Exploring Urban Lifestyles Using a Nonparametric Temporal Graphical Model*

Shoaib Jameel[1], Yi Liao[2], Wai Lam[2], Steven Schockaert[1], Xing Xie[3]
[1]School of Computer Science and Informatics, Cardiff University
[2]Dept. of Systems Engineering and Engineering Management, The Chinese University of Hong Kong
[3]Microsoft Research
{jameels1,schockaerts1}@cardiff.ac.uk {yliao, wlam}@se.cuhk.edu.hk
xingx@microsoft.com

ABSTRACT

We propose a new unsupervised nonparametric temporal topic model to discover lifestyle patterns from location-based social networks. By relating the textual content, time stamps, and venue categories associated to user check-ins, our framework detects the predominant lifestyle patterns in a given geographic region. The temporal component of our model allows us to analyse the evolution of lifestyle patterns throughout the year. We provide examples of interesting patterns that have been discovered by our model, and we show that our model compares favourably to existing approaches in terms of lifestyle pattern quality and computation time. We also quantitatively show that our model outperforms existing methods in a time stamp prediction task.

Keywords

Graphical models, location-based social networks, text mining, lifestyle patterns, urban computing

1. INTRODUCTION

> *A city is more than a place in space, it is a drama in time.* – Patrick Geddes

The lifestyles of different groups of people can differ considerably, and this is reflected in the types of places they visit. For example, an office worker may go to the office in the morning, visit a sandwich shop at midday, and go home in the evening. Moreover, these lifestyle patterns are not fixed, as people constantly adapt to changes in their environment [15, 9, 36], e.g., the changing weather across different seasons. Our aim in this paper is to characterize the predominant lifestyle patterns in a given geographic region, and to analyse how they evolve throughout the year.

In this paper, we will characterize lifestyle patterns from information that is shared publicly on location-based social networks (LBSNs) such as Facebook Places and Foursquare, which enable people to share information about the places they visit. A check-in or digital footprint, in this context, is a record of the venue which a user has visited along with a time stamp of that visit. Several authors have proposed models for analysing check-in data to get insights into human mobility patterns [35, 32]. For example, [35] proposed a hybrid model, based on a combination of two existing topic models, that combines check-in information with a friendship graph. While their method was able to discover some interesting results, it is computationally demanding and cannot discover interpretable patterns from small datasets. Most importantly, the model lacks a temporal component, which makes it harder to distinguish between some lifestyle patterns (e.g., office workers who go to the pub just after work, compared to students who go to the pub late in the evening) and means that it cannot be used to analyse changes over time.

In this paper, we present a new statistical model for discovering interpretable lifestyle patterns from check-in data in which we address the aforementioned issues. By relating the textual content, time stamps and venue categories associated with user check-ins, our framework detects the predominant lifestyle patterns for a given geographic region, where a lifestyle pattern is characterised by the kinds of places that are frequented by a given group of people at different times in the day and different days of the week. To tackle the pattern discovery problem, we develop an unsupervised nonparametric temporal topic model. The temporal component of our model allows us to analyse the evolution of lifestyle patterns throughout the year. To deal with sparsity issues in small datasets, we abstract away from raw check-in data, and instead consider latent venue types and discrete temporal units. This allows us, among others, to study lifestyle patterns at a finer spatial granularity or in areas of the world where the uptake of LBSNs is relatively limited.

*The work described in this paper is substantially supported by grants from the Research Grant Council of the Hong Kong Special Administrative Region, China (Project Codes: 14203414) and the Direct Grant of the Faculty of Engineering, CUHK (Project Code: 4055034). This work is also supported by ERC Starting Grant 637277. We thank Dr. Zhiyuan Cheng, Google for providing the raw check-in dataset and Prasanta Bhattacharya of the National University of Singapore for giving us valuable information about crawling data. This work was done when the first author was a Postdoctoral Fellow at the Chinese University of Hong Kong.

ICTIR '16, September 12-16, 2016, Newark, DE, USA

DOI: http://dx.doi.org/10.1145/2970398.2970401

As we illustrate in Section 4, our method can be used to gain insight into the lifestyles of different groups of people in a given region. Among others, the obtained lifestyle patterns could provide valuable information for advertising, by informing marketers about when and where they can most effectively reach a particular target audience [22, 9]. The temporal component of the lifestyle patterns would also be useful for developing time-sensitive recommendation systems (e.g., not recommending sandwich shops to office workers during the weekend). Developing models of people's lifestyles could furthermore be useful for the field of medical science [1], although the fact that LBSNs only cover particular demographics would constitute an important drawback for such applications. While this could be alleviated by applying our model to data sources that have a wider coverage, such as credit card transaction data or mobile phone traces [35], a lack of access to such data sources means that we have not considered this in our experiments. It is important to note that the use of check-in data is mostly a matter of convenience. We indeed expect that other sources of data, such as credit card transactions and mobile phone logs would enable more detailed predictions. However, we believe that the model we propose could be straightforwardly applied to such data sources. The advantage of using check-in data is that this information is publicly available, which we believe makes it a suitable information source for developing and evaluating the model.

2. RELATED WORK

The most closely related work is [35], where the authors propose an approach that combines the relational topic model (`RTM`) [7] with the hierarchical latent Dirichlet allocation (`hLDA`) model [4] to discover so-called urban lifestyle patterns from LBSNs. Specifically, the output of this method is a hierarchical topical pattern, which closely resembles the tree structure generated by the `hLDA` model. Each node of this tree corresponds to a topic, where the further down the tree we go, the more specific these topics become. The topics are called living patterns in [35] and intuitively correspond to soft clusters of related venues, encoded as probability distributions over footprints. Each branch of the tree represents a lifestyle pattern. In other words, a lifestyle pattern corresponds to a set of living patterns, some of which are shared with many other lifestyle patterns (viz. those corresponding to nodes close to the root) while others may be more specific. The `RTM` component of the model is used to take into account a social friendship graph, capturing the intuition that friends tend to have similar lifestyles. As mentioned in the introduction, the main limitations of the approach from [35], which we address in our model, are the fact that it is computationally demanding, cannot discover interpretable patterns from small datasets and does not have a temporal component. In contrast to the approach from [35], our model generates flat patterns. The main reason why a tree structure is used in [35] is to capture correlations between different lifestyle patterns. However, this tree structure makes the model computationally too demanding, and as we will explain in the next section, in our model we can still capture correlations between lifestyles patterns, by using a two-level topic structure. Another important difference is that our model has a temporal component, which requires a considerably different posterior inference scheme than the one that was proposed in [35].

Some authors have already studied topic models with a temporal component. For example, [5] proposes a Markovian temporal topic model, which generates topics based on the change in co-occurrence information over time. A non-Markovian temporal topic model has been proposed in [31], based on the assumption that document time stamps are generated from a Beta distribution. Several temporal topic models that are specifically aimed at social network data have recently been proposed. Yin et al. [34] proposed topic models to capture user behaviour in social media. Their underlying assumption is that users' intrinsic interests depend on the temporal context. Some nonparametric temporal topic models have been proposed as well [10]. Temporal features have also been used in [28], which focuses on linking Twitter posts to external documents to improve the quality of topic models. However, to the best of our knowledge, temporal topic models have not yet been considered for studying lifestyle patterns. The reason why we need a new model to capture lifestyle patterns is that existing temporal topic models are aimed at modelling documents (which are associated with single time stamps) whereas we need to model user behaviour (which is associated with sequences of time stamps), and because lifestyles are best modelled using two levels of topics, as we will discuss below.

