CIVE648 Project: Bayesian Poisson Tensor Factorization for Learning Paratransit Mobility Patterns

Dingyi Zhuang[a], Lijun Sun[a,*]

[a] *Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, Quebec H3A 0C3, Canada*

Abstract

Increasing transportation accessibility lacks equivalent research attention on paratransit for special population, like the disabled. Because understanding mobility patterns for customers is the first step towards a humanized transit service. We suggest unveiling travel behaviors of the disabled via Bayesian Poisson Tensor Factorization (BPTF), a powerful tool to analyze latent patterns of sparse, dispersed and discrete paratransit trip records. Temporal and spatial features including multi-peaks, weekday-weekend indifference and regular trip purpose are discovered. We wish our report can serve as trial to get into the travel behaviors of the special population and invoke more welfare and attention on their transit needs.

Keywords: Urban Computing, Human Mobility, Tensor Factorization

1. Introduction

1.1. Background

As public transport and on-demand mobility services benefit citizens' daily lives these years, how to provide accessible transportation services for special population, e.g. the disabled, lacks equivalent attention. Currently, accessible paratransit transportation offers conventional transportation services such as buses and other public transit provided by local transit organizations (e.g., Toronto Transit Commission (TTC), London Transit Commission, North Bay Transit). Also, it provides license taxis (including both accessible and regular taxis) for more flexible routine (Ontario, 2019). These intercommunicated transportation methods extend the physical limits of the disabled group, further more, provides us opportunities to analyze the travel behaviors of the disabled. Unfortunately, there rises little attention on the mobility analysis or travel behavior researches on the disabled. This might attribute to the dispersion, sparsity and discreteness of paratransit data and the relatively fixed purpose of traveling, e.g. work or study. Thus our study majorly focuses on how to deal with dispersed and sparse count data and discover their interpretable hidden patterns.

*Instructor

Email addresses: dingyi.zhuang@mail.mcgill.ca (Dingyi Zhuang), lijun.sun@mcgill.ca (Lijun Sun)

1.2. Related work

As mentioned above, few works are related to the mobility patterns of paratransit services. However, some surveys well summarize the benefits of analyzing paratransit transportation. Odufuwa (2007) focused on the tendency of the disabled people to be restricted to certain areas, travelling accompanied with relatives that provides assistance where needed and expose to travel difficulties. They revealed that determinants of traveling in company of relatives differs by the extent or severity of the physical disability, travel environment, available modes, distance, purpose of travel and cost of travel and suggested the need to review and retool the operational pattern of public transportation services and the planning as well as implementation of friendly travel environmental policy. Wallace (1997) appeals for the attention on the study of the paratransit customers since 1997. They studied demographic and other characteristics of paratransit customers in southeastern Michigan, and presented the development of a causal model reflecting customer satisfaction with paratransit service. Such models can help researchers and transit operators gauge the potential of improving customer satisfaction through system changes, such as the addition of advanced public transportation systems.

There are many models to unveil the hidden patterns of data, among them tensor factorization is one of the most powerful tool (Kolda and Bader, 2009). Although urban paratransit data is large, sparse and dispersed, tensor factorization can introduces assumption on the data in order to decompose the input tensor into latent factors (Acharya et al., 2015; Matsubara et al., 2012). Pattern recognition can then performed upon the latent factors (Yu et al., 2016; Deng et al., 2016; Yu and Liu, 2016; Harris et al., 2019). Different data types need different factorization schema. Targeting at sparse count data, Bayesian Poisson Tensor Factorization assumes Poisson distribution on the input data, which can perform scalable and interpretable factorization for count inputs (Schein et al., 2015; Cemgil, 2009; Gopalan et al., 2013; Paisley et al., 2014). Thus this model is suitable for our case.

1.3. Motivation

To conclude, we find that mobility patterns of special population lacks attention and Bayesian Poisson Tensor Factorization is a powerful tool to analyze their sparse, dispersed and discrete trip records. Thus we perform BPTF on paratransit trips data to reveal the spatial-temporal mobility patterns of the disabled people.

