

On-line Evolutionary Sentiment Topic Analysis Modeling

YongHeng Chen^{1, 2}, ChunYan Yin¹, YaoJin Lin², Wanli Zuo³

¹ School of Information Engineering, Lingnan Normal University,
Zhanjiang, Guangdong, China,

² College of Computer Science, Minnan Normal University,
Zhangzhou, Fujian 363000, China,

Key Laboratory of Data Science and Intelligence Application, Fujian Province University
E-mail: {YH_chen,yjlin}@mnnu.edu.cn

³ College of Computer Science and Technology, Jilin University
Changchun, Jilin 130012, China
E-mail: wanli@jlu.edu.cn

Received 3 June 2017

Accepted 6 January 2018

Abstract

As the rapid booming of reviews, a valid sentiment analysis model will significantly boost the review recommendation system's capability, and present more constructive information for consumers. Topic probabilistic models have already shown many advantages for detecting potential structure of topics and sentiments in reviews corpus. However, most reviews are presented through time-dependent data streams and some respects of the potential structure are unfixed and time-varying, such as topic number and word probability distribution. In this paper, a novel probabilistic topic modelling framework is proposed, called on-line evolutionary sentiment/topic modeling (OESTM), which has the capacity for achieving the optimization of the aforementioned aspects. Firstly, OESTM depends on an improved non-parametric Bayesian model for estimating the best number of topics that can perfectly explain the current time-slice, and analyzes these latent topics and sentiment polarities simultaneously. Secondly, OESTM implements the birth, death and inheritance for detected topics through the transfer of parameters from previous time slices to the updated time slice. The experiments show that significant improvements have been achieved by the proposed model with respect to other state-of-the-art models.

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

1. Introduction

Reviews imply a lot of valuable sentiment information of consumers on a variety of topics of the product or service. For example, there are praises and complaints over topics, such as the taste of food or the hygiene conditions within restaurant reviews. This information implied within these reviews directly affects the purchase decision-making of con-

sumers. It is more important for company to understand the experience on use of the product, and optimize their produces or services quality. However, it is quite challenging to detect topics and sentiments over the large textual data sets by human being. This has led to the development for the topic-based sentiment mining and retrieval techniques in recent years as a new research field, which aims to automatically detect attitudes and emotions with respect to certain

latent topic implied within reviews.

As a core task of topic-based sentiment mining, topic models, that can detect latent topics clustering, have been explored recently. Topic models consist mainly of parametric Bayesian and non-parametric Bayesian methods. Latent Dirichlet Allocation (LDA) ¹ is one of the basic and most generally model for parametric Bayesian. LDA extracts and clusters semantically related topics by co-occurrence information of term within a document collection. But LDA can only detect a predefined number of topics. However, reviews often come as time-dependent data streams. The number of topics should be flexibly and automatically learned. So we assume 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 process (DP), one for each group, where the well-known clustering property of the Dirichlet process provides a nonparametric prior for the number of mixture components within each group. In this paper, instead of modeling each document as a single data point, we model each document as a Dirichlet process. In this setting, each word is a data point and thus will be associated with a topic sampled from the random measure. The random measure thus represents the document-specific mixing vector over a potentially infinite number of topics. To share the set of topics across documents, Teh et al. introduced the Hierarchical Dirichlet Process (HDP), which is a typical non-parametric Bayesian model and has ability to estimate the best number of mixture components (topics) ². In HDP, the document-specific random measures are tied together by modeling the base measure itself as a random measure sampled from a DP. The discreteness of the based measure ensures sharing of the topics between all the groups.

But HDP itself is a static model and not considering time-stamp information embed in reviews. If topics are extracted independently through static HDP for each grouped dataset according to time slice, the evolutionary information will be lost. Sentiment analysis also plays an important role in topic-based sentiment mining tasks. Earlier researches mainly use categorization approaches, and have

made some achievements. But the basic premise is the need for training data collection, which is based on a substantial amount of time and energy. More importantly, most reviews are not labeled, which puts forward the problem for traditional approaches. The appearance of LDA topic model provides new opportunity for solving these problems. Sentiment analysis approaches based on LDA have the benefit of LDA model, namely that these approaches can train data collection free, and identify topic and sentiment precisely. Meanwhile, these approaches also inherit the disadvantage of LDA model, i.e., the number of topics cannot be learned flexibly and automatically.

Inspired by the above discussions, we propose an on-line evolutionary sentiment/topic model (OESTM) to jointly exploit the sentiment and topics property for continuous reviews. The motivation of OESTM is to detect and track dynamic sentiment and topics simultaneously over time. OESTM has the capacity for estimating the best number of topics through adding a sentiment level to the Hierarchical Dirichlet Process (HDP) topic model, and then control the topics' birth, death and inheritance by proposed Time-dependent (Chinese restaurant franchise process) CRFP which adds time decay dependencies of historical epochs to the current epochs. Compared with the existing sentiment-topic models, the biggest difference of OESTM is that OESTM can determine topic number automatically. Furthermore, we implement OESTM with the collapsed Gibbs sampling algorithm. The experiment results show that OESTM can effectively detect and track dynamic sentiment and topic.

Compared with the existing methods, the contributions of this work are fourfold:

- The proposed OESTM can effectively and jointly exploit the topic and sentiment information of social media reviews.
- In order to provide the more flexible method to select the number of topics, OESTM first fuses non-parametric HDP to topic/sentiment analysis model, which models each review document as a Dirichlet Process for topic discovery.
- In order to track the evolution of the topics and sentiment, OESTM employs the presented Time-

dependent CRFP to accomplish the model's regeneration.

- The main purpose of the sentiment and topic mixture model is to extract the sentiment and topics from social media reviews. We apply our OESTM model to discover the dynamic sentiment and topics with real social media data. We compare the performance of OESTM with ASUM, JST, HDP and LDA. The experimental results show that OESTM outperform these model in terms of generalization performance, model's complexity, and sentiment classification accuracy, which indicates the effectiveness of our dynamic non-parametric model.

2. Related Work

There are two research fields related specifically to this paper: topic-based sentiment analysis and non-parametric Dirichlet Process. In recent years, a great deal of interest has been attracted to sentiment analysis, as the amount of product/service review grows rapidly. 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 opinions and emotions detection³. Liu et al. presented an analysis framework to compare consumer sentiment polarities score of multiple produces by a supervised pattern discovery method⁴.

To determine sentiment polarities of the document, Pang et al. proposed a machine learning method, which adopts text categorization techniques and minimum cuts in graphs⁵. 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 sentiment⁶. 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 cell phone, without having to

read an overall product evaluation.

Motivated by this observation, researchers consider combining topic extraction to sentiment analysis, which is called topic-based sentiment analysis. As one of the state-of-the-art methods of topic model, Latent Dirichlet Allocation (LDA) model has gained popularity, which is a hierarchical Bayesian network. LDA builds robust topics' summaries in accordance with the multinomial probability distribution over words for each topic, and can further deduce the discrete distributions over topics for each document. In order to use time information to improve topics discovery, TOT proposed by Wang et al. models time jointly with word co-occurrence patterns based on LDA in an off-line fashion⁷. Meo et al. presented a matching algorithm, that allows dynamically and autonomously managing the evolution^{9,8}. Nonetheless, above-mentioned topic models have no capabilities of working in an on-line fashion. This stimulated researchers to search for an optimized model, and ultimately several online topic models have been proposed^{10,11,12}. Using temporal streams information, the dataset is divided by predefined time slice. At each time slice, documents are supposed to be exchangeable. But it is not true between documents across time slice. This core idea will be inherited by this paper. Based LDA and its extended models, many topic-based sentiment analysis models are proposed. Joint sentiment/topic model (JST)¹³ and Aspect and sentiment unification model (ASUM)¹⁴ are representatives of these. JST can implement the detection of sentiment and topic simultaneously based on LDA model. ASUM constrains the words in a single sentence to come from same polarity, which is called sentence-level JST model. 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)¹⁵ and dynamic joint sentiment-topic model¹⁶ are the rarely work to detect and track dynamic topic and sentiment based on probability topic model. 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 ap-

proach, 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.

