ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from http://www.i6doc.com/en/.

Learning human behaviors and lifestyle by capturing temporal relations in mobility patterns

Eyal Ben Zion and Boaz Lerner

Ben-Gurion University of the Negev, Beer Sheva, Israel
{benzione, boaz}@bgu.ac.il

Abstract. Many applications benefit from learning human behaviors and lifestyle. Different trajectories can represent a behavior, and previous behaviors and trajectories can influence decisions on further behaviors and on visiting future places and taking familiar or new trajectories. To more accurately explain and predict personal behavior, we extend a topic model to capture temporal relations among previous trajectories/weeks and current ones. In addition, we show how different trajectories may have the same latent cause, which we relate to lifestyle. The code for our algorithm is available online.

1 Introduction

Determining a user's learning mobility pattern (LMP) is challenging because it involves many aspects in the user's life and levels of knowledge combined with a high level of uncertainty. This challenge is historically connected to systems optimization, e.g., in predicting the density of cellular users to perform resource reservation and handoff prioritization in cellular networks [1]. As information collected from cellphones has become increasingly personal, models have become more user-specific. Thus, predicting a mobility pattern is at a higher level of learning than finding the geographic coordinates of locations, enabling prediction of significant places for individuals [2].

Several studies in LMPs from mobile phone data have used the latent Dirichlet allocation (LDA) model [3]. The LDA can infer the function (LDA "topic") of a region (LDA "document") in a city, e.g., educational locations or business districts using human mobility patterns based on entrance/exit from a region (LDA "words") and categories of points of interest, e.g., restaurants and shopping malls [4]. It can also infer [5] on potential and intrinsic relations among geographic locations (topics) in accordance with user trajectories (documents) using location records (words) a user shares in location-based social networking services[1]. Also, the LDA model can be extended to capture temporal changes when the previous word impacts the current one [6] or when the previous topic impacts the current word or topic [7].

To learn mobility patterns by LDA, previous research found interactions among a user's significant places [8], but not among trajectories comprising significant places, which reflect user behavior. Because trajectories occurring in previous weeks can impact a current trajectory and behavior, we learn behavior by capturing temporal relations in mobility patterns. We extend the LDA to

[1] http://foursquare.com/

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

capture the time influence on user mobility patterns using temporal models that relax the model's assumptions of "bag of words" and document order, which cannot capture dependencies among mobility patterns. After modifying the raw data of a user's locations, which consist of latitude, longitude, and timestamps, into a corpus of documents with a trajectory as a "word" and a week as a "document" (in the context of LDA), we apply temporal modeling to cellular-phone user data and demonstrate the advantage of the extended models over the original, which does not capture dependencies. Further, from the learned mobility patterns we infer their latent causes, which are related to lifestyle.

2 Latent Dirichlet allocation

An LDA is a topic model [3] in which a document comprises several topics, where a topic has a probability distribution over words. This model finds a latent topic $z_{d,j} \in \{1, \ldots, K\}$ for the observed word $w_{d,j} = t$ from a vocabulary $V = \{1, \ldots, N\}$ of words for the jth token in the dth document, $\boldsymbol{w}_d$, in a document corpus $W = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_M\}$. It makes a "bag of words" assumption, i.e., neither the word order in each document, nor document order in the corpus is important. It finds the latent topic by computing co-occurrences of words in the corpus [3] and has three inputs: two hyperparameters of the Dirichlet distributions, α and β, and the number of latent topics, K. In this study, we see the LDA model as a Bayesian network (BN) over a set of nodes (variables) $\mathcal{V} = \{v_1, \ldots, v_n\}$. We define [9]: 1) The parents of node $v_\imath$, $Pa(v_\imath)$, are the nodes $v_\jmath$ that have a direct edge to $v_\imath$ ($v_\jmath \rightarrow v_\imath$); 2) The descendants of $v_\imath$, $D(v_\imath)$, are the nodes that are reachable from $v_\imath$ by following directed edges; 3) The Markov blanket of node $v_\imath$, $MB(v_\imath)$, consists of the parents, children, and parents of children of $v_\imath$; and 4) The children and parents of children of $v_\imath$ are $Pa^C(v_\imath) = MB(v_\imath) - Pa(v_\imath)$.

3 Extending LDA to capture temporal relations

To solve the LMP problem, we extend the LDA model to identify human behaviors and lifestyle (LS) by allowing a previous trajectory and even a previous week to affect the current trajectory. In our extension, words, w, are the trajectories, documents, d, are the weeks, and topics, z, are LSs.

