

Mobility Pattern Discovery from Automobile Traces with A Temporal-Spatial Model

Yuxi Zhang, Jiawei Hu, Ya Zhang*

Big Data and Business Innovation Laboratory
Shanghai Jiao Tong University, Shanghai 200240, China
{snow_sword, jiawei.hu, ya.zhang}@sjtu.edu.cn

Abstract

An increasing number of automobiles are nowadays equipped with GPS-enabled devices, which enables the collection of their moving traces in real-time at large scale. Hidden behind the numerous and disordered traveling traces are some inherent mobility patterns of the corresponding regions. However, it is hard to reveal the patterns directly from the traces due to the dynamics of user mobility. On the one hand, the mobility patterns differ from time to time and from region to region. On the other hand, drivers of different types may have distinct driving habits. In this paper, attempting to discovering the mobility patterns in both temporal and spatial dimensions, we extend the Latent Dirichlet Allocation model and propose a novel probabilistic model called Temporal-Spatial Latent Dirichlet Allocation (TS-LDA). We validate the model with a real-world GPS trajectory data set. The experimental results have revealed interesting mobility patterns.

Introduction

Mobility have become essential for urban lives in modern society due to continuous expanding of urban population density. With the increasing number of GPS-equipped automobiles and mobile devices, collecting moving traces of human beings at large scale in real-time becomes possible. It opens up great opportunities to study human mobility. As a result, analysis of the massive GPS trajectories is gaining increasing popularity in recent years, with a diverse set of goals such as mobility pattern discovering, location prediction, and region clustering.

In this paper, we focus our analysis on discovering inherent mobility patterns of individuals hidden behind the GPS trajectories. Individuals seldom move randomly and usually with a clear goal. Thus, each trip is usually associated with a pre-defined destination, such as office, home or shopping center. For people who live regular lives, there must be some inherent mobility patterns. For example, people who living the suburb may probably drive to the downtown in the morning and return home in the late afternoon or evening every weekday. The purpose of this study is to discovering

the *semantic* time intervals and regions with similar trajectory patterns, and furthermore reveal such inherent mobility patterns hidden behind the GPS trajectories, including finding out the traffic flow tendency in different time intervals and classifying all the drivers into different types according to their driving habits. Outcome of this analysis is expected to benefit several parties of urban lives. It offers supporting references to the government for optimizing urban infrastructure and transportation rules to improve urban mobility. For urban citizens, it can not only help them make more informed driving plans, but also provide personalized driving assistant based on their driving habits learned from their past driving traces. Advertisers can leverage the mobility patterns for behavior targeting and deliver personalized advertisement to individuals.

To our best knowledge, there have been limited researches on this topic. Geo topic model was proposed for location recommendation based on human activity and interests (Kurashima et al. 2013), while this model did not take the temporal factor into consideration, which is an influential component of human mobility. Giannotti et al. proposed a clustering method to analyze the mobility behavior (Giannotti et al. 2009). Although this study used both temporal and spatial features to describe trajectories, they ignored the *semantic* meanings of time and positions. For instances, 7 a.m. to 9 a.m. and 4 p.m. to 6 p.m. are usually considered morning traffic peak and afternoon traffic peak, respectively. The daytime in weekends is considered as a different temporal type than that of weekdays. Similarly, regions also demonstrate semantic patterns such as city center and suburb. Different from the previous researches, we aim to establish a thematic structure over time and positions separately and mine the hidden temporal and spatial patterns, which can be furthermore applied to describe human mobility patterns.

Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003) was proved successful in mining the hidden topics in large archives of documents, and its applications have extended far beyond the natural language processing. A DMR-based topic model (Mimno and McCallum 2012) was applied to discovering urban region functions combining human mobility data and POIs (Yuan, Zheng, and Xie 2012). However, they regarded the combination of time and position as a whole, rather than model them on separate topics.

A coupled-LDA model was proposed to extracting separate topics on multi-dimensional features in mining IPTV user behaviors (Chen, Zhang, and Zha 2013). This model does not fit our problem well due to the difference in human behavior descriptions. In Chen’s work, the observed features, program and timestamp, belong to two independent topics, program interest and temporal factor, separately. However, in our case, both the origin and the destination in a trip follow the same topic distribution. Chen et al. did not take this situation into their consideration. In this paper, we extend the LDA model and proposed a Temporal-Spatial model for human mobility pattern analysis. The model considers both temporal and spatial features and establishes high-level thematic structures separately for the two set of features. Based on this model, we reveal the hidden temporal and spatial patterns. We further analyze human mobility patterns based on the combination of temporal and spatial patterns, and categorize drivers into different types. Finally, we show that the model is able to capture the regular transition patterns of each type of drivers given a specified time interval.

