

On-Line Labeled Topic Model

YongHeng Chen^{*}, Yaojin Lin and Hao Yue

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

Abstract: A large number of electronic documents are labeled using human-interpretable annotations. High-efficiency text mining on such data set requires generative model that can flexibly comprehend the significance of observed labels while simultaneously uncovering topics within unlabeled documents. This paper presents a novel and generalized on-line labeled topic model (OLT) tracking the time development of extracted topics through a structured multi-labeled data set. Our topic model has an incrementally updated principle based on time slices in an on-line fashion, and can detect dynamic trending for labeled topics in parallel. Empirical results are presented to demonstrate lower perplexity and high performance of our proposed model when compared with other models.

Keywords: Bayesian models, Gibbs sampling, topic modeling, variational expectation-maximization.

1. INTRODUCTION

As e-commerce gets ever-increasingly fashionable, people need a valid method to obtain information from a large number of electronic documents. In general, the features of electronic documents are employed to classify the corpus.

Generative models have been usually employed to find the subgroups (topics) included in the document. Hofmann proposed a Probabilistic Latent Semantic Indexing (PLSI) model. PLSI is a generative model, which adds a new topic parameter between document and words [1]. Latent Dirichlet Allocation (LDA) was proposed by Blei *et al.*, which can construct the probability distributions between document and topic, topic and words, *etc.* [2].

While LDA is sufficient to model multi-topics per document, this model is not suitable for labeled corpora since, as an unsupervised model, it gives unobvious pattern of integrating a set of labels into its generating process. For integrating the set of labels, several models based on LDA have been proposed [3].

If employing LDA model to get this purpose, this model is commonly used to optimize the capability of a collection of potential topics, such as proposed in [4, 5], instead of modeling the supervised labels set into model's learning analysis. Latent Semantic Indexing [6] and related methods [7] are also popular unsupervised approaches. While unsupervised label topics are suited to acquire wider patterns in corpus, the obtained topics don't usually in accordance with human provided labels.

On the contrary, the supervised LDA model emphasizes the prediction issue through deducing the most predictive potential topics of document paired with a response [8]. However, these models are single labeled document

supervision and learning algorithm and cannot be applied to multi-labeled corpus.

Recently, a multi-Labeled generative model referred as Labeled Latent Dirichlet Allocation (L-LDA) has been proposed by Daniel Ramag *et al.* in 2009, which bounds LDA model by demonstrating a one-to-one communication between this model's topics and labels [9]. L-LDA considers that every document is tagged using a collection of provided labels, and these labels are important to construct the words based on pre-label probability distributions [9]. This can enable L-LDA to directly learn word-tag correspondences. However, the presence of any potential topics is not assumed by L-LDA. Besides, L-LDA ignores the difference between the topics of computer recognition and artificial labels, which leads to model's insufficient fitness with document's data and poor generalization ability.

Moreover, big data analysis with the above-mentioned topic models can be computationally difficult [10]. A primary study challenge for topic modeling is to efficiently serve models to big corpora [11, 12]. These motivated researchers looked for an optimization model, and ultimately several On-line topic models have been proposed [13]. Alsumait *et al.* (2008) proposed On-line Labeled Topic Model (OLT) that copes with documents in an on-line way through resampling topic distributions for documents from the new stream updating parameters. OLT uses collapsed Gibbs sampling for approximate inference [14]. Hoffman *et al.* (2010) adopted an On-line LDA variational Bayes as the similar posterior inference for LDA [15].

