Phrase-based Topic and Sentiment Detection and Tracking Model using Incremental HDP

YongHeng Chen[1, 2], YaoJin Lin[1, 2], WanLi Zuo[3]

[1]College of Computer Science, Minnan Normal University, zhangzhou363000, China
[2]Key Laboratory of Data Science and Intelligence Application, Fujian Province University
[3]College of Computer Science and Technology, Jilin University, Changchun, China
*Corresponding author: YongHeng Chen
[e-mail: yh_chen@mnnu.edu.cn]

Received April 28, 2017; revised July 8, 2017; revised July 17, 2017; accepted August 25, 2017; published December 31, 2017

Abstract

Sentiments can profoundly affect individual behavior as well as decision-making. Confronted with the ever-increasing amount of review information available online, it is desirable to provide an effective sentiment model to both detect and organize the available information to improve understanding, and to present the information in a more constructive way for consumers. This study developed a unified phrase-based topic and sentiment detection model, combined with a tracking model using incremental hierarchical dirichlet allocation (PTSM_IHDP). This model was proposed to discover the evolutionary trend of topic-based sentiments from online reviews. PTSM_IHDP model firstly assumed that each review document has been composed by a series of independent phrases, which can be represented as both topic information and sentiment information. PTSM_IHDP model secondly depended on an improved time-dependency non-parametric Bayesian model, integrating incremental hierarchical dirichlet allocation, to estimate the optimal number of topics by incrementally building an up-to-date model. To evaluate the effectiveness of our model, we tested our model on a collected dataset, and compared the result with the predictions of traditional models. The results demonstrate the effectiveness and advantages of our model compared to several state-of-the-art methods.

Keywords: topic minding, sentiment analysis, nonparametric Bayesian statistics, Markov chain Monte Carlo

YongHeng Chen's word was supported by the National Natural Science Foundation of China under Grant No. 60373099, No. 60973040, and No. 61303131;

https://doi.org/10.3837/tiis.2017.12.012 **ISSN : 1976-7277**

1. Introduction

With the sheer overwhelming amount of reviews available online, consumers can turn to online reviews to seek advice, while companies see such reviews as a valuable source of consumer feedback, e.g. providing feedback about the waiting time at a restaurant or the noise level of a vacuum cleaner. The sentiment information about a specific topic, implied within these reviews, directly affects the purchase decision-making of consumers. It is therefore important for a company to understand the consumer experience of using a product, and to optimize their products or services quality accordingly. However, revealing sentiments and opinions via manual analysis of a large volume of textual data is rather time-consuming.

Hence, an efficient and convenient method for both analysis and organization of sentiment information from reviews is required, particularly one that is accessible for those who are not familiar with computer science or informatics techniques. Statistical admixture topic models have been proven to be a very useful tool to attain that goal. Latent Dirichlet Allocation (LDA) [1] is one of the basic and most general models for parametric Bayesian statistics and is a popular topic modeling method developed to automatically extract a set of semantic themes from large collections of documents. LDA clusters semantically relate topics by co-occurrence information of terms within a document collection. However, LDA can only detect a predefined number of topics; while reviews often are time-dependent data streams. Consequently, the number of topics should be flexible and automatically learned. Therefore, we assumed that the number of mixture components (topics) is unknown a priori and is to be inferred from the data. In this setting, it is natural to consider sets of Dirichlet processes (DP), where the well-known clustering property of the Dirichlet process provides a non-parametric prior for the number of mixture components within each group. Instead of modeling each document as a single data point, we modeled each document as a Dirichlet processin this study. In this setting, each word presented a data point and was thus associated with a topic sampled from random measure. The random measure consequently represented a document-specific mixing vector over a potentially infinite number of topics. To share a set of topics across documents, Tehet al. introduced the Hierarchical Dirichlet Process (HDP), which is a typical non-parametric Bayesian model and has the ability to estimate the optimal number of mixture components (topics) [2]. In HDP, document-specific random measures are linked by modeling the base measure itself as a random measure sampled from a DP. The discreteness of the based measure ensures the sharing of the topics between all groups. However, HDP itself is a static model and cannot consider time-stamp information embed in reviews. If topics are independently extracted through static HDP for each grouped dataset according to a specific time slice, the evolutionary information will be lost.

As a core component of any recommendation system, sentiment analysis has long been explored. Earlier studies mainly used categorization approaches, and have made noteworthy achievements. However, the basic premise of using these approaches is the requirement for the collection of training data, which requires a substantial amount of time and energy. More

importantly, most reviews are not labeled, which introduced a severe problem for traditional approaches. Otherwise, sentiment polarities are relevant for topics. Therefore, analyzing both topic and sentiment simultaneously will help consumers to better understand the content of the reviews. Therefore, LDA-based sentiment analysis approaches are becoming increasingly popular, integrating the benefits of the LDA model: namely that these approaches can train data collection free, precisely identifying topic and sentiment. Furthermore, these approaches also inherit the main disadvantage of the LDA model, i.e., the number of topics cannot be flexibly and automatically learned. In general, single terms communicate little information relative to phrases. Numerous verbs or prepositions are even meaningless without related words. More importantly, in short comments, such as on Amazon, most opinions are conveyed in the form of concise phrases, such as "adequate quality", or "good service". Therefore, the bag-of-words assumption is not sufficient to meet the demands of detecting topics from a large volume of textual data. This challenge has attracted significant effort for exploiting phrased LDA to extract topics and sentiment [27, 28, 29].

Considering all these associated problems, and to track the trends of latent topics with sentiment polarity for reviews, an excellent evolutionary model of topic–based sentiment analysis should possess the following three characteristics. (1) The number of topics can be automatically determined; (2) Topics should be allowed to evolve over time; (3) Topics and topic-related sentiments can be detected simultaneously at the level of the single phrase. In this paper, we propose the PTSM_ IHDP, which is an on-line evolutionary sentiment/topic model for continuous reviews, which can both detect and track topic-based sentiments by a time-dependent HDP for review datasets, and has the capacity to estimate the best number of topics for each time slice, while controlling the birth, death, and inheritance of the topic by calculating historical influences from a previous time slices to the updated time slice.

2. Related Work

In this section, we briefly review previous methods which are most related to our current study, including topic-based sentiment analysis and non-parametric Dirichlet Process.

- Topic-based sentiment analysis

Sentiment analysis has drawn much attention for extracting sentiments from underlying review content streams. Typical early studies concentrated mostly on sentiment classification, which is composed of detecting opinions and sentiment polarities. Through introducing a method that combined Conditional Random Fields (*CRF*) and a variation of AutoSlog, Choi et al. implemented opinion and emotion detection [3]. Liu et al. presented an analysis framework to compare a consumer sentiment polarity score of multiple produces with a supervised pattern discovery method [4]. To determine sentiment polarities of the document, Pang et al. proposed a machine-learning method, adopting text categorization techniques and minimum cuts in graphs [5]. In a different study, Pang et al. achieved sentiment classification, where a review can be either positive or negative, at the level of the document using machine-learning

techniques via the overall sentiment [6]. However, these studies merely focused on to sentiment classification, and did not take the latent topics embedded in the document into account, thus providing insufficient information for consumers, likely leading to inapplicable results. For example, consumers just want to know merits and faults (sentiment) about the battery life (topic) of a cellphone, without having to read an overall product evaluation. Motivated by this observation, researchers considered combining topic extraction with sentiment analysis, which is called topic-based sentiment analysis. As one of the state-of-the-art methods of this topic model, the Latent Dirichlet Allocation (*LDA*) model has gained popularity, which is a hierarchical Bayesian network. *LDA* builds robust topic summaries in accordance with the multinomial probability distribution over words for each topic, and can further deduce discrete distributions over topics for each document. To use timestamps to improve topics discovery, the Topics over time (TOT) model proposed by Wang et al. jointly modeled time with word co-occurrence patterns based on *LDA* in an off-line fashion [7]. Nonetheless, the above-mentioned topic models have no capabilities for working in an on-line fashion. These stimulated studies searched for an optimized model, and several online topic models have ultimately been proposed [8, 9, 10]. Using temporal streams information, the dataset can be divided by a predefined time slice. At each time slice, documents are supposed to be exchangeable. However, this is not true between documents and across time slices. This core idea will be inherited by this study.