So far, however, great mass of sentiment analysis models are based on LDA, which also inherit the defect of LDA model that the number of topics must be pre-determined. It is insufficient for the dynamic and massive social media data. The question can be resolved through replacing Dirichlet allocation by nonparametric Bayesian process^{17,18,19,20}. Dirichlet Process is a typical method for 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. Document could be modeled as a DP, and each word in document d is a target object that is created by the distribution words over a topic sampled from the distribution of document-based mixing vector over infinite number of topics. To allow sharing data among the collection of topics across documents, another non-parametric model, Hierarchical Dirichlet Process, was proposed, which used Dirichlet Processes as the Bayesian prior to solve the topics number determination problem.

Many models that integrating time information based nonparametric Bayesian have recently been proposed to improve topic discovery^{21,22}. In order to implement dynamically clustering analysis of topics, some nonparametric Bayesian-extended models have been proposed on others^{23,24}. But, actually there're several important differences between OESTM and the aforementioned models as following: (1)OESTM is the first dynamic sentiment-topic mixture model based on non-parametric HDP topic model; (2)In order to track the trend of the detected topics, OESTM first uses Time-dependent CRFP to realize the development and change of the topics over time; (3)OESTM implements Gibbs Sampling Process to obtain parameters at each time slice rather than executes global deduction, which is effective for updating timely evolution.

3. Methodology

3.1. Hierarchical Dirichlet Process

Before presenting the on-line evolutionary sentiment/topic modeling (OESTM), let us review the basic Hierarchical Dirichlet Process (HDP). The graphical representations of HDP and OESTM models are shown in Figure 1.

HDP (see Fig.1(a)) assumes document as x_d with $d \in \{1, \dots, D\}$, and N_d as the size of document d , there is a local random probability measures θ_d to denote topics distribution of document d . The random probability measure G_0 is a global topic distribution shared by all the documents collection, which is distributed as a Dirichlet Process with concentration parameter γ and base probability measure H . Each document d is generated based on local random measures θ that are also distributed as Dirichlet Process and conditionally independent given G_0 with concentration parameter α and base probability measure G_0 . Let $k_{d1}, k_{d2} \dots$ be independent random variables distributed as local measures θ . Each k_{di} is topic assignment of a single observation i^{th} word within the d^{th} document. Then the word x_{di} is generated from the conditional distribution $F(k_{di})$ given k_{di} . In order to simplify infer of the sampling process, F is often selected for multinomial distribution, and then forms conjugate distribution with base measure H . The likelihood is given by:

$$\begin{aligned} G_0 | \{\gamma, H\} &\sim DP(\gamma, H) \\ \theta_d | \{\alpha, G_0\} &\sim DP(\alpha, G_0) \\ k_{di} | \theta_d &\sim \theta_d & d \in \{1, \dots, D\} \\ x_{di} | k_{di} &\sim F(k_{di}) & i \in \{1, \dots, N_d\} \end{aligned} \quad (1)$$

The Hierarchical Dirichlet Process can readily be extended to more than two levels. That is, the base measure H can itself be a draw from a DP, and the hierarchy can be extended for as many levels as are deemed useful. In general, we obtain a tree in which a DP is associated with each node, in which the children of a given node are conditionally independent given their parent, and in which the draw from the DP at a given node serves as a base measure for its children.

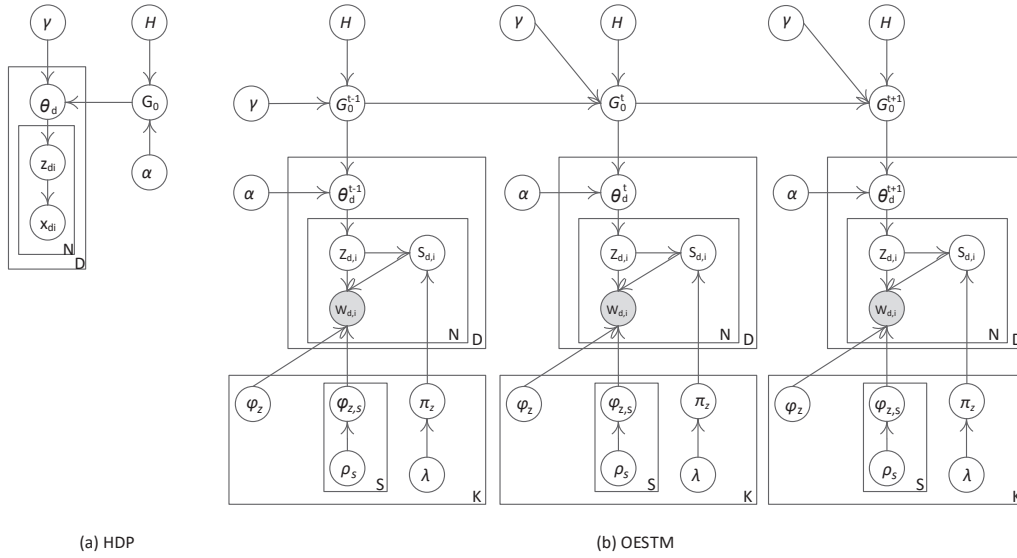


Fig. 1. OESTM is shown with one inspiration model: (a) HDP model (b) OESTM model.

3.2. On-line Evolutionary Sentiment/topic Modeling

While HDP has the capability of determining the appropriate number of latent topics, it is not adequate for tracking the trend of topics; it does not have the capability to combine sentiment labels into training procedure. This stimulates us to propose OESTM (see Fig.1(b)). The reviews dataset x will be divided according to time slice, $x = \{x^1, x^2, \dots, x^T\}$, 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, $x^t = \{x_1^t, x_2^t, \dots, x_{D_t}^t\}$, where D_t denotes the number of reviews within time slice t . Each review is presented with a series of words $x_d^t = (x_{di}^t)_{i=1}^{N_d^t}$, where N_d^t is the number of words within review x_d^t . Let us suppose a review is presented with the probability distribution over infinite topics, and topic is probability distribution over words. The target for on-line evolutionary sentiment/topic modeling 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 different time slices.

In order to further integrate the time information into HDP, base measure G_0 should be dynamically

calculated for each time slice, that is, the number of mixture components at each time point is unbounded; the components themselves can retain, die out or emerge over time; and the actual parameterization of each component can also evolve over time in a Markovian fashion. Through considering previous time slices, the base measure G_0^t at current time slice can be obtained as follows:

$$G_0^t | G_0, \gamma, \phi_{1:k} \sim DP(\gamma + \sum_{k=1}^{\Delta} E(v, \delta) \cdot d_k^{t-\delta}, \frac{G_0}{1 + \sum_{\delta=0}^{\Delta} E(v, \delta)} + \frac{\sum_{\delta=0}^{\Delta} E(v, \delta) G_0^{t-\delta}}{1 + \sum_{\delta=0}^{\Delta} E(v, \delta)}) \quad (2)$$

where decay function $E(v, \delta)$ represents the function of exponential kernel, $E(v, \delta) = \exp(-\delta/v)$, that manages the weight of topic k at time slice $t - \delta$. v and Δ define the decay factor of the time-decaying kernel and time windows that influences current time slice. Each epoch is independent when $\Delta=0$, and time is ignored when $\Delta = T$ and $v = \infty$. In between, the values of these two parameters affect the expected life span of a given component. The larger value of Δ and v , the longer expected life span of the topic, and vice versa. If we let d_k^t denotes the number of parameters in epoch t associated with component ϕ_k , then $\sum_{\delta} E(v, \delta) d_k^{t-\delta}$, the prior

weight of component k at epoch t . Furthermore, in order to incorporate sentiment polarity labels to realize sentiment classification, our model constructs the relation between topics and sentiment labels on the basis of some ideas of JST¹³. OESTM model implements sentiment analysis by adding an additional sentiment layer between the document and topic. Hence, OESTM is a four layer model, where sentiment labels are associated with documents, under which topics are associated with sentiment labels and words are associated with both sentiment labels and topics.