To consider temporal relations in the LMP problem, models that capture different dependencies in user mobility patterns are considered. Each combination with a dotted line in Figure 1 is one of 15 possible temporal models that extend the LDA. In Table 1, the symbols $(w_p)z$ represent a model in which the previous trajectory node, w_p, is another parent to the LS node, z. The symbols $(d_p)z-(w_p)w$ represent a model in which the previous week node, d_p, is another parent of the LS node, z; and also the previous trajectory node, w_p, is another parent of the current trajectory node, w. The symbols $(d_p w_p)zw$ represent a model in which the previous week node, d_p, and previous trajectory node, w_p, are additional parents of the LS node, z and of the current trajectory node, w. Algorithm 1 iteratively extends the Gibbs sampling for the LDA [3] to LSs by:

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

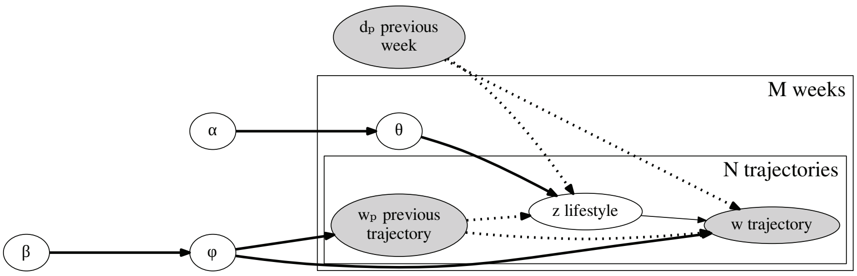

Fig. 1: Temporal models extending LDA – dotted lines represent possible edges.

Model #	Model symbol	Model #	Model symbol
1	original LDA	9	$(w_p)z - (d_p)w$
2	$(d_p)z$	10	$(d_p w_p)w$
3	$(d_p)w$	11	$(w_p)zw$
4	$(w_p)z$	12	$(d_p w_p)z - (d_p)w$
5	$(w_p)w$ [[6]]	13	$(d_p)z - (d_p w_p)w$
6	$(d_p)zw$	14	$(d_p w_p)z - (w_p)w$
7	$(d_p w_p)z$	15	$(w_p)z - (d_p w_p)w$
8	$(d_p)z - (w_p)w$	16	$(d_p w_p)zw$

Table 1: Fifteen temporal extensions of the original LDA model.

$$P(z_{d,j} = k | \boldsymbol{Z}_{-(d,j)}, \boldsymbol{MB}(z)) = \frac{n^k_{Pa(z)} + \alpha_k - 1}{\Sigma^K_{k`=1} n^{k`}_{Pa(z)} + \alpha_{k`} - 1} \cdot \frac{n_k^{Pa^C(z).t} + \beta_t - 1}{\Sigma^N_{t`=1} n_k^{Pa^C(z).t`} + \beta_{t`} - 1},$$

where $\boldsymbol{MB}(z)$, $Pa(z)$, and $Pa^C(z)$ are the Markov blanket, parents, and children and parents' children of LS z, respectively; $\boldsymbol{Z}_{-(d,j)}$ is the distribution over LSs without LS $z_{d,j}$ for the jth day in the dth week; $n^k_{Pa(z)}$ is the number of times LS k has been observed in z for a parent configuration; and $n_k^{Pa^C(z).t}$ is the number of times a trajectory t has been assigned LS k in z, conditioned on a configuration of z's children and parents' children. Following Gibbs sampling, the predictive distributions, $\hat{\boldsymbol{\varphi}}$ and $\hat{\vartheta}$ (the trajectory and LS-mixture components) are:

$$\hat{\varphi}_{k,Pa^C(z).t} = (n_k^{Pa^C(z).t} + \beta_t)/(\Sigma^N_{t`=1} n_k^{Pa^C(z).t`} + \beta_{t`}).$$

$$\hat{\vartheta}_{Pa(z),k} = (n^k_{Pa(z)} + \alpha_k)/(\Sigma^K_{k`=1} n^{k`}_{Pa(z)} + \alpha_{k`}).$$

4 Methodology

We used data from the Google location app[2] that tracks users in non-uniform intervals of time and records latitude, longitude, and timestamps. Since the LDA's input should be documents (weeks), we modified the raw data in four stages. First, because the recorded latitude and longitude are not always precise,

