

Sentiment-Based Topic Suggestion for Micro-Reviews

Ziyu Lu*, Nikos Mamoulis[‡], Evaggelia Pitoura[‡] and Panayiotis Tsaparas[‡]

* The University of Hong Kong
zylu@cs.hku.hk

[‡] University of Ioannina
{nikos,pitoura,tsap}@cs.uoi.gr

Abstract

Location-based social sites, such as Foursquare or Yelp, are gaining increasing popularity. These sites allow users to check in at venues and leave a short commentary in the form of a *micro-review*. Micro-reviews are rich in content as they offer a distilled and concise account of user experience. In this paper we consider the problem of predicting the topic of a micro-review by a user who visits a new venue. Such a prediction can help users make informed decisions, and also help venue owners personalize users' experiences. However, topic modeling for micro-reviews is particularly difficult, due to their short and fragmented nature. We address this issue using pooling strategies, which aggregate micro-reviews at the venue or user level, and we propose novel probabilistic models based on Latent Dirichlet Allocation (LDA) for extracting the topics related to a user-venue pair. Our best topic model integrates influences from both venue inherent properties and user preferences, considering at the same the sentiment orientation of the users. Experimental results on real datasets demonstrate the superiority of this model compared to simpler models and previous work; they also show that venue-inherent properties have higher influence on the topics of micro-reviews.

Introduction

In the past few years, location-based social sites, such as Foursquare¹, Yelp² and Facebook places³ have emerged as prime online destinations, where users can record their footprints via check-ins, as well as their experience through *micro-reviews*. Micro-reviews, or tips, accompany a check-in at a venue, and they contain a short commentary on the venue. Tips may offer information about the venue, opinions on what is good, or advice for new customers. They are very targeted and concise, and they provide a distilled account of the experience of the users in the venue. They are a fast-growing corpus, and they have recently attracted considerable research interest (Aggarwal, Almeida, and Kumaraguru 2013; Moraes et al. 2013; Nguyen, Lauw, and Tsaparas 2015) for the rich content they contain.

In this paper, we consider the problem of predicting the topic that a user will comment on in her tip when checking in to a new venue. This is a problem of great practical importance for both users and venues. For venues, knowing in advance the aspect of the venue that a user will most likely focus on, allows them to offer a personalized experience to the user. For example, if we can predict that when visiting a restaurant a specific user is likely to comment on the quality of the service, the manager of the venue can make sure to fine-tune their service in order to meet the customer's needs. Furthermore, exposing the topic prediction to the users provides fine-grain information about the venue. For example, for a user who is a wine enthusiast, recommending a restaurant, and predicting that the user is likely to comment on the extensive wine selection of the place adds texture and context to the recommendation.

Similar problems have been considered in the context of full-text reviews, where the goal is to estimate the rating of specific aspects of venues (Wang, Lu, and Zhai 2011; Hai et al. 2014; Wang, Lu, and Zhai 2010; Moghaddam and Ester 2011). These approaches rely on generative models that extract latent topical aspects and their ratings. However, applying such techniques to micro-reviews is not straightforward. Micro-reviews have special characteristics, very different from those of reviews. First, they are very short (up to 200 characters), and provide very limited context and few word co-occurrences (Lin et al. 2014). Second, due to the length limitation, the expression is very dense, and the text is often fragmented and poorly structured. Finally, micro-reviews often contain diverse pieces of information stitched together in a few sentences, resulting in incoherent semantics. These characteristics make opinion mining and analysis on micro-reviews harder compared to the full-text reviews. To the best of our knowledge, we are the first to consider the problem of topic prediction for micro-reviews.

In order to tackle the limited and incomplete characteristics of micro-reviews, we use *pooling* strategies for document collaborative filtering (Mehrotra et al. 2013; Weng et al. 2010) and *contextualization* (Tang, Zhang, and Mei 2013) to integrate multiple contexts. We aggregate all micro-reviews of a venue and a user into a single venue-document and user-document respectively, and we consider probabilistic topic models on the aggregated documents for the problem of topic prediction. We first define two simple mod-

Copyright © 2016, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹<https://foursquare.com/>

²<http://www.yelp.com/>

³<https://www.facebook.com/places/>

els that apply Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003) individually on the venue and user documents. Building on these models, we propose two novel models, Authority-LDA (ALDA) and Authority-Sentiment LDA (ASLDA), which integrate both venue inherent properties and user personalized preferences. In the ASLDA model we add a sentiment layer to cluster topics into different sentiment groups based on the assumption that users might have personalized sentiment orientation (e.g. some users tend to be negative), and they are more likely to comment on aspects (topics) that match their sentiment orientation (e.g., negative users will tend to comment on negative aspects). The ASLDA model can predict not just the topic of a future tip, but also the most probable sentiment orientation. We evaluate the proposed models on datasets from two real location-based social sites Foursquare and Yelp. Experimental results show that our methods outperform other competitor approaches.

In summary, in this paper we make the following contributions:

- We define the problem of topic prediction for micro-reviews when a user checks in an unvisited venue. To the best of our knowledge, this is the first work that deals with topic prediction and suggestion in the context of micro-reviews at location-based social sites.
- We define four probabilistic models for the problem, including two novel probabilistic models that leverage both venue inherent aspects and user personalized preferences. The Authority-Sentiment LDA (ASLDA) model introduces a novel way to incorporate the influences from the user sentiment orientation in the topic prediction, and it is able to predict the sentiment orientation of a future tip.
- We evaluate the proposed methods for topic prediction on four datasets from two real location-based social sites Foursquare and Yelp, and compare against other approaches.