Based on *LDA* and its extended models, many topic-based sentiment analysis models have been proposed. Joint sentiment/topic model (JST) [11] and Aspect and sentiment unification model (ASUM) [12] are representative of these. *JST* can simultaneously implement the detection of sentiments and topics based on the *LDA* model. The *ASUM* model constrains the words in a single sentence that originate from the same polarity, which is called the sentence-level JST model. Zhan developed a topic model [27] defined on opinion phrases, in which they proposed a semantic dependent word pair generative model for pairs of nouns and adjectives for each sentence, then applied the LDA model to reveal topics from opinion phrases at the sentence level. Lu et al. [28] described that applying preprocessed reviews can potentially boost the quality of models for topic detection. The authors assumed that each review can be parsed into phrases of the format < *head term, modifier term*>, where the head term is a topic or a feature, and the modifier expresses some sentiment towards this topic. Then, they adopted the PLSI model to detect most topics of an online product. Since social media data are produced continuously by many uncontrolled users, the dynamic nature of such data requires the sentiment and topic analysis model to be updated dynamically. Time-aware Topic-Sentiment (TTS) [30] and dynamic joint sentiment-topic model (dJST) [31] are the rarely work to detect and track dynamic topic and sentiment based on probability topic model, where TTS and dJST are both using word-based latent Dirichlet allocation (LDA). However, TTS had jointly modeled time, word co-occurrence and sentiment with no Markov dependencies such that it treated time as an observed continuous variable. This approach, however, works offline, as the whole batch of documents is used once to construct the model. This feature does not suit the online setting where text streams continuously arrive with time.

Otherwise, since LDA are parametric probabilistic model, both of them require determining the topic number beforehand. It is insufficient for the dynamic and massive social media data.

- Dirichlet Process Mixture Model

The selection of the number of topics can have a significant impact on how well a given model fits the data, and its ability to generalize on the training data. The above analysis naturally led us to consider non-parametric Bayesian models [13, 14, 15] to alleviate this problem. The Dirichlet Process is a typical method for a nonparametric Bayesian process, which is represented by $DP\ (G_0,\ \alpha)$, where G_0 is a base measure parameter and α is a concentration parameter. A document canbe modeled as a *DP*, and each word in that document d would become a target object that is created by the distribution words over a topic sampled from the random measure $G_d \in DP\ (G_0\ ,\alpha)$. The random measure *G* denotes the distribution of the document-based mixing vector over an infinite number of topics. To allow the sharing of data among the collection of topics across documents, another non-parametric model was proposed: the Hierarchical Dirichlet Process, which uses Dirichlet processes as the Bayesian prior to solve the topics number determination problem. Many models that integrated time information based nonparametric Bayesian have recently been proposed to improve topic discovery [16, 17]. To dynamically implement clustering analysis of topics, several nonparametric Bayesian-extended models have been proposed [18, 19].

Our PTSM_IHDP model differs from other models because: (1) PTSM_IHDP is directing proposed against a specific field of topic-based sentiment analysis, using a phrase-based method to construct both topic word and sentiment word; (2) PTSM_IHDP implements Gibbs Sampling to obtain parameters at each time slice rather than executing global deduction, which is effective for updating a timely evolution; (3) To track the trend of the detected topics, PTSM_IHDP proposed time-dependency HDP to realize the development and change of the same topic over time.

3. Methodology

3.1 Phrase-based Topic and Sentiment Model based IHDP

While HDP has the capability of determining the appropriate number of latent topics, it is not adequate for tracking a trend of topics and for detecting topics of short comments; it does not have the capability to provide means of combining sentiment labels into the training procedure. This stimulated us to propose the PTSM_IHDP (see **Fig. 1**). A few notions will first be introduced. The dataset of continuous reviews x will be divided according to the time slice, x= $\{x^1, x^2 \ldots x^T\}$, where *T* denotes the number of time slices and x^t represents the dataset of reviews that included publishing times in the time slice t. Moreover, $x^t = \{x_1^t, x_2^t \ldots x_{D_t}^t\}$, where D_t denotes the number of reviews within time slice t. Each review is presented with a series of words $x_d^t = (x_{di}^t)_{N_d^t i=1}$, where N_d^t represents the number of words within review x_d^t. Supposing a review is presented with the probability distribution over infinite topics, and that topic

represents a probability distribution over words, the target for PTSM_IHDP is to estimate the number of topics, and track the development of each topic by analyzing the change of the word distribution and sentiment polarity of the topic at the level of phrase for different time slices.

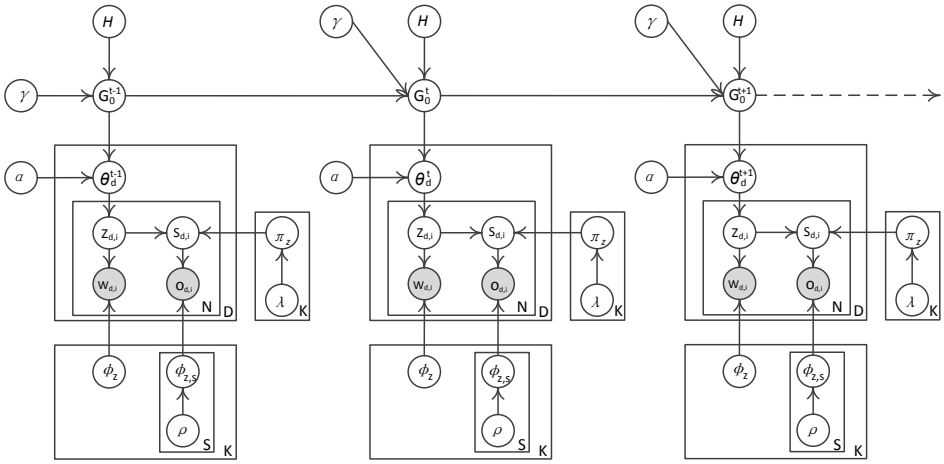


Fig. 1. Phrase-based topic and sentiment detection as well as tracking model using incremental hierarchical dirichlet allocation (PTSM_IHDP)

To further integrate the time information into HDP, the base measure G_0 should be dynamically calculated for each time slice, i.e., the number of mixture components at each time point is unbounded; the components themselves can remain, die out, or emerge over time; and the actual parameterization of each component can also evolve over time via Markovian fashion. Through considering previous Δ time slices, the base measure G_0^t at the current time slice can be incrementally obtained as follows:

$$G_0^t \mid G_0, g, f_{1:k} \sim DP\left(g + \sum_{M}\sum_{d=1}^{D} E(v,d) \times d_k^{t-d}, \frac{H}{1+\sum_{d=0}^{D} E(v,d)} + \frac{\sum_{d=0}^{D} E(v,d) G_0^{t-d}}{1+\sum_{d=0}^{D} E(v,d)}\right) \quad (1)$$

where the decay function $E(v,\delta)$ represents the function of the exponential kernel, $E(v,\delta)=exp(-\delta/v)$, that manages the weight of topic k at time slice t-δ. v and Δ define the decay factor of the time-decaying kernel and the time windows that influences the current time slice. Each epoch is independent when Δ=0, and time can be ignored when Δ = T and v = ∞. In between, the values of these two parameters affect the expected life span of a given component. The larger the values of Δ and v, the longer the expected life span of the topic, and vice versa. $d_k^{t-\delta}$ denotes the number of parameters θ in time slice t-δ that are associated with φ_k. M is the normalization factor, which can been obtained via $1+\sum_{\delta}^{\Delta} E(v,\delta)$. Furthermore, to incorporate sentiment polarity labels to realize sentiment classification, our model constructed the relationships between topics and sentiment labels on the basis of some ideas of ILDA [29]. PTSM_IHDP preprocesses the review into a bag-of-phrases <w_n, o_n>leading to two observed

variables w_n (topic word) and o_n (sentiment word), and implementing sentiment analysis by adding an additional sentiment layer. The formal definition of the generative process in the PTSM_IHDP model corresponding to the graphical model is as follows:

(1)Generate the global topic distribution at time slice t: $G_0^t \mid G_0, \gamma, \phi_{1:k}$
(2)Generate the word-topic distribution for each topic: $\varphi_k \sim H$
(3)Generate sentiment distributions for each topic: $\varphi_{k,s} \sim Dir(\rho_s)$
(4)For each document d:
(4.1)Generate local topic distribution of document: $\theta_d^t \sim DP(\alpha, G_0^t)$
(4.2)For the i^{th} phrase in document d:
(4.2.1)Draw a topic assignment: $z_{d,i} \sim Mult(\theta_d^t)$
(4.2.2)Generate the sentiment distribution: $\pi_z \sim Dir(\lambda)$
(4.2.3)Draw a sentiment assignment: $s_{d,i} \sim Mult(\pi_{z_{d,i}})$
(4.2.4)Generate the topic word: $w_{d,i} \sim \varphi_{z_{d,i}}$
(4.2.5) Generate the sentiment word: $o_{d,i} \sim \varphi_{z_{d,i}, s_{d,i}}$

3.2 Incremental Hierarchical Dirichlet Allocation

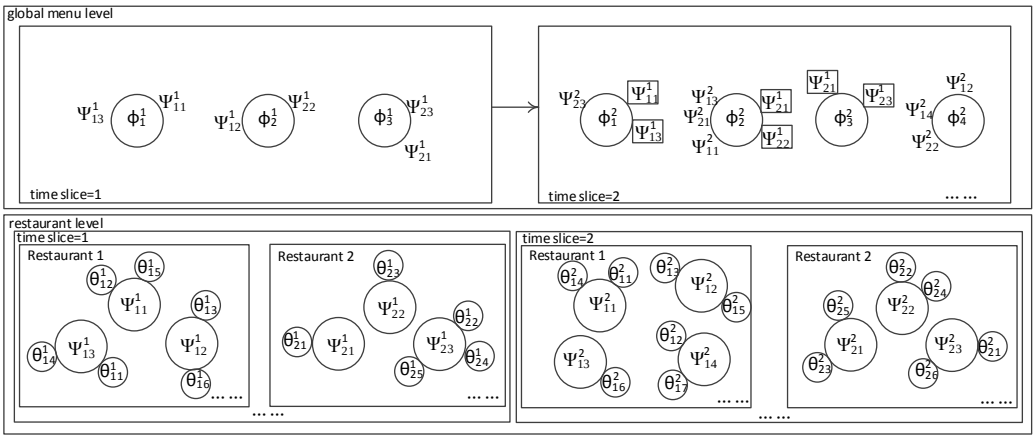


Fig. 2. Time-dependent CRFP

CRP is a discrete-timestochastic process, analogous to seated customers at tables in a Chinese restaurant. The Chinese restaurant franchise process (CRFP) is based on and extended Chinese restaurant process to allow for multiple restaurants, which share a set of dishes, which is a random process that produces an interchangeable division of data points and allows multiple data points to share a set of topics. CRFP is typically employed to simulate the HDP process. In this process, each restaurant maintains its set of tables, while sharing the same set of mixtures. A customer at one restaurant can chose to sit at an existing table with a probability proportional to the number of customers sitting on this table, or starting a new table at a certain probability and chose its dish from a global distribution. Due to expansibility and hierarchy, CRFP is widespread applied in non-parametric models. Although CRFP has the capacity of constructing data points by using s set of topics and allowing the number to topics to be infinite, it is a static process and cannot track the development of either latent topics or the probability distribution of words. This paper also used CRFP to construct a mixture model of the grouped

data. However, we replaced the first level of CRFP with a novel time-dependent random process, which could simulate incremental HDP. The modified CRFP was named a time-dependent CRFP, which can estimate the best number of topics for the current time-slice by considering influences from a previous set of time slices.

A wide array of metaphors and conceptions are employed in CRFP. An example for time-dependent CRFP is shown in **Fig. 2**. At the restaurant level, each restaurant is denoted by a rectangle and consumers (small circles) sit at different dining-tables (big circles) associated with a dish in this restaurant. A restaurant is a metaphor for a review document, while consumers correspond to phrases. At global menu level, the set of dish (topic), is served in common for all of the restaurants. x^t_{di} represents the i^{th} consumers in restaurant d for time slice t. θ^t_{di} represents the dish enjoyed by this customer, Ψ_{dj} represents the dish for j^{th} table and φ^t_k represents dish k on the menu. To record the relationship among consumers, tables, and dishes, this study provides two index variables. b^t_{di} represents the index of tables for the i^{th} consumer in restaurant d, and k^t_{dj} represents the index of dishes for j^{th} table in restaurant d for time slice *t*. Consequently, we have $\Psi^t_{db^t_{di}} = \theta^t_{di}$ and $\phi^t_{k^t_{dj}} = \Psi^t_{dj}$. This study used a time-dependent CRFP to simulate incremental HDP, and implemented the assignment of customers to dinning-tables and the relationship between tables and dishes, in which the global menus of each time slice were tied over time.

Table assignment

At time slice *t*, a customer enters restaurant *d*. This customer will pick the j^{th} dining-table with the following probability:

$$n^t_{dj} / n^t_d - 1 + \alpha \tag{2}$$

where n^t_{dj} and n^t_d denote the number of customer around the j^{th} dining table and in restaurant *d*, respectively, enjoying dish $\psi^t_{db^t_{dj}}$ ordered from the global menu by the first customer who sits at that table. α is a parameter governing the likelihood of choosing a new table. Alternatively, this customer can select a new table with the following probability:

$$\alpha / n^t_d - 1 + \alpha \tag{3}$$

Dish assignment

We will first provide some notations that have been adopted for dish assignment. Customers in the restaurant sit at different tables and each table is associated with a dish according to the dish menu. Let Nt^t_i represent the number of dining-tables within restaurant *i* at time slice *t*. The number of dining-tables that have ordered dish k for all of the restaurants is represented by d^t_k at time slice t, which can be calculated as follows:

$$d_k^t = \sum_{i=1}^{D_t} \sum_{j=1}^{Nt_i^t} 1(\Psi_{ib_{ij}^t}^t = k) \tag{4}$$

To integrate historical influences from a previous time slice, another parameter $d_k^{'t}$ has been defined as follows:

$$d_k^{'t} = \sum_{\delta=1}^{\Delta} E(v, \delta) \cdot d_k^{t-\delta} \tag{5}$$

The probability for a customer to pick a new table and chooses dish k that has been ordered by previous customs from the global menu at time slice t, is as follows:

$$\frac{d_k^{'t} + d_k^t}{\sum_{k=1}^{K_t} d_k^t + d_k^{'t} + \gamma} \tag{6}$$

where K_t is the number of dishes at time slice *t*.

If this dish is ordered by consumers in the previous Δ time slices, but has not yet been ordered by consumers at time slice *t*, this changes the distribution of this dish in Markova fashion: $\varphi_k^t | \varphi_k^{t-1} \sim P(.|\varphi_k^{t-1})$ (i.e. improve the probability distribution of topic). The probability is as follows:

$$\frac{d_k^{'t}}{\sum_{s=1}^{K_t} d_s^t + d_s^{'t} + \gamma} P(. \mid \phi_k^{t-1}) \tag{7}$$

If this dish has not been ordered at any time slices, i.e., it is a new dish, the number of dish K_t increments by one and the customer can select a new dish $\varphi_k \sim H$. The probability is as follows:

$$\frac{\gamma}{\sum_{s=1}^{K_t} d_s^t + d_s^{'t} + \gamma} H \tag{8}$$

4. Approximating Posterior Inference

In this section, we used the collapsed Gibbs sampling algorithm for posterior sampling of the assignments at the tables, and the dishes that have been served at a specific table in each restaurant. To perform Gibbs sampling, Markov Chain Monte Carlo (MCMC) [20] based on time-dependent CRFP was constructed to integrate parameters b_{di}^t, k_{dj}^t, φ_k^t and s_{di}^t into a joint probability distribution and states converges to a sample from this joint probability distribution.