Some works have extended topic models with both temporal and spatial features. For example, Zhou et al. [38] proposed a location-time constrained topic model to capture Twitter events. In [16], the authors study urban dynamics using the `LDA` model [6], analysing the temporal and spatial nature of topics in a post-hoc step. Bauer et al. [2] proposed a topic model that considers both spatial and temporal features to model the topics discussed by Foursquare users, aiming to provide insights into the cultural idiosyncrasies of different cities. Temporal and spatial features are also prominent in approaches that study human mobility patterns. For example, [23] studies the structural and temporal changes in the spatial network formed by human movement patterns. Kim et al. [14] used the `LDA` model to discover topic-based place semantics without using any predefined semantic categories. In addition, the authors studied the temporal dynamics of the place semantics.

The idea of two-level topic models is also used in the Pachinko Allocation Model (`PAM`) [20], which generates super and sub-topics that can be represented as a directed acyclic graph. A nonparametric extension to the `PAM` model has been proposed in [19]. A nonparametric extension of `PAM` for short texts has been proposed in [26]. In [25] and [12], graphical models have been proposed that generate a hierarchy of topics, aimed respectively at document co-clustering and topic segmentation. In [37], a hierarchical topic structure is proposed consisting of different levels of the Hierarchical Dirichlet Processes (`HDP`) [27] model. Our model is significantly different from the above graphical models, which do not consider the change of topics over time. Moreover, in contrast to the `PAM` model, our model does not impose any arbitrary relationships on the hierarchy.

3. OUR FRAMEWORK

A lifestyle pattern in our model corresponds to a probability distribution over lifestyle topics, which are in turn probability distributions over abstract footprints (see below). Lifestyle topics play a similar role as the living patterns in [35], but they are based on latent venue categories

rather than individual venues and they incorporate a temporal component. Intuitively, a lifestyle topic models one aspect of a given lifestyle. By describing lifestyles in terms of such topics, we allow the model to make explicit the commonalities between different lifestyles. This high-level view makes lifestyle patterns easier to interpret and enables us to more clearly describe how two lifestyle patterns differ or how a given lifestyle pattern changes over time. Moreover, the fact that lifestyle patterns can share lifestyle topics means that fewer parameters have to be learned, leading to more robust results.

An important issue in dealing with check-in data is sparsity, as we may only have limited information about some venues. This problem is exacerbated by our use of a temporal component, which requires sufficient numbers of check-ins for different times of the day, days of the week and months of the year. To deal with this issue, we abstract away from specific venues and specific time stamps, as we explain in Section 3.1. This design makes our model more reliable, essentially avoiding overfitting by estimating topics from the abstract check-in behaviour of many users. In Section 3.2, we then introduce the graphical model which relates the resulting abstract footprints to lifestyle patterns. Section 3.3 explains how we can do posterior inference in this model to discover these lifestyle patterns.

3.1 Dealing with Sparsity

The input to our method consists of a set of check-ins for each user, along with the user comments associated with the check-ins and the tags and categories associated with each venue. Rather than estimating lifestyles from the check-in data directly, we first convert each check-in to an abstract footprint, in which the specific name of the venue is replaced by a venue category. While we could use the categories provided by Foursquare and similar services, the taxonomies which are used by these LBSNs are not always sufficiently fine-grained. Moreover, these taxonomies are LBSN-specific, which causes problems when we want to integrate check-in data from different LBSNs.

To obtain a suitable category structure, we first represent each venue as a bag of words, consisting of (i) the tags associated with each venue, (ii) the names of the categories assigned to the venue, and (iii) the nouns and adjectives occurring in the user comments associated with the venue. To generate the venue categories that will be used in the abstract footprints, we then use the Hierarchical Dirichlet Process (`HDP`) model, a nonparametric counterpart of the well-known Latent Dirichlet Allocation (`LDA`) method in which the appropriate number of topics is chosen automatically based on the characteristics of the data. The output of the `HDP` method consists of a number of latent topics, which we will interpret as venue categories. Note that each venue category is thus represented as a probability distribution over words. Finally, we assign a label to each of the discovered venue categories. To this end, we select the most representative word from each probability distribution using the PMI based technique from [18]. In cases where there is more than one category with the same label, the second most representative word is added to the label. While this method is simpler than some other existing label selection methods [17, 21, 18], we found it to yield good results in this context.

The time stamp associated with check-in records will be used in two ways. First we will use it to analyse how lifestyle

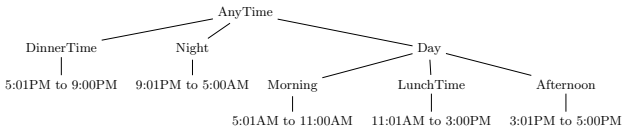

Figure 1: Intra-day temporal concept ontology

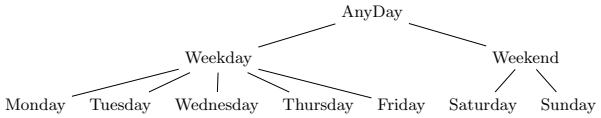

Figure 2: Intra-week temporal concept ontology

topics evolve throughout the year. Second, the lifestyle topics themselves will have a temporal component, which relates venues to times of the day and days of the week. For example, we may want to express that a given group of people tends to go to the pub in the evening on weekdays. For the latter purpose, we consider the discrete temporal units shown in Figures 1 and 2. Note that the boundaries of units such as Morning are necessarily somewhat arbitrary. Although we could learn appropriate boundaries from data (e.g., making the boundaries dependent on the geographic region), the definitions in Figures 1 and 2 will be sufficient to explain the main principles. The aim of these temporal units is to allow us to express information about the context of typical visits. This idea could be extended to capture weather information (e.g., allowing us to express that some group of people tends to go to the beach on sunny days) or particular types of recurring events (e.g., train strikes, bank holidays, school holidays, sales periods). Including such additional context factors offers scope for making the proposed model more useful in practice, but it does not affect any of the technical details.

In summary, we represent each check-in record as an abstract footprint, which is composed of a venue category, an intra-day time unit and an intra-week time unit. An example of an abstract footprint is `[Restaurant:(LunchTime, Weekday)]`.

3.2 Lifestyle Pattern Discovery

For each user, we construct a user activity document, consisting of the abstract footprints that correspond to the user's check-ins. In this section, we propose a novel Bayesian nonparametric graphical model for discovering lifestyle patterns from a given set of such user activity documents. Specifically, the model will discover lifestyle topics, which intuitively correspond to soft clusters of abstract footprints, and lifestyle patterns, which intuitively correspond to soft clusters of lifestyle topics. At the same time, lifestyle patterns will correspond to hard clusters of users. An example of a lifestyle topic is:

$$\{[\texttt{Office:(Weekday,Morning)}\ 0.5],\ [\texttt{Restaurant:(Weekday,LunchTime)}\ 0.2],\ [\texttt{Gym:(Weekday,DinnerTime)}\ 0.1], ...\}$$

where the probabilities reflect the proportions of check-ins that correspond to each abstract footprint. Note that lifestyle topics are similar to the topics that are considered in `LDA`, but instead of considering distributions over unigrams, we

consider distributions over abstract footprints. Lifestyle patterns capture the intuition that there will typically be several lifestyle topics that apply to a given user. For example, an office worker who enjoys going for a hike in the weekend may be modelled as a mixture of an "office-worker" topic and an "outdoor" topic. In the model, each user will be assigned to one of the lifestyle patterns. The output can thus be interpreted as a clustering of users, where each cluster corresponds to a lifestyle pattern.

Each user activity document corresponds to a single lifestyle pattern, but is associated with a mixture of different lifestyle topics. Therefore, the lifestyle topics are sampled from a Dirichlet process whereas hierarchical Dirichlet processes are used to generate samples from the lifestyle topics. To generate the abstract footprints of a user activity document, we first sample a lifestyle pattern. Given that lifestyle pattern, we then sample a sequence of lifestyle topics for the document, each time also sampling an abstract footprint from the lifestyle topic.