[2]http://www.google.com/maps/

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

Algorithm 1 A generative temporal model to extend LDA

```
procedure LDA(α, β, M, K) (K – # of lifestyles; M – # of weeks)
    for each combination of Pa(z) do
        Sample the lifestyle-mixture components: θ_{Pa(z)} ∽ Dirichlet(α)
    end for
    for k = 1 to K do
        Sample trajectory selection components: φ_k ∽ Dirichlet(β_k)
    end for
    for d = 1 to M and j = 1 to Week length do
        Sample a lifestyle: z_{d,j} ∽ Multinomial(θ_{Pa(z)})
        Sample a trajectory: w_{d,j} ∽ Multinomial(φ_{z_{d,j}, Pa^C(z)})
    end for
    return corpus of weeks
end procedure
```

when places that are close to one another have the same semantic meaning, we grouped them. By converting the geographic map into a grid ($10X10m^2$)(this cell size performed best in a preliminary study) and rounding the latitude and longitude according to: $new_x = round(old_x * 10,000/5) * (5/10,000)$, (10,000 and 5 are parameters to create the grid cell size), we defined every grid cell as a stop area, and every longitude and latitude point was related to a stop area by geographical position. Second, we filtered non-significant places, (places visited infrequently, i.e., less than twice in a week). If this place was suspected as rare in at least 93% of the weeks, it was labeled *Other*. For the remaining (significant) stop areas, the user supplied their meaning, e.g., "Home", "Work", "Sport". In this process, the user could group close stop areas with the same meaning, which is specific to him. Third, we limited trajectories (places a user visited sequentially in a given interval of time) to the length of a day (a basic time period in a user's life). Thus, to form a trajectory, we created a vector with 24 slots, each representing one hour. Each slot was assigned a letter for the semantic meaning of the most frequent stop area at this specific hour on this specific day. Thus, the daily trajectory was defined as a string of 24 letters. If there was no record at a specific hour, this hour was assigned the semantic value of *No Record*. To reduce the noise in the trajectories (the number of possible trajectories is the number of stop areas to the power of 24), they were clustered using hierarchical clustering with the edit distance (measuring the distance between two strings that are 24 dimensional daily trajectories). Fourth, the representation of documents in the LDA model are weeks (another basic time unit in the user's life). Thus, a series of seven sequential trajectories was grouped in a week, beginning on a Sunday and ending on a Saturday. To reduce the noise in the weeks, they were clustered using hierarchical clustering with the KL-divergence as a metric because weeks are represented as trajectory distributions. We used 100 and 20 clusters for clustering the trajectories and weeks, respectively, in accordance with our data size (see below) and because these values gave good results in a preliminary study.

We examined the 16 models of Table 1 (the LDA and its 15 extensions) with seven cellular users over periods of 105, 77, 73, 109, 147, 70, and 79 weeks having 12, 14, 11, 11, 9, 10, and 15 significant stop areas, respectively. We used a sliding

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from http://www.i6doc.com/en/.

window size of 75, 47, 43, 79, 117, 40, and 49 weeks, respectively, with the last five weeks used for testing ($\tilde{m}$) and the rest were used for training (m); thereby, we could create sevenfold datasets with independent test sets each user.

We calculated the models' log-likelihood

$$LL = \log P(\boldsymbol{W}_{\tilde{m}}|\mathcal{M}) = \sum_{t=1}^{N} n_{\tilde{m}}^{t} \cdot \log(\sum_{k=1}^{K} \hat{\vartheta}_{Pa(z),k} \cdot \hat{\varphi}_{k,Pa^{C}(z).t}),$$

where $n_{\tilde{m}}^{t}$ is the number of times a trajectory t has been observed in tested weeks $\tilde{m}$, and the perplexity (our performance measure) using the test set is $\exp\{-LL(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\varphi}})/n_{\tilde{m}}\}$. We tested the perplexity of the 16 models using 15 initializations of the Gibbs algorithm and chose the best model according to a repeated Friedman test. As expected [3], the perplexity decreases with the number of LSs. The complexity of the model is $[(N-1) \cdot |Pa(w)| \cdot (K-1) \cdot |Pa(z)|]$.

#	User 1	User 2	User 3	User 4	User 5	User 6	User 7
2	ND	[4, 7, 9, 11-16]	ND	[2, 4]	ND	ND	ND
6	[1-5, 7]	[4, 7, 11-12, 14-16]	[1-2]	[1-4, 7, 9]	[1-2, 4, 7]	[2, 4, 7, 9]	[1, 5]
10	[1-5, 7-8]	[1, 4-5, 7-9, 11-12, 14-16]	[1-2, 5]	[1-2, 4, 9]	[2, 4]	[2, 4-5, 7-8]	[1, 3, 5]
14	[1-5, 7]	ND	[1, 5]	[1-5]	[1-4]	[2-4, 7-9]	[3, 5]
18	[2-5]	ND	[1-3, 5]	[1-4]	[1-4]	[1-5, 7-8]	[3, 5]

Table 2: Best models according to a Friedman test (ND is "no difference").