Related Work

Micro-reviews is a relatively new corpus that only recently drew the attention of the research community. There is work on micro-reviews on spam detection (Aggarwal, Almeida, and Kumaraguru 2013), polarity analysis (Moraes et al. 2013), micro-review summarization (Nguyen, Lauw, and Tsaparas 2015). To the best of our knowledge we are the first to consider the problem of topic prediction for micro-reviews.

Topic modeling algorithms have been widely adopted in text mining (Blei, Ng, and Jordan 2003; Rosen-Zvi et al. 2004). One of the first such models, proposed by Blei et al. (Blei, Ng, and Jordan 2003), is Latent Dirichlet allocation (LDA). Many topic models based on LDA have been developed to address review mining problems. For example, Moghaddam and Ester (Moghaddam and Ester 2011) introduced an Interdependent Latent Dirichlet Allocation (ILDA) model to infer latent aspects and their ratings for online product reviews. Lin and He (Lin and He 2009) proposed a joint sentiment-topic model (JST) for sentiment analysis of movie reviews, by extending LDA with a new sentiment

layer. JST is based on the assumption that topic generation depends on sentiments, and word generation depends on sentiment-topic pairs.

However, (Lin et al. 2014) showed that the characteristics of short text reduce the effectiveness of topic modeling methods. Micro-reviews in location-based social sites are very short, and have a relatively small vocabulary and a broad range of topics. The probability of word co-occurrence in the micro-reviews is very small, compromising the performance of topic models originally designed for long reviews. In order to address this data sparsity problem, heuristics such as *document pooling* (Mehrotra et al. 2013; Weng et al. 2010) or *contextualization* (Tang, Zhang, and Mei 2013) have been proposed to improve the performance of topic modeling on short text. For instance, Mehrotra et al. (Mehrotra et al. 2013) proposed to aggregate all documents by the same author or all documents with specific hashtags and form pooling documents on which topic modeling can be applied effectively. Contextualized topic models are proposed to integrate particular types of contexts into classical models like LDA, by introducing either additional layers to the topic model (Jo and Oh 2011; Rosen-Zvi et al. 2004; Lin and He 2009) or use a coin-flipping selection process to select among contexts (Paul and Girju 2010; Tang, Zhang, and Mei 2013; Zhao et al. 2011). The author-topic model (AT) of (Rosen-Zvi et al. 2004), which utilizes authorship information for modeling scientific publications, can also be viewed as a contextualized topic model. Tang et al. (Tang, Zhang, and Mei 2013) proposed a model, which formulates different types of contexts as multiple views of the partition of the corpus and uses voting to determine consensus topics.

Our models adopt both pooling methods and contextualization in order to facilitate topic discovery for micro-reviews in location-based social sites. We aggregate micro-reviews on the same venue or micro-reviews by the same author to construct aggregated pooling documents. To the aggregated documents, we add additional context such as authority information and the sentiment orientation of users to improve latent topic learning. Prior work on sentiment-topic models (Lin and He 2009; Moghaddam and Ester 2011) introduced sentiment as a latent (unknown) variable based on some assumptions of the dependencies between sentiment variable and topic variable, and then jointly learned the sentiments and topics. On the other hand, in our work, we assume that the sentiment information is observed (from an existing sentiment lexicon for short text) and utilize the sentiment orientation of users to enhance the process of topic discovery.

Problem Definition

In this section we introduce some terminology and define our problem.

A location-based social site consists of a set of users, a set of venues, and a collection of micro-reviews. Formally, we will use V to denote the set of venues in the site, and A to denote the set of all users (authors) in the site. A *tip* t is a short piece of text written by a user $u \in A$ for a venue $v \in V$. A tip is defined over a vocabulary W consisting of the union of all terms that appear in all tips in our collection. We

assume that stop-words have been removed and that terms have been stemmed. We define a *micro-review* $r = \langle u, v, t \rangle$ as a triplet consisting of a user $u \in U$, a venue v , and a tip t that was left from the user u about the venue v . We use R to denote the set of all micro-reviews.

As tips consist of short text, studying them individually is not very informative. We thus use pooling methods to construct aggregated documents for a venue or a user. For a venue v , we use A_v to define the set of all users that have written a micro-review for venue v , and R_v to define the collection of all micro-reviews for venue v . We use d_v to denote the *venue-document* defined by taking the union of all the tips in R_v . We use $W_d = \{w_1, w_2, \dots, w_m\}$ to denote the vocabulary of the document d_v . In a symmetric fashion, for a user u we define the venue set V_u , the micro-review set R_u , the *user-document* d_u and the vocabulary W_u .

Given a collection of (user or venue) documents $D = \{d_1, d_2, \dots\}$, using topic-modeling techniques we can extract a set of K *latent topics* $Z = \{z_1, \dots, z_k\}$. Each topic z_i is defined as a distribution over the vocabulary W . Our goal is given user-venue pair (u, v) for which there is currently no micro-review r , to predict the latent topic of that tip. Formally, we define our problem as follows.

Problem Definition: Given social site consisting of a collection of users A , venues V , and micro-reviews R , a set of latent topics Z , a user-venue pair (u, v) , and a number N , identify a set of N latent topics $Z_N \subset Z$ that the user u is most likely to comment on about venue v .