Figure 3 depicts the model in plate notation, where plates signify repetition of variables and latent variables are represented in unshaded circles. Note that our model captures various interactions of the user with respect to the location, time, activity, and comments all in a unified model. In addition, interactions between topic hierarchies are also modelled in the same graphical model, which makes our task more challenging. The number of repetitions for the inner larger plate corresponds to the number $|C|$ of abstract footprints in a given user's activity document, while the number of repetitions for the outer plate corresponds to the number of users $|U|$. The observed variables correspond to the abstract footprints f_i and their time stamp t_i.

The variable ω is the parameter or hyperparameter of the Dirichlet process associated with the lifestyle patterns. The variable α represents the distribution over lifestyle patterns. Each user corresponds to a specific lifestyle pattern l, which is drawn from the distribution α. The lifestyle pattern l is used to determine a probability distribution θ over lifestyle topics for the user. Specifically, θ is obtained from a Dirichlet process mixture, based on a corpus-wide distribution of lifestyle topics π and a vector τ_l of hyperparameters that is dependent on the chosen lifestyle pattern l. These hyperparameters are encoded for each lifestyle pattern l in the matrix τ. Note that because of the two-level topic structure, a matrix of hyperparameters is needed, as opposed to a vector in traditional `HDP` models. The variable γ is the hyperparameter of π. The distribution θ is subsequently used to assign a lifestyle topic z_i to each abstract footprint in the corresponding user activity document. The matrix Ψ defines each of the lifestyle topics as a probability distribution over abstract footprints. It depends on the vector ρ, whose dimension is equal to the number of footprints in the vocabulary. In particular, ρ determines a Dirichlet distribution, so a sample from ρ is a distribution over footprints.

The remaining variables of our model aim to capture how the prevalence of lifestyle topics evolves throughout the year. The observed variable t_i in Figure 3 corresponds to the time stamp associated with abstract footprint f_i, at a given level of granularity, e.g., t_i could be the month or season in which the check-in occurred. While we will only consider discrete time units in the analysis of the lifestyle pattern, assuming a continuous distribution in the underlying model allows for a more faithful analysis [10, 31, 29]. The discrete time component t_i of footprint f_i is modelled using a hierarchical Dirichlet process mixture of Gaussians, which is a distribution suitable to model multimodal variation on an unbounded timeframe. In addition, it can maintain tractable inference. To allow a flexible distribution over time, we use a Dirichlet process as the mixing measure. In particular, for each lifestyle topic z_i, a distribution μ_{z_i} of temporal components is obtained from a Dirichlet process with parameters ξ and κ, where κ depends on the hyperparameter λ. Each of these temporal components μ_{z_i} is associated with a Gaussian η_{z_i}, which has been sampled from a Normal-Inverse Gamma distribution (whose four parameters are denoted by ϵ in the graphical model for simplicity). Note that the components of the mixtures of Gaussians are thus shared among different lifestyle topics, which allows us to model correlations between the evolution of different topics.

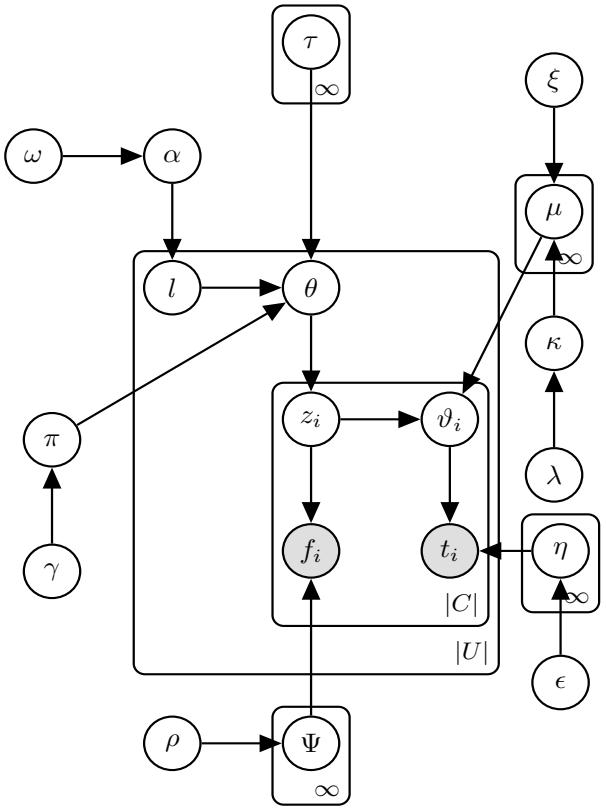

Figure 3: Graphical representation of our model.

The hyperparameters ω, γ, and τ are estimated from the data. To this end, we place hyper-hyperpriors or hyper-hyperparameters on these hyperparameters (not shown in Figure 3). The idea of using hyper-hyperpriors has already been adopted in some topic models such as [4, 27], where the aim is to find the posterior of the hyperparameters based on the data characteristics rather than explicitly specifying the hyperparameter values. However, we manually set weak hyperparameter values for the priors ρ, ξ, λ, and ϵ, as we found that automatically inferring the values of these parameters did not affect the results much in this case, while setting these values manually makes inference more efficient. The complete generative process of our model is as follows:

1. Draw a global base distribution over latent lifestyle patterns $\alpha|\omega \sim \mathbf{GEM}(\omega)$[1].
2. Draw a global base distribution over time component distributions $\kappa|\lambda \sim \mathbf{GEM}(\lambda)$
3. For each lifestyle topic $z = 1, 2, \cdots$
   (a) Draw a distribution over abstract footprints $\Psi_z|\rho \sim \mathbf{Dirichlet}(\rho)$
   (b) Draw a distribution over time components $\mu_z|\xi, \kappa \sim \mathbf{DP}(\xi, \kappa)$
4. For each time component $t = 1, 2, \cdots$
   (a) Draw a distribution over time $\eta_t|\epsilon \sim$ **Normal-inverse Gamma**(ϵ)
5. For each user document $u = 1, 2, \cdots, |U|$
   (a) Draw the lifestyle pattern $l|\omega \sim \omega$
   (b) Draw a distribution over lifestyle topics $\theta|l, \pi, \tau \sim \mathbf{DP}(\pi, \tau_l)$
   (c) For each abstract footprint $f_i, i = 1, 2, \cdots, |C|$ in the user activity document
      i. Draw the lifestyle topic $z_i|\theta \sim \theta$
      ii. Draw the abstract footprint $f_i|\Psi_{z_i} \sim \mathbf{Multinomial}(\Psi_{z_i})$
      iii. Draw a time component indicator $\vartheta_i|\mu_{z_i} \sim \mu_{z_i}$
      iv. Draw a time-stamp $t_i|\eta \sim \eta_{\vartheta_i}$

3.3 Posterior Inference

We use Gibbs sampling to perform posterior inference. As in some existing nonparametric topic models [10], we adopt a marginalization technique to speed up posterior inference, where we marginalize over the temporal distributions in order to sample lifestyle topics, and then sample the temporal distribution conditioned on the lifestyle topics. Note that this only has a minimal impact on the quality of the desired posterior [10]. In order to compute the posterior of the hyper-hyperpriors, we interleave the Metropolis Hastings (MH) steps between iterations of the Gibbs sampler to obtain new values of the hyper-hyperpriors. After some iterations of the sampler, these values are computed and updated and the sampler moves forward with the updated values. Computing the posterior of the hyper-hyperpriors using Gamma distributions has been discussed in [27].