The formal definition of the generative process in OESTM model corresponding to the graphical model is as follows:

(1)Generate the global topic distribution at time slice $t: G_0^t | G_0, \gamma, \phi_{1:k}$

(2)Generate the neural words distribution for each topic: $\phi_k \sim H$

(3)Generate sentiment words distribution given topic and sentiment: $\phi_{k,s} \sim Dir(\rho)$

(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 word in document d :

(4.2.1)Draw an topic assignment: $z_{d,i} \sim Mult(\theta_d^t)$

(4.2.2)Generate sentiment distribution of topic: $\pi_z \sim Dir(\lambda)$

(4.2.3)Draw a sentiment assignment: $s_{d,i} \sim Mult(\pi_z)$

(4.2.4) $w_{d,i} \sim \phi_z, \phi_{z,s}$

3.3. Time-dependent CRFP

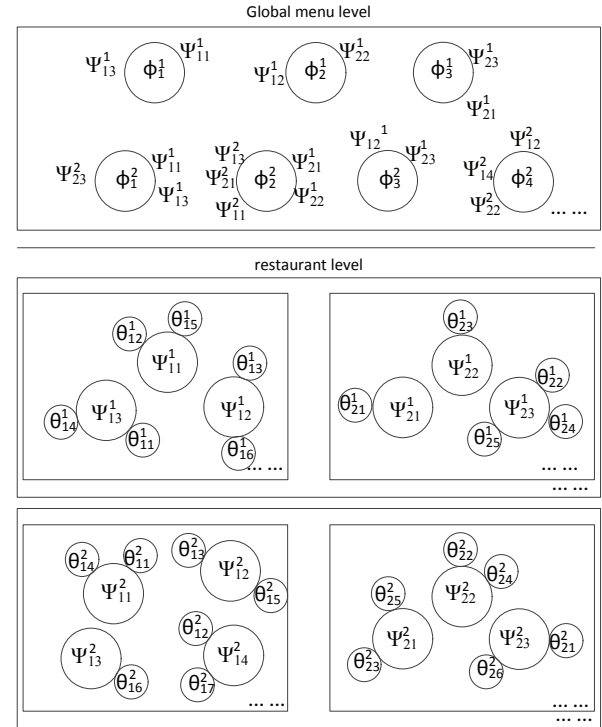


Fig. 2. Time-dependent CRFP.

Chinese restaurant process (CRP) is a metaphor, which is closely connected to Dirichlet Processes, and therefore useful in applications of nonparametric Bayesian methods including Bayesian statistics. CRP is a discrete-time stochastic process, analogous to seating customers at tables in a Chinese restaurant. Chinese restaurant franchise process (CRFP) based on Chinese restaurant process extended Chinese restaurant process to allow multiple restaurants which share a set of dishes, which is a random process that products an interchangeable division of data points and allows multiple data points to share a set of topics³². CRFP is usually employed to simulate HDP process. In this metaphor each restaurant maintains its set of tables but shares the same set of mixtures. A customer at restaurant can chose to sit at an existing table with a probability proportional to the number of customers sitting on this table, or start a new table with probability and chose its dish from a global distribution. As the expansibility and hierarchy, CRFP is widespread applied in

non-parametric models. Although CRFP has the capacity of constructing data point by using a set of topics and allowing topics' number to be infinite, it is static process and cannot track the development of the latent topics and probability distribution of words. This paper uses CRFP to construct a mixture model of grouped data. However, we replace the first level of CRFP with a novel time-dependent random process. The modified CRFP is called time-dependent CRFP, which can estimate the optimal number of topics for current time slice by considering influences from a previous time slices.

A wide array of metaphors and conceptions are employed in CRFP. An example for time-dependent CRFP is shown in Figure 2. A document is represented as a restaurant, and words are represented as customers in the restaurants. Words that convey homogeneous semantic theme are grouped together as customers of similar taste sit at the same table. A dish is chosen for each table from the global menu of dish tree, which corresponds to topic assignment for each group of words. At restaurant level, each restaurant is denoted by a rectangle and consumers (small circles) sit around different dining-tables (big circles) associated with a dish in this restaurant. At global menu level, the set of dish (topic), is served in common for all of restaurants. x_{di}^t represents the i^{th} consumers in restaurant d for time slice t . θ_{di}^t represents the dish enjoyed by this customer, ψ_{dj}^t represents the dish for j^{th} table and ϕ_k^t represents dish k on menu. In order to record the relation among consumers, tables and dishes, this paper gives two index variables. b_{di}^t represents index of table, and k_{dj}^t represents index of dishes in restaurant d for time slice t . So we have $\Psi_{db_{di}^t}^t = \theta_{di}^t$ and $\phi_{k_{dj}^t}^t = \psi_{dj}^t$. This paper uses time-dependent CRFP to implement the assignment of customer to dinning-table and the relation between tables and dishes.

Table assignment

At time slice t , the i^{th} customer comes in restaurant d . This customer can pick j^{th} dining-table with probability:

$$n_{dj}^t / (n_d^t - 1 + \alpha) \quad (3)$$

where n_{dj}^t and n_d^t denote respectively the number of customer around dining table j and in restaurant d ,

enjoying dish $\psi_{db_{dj}^t}^t$ ordered from global menu by the first customer who sits at that table. α is a parameter governing the likelihood of choosing of a new table. Alternatively, this customer can select a new table with probability:

$$\alpha / (n_d^t - 1 + \alpha) \quad (4)$$

Dish assignment

We firstly give some notations that will be adopted for dish assignment. Customers in the restaurant sit around different tables and each table is associated with a dish (topic) ψ according to the dish menu (global topics). Let Nt_i^t represents 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 restaurants is represented as d_k^t at time slice t , which is calculated as follows:

$$d_k^t = \sum_{i=1}^{D_t} \sum_{j=1}^{Nt_i^t} 1[\Psi_{ib_{ij}^t}^t = k] \quad (5)$$

where the $[\Psi_{ib_{ij}^t}^t = k]$ is a conditional expression. If the condition is met, the number of dining-tables that have ordered dish k is added one.

In order to integrate the historical influences from a previous time slices, another parameter d_k^t is defined as follows:

$$d_k^t = \sum_{\delta=1}^{\Delta} E(v, \delta) \cdot d_k^{t-\delta} \quad (6)$$

So the popularity of a topic at epoch t depends both on its usage at this epoch, d_k^t as well as its historic usage at the proceedings epochs, $d_k^{t-\delta}$. If the customer picks a new table and chooses a dish k ordered by previous customs from the menu at time slice t , the probability is as follows:

$$\frac{d_k^t + d_k^{t-\gamma}}{\sum_{s=1}^{K_t} d_s^t + d_s^{t-\gamma} + \gamma} \quad (7)$$

where K_t is the number of dishes at time slice t . This means that a topic is considered dead only when it is unused for a consecutive epochs. For simplicity, we let d_k^t for newly-born topics at epoch t , and d_k^t for topics available to be used (i.e having $d_k^t > 0$) but not yet used in any document at epoch t .

If this dish is ordered by consumers in Δ previous time slices but not yet ordered by consumers at

t time slice, then changing the distribution of this dish in Markova fashion: $\phi_k^t | \phi_k^{t-1} \sim p(\cdot | \phi_k^{t-1})$ (i.e. improve the word probability distribution of topic). The probability is as follows:

$$\frac{d_k^t}{\sum_{s=1}^{K_t} d_s^t + d_s^t + \gamma} P(\cdot | \phi_k^{t-1}) \quad (8)$$

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

$$\frac{\gamma}{\sum_{s=1}^{K_t} d_s^t + d_s^t + \gamma} H \quad (9)$$

More formally, getting aforementioned formula together, we have

$$\begin{aligned} \theta_{di}^t | \theta_{di}^t, \alpha, \psi^{t-\Delta:t} &\sim \sum_{j=1}^{j=B_d^t} \frac{n_{dj}^t}{n_d^t - 1 + \alpha} \xi_{\psi_{dj}^t} + \frac{\alpha}{n_d^t - 1 + \alpha} \xi_{\psi_{dj}^{new}} \\ \psi_{dj}^{new} | \psi, \gamma &\sim \sum_{k: d_k^t > 0} \frac{d_k^t}{\sum_{s=1}^{K_t} d_s^t + d_s^t + \gamma} \xi_{\phi_k^t} \\ &+ \sum_{k: d_k^t = 0} \frac{d_k^t}{\sum_{s=1}^{K_t} d_s^t + d_s^t + \gamma} P(\cdot | \phi_k^{t-1}) + \frac{\gamma}{\sum_{s=1}^{K_t} d_s^t + d_s^t + \gamma} H \end{aligned} \quad (10)$$

where B_d^t represents the number of dinning-tables in restaurant d at time slice t and ξ are probability measures concentrated at ψ and ϕ .