LS 1	LS 2	LS 3
[Home, 0]	[Friends, 0] → [Study, 11] → [Other, 12] → [Study, 20] → [Friends, 22] → [Other, 23]	[Parents Home, 0] → [Home, 15] → [Parents Home, 17]
[Home, 0] → [Study, 14] → [Other, 18] → [Home, 20]	[Friends, 0] → [Study, 12] → [Other, 15]	[Parents Home, 0] → [Other, 20] → [Home, 21]
[Friends, 0] → [Home, 2] → [Study, 10] → [Home, 15] → [Study, 19] → [Home, 22]	[Friends, 0] → [Home, 3]	[Home, 0] → [Friends, 18] → [Other, 19]
[Home, 0] → [Other, 17] → [Bus Station, 18] → [Parents Home, 19] → [Bus Station, 23]	[Friends, 0] → [Study, 15] → [Other, 18] → [Friends, 23]	[Parents Home, 0] → [Other, 13] → [Parents Home, 17] → [Other, 22]

Table 3: Randomly selected trajectories of the four most probable trajectory clusters learned using Model 4 for three (of ten) LSs of User 5. A trajectory is represented by its significant stop areas and the corresponding hours of the stop.

Table 2 shows statistically significant models among the 16 for increasing numbers (in $[2-18]$) of LSs learned and the seven users. We use "ND" when there was no statistical difference among the models. We use a performance score that measures the complexity of the learning task by the inverse number of models of the 16 that are statistically significant to the others. A large value of this score reflects ease in "being a good model", as more models are good or consider the task easy. Then we identified two groups of users: Group 1 of Users 1, 2, 4, and 7; and Group 2 of Users 3, 5, and 6. While the score of Group 1 shows a peak for a moderate number of LSs (usually at 10), that of Group 2 shows an increase from a very low number of LSs to a plateau for this number (Users 3 and 5) or to a higher number of LSs monotonically (User 6), and generally with a lower peak value than that of Group 1. Since the number of LSs is a measure of model complexity, we may infer that Group 1 is of users with more stable

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

life (these users need a model with moderate complexity; otherwise, it either underfits or overfits the data), whereas users of Group 2 show less stability and more uncertainty, requiring models with higher complexities to reach reasonable performance. Checking our records, we observed that Group 1 includes working individuals who usually have families, whereas Group 2 includes either students (Users 5 and 6) or a person who changed addresses during the data collection period (User 3). Extending the LDA algorithm by modeling temporal relations provided a clear advantage in some cases (Users 2 and 6 for few LSs and User 7 for many LSs), a slight advantage in some other cases (some models for Users 1, 4, and 5), and no advantage for User 3.

Finally, Table 3 shows trajectories in the four most probable clusters for three LSs of User 5 (a nonworking student). Each stop area of a trajectory is derived from a semantic dictionary of stop areas for this user. Roughly speaking, the three LSs reflect a routine of traveling between home and place of studies (LS1), going out to/with friends (LS2), and visiting parents and related places (LS3).

5 Conclusions

This study proposes an approach to solve the LMP problem by extending the LDA algorithm to learn temporal relations. This approach shows how significant visited places and sequences of these places are important to understanding user behavior and are related to lifestyle. These temporal models are also applicable to other time-related problems, e.g., intelligent service-based human mobility that requires user clustering based on lifestyle distributions over trajectories.

References

[1] W. Soh and H. Kim. QoS provisioning in cellular networks based on mobility prediction techniques. *Communications Magazine, IEEE*, 41(1):86–92, 2003.

[2] D. Ashbrook and T. Starner. Using GPS to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5):275–286, 2003.

[3] T. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5228–5235, 2004.

[4] J. Yuan, Z. Yu, and X. Xing. Discovering regions of different functions in a city using human mobility and POIs. In *Proc. of the 18th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 186–194, Beijing, China, 2012.

[5] X. Long, J. Lei, and J. Joshi. Exploring trajectory-driven local geographic topics in FourSquare. In *Proc. of the 2012 ACM Conf. on Ubiquitous Computing*, pages 927–934, Pittsburgh, PA, USA, 2012.

[6] X. Wang, A. McCallum, and X. Wei. Topical n-grams: phrase and topic discovery, with an application to information retrieval. In *The 7th IEEE International Conference on Data Mining (ICDM)*, pages 697–702, Omaha, NE, USA, 2007.

[7] N. Barbieri, G. Manco, E. Ritacco, M. Carnuccio, and A. Bevacqua. Probabilistic topic models for sequence data. *Machine Learning*, 93(1):5–29, 2013.

[8] K. Farrahi and D. Gatica-Perez. Extracting mobile behavioral patterns with the distant n-gram topic model. In *16th Int. Symposium on Wearable Computers*, pages 1–8, Newcastle, UK, 2012.

[9] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search.* MIT Press, 2000.