However, in order to identify novel topics and analyze their evolution, conventional OLT model needs to measure the word distribution between each topic t before and after an update using Kullback Leibler or Jensen-Shannon divergence [14]. With the increase of identified topics, the performance of topics will be greatly affected. In addition,

^{*}Address correspondence to this author at the China College of Computer Science, Minnan Normal University, Zhangzhou, Fujian 363000, China; Tel: +86 15260656959; Fax: +86 0596 2526960; E-mail: cyh771@163.com

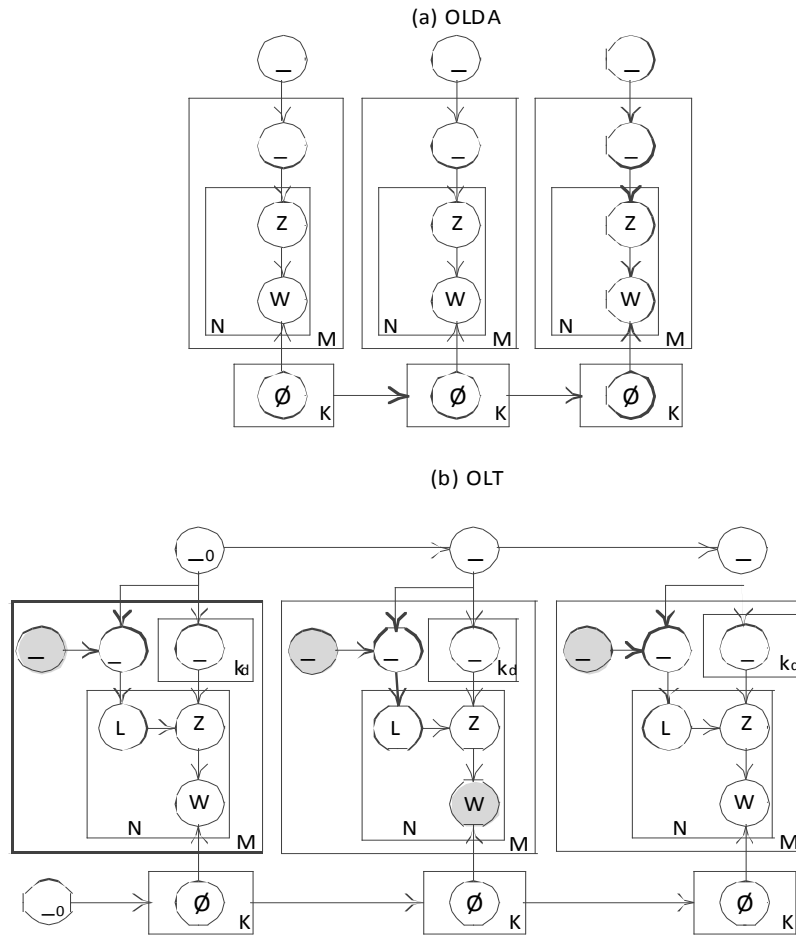


Fig. (1). OLT is shown with one inspiration models: (a) OLDA model (b) OLT model.

OLDA model only considers static vocabulary across time [16, 17].

In this paper, on-line partially OLT is presented, which is based on optimized L-LDA model and modified OLDA model. OLT model takes advantage of the unsupervised learning principle of topic model, but with the restriction that matches discovered topics with defined labels. OLT model considers that each document may employ topics that are classified in the set of connected document’s labels, and finds the development of the extracted topics only through comparing topics associated with the same label in an on-line manner. In order to further improve the performance, OLT model assigns the number of distinct types of labels to threads and adopts multithread technology to implement parallel computing. OLT model also has the ability to predict a label for unlabeled documents.

2. METHODOLOGY

2.1. Models with Topics

Before presenting the OLT model, let us review the basic LDA model. A glossary of notations used in the paper is summarized in Table 1, and the graphical model representations of our OLT model is shown in Fig. (1).

Table 1. SYMBOL.