The remainder of this report is structured as follows. In Section 2, we derive the details of BPTF model. In Section 3, we introduce the paratransit data and perform BPTF to discover the spatial and temporal patterns. And we conclude and discuss in Section 4.

2. Methods

Tensor factorization methods can decompose an observed M-way tensor $\mathbf{Y}$ into M latent factor matrices that provide low-dimension representations of origin tensor. As Kolda and Bader (2009) mentioned, the most common decomposition methods are Tucker decomposition (Tucker, 1966) and the Canonical Polyadic (CP) decomposition (Harshman et al., 1970). However, they are not designed for sparse count data as the sparsity and discrete features make it harder to reconstruct original tensor with mean square error loss. To address this issue, we follow the work of Schein et al. (2015) by using Tensor Bayesian Tensor Factorization. My work here is to derive the model using the variational inference myself, as the authors did not provide the detailed derivation.

2.1. Notations

We introduce some notations which will be used later:

- Let observation $\mathbf{Y}$ be a three-way positive count tensor of size $O \times D \times T$, where O, D, T are the dimensions for origin, destination and timestamp.
- Let lower case y_{ijt} stands for the element in $\mathbf{Y}$, where i, j, t are the specific origin, destination and timestamp respectively.
- Let K be the latent dimension of latent factor

2.2. Bayesian Poisson Tensor Factorization

The most import assumption of BPTF model is drawing Poisson distribution on the observed count tensor $\mathbf{Y}$. And then tensor decomposition along with variational inference are applied to fit the parameters of Poisson distribution. We follow the same three steps.

Firstly, we assume

$$y_{ijt} \sim Poisson(\lambda_{ijt}) \tag{1}$$

The CP factorization formula treats each observed count y_{ijt} as

$$y_{ijt} \approx \hat{y}_{ijt} \equiv \sum_{k=1}^{K} \theta_{ik}^{(1)} \theta_{jk}^{(2)} \theta_{tk}^{(3)} \approx \lambda_{ijt} \tag{2}$$

for $i \in [O]$, $j \in [D]$, and $t \in [T]$ where the $\theta_{ik}^{(1)}, \theta_{jk}^{(2)}, \theta_{tk}^{(3)}$ are the corresponding latent factors and $\hat{y}_{ijt}$ is the reconstruction of y_{ijt}. The set of all factors used to model $\mathbf{Y}$ can be aggregated into three latent factor matrices, for example, $\Theta^{(1)} \equiv ((\theta_{ik}^{(1)})_{i=1}^{N})_{k=1}^{K}$.

In the probabilistic way, we can approximate the reconstruction with the parameter λ_{ijt} of Poisson distribution, which means

$$y_{ijt} \sim Poisson(\sum_{k=1}^{K} \theta_{ik}^{(1)} \theta_{jk}^{(2)} \theta_{tk}^{(3)}) \tag{3}$$

Then the process of decomposing $\mathbf{Y}$ into its latent factor matrices is known as Poisson Tensor Factorization (PTF), and can be achieved via maximum likelihood estimation (MLE) of $\Theta^{(1)}, \Theta^{(2)}, \Theta^{(3)}$. PTF often yields better latent factor estimation for sparse count data than those obtained by assuming the count is drawn from Gaussian distribution (Cemgil, 2009; Chi and Kolda, 2012; Schein et al., 2015).

In BPTF, we follow the most assumption of PTF but we impose prior distributions on the latent factors and perform Bayesian inference to obtain the point estimate, rather than MLE schema. Since the Gamma distribution is the conjugate prior for a Poisson likelihood, BPTF typically imposes Gamma priors on the latent factors. Here three latent factors have their own Gamma priors, e.g. $\theta_{ik}^{(1)}$,