TS-LDA Model

In natural language processing and text mining, the LDA model has been proved effective for mining the unobserved thematic structure in large archives of documents. There are 3 key concepts in LDA model: document, word and topic. When it comes to driving mobility problem, we need to correspondingly define these key concepts. Yuan et al. (Yuan, Zheng, and Xie 2012) regard the geographic regions as documents and the region functions as the hidden topic. In addition, they use a triple (origin, destination and leaving time) to characterize a transition in a certain region, and regard it as the word in the document. In our study, we use the same way to characterize the transitions. But the differences lie in that we treat the users as the documents and the driving patterns as the hidden topics. Therefore, the driving transition is the *word* of users. Table 1 shows the analogy relationships among these three works.

Text analysis	Yuan’s work	Our work
Documents	Regions	Users
Topics	Region functions	Driving patterns
Words	Region transitions	Driving transitions

Table 1: Analogy relationship among text analysis, Yuan’s work and our work.

In the original LDA model, positions and time are treated as a whole, and thematic structures are built on the position-time couple. However, this method ignores the intrinsic semantics of temporal and spatial features. In our case, we use a triple (origin, destination, leaving time) to characterize a piece of driving trajectory. We attempt to separately build two high-level thematic structures – spatial pattern and temporal pattern. Figure 1 shows the graphical model of TS-LDA. Table 2 lists the annotations used in this model.

In the TS-LDA model, a driving transition is generated by four steps:

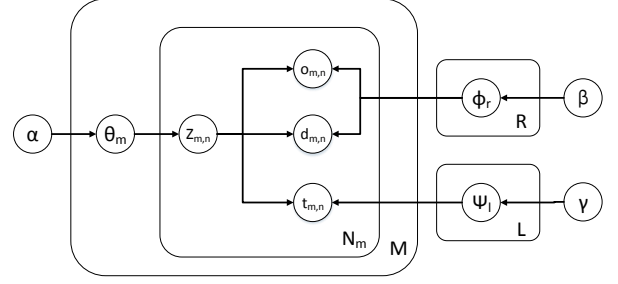


Figure 1: The Graphical Model of TS-LDA

1. Choose $\theta_m \sim \text{Dir}(\alpha)$, where $m \in \{1, \dots, M\}$;
2. Choose $\phi_r \sim \text{Dir}(\beta)$, where $r \in \{1, \dots, R\}$;
3. Choose $\psi_l \sim \text{Dir}(\gamma)$, where $l \in \{1, \dots, L\}$;
4. For each user m , where $m \in \{1, \dots, M\}$:
 - (a) Choose a driving mobility pattern
 $z_{m,n} = (z_{m,n,1}, z_{m,n,2}, z_{m,n,3}) \sim \text{Multinomial}(\theta_m)$;
 - (b) Choose an origin position
 $o_{m,n} \sim \text{Multinomial}(\phi_{z_{m,n,1}})$;
 - (c) Choose a destination $o_{m,n} \sim \text{Multinomial}(\phi_{z_{m,n,2}})$;
 - (d) Choose a leaving timestamp
 $t_{m,n} \sim \text{Multinomial}(\psi_{z_{m,n,3}})$;

Our purpose is to minimize the log likelihood:

$$\begin{aligned}
LL(\theta, \phi, \psi; \alpha, \beta, \gamma) &= -\log P(O, D, T | \theta, \phi, \psi; \alpha, \beta, \gamma) \\
&= -\log \prod_{m=1}^M \prod_{n=1}^{N_m} \sum_{z_{m,n}} P(o_{m,n}, d_{m,n}, t_{m,n}, z_{m,n} | \theta, \phi, \psi; \alpha, \beta, \gamma) \\
&= -\log \prod_{m,n} \sum_{z_{m,n}} P(z_{m,n} | \theta_m) \cdot P(o_{m,n} | \phi_{z_{m,n,1}}) \\
&\quad \cdot P(d_{m,n} | \phi_{z_{m,n,2}}) \cdot P(t_{m,n} | \psi_{z_{m,n,3}}) \\
&= -\sum_{m=1}^M \sum_{n=1}^{N_m} \log \sum_{k=1}^K \theta_m(k) \cdot \phi_{k_1}(o_{m,n}) \cdot \phi_{k_2}(d_{m,n}) \cdot \phi_{k_3}(t_{m,n}).
\end{aligned} \tag{1}$$