In our model, there is an unbounded number of lifestyle patterns and an unbounded number of lifestyle topics. Note that because of the two-level topic structure, existing posterior inference methods such as those used in the `HDP` model or in [10] cannot be directly applied. In our model, the Dirichlet process which is responsible for generating the lifestyle patterns generates an infinite number of `HDP`s. Each of these `HDP`s is comprised of a time `HDP` and a lifestyle topic `HDP`, which are responsible for generating temporal assignments and lifestyle topics assignments respectively. The Gibbs sampling algorithm also has to take into account that topic distributions might change over time. Although the sampling mechanism which we propose is similar to sampling in other temporal topic models [29, 31, 10], an important difference is that each of the abstract footprints have their own time stamp. Therefore, our Gibbs sampler needs to sample lifestyle topics in the usual way, but in addition, it has to compute the change in the lifestyle topic patterns over time. This sampling model is also different from [35], which borrows the Gibbs sampling procedure of the `hLDA` model [4], where the sampler determines the topic hierarchies.

To enable Gibbs sampling, we need to derive the sampling equations to generate the lifestyle patterns and topics. For sampling the lifestyle patterns, we can use the Gibbs sampling method of the stick breaking construction of the Dirichlet process mixture model. Specifically, the method to find out the lifestyle patterns follows the Chinese Restaurant Process (CRP) scheme of the Dirichlet process mixture models. We refer to [11] for a detailed explanation of this standard sampling procedure. Note that we can only give a brief overview of the sampling equations of our model in this section. Some of the derivations are not shown here, as they can easily be derived from the derivations of the existing `HDP` model, as both models make an exchangeability assumption. For example, sampling the abstract footprint in a lifestyle topic, conditioned on time, closely resembles the `HDP` based sampling mechanism. The only difference is that the vector of hyperparameters τ_l depends on the lifestyle pattern l assigned to that document.

To generate the lifestyle topics given a lifestyle pattern l, we use a sampling mechanism with two types of `HDP`s, one for the abstract footprints and another for the temporal component, the latter being associated with a Gaussian distribution over time. Accordingly, each abstract footprint will have two assignments: a lifestyle topic assignment and a temporal assignment. In each iteration, both assignments need to be sampled.

For abstract footprints, we denote the global pool of lifestyle topics associated with lifestyle pattern l as $\mathbf{k}_l$. This global pool is reminiscent of the set of available dishes in the Chinese restaurant Franchise (CRF) analogy for the `HDP` model. In addition to the global pool of lifestyle topics, we have a pool of local lifestyle topic assignments $\mathbf{z}_l$, specifying a lifestyle topic z^u_{il} for each abstract footprint f^u_i in each user activity document u. This pool of assignments is reminiscent of the tables in the Chinese restaurant analogy. The mechanism is that we first sample a topic k_l from the global pool of lifestyle topics, and then k_l is added to one of the local assignments in the pool $\mathbf{z}_l$, which corresponds to assigning a lifestyle topic to one of the abstract footprints in the corresponding user activity document. Each local assignment can involve several topics k_l, as a user activity document is considered to be a mixture of different lifestyle topics.

Let $h^{\neg f^u_i}_{k_l}$ denote the conditional density associated with the abstract footprint f^u_i in a lifestyle pattern l, given the mixture component k_l and all lifestyle topic assignments except f^u_i. In other words, $h^{\neg f^u_i}_{k_l}(f^u_i)$ is the probability that f^u_i belongs to the topic k_l. Throughout this section, whenever we place the $\neg$ symbol, it means that we are ignoring that variable in the counts. Marginal counts are represented with dots. Let h denote the probability density function of H, where H is a base measure. Let g denote one of the components from $G(\Psi)$, which is the emission distribution. As in the usual topic modeling approach, H can be assumed to be a Dirichlet distribution over the vocabulary and G

[1]Recall that **GEM** is a one-parameter stochastic process obtained from the $\mathbf{Beta}(a, b)$ distribution where $a = 1$ [24].

is a topic-word multinomial distribution. We estimate the probability $h_{k_l}^{\neg f_i^u}(f_i^u)$ as follows:

$$\frac{\int h_l(f_i^u|\Psi_{k_l})\prod_{k_l'}\prod_{(u',i')\neq(u,i),z_{i'l}^{u'}=k_l'} h_l(f_{i'}^{u'}|\Psi_{k_l'})g(\Psi_{k_l'})\mathrm{d}\Psi_{k_l'}}{\int\prod_{k_l'}\prod_{(u',i')\neq(u,i),z_{i'l}^{u'}=k_l'} h_l(f_{i'}^{u'}|\Psi_{k_l'})g(\Psi_{k_l'})\mathrm{d}\Psi_{k_l'}} \tag{1}$$

where z_{il}^u is the lifestyle topic that has been assigned to the abstract footprint f_i^u for a lifestyle pattern l. It can be shown that:

$$h_{k_l}^{\neg f_i^u}(f_i^u) = \frac{\rho + n_{lzv}^{\neg ui}}{V\rho + n_{lz.}^{\neg ui}} \tag{2}$$

where n_{lzv} is the number of times an abstract footprint v has been sampled in a topic $z = k_l$ and $n_{lz.}$ denotes the number of abstract footprints belonging to the topic $z = k_l$. The form of $g(\Psi_{k_l})$ is as follows:

$$g(\Psi_{k_l}) = \text{Dirichlet}(\rho) = \frac{\Gamma(\sum_{v=1}^V \rho_v)}{\prod_{v=1}^V \Gamma(\rho_v)}\prod_{v=1}^V \Psi_{lzv}^{\rho_v - 1} \tag{3}$$

where Γ denotes the Gamma function. We also need to derive the link between the pool of local lifestyle topic assignments $\mathbf{z}_l$ and the global pool of lifestyle topics $\mathbf{k}_l$. Let $\mathbf{f}_u^z$ be the set of all abstract footprints assigned with lifestyle topic z in the user activity document u. Changing the topic assignment of all abstract footprints in $\mathbf{f}_u^z$ changes the component membership of all data items associated with that topic. Let us write $h_{k_l}^{\neg \mathbf{f}_u^z}(\mathbf{f}_u^z)$ for the likelihood of the assignment of $\mathbf{f}_u^z$ to topic z, given all data items associated with mixture component k_l leaving out $\mathbf{f}_u^z$. It is evaluated as follows:

$$\frac{\int\prod_{k_l}\prod_{f_i^u\in\mathbf{f}_u^z} h_l(f_i^u|\Psi_{k_l})\prod_{f_{i'}^{u'}\notin\mathbf{f}_u^z, z_{i'l}^{u'}=k_l} h_l(f_{i'}^{u'}|\Psi_{k_l})g(\Psi_{k_l})\mathrm{d}\Psi_{k_l}}{\prod_{f_{i'}^{u'}\notin\mathbf{f}_u^z, z_{i'l}^{u'}=k_l} h_l(f_{i'}^{u'}|\Psi_{k_l})g(\Psi_{k_l})\mathrm{d}\Psi_{k_l}} \tag{4}$$

It can be shown that the above equation can be written as:

$$\frac{\prod_{v=1}^V \Gamma(\rho + n_{lzv})}{\Gamma(V\rho + n_{lz.})}\cdot\frac{\Gamma(V\rho + n_{lz.}^{\neg f_i^u})}{\prod_v \Gamma(\rho + n_{lzv}^{\neg f_i^u})} \tag{5}$$

where n_{lzv} is the number of times an abstract footprint v has been sampled in a topic $z = k_l$, where z is one of the topics from the global pool of lifestyle topics $\mathbf{k}_l$ associated with lifestyle pattern l; $n_{lzv}^{\neg f_i^u}$ denotes the number of times an abstract footprint v has been sampled excluding the current abstract footprint for user u from the count, V denotes the number of unique abstract footprints in the entire dataset and $n_{lk.}$ denotes the number of abstract footprints associated with lifestyle topic $z = k_l$. Let $n_{lzv}^{\neg uz}$ denote the number of abstract footprints with $z = k_l$ as one of the topics from the global pool of lifestyle topics $\mathbf{k}_l$ associated with lifestyle pattern l excluding the current topic assignment.