$$\theta_{ik}^{(1)} \sim Gamma(\alpha, \alpha\beta^{(1)}) \tag{4}$$

The Gamma distribution is parameterized by a shape parameter $\alpha > 0$ and a rate parameter $\alpha\beta > 0$. The expectation of $\theta_{ik}^{(1)}$ should be $\mathbb{E}[\theta_{ik}^{(1)}] = \frac{\alpha}{\alpha\beta^{(1)}} = \frac{1}{\beta^{(1)}}$ (algorithmic expectation) and $\mathbb{G}[\theta_{ik}^{(1)}] = \mathbb{E}[ln(\theta_{ik}^{(1)})] = \frac{exp(\Psi(\alpha))}{\alpha\beta^{(1)}}$ (geometric expectation), with $\Psi(\cdot)$ as the digamma function. Thus, when shape parameter $a \ll 1$ and rata parameter b is small, the Gamma distribution concentrates most of its mass near zero yet maintains a heavy tail and can therefore be used as a sparsity-included prior (Cemgil, 2009; Gopalan et al., 2013).

With PTF and Bayesian priors introduced, we continue on Bayesian inference in BPTF. Given the definition of Equation 4, we can draw the similar Gamma distribution for $\theta_{jk}^{(2)}$ and $\theta_{tk}^{(3)}$. It can be known that their prior expectation is completely determined by $\beta^{(1)}, \beta^{(2)}$ and $\beta^{(3)}$, which can be inferred from the data (Cemgil, 2009; Liang et al., 2014). The shape parameter α here detremines the sparsity of latent factor matrices, we follow the work of Schein et al. (2015) by setting $\alpha = 0.1$ to encourages sparsity and hence promote the interpretability of the factors.

2.3. Variational Update

The target of BPTF is to reconstruct the parameter of Poisson parameter λ_{ijt} instead of direct MLE point estimate (although we try to minimize their difference), which means Bayesian inference should be applied to update related parameters. Given an observed tensor $\mathbf{Y}$ and the model hyperparameter set $\mathcal{H} \equiv \{\alpha, \beta^{(1)}, \beta^{(2)}, \beta^{(3)}\}$, our posterior is:

$$P(\Theta^{(1)}, \Theta^{(2)}, \Theta^{(3)}|\mathbf{Y}, \mathcal{H})$$

However, the posterior distribution for BPTF is analytically intractable and has to be approximated. While variational inference turns distribution approximation into optimization algorithm. It first introduces a parametric family of distributions Q over the latent variables $\{\theta^{(1)}, \theta^{(2)}, \theta^{(3)}\}$, indexed by the values of a set of *variational parameters* S. The function of Q is to facilitate the optimization of S. Here we use a fully factorized *mean-field approximation* where we define every element in $\{\theta^{(1)}, \theta^{(2)}, \theta^{(3)}\}$ follows identical and independent distribution Q and their product forms $O \times D \times T \times K$ independent Gamma distributions. For example, consider $\theta_{ik}^{(1)}$:

$$\theta_{ik}^{(1)} \sim Q(S_{ik}^{(1)}) = Gamma(\gamma_{ik}^{(1)}, \delta_{ik}^{(1)}) \tag{5}$$

Where $S^{(1)} \equiv ((\gamma_{ik}^{(1)}, \delta_{ik}^{(1)})_{i=1}^{O})_{k=1}^{K}$. The full set of variational parameters is thus $S \equiv \{S^{(1)}, S^{(2)}, S^{(3)}\}$. The form of Q is similar to that in Bayesian Poisson Matrix Factorization (Cemgil, 2009; Paisley et al., 2014; Gopalan et al., 2013). It might seems confusing to introduce the Gamma distribution again on latent factors, e.g. $\theta_{ik}^{(1)}$, but with different parameters $\gamma_{ik}^{(1)}, \delta_{ik}^{(1)}$ rather than α and $\beta^{(1)}$. This is because $\alpha, \beta^{(1)}$ are the hyperparameters we want to learn from the data, but we use variational parameters $\gamma_{ik}^{(1)}, \delta_{ik}^{(1)}$ to approximate.