SYMBOL	DESCRIPTION
K	number of topics
M	Total number of documents
N	Total number of unique words
N_d	number of word tokens in document d
$_d$	K -vector of dirichlet priors for document d
$_k$	V -vector of dirichlet priors for topic k
$_d$	the multinomial distribution of topics specific to the document d
\emptyset_k	the multinomial distribution of words specific to the topic k
$_d$	the multinomial distribution of labels specific to document d
$_$	a sparse binary vector of usable labels
L	a space of label
Z_{di}	the topic associated with the i th token in the document d
W_{di}	the i th token in document d
t_{di}	the timestamp associated with the i th token in the document d

LDA model’s regarding words as interchangeable is a simplification that is in compliance with the target of recognizing the semantic topics for each document. For a large number of sets of interest, but nonetheless, the supposition of

interchangeable documents is improper. For many data sets, such as news articles, research papers, the sequence of the documents implies a trending collection of topics. OLDA model in Fig. (1a), treats the temporal sequencing information and regards that the data sets are distributed by time slices, and documents as interchangeable within each time slice. OLDA model extends LDA model to operate in an on-line way. It models the newly arrived data sets within each slice by a k components topic model, where the generated model associated with slice t-1 is employed as the priors for LDA model at time slice t.

OLDA model regards hyper-parameters β as the prior observed enumerations on quantity of times words that are sampled from a topic [13]. Therefore, LDA model runs on data set at current time, and then constructs the count of words in topics that can be employed as the priors for the successive data stream.

2.2. On-line Labeled Topic Model

However, OLDA has the capacity to find topics included in documents corresponding to a sequential time slice. It cannot model corpus with labels, since, as an unsupervised model, it does not have the capabilities to provide a mean of incorporating meta-data into generating the proceeding. This simulated us to optimize OLDA to integrate on-line labeled document collection.

Through overcoming the disadvantages of L-LDA model and integrating it with optimized OLDA model, we proposed on-line labeled topic model (OLT) (see Fig. (1b)), which is a generative model for sequentially organized data set of labeled-documents. OLT model has the ability to update itself incrementally based on time slices in an on-line fashion. On the application side, it finds potential under-dimensional composition of document and study dynamic evolution for the classified topics by labels in parallel. So, in conclusion, we require an OLT model to: (1) update the model and transact the input when labeled-documents within an emerging stream appear; (2) control excessive growth over time; (3) measure the evolution of captured topics in parallel.

Table 2. Work flow of the static portion.

For each topic $k \in \{1, \dots, K\}$:
 -Generate $\phi_k = (\phi_{k,1}, \dots, \phi_{k,v})^T \sim \text{Dir}(\cdot | \cdot)$
 For each document d
 -Select labels of documents \wedge_d
 -Generate $\theta^{(d)} = (\theta_{d,1}, \dots, \theta_{d,L})^T \sim \text{Dir}(\cdot | \wedge_d)$
 -Generate topic distribution
 $\alpha^{(d)} = (\alpha_{d,1}, \dots, \alpha_{d,K})^T \sim \text{Dir}(\cdot | \cdot, \wedge_d)$, where $\wedge_{(0)} = 1$
 -For each word $i \in \{1, \dots, N_d\}$
 Generate $z_i \sim \text{Multinomial}(\alpha_{d,1}, \dots, \alpha_{d,K})$
 Generate $x_i \sim \text{Multinomial}(\theta^{(d)})$
 Generate $w_i \sim \text{Multinomial}(\phi_{k,z_i}, \dots, \phi_{k,v})$

OLT model is separated with time slice, a disconnection period, e.g. a day, or a year and documents are distributed in different time slices according to their time. The received sequence of documents within a time slice T are expressed as $S_T = \{d_1, \dots, d_N\}$, where N is the number of received documents within T. In the section below, there are two parts corresponding to generative process: firstly, the static portion that has been shown in Table 2 within OLT model, considers a single time slice only, and secondly, the dynamic portion

shown in Table 3 within OLT model, is the migration of model over time.