In order to sample the temporal dynamics of the lifestyle topics, we define the conditional distribution of the lifestyle topics given the temporal information. This definition is as follows:

$$P(z_i^u|\mathbf{z}^{\neg ui}, \mathbf{k}, \mathbf{t}, l)P(f_i^u, t_i^u|z_i^u, \mathbf{f}^{\neg ui}, \mathbf{t}^{\neg ui}, \mathbf{z}^{\neg ui}, \boldsymbol{\vartheta}^{\neg ui}, l) \tag{6}$$

Let m_{luk} denote the number of lifestyle topics associated with k_l in the user activity document u. Let n_{luk} denote the number of abstract footprints associated with the global topic k_l in u. Let $m_{l..}$ denote the total number of lifestyle topics associated with k_l in all user activity documents. Let $\mathbf{r}$ denote the global pool of time components. Let R denote the number of time related components which can increase or decrease. Let $c_{lk\vartheta}$ denote the number of abstract footprints associated with lifestyle topic k_l drawn from the global pool and time component ϑ. Let e_r denote the total number of time components associated with r. Then (6) can be written as:

$$\begin{cases} n_{luk}^{\neg ui}\frac{\rho + n_{lzv}^{\neg ui}}{V\rho + n_{lz.}^{\neg ui}}\sum_{\vartheta_i}\frac{c_{lk\vartheta}^{\neg ui}}{c_{lk.}^{\neg ui}+\xi}S_r^{\mathbf{t}}(t_i^u) + \\ \frac{\xi}{c_{lk.}^{\neg ui}+\xi}\left(\sum_{r=1}^R \frac{e_r}{e_.+\lambda}S_r^{\mathbf{t}^{\neg ui}}(t_i^u) + \frac{\lambda}{e_.+\lambda}S_{r_{\text{new}}}(t_i^u)\right) \\ \text{if it is an existing lifestyle topic} \\ \tau_l\left(\sum_{k_l=1}^K \frac{m_{luk}}{m_{l..}+\gamma}\frac{\rho + n_{lzv}^{\neg ui}}{V\rho + n_{lz.}^{\neg ui}} + \frac{\gamma}{m_{l..}+\gamma}\frac{\Gamma(V\rho)}{\prod_v\Gamma(\rho)}\frac{\Gamma(\rho+1)\prod_{w\neq v}\Gamma(\rho)}{\Gamma(V\rho+1)}\right) \\ \sum_{\vartheta_i}\frac{c_{lk\vartheta}^{\neg ui}}{c_{lk.}^{\neg ui}+\xi}S_r^{\mathbf{t}}(t_i^u) + \frac{\xi}{c_{lk.}^{\neg ui}+\xi}\left(\sum_{r=1}^R \frac{e_r}{e_.+\lambda}S_r^{\mathbf{t}^{\neg ui}}(t_i^u) + \right. \\ \left.\frac{\lambda}{e_.+\lambda}S_{r_{\text{new}}}(t_i^u)\right) \text{ if a new lifestyle topic is drawn} \end{cases}$$

Finally, $S_r^{\mathbf{t}^{\neg ui}}(t_i^u)$ is the posterior predictive time distribution, defined as follows:

$$\int_0^\infty \frac{1}{\overline{\sigma}}\exp(-\frac{1}{2\overline{\sigma}^2}n(\overline{\mu}-\overline{x})^2)\overline{\sigma}^{-\nu-2}\exp(-\frac{\nu s^2}{2\overline{\sigma}^2})\mathrm{d}\overline{\sigma}^2 \tag{7}$$

where $\overline{\mu}$ is the uninformative location mean prior, $\overline{\sigma}$ is the uninformative scale prior, $\overline{x}$ is the sample mean, n is the total number of sample points, ν is the number of degrees of freedom and s is the sample variance.

In order to resample the component memberships of all the data items associated with the lifestyle topic z, we marginalize over the time assignments of all the abstract footprints, as the sampler will otherwise be too slow to converge. In order to speed up this process, we adopt a selective block based procedure where we conduct approximation only for a subset of the abstract footprints associated with that lifestyle topic. We write the estimates as $P_{\text{est}}(\mathbf{t})$, which will be used to estimate the true Gibbs sampling probabilities. The process can be written as:

$$P(k_{luz=\text{new}}|z_i^u, \mathbf{f}^{\neg uk}, \mathbf{t}^{\neg uk}, \mathbf{z}^{\neg uk}, \boldsymbol{\vartheta}^{\neg uk}, l) \propto$$
$$\begin{cases} m_{l.k}^{\neg uz=k_l}\frac{\prod_{v=1}^V\Gamma(\rho+n_{lkv}^{\neg uz=k_l})}{\Gamma(V\rho+n_{lk.}^{\neg uz=k_l})}\cdot\frac{\Gamma(V\rho+n_{lk.}^{\neg uz=k_l})}{\prod_v\Gamma(\rho+n_{lkv}^{\neg uz=k_l})}P_{\text{est}}(\mathbf{t}) & k_l \text{ is existing} \\ \gamma\frac{\Gamma(V\rho)\prod_v\Gamma(\rho+n_{l.v}^{\neg uz=k_{\text{new}}})}{\Gamma(V\rho+n_{l..}^{\neg uz=k_{\text{new}}})\prod_v\Gamma(\rho)}P_{\text{est}}(\mathbf{t}) & k_l = k_{\text{new}} \end{cases} \tag{8}$$

4. EXPERIMENTAL RESULTS

In this section we present a number of qualitative and quantitative results. In particular, we provide several examples of lifestyle patterns that have been discovered by our method (Section 4.1), and we illustrate how the model can be used to analyse the temporal variation in the importance of lifestyle topics. We present quantitative results related to lifestyle pattern quality discovery (Section 4.2). We then present an analysis of the required running time (Section

4.3). We also quantitatively compare our method against state-of-the-art comparative models on a time stamp prediction task (Section 4.4).

The raw dataset used in the experiments has been obtained from the lead author of [8], who initially obtained it from Foursquare check-ins reported on Twitter over a period of five months. This dataset contains check-in data, including any associated user comments, but only contains the IDs of the venues. Using these IDs, we have crawled additional venue description data from Foursquare using its public API, collecting the tags, categories, name and location of each venue. We refer interested readers to find the statistics and additional details about the dataset from [8]. We have considered the following countries in our experiments: USA (approximately 8 million check-ins, 93,501 users and 879,476 venues), India (approximately 800,000 check-ins, 2220 users and 12,277 venues), Singapore (approximately 500,000 check-ins, 1207 users and 18,082 venues), Hong Kong (approximately 400,000 check-ins, 3788 users and 5282 venues), Australia (approximately 1 million check-ins, 1000 users and 30,880 venues) and Indonesia (approximately 6 million check-ins, 69,805 users and 302,725 venues). Note that the friendship graph that was used in [35] is no longer publicly available, due to a change in the sharing policies adopted by the content providers.

In all experiments, the Gibbs sampler for our model is run for 1000 iterations, which we empirically found was sufficient for the parameters of the model to converge. We have set $\rho = 0.01$, $\xi = 0.02$, $\epsilon = 0.01$, and $\lambda = 0.05$ which represent weak prior values as we have found experimentally that selecting different weak prior values did not have much impact on the results. The following Gamma priors have been placed on the hyperparameter values: $\omega = \text{Gamma}(1.0, 1.0)$, $\tau = \text{Gamma}(1.0, 10.0)$ and $\gamma = \text{Gamma}(1.0, 1.0)$. While these priors have parameters, such hyper-hyperparameters have a much weaker impact on the inference results than fixing the original hyperparameters for these distributions [4, 3].