Proposed models

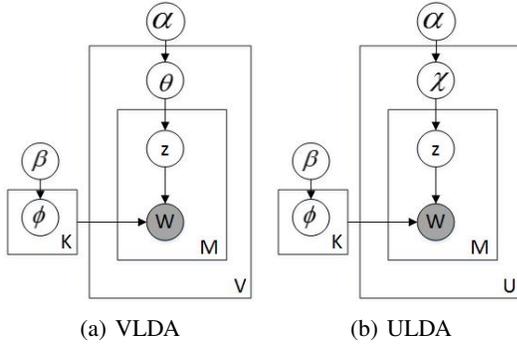


Figure 1: LDA models

LDA model

The LDA model was proposed in (Blei, Ng, and Jordan 2003). It represents each document as a multinomial distribution over K latent topics, and each topic as a multinomial distribution over terms (words). We applied LDA on our pooling documents (venue or user documents). For each type (venue/user) of document collection, we derived two distributions. Using the venue document collection, we derived a venue-topic distribution, and a topic-word distribution. This model is denoted as Venue-LDA model (VLDA). Similarly, using the user document collection, we extracted a

user-topic distribution and a topic-word distribution to form a User-LDA model (ULDA). Each of these two models captures the different influences that the venues or users, respectively, have on the topics of the tip to be given by the target user to the target venue. The graphical representations of VLDA and ULDA are shown in Figure 1. In the diagrams, M is the number of terms in a venue/user document, V is the number of venues, U is the number of users, and K is the number of topics (aspects) z .

In both models, ϕ is the topic-term distribution; θ is the venue-topic distribution, and χ is the user-topic distribution in Figure 1(a) and Figure 1(b), respectively; α and β are prior parameters. The generative process of LDA models on venue documents (VLDA) or user documents (ULDA) are as follows:

1. Sample θ (VLDA) or χ (ULDA) from Dirichlet priors $Dir(\alpha)$.
2. For each topic z , sample ϕ_z from Dirichlet priors β_z .
3. For each term w_{di} in the (venue or user) document d ,
 - draw a topic z_{di} from $Dir(\theta_d)$ (VLDA) or $Dir(\chi_d)$ (ULDA)
 - draw a term w_{di} from $Dir(\phi_{z_{di}})$

Parameter Estimation. The probability of document collection D is defined as follows:

$$p(D|\alpha, \beta) = \prod_{d=1}^N \int p(\theta|\alpha) \left(\prod_{m=1}^M \sum_z p(z|\theta) p(w_{dm}|z, \beta) \right) d\theta_d$$

For the venue document collection, each document is a venue document and $N = V$. For the user document collection, each document is a user document and $N = U$, and θ is replaced by χ .

We use Gibbs sampling (Griffiths and Steyvers 2004) to perform approximate inference, and to estimate the unknown parameters $\{\theta, \phi\}$. The conditional distribution for Gibbs sampling is as follows:

$$p(z_{di} | \mathbf{z}^{-di}, \mathbf{w}, d) \propto \frac{n_{dz_{di}}^{-di} + \alpha_{z_{di}}}{\sum_z (n_{dz}^{-di} + \alpha_z)} \times \frac{n_{z_{di}w_{di}}^{-di} + \beta_{w_{di}}}{\sum_w (n_{z_{di}w}^{-di} + \beta_w)}$$

where n_{dz} is the number of times that topic z has been sampled from the multinomial distribution to the document d . n_{zw} is the number of times that term w has been sampled to topic z . A superscript (e.g., $-di$), denotes a quantity, excluding the specified instance (e.g. the i_{th} word in document d).

After sampling for a sufficient number of iterations, θ or χ and ϕ are calculated as follows:

$$\begin{aligned} \hat{\theta}_{vz} &= \frac{n_{vz} + \alpha_z}{\sum_{z'} n_{vz'} + \alpha_{z'}} & \hat{\chi}_{uz} &= \frac{n_{uz} + \alpha_z}{\sum_{z'} n_{uz'} + \alpha_{z'}} \\ \hat{\phi}_{zw} &= \frac{n_{zw} + \beta_w}{\sum_{w'} n_{zw'} + \beta_{w'}} \end{aligned} \quad (1)$$

Authority-LDA model

VLDA and ULDA models only consider influences from one side (either venue or user). We propose a new Authority-LDA model (ALDA) which integrates influences from both

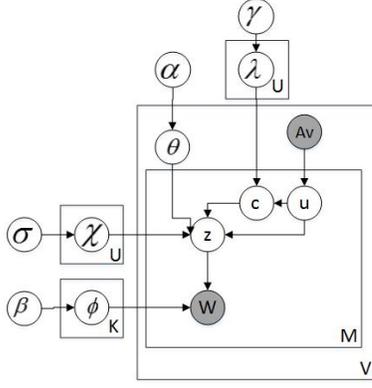


Figure 2: Authority-LDA model

Table 1: Notations used in ALDA

Symbols	Description
V, U	the number of venues, the number of users
K	the number of topics
M	the number of words in a document
A_v	the set of users for the venue document d_v
u, z	the user variable, the latent topic variable
c	the switch variable
θ_v	distribution of topics specific to venue v
χ_u	distribution of topics specific to user u
ϕ_z	distribution of words specific to topic z
α, σ, β	Dirichlet priors for θ, χ, ϕ
λ_u	the parameter of Bernoulli distribution specific to user u for sampling the binary switch c
γ	Beta prior for λ , where $\gamma = \{\gamma, \gamma'\}$

users and venues. ALDA employs venue-wise pooling to construct venue documents for the representation of venue influences and leverages the author information to represent user influences as in Author-Topic (AT) modeling (Rosen-Zvi et al. 2004). Then the latent topics depend on both the inherent aspects of venues and personal preferences of users.

We use a mixing parameter λ to control the weights of influence from both sides. The parameter λ is follows a Bernoulli distribution which samples a binary variable c that switches between influence from venue inherent aspects and user preferences. In other words, when a user u comments on a venue, we assume that the tip is influenced by the user's personal preferences with probability λ_u ($c=1$) and by the inherent aspects of venue with probability $1-\lambda_u$ ($c=0$). The latent topics are still multinomial distributions over terms. Figure 2 shows a graphical representation of the ALDA model, while Table 1 summarizes the symbols used in ALDA.