Thus, the variational parameters are then fit so as to yield the closest member of Q to extract posterior, known as the *variational distribution*. Specifically, the algorithm sets the values of S to those that minimize the KL divergence of the exact posterior from Q. It can be shown that these values are the same as those that maximize a lower bound on $P(\mathbf{Y}|\mathcal{H})$, known as the *evidence lower bound* (ELBO):

$$\mathcal{B}(S) = \mathbb{E}_Q[ln(P(\mathbf{Y}, \Theta^{(1)}, \Theta^{(2)}, \Theta^{(3)}|\mathcal{H}))] + H(Q) \tag{6}$$

Where H(Q) is the entropy of Q. When Q is a fully factorized approximation, finding values S that maximize the ELBO in Equation 6 can be achieved by performing coordinate ascent, iteratively updating each variational parameter, while holding the others fixed, until convergence (defined by relative change in ELBO). The update equation for each parameter can be derived using an auxiliary variable with multinomial distribution (Cemgil, 2009; Paisley et al., 2014; Gopalan et al., 2013):

$$\begin{gathered}\text{Set } \lambda = \textstyle\sum_i \sum_j \sum_t \lambda_{ijt}, x \sim Poisson(\lambda) \\ (x_{111}, ..., x_{ijt})|x \sim Multinomial(\tfrac{\lambda_{111}}{\lambda}, ..., \tfrac{\lambda_{ijt}}{\lambda}) \\ \text{Distribution of } x_{ijt} \equiv \text{Distribution of } y_{ijt}\end{gathered}$$

According to the derivation by Kurt (2018), we can obtain the update equation of variational parameters using the auxiliary multinomial distribution:

$$\gamma_{ik}^{(1)} = \alpha + \sum_{i,j,t} y_{ijt} \frac{\mathbb{G}_Q[\theta_{ik}^{(1)}\theta_{jk}^{(2)}\theta_{tk}^{(3)}]}{\sum_{k=1}^{K} \mathbb{G}_Q[\theta_{ik}^{(1)}\theta_{jk}^{(2)}\theta_{tk}^{(3)}]} \tag{7}$$

$$\delta_{ik}^{(1)} = \alpha\beta^{(1)} + \sum_{j,t} \mathbb{E}_Q[\theta_{jk}^{(2)}\theta_{tk}^{(3)}] \tag{8}$$

Since Q is fully factorized, each expectation of a product can be factorized into a product of individual expectations, which, e.g. for $\theta_{ik}^{(1)}$ are

$$\mathbb{E}_Q[\theta_{ik}^{(1)}] = \frac{\gamma_{ik}^{(1)}}{\delta_{ik}^{(1)}} \qquad \mathbb{G}_Q[\theta_{ik}^{(1)}] = \frac{exp(\Psi(\gamma_{ik}^{(1)}))}{\delta_{ik}^{(1)}} \tag{9}$$

Each expectation can be cached to improve efficiency. Note that the summand in Equation 8 need only be computed for those values of j and t for which $y_{ijt} > 0$; observed $\mathbf{Y}$ is very sparse, inference is efficient even for very large tensors.

The hyperparameters $\beta^{(1)}, \beta^{(2)}$ and $\beta^{(3)}$ can be optimized via an empirical Bayesian inference method, in which each hyperparameter is iteratively updated along with the variational parameters according to the following update equation:

$$\beta^{(1)} = \frac{1}{Mean(\mathbb{E}_Q[\theta_{ik}^{(1)}])} \tag{10}$$

Update equations (7), (8) and (10) completely specify the variational inference algorithm for BPTF.

3. Results

We implement BPTF model on paratransit records of the disabled provided by TTC, which is a dispersed sparse dataset. Based on that, we discover mobility patterns to better understand their travel behavior from spatial and temporal aspects.

3.1. Data description

The paratransit data contains two-year paratransit trip records in Toronto from May 16th, 2017 to May 16th, 2019. They are collected by *HASTUS-OnDemand* program by TTC which offers both fixed-routine and on-demand mobility services and real-time scheduling. There are 9966487 paratransit trips provided, along with many detailed labels describing the paratransit services. Since we focus on the spatial-temporal mobility patterns, only the origins, destinations and timestamps are extracted. It leaves a problem about how to choose the research objects, whether the individual mobility patterns or the regional Origin-Destination (O-D) mobility patterns. Because paratransit data majorly record the trips of regular users, the patterns might be fixed if we analyze the individual-level mobility. Therefore, we take regions as our objects.