In our formalization of OLT model, a document d contains a multi-set of words w_d from a dynamic vocabulary V and a collection of labels \wedge_d from labels L (indexed by $1 \dots L$), each of the label is related with topics K_1 (indexed by $1 \dots K_1$) and where every topic $\phi_{1,k}$ is expressed by a multinomial distribution over w_d constructed by a symmetric Dirichlet prior β . Suppose, in the case of a general latent label L that can be used to calculate documents within the cluster, latent topics that have not been stuck with any label are optionally modeled. OLT model supposes that every topic can only participate in one label. In OLT model, we presume that a document d is constructed as follows: an explicit sub-collection of usable labels, expressed by a binary vector \wedge_d ; a document-explicit mix $\theta_{d,j}$ on topics $1 \dots K_j$ is constructed by a Dirichlet prior with hyper parameters α for every label j included in \wedge_d ; every word w within the document d is constructed by topic-word distribution of label, i.e. a label l from Ψ_d is selected firstly and then a topic z is chosen from $\theta_{d,l}$ that have correlation with that label l. Therefore, we can select word w through the $\phi_{1,k}$.

Table 3. Work flow of the dynamic portion.

set initial values α_0, θ_0 ; topic number K ; contribution factor ρ ;
 window size $|k|$;
 Iterative step for new time slice T_{new} :
 1 append T_{new} to sliding window;
 2 if number of time slices sliding window contains less than $|k|$
 3 update vocabulary in sliding window by adding different words within T_{new}
 4 else
 5 remove time slice that is first old in sliding window
 6 update vocabulary in sliding window by
 (a) removing words in documents within the removed time slice
 (b) adding new words within T_{new}
 7 Calculate priors α' and θ'
 8 Resample z using α' and θ' for documents in sliding windows

For the dynamic portion of OLT model, κ is represented as a sliding window that comprises a fixed number of time slices and preserves a dynamic vocabulary, where each word relates to the total number of documents contained in current sliding window [13]. In order to keep κ constant, the documents partitioned into first old time slice, are deleted when documents within a new time slice T appear. In this process, vocabulary will be updated by decreasing the total number related with words in documents contained within time slice T-1 by one. If the total number equals to zero, the corresponding word will be deleted. Further update of vocabulary is performed on words in documents within time slices T by the reverse operation. This serves two purposes. The first cause is that the consecutive model will grow indefinitely over time and become less sensitive to modified topic if documents partitioned into different time slices are all saved. So, we use sliding window to control excessive growth over time. The second cause is optimizing OLDA model. Because a fixed vocabulary is considered by OLDA, this assumption is not rational for a practical OLT model, where it is impractical to pre-compute the vocabulary in advance. Therefore, OLT model optimizes OLDA by re-generating vocabulary when adding documents in new time slice to sliding window.

When documents for time slice T+1 arrive, the current OLT model is incrementally updated by representing topic distributions z for all documents within time slices included in sliding window κ, employing ψ, θ and Ø inferred according to previous model in time slice T to construct hyper-parameters α' and β', i.e. the Dirichlet priors, for current OLT model. Let ρ be a contribution which determines its degree of contribution of previously known parameters by computing the priors of the successive new mode [13]. The Dirichlet hyper-parameters of a topic t in current time slice T+1 can be obtained as follows:

$$\alpha_t^{T+1} = \alpha_0 \times \frac{D_T}{N_T} \times n_{t..}^{(ki)}$$

$$\beta_t^{T+1} = \beta_0 \times (1 + \rho \times (\frac{V_{T+1}}{N_T} \times n_{t..}^{(vi)} - 1))$$

where α_t^{T+1} and β_t^{T+1} are natural parameters for topic t in time slice T+1; N_T and D_T are the number of documents included into sliding {T-|κ|,...,T} and the number of tokens contained in those documents, respectively; V_T is the number of vocabulary in sliding {T-|κ|,...,T}; n_{t..}^(ki) and n_{t..}^(vi) are the total number of topics t related with documents in slice T and the total number of words contained in those documents related with topic t.

In general, tracking the evolution of captured topics only needs taking into consideration topics associated with same label. For example, if we want to measure the trending of football topic within some time slice, we will only calculate the distance of the captured topics associated with label of sport within that time slice using KL or JS divergence. So, in order to further improve performance of tracking topics' evaluation, OLT model can assign the detected topics according to the associated label and adopt multithread technology to implement parallel computing.