Note that the latent topics z depend on both the venue-topic distribution θ and the user-topic distribution χ . A_v is the set of users who have written micro-reviews about venue v , namely the authority users for the venue document v . ϕ represents topic-term distribution. M is the number of terms in the venue documents, V is the number of venues, U is the number of users, and K is the number of topics (aspects) z . The generative process of the proposed Authority-LDA

model is as follows:

- For each topic z , draw ϕ_z from $Dir(\beta)$.
- For each user u ,
 - Draw χ_u from $Dir(\sigma)$ and λ_u from $Beta(\gamma)$
- For each venue document d_v ,
 - Draw θ_v from $Dir(\alpha)$
 - For each term $w_{d_v,i}$ in the venue document d_v ,
 - Draw a user u from A_v uniformly,
 - Draw switch $c \sim Bernoulli(\lambda_u)$
 - If $c=0$
 - Draw a topic $z_{d_v,i}$ from $Dir(\theta_v)$
 - If $c=1$
 - Draw a topic $z_{d_v,i}$ from $Dir(\chi_u)$
 - draw a term $w_{d_v,i}$ from $Dir(\phi_{z_{d_v,i}})$

Parameter Estimation. We also use Gibbs sampling to estimate the unknown parameters $\{\theta, \chi, \phi, \lambda\}$. In the Gibbs sampling procedure, we first compute the posterior distribution on u, c and z and then estimate $\{\theta, \chi, \phi, \lambda\}$. The posterior distribution of the hidden variables for each word $w_{d_v,i}$ is calculated as follows:

$$P(u_{d_v,i} = u, z_{d_v,i} = z, c_{d_v,i} = 1 | \mathbf{u}^{-d_v,i}, \mathbf{z}^{-d_v,i}, \mathbf{c}^{-d_v,i}, \mathbf{w}, A_v) \propto \frac{n_{uc}^{-d_v,i}(1) + \gamma}{n_{uc}^{-d_v,i} + \gamma + \gamma'} \times \frac{n_{uz}^{-d_v,i} + \sigma}{\sum_{z'} (n_{uz'}^{-d_v,i} + \sigma)} \times \frac{n_{zw}^{-d_v,i} + \beta}{\sum_{w'} (n_{zw'}^{-d_v,i} + \beta)}$$

$$P(u_{d_v,i} = u, z_{d_v,i} = z, c_{d_v,i} = 0 | \mathbf{u}^{-d_v,i}, \mathbf{z}^{-d_v,i}, \mathbf{c}^{-d_v,i}, \mathbf{w}, A_v) \propto \frac{n_{uc}^{-d_v,i}(0) + \gamma'}{n_{uc}^{-d_v,i} + \gamma + \gamma'} \times \frac{n_{uz}^{-d_v,i} + \alpha}{\sum_{z'} (n_{uz'}^{-d_v,i} + \alpha)} \times \frac{n_{zw}^{-d_v,i} + \beta}{\sum_{w'} (n_{zw'}^{-d_v,i} + \beta)}$$

where $n_{uc}(1)$ and $n_{uc}(0)$ is the number of times that $c=1$ and $c=0$, respectively, has been sampled for user u . n_{uc} equals $n_{uc}(1) + n_{uc}(0)$. n_{vz} is the number of times that topic z has been sampled from the distribution θ_v specific to venue v and n_{uz} is the number of times that topic z has been sampled from the distribution χ_u . n_{zw} is the number of times that term w has been sampled from the distribution ϕ_z specific to topic z . Superscript $-d_v,i$ again denotes a quantity excluding the current instance d_v,i .

After Gibbs sampling, $\{\theta, \chi, \phi, \lambda\}$ can be estimated as follows:

$$\hat{\theta}_{vz} = \frac{n_{vz} + \alpha_z}{\sum_{z'} n_{vz'} + \alpha_z} \quad \hat{\chi}_{uz} = \frac{n_{uz} + \sigma_z}{\sum_{z'} n_{uz'} + \sigma_z}$$

$$\hat{\phi}_{zw} = \frac{n_{zw} + \beta_w}{\sum_{w'} n_{zw'} + \beta_w} \quad \hat{\lambda}_u = \frac{n_{uc}(1) + \gamma}{n_{uc} + \gamma + \gamma'} \quad (2)$$

Authority-Sentiment-LDA model

Quite frequently, the commenting behavior of users is affected by their sentiment. For example, there exist *negative users* who tend to comment on negative aspects of products; at the same time they do not bother to post their opinions for positive or neutral aspects. Similarly for *positive users*. Motivated by this observation, we label the users based on their *sentiment orientation*, i.e., tendency to give positive or negative comments. We then predict the tip aspects for the target venue, taking into consideration the user sentiments.

on θ_v and χ_u . The probability distribution of the detected topics given a query pair (v, u) can then be computed as follows:

$$p(\mathbf{z}|u, v) \propto \lambda_u \chi_u + (1 - \lambda_u) \theta_v \quad (5)$$

λ_u is a personalized parameter to weight the influences from the venue v and user u as defined in Equation 2.

Suggestion by ASLDA

The Authority-sentiment-LDA model considers the sentiment orientation of users and identify topics from the venues’ inherent properties and the users’ sentiment-based commenting preferences. The detected topics are interdependently influenced by the venue-topic distribution θ and the user-topic distribution which is inferred from user-sentiment distribution χ and sentiment-topic distribution π . Given a query pair (v, u) , the probability of the detected topics can be estimated as follows:

$$p(\mathbf{z}|u, v) \propto \lambda_u \sum_s \chi_{u,s} \pi_s + (1 - \lambda_u) \theta_v \quad (6)$$

where λ_u is a smoothing parameter to weight the influences from the venue and user as in Equation 3.