To infer the posterior probability distribution of these, the conditional probability topic word x_{di}^t and x_{db}^t (i.e. all consumers pick the b^{th} dining-table in restaurant d at time slice t) of each phrase should be firstly solved. The prior Dirichlet distribution H samples topic φ_k^t through probability $h(\varphi_k^t|\eta)$, and multinomial distribution samples topic word x_{di}^t from topic φ_k^t through $f(x_{di}^t|\varphi_k^t)$. Given all the previous phrases except for the considered i^{th} phrase in document d at time slice t, the conditional posterior for x_{di}^t is:

$$f_k^{-x_{di}^t}(x_{di}^t) = p(x_{di}^t \mid -x_{di}^t, b, k) = \frac{p(x_{di}^t \mid b, k)}{p(-x_{di}^t \mid b, k)} = \frac{\int f(x_{di}^t \mid \phi_k^t) \prod_{d',i' \neq di, z_{d',i'}=k} f(x_{d',i'} \mid \phi_k^t) h(\phi_k^t \mid \eta) d(\phi_k^t)}{\int \prod_{d',i' \neq di, z_{d',i'}=k} f(x_{d',i'} \mid \phi_k^t) h(\phi_k^t \mid \eta) d(\phi_k^t)} \tag{9}$$

Since both the base measure H and multinomial distribution (i.e. topic word distribution) are conjugated distributions, the above formula can be simplified to:

$$f_k^{-x_{di}^t}(x_{di}^t = v) = \frac{n_k^{-x_{di,v}^t} + \eta}{n_k^{-x_{di}^t} + V\eta} \tag{10}$$

Where $n_k^{-x_{di,v}^t}$ is the topic word count of v in topic k except for x_{di}^t, where $n_k^{-x_{di}^t}$ is the number of topic words in topic k except for word x_{di}^t and V as the length of word vocabulary. Given all words except for the topic words x_{db}^t, the conditional posterior for x_{db}^t is:

$$f_k^{-x_{db}^t}(x_{db}^t) = \frac{n_k^{-x_{db}^t} + V\eta}{n_k^{-x_{db}^t} + n^{-x_{db}^t} + V\eta} \frac{\prod_v \Gamma(n_k^{-x_{db,v}^t} + n^{x_{db,v}^t} + \eta)}{\prod_v \Gamma(n_k^{x_{db,v}^t} + \eta)} \tag{11}$$

Where $n_k^{-x_{db}^t}$ is the topic word count of topic k except for x_{db}^t, $n^{-x_{db}^t}$ is the topic word count except for x_{db}^t. $n_k^{-x_{db,v}^t}$ is the number of the topic word v assigned to topic k except for x_{db}^t.

According to a time-dependent CRFP and sentiment analysis requirement, we employed the four stages of the inference process.

Sampling table b

For each customer at time slice t, the distribution of table b_{di}^t with the specific x_{di}^t is concerned with the number of consumers around this table, which is given by:

$$p(b_{di}^{t} = b \mid -b_{di}^{t}, k_{t-\Delta:t}, x_{di}^{t}) \propto \begin{cases} n_{db}^{-x_{di}^{t}} \cdot f_{k_{db}^{t}}^{-x_{di}^{t}}(x_{di}^{t}) \\ \alpha P(k_{db^{new}}^{t} = k \mid k_{t-\Delta:t}^{-tdi}, -b_{di}^{t}, x_{di}^{t}) \end{cases} \tag{12}$$

Where $n_{db}^{-x_{di}^{t}}$ is the topic word count of b_{di}^{t} except for x_{di}^{t} . The probability for sitting on a new table can be constructed by marginalizing over all available dishes, which leads to the following equation:

$$P(k_{db^{new}}^{t} = k \mid k_{t-D:t}^{-x_{di}^{t}}, -b_{di}^{t}, x_{di}^{t}) \propto \begin{cases} (d_{k}^{-tdb^{new}} + d_{k}^{'t}) f_{k}^{-x_{di}^{t}}(x_{di}^{t}) & k \text{ is being used} : d_{k}^{-tdb^{new}} > 0 \\ d_{k}^{'t} f_{k}^{-x_{di}^{t}}(x_{di}^{t}) & k \text{ is available but not used} : d_{k}^{'t} > 0 \\ \gamma f_{k_{new}}^{-x_{di}^{t}}(x_{di}^{t}) & k \text{ is a new topic} \end{cases} \tag{13}$$

Sampling topic *k*

Once the assignment of tables is complete, the posterior sampling dish k can be implemented. The process of sampling dish k_{dj}^{t} is similar to the above equation; however, we need to consider the probability of small groups of words (such as consumers on a given table). The conditional probability is estimated as follows:

$$P(k_{db}^{t} = k \mid k_{t-D:t}^{-tdi}, x_{db}^{t}) \propto \begin{cases} (d_{k}^{-tdb^{new}} + d_{k}^{'t}) f_{k}^{-x_{db}^{t}}(x_{db}^{t}) & k \text{ is being used} : d_{k}^{-tdb^{new}} > 0 \\ d_{k}^{'t} f_{k}^{-x_{db}^{t}}(x_{db}^{t}) & k \text{ is available but not used} : d_{k}^{'t} > 0 \\ \gamma f_{k}^{-x_{db}^{t}}(x_{db}^{t}) & k \text{ is a new topic} \end{cases} \tag{14}$$

Sampling φ_k

Given *b*, *k*, and observed *x*, the posterior conditional probability distribution of every φ_{k}^{t} only depends on whether all consumers enjoyed their dish, which can be estimated as follows:

$$P(\phi_{k}^{t} \mid t, k, x, \phi_{k}^{-t}) \propto h(\phi_{k}^{t} \mid \eta) \prod_{di:k_{db_{di}}=k} f(x_{di}^{t} \mid \phi_{k}^{t}) \tag{15}$$

Sampling *s*

Following topic inference, a sentiment needs to be chosen for the phrase under this topic. Let s_{di}^{t} denote the sentiment polarity for the i^{th} phrase in document d at time slice t. The posterior conditional probability of assigning sentiment polarity s_{di}^{t} can be drawn by:

$$P(s_{di}^{t} = s \mid k) \propto \frac{n_{kso}^{-o_{di}^{t}} + \rho_{s}}{n_{ks}^{-o_{di}^{t}} + V \cdot S} \cdot \frac{n_{ks}^{-o_{di}^{t}} + \lambda}{n_{k}^{-o_{di}^{t}} + S \cdot \lambda} \tag{16}$$

Where $n_{kso}^{-o_{di}^{t}}$ is the number of sentiment word o that has been assigned to sentiments under topic k except for o_{di}^{t} , $n_{ks}^{-o_{di}^{t}}$ is the number of sentiment words that have been assigned to sentiments under topic k except for o_{di}^{t} , $n_{k}^{-o_{di}^{t}}$ is the sentiment word count of topic k except for o_{di}^{t} .

5. Experiments

5.1 Dataset Presetting

To accomplish experiments, we use two different review datasets. The first dataset was composed of restaurant reviews from the website Yelp.com. The second dataset was the collection of hotel reviews that has been used previously [22]. These datasets were preprocessed by (1) deleting stop-words and non-English alphabets; (2) deleting low frequencies with appearances be low six times, as well as short reviews that are shorter than seven words; (3) adopting a Snowball algorithm for word stemming; (4) Stanford Dependency Parser [31] was used to extract opinion phrases, which is a widely used parser in the area of text mining.

Table 1. The properties of the data sets and sentiment lexicon

Sentiment lexicon	# of polarity words(pos./neg.)
paradiams	21 / 21
Paradiams+MI	41 / 41
MPQA	1335 / 2214

Corpus	Reviews	Words	Phrase
Restaurant	41,715	3,616,286	352,627
Hotel	34,157	4,287,183	407,842

Sentiment analysis is far more challenging than topics detection, because consumers express their attitudes is a subtle manner; however, topic detection is simply implemented on the basis of the co-occurrence of words. One technique to increase the accurateness of sentiment analysis was to integrate prior data, i.e., a sentiment lexicon. In this section, the three sentiment lexical, *Paradiams* [6], *Mutual Information* (*MI*) [23], and *MPQA*[24], have been used to improve sentiment classification accuracy. **Table 1** shows the properties of the datasets and sentiment lexicon information adopted for our experiments.