In order to learn the various distributions, we perform approximate inference using Gibbs sampling for our model. The following part shows some key steps in the inference. The total probability may be expressed as follows:

$$\begin{aligned}
P(O, D, T, Z, \theta, \phi, \psi; \alpha, \beta, \gamma) &= P(O, D | \phi, Z) P(T | \psi, Z) P(Z | \theta) P(\theta; \alpha) P(\phi; \beta) P(\psi; \gamma).
\end{aligned} \tag{2}$$

Ingrate out θ, ϕ, ψ , we obtain:

$$\begin{aligned}
P(O, D, T, Z; \alpha, \beta, \gamma) &= \int_{\theta} P(\theta; \alpha) P(Z | \theta) d\theta \\
&\quad \cdot \int_{\phi} P(\phi; \beta) P(O, D | \phi, Z) d\phi \cdot \int_{\psi} P(\psi; \gamma) P(T | \psi, Z) d\psi
\end{aligned}$$

Symbol	Description	Type
M	number of documents	integer
R	number of region patterns	integer
L	number of time patterns	integer
K	$= R^2 L$, number of driving mobility patterns	integer
V_r	number of regions	integer
V_t	number of timestamps	integer
N_m	number of transitions of user m	integer
θ_m	the multinomial distribution of driving mobility patterns for user m	K -dim vector
ϕ_r	the multinomial distribution of region pattern r	V_r -dim vector
ψ_l	the multinomial distribution of time pattern l	V_t -dim vector
$Z = (z_{m,n})$	identity of driving mobility patterns of driving transitions n in user m	$[1, K]$
$O = (o_{m,n})$	identity of the origin position of driving transitions n in user m	$[1, V_r]$
$D = (d_{m,n})$	identity of the destination of driving transitions n in user m	$[1, V_r]$
$T = (t_{m,n})$	identity of the leaving timestamp of driving transitions n in user m	$[1, V_t]$
α, β, γ	Dirichlet priors to the multinomial distribution θ_m, ϕ_r, ψ_l respectively	vector
$n_u(m, k)$	number of driving transitions of user m assigned to driving mobility pattern k	integer
$n_r(r, v_r)$	number of position (origin or destination) v_r assigned to region pattern r	integer
$n_t(l, v_t)$	number of timestamp v_t assigned to time pattern l	integer
$n_{u,r,t}^{-(m,n)}$	the same meaning as n_u, n_r, n_t but with the n th driving transitions of user m excluded.	integer

Table 2: Annotations used in TS-LDA model.

$$\begin{aligned}
&= \left(\prod_{j=1}^M \frac{\Gamma(\sum_{i=1}^K \alpha_i)}{\prod_{i=1}^K \Gamma(\alpha_i)} \frac{\prod_{i=1}^K \Gamma(n_u(i, j) + \alpha_i)}{\Gamma(\sum_{i=1}^K (n_u(i, j) + \alpha_i))} \right) \\
&\cdot \left(\prod_{i=1}^R \frac{\Gamma(\sum_{r=1}^{V_r} \beta_r)}{\prod_{r=1}^{V_r} \Gamma(\beta_r)} \frac{\prod_{r=1}^{V_r} \Gamma(n_r(i, r) + \beta_r)}{\Gamma(\sum_{r=1}^{V_r} (n_r(i, r) + \beta_r))} \right) \\
&\cdot \left(\prod_{i=1}^L \frac{\Gamma(\sum_{t=1}^{V_t} \gamma_t)}{\prod_{t=1}^{V_t} \Gamma(\gamma_t)} \frac{\prod_{t=1}^{V_t} \Gamma(n_t(i, t) + \gamma_t)}{\Gamma(\sum_{t=1}^{V_t} (n_t(i, t) + \gamma_t))} \right). \quad (3)
\end{aligned}$$