In addition, after getting the predicted topic z , we can obtain the corresponding sentiment orientation probability distribution π_z . Sorted on the probability values, we can use the sentiment orientation s with highest probability π_{sz} as the most probable sentiment orientation.

Experiments

Data

We used datasets from two popular sites (Foursquare and Yelp). The first two are extracted from the public Yelp tip dataset⁴. We split the venues in this dataset by state (there are very few records for each city, therefore we did not choose to split at the city level). Then, we picked the two states which have the most records: AZ (Arizona) and NV (Nevada). We also constructed two datasets from Foursquare, by crawling user-venue-tip micro-reviews records from two cities, respectively SF (san francisco) and NY (New York).

We preprocessed the micro-reviews in the four datasets, by replacing special symbols in them, tokenizing tips, stemming and removing stopwords⁵. We also cleaned terms which have less than three occurrences in the dataset, as well as venues and users which have only one record. Table 2 shows the statistics of the four datasets, after these preprocessing and cleaning steps.

Table 2: Statistics (after preprocessing)

Dataset	SF	NY	Yelp-AZ	Yelp-NV
number of venues $ V $	10476	9668	6200	5640
number of users $ A $	5118	5315	7351	8970
number of terms $ W $	21705	21261	5617	6621

⁴http://www.yelp.com/dataset_challenge

⁵<http://jmlr.org/papers/volume5/lewis04a/all-smart-stop-list/english.stop>

For each dataset, we mark 80% of the data as the training set and use the remaining user-venue-tip data as test set. Then, for each dataset, we use pooling strategies to generate the venue document collection and user document collection by aggregating all tips for a venue/user in training set as a venue/user document (see the definition of venue/user documents in Problem Definition).

Compared Methods

The list of compared methods is as follows:

- **Base LDA models:** These are the VLDA and ULDA models. VLDA applies LDA on the venue document collection (Figure 1(a)) while ULDA applies on the user document collection (Figure 1(b)).
- **Authority-LDA model (ALDA):** ALDA considers influences from both venue inherent characteristics and user personalized preferences. A personalized mixture parameter λ_u has been learned to weight the two influences.
- **Authority-sentiment-LDA model (ASLDA):** This model uses a sentiment layer to enhance ALDA. The intuition is that users have personalized sentiment-orientated commenting behaviors and they trend to comment on sentiment oriented aspects.
- **Author topic models (AT):** The author-topic model was proposed in (Rosen-Zvi et al. 2004) to identify the latent topics from author behaviors.
- **CA-LDA:** CA-LDA (content-aware LDA) is proposed in (Yin et al. 2013). The model is originally defined to recommend points of interest (POIs) based on content and location context, given a target (user,city) pair. Different from other POI recommendation works, this model accepts a pair of inputs, e.g. a user and a city. Therefore, we can use the CA-LDA model via replacing the query location variable by the query venue.
- **mLDA:** mLDA (Tang, Zhang, and Mei 2013) is a contextualized model to discovery consensus topics from multiple contexts. The type of contexts used are venue and user.
- **JST:** JST is a joint sentiment-topic model to identify the sentiment-topic pairs in (Lin and He 2009). The difference between JST and our ASLDA model is that JST adds the sentiment as a latent variable, and learns the sentiment-topic pairs jointly. For completeness, we tested two versions of JST: vJST is applied on venue documents and uJST is applied on user documents.

For JST, CA-LDA and mLDA, we use the same parameter settings as those used in the papers where these methods were introduced (i.e., (Yin et al. 2013), (Lin and He 2009) and (Tang, Zhang, and Mei 2013)). In particular, for both vJST and uJST, $\alpha = 50/K$, $\beta = 0.01$, while γ is set to 0.01 for positive sentiment labels and 5 for negative sentiment. For CA-LDA, $\alpha = \alpha' = 50/K$, $\beta = 0.01$, $\gamma = \gamma' = 0.5$. For mLDA, $\gamma = 1$, $\delta = 1$ and $\eta = 0.01$. Likewise, in models VLDA, ULDA, AT, ALDA and ASLDA, parameters $\alpha, \sigma, \eta, \beta, \gamma$ were set to their typical values ($\alpha = \sigma = \eta = 50/K$, $\beta = 0.01$, and $\gamma = \gamma' = 0.5$), as

in previous work (Rosen-Zvi et al. 2004; Yin et al. 2013; Lin and He 2009).

Evaluation metrics

We use two quantitative metrics for model evaluation.

Perplexity measures how well the model fits the test data and evaluates the predictive power of a topic model. A lower perplexity score indicates stronger predictive power. For a test set of micro-reviews R_{test} , the perplexity is as follows:

$$perplexity(R_{test}) = \exp \left\{ -\frac{\sum_{r \in R_{test}} \log P(\mathbf{w}_r)}{\sum_{r \in R_{test}} M_r} \right\}$$

where M_r is the number of words in micro-review r .

Topic Coherence: PMI (point-wise mutual information) measures the semantic coherence of learned topics (Newman et al. 2010); the PMI score for the given topic z is calculated as the average relatedness of each pair of words in the set of top M representative words of a given topic:

$$PMI(z) = \frac{2}{M \times (M - 1)} \sum_{1 \leq i < j \leq M} \log \frac{p(w_i, w_j)}{p(w_i)p(w_j)}$$

where $p(w)$ is the probability of word w to appear in the corpus and $p(w_i, w_j)$ is the joint probability of w_i and w_j . For each method, we average over PMI scores for all the learned topics. For the shown results, we set $M=10$, namely the top 10 probable words. Higher PMI values are better since they indicate a coherent topic.