4. Approximate Posterior Inference

In this section, the collapsed Gibbs sampling algorithm is utilized^{25,26} for posterior sampling the assignments of the tables, and the dishes that serve a specific table in each restaurant. For performing Gibbs sampling, Markov Chain Monte Carlo (MCMC) based on time-dependent CRFP is constructed to integrate out parameters b_{di}^t , k_{dj}^t , ϕ_k^t and s_{di}^t into joint probability distribution and states converges to a sample from this joint probability distribution, where b_{di}^t represents index of table assigned to customer x_{di}^t , k_{dj}^t represents index of dish enjoyed by table j, the posterior probability distribution of k, and the sentiment assignment of customer x_{di}^t respectively in restaurant d at time slice t. In order to use Gibbs sampling algorithm for posterior inference, we add a superscript i to a variable, indicate

the same quantity it is added to without the contribution of object i. For example $n_{db}^{-t, di}$ is the number of customers sitting on table b in document d in epoch t without the contribution of word x_{di}^t .

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

$$f_k^{-x_{di}^t}(x_{di}^t) = \frac{p(x_{di}^t | -x_{di}^t, b, k) = \frac{\int f(x_{di}^t | \phi_k^t) \prod_{d' \neq di, z_{d' t} = k} f(x_{d' t}^t | \phi_k^t) h(\phi_k^t | \eta) d(\phi_k^t)}{\int \prod_{d' \neq di, z_{d' t} = k} f(x_{d' t}^t | \phi_k^t) h(\phi_k^t | \eta) d(\phi_k^t)}} \quad (11)$$

As base measure H and multinomial distribution (i.e. word distribution with topic) is conjugate distribution, the above formula can be simplified as:

$$f_k^{-x_{di}^t}(x_{di}^t = v) = \frac{n_k^{-x_{di}^t, v} + \eta}{n_k^{-x_{di}^t} + V\eta} \quad (12)$$

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

$$f_k^{-x_{db}^t}(x_{db}^t) = \frac{n_k^{-x_{db}^t} + V\eta}{n_k^{-x_{db}^t} + n^{-x_{db}^t} + V\eta} \frac{\prod_v \Gamma(n_k^{-x_{db}^t, v} + n^{-x_{db}^t, v} + \eta)}{\prod_v \Gamma(n_k^{-x_{db}^t, v} + \eta)} \quad (13)$$

According to time-dependent CRFP and sentiment analysis requirement, we employ the four stages of 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:

$$p(b_{di}^t = b | -b_{di}^t, k_{t-\Delta t}^t, x_{di}^t) \propto \begin{cases} n_{db}^{-t, di} \cdot f_{k_{db}^t}^{-x_{di}^t}(x_{di}^t) \\ \alpha P(k_{db}^{new} = k | k_{t-\Delta t}^t, -b_{di}^t, x_{di}^t) \end{cases} \quad (14)$$

Several points are in order to explain equation (14). There are two choices for word: either to sit on an existing table, or to sit on a new table and choose a new topic. In the second case, we need to sample a topic for this new table which leads to the equation (15), where the probability to sit on a new table can be constructed by marginalizing over all available dishes.

$$P(k_{db}^{new} = k | k_{t-D:t}^{-tdi}, -b_{di}^t, x_{di}^t) \propto \begin{cases} (d_k^{-tdb^{new}} + d_k^t) f_k^{-x_{di}^t}(x_{di}^t) \\ k \text{ is being used : } d_k^{-tdb^{new}} > 0 \\ d_k^t f_k^{-x_{di}^t}(x_{di}^t) \\ k \text{ is available but not used : } d_k^t > 0 \\ \gamma f_{k_{new}}^{-x_{di}^t}(x_{di}^t) \\ k \text{ is a new topic} \end{cases} \quad (15)$$

Sampling topic k Once the assignment of tables is complete, the posterior sampling dish k_{dj}^t is similar as the above equation, but we need to consider the probability of small groups of words (like consumers on a given table). The conditional probability is estimated as follows:

$$P(k_{db}^t = k | k_{t-D:t}^{-tdi}, x_{db}^t) \propto \begin{cases} (d_k^{-tdb^{new}} + d_k^t) f_k^{-x_{db}^t}(x_{db}^t) \\ k \text{ is being used : } d_k^{-tdb^{new}} > 0 \\ d_k^t f_k^{-x_{db}^t}(x_{db}^t) \\ k \text{ is available but not used : } d_k^t > 0 \\ \gamma f_k^{-x_{db}^t}(x_{db}^t) \\ k \text{ is a new topic} \end{cases} \quad (16)$$

Sampling topic is important as it potentially changes the membership of all data sitting at table and leads to a well-mixed MCMC.

Sampling ϕ_k Given b , k and observed x , the posterior conditional probability distribution of every

ϕ_k only depends on whether all consumers enjoyed their dish, which is estimated as follows:

$$P(\phi_k^t | t, k, x, \phi_k^{-t}) \propto h(\phi_k^t | \eta) \prod_{di: k_{db_{di}} = k} f(x_{di}^t | \phi_k^t) \quad (17)$$

Sampling s Let s_{di}^t denote the sentiment polarity for i^{th} word in document d at time slice t . After topic inference, a sentiment needs to be chosen for the very word under this topic. That is we then sample the sentiment polarity s_{di}^t for every customer enjoying this dish after dish inference. We apply a Dirichlet allocation as the prior for sentiment distribution. Under the aspect k at epoch t , the sentiment inference can be made as equation (18), where $n_{kss}^{-t, x_{di}}$ is the number of word x has been assigned to sentiment s under topic k except x_{di} , $n_{ks}^{-t, x_{di}}$ the number of words have been assigned to sentiment s under topic k except x_{di} , $n_k^{-t, x_{di}}$ is the word count of topic k except x_{di} , and S is the number of sentiment.

$$p(s_{di}^t = s | k) \propto \frac{n_{kss}^{-t, x_{di}} + \rho_s}{n_{ks}^{-t, x_{di}} + V \cdot S} \cdot \frac{n_{ks}^{-t, x_{di}} + \lambda}{n_k^{-t, x_{di}} + S \cdot \lambda} \quad (18)$$

OESTM model adopts indirect method of MCMC sampling algorithm to infer distribution parameters θ , ϕ and π . Using these distribution parameters, the latent topics, the sentiment polarity and represent words of topic can be mined.

5. Experiments

5.1. Datasets Presetting

To accomplish experiments we use two different reviews 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 previously²⁷. These datasets were preprocessed by (1) deleting stop-words and non-English alphabets; (2) deleting low frequency with appearances be low six times, as well as short reviews that are shorter than seven words; (3) adopting Snowball algorithm to stemming for words.

Sentiment analysis is much more challenging than topics detection, because consumers express

their attitudes through subtle manner, but topic detection is simply implemented on the basis of words co-occurrence. One technique increasing the accurateness of sentiment analysis is to integrate prior data, i.e., sentiment lexicon. In this section, three sentiment lexicons, Paradiams²⁸, Mutual Information (MI)²⁹ and MPQA³⁰, will be used to improve the sentiment classification accuracy. Table 1 shows the properties of datasets and sentiment lexicon information adopted in our experiments.

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	Sentences
Restaurant	41,715	3,616,286	199,265
			Sentence length
			≤4 ≤12 ≤20
			26.00% 79.05% 93.00%
Hotel	34,157	4,287,183	239,639
			Sentence length
			≤4 ≤12 ≤20
			34.00% 82.14% 95.86%

Unless otherwise stated, in this experiment these parameters were set according to the following values: the hyper parameter η of base probability measure H was set as 10; concentration parameter γ and α were respectively obtained by vague gamma prior, $\gamma \propto \Gamma(1, 0.1)$, $\alpha \propto \Gamma(1, 1)$. Continuous time slices $\Delta=4$. The number of time slices is set as 20, $T=20$. To enable comparison to other models, parameters that are LDA-based models were set to the following: Dirichlet hyper parameter $\alpha = 0.5, \beta = 0.02$. The hyper parameters λ and ρ were set as 1.0 and $\{10^{-7}, 0.01, 2.5\}$. The parameter v in exponential kernel is set as 0.5.