In a single sampling iteration, we sample an assignment $z_{m,n}$ for a driving transition $(o_{m,n}, d_{m,n}, t_{m,n})$, conditioned on other assignments $Z^{-(m,n)}$ for all driving transitions except for $(o_{m,n}, d_{m,n}, t_{m,n})$. Assuming $(o_{m,n}, d_{m,n}, t_{m,n}) = (w_1, w_2, w_3)$, the conditional distribution may be expressed as:

$$\begin{aligned}
&P(z_{m,n} = k | Z^{-(m,n)}, O, D, T; \alpha, \beta, \gamma) \\
&\propto (n_u^{-(m,n)}(m, k) + \alpha_k) \frac{n_r^{-(m,n)}(k_1, w_1) + \beta_{w_1}}{\sum_{v_r=1}^{V_r} (n_r^{-(m,n)}(k_1, v_r) + \beta_{v_r})} \\
&\cdot \frac{n_r^{-(m,n)}(k_2, w_2) + \beta_{w_2} + c_r}{\sum_{v_r=1}^{V_r} (n_r^{-(m,n)}(k_2, v_r) + \beta_{v_r}) + c_k} \\
&\cdot \frac{n_t^{-(m,n)}(k_3, w_3) + \gamma_{w_3}}{\sum_{v_t=1}^{V_t} (n_t^{-(m,n)}(k_3, v_t) + \gamma_{v_t})}, \quad (4)
\end{aligned}$$

where

$$\begin{aligned}
k &= (k_1, k_2, k_3) \\
c_r &= \begin{cases} 1 & \text{, if } k_1 = k_2 \text{ and } w_1 = w_2; \\ 0 & \text{, otherwise.} \end{cases} \\
c_k &= \begin{cases} 1 & \text{, if } k_1 = k_2; \\ 0 & \text{, if } k_1 \neq k_2. \end{cases}
\end{aligned}$$

Using Bayes rule, we can obtain the posterior distribution of the multinomial parameters θ_m, ϕ_r and ψ_l :

$$\begin{aligned}
P(\theta_m | Z, \alpha) &= \text{Dirichlet}(\mathbf{n}_u(m, \cdot) + \alpha) \\
P(\phi_r | Z, \beta) &= \text{Dirichlet}(\mathbf{n}_r(r, \cdot) + \beta) \\
P(\psi_l | Z, \gamma) &= \text{Dirichlet}(\mathbf{n}_t(l, \cdot) + \gamma)
\end{aligned} \quad (5)$$

Therefore, using the expectation of the Dirichlet distribution, ϕ_r and ψ_l can be derived as follows.

$$\begin{aligned}
\theta_m(k) &= \frac{n_u(m, k) + \alpha_k}{\sum_{i=1}^K (n_u(m, i) + \alpha_i)} \\
\phi_r(v_r) &= \frac{n_r(r, v_r) + \beta_{v_r}}{\sum_{i=1}^{V_r} (n_r(r, i) + \beta_i)} \\
\psi_l(v_t) &= \frac{n_t(l, v_t) + \gamma_{v_t}}{\sum_{i=1}^{V_t} (n_t(l, i) + \gamma_i)}
\end{aligned} \quad (6)$$

Note that ϕ_r is the region distribution for a certain region pattern r . We can use Bayes rule again to easily get the region clustering results. Clustering for timestamps is similar given ψ .

Data Set

The data set for our experiments is provided by an operational telematics company. It contains GPS trajectory data collected from automobiles in Shanghai from June 2012 to September 2012. In Shanghai urban area (121.31°N to 121.66°N, 31.11°E to 31.36°E), there are 4005 user records with 164,656 driving tracks.

We firstly convert the automobile GPS data into driving transitions. For each user, if there are not any GPS records in a certain time (here we choose 30 minutes) and the velocity in last GPS record is zero, we treat the position in last GPS record as the destination of this transition, and the next GPS record as the origin of the next transition. Furthermore, if

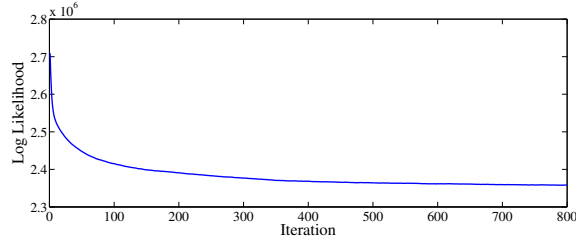


Figure 2: Log Likelihood

the origin and the destination in one transition are the same position, we consider it as an invalid transition.