Experimental Results

Figure 4 shows the perplexity scores of all methods on the four datasets. Observe that our proposed models have lower perplexity results in all cases. ASLDA has significantly lower perplexity compared to all other methods, while ALDA is marginally better than VLDA and clearly better than the remaining models. When comparing the base LDA models, which consider influences from either venue inherent properties or user preferences, VLDA has better performance than ULDA. This indicates that venue-inherent properties have higher influence on the topics of micro-reviews. Still, models that consider both venue-inherent properties and user personalized preferences (i.e., ASLDA and ALDA) outperform base LDA models. This demonstrates that topics of micro-reviews are related with both venue inherent properties and user personalized preferences, when a user is commenting on a venue. Considering only one of the two factors is not sufficient for accurate prediction.

Note that both vJST and uJST do not have good performance. vJST has higher perplexity scores than VLDA and uJST has the highest perplexity scores compared to all methods. This indicates that introducing an additional latent sentiment layer into topic models is not appropriate for micro-reviews because it escalates the data sparsity problem. Table 3 shows the PMI scores of all models for all datasets for $K = 10$ topics. PMI measures topic coherence of the learned topics; higher PMI values are better. Observe that the PMI results also show that ASLDA and ALDA outperform the other approaches. The superiority of ASLDA is not

Table 3: PMI scores for four datasets

Method	SF	NY	Yelp-AZ	Yelp-NV
ASLDA	3.16	2.55	3.11	3.22
ALDA	2.91	2.49	3.03	2.91
VLDA	2.89	2.47	2.86	2.88
ULDA	2.86	2.42	2.79	2.85
AT	2.67	2.42	2.84	2.73
CA-LDA	2.80	2.44	2.72	2.82
mLDA	2.66	2.49	2.43	2.70
vJST	2.22	2.31	2.20	2.25
uJST	2.19	2.20	2.03	2.31

as prominent here as in the case of perplexity scores (about 10% improvement over base models VLDA and ULDA). When changing the number of topics, the relative performance of the models in terms of PMI scores is similar (i.e., ASLDA is always superior).

Results for Topic Suggestion

In this section, we manually compare some of the results by ASLDA, ALDA, VLDA and ULDA, in order to gain more insights into their performance. Refer to the prediction formulae (4, 5 and 6) for all these four models. We consider user-venue pairs from the test data (not used for training), and we compare the top topic suggested by the different methods against the actual tip left by the user for the venue.

We use Foursquare-NY dataset at $K=10$ to generate the results that follow. Tables 4 and 5 show for different user-venue pairs the actual tip left by the user for this venue, and the top topic id predicted by the different methods based on the scores from the prediction functions. In order to understand the semantics of the different topics we list the top-10 words for example topics in Tables 6 and 7. The words are sorted based on the topic-word probabilities which are appropriately normalized as in (Srivastava and Sahami 2009) to prevent common words from ranking very high. Table 6 lists the top-10 words of each suggested topic by ASLDA and ALDA, while Table 7 lists the top-10 words of each suggested topic by ULDA and VLDA.

In order to investigate the influence from the venue-side and the user-side, we give some examples for the same venue with different users as venue-based examples in Table 4. The examples for the same user at different venues are user-based examples in Table 5. There are two example venues (venues 3158 and 3164). For venue 3158, there are three test tips from 3 different users (users 4410, 1591, and 1853). There are two test tips for venue 3164 from user 745 and user 4440. In Table 5, we list the test tips for user 4078, and all his/her tips for four different venues.

Our first observation is that **(1) the suggested topics based on venue inherent properties (VLDA) are different from those based on user preferences (ULDA). In general, the tip topics are more related to the venue inherent properties.** As VLDA model considers only venue effects, the suggested results are the same for a venue ignoring user personalization. Similarly, the results of ULDA for a user are the same for different venues. In Table 4, venue 3158 is a bar and venue 3164 is also a nightlife place (club). The sug-

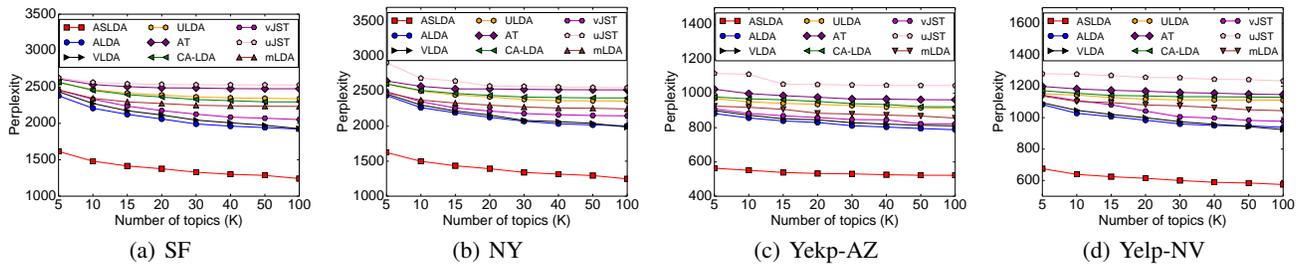


Figure 4: Perplexity results

Table 4: Venue-based examples for topic suggestion

vid	uid	Tip content	ASLDA	ALDA	VLDA	ULDA
3158	4410	Order food from Bianca’s next door and they’ll bring it to you at the bar, like it was table service.	9	8	0	1
	1591	Sparkling wine: Blanc De Blanc. Cocktail that tastes like sprite: French75.	9	8	0	7
	1853	go downstairs to party. stay upstairs to chill.	5+	8	0	4
3164	745	Love the glass slipper and the Robert Johnston. Strong drinks!!! Love	5+	8	0	8
	4440	To get to this swanky lounge Stormy.	5+	5	0	8