Since TTC already divided Toronto city into more than 300K grids to depict the trips, we take these grids along with their numbering as our origins and destinations index. However, if we directly construct an O-D matrix with 300K rows and 300K columns, the computation will be intractable as it requires at least $300,000 \times 300,000 \times 4Byte = 360GB$ memory. What's more, the variance-to-mean ratio for two years records reaches 10.53, which means the trip data are extremely dispersed and sparse. To make the results more interpretable, we only use the most popular 100 grids and the O-D flow among them as our spatial information and divide temporal resolutions into 15min, 30min and 60min to test the performance of BPTF. Therefore, our input to the model will be a $100 \times 100 \times T$ tensor recording the number of trips, where T dimension depends on different temporal resolutions.

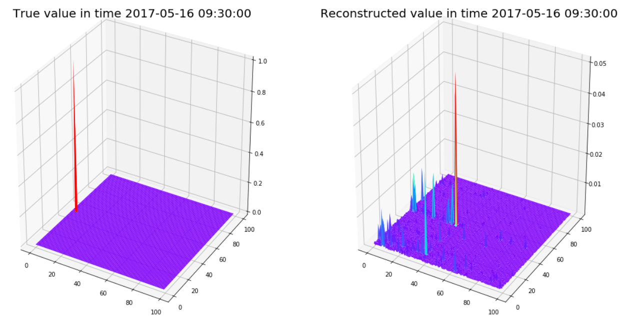
3.2. Temporal patterns

We firstly compare the selection of different time resolutions, then discover the temporal patterns of peak time and non-peak time under given time resolution and finally compare the weekday and weekend mobility difference.

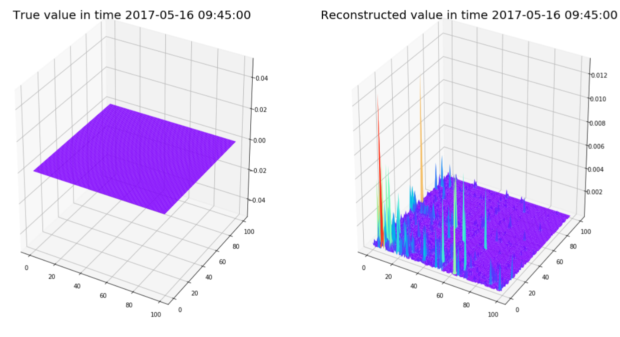
As shown in Table 1, 15min resolution can converge fast and achieve a higher Evidence Lower Bound. However, with finer granularity, the runtime for each iteration also increases. After comparison, we choose 15min as our temporal resolution to discover the temporal patterns.

	15min	30min	1h
Iterations	**30**	49	67
ELBO	**3883131.64**	1735944.21	3746764.34
Time/iter	0.58s	**0.36s**	0.54s