In addition, in order to make different transitions share more commons (i.e. increase the repentance of *words*), we divide the whole urban area into many cell regions, which size is 250 meters by 250 meters (approximately 0.0025° in longitude or latitude), and then label all the cell regions in sequence. In temporal aspect, we use weekday along with hour as the tag of time, ignoring the exact minute and second as well as the date. Therefore, a transition is described by two cell region labels and one time tag.

Experimental Results

We first perform a convergence analysis on our model. From Figure 2, we see that the log likelihood tend to be convergent after a certain iterations, demonstrating that the model does converge.

In the rest of this sections, we firstly present the temporal and spatial patterns derived from the multinomial parameter ϕ and ψ . Then, we present the statistic results on all the trajectory data using obtained temporal and spatial patterns. For comparison, we experiment with a spatial-only LDA model, which ignores the effect of time, and present the statistic results in the same way. Finally, we show several mobility patterns of different driver categories.

Temporal Patterns

Figure 3 shows the temporal patterns generated by TS-LDA model, where different colors indicate different categories. From this figure, we can see that this model well clusters the timestamps to eight distinct categories. Each time pattern has a clear meaning in the real world.

Category 2 and 3 represent the morning peak and evening peak respectively. Category 7 shows that timestamps in the weekend belong to the same time pattern. Category 1, 6, 0 and 4 represent the morning, noon, afternoon, evening in weekdays respectively. Finally, Category 5 shows that all the timestamps between 10 p.m. and 7 a.m. belong to the same cluster *night*.

Spatial Patterns

Figure 4 shows the spatial patterns generated by TS-LDA model, where different colors indicate different regions. These regions are basically clustered on geographic positions, because users often drive their car in a certain area.

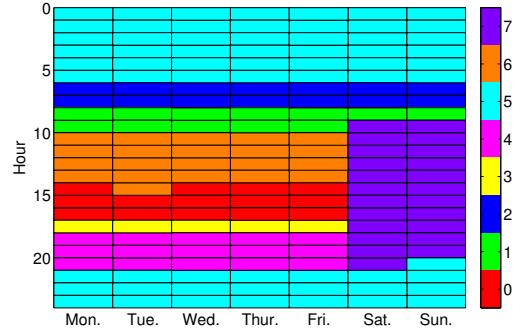


Figure 3: Temporal Patterns

Furthermore, these clustered regions share some commons with administrative districts (Figure 5) as follows:

- Region 5: This region contains District A, B, E and the northern part of District C. This area is the downtown of Shanghai, which is the most prosperous region and regarded as the CBD of Shanghai. Therefore, it is plausible to cluster them into same region.
- Region 8: This region contains District L and the middle and southern part of District C. Different from the northern part, the middle and southern part of C is further from the downtown, and not so prosperous as the northern part. It is more like District L, which has more residential communities and parks.
- Region 3: This region contains District D and part of District F. This area is an economic technical development area. Furthermore, the Hongqiao airport and train station locate in this area, so it is also a transportation center.
- Region 7: This region contains the main part of District F. Religion atmosphere is the most distinctive character in this region, for many famous temples and mosques locate here.
- Region 2: This region contains District G, H and I. This region is basically an education and entertainment area. Many famous universities and colleges locate here. There are also many parks and some sports stadiums.
- Region 0, 1, 6: District J consists of these three regions, thanks to its really large area. Nevertheless, these three regions do have some differences. Region 6 is corresponding to Lujiazui area, which is another commercial center in Shanghai, while Region 0 and 1 are basically residential areas.
- Region 4: This region is the suburb residential area with many parks, corresponding to District M.
- Region 9: This region is the suburb industry area with many factories, corresponding to District K.

User Mobility Pattern

We firstly show the statistic results of user mobility transitions. Figure 6a – 6h show the results in TS-LDA model, while in comparison, we show the results of spatial-only LDA model in Figure 6i. The color in Figure 6 represents the

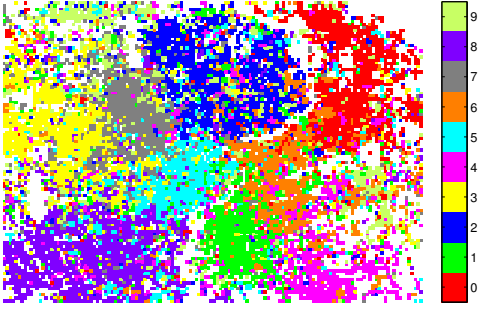


Figure 4: Region Patterns

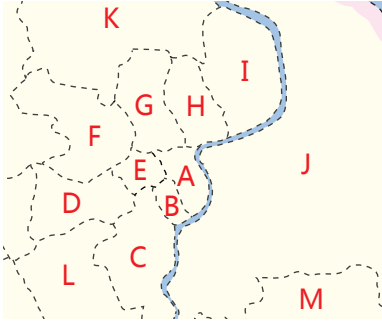


Figure 5: Shanghai Administrative Districts

percentage of a certain transition type (origin region, destination region, temporal pattern) among all the transitions. The darker is the color, the higher is the percentage.

