. We utilized the same scheme for the data collection for a fair comparison with past research [7]. The location data for each day was divided into 48 time segments, i.e., $S = 48$; each time segment was 30 minutes long.

3.2 Creating one predictor in several cities including land attributes

In this experiment, we used land attributes, weather, and holiday or weekday as the external factors. We used the same data obtained by the crowd sourcing used in [7] as the features of the land attributes. We obtained 120,000 valid answers from 1500 crowd workers. The land attributes were composed of ten items and the number of dimensions of this factor was 10. The value of each item was set to the ratio of answers obtained from crowd sourcing. The weather features were created from information obtained from the Japan Meteorological Agency's website[2]. The Meteorological Agency's home page provides 13 weather categories and we converted them into 4 weather dimensions: sunny, cloudy, rainy, and bad conditions. The feature of holidays or weekdays, regarding public holidays and weekends as "holidays", was used as a two-dimensional indication.

Here we explain how to create external feature vector. For the bilinear model, we created a time feature vector $t \in \mathbb{R}^{48}$ and an external feature vectors $d \in \mathbb{R}^{16}$ simply by concatenating three external factors. This feature design does not consider the interaction across the external factors. As for the multilinear model, we created a time feature vector t, land attributes feature $d_2 \in \mathbb{R}^{10}$, and $d_3 \in \mathbb{R}^8$, which is the Kronecker product of the weather factor and holiday or weekday factor. Note that the dimensions of d_3 are the product of the dimensions of two factors (2×4).

In this experiment, we used only 14 days worth of data, from July 1, 2013 to July 14, 2013, to keep down memory usage. We used two data patterns; 10 or 20 POIs selected from the 300 POIs as training data and POIs not used in the training data were used as test data.

We used the same three evaluation indices that were used in [7]: mean absolute error (MAE), mean negative log-likelihood (MNLL), and mean absolute peak error (MAPE). When the test data were set to y and when the predictive value was set to λ, $\mathrm{MAE} = \sum_{m=1}^{M} \sum_{s=1}^{S} (|y_{m,s} - \lambda_{m,s}|)/MS$, $\mathrm{MNLL} = \sum_{m=1}^{M} \sum_{s=1}^{S} (-\ln \mathrm{Pois}(y_{m,s}|\lambda_{m,s}))/MS$, $\mathrm{MAPE} = \sum_{m=1}^{M} |\mathrm{Peak}(y_{m,s}) - \mathrm{Peak}(\lambda_{m,s})|/M$, where m is the index of days and s is the index of the time segment in a day. The Peak(y) function calculates the time segment at which y reaches a peak during the day and is defined as $\mathrm{Peak}(y_{m,s}) = \mathrm{argmax}_s(y_{m,s})$. The average mean and standard deviation from the five-fold cross-validation for all POIs were used as the criterion.

Table 1 lists the results for the two models that take into account the difference in external factors. Compared with the existing model with a small sample, the proposed model improved MAE, MNLL, and MAPE by 24.2%, 39.1%, and 40.7%.

Table 1: Comparison of models with land attributes.

train POIs		10 POIs		20 POIs	
model	index	mean	stdev	mean	stdev
bilinear	MAE	216.5	49.1	183.5	5.5
multilinear	MAE	**164.2**	**37.6**	**128.2**	**9.5**
bilinear	MNLL	76.6	29.4	58.8	4.0
multilinear	MNLL	**46.7**	**18.8**	**31.8**	**4.7**
bilinear	MAPE	216.7	21.9	208.4	5.6
multilinear	MAPE	**128.6**	**8.3**	**115.8**	**2.3**

3.3 Evaluation on multilinear vs bilinear

In this experiment, we utilized three external factors: days of the week, holidays or weekdays, and weather. We build these feature

---

[1] https://emg.yahoo.co.jp/

[2] http://www.jma.go.jp/jma/index.html

vectors for the multilinear model: a time feature $t \in \mathbb{R}^{48}$, weather feature $d_2 \in \mathbb{R}^4$, and $d_3 \in \mathbb{R}^{14}$, which is the Kronecker product of the factors on day of the week, and the holiday or weekday.

The following two models were tested for comparison: a low-rank bilinear model (bilinearLR) and bilinear model optimized by nuclear norm minimization using the Frank-Wolf algorithm (bilinearFW). With regards to the feature vector for these models, we created a time feature vector $t \in \mathbb{R}^{48}$ and external feature vector $d \in \mathbb{R}^{56}$, which was the Kronecker product of all external factors. This feature design enabled us to verify whether our multilinear model, whose optimization promotes low-rank for each mode (factor) of a parameter, is useful or not.

We used three patterns of days of training data, i.e., 30, 90, and 180 days, and tested with 180 days for each of the days of the training pattern. We used three indicators: MAE, MNLL, and MAPE. We conducted five-fold cross-validation on each POI for three patterns of training data and compared the average values on all 300 POIs.

Table 2 lists the results for the models considering the product of external factors. Our multilinear model improved MAE, MNLL, and MAPE by 16.7%, 15.6%, and 23.2% against the existing model.