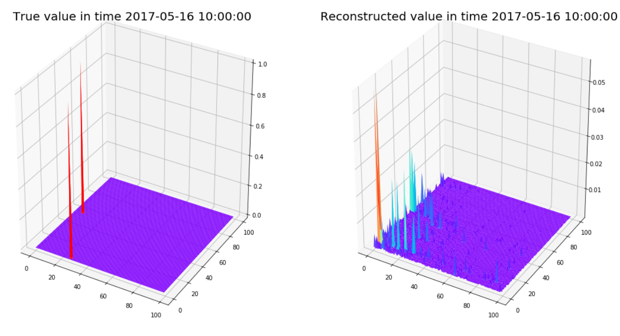
Table 1: Temporal resolution comparison



i 9:30 am



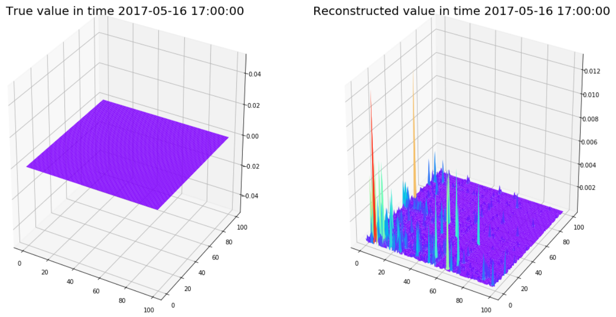
ii 9:45 am



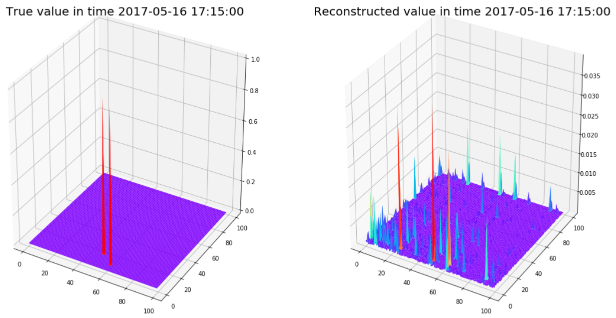
iii 10:00 am

Figure 1: Morning peak temporal patterns

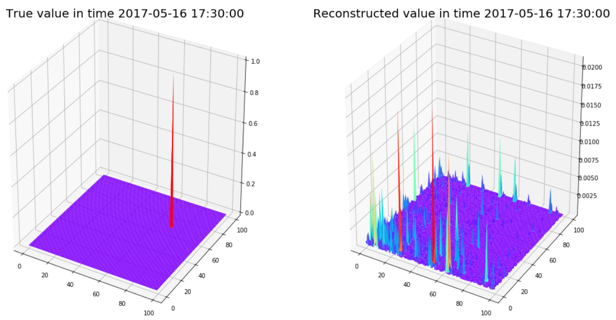
We then analyze the temporal patterns of peak time and non-peak time. Without loss of generality, we choose the May 16th, 2017—the first recorded day—as the observation and analyze the fluctuation of trips.



i 17:00 pm

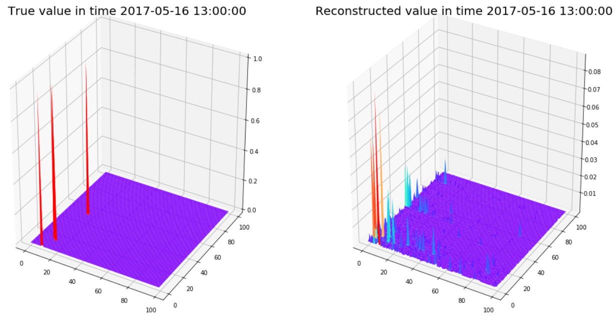


ii 17:15 pm

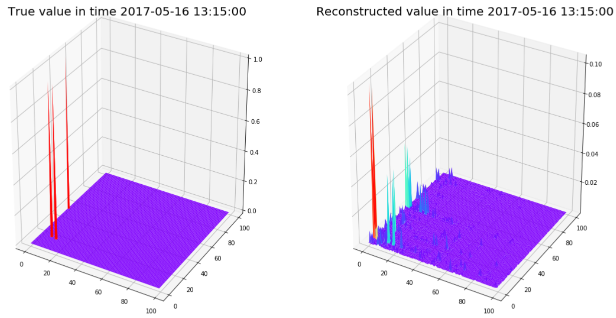


iii 17:30 pm

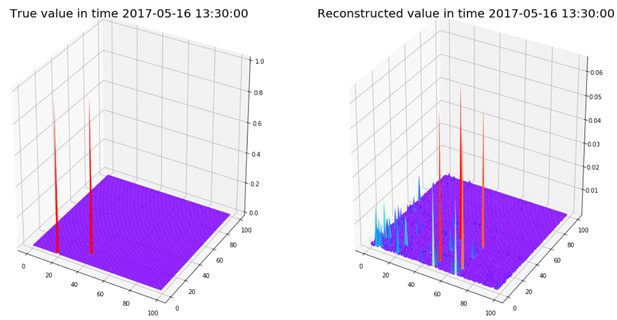
Figure 2: Evening peak temporal patterns