From Figure 6i, we can see that inner-regional transitions play a dominant role among all the transitions. However, the distribution of inter-regional transitions is very sparse. Due to the lack of temporal feature, we can hardly find some distinct patterns in spatial-only LDA.

On the other hand, considering the effect of time, we can mine some meaningful patterns in Figure 6:

- The percentages on the diagonal are larger than off-diagonal ones in all eight figures, which means transitions inside one region play a dominant role in most of the regions all the time.
- In the morning peak (Figure 6a), Region 1, 2, 3 are the major origin areas, while Region 5, 6, 7 are the major destination areas. This statistic result well match the fact that Region 1, 2, 3 are suburb areas, while Region 5, 6, 7 are the downtown areas. Particularly, Region 5 is the CBD district in Shanghai, which is the most popular destination in the morning peak.
- The transitions in the evening peak (Figure 6e) are just the opposite of which in the morning peak. Region 5, 6, 7 become popular origins, while Region 1, 2, 3 are the main destinations. In addition, Region 5 is the only popular origin at night, possibly due to the beautiful night scenes in The Bund attracting many tourists to stay there until night.

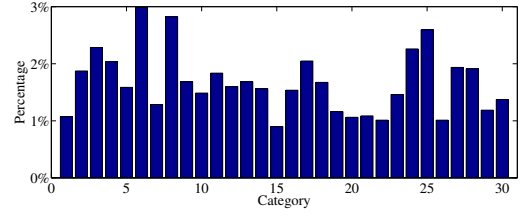


Figure 7: The user percentage in each category.

- In the morning (Figure 6b) and afternoon (Figure 6d), the number of transitions is comparatively small, which means these two time intervals are working time with not many people driving in the city, while in the noon, the number becomes larger, indicating that it is time for people to drive out for lunch.
- The transitions in the weekends looks a little complicated and disorder. The number of transition is the largest among all the eight temporal patterns, and there is no dominant origin or destination, because the driving motivations are usually irregular for most people in the weekends. That is why there are not any obvious patterns in Figure 6h.

We try to group users based on their mobility patterns. We treat θ_m as the K -dimensional feature of user m and use K-means clustering method with Pearson correlation distance on θ . We divide all the users into 30 categories. Figure 7 shows the driver percentage in each category.

Here we choose four typical and comparable user types. Drivers in both of first two types (Row 1, 2 in Figure 8) live in Region 7 (Putuo District). However, they are distinctly two different types. Drivers belonging to Row 1 basically move inside Region 7, which indicates that their working places are not too far away from their home, so their mobility pattern is local movement. On the contrary, drivers in Row 2 usually drive far away from their home. Their working places are widely spread all over the west of Shanghai. Although they may work in different areas, their working hours and driving patterns are similar.

On the other hand, drivers in Row 3 and Row 4 are two another types, who largely drives in the eastern part of Shanghai. Drivers in Row 3 lives in the suburb area, Region 0, and the downtown of Shanghai, Region 6, is a major working place for them. However, drivers in Row 4 live an opposite life. They live in the downtown, Region 6, and working in the suburbs, Region 0, 1, 4. Their mobility pattern is driving to suburb in the morning, and back to downtown in the evening.

According to previous results and analysis, we can see that our model clearly divide drivers into different types, and successfully mine out their mobility patterns, which combine both temporal and spatial features.