4.1 Illustrative Example

Figure 4 displays some of the lifestyle patterns that have been discovered from the USA check-in data by our approach. In the figures, each dot corresponds to a lifestyle pattern while each text box shows the most prominent abstract footprints of a given lifestyle topic. Arrows in the diagram indicate which lifestyle topics are associated with which lifestyle patterns. Since different lifestyle patterns can share the same lifestyle topics, overlap between lifestyle patterns is made clear. To generate these diagrams from the graphical model, we have adopted the following procedure. For each lifestyle pattern x, we rank the lifestyle topics y according to the corresponding Dirichlet parameter τ_{xy}. After the model is trained, it outputs the change in topic distributions over time. We can use this information to present results based on any desired temporal granularity level, such as seasons, by doing a post-hoc analysis. For example, some lifestyle topics will have a high probability in some seasons and a much lower probability in other seasons. Figure 4 highlights some topics which are specific to summer and some topics which are specific to winter.

The lifestyle patterns in Figure 4 intuitively correspond to students, researchers and office workers. The figure makes explicit the common aspects of their lifestyles, e.g., every

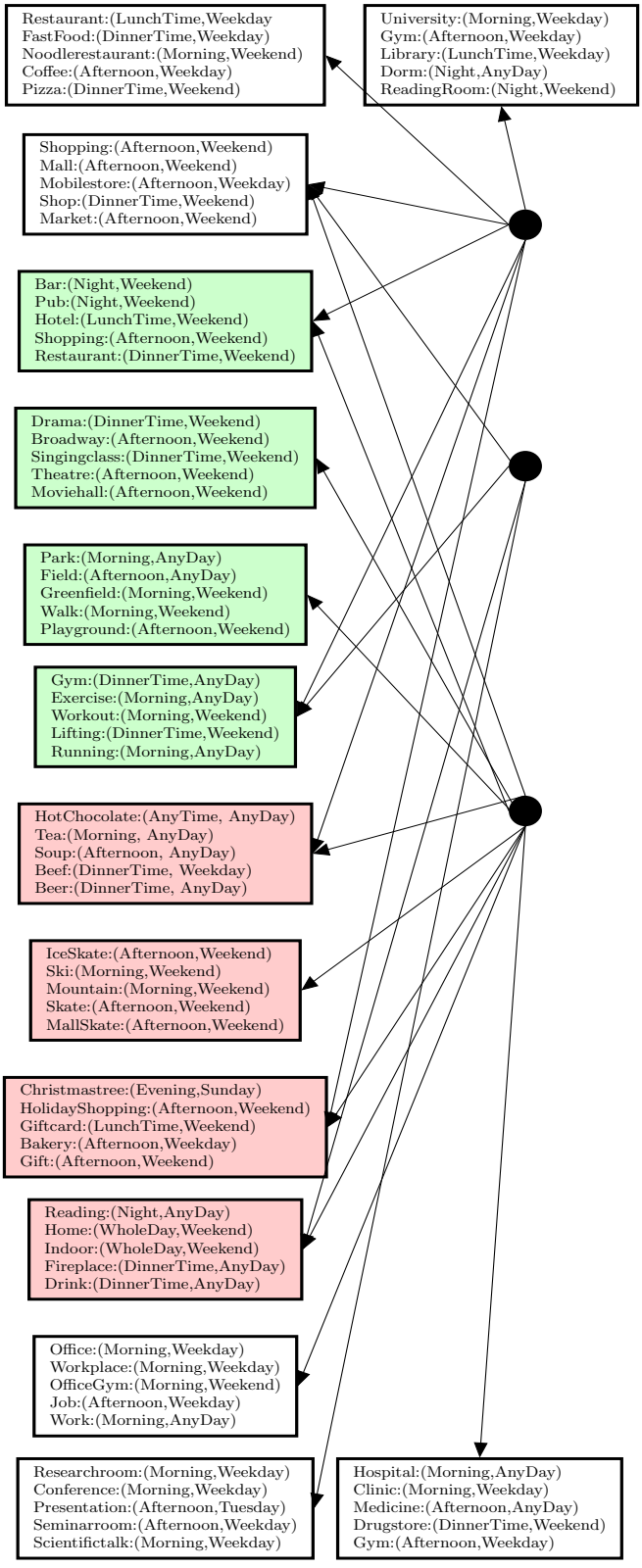

Figure 4: Combined lifestyle patterns of some people in the USA. In the figure, the coloured light green rectangles represent the living patterns in summer and coloured light red rectangles represent the living patterns during the winter season.

group is associated with a shopping related lifestyle topic. Furthermore, the summer and winter specific topics clearly show how people's lifestyles change based on the seasons, e.g., researchers and office workers go to the park in summer while rather staying at home and reading a book in winter. Other winter specific topics are related to hot drinks, winter sports and Christmas shopping.

In [35], lifestyle patterns are modelled as branches of a tree, where common topics appear towards the top of the tree, while topics that are specific to particular groups appear further down. While this tree representation captures commonalities between different lifestyles in a natural way, it relies on the rather restrictive assumption that a single hierarchical clustering of individuals can be found that explains all commonalities between people's lifestyle patterns. In contrast, in our model commonalities are modelled by having topics that are connected to many lifestyle patterns. This allows us to model various forms of overlap between groups of users. For example, a university related topic may apply to both students and professors (but not office workers), while an office environment related topic may apply to both professors and office workers (but not students). A single tree representation would not be able to capture both types of commonality.

4.2 Lifestyle Pattern Quality Evaluation

We evaluate the quality of the discovered lifestyle patterns by comparing our model with temporal as well as non-temporal topic models. One recent work that can discover lifestyle patterns is the one described in [35], which uses the `hLDA` model to generate a lifestyle spectrum (i.e. a tree of lifestyle patterns). We cannot directly apply the full model from [35] because we do not have access to the friendship connections it is based on. However, we can compare the lifestyle pattern quality of our model with the `hLDA` component of their model. As comparative models, we also use Latent Dirichlet Allocation (`LDA`) [6], the Biterm topic model (`BTM`), which is a topic model suited for short texts [33], `PAM` model [20], `nPAM` model [19], the topics over time model (`TOT`), described in [31], and the nonparametric topics over time model (`npTOT`), described in [10]. These latter two models can be regarded as the state-of-the-art temporal topic models. Note that using these comparative models for our task required us to make some adaptations. In particular, the original `TOT` model assumes that each document has a unique time stamp, which applies to all the words in that document. Our user activity documents, on the other hand, contain different footprints, each with their own time stamp, i.e., the time stamp associated with words is not constant for all words of the document. We modified the Gibbs samplers for the `TOT` and `npTOT` models to allow each word to be associated with a different time stamp. To analyse the usefulness of using a mixture of Gaussians, we also consider a variant of our model where the mixture of Gaussians is replaced by a single Gaussian distribution. We call this model `Variant` in our experimental results. Publicly shared implementations have been used for all comparative models, except for `npTOT` and `nPAM`.

We use the perplexity metric to quantify the lifestyle pattern quality. In order to compute the perplexity for our model we followed the technical details from [30]. Lower perplexity results represent better generalization ability of the model. In this experiment, the original data was randomly split into training and testing sets. We learn the parameters of the model using the training data (75%), and report the perplexity results on the held-out data (25%). For the parametric topic models, we use a tuning set to determine the number of topics following the tuning procedure described in [13]. Our objective is to compare how well our model has learned all parameters and how it performs in terms of its generalization ability.