Table 5: User examples for topic suggestion

uid	vid	Tip content	ASLDA	ALDA	VLDA	ULDA
4078	590	Everything’s good here. For the best value, stick with 64oz Lagrrr. You can spend big \$\$\$ in this place.	5+	5	7	8
	4157	Very underwhelming. The “Korean Burger” is topped with pico de galo. All of the ingredients tasted very cheap.	0-	3	1	8
	8039	Avoid this place, they are a money grab.	0-	5	2	8
	7038	Good quality but pricey for what you get. You’ll need an extra slice on top of whatever you regularly order.	0-	5	6	8

Table 6: Top-10 words for example topics in ALDA and ASLDA

	ASLDA			ALDA		
0	5	9	3	5	8	
pricey	place	night	chees	place	night	
food	love	bartend	burger	staff	bar	
neighbourhood	price	music	sandwich	service	music	
service	service	live	fri	price	room	
bad	staff	bar	pizza	nice	live	
money	food	beer	food	food	awesome	
pizza	good	happy-hour	spot	time	cocktail	
space	awesome	cocktail	chicken	bad	space	
money	speical	table	bacon	friendly	happy-hour	
place	parti	play	delicis	people	drink	

gested results from VLDA are all from topic 0 (nightlife). However, the results from ULDA are very different. For 3158 they are about fastfood (topic 1 of ULDA) and deserts (topic 7 of ULDA). In Table 5, we see the suggested results from ULDA for user 4078. ULDA suggests topic 8 (food

and service) which is relatively general, indicating that user 4078 often comments on food and service of a venue. The results of VLDA are more specific, but they do not always capture the tip content. For example, for venue 8039, the tip is “Avoid this place, they are a money crab”. The suggested topic of VLDA is 2, due to the word “avoid”, which is in the top-10 words of topic 2 for VLDA (Table 7). However, this topic does not appropriately capture the content of the tip which is about the price.

This leads us to our second observation that **(2) both venue inherent and user preferences are essential for topic suggestions**. ASLDA and ALDA can perform weighted integration of effects from both venue properties and user preferences leading to significant improvement in the results. For example, for the tip “Good quality but pricey for what you get. You need an extra slice on top of whatever you regularly order:” in Table 5 both ULDA and VLDA output general topics that fail to capture that the tip is about price and value. The topics of ASLDA and ALDA capture this concept. Furthermore, the parameter λ allows to personalize the results per user. In Table 4, ASLDA and ALDA output different results for different users for the same venue.

Table 7: Top-10 words for example topics in VLDA and ULDA

VLDA					ULDA			
0	1	2	6	7	1	4	7	8
free	chees	wait	pizza	check	order	free	good	food
night	amaz	time	good	shop	chicken	beer	ice	place
music	brunch	better	wine	store	fri	bar	chocol	good
bar	egg	bad	nyc	day	pork	night	definit	service
watch	order	long	excel	find	sandwich	drink	cake	nice
Sunday	bacon	advoid	favorit	buy	taco	happy-hour	thing	staff
crowd	menu	seat	slice	close	soup	bartend	love	price
room	sandwich	bathroom	small	people	menu	music	bread	bad
live	grill	leave	spoy	bring	rice	good	recommend	friendli
weekend	delici	call	city	play	burger	order	sweet	pretti

In addition, **(3) ASLDA and ALDA produce better topics**. The suggested topics are more specific, and better capture the content of the tips. For example, the suggested topic from ALDA for the test tips in Table 4 is topic 8, which includes words belonging to “nightlife”, but also includes words that do not appear in the VLDA result, such as, “cocktail” and “happy-hour”. The superior performance of ALDA and ALSDA, which integrate both venue inherent properties and user preferences implies the essence of considering both these factors.

Moreover, **(4) users have sentiment-oriented topic preferences**. ASLDA adds the sentiment layer to model user sentiment orientation and finds topics also based on this factor. In our experiments, for each listed topic, we used its highest probability value of sentiment π_z as its possible sentiment orientation. Topic numbers with “+” indicate that the sentiment orientation of the topic is more likely to be positive while “-” means that the sentiment orientation of the topic is more likely to be negative. The absence of a sign next to a topic number indicates that the sentiment orientation is neutral. For example, most of the tips by user 4078 in Table 5 are about price and food. Most of the opinions by this user for price at different venues are negative, like “pricy” and “expensive”, mapping to the topic 0-. On the other hand, the first positive tip can be mapped to the 5+ topic. This demonstrates that our ASLDA model can find different sentiment orientations toward different topics. The results also indicate that the user is concerned about price, and she is budget-conscious. Also, 3 of 4 tips from the user are negative. This indicates that the user belongs to the class of “negative users”, who are more likely to comment on negative points in their preferences.

Table 8: Sentiment Classification Accuracy (%)

Dataset	ASLDA	vJST	uJST
SF	40.38	36.07	39.73
NY	53.13	40.63	52.59

We also conducted a sentiment classification experiment on ASLDA and JST to evaluate the sentiment analysis performance of ASLDA as in (Lin and He 2009). Given a query pair (v, u) , we define the predicted sentiment label of ASLDA as the sentiment orientation s with the high-

est sentiment orientation probability π_{sz} for the topic z obtained from Equation 6. In addition, to compare with uJST and vJST, we define the suggested sentiment label to be the sentiment s with the highest conditional probability $p(s|d)$ given the document u and v respectively. We collected manual labels for the test tips from human labelers as ground truth. Foursquare datasets are used as examples to show the sentiment classification accuracy results in Table 8. We can see that ASLDA outperforms both vJST and uJST. This indicates that our proposed ASLDA model has better performance in sentiment analysis. In addition, uJST (user documents) works better than vJST (venue documents), showing that sentiment orientation of micro-reviews is more affected by the user side, in line with our assumption in ASLDA. In order to measure the statistical significance of our methods, we performed paired t-test for the predicted sentiment labels of ASLDA and uJST (best version of JST). The p-values are both less than 0.05 ($1.72 e^{-7}$ for SF dataset and 0.016 for NY dataset).