Table 2: Comparison of bilinear and multilinear models.

train data		30 days		90 days		180 days	
model	index	mean	stdev	mean	stdev	mean	stdev
bilinearLR	MAE	78.0	8.6	66.8	11.5	59.8	8.6
bilinearFW	MAE	80.6	7.9	62.9	8.5	55.3	6.0
multilinear	MAE	**64.9**	**6.5**	**57.5**	**6.4**	**53.0**	**3.8**
bilinearLR	MNLL	17.2	4.7	13.9	3.9	12.3	3.6
bilinearFW	MNLL	36.5	12.3	22.4	7.7	17.1	6.8
multilinear	MNLL	**14.5**	**2.1**	**12.3**	**2.3**	**10.9**	**1.5**
bilinearLR	MAPE	109.4	22.9	86.9	20.4	84.1	24.8
bilinearFW	MAPE	101.3	23.3	78.4	15.8	71.9	12.4
multilinear	MAPE	**84.0**	**24.4**	**66.0**	**13.4**	**63.0**	**11.4**

4 DISCUSSION

Creating one predictor in several cities including land attributes. In the first experiment, the multilinear model performed better than the bilinear model in all evaluation indicators. This probably resulted from the fact that external factors were not multiplied together in the bilinear model but were multiplied together in the multilinear model. These results show that by multiplying the factors, we can predict urban dynamics precisely. Multiplication of factors enables the urban dynamics to be expressed in a form that fits each situation rather their than being represented by a linear sum of factors; which is a quite natural result.

Evaluating multilinear model against bilinear models. In the second experiment, the mean of the evaluation value of the multilinear model was the highest among all evaluation indicators and the bilnearFW showed the second performance. When we focus on the results of the bilinearFW and bilinearLR, we can see that the bilnearFW, which employs a convex optimization, showed more stable and better performance than bilinearLR, which employs a non-convex optimization. This result implicitly indicates that our multilinear model with convex optimization is a valid approach compared to that with non-convex one.

Then we compare our multilinear model with bilinear models. As explained in the experimental section, the external feature vector for bilinear models was the Kronecker product of all external factors. This means even the bilinear model in this experiment could

account for interactions of external factors; the difference between our multilinear model and the bilinear models was that the multilinear model imposed low-rank constraints on each feature mode. Therefore, that our multilinear model performed the best indicates that promoting low-rank for each mode of the parameter tensor can resolve the overfitting issue. Note that among the predictive models, ours is the only one to which this approach can be applied.

5 CONCLUSION

In this paper, to enhance the power of the predictive approach to population dynamics with multiple contexts, we devised a novel predictive model with multilinear Poisson regression parameterized by a tensor. Compared with the recent advances in the predictive approach that use bilinear models, ours provides a more systematic way to design features even when multiple contexts are exploited. To tackle the overfitting issues raised by a tensor with large number of parameters, the tensor parameter is regularized with trace norms and optimized via convex matrix optimization, so our model captures multiple contexts and at the same time avoids local optima.

Experimental results using long-term and short-term population scale smartphone location datasets showed that our model outperforms other state-of-the-art models. In contrast to the conventional models of the predictive approach, our model is able to optimize the single multilinear Poisson regression for multiple areas efficiently and it is statistically strong even on short-term datasets. These features allow it to detect sudden changes due to events that occur in cities (e.g. opening of train lines). In order to consider much more contexts, how to improve the high memory cost resulted from introducing a tensor parameter is a topic for our future work.

Acknowledgement

REFERENCES

[1] Zipei Fan, Xuan Song, and Ryosuke Shibasaki. 2014. CitySpectrum: a non-negative tensor factorization approach. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*.
[2] Seung-Taek Park and Wei Chu. 2009. Pairwise preference regression for cold-start recommendation. In *Proceedings of the third ACM conference on Recommender systems*.
[3] R Pulselli, P Ramono, Carlo Ratti, and E Tiezzi. 2008. Computing urban mobile landscapes through monitoring population density based on cellphone chatting. *Int. J. of Design and Nature and Ecodynamics* (2008).
[4] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. 2010. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review* (2010).
[5] Steffen Rendle. 2010. Factorization Machines. In *Proceedings of the 2010 IEEE International Conference on Data Mining*.
[6] Bernardino Romera-Paredes, Hane Aung, Nadia Bianchi-Berthouze, and Massimiliano Pontil. 2013. Multilinear multitask learning. In *Proceedings of the 30th International Conference on Machine Learning*.
[7] Masamichi Shimosaka, Keisuke Maeda, Takeshi Tsukiji, and Kota Tsubouchi. 2015. Forecasting urban dynamics with mobility logs by bilinear Poisson regression. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*.
[8] Marco Signoretto, Lieven De Lathauwer, and Johan AK Suykens. 2010. Nuclear norms for tensors and their use for convex multilinear estimation. *Submitted to Linear Algebra and Its Applications* (2010).
[9] Koh Takeuchi, Ryota Tomioka, Katsuhiko Ishiguro, Akisato Kimura, and Hiroshi Sawada. 2013. Non-negative multiple tensor factorization. In *IEEE 13th International Conference on Data Mining*.