5.2. Perplexity

The density measurement, expressing the potential configuration of data, is the intention of document modeling. Measuring the model's universal performance on formerly unobserved document is general method to estimate that. Perplexity is a canonical measure of goodness that is 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 will employ perplexity of the statistical model on test review datasets to measure the generalization performance when the performance of the model starts to steady state. The lower perplexity manifests the better generalization performance will be. We classified the data into 80% for training set and 20% for testing set, where classification proportion is consistently across time slices. Formally, given the testing dataset, the perplexity value can be calculated as follows.

$$perplexity(x) = \exp\left(-\frac{\sum_j \sum_i N_j \log_2 p(x_{ji})}{\sum_j N_j}\right) \quad (19)$$

Four models (LDA, ASUM, HDP and OESTM) will be adopted over two datasets to compare the perplexity performance. Firstly, the number of topics for LDA and ASUM was set to 20. The first two rows of Figure 3 illustrate 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 to 100. The second two rows of Figure 3 shows the result of the perplexity as a function of the number of topics. Because nonparametric Bayesian models, HDP and OESTM, are irrelevant to the number of topics, the values of perplexity for them are fixed.

As shown in Figure 3, HDP and OESTM models can effectively work for documents clustering than LDA and ASUM models, i.e., have better generalization performance and presents lower perplexity value. From the results it can also be seen that for LDA and ASUM model, picking the right number of topics is key to getting good performance. When the number of topics is too small, the result suffers from under-fitting. However, blindly increasing the number of topics could on the other hand make over-fitting. On the other hand, the number of topics obtained intelligently under non-parametric OESTM and HDP model is consistent with this range of the best-fitting LDA and ASUM model. Otherwise, OESTM model is slightly better than HDP, which implies that OESTM can well find topics through

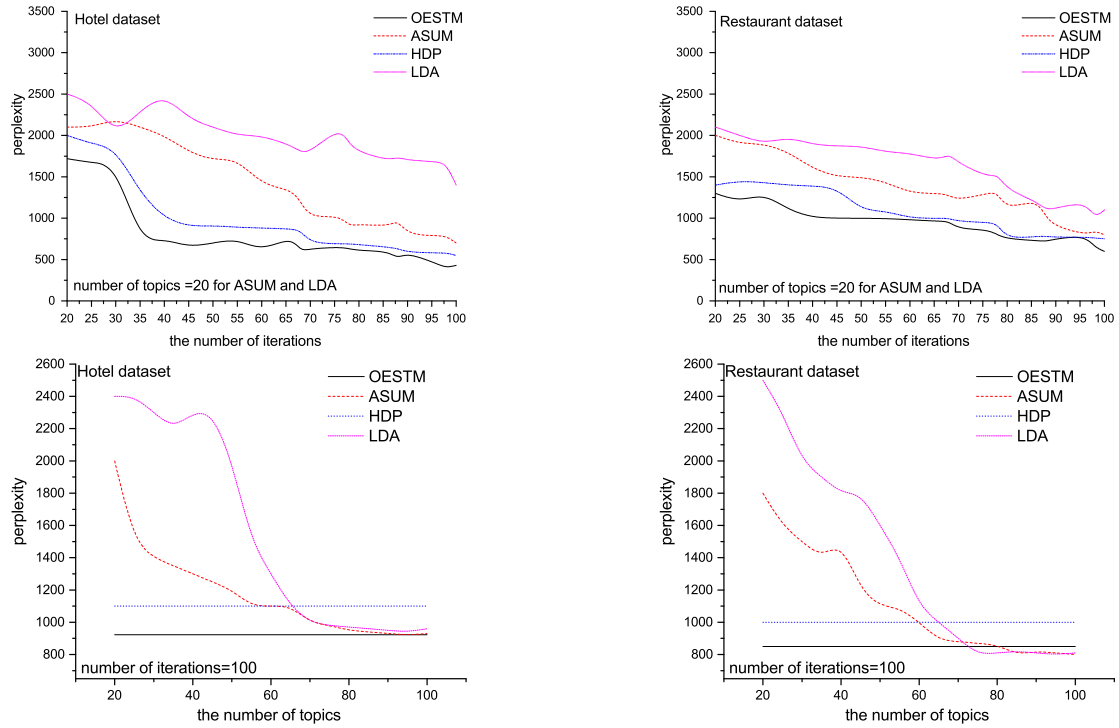


Fig. 3. perplexity score on two datasets against different models. The first two rows are perplexity for different numbers of iterations. The last two rows are perplexity for different numbers of topics.

incorporating information of time and sentiment to provide a better prior for the content of emerging documents. However, perplexity values of OESTM presents overall the similar results with HDP model.

5.3. Complexity

Nonparametric Bayesian methods are often used to sidestep model selection and integrate over all instances (and all complexities) of a model at hand (e.g., the number of clusters). The model, though hidden and random, still lurks in the background. Here we study 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 will measure the Complexity of model to evaluate non-parametric Bayesian models³¹. To implement the complexity of model the definition of a topic's complexity will firstly be given. The complexity of a topic is in proportion to the number of words allocated to this topic, i.e., the complexity of a topic is zero if no unique words

are allocated to this topic, otherwise, the number of unique words allocated to this topic. So we express the complexity of a topic k as follows:

$$Complexity_k = \sum_d 1[(\sum_n 1[K_{di} = k]) > 0] \quad (20)$$

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

$$Complexity = \#topics + \sum_k Complexity_k \quad (21)$$

The complexity analysis considers the number of topics employed to describe the dataset according to Equation 20. A higher complexity of a model shows this model need more topics to express the dataset - that is, the dataset is classified into more dimensions. So a lower complexity manifests the better model will be, on condition that the generated experiment results for perplexity is alike. Figure 4 illustrates the result of the complexity as a function

of the number of iterations of the Gibbs sampler for the two different datasets. As shown in Figure 4, OESTM model has lower average complexity and is better than the HDP in all cases. In spite of the advantage of OESTM model in terms of perplexity might be small than HDP, but taking account of average complexity, the overall effect of OESTM model is more superior.

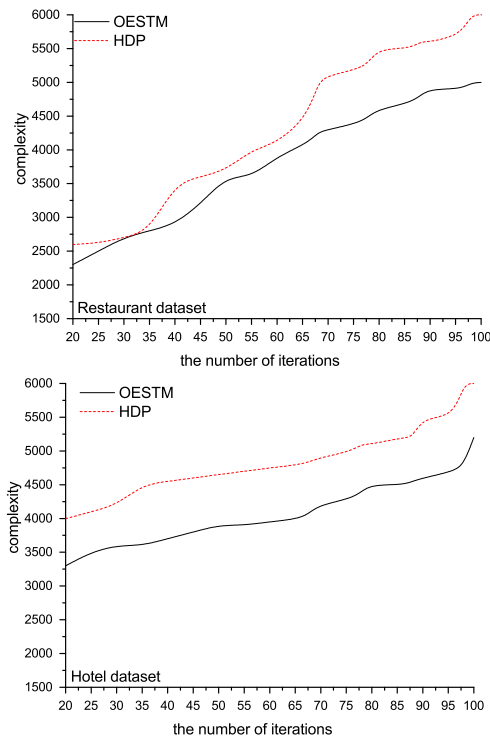


Fig. 4. Model complexity comparison of HDP and OESTM models.

5.4. Sentiment Classification

Three sentiment polarity labels, positive, negative and neutral, are selected to associate to all words. $l(x)$ is used to express the polarity label for word x . $l(x)=1$ if label is positive, -1 negative and 0 neutral. Firstly, each word term matches with sentiments lexicon. The sentiment polarity label will be assigned to a word, if that word matches with one of words in sentiments lexicon. Otherwise, a randomly sentiment polarity label from three labels is selected for a word.

After posterior sampling the assignments, i.e., MCMC reaches stead state, each word within a doc-

ument is attached a sentiment polarity label using a sentiment polarity of a local topic that this word has been assigned. The sentiment polarity of local topics can be further obtained according sentiment distribution π within a local topic. The document sentiment value can be calculated as follows:

$$s_d^t = \sum_{x \in x_d^t} l(x) \varphi_{z_d, x} s_{dx}^t \quad s_d^t \in [-1, 1] \quad (22)$$

Where s_d^t denotes the sentiment value of document d . if this value is less than 0 , the document is categorized as negative. If this value is more than 0 , the document is positive. Otherwise, the document is neutral. Restaurant and Hotel corpus uses 5-stars rating system. We assume reviews with 1 or 2-stars are negative. Reviews are considered as positive with 4 or 5-stars. Reviews with 3-stars will not be considered, i.e., reviews are being categorized either as positive or negative, without the alternative of neutral. In this subsection, we measure the sentiment classification accuracy in terms of different sentiment lexicons for three sentiment models, JST, ASUM and OESTM.