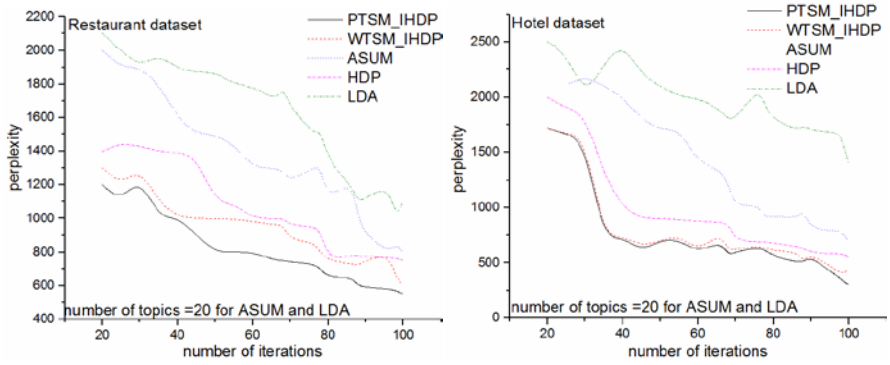
Unless otherwise stated, the parameters in this experiment were set according to the following values: the hyper parameter η of base probability measure H was set to10; concentration parameters γ and α were respectively obtained via vague gamma prior,

$r \propto \Gamma(1,0,1)$, $\alpha \propto \Gamma(1,1)$. Continuous time slices had Δ=4. The number of time slices was set to 20, T=20. To enable comparison to other models, parameters that are LDA-based models were set to the following: Dirichlet hyper parameter α=0.5, ß=0.02. The hyper parameters λ and ρ were set to1.0 and $\{10^{-7}, 0.01, 2.5\}$. The parameter v in the exponential kernel was set to 0.5.

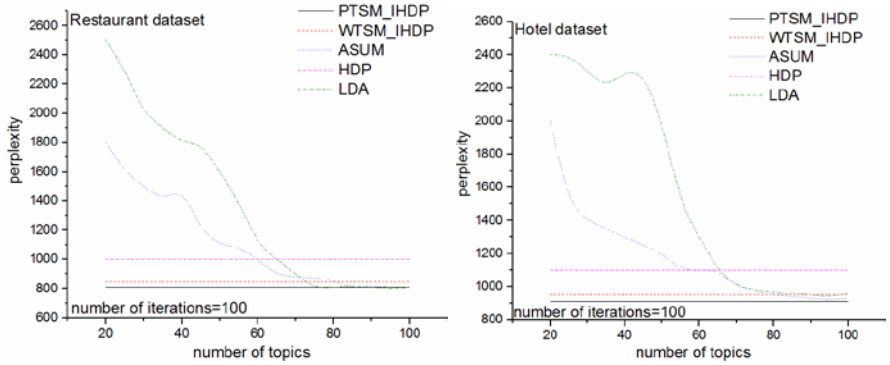
5.2 Perplexity

Perplexity is a canonical measure of goodness that has been used in language modeling to measure the likelihood of a held-out test data to be generated from the potential distributions of the model. In this subsection, we employed perplexity of the statistical model on test review datasets to measure the generalization performance when the performance of the model started to follow a steady state. Lower perplexity manifested better generalization performance. We classified the data into 80% for the training set and 20% for the testing set, where classification proportion were consistent across time slices. Formally, given the testing dataset, the perplexity value could be calculated as Eq. (10).

$$perplexity(x) = exp(-\sum_{j}^{J}\sum_{i}^{N_j} log_2 p(x_{ji}) \;/\; \sum_{j}^{J} N_j) \tag{17}$$



(a) Perplexity for different numbers of iterations



(b) Perplexity for different numbers of topics

Fig. 3. Perplexity score of two datasets against different models.

To test the optimization of our phrase-based relative word-based method, we modified our model to merge w and o and constructed word at word level, with the following acronym: WTSM_IHDP. Five models (LDA, ASUM, HDP, WTSM_IHDP, and PTSM_IHDP) have been adopted over both datasets to compare the perplexity of the performance. Firstly, the number of topics for LDA and ASUM was set to 20. **Fig. 3(a)** illustrates the result of the perplexity as a function of the number of iterations of the Gibbs sampler. Secondly, the number of iterations for four models was set to100. **Fig. 3(b)** shows the result of the perplexity as a function of the number of topics. Because nonparametric Bayesian models, HDP, WTSM_IHDP, and PTSM_IHDP, are irrelevant to the number of topics, the resulting values of their perplexity are fixed. As shown in **Fig. 3**, HDP, WTSM_IHDP, and PTSM_IHDP models can work more effectively for document clustering than the LDA and ASUM models, i.e., have better generalization performance and present lower perplexity values. The major reason is that non-parametric HDP, WTSM_IHDP, and PTSM_IHDP models can intelligently determine the number of topics rather than do this predefined, thus avoiding both under-fitting and over-fitting of information. More important, PTSM_IHDP has better generalization performance than WTSM_IHDP. This shows that the phrase-based method can provide better generalization performance, comparing a word-based method for short comments.

5.3 Complexity

Nonparametric Bayesian methods have often been used to side step model selection and integrate over all instances (and all complexities) of a given model (e.g., the number of clusters). The model, although hidden and random, still remains in the background. Here, we studied its posterior distribution with the desideratum that between two equally good predictive distributions, a simpler model (or a posterior peaked at a simpler model) is preferred. In this subsection, we measured the complexity of the model to evaluate non-parametric Bayesian models [25]. To implement the complexity of the model, the definition of topic complexity will firstly be presented. The complexity of a topic is proportional to the number of words allocated to this topic, i.e., the complexity of a topic is zero if no unique words would be allocated to this topic; otherwise, it would be the number of unique words allocated to this topic. Therefore, we expressed the complexity of a topic k as follows:

$$Complexity_k = \sum_d 1[(\sum_n 1[K_{di} = k]) > 0] \tag{17}$$

where K_{di} is the topic assignment for the i^{th} word in document d. For the posterior topic allocation of the Gibbs sample, the complexity of the model can be given by the sum of all complexities of the topic and the number of topics, which can been computed as follows:

$$Complexity = \#topics + \sum_k Complexity_k \tag{19}$$

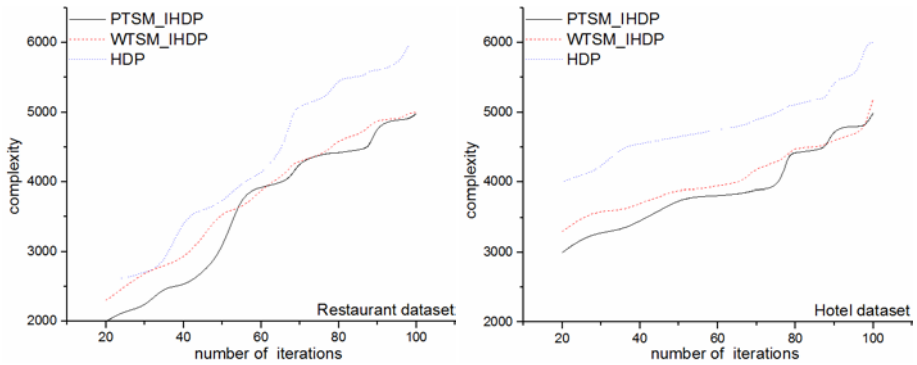


Fig. 4. Model complexity comparison of HDP and PTSM_IHDP models.