Table 1: Perplexity results for the raw dataset.

	`LDA`	`BTM`	`PAM`	`nPAM`	`TOT`	`npTOT`	`hLDA`	`Variant`	`Our`
US	32374	49832	23450	31245	23982	18934	45938	21758	**13029**
Ind	198	215	212	254	173	154	213	153	**132**
Sin	213	324	255	301	243	199	243	176	**154**
HK	156	214	201	205	143	140	197	134	**121**
Aus	5483	6847	5223	5558	5321	4382	6785	5201	**4219**
Indo	15032	20543	18503	21504	18914	13754	24543	20101	**12433**

Table 2: Perplexity results for the abstract footprint dataset.

	`LDA`	`BTM`	`PAM`	`nPAM`	`TOT`	`npTOT`	`hLDA`	`Variant`	`Our`
US	31832	54321	22251	20432	21514	15983	42874	21203	**12908**
Ind	143	178	155	163	154	143	175	**119**	121
Sin	209	301	201	167	198	154	214	216	**122**
HK	123	212	175	134	127	132	154	123	**109**
Aus	4322	5487	5111	4858	4274	4132	5187	4154	**4091**
Indo	13493	18732	17233	20876	15786	11876	21435	21492	**11234**

Tables 1 and 2 report the perplexity results for the raw and abstract footprint data respectively. The non-temporal topic models do not use the timestamps in the user activity document during parameter estimation. We apply them in a traditional setting where the entire user document is the input to the model.

From the results in Tables 1 and 2, we can see that our model has the lowest perplexity on the held-out data in most of the datasets. It consistently performs better in the raw dataset showing its robustness in handling noisy data. For the abstract footprint data in Table 2, our model also performs best overall, although the `Variant` model is slightly better in the India dataset. We also observe that using abstract footprints improves the perplexity of most models, with the `Variant` model on the Singapore and Indonesia datasets being an exception. This clearly shows the usefulness of the abstract footprints for dealing with sparsity, regardless of the specific model being used.

4.3 Running Time Comparison

In this section, we compare the running time of our method with a number of existing methods. We present results in terms of the number of CPU hours spent on generating the topic model (based on the abstract footprints). We used a single-threaded implementation in the C programming language for all models. The models were run on an Intel Core i7-3820 3.6GHz machine with 64GB of primary memory. The number of iterations of the Gibbs sampler was set to 1000 in all cases.

As the results in Table 3 show, the latent topic representations of our model can be found considerably faster than those of the `hLDA` model. In this comparison, we applied `hLDA` with three levels, which we consider an absolute minimum for describing lifestyle patterns. The running time of the `hLDA` model, however, depends crucially on the number of levels, and the model quickly becomes intractable as this number is increased. The reason why our model generates latent representations faster than `hLDA` is that we have a

Table 3: Running time performance in CPU hours.

	US	Ind	Sin	HK	Aus	Indo
hLDA	5.50	0.55	1.02	0.23	3.45	4.32
nPAM	6.02	1.03	0.55	0.31	2.33	4.11
npTOT	3.24	0.54	0.56	0.20	1.45	2.23
Variant	2.25	0.29	0.35	0.14	1.33	2.11
Our	2.29	0.28	0.35	0.14	1.32	2.11

much simpler CRF representation, which speeds up the posterior inference computation. The hierarchical tree generation, which arranges topics based on their commonalities also takes up computational resources, whereas similar commonalities can easily be obtained as a post-hoc analysis.

Table 3 also shows the running time of npTOT, which is again slower than that of our model. In particular, the npTOT model took a considerable amount of time to estimate the temporal distribution of every abstract footprint in the document, as it was not designed with this kind of information in mind. The Variant model, which is a simpler version derived from our model, matches the running speed of our model and is in some cases slightly better.

4.4 Time stamp Prediction

In this experiment we consider time stamp prediction, where the objective is to predict the month of a given check-in, given its abstract footprint. We quantitatively compare our model with TOT and npTOT. Existing non-temporal topic models such LDA and its proposed extensions cannot solve this task because they are not designed to handle different time stamps in the same document and they do not incorporate time stamp related meta data in their graphical model. We have split the collection of user activity documents for each country into a training set (approximately 75% of the data) and a test set. The model is learned using the training set, where the time stamps are visible. In the test set, we hide the time stamps and try to predict the correct month. We did not randomly split this data, but the testing set has later timestamps, where we intend to predict the future events given the past.

To select the number of topics for the TOT model, we used 25% of the training set as a tuning set and trained the model on the remaining 75% of the training set. We then selected the number of topics that led to the highest accuracy. Using the learned model we can predict the time stamp of an abstract footprint by choosing the time stamp the maximizes the posterior.

We measure the performance of the models using three metrics described in [31]. Specifically, we use the L1 error measure, denoted as L1 in our results, and its expected value, denoted as E(L1). We also compute accuracy. The L1 and E(L1) error measures are to be minimized while accuracy is to be maximized. Table 4 shows the micro-average results for two different configurations: one configuration in which the abstract footprints were used (as elsewhere throughout this paper) and one configuration where the actual venue was used instead of the latent venue category. In this way, we can see to what extent the use of abstract venue categories helps each of the models. The results clearly show that our model outperforms the comparative models. The results again show the usefulness of the abstract footprints as a mechanism for addressing sparsity. The better performance of our model can be explained by the fact that it has more parameters, which helps it fit the data better. Note that the issues of overfitting and underfitting are avoided because the nonparametric nature of our model means that it automatically chooses the complexity of the model based on the data characteristics. It can outperform npTOT, which is also a nonparametric model, because our model generates two levels of topics, which enables it to better exploit the correlations between different topics.

The considered time stamp prediction task is a very hard task, which we believe cannot be adequately solved by simple methods. Even for humans, for most check-in records we can only guess in which month they have taken place, but there are some check-ins where reasonable estimates can be made (e.g., check-ins at a beach are more likely to be in summer). We certainly don't claim that our method can solve this task. However, we do believe that it allows us to quantitatively show that our model captures temporal trends in a better way than the baselines. The fact that the accuracy scores are only marginally better than random guessing is simply a consequence of the fact that for many venue types, the distribution over months is fairly uniform.

The aim of the perplexity and time stamp tasks is simply to quantitatively show that our model is capturing certain aspects of lifestyle patterns better than existing methods. The fact that we have to use these tasks, rather than tasks which would directly evaluate the quality of the lifestyle patterns, is because this notion of quality is so subjective, and because testing the effectiveness of deploying such a model in real-world applications (e.g. advertising campaigns) would be extremely challenging for which labeled data sets are difficult to obtain.

5. CONCLUSIONS

We have proposed a text mining framework for analysing the lifestyles of users of location-based social networks such as Foursquare. In particular, our framework is based on a novel topic modelling approach, in which we explicitly address the sparsity of check-in data and incorporate a temporal component for analysing how lifestyle patterns change throughout the year. The output of our method consists of a set of lifestyle patterns, each of which corresponds to a probability distribution over lifestyle topics, the latter intuitively corresponding to soft clusters of related venues. The nonparametric nature of our model means that we do not have to specify the number of lifestyle patterns and topics a priori. Our experimental results show that useful lifestyle patterns can indeed be obtained in this way, and that the model can capture the dynamics of lifestyle topics more faithfully than existing topic models.

6. REFERENCES

[1] A. Agarwal, N. R. Desai, R. Ruffoli, and A. Carpi. Lifestyle and testicular dysfunction: a brief update. *Biomedicine & Pharmacotherapy*, 62(8):550–553, 2008.