Table 2 presents the predictive results of sentiment classification accuracy. It can be observed from Table 2 that through incorporating only 21 pos. and 21 neg. paradigm words, OESTM model merely acquired a relatively poor 74.3% overall accuracy and JST and ASUM acquire 66.8% and 73.5% respectively based on Restaurant dataset. Similar results can be observed for Hotel dataset. By combining the top 20 words based on MI scores with paradigm words, it can be illustrated that there's lots of improvement in classification accuracy with 2%, 2% and 14% for JST, ASUM and OESTM respectively. But classification accuracy is 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 decrease around 1%, 4% and 8% respectively for JST, ASUM and OESTM based on Restaurant dataset. Similar experiment results can be found for Hotel dataset. As shown in it, in all settings, the accuracy of sen-

Table 2. Sentiment classification accuracy comparison.

Sentiment lexicon									
paradiams									
Paradiams+MI									
MPQA									

Dataset	JST(%)			ASUM(%)			OESTM(%)		
Restaurant	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
	70.2	78.7	74.54	74.6	84.3	79.45	86.6	88.6	87.6
	68.2	78.8	73.5	72.4	80.7	76.55	78.4	84.7	81.55

Dataset	JST(%)			ASUM(%)			OESTM(%)		
Hotel	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
	72.3	80.7	76.5	76.4	86.5	81.45	87.9	89.6	88.75
	68.6	79.8	74.2	72.4	81.4	76.9	74.6	86.7	80.65

timent classification of OESTM model always performs better than JST and ASUM sentiment models.

5.5. Hyperparameter Sensitivity

There are two principle parameters defined in OESTM model, namely ν and η , where ν defines the time decaying kernel, and η is the hyper parameter of base probability measure H . To assess the sensitivity of OESTM to hyper parameters' settings, we conducted a sensitivity analysis in which we hold all hyper parameters fixed at their default values, and vary one of them. Held-out likelihood (LL) is widely used in the topic modeling community to compare how well the trained model explains the held out data. In this subsection, we use parameter variations as a proposal for calculating the test LL based on Restaurant dataset for OESTM. We should note here that the order of the process can be safely set to T , however, to reduce computation, we can set Δ to cover the support of the time-decaying kernel, i.e., we can choose Δ such that $E(\nu, \Delta)$ is smaller than a threshold, say .001. The results are shown in Figure 5.

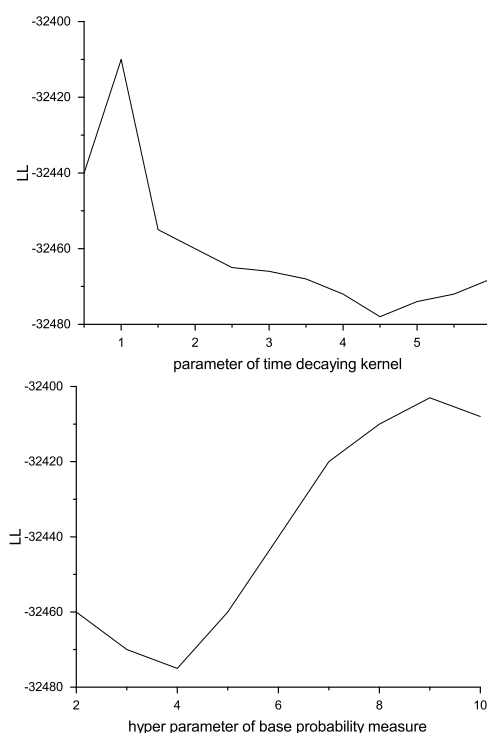


Fig. 5. Held-out Likelihood for different parameters.

Firstly, While varying ν , we fixed $\Delta = T$ to avoid biasing the result. We noticed that when $\nu = 6$, some topics weren't born and were modeled as a contin-

uation of other related topics. v depends on the application and the nature of the data. In the future, we plan to place a discrete prior over v and sample it as well. Secondly, the best setting for the variance of base measure is from [5, 10], which results in topics with reasonably sparse word distributions.

5.6. Evolutionary Senti-Topic Discovery

OESTM is proposed to produce topics coupled with a sentiment for each time slice, easier facilitating customers to reveal how development and change of topics and topic sentiment scores developed over time. In this experiment, the probability distribution for words given topic k , i.e. neural words that are assigned to topic k was estimated using ϕ_k , the distribution for words given topic k and sentiment label s was estimated using $\phi_{k,s}$. In order to track the sentiment trend for each topic, the sentiment scores of topic k with sentiment polarities for time slice t are defined as following.

$$s_k^t = \begin{cases} \sum_{x \in x_k^t} \phi_{k,\text{pos.}} \text{ positive score} \\ - \sum_{x \in x_k^t} \phi_{k,\text{neg.}} \text{ negative score} \end{cases} \quad (23)$$

Where x_k^t is the words set that are assigned to topic k . Average sentiment score for each topic is the sum of score for positive or negative. Then, the sentiment scores for each topic series form the sentiment time series $\{\dots, s_k^{t-1}, s_k^t, s_k^{t+1}, \dots\}$.

In this subsection, the topic word probability distributions and the sentiment word probability distributions of topics for different time slice will be firstly obtained by OESTM. Then each topic' sentiment score for different time slices will be calculated through equation 23. Four example topics were used: the first three are selected by the probability values, ranging from high to low, while the last one is selected randomly. This is coupled with topic and senti-topic words probability distributions of three time slices in the left of Figures 6 and 7. The first five topic words and ten senti-topic words are picked attached probability. The emotional changes of left four topics, reflected by sentiment score, for different time slices are shown at the right of Figure 6 and Figure 7.

For example, the first topic word for the extracted topic 1 based on Restaurant dataset is “meat” with probability 0.181373, and the senti-topic words coupled with positive and negative label at time slice 12 are “good” and “dry” with probability 0.161267 and 0.152256 respectively. Topic 4 based on Restaurant dataset is emerging at time slice 10, which is inherited at time slice 13. After time slice 13 this topic is not be identified and will die at time slice $13+\Delta$. This proves the efficacy of OESTM in achieving the birth, death and inherit of topics. The topic 4 based on Hotel dataset is born at time slice 3, which is observed from sentiment score graph, but do not been identified at time slices 4, 6, 9, 14 and 17. This topic has no died, because it is used for subsequent consecutive Δ time slices, rather than been inherited through considering influence from previous time slices to the updated time slice at time slices 5, 7, 8, 10, 11, 12 and 13. This proves more that our proposed model has capacity to inherit topics that are identified at previous consecutive Δ time slices.

The sentiment score reflects the topic' emotional status. For example, the sentiment average score for topic 2 based on Restaurant dataset at time slice 7 is -0.49, which indicates that a lot of negative feedback during this period on health topic have been received. This can attract the attention of the company, and supervise and urge it to improve related service.

6. Conclusions

In this paper we addressed the problem of modeling time-dependent review corpus. On-line evolutionary sentiment/topic modeling (OESTM) is proposed, which can adapt number of topics, the topic words (without sentiment information) distributions of topics, the senti-topic words (attached sentiment polarity labels) distributions of topics, and track the topics' emotional development over time slice. As far as we know, we are the first to deal with a time-dependent reviews through non-parametric Bayesian topic model in order to implement topic-based sentiment analysis. At each time slice, a topic-clustering with the estimated best cluster number based on time-dependent CRFP, smoothing with

		Sentiment words		
topics				
{topic words}		time slice=12	time slice=13	time slice=14
Topic 1	Pos.	[.161267]good	[.159117]good	[.176476]delicious
[.181373]		[.116727]thin	[.099280]delicious	[.122461]tasty
meat		[.046905]fresh	[.048620]favorite	[.051856]crisp
[.180249]		[.030135]tasty	[.031287]crisp	[.028462]moist
taste		[.023433]pretty	[.025345]tasty	[.028462]tender
[.095320]		[.017563]crisp	[.025345]pretty	[.028462]fresh
flavor		[.017563]delicious	[.018025]fresh	[.018317]many
[.095320]		[.011377]better	[.018025]super	[.018317]chicken
taste		[.011377]dainty	[.010998]dainty	[.014727]nice
[.089622]		[.006203]pleasantly	[.010998]mended	[.006301]better
salti	Neg.	[.152256]dry	[.186564]nasty	[.093328]nasty
		[.049259]too	[.069212]too	[.063519]sauc
		[.032073]complains	[.047342]greasy	[.063519]unworth
		[.032073]salty	[.030025]spicy	[.030031]salty
		[.024710]pork	[.030025]bland	[.030031]dry
		[.016991]little	[.024218]salty	[.023185]lack
		[.01302925]bad	[.016953]undercooked	[.018255]hard
		[.012252]staleness	[.016953]disappoint	[.009817]tough
		[.012252]undercooked	[.012101]pork	[.009817]little
		[.012252]lack	[.012101]staleness	[.006605]crust