i 13:00 pm



ii 13:15 pm



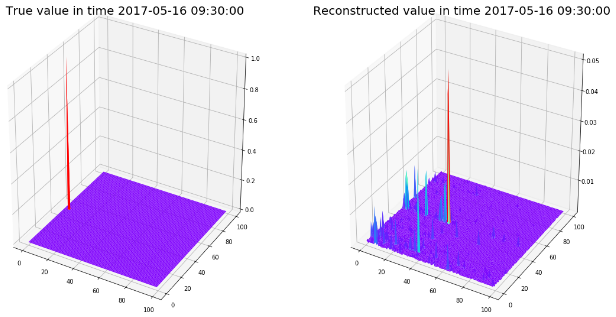
iii 13:30 pm

Figure 3: Non-peak period temporal patterns

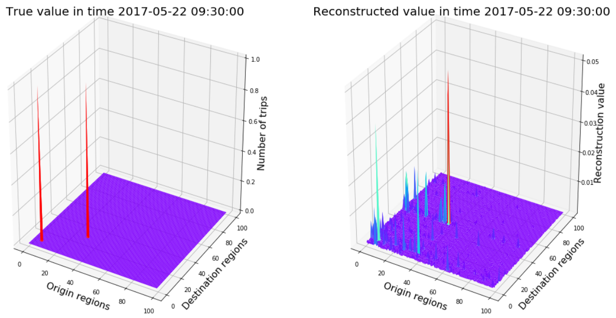
We choose 9:30 10:00 and 17:00 17:30 as morning peaks and evening peaks periods, and select 13:00 13:30 as non-peak period.

As shown in Figure 1 and Figure 2, trips appears less than our expectation during so-called 'peaks' periods, even less than the some of non-peak periods, e.g. Figure 2i and Figure 3i. This means that even though 'peaks' do exist, they might not form sharp morning and evening peaks, instead, there might exist other peaks like 'noon-peaks'. Those peaks can explain different lifestyles of the disabled. They might work and study in regularly like others, but might have some other activities, e.g. treatment, during the lunch time to cause the 'noon-peaks'.

The power of BPTF is also unveiled, where we can infer the trip possibility for some time periods even without true trips as Figure 1ii and Figure 2i . This can illustrate the scalability and robustness of BPTF model for such sparse data. However, the reconstruction values are the λ_{ijt} in Equation 1 that we hope to approximate to the true trip records, which are more than 10 times less than the true ones. These blames to the sparsity of data because even we choose the most popular 100 regions, only 106794 records are fed into the model. This means only 10 trips are recorded among 100 regions every 15min on average. If we use the original



i Tuesday temporal pattern



ii Sunday temporal pattern

Figure 4: Weekday and weekend temporal patterns

100,000 regions as input, the result will get even worse. This can somehow explain why previous researchers are not interested in the mobility patterns of the disabled, where their traveling records do not contain enough information for interesting discovery.

We then analyze the temporal patterns change in a daily basis. From Figure 4 we can find that The temporal patterns fluctuate little between weekdays and weekends. This attributes to the low traffic volume of the customers, leading to similar trip requests in weekdays and weekends.

3.3. Popular regions analysis

Besides temporal patterns, we try to look into what information we can obtain from spatial aspect. Our idea is simple, we find the most 10 popular regions that occur most frequently in the trip records. For anonymous reason, we do not post any information related to the customers, like the specific postal codes or coordinates of the regions. Because the inflow (regions as trip destinations) and outflow (regions as trip origins) output the same 10 regions, here we use the outflow number as trip number to sort 10 regions, as Table 2 shows.

Shown in Table 2, most popular regions are all about residence, medical services and shops (collected from

Region index	Trip number	Land use
51361	121564	Residence
2	93624	Residence
520	91260	Restaurants and shops
106	77554	Residence
224181	70648	Medical services
3173	67017	Medical services
768	47721	Shops
2510	47423	Residence
4980	47188	Medical services
386	44406	Medical services

Table 2: Land use information of 10 most popular regions

Google Maps). This reflects that most the disabled people have relatively regular travel purpose, they use the paratransit services to transit between home and medical institutions. Since the frequency to transit to/from restaurants and shops is so high, even higher than some of returning to residence records, it is reasonable to assume these restaurants and shops are their workplaces. Thus the spatial patterns of the disabled are the same as the non-disabled people whose lives are also simply companies and home. To verify, more case studies need to be conducted and surveys to be performed.