Related Works

Mobility analysis: Recently, the increasing availability of large scale trajectory data have raised a number of researches on mobility pattern analysis. Some researches focuses on

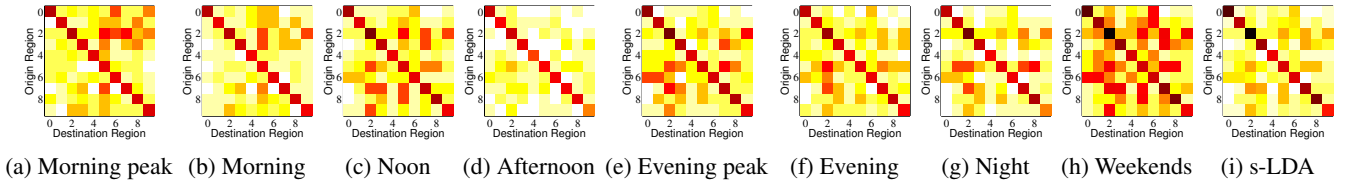


Figure 6: Statistic results of user mobility. Figure 6a – 6h are the results in TS-LDA, corresponding to temporal pattern 2, 1, 6, 0, 3, 4, 5, 7 respectively, while the Figure 6i is in spatial-only LDA. The darker color represents the higher percentage.

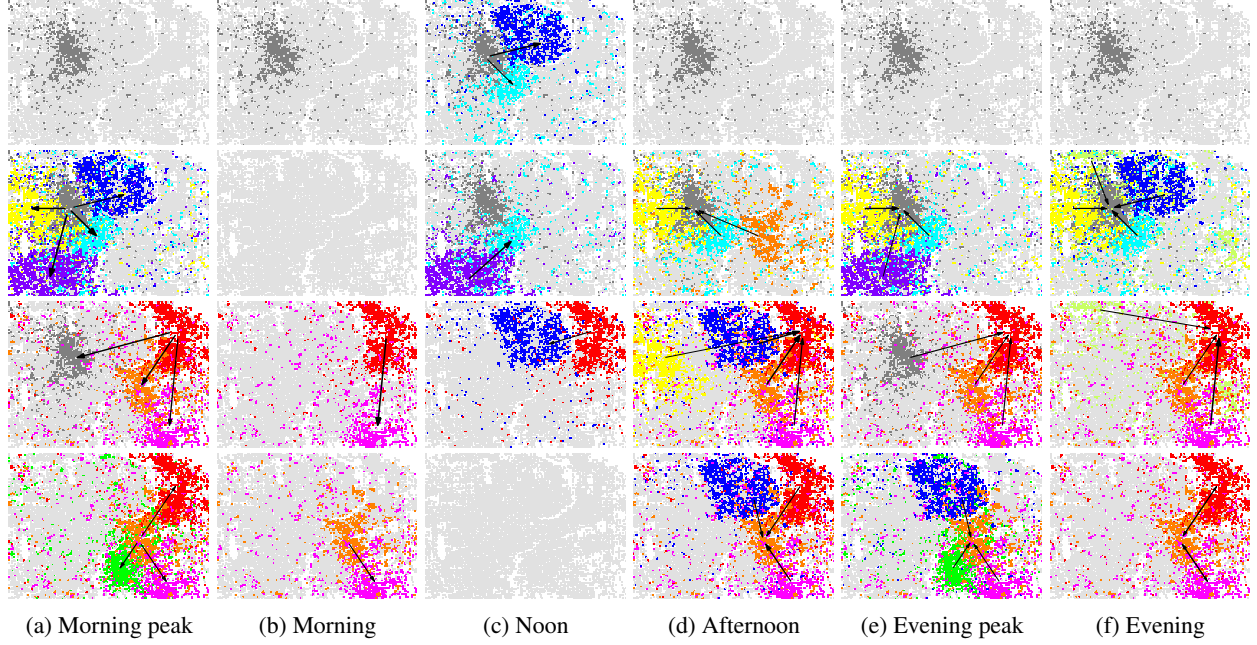


Figure 8: Each row represents the mobility patterns of a certain user type in different temporal patterns. The two figures in one column share the same temporal pattern. The temporal patterns from left to right are pattern 2, 1, 6, 0, 3, 4, representing morning peak, morning, noon, afternoon, evening peak, evening respectively. Arrows in the figures represent the most frequent transitions of each user in the given temporal pattern. The wider is the arrow, the higher probability is this transition pattern. Here we only draw the inter-regional transitions with probability over a threshold.

discovering the hidden information behind the trajectory data. Giannotti et al. proposed a query and mining model on trajectory clustering and mobility analysis (Giannotti et al. 2011). Cho et al. focused on reveal the correlation between user movements and their social networks (Cho, Myers, and Leskovec 2011). On the other hand, some other of them focused on mobility prediction. Song et al. focused on prediction of human emergency behavior after severe disaster (Song et al. 2014). Kalman filtering was used on location prediction in public ATM networks (Liu, Bahl, and Chlamtac 1998). Backstrom et al. proposed a prediction method based on friend and IP address (Backstrom, Sun, and Marlow 2010). In addition, there are many other novel topics using GPS trajectory data (Trasarti et al. 2011), (Ong et al. 2011), (Yuan, Zheng, and Xie 2012).