3. APPROXIMATE VARIATION INFERENCE

In this section, assuming there is a sliding window κ that comprises a fixed number of time slices, {T-|κ|-1,...,T-1}, after a new time slice T+1 arrives, we distract our emphasis from parameters deduction to extraction within time slice T+1, i.e., θ, Ø and ψ.

Similar to the original LDA model, exact inference for our model is intractable. A variety of algorithms have been used to estimate the parameters of topic models, from basic expectation-maximization, to approximate inference expectation propagation, and Gibbs sampling. We used Gibbs Sampling here for posterior inference over parameters, since it is easy to extend and it has been proved to be quite effective in avoiding local optima.

Our main purpose is extracting the joint probability of the observed word based on unobserved label and topic allocations. This likelihood probability can be computed as follows: p(z,w,l|α, β,Λ)= p(w|z,β)P(z,l|α,Λ). Later, this joint probability can be extracted to derive efficient updates for parameters θ, Ø and ψ. First, the probability p(w|z,β) can be decomposed into ∫p(w,Ø|z,β)dØ. The detailed derivation process of it can be computed as follows:

$$\int p(w, \varphi | z, \beta) d\varphi$$

$$= \int p(w | z, \beta, \varphi) p(\varphi | \beta) d\varphi$$

$$= \int \prod_{k=1}^K \left(\prod_{v=1}^V \varphi_{k,v}^{n_{k,v}} \right) \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \left(\prod_{v=1}^V \varphi_{k,v}^{\beta_v-1} \right) d\varphi \tag{1}$$

$$= \int \prod_{k=1}^K \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \prod_{v=1}^V \varphi_{k,v}^{n_{k,v} + \beta_v - 1} d\varphi$$

$$= \prod_{k=1}^K \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \frac{\prod_{v=1}^V \Gamma(\beta_v + n_{k,v}^{(vi)})}{\Gamma(\sum_{v=1}^V \beta_v + n_{k,v}^{(vi)})} d\varphi$$

We expect further expansion of remaining joint probability P(z,l|α,Λ) which is shown as P(z,l|α,Λ)= P(l|α,Λ)P(z|l,α). Each half of this formula will be probed further in turn. Firstly, let's consider probability P(l|α,Λ), which, on account of Λ is supposed to be known, and could be decomposed into:

$$P(l | \alpha, \Lambda)$$

$$= P(l | \Lambda, \psi) P(\psi | \alpha, \Lambda)$$

$$= \int_{\Phi} \prod_{d=1}^M P(\psi_d | \alpha, \Lambda_d) \prod_{i=1}^{W_d} P(l_{d,i} | \Lambda_d, \psi_d) d\Phi \tag{2}$$

$$= \prod_{d=1}^M \prod_{j \in \Lambda_d} \frac{\Delta(n_{d,j}, \dots + \alpha)}{\Delta(\alpha)}$$

$$P(z | l, \alpha)$$

$$= \int_{\theta} P(z | l, \theta) P(\theta | \alpha) d\theta$$

$$= \int_{\theta} \prod_{d=1}^M \prod_{i=1}^{M_d} P(z_{d,i} | l_{d,i}, \theta_{d,i}) P(\theta | \alpha) d\theta \tag{3}$$

$$= \int_{\theta} \prod_{d=1}^M \prod_{i=1}^{M_d} P(z_{d,i} | l_{d,i}, \theta_{d,i}) P(\theta | \alpha) d\theta$$

$$= \int_{\theta} \prod_{d=1}^M \prod_{i=1}^{M_d} \theta_{d,i} P(\theta | \alpha) d\theta$$

$$= \int_{\theta} \prod_{d=1}^M \prod_{j \in \Lambda_d} \prod_{k=1}^{K_j} (\theta_{d,j,k})^{n_{d,j,k}} P(\theta | \alpha) d\theta$$

$$= \prod_{d=1}^M \prod_{j \in \Lambda_d} \frac{\Delta(n_{d,j}, \dots + \alpha)}{\Delta(\alpha)}$$