4. Discussion

In this report, motivated by the lack of research on special population mobility, we implemented Bayesian Poisson Tensor Factorization on TTC paratransit data to discover their spatial-temporal patterns. For temporal patterns, we discover that 'peaks' are not strictly defined in the morning and evening, instead, other time periods like noon and afternoon will also form peaks. Besides, weekday and weekends show no obvious difference, which attributes to low traffic volume of the customers. For spatial patterns, we find that most popular regions are related to residence, work and medical services, which means the traveling behaviors are the same as the non-disabled ones. Therefore, previous researches about public transport mobility patterns, on-demand mobility services and individual travel behaviors can be transplanted to analyze paratransit mobility. However, to be noticed that the sparse and dispersed features of paratransit trip records—there may be just one record in an hour—makes the hidden patterns hard to unveil and interpret. Our work offers a trial on a marginal research field. We wish there can be more attention on the researches of the disabled and how their transits can be improved.

We would like to thank Toronto Transit Commission to offer the data and Prof. Lijun Sun's inspiring opinions and lectures on conducting the researches.

References

Acharya, A., Ghosh, J., Zhou, M., 2015. Nonparametric bayesian factor analysis for dynamic count matrices. arXiv preprint arXiv:1512.08996 .

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

Chi, E.C., Kolda, T.G., 2012. On tensors, sparsity, and nonnegative factorizations. SIAM Journal on Matrix Analysis and Applications 33, 1272–1299.

Deng, D., Shahabi, C., Demiryurek, U., Zhu, L., Yu, R., Liu, Y., 2016. Latent space model for road networks to predict time-varying traffic, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM. pp. 1525–1534.

Gopalan, P., Hofman, J.M., Blei, D.M., 2013. Scalable recommendation with poisson factorization. arXiv preprint arXiv:1311.1704 .

Harris, K.D., Aravkin, A., Rao, R., Brunton, B.W., 2019. Time-varying autoregression with low rank tensors. arXiv preprint arXiv:1905.08389 .

Harshman, R.A., et al., 1970. Foundations of the parafac procedure: Models and conditions for an" explanatory" multimodal factor analysis .

Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. SIAM review 51, 455–500.

Kurt, B., 2018. Bayesian poison tensor factorization for dummies. https://github.com/bariskurt/bptf. Accessed December 17, 2019.

Liang, D., Paisley, J.W., Ellis, D., et al., 2014. Codebook-based scalable music tagging with poisson matrix factorization., in: ISMIR, pp. 167–172.

Matsubara, Y., Sakurai, Y., Faloutsos, C., Iwata, T., Yoshikawa, M., 2012. Fast mining and forecasting of complex time-stamped events, in: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM. pp. 271–279.

Odufuwa, B.O., 2007. Towards sustainable public transport for disabled people in nigerian cities. Studies on Home and Community Science 1, 93–101.

Ontario, 2019. How to provide accessible transportation services. https://www.ontario.ca/page/how-provide-accessible-transportation-services. Accessed December 17, 2019.

Paisley, J.W., Blei, D.M., Jordan, M.I., 2014. Bayesian nonnegative matrix factorization with stochastic variational inference.

Schein, A., Paisley, J., Blei, D.M., Wallach, H., 2015. Bayesian poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM. pp. 1045–1054.

Tucker, L.R., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika 31, 279–311.

Wallace, R.R., 1997. Part 2: Paratransit: Paratransit customer: Modeling elements of satisfaction with service. Transportation Research Record 1571, 57–66.

Yu, H.F., Rao, N., Dhillon, I.S., 2016. Temporal regularized matrix factorization for high-dimensional time series prediction, in: Advances in neural information processing systems, pp. 847–855.

Yu, R., Liu, Y., 2016. Learning from multiway data: Simple and efficient tensor regression, in: International Conference on Machine Learning, pp. 373–381.