The complexity analysis considers the number of employed topics to describe the dataset. A higher complexity of a model shows that this model requires more topics to express the dataset, i.e., the dataset was classified into more dimensions. Therefore, a lower complexity manifests a better model, under the condition that the generated experimental results have a similar perplexity. **Fig. 4** illustrates the result of the complexity as a function of the number of iterations of the Gibbs sampler for both different datasets. As shown in Figure 4, PTSM_IHDP and WTSM_IHDP model have lower average complexity and outperformed the HDP in all cases. Taking account of the average complexity, the overall effect of the PTSM_IHDP model is more superior to that of WTSM_IHDP.

5.4 Sentiment Classification

We used three sentiment polarity labels: positive, negative, and neutral, associated to sentiment words for all phrases. *l(x)* was used to express the polarity label for the sentiment word x. *l(x)=1* if the label is positive, -1 if it is negative, and 0 if it is neural. Firstly, each sentiment word term matched the sentiments lexicon. The sentiment polarity label will be assigned to a sentiment word, if that sentiment word matched one of the words in the sentiments lexicon. Otherwise, one of three randomly sentiment polarity labels was selected for a sentiment word. After posterior sampling of the assignments, each sentiment word within a document had a sentiment polarity label attached. The sentiment polarity of local topics can be further obtained according to sentiment distribution within a local topic. The document sentiment value could be calculated as follows:

$$s_d^t = \sum_{x \in x_d^t} l(x)\varphi_{z_{d,x}, s_{dx}^t} \qquad s_d^t \in [-1, 1] \tag{20}$$

Where s_d^t denotes the sentiment value of document *d*; if this value is below 0, the document is categorized as negative. If this value is more than 0, the document is categorized as positive. Otherwise, the document is categorized as neutral. Restaurant and hotel corpuses use a 5-stars rating system. We assumed reviews with 1 or 2 stars to be negative. Reviews were considered

positive if they had 4 or 5 stars. Reviews with 3 stars have not been considered, i.e., reviews are being categorized as either positive or negative, without a neutral alternative. In this subsection, we measured the sentiment classification accuracy in terms of different sentiment lexical for the four sentiment models: JST, ASUM, WTSM_IHDP, and PTSM_IHDP.

Table 2. Sentiment classification accuracy comparison

Sentiment lexicon

paradiams
Paradiams+MI
MPQA

Restaurant dataset

JST(%)			ASUM(%)			WTSM_IHDP(%)			PTSM_IHDP(%)		
pos.	neg.	overall	pos.	neg.	overall	pos.	neg.	overall	pos.	neg.	overall
63.4	70.2	66.8	70.4	76.6	73.5	70.6	78.0	74.3	76.2	74.6	75.4
70.2	78.7	74.54	74.6	84.3	79.45	86.6	88.6	87.6	88.4	89.6	89
68.2	78.8	73.5	72.4	80.7	76.55	78.4	84.7	81.55	84.5	87.3	85.9

Hotel dataset

JST(%)			ASUM(%)			WTSM_IHDP(%)			PTSM_IHDP(%)		
pos.	neg.	overall	pos.	neg.	overall	pos.	neg.	overall	pos.	neg.	overall
66.6	74.6	70.6	68.3	75.4	71.85	71.4	78.8	75.1	72.5	80.7	76.6
72.3	80.7	76.5	76.4	86.5	81.45	87.9	89.6	88.75	84.3	92.9	88.6
68.6	79.8	74.2	72.4	81.4	76.9	74.6	86.7	80.65	82.2	90.4	86.3

Table 2 presents the predictive results of sentiment classification accuracy. Via incorporating only 21 positive and 21 negative paradigm words, both the PTSM_IHDP and WTSM_IHDP model acquired a relatively poor 75.4% and 74.3% overall accuracy, while JST and ASUM acquired 66.8% and 73.5%, respectively, based on the restaurant dataset. Similar results were obtained for the hotel dataset. By combining the top 20 words based on MI scores with paradigm words, it could be illustrated that significant improvement could be obtained in classification accuracy with 8%, 8%, 14%, and 14% for JST, ASUM, WTSM_IHDP, and PTSM_IHDP, respectively, using the restaurant dataset. However, classification accuracy was not proportional to the number of sentiment polarity words. Table 2 shows that incorporating the selected words in the MPQA sentiment lexicon caused the deterioration of classification accuracy, leading to impairment of the performance. Classification accuracy decreased by approximately 1%, 3%, 6%, and 4%, respectively for JST, ASUM, WTSM_IHDP, and PTSM_IHDP of the restaurant dataset. Similar experimental results were found for the hotel dataset. This indicates that in all settings, the accuracy of sentiment classification of the PTSM_IHDP model preformed best, while WTSM_IHDP also outperformed other traditional models.

5.5 Evolutionary Senti-Topic Discovery

PTSM_IHDP was proposed to produce topics coupled with sentiment for each time slice, easier facilitating customers to reveal how development and change of topics and topic sentiment scores developed over time. The probability distribution for topic words was estimated via φ_k, the distribution for words given topic k and sentiment label s was estimated via $\varphi_{k,s}$. To track the sentiment trend for each topic, the sentiment scores of topic k with sentiment polarities for time slice t were defined as following:

$$s_k^t = \begin{cases} \sum_{o \in o_k^t} \phi_{k,\,pos.} & positive\ score \\ -\sum_{o \in o_k^t} \phi_{k,\,neg.} & negative\ score \end{cases} \tag{21}$$

Where o_k^t are the sentiment words set that are assigned to topic *k*. the average sentiment score for each topic was the sum of score for positive or negative. Then, the sentiment scores for each topic series could be obtained form the sentiment time series $\{\ldots, s_k^{t-1}, s_k^t, s_k^{t+1} \ldots\}$. In this subsection, the topic word probability distributions and the sentiment word probability distributions of topics for different time slices could first be obtained by PTSM_IHDP. Then, the sentiment scores of each topic for different time slices could be calculated.

Three example topics were used: the first two were selected by the probability values, ranging from high to low, while the last one was selected randomly. This was coupled with topic word probability distributions based on restaurant and hotel datasets and topic' emotional development reflected by sentiment scores of different time slices that ranged at the upper and at the bottom of **Fig. 5** and **Fig. 6**. The upper part includes the first five sentiment words that have a probability value attached under two different sentiment labels (positive and negative) at different time slices. The lower portions show the changing of sentiment scores under different time slices for sampled topics. E.g., the first topic word for the extracted topic 1 based on restaurant dataset was "meat" with a probability of 0.17432, and the first sentiment word coupled with positive and negative label at time slice 12 was the word "good" and "flesh" with probabilities of 0.14363 and 0.132734, respectively. Topic 3 based on the restaurant dataset was emerging at time slice 10, and was inherited at time slice 13. After time slice 13, this topic has not been identified and died at time slice 13+Δ. This proves the efficacy of PTSM_IHDP in achieving the birth, death, and inheritance of topics. Topic 3 is based on the hotel dataset and is born at time slice 3, which could be observed from the sentiment score graph, but has not been identified at time slices 4, 6, 9, 14, and 17. This topic had not died, because it was used for subsequent consecutive Δ time slices, rather than having been inherited through considering the influence from previous time slices to the updated time slice at time slices 5, 7, 8, 10, 11, 12, and 13. This further indicates that our proposed model has the capacity to inherit topics that have been identified at previous consecutive Δ time slices. The sentiment score reflected the emotional status of topic. For example, the sentiment average value for topic 1 based on the restaurant dataset at time slice 5 was -0.40, indicating that a lot of negative feedback during this period on topic 1 have been received. This can attract the attention of the company, inform, and urge it to improve the related service.