The ability of ASLDA to measure user sentiment-oriented topics can be very useful in practice. The results of the model can facilitate personalization of the user experience. For instance, we can observe that user 4078 is a “negative user” with respect to the price aspect. Thus, it is better to recommend to the user restaurants for lower or medium budgets (and ordinary food categories). As another example, for a new customer who has booked hotel for the first time, the hotel manager can use ASLDA to predict the user’ sentiment-oriented topics for the hotel. Then the hotel manager can try to improve the aspects that the customer is expected to rate negatively (e.g. offer the customer a quiet room at the same price if the customer tends to give negative comments on noise).

Conclusions

In this paper, we proposed four generative models for topic suggestion on micro-reviews in location-based social sites. In order to address the data sparsity problem due to the shortness of micro-review documents, we used document pooling strategies to aggregate all micro-reviews, forming venue-level and user-level documents. We applied LDA on venue-level documents and user-level documents, to derive models VLDA and ULDA, respectively. VLDA discovers topics based on the inherent aspects of venues, and ULDA

models identified topics from user preferences. The experimental results demonstrate that venue inherent properties have larger effect than user preferences for topic suggestion on micro-reviews. However, employing both these data sources in topic prediction is essential. In view of this, we introduced the ALDA model which integrates influences from both venues and user preferences, and used a mixture parameter to weight the effects. The ALDA model can learn more coherent topics and the suggested topics are more related to the true tip content. Going one step further, we proposed a ASLDA model, which also considers sentiment-orientation of user preferences to suggest topics for micro-reviews. We gave examples of topic suggestion by all these models which indicate the superiority of the ASLDA model. We believe that the ASLDA model finds important application in recommender systems toward predicting the focused aspects of a customer on a specific venue. In addition, sentiment orientation based topic suggestion can be used to improve personalized services in practice.

In the future, we plan to investigate the integration of other larger sentiment lexicons (e.g., SentiWordNet) for sentiment context expansion. Another direction is to integrate social information for the cold-start problem in topic suggestion (there is no social information in our datasets). For example, for a new user who has no check-ins or very few check-ins, we can suggest topics preferred by friends.

Acknowledgments

This work was supported by grant 715413E from Hong Kong RGC and by Marie Curie Reintegration Grant projects titled JMUGCS and LBSKQ which have received research funding from the European Union.

References

Aggarwal, A.; Almeida, J.; and Kumaraguru, P. 2013. Detection of spam tipping behaviour on foursquare. In *Proceedings of the 22Nd International Conference on World Wide Web Companion*, WWW '13 Companion.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3:993–1022.

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

Hai, Z.; Cong, G.; Chang, K.; Liu, W.; and Cheng, P. 2014. Coarse-to-fine review selection via supervised joint aspect and sentiment model. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '14.

Hu, M., and Liu, B. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04.

Jo, Y., and Oh, A. H. 2011. Aspect and sentiment unification model for online review analysis. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11.

Lin, C., and He, Y. 2009. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09.

Lin, T.; Tian, W.; Mei, Q.; and Cheng, H. 2014. The dual-sparse topic model: Mining focused topics and focused terms in short text. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14.

Mehrotra, R.; Sanner, S.; Buntine, W.; and Xie, L. 2013. Improving lda topic models for microblogs via tweet pooling and automatic labeling. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '13.

Moghaddam, S., and Ester, M. 2011. Ilda: Interdependent lda model for learning latent aspects and their ratings from online product reviews. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11.

Moraes, F.; Vasconcelos, M.; Prado, P.; Almeida, J.; and Gonçalves, M. 2013. Polarity analysis of micro reviews in foursquare. In *Proceedings of the 19th Brazilian Symposium on Multimedia and the Web*, WebMedia '13.

Newman, D.; Lau, J. H.; Grieser, K.; and Baldwin, T. 2010. Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, 100–108.

Nguyen, T.-S.; Lauw, H. W.; and Tsapras, P. 2015. Review synthesis for micro-review summarization. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, WSDM '15.

Paul, M., and Girju, R. 2010. A two-dimensional topic-aspect model for discovering multi-faceted topics. In *the proceedings of the 24th AAAI Conference on Artificial Intelligence*, AAAI'10.

Rosen-Zvi, M.; Griffiths, T.; Steyvers, M.; and Smyth, P. 2004. The author-topic model for authors and documents. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, UAI '04, 487–494.

Srivastava, A., and Sahami, M. 2009. *Text Mining: Classification, Clustering, and Applications*. Chapman & Hall/CRC, 1st edition.

Tang, J.; Zhang, M.; and Mei, Q. 2013. One theme in all views: Modeling consensus topics in multiple contexts. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13.

Wang, H.; Lu, Y.; and Zhai, C. 2010. Latent aspect rating analysis on review text data: A rating regression approach. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, 783–792.

Wang, H.; Lu, Y.; and Zhai, C. 2011. Latent aspect rating analysis without aspect keyword supervision. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11.

Weng, J.; Lim, E.-P.; Jiang, J.; and He, Q. 2010. Twitterrank: Finding topic-sensitive influential twitterers. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, 261–270.

Yin, H.; Sun, Y.; Cui, B.; Hu, Z.; and Chen, L. 2013. Lcars: A location-content-aware recommender system. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13.

Zhao, W. X.; Jiang, J.; Weng, J.; He, J.; Lim, E.-P.; Yan, H.; and Li, X. 2011. Comparing twitter and traditional media using topic models. In *Proceedings of the 33rd European Conference on Advances in Information Retrieval*, ECIR'11.