Topic Model: The topic model is also a very hot research area. The Latent Dirichlet Allocation model proved to be very successful in mining the hidden topics of large archives documents (Blei, Ng, and Jordan 2003). Ram-

age et al. proposed the Labeled LDA for credit attribution in multi-labeled documents (Ramage et al. 2009). Chen et al. extended the LDA model into Coupled LDA to mining the IPTV user behaviors (Chen, Zhang, and Zha 2013). In computer vision area, spatial-temporal features have already raised researchers attentions when applying LDA model (Wang and Grimson 2008), (Niebles, Wang, and Fei-Fei 2008).

Conclusion

In this paper, we propose the TS-LDA model for discovering mobility patterns, which separately consider the effects of time and position. This model not only extract the hidden temporal and spatial patterns successfully, but also categories the drivers into different types according to their driving mobility habits. This model also reveal some interesting characters on both regions and drivers, which can offer some guides on traveling and help drivers better make their driving plan.

References

- Backstrom, L.; Sun, E.; and Marlow, C. 2010. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World wide web*, 61–70. ACM.
- Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3:993–1022.
- Chen, W.; Zhang, Y.; and Zha, H. 2013. Mining iptv user behaviors with a coupled lda model. In *Broadband Multimedia Systems and Broadcasting (BMSB), 2013 IEEE International Symposium on*, 1–6. IEEE.
- Cho, E.; Myers, S. A.; and Leskovec, J. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1082–1090. ACM.
- Giannotti, F.; Nanni, M.; Pedreschi, D.; Renso, C.; and Trasarti, R. 2009. Mining mobility behavior from trajectory data. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4, 948–951. IEEE.
- Giannotti, F.; Nanni, M.; Pedreschi, D.; Pinelli, F.; Renso, C.; Rinzivillo, S.; and Trasarti, R. 2011. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *The VLDB Journal/The International Journal on Very Large Data Bases* 20(5):695–719.
- Kurashima, T.; Iwata, T.; Hoshide, T.; Takaya, N.; and Fujimura, K. 2013. Geo topic model: joint modeling of user's activity area and interests for location recommendation. In *Proceedings of the sixth ACM international conference on Web search and data mining*, 375–384. ACM.
- Liu, T.; Bahl, P.; and Chlamtac, I. 1998. Mobility modeling, location tracking, and trajectory prediction in wireless atm networks. *Selected Areas in Communications, IEEE Journal on* 16(6):922–936.
- Mimno, D., and McCallum, A. 2012. Topic models conditioned on arbitrary features with dirichlet-multinomial regression. *arXiv preprint arXiv:1206.3278*.
- Niebles, J. C.; Wang, H.; and Fei-Fei, L. 2008. Unsupervised learning of human action categories using spatial-temporal words. *International journal of computer vision* 79(3):299–318.
- Ong, R.; Pinelli, F.; Trasarti, R.; Nanni, M.; Renso, C.; Rinzivillo, S.; and Giannotti, F. 2011. Traffic jams detection using flock mining. In *Machine Learning and Knowledge Discovery in Databases*. Springer. 650–653.
- Ramage, D.; Hall, D.; Nallapati, R.; and Manning, C. D. 2009. Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, 248–256. Association for Computational Linguistics.
- Song, X.; Zhang, Q.; Sekimoto, Y.; and Shibasaki, R. 2014. Prediction of human emergency behavior and their mobility following large-scale disaster. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 5–14. ACM.
- Trasarti, R.; Pinelli, F.; Nanni, M.; and Giannotti, F. 2011. Mining mobility user profiles for car pooling. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1190–1198. ACM.
- Wang, X., and Grimson, E. 2008. Spatial latent dirichlet allocation. In *Advances in Neural Information Processing Systems*, 1577–1584.
- Yuan, J.; Zheng, Y.; and Xie, X. 2012. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 186–194. ACM.