topics {topic words}	Sentiment words	time slice=12	time slice=13	time slice=14
Topic 1 [.174320]meat [.160132]taste [.089274]flavor [.086254]taste [.084267]salti	Pos.	[.143630]good [.132734]fresh [.082742]tasty [.062437]pretty [.028363]pleasantly	[.096237]crisp [.042315]tasty [.039324]pretty [.013413]fresh [.012767]super	[.104300]tasty [.073428]crisp [.049560]Tender [.023520]fresh [.013241]nice
	Neg.	[.143621]dry [.064325]complains [.025329]pork [.014521]little [.013960]staleness	[.048345]too [.026455]bland [.014854]undercooked [.014220]disappoint [.011222]staleness	[.083428]nasty [.044534]dry [.023421]lack [.013437]tough [.009345]crust
Topic 2 [.149433]healt [.130322]dishware [.087432]sanitary [.083481]cup [.079285]table	Pos.	[.153542]clean [.120022]atmospher [.083456]bright [.053421]fun [.009674]capacious	[.124354]clean [.083427]quiet [.083002]capacious [.067543]joyful [.034326]well	[.125224]clean [.104554]happy [.084534]spacious [.058434]better [.038454]resort
	Neg.	[.142724]dirty [.100024]noise [.074353]smoking [.058455]music [.024540]talk	[.153479]ugly [.126836]dirty [.118466]noise [.082458]badly [.048725]unhappy	[.094564]worse [.092546]badly [.084369]loud [.036234]many [.008453]messy
(random selected) Topic 3 [.116405]roach [.090001]health [.070442]bug [.065302]dish [.044371]sour	Pos.		[.063456]feel [.054358]well [.010445]cost [.008045]good [.005295]swtte	
	Neg.		[.158345]disgusting [.126540]droppings [.084524]split [.069739]angry [.027569]outrageous	

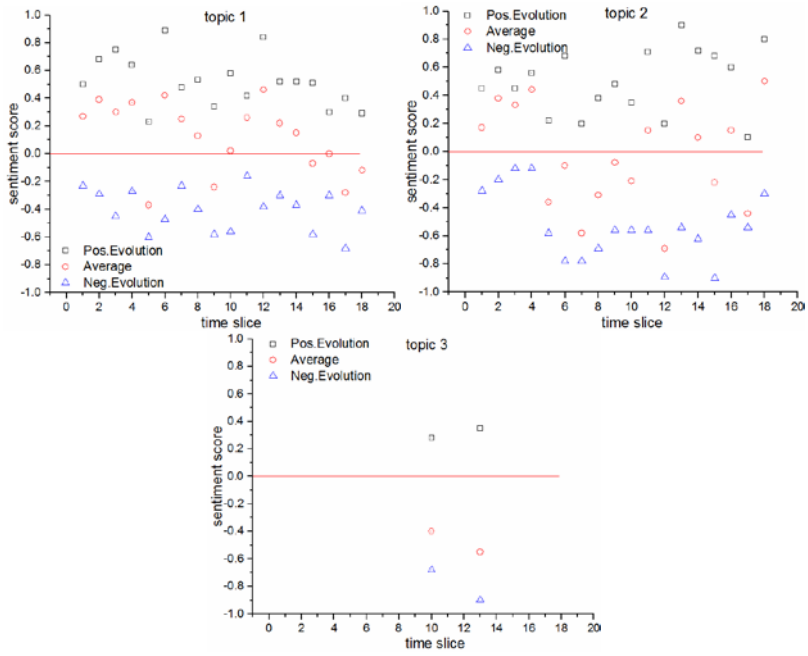


Fig. 5. Evolutionary senti-topic discovery based on Restaurant dataset

topics {topic words}	Sentiment words time slice=8	time slice=9	time slice=10
Topic 1 [.126521]location [.112113]traffic [.090282]station [.082302]bus [.082302]taxi	Pos. [.137450]convenient [.084534]cetral [.079268]travel [.048497]superiorly [.047432]advantageous	[.157842]facilitate [.136285]shopping [.114574]cetral [.094479]convenient [.010041]superiorly	[.150054]facilitate [.137834]cetral [.118438]convenient [.090032]short [.034579]advantageous
	Neg. [.142248]heavy [.105455]crowded [.096490]inconvenience [.074365]expensive [.036584]traffic	[.135440]inconvenience [.094565]crowded [.084547]expensive [.064349]jam [.038353]heavy	[.123532]inconvenience [.107732]difficult [.068625]crowded [.039853]remote [.0223421]badly
Topic 2 [.128753]curtains [.112325]floor [.081232]beds [.069234]restroom [.069234]toilet	Pos. [.100087]clean [.097322]fresh [.079001]dirty [.056802]nice [.027620]new	[.122001]clean [.089703]fresh [.071200]dirty [.051248]washed [.049258]timely	[.090303]clean [.079620]washed [.078291]better [.066313]professional [.023540]fast
	Neg. [.069007]filthy [.063429]dust [.054723]carpets [.050022]worn [.032370]sanitation	[.123420]dust [.118023]lack [.093518]roach [.070038]long [.063528]poor	[.100833]dust [.063464]poor [.058450]mosquito [.039530]old [.099579]disappointed
(random selected) Topic 3 [.098623]broadband [.087723]internel [.078832]computer [.068322]router [.060231]speed	Pos. [.102842]fast [.072496]enough [.059353]access [.032922]faster [.030014]wireless		[.102505]fast [.100512]reliable [.088425]access [.082342]wifi [.055823]satisfactory
	Neg. [.104872]slow [.097690]lack [.050703]problem [.049000]instability [.040152]charge		[.107324]interrupt [.086446]charge [.079344]inconvenien [.058008]trouble [.027231]instability

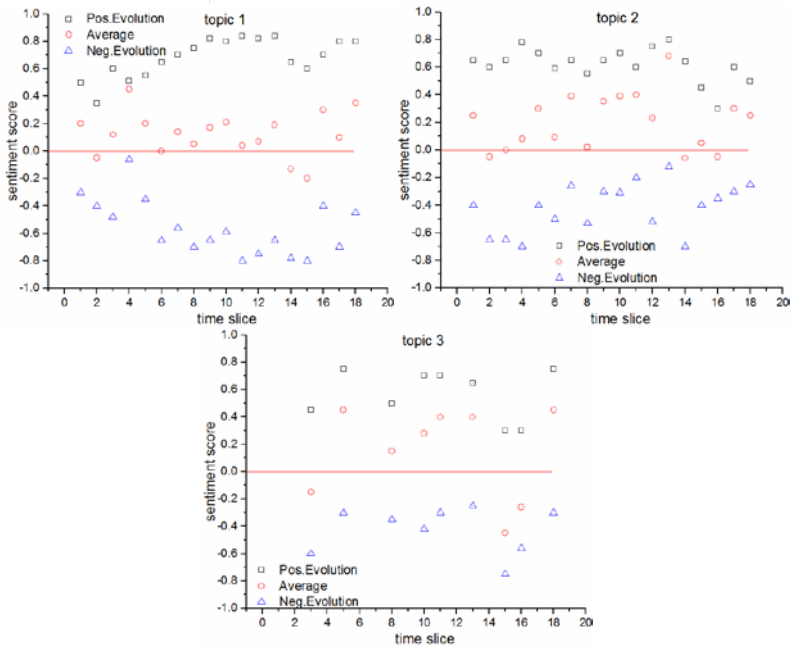


Fig. 6. Evolutionary senti-topic discovery based on Hotel dataset

6. Conclusion

In this study we addressed the problem of modeling a time-dependent review corpus for simultaneous analysis of topics and sentiment. A phrase-based topic and sentiment detection and tracking model that uses incremental hierarchical dirichlet allocation (PTSM_IHDP) was proposed, which can determine the topic number automatically via a non-parametric Bayesian topic model. We furthermore used phrase-based methods to construct topic words and sentiment words, and track the emotional development of the topics by a proposed time-dependent CRFP. A collapsed Gibbs sampling algorithm was utilized to infer parameters. Experiments have been conducted to evaluate the performance of PTSM_IHDP based on two real world datasets. The preliminary results demonstrated superiority of our proposed model over several state-of-the-art methods on generalization performance, lower average complexity, and better sentiment classification accuracy. The details of typical senti-topic discovery have also illustrated that PTSM_IHDP can effectively detect and track dynamic sentiment and topics.

As semantic information inherently consists of data with different modalities, we plan to extend our model to study and explore the interactions between multi-modal data for the sentiment topic analysis. Furthermore, one of the limitations of our model is that it requires setting the time span of each epoch. In the future works, we will consider some other time dependency modes to model the sentiment topic dynamics.