		Sentiment words		
topics				
{topic words}		time slice=12	time slice=13	time slice=14
Topic 2	Pos.	[.143600]clean	[.130775]clean	[.172245]clean
[.110833]		[.124085]quiet	[.122300]bright	[.156782]quiet
healt		[.094502]atmospher	[.091423]quiet	[.111366]happy
[.170562]		[.071260]well	[.076745]resort	[.094653]spacious
dishware		[.052693]bright	[.076745]capacious	[.078231]well
[.098240]		[.052693]spacious	[.050057]joyful	[.048222]better
sanitary		[.034245]date	[.050057]well	[.048222]resort
[.092341]		[.034245]fun	[.018736]great	[.048222]nice
cup		[.019221]great	[.018736]nice	[.020012]bright
[.089225]		[.009144]capacious	[.009325]best	[.020012]fun
table	Neg.	[.136729]dirty	[.141611]ugly	[.120013]ugly
		[.129437]ugly	[.140078]dirty	[.112342]dirty
		[.110985]crowd	[.133781]noise	[.087932]worse
		[.094551]noise	[.095820]badly	[.086542]badly
		[.069723]smoking	[.088570]smoking	[.076320]noise
		[.050956]music	[.088570]loud	[.076320]loud
		[.050956]talk	[.088570]messy	[.076320]smoking
		[.050956]look	[.041955]talk	[.021917]unhappy
		[.013000]loud	[.021063]unhappy	[.009777]many
		[.013000]messy	[.021063]many	[.009777]messy

		Sentiment words		
topics				
{topic words}		time slice=12	time slice=13	time slice=14
Topic 3	Pos.	[.160128]friendly	[.150057]friendly	[.124361]comfortable
[.201780]		[.100231]comfortable	[.137392]better	[.118435]friendly
serve		[.100001]helpful	[.111312]comfortable	[.092352]courteous
[.190361]		[.096931]nice	[.111312]satisfactory	[.063421]extremely
attendant		[.096931]feel	[.092351]courteous	[.063421]attentive
[.182125]		[.040013]pleasant	[.055402]fast	[.063421]professional
staff		[.040013]fast	[.036937]short	[.023342]helpful
[.182125]		[.040013]short	[.036937]helpful	[.008321]pleasant
waiter		[.009365]satisfactory	[.023432]professional	[.008321]fast
[.099455]		[.009231]courteous	[.012925]good	[.004235]nice
upkeep	Neg.	[.110763]slow	[.099961]rude	[.117635]rude
		[.083763]rude	[.099384]slow	[.108632]time
		[.083762]badly	[.089342]time	[.108632]slow
		[.070387]wait	[.070093]long	[.090983]attitude
		[.070387]rude	[.040032]rude	[.080371]badly
		[.059973]long	[.009127]attitude	[.060991]wait
		[.040834]poor	[.009127]cold	[.060991]long
		[.040834]attitude	[.009127]wait	[.053872]poor
		[.040834]very	[.005325]badly	[.028884]express
		[.009322]express	[.005325]very	[.007678]unprofess.

		Sentiment words		
topics				
{topic words}		time slice=12	time slice=13	time slice=14
Topic 4	Pos.		[.047521]feel	
(random selected)			[.047521]worth	
[.098235]			[.037635]well	
roach			[.009466]cost	
[.078232]			[.007873]cheap	
health			[.007873]people	
[.067323]			[.007873]good	
bug			[.004873]swtte	
[.067323]			[.000225]inexpensive	
dish			[.000225]juicy	
[.046345]	Neg.		[.212319]sick	
sour			[.179835]too	
			[.130021]disgusting	
			[.130021]droppings	
			[.079873]split	
			[.053493]angry	
			[.033423]outrageous	
			[.033423]unclean	
			[.033423]bad	
			[.020025]unworthy	

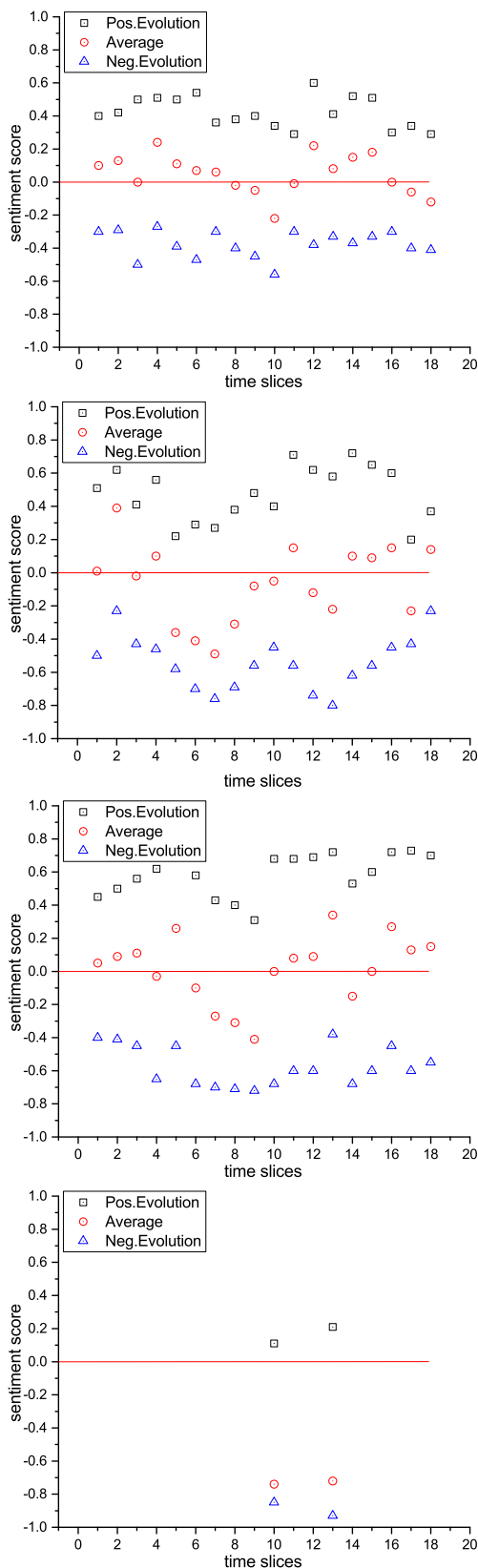


Fig. 6. Evolutionary senti-topic discovery based on Restaurant dataset. Left are discovered topics, topic words and senti-topic words. Right are the sentiment emotional changes of left topic.