[2] S. Bauer, A. Noulas, D. O. Séaghdha, S. Clark, and C. Mascolo. Talking places: Modelling and analysing linguistic content in Foursquare. In *SocialCom*, pages 348–357, 2012.

[3] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. John Wiley & Sons, 1994 (ISBN: 0-471-92416-4).

[4] D. M. Blei, T. L. Griffiths, and M. I. Jordan. The nested Chinese restaurant process and Bayesian

Table 4: Time stamp prediction performance.

	Raw Footprints												Abstract Footprints											
	TOT			npTOT			Variant			Our			TOT			npTOT			Variant			Our		
	L1	E(L1)	ACC.	L1	E(L1)	ACC.	L1	E(L1)	ACC.	L1	E(L1)	ACC.	L1	E(L1)	ACC.	L1	E(L1)	ACC.	L1	E(L1)	ACC.	L1	E(L1)	ACC.
US	3.12	3.42	0.23	2.87	3.01	0.26	2.65	2.83	0.29	2.54	2.59	0.24	2.98	3.01	0.29	2.85	2.92	0.28	2.58	2.89	0.25	2.43	2.39	0.39
Ind	2.28	2.31	0.24	2.13	2.34	0.26	2.09	2.29	0.29	1.89	2.04	0.24	2.11	2.15	0.26	2.14	2.33	0.28	2.14	2.16	0.29	1.77	1.89	0.39
Sin	3.45	3.54	0.24	3.21	3.25	0.28	2.61	2.62	0.29	2.55	2.54	0.21	3.44	3.45	0.28	3.09	3.19	0.29	2.88	2.91	0.34	2.31	2.45	0.34
HK	1.99	2.33	0.23	1.98	2.11	0.31	1.72	1.67	0.31	1.54	1.59	0.21	1.87	1.92	0.31	1.78	2.19	0.33	1.58	1.65	0.31	1.49	1.59	0.31
Aus	3.12	3.45	0.25	2.91	2.95	0.26	2.92	2.96	0.34	2.86	2.92	0.24	3.1	3.15	0.31	2.87	3.01	0.31	2.89	3.01	0.33	2.85	3.01	0.36
Indo	2.11	2.31	0.29	2.15	2.21	0.32	2.11	2.19	0.36	2.01	2.11	0.26	2.02	2.19	0.31	2.01	2.19	0.33	2.00	2.21	0.34	1.99	2.06	0.39

nonparametric inference of topic hierarchies. *JACM*, 57(2):7, 2010.

[5] D. M. Blei and J. D. Lafferty. Dynamic topic models. In *ICML*, pages 113–120, 2006.

[6] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *JMLR*, 3:993–1022, 2003.

[7] J. Chang and D. M. Blei. Relational topic models for document networks. In *AISTATS*, pages 81–88, 2009.

[8] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui. Exploring millions of footprints in location sharing services. *ICWSM*, 2011:81–88, 2011.

[9] Y. Dingqi. *Understanding Human Dynamics from Large-Scale Location-Centric Social Media Data: Analysis and Applications*. PhD thesis, 2015.

[10] A. Dubey, A. Hefny, S. Williamson, and E. P. Xing. A nonparametric mixture model for topic modeling over time. In *SDM*, pages 530–538, 2013.

[11] H. Ishwaran and L. F. James. Gibbs sampling methods for stick-breaking priors. *JASA*, 96(453), 2001.

[12] S. Jameel and W. Lam. An unsupervised topic segmentation model incorporating word order. In *SIGIR*, pages 203–212, 2013.

[13] S. Jameel, W. Lam, and L. Bing. Supervised topic models with word order structure for document classification and retrieval learning. *Information Retrieval Journal*, 18(4):283–330, 2015.

[14] E. Kim, H. Ihm, and S.-H. Myaeng. Topic-based place semantics discovered from microblogging text messages. In *WWW*, pages 561–562, 2014.

[15] K. B. Kirk and J. J. Thomas. The lifestyle project. *Journal of Geoscience Education*, 51(5):496–499, 2003.

[16] F. Kling and A. Pozdnoukhov. When a city tells a story: Urban topic analysis. In *SIGSPATIAL*, pages 482–485, 2012.

[17] J. H. Lau, K. Grieser, D. Newman, and T. Baldwin. Automatic labelling of topic models. In *ACL-HLT*, pages 1536–1545, 2011.

[18] J. H. Lau, D. Newman, S. Karimi, and T. Baldwin. Best topic word selection for topic labelling. In *ACL*, pages 605–613, 2010.

[19] W. Li, D. Blei, and A. McCallum. Nonparametric Bayes Pachinko allocation. 2007.

[20] W. Li and A. McCallum. Pachinko allocation: DAG-structured mixture models of topic correlations. In *ICML*, pages 577–584, 2006.

[21] Q. Mei, X. Shen, and C. Zhai. Automatic labeling of multinomial topic models. In *KDD*, pages 490–499, 2007.

[22] R. D. Michman, E. M. Mazze, and A. J. Greco. *Lifestyle marketing: reaching the new American consumer*. Greenwood Publishing Group, 2003.

[23] A. Noulas, B. Shaw, R. Lambiotte, and C. Mascolo. Topological properties and temporal dynamics of place networks in urban environments. In *WWW*, pages 431–441, 2015.

[24] J. Pitman. Poisson–Dirichlet and GEM invariant distributions for split-and-merge transformations of an interval partition. *CPC*, 11(05):501–514, 2002.

[25] M. M. Shafiei and E. E. Milios. Latent Dirichlet co-clustering. In *ICDM*, pages 542–551, 2006.

[26] D. Shepard. Nonparametric Bayes Pachinko allocation for super-event detection in Twitter. In *TENCON*, pages 1–5, 2014.

[27] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet processes. *JASA*, 101(476), 2006.

[28] J. Vosecky, D. Jiang, K. W.-T. Leung, K. Xing, and W. Ng. Integrating social and auxiliary semantics for multifaceted topic modeling in Twitter. *ACM TOIT*, 14(4):27:1–27:24, 2014.

[29] D. D. Walker, K. Seppi, and E. K. Ringger. Topics over nonparametric time: A supervised topic model using Bayesian nonparametric density estimation. In *UAI*, 2012.

[30] H. M. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno. Evaluation methods for topic models. In *ICML*, pages 1105–1112, 2009.

[31] X. Wang and A. McCallum. Topics over time: A non-Markov continuous-time model of topical trends. In *KDD*, pages 424–433, 2006.

[32] L. Wu, Z. S. Ye Zhi, and Y. Liu. Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS ONE*, 9(5):e97010:993–1022, 2014.

[33] X. Yan, J. Guo, Y. Lan, and X. Cheng. A biterm topic model for short texts. In *WWW*, pages 1445–1456, 2013.

[34] H. Yin, B. Cui, L. Chen, Z. Hu, and X. Zhou. Dynamic user modeling in social media systems. *ACM TOIS*, 33(3):10:1–10:44, 2015.

[35] N. J. Yuan, F. Zhang, D. Lian, K. Zheng, S. Yu, and X. Xie. We know how you live: exploring the spectrum of urban lifestyles. In *OSN*, pages 3–14, 2013.

[36] Q. Yuan, G. Cong, K. Zhao, Z. Ma, and A. Sun. Who, Where, When, and What: A nonparametric Bayesian approach to context-aware recommendation and search for Twitter users. *TOIS*, 33(1):2:1–2:33, 2015.

[37] E. Zavitsanos, G. Paliouras, and G. A. Vouros. Non-parametric estimation of topic hierarchies from texts with hierarchical Dirichlet processes. *JMLR*, 12:2749–2775, 2011.

[38] X. Zhou and L. Chen. Event detection over Twitter social media streams. *VLDBJ*, 23(3):381–400, 2014.