Reference

[1] DM Blei, A. Y. Ng and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, May, 2003. Article (CrossRef Link)

[2] Y.W Teh, M.I Jordan, MJ Beal and DM Blei, "Hierarchical Dirichlet Processes," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1566-1581, Dec., 2006. Article (CrossRef Link)

[3] Choi, Yejin and et al, "Identifying sources of opinions with conditional random fields and extraction patterns," in *Proc. of Conference on Human Language Technology and Empirical Methods in Natural Language Processing Association for Computational Linguistics*, pp. 355-362, Oct., 2005. Article (CrossRef Link)

[4] B. Liu, M. Hu, and J. Cheng, "Opinion observer: analyzing and comparing opinions on the Web," in *Proc. of International Conference on World Wide Web ACM*, pp. 342-351, Sep., 2005. Article (CrossRef Link)

[5] Bo Pang and Lillian Lee, "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts," *Meeting on Association for Computational Linguistics Association for Computational Linguistics*, no. 271, July 21-26, 2004. Article (CrossRef Link)

[6] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up: sentiment classification using machine learning techniques," in *Proc. of Conference on Empirical Methods in Natural Language Processing Association for Computational Linguistics*, vol. 10, pp. 79-86, 2002. Article (CrossRef Link)

[7] X. Wang and A. Mccallum, "Topics over time: a non-Markov continuous-time model of topical trends," in *Proc. of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 424-433, Aug. 20-23, 2006. Article (CrossRef Link)

[8] L. Alsumait, D. Barbara and C. Domeniconi, "On-line LDA: Adaptive Topic Models for Mining

Text Streams with Applications to Topic Detection and Tracking," *IEEE International Conference on Data Mining IEEE Computer Society*, pp. 3-12, Dec. 15-19, 2008. Article (CrossRef Link)
[9] L. Sato and H. Nakagawa, "Stochastic Divergence Minimization for Online Collapsed Variational Bayes Zero Inference of Latent Dirichlet Allocation," in *Proc. of ACM SIGKDD International Conference ACM*, pp. 1035-1044, 2015. Article (CrossRef Link)
[10] K. Sasaki, T. Yoshikawa and T. Furuhashi, "Twitter-TTM: An efficient online topic modeling for Twitter considering dynamics of user interests and topic trends," in *Proc. of International Symposium on Soft Computing and Intelligent Systems IEEE*, pp. 440-445, Dec. 3-6, 2014. Article (CrossRef Link)
[11] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," in *Proc. of ACM Conference on Information and Knowledge Management ACM*, vol. 217, no. 4, pp. 375-384, Nov. 2-6, 2009. Article (CrossRef Link)
[12] Y. Jo and A. H. Oh, "Aspect and sentiment unification model for online review analysis," in *Proc. of International Conference on Web Search and Web Data Mining*, vol. 81, no. 6, pp. 815-824, Feb. 9-12, 2011. Article (CrossRef Link)
[13] C. Lin, Y. He, R. Everson and S. Rüger, "Weakly Supervised Joint Sentiment-Topic Detection from Text," *IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no. 6, pp. 1134-1145, Jun., 2012. Article (CrossRef Link)
[14] A. Lijoi, R. H. Mena and I. Prünster, "Bayesian Nonparametric Analysis for a Generalized Dirichlet Process Prior," *Statistical Inference for Stochastic Processes*, vol. 8, no. 3, pp. 283-309, Dec., 2005. Article (CrossRef Link)
[15] R. M. Neal, "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 249-265, Jun., 2000. Article (CrossRef Link)
[16] K. Yu and P. M. Djuri, "Dirichlet process mixture models for time-dependent clustering," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4383-4387, March 20-25, 2016. Article (CrossRef Link)
[17] L. Ren, D. B. Dunson and L. Carin, "The dynamic hierarchical Dirichlet process," in *Proc. of the international conference on Machine learning*, pp.824-831, July 5-9, 2008. Article (CrossRef Link)
[18] T. Xu, Z. Zhang, P. S. Yu and B. Long, "Dirichlet Process Based Evolutionary Clustering," in *Proc. of IEEE International Conference on Data Mining*, pp. 648-657, Dec. 15-19, 2008. Article (CrossRef Link)
[19] T. Xu, Z. Zhang, P. S. Yu and B. Long, "Evolutionary Clustering by Hierarchical Dirichlet Process with Hidden Markov State," in *Proc. of IEEE International Conference on Data Mining*, pp. 658-667, Dec. 15-19, 2008. Article (CrossRef Link)
[20] D. Sorensen and D. Gianola, "Implementation and Analysis of MCMC Samples," *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*, pp. 539-560, 2002. Article (CrossRef Link)
[21] T. L. Griffiths and M. Steyvers, "Finding scientific topics," in *Proc. of the National Academy of Sciences of the United States of America*, vol. 101, pp. 5228-5235, 12 Nov., 2014. Article (CrossRef Link)
[22] K. Ganesan and C. Zhai, "Opinion-based entity ranking," *Information Retrieval Journal*, vol.15, no. 2, pp. 116-150, Apr., 2012. Article (CrossRef Link)
[23] T. Wilson, J. Wiebe and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," *International Journal of Computer Applications*, vol. 7, no. 5, pp. 347-354, Oct. 6-8, 2005. Article (CrossRef Link)
[24] F. Maes, A. Collignon and et al, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187-198, Apr., 1997. Article (CrossRef Link)
[25] S. Mousavi, K. Welch and et al, "Non-equilibrium split Hopkinson pressure bar procedure for non-parametric identification of complex modulus," *International Journal of Impact Engineering*, vol. 31, no. 9, pp. 1133-1151,Oct., 2005. Article (CrossRef Link)

[26] M. Zhang and B. Kang, "Visual Tracking Algorithm Based on Probabilistic Graphical Model," *International Journal of Signal Processing Image Processing and Pattern Recognition*, vol. 8, no. 9, pp.157-166, Sep., 2015. Article (CrossRef Link)
[27] T. J. Zhan and C. H. Li, "Semantic dependent word pairs generative model for fine-grained product feature mining," *Asia conference on Advances in knowledge discovery and data mining*, vol. 1, pp. 460-475, May 24-27, 2011. Article (CrossRef Link)
[28] Lu Yue, C. X. Zhai and N. Sundaresan, "Rated aspect summarization of short comments." in *Proc. of the international conference on World Wide Web*, pp. 131-140, Apr. 20-24, 2009. Article (CrossRef Link)
[29] S. Moghaddam and M. Ester, "ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews," in *Proc. of International ACM SIGIR Conference on Research and development in Information Retrieval*, pp. 65-674, July 24-28, 2011. Article (CrossRef Link)
[30] M. Dermouche, J. Velcin and et al, "A Joint Model for Topic-Sentiment Evolution over Time," in *Proc. of IEEE International Conference on Data Mining IEEE*, pp. 773-778, Dec. 14-17, 2014. Article (CrossRef Link)
[31] Y. He, C. Lin, W. Gao and et al, "Dynamic joint sentiment-topic model," *Acm Transactions on Intelligent Systems and Technology*, vol. 5, no. 1, pp. 1-21, Dec., 2014. Article (CrossRef Link)

Yongheng cheng was born in Heilongjiang of China in Dec 1980 and received the Ph.D. degree at the Department of Computer Science and technology, Jilin University. His current main research interests include Data Mining, Web Intelligence and Ontology Engineering and Information integration. He is a member of System Software Committee of China's Computer Federation. More than 20 papers of him were published in key Chinese journals or international conferences, 10 of which are cited by SCI/EI.

Yaojin Lin received the Ph.D. at Hefei University of Technology, and a Professor in the Department of Computer and Engineering, Minnan Normal University. His research interests include data mining, granular computing.

Wan-Li Zuo was born in Jilin of China in Dec 1957. He is a professor and doctoral supervisor at Department of Computer Science and technology, Jilin University. Main research area covers Database Theory, Machine Learning, Data Mining and Web Mining, Web Search Engines, Web Intelligence.