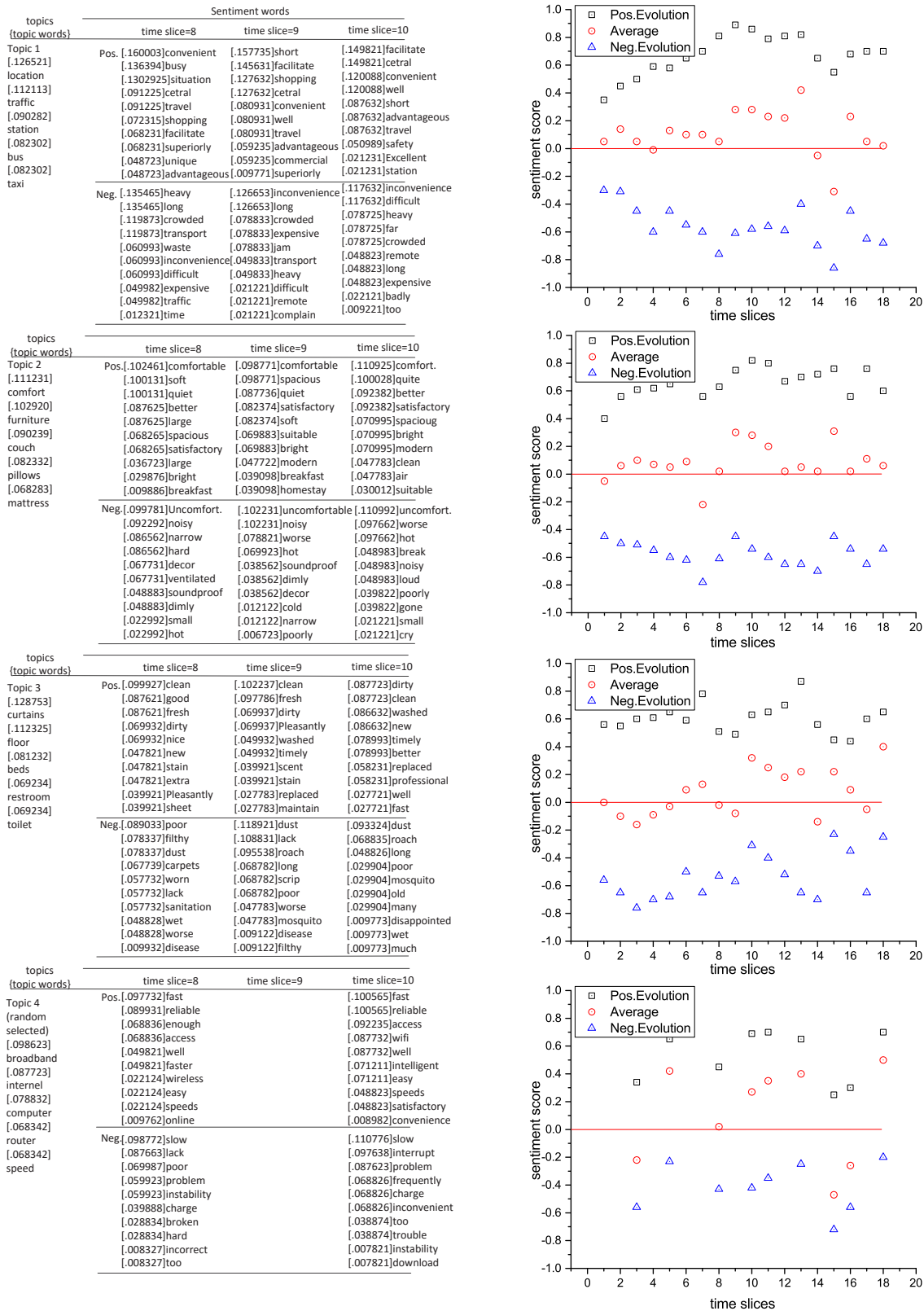


Fig. 7. Evolutionary senti-topic discovery based on Hotel dataset. Left are discovered topics, topic words and senti-topic words. Right are the sentiment emotional changes of left topic.

the topics' inheritance by considering historical influences from a previous time slices and sentiment analyzing for latent topics are automatically implemented. A collapsed Gibbs sampling algorithm is utilized to infer parameters. To evaluate the effectiveness of our proposed algorithm, we collect a real-world dataset to conduct various experiments. The preliminary results showed superiority of our proposed model over several state-of-the-art methods on generalization performance, lower complexity, accurate sentiment classification.

One of the limitations of our model is that it requires setting the time span of each epoch. In the future, we will consider other time dependency modes to optimize dynamic parameter inference.

Acknowledgments

This work is supported by the National Natural Science Foundation of China No.61303131, No. 61672272, No. 60973040;

1. D. M. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, 3, 993-1022(2003)
2. Y. Teh, M. Jordan, M. Beal, and D. Blei, "Hierarchical Dirichlet Processes," *Journal of the American Statistical Association*, 101,1566-1581(2006)
3. Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan, "Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns," in the confe. on HLT and EMNLP, 355-362(2005)
4. B. Liu, M. Hu, and J. Cheng, "Opinion Observer: Analyzing and Comparing Opinions on the Web," in confe. of World Wide Web, 342-351(2005)
5. B. Pang and L. Lee, "A sentimental education: Sentiment analysis using sub-jectivity summarization based on minimum cuts," in *Proc. of ACL*, 271-278(2004)
6. B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up: sentiment classification using machine learning techniques," in *EMNLP of the ACL on Empirical methods in natural language processing*, 79-86(2002)
7. Wang, Xuerui, and A. Mccallum, "Topics over time:a non-Markov continuous-time model of topical trends," *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ACM*, 424-433(2006).
8. P. D. Meo, F. Messina, D. Rosaci, and G. M. L. Sarn, "Forming Time-Stable Homogeneous Groups into Online Social Networks," *Information Sciences* (2017).
9. P. D. Meo, E. Ferrara, D. Rosaci, and G. M. L. Sarn, "Trust and Compactness in Social Network Groups," *IEEE Transactions on Cybernetics* 45.2(2015):205.
10. L. Alsumait and C. Domeniconi, "On-Line LDA: Adaptive Topic Models for Mining Text Streams with Applications to Topic Detection and Tracking," in *The IEEE International Confe. on Data Mining series*, 3-12(2008)
11. M. Hoffman, D.M. Blei and F. Bach, "Online learning for latent dirichlet allocation," in *Advances in Neural Information Processing Systems*, 23,856864(2010)
12. J. H Lau, Collier N, Baldwin T, "On-line Trend Analysis with Topic Models: #twitter Trends Detection Topic Model Online," in *COLING*, 1519-1534(2012)
13. C. Lin, and Y. He, "Joint sentiment/topic model for sentiment analysis," In *ACM confe. on Information and knowledge management*, 375384(2009)
14. Y Jo and AH Oh,"Aspect and sentiment unification model for online review analysis," in *ACM confe. on Web search and data mining*, 815824 (2011)
15. M. Dermouche, J. Velcin, L. Khouas, and S. Loudcher, "A Joint Model for Topic-Sentiment Evolution over Time," *IEEE International Conference on Data Mining IEEE*, 773-778(2014).
16. Y. He, C. Lin, W. Gao, and K. F. Wong, "Dynamic joint sentiment-topic model," *Acm Transactions on Intelligent Systems and Technology* 5.1(2014):1-21.
17. C. Lin, Y. He, R. Everson, and S. Rger, "Weakly supervised joint senti-ment-topic detection from text," *Knowledge and Data Engineering*, 24, 11341145(2002)
18. T.S. Ferguson, "A bayesian analysis of some non-parametric problems," *the Annals of Statistics*, 1,209-230(1973)
19. R. M. Neal, "Markov chain sampling methods for dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, 9,249-265(2000)
20. A. Ahmed, E. P. Xing, "Dynamic Non-Parametric Mixture Models and The Recurrent Chinese Restaurant Process: with Applications to Evolutionary Clustering," *Siam International Conference on Data Mining*, 219-230(2008).
21. X. Z. Zoubin, X. Zhu, Z. Ghahramani and J. Lafferty, "Time-sensitive dirichlet process mixture models," *Technical report*(2005)
22. L. Ren, D. B. Dunson and L. Carin, "The dynamic hierarchical dirichlet process," In *International Confe. on Machine Learning*, 824831(2008)
23. T. Xu, Z. M. Zhang, P. S. Yu and B. Long, "Dirichlet process based evolution-ary clustering," In *The IEEE International Confe. on Data Mining series*, 648657(2008)

24. T. Xu, Z. M. Zhang, P. S. Yu and B. Long, "Evolutionary clustering by hierarchical dirichlet process with hidden markov state," In The IEEE International Confe. on Data Mining series, 658667(2008)
25. C. Andrieu and N. D. Freitas, "An Introduction to MCMC for Machine Learning," Journal of Machine Learning, 50,5-43(2003)
26. T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proceedings of the National Academy of Science, 5228-5235(2004)
27. K. Ganesan and C. Zhai, "Opinion-based entity ranking," Information retrieval, 15,116-150(2012)
28. B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up:sentiment classification using machine learning techniques," Proceedings of the ACL confe. on Empirical methods in natural language processing, 79-86(2002)
29. T. Wilson, J. Wiebe and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in Proceedings of the Confe. on Human Language Technology and Empirical Methods in Natural Language Processing, 347-354(2005)
30. F. Maes, A. Collignon and D. Vandermeulen, "Multimodality Image Registration by Maximization of Mutual Information," IEEE Trans. on Medical Imaging, 17, 187-198(1997)
31. C. Wang and D. M. Blei, "Decoupling Sparsity and Smoothness in the Discrete Hierarchical Dirichlet Process," In Advances in neural information processing systems, 982-1989(2009)
32. W. T. Freeman, A. S. Willsky, E. B. Sudderth, "Graphical models for visual object recognition and tracking," PhD thesis, Massachusetts Institute of Technology(2006)