Here, we introduce the notation in which ∇(x) is defined as follows:

$$\Delta(x) = \frac{\prod_{k=1}^{\dim x} \Gamma(x_k)}{\Gamma(\sum_{k=1}^{\dim x} x_k)}$$

We can obtain equation (4) based on equations (1), (2) and (3), which is required by Gibbs sampling.

$$p(l_{d,i} = j, z_{d,i} = k | l_{-d,i}, z_{-d,i}, \mathbf{w}, i = v, \alpha, \beta)$$

$$\propto l[j \in \Lambda_d \wedge k \in 1..K_j] \left(\frac{\alpha + n_{d,j,k}^{(-d,i)}}{k_j \alpha + n_{d,j,\dots}^{(-d,i)}} \right) \tag{4}$$

$$\left(\frac{\beta + n_{\dots,j,k,t}^{(-d,i)}}{V\beta + n_{\dots,j,k,\dots}^{(-d,i)}} \right) \left(\frac{n_{d,j,\dots}^{(-d,i)} + \alpha_j}{n_{d,\dots,\dots}^{(-d,i)} + \sum_{j \in \Lambda_d} \alpha_j} \right)$$

The symbol $n^{(-d, i)}$ refers to the relevant quantity eliminating the current allocation of topic z and label l in document d at position i .

4. DISCOVERY AND TRACKING OF TOPIC

OLT model has the ability to detect topics and track their trending over time slice. After new time slice is added into window sliding, the word distribution in detected topics influences. To discover these novel topics and track the evolution of captured topics, we require computing the degree of change, and similarity of topics before and after an update using the Jensen-Shannon (JS) divergence. Two topic distributions, P and Q , are calculated using JS divergence as follows:

$$D_{JS} = \frac{1}{2}D_{KL}(P \| R) + \frac{1}{2}D_{KL}(Q \| R) \quad R = \frac{1}{2}(P + Q)$$

where D_{KL} is Kullback Leibler divergence.

However, only using the detection of the similarity does not satisfy the requirements for measuring the evolution of the relationship between topics. In labeled data sets, topics are closely related to the label. So, in this paper, OLT model only considers the relationship evolution between topics that have the same label. This means that we need to explicitly know the latent response label of topics.

We can label an extracted topic by label statistics of words connecting it. Considering that each word has disparate degree of importance associated with a label, tfidf is employed. The potential label for a topic is computed as follows:

$$WL_{v,l} = \frac{\sum_{i \in L} \sum_{v \in d} n_v \times tfidf_v}{\sum_v \phi_v \times WL_v}$$

The label for unlabeled document can be computed through label of topics included in it. Finally, during detecting the evolution of topics between adjacent time slice, OLT model assigns the number of distinct types of labels to threads and employs multithread technology to implement parallel computing.

Topic evolution process is demonstrated as follows:

1. Calculating the label of topics $\{\theta_1^{t+1}, \theta_2^{t+1} \dots \theta_K^{t+1}\}$ detected at $t+1$ time slice;
2. Allocating these topics to N threads according to the different labels, N is the number of different types of labels;
3. N threads parallel implement
4. pool.SubmitJob

For each detected topic at $t+1$ time slice, $k=1$ to K

- (a) Compute JS distance between θ_k^{t+1} and topics having the same label with θ_k^{t+1} ;
- (b) If JS divergence is less than a specific threshold, two topics have the evolution relationship;
- (c) Classify a topic as being novel if the JS divergence exceeds a threshold;

5. pool.Sync ()

5. EXPERIMENTS

In this section, Enron corpus and Xinhua News Agency will be employed to measure our proposed model. Enron corpus contains 1,227,250 emails with 151 authors. Xinhua News data set contains news attached with time stamp. These data sets were preprocessed for deleting extremely common words, and non-English alphabets.

5.1. Perplexity

The primary target of text modeling is density measurement that represents the latent data organization. Perplexity is an ordinary method to estimate this. The higher the likelihood is, the better the performance will be accomplished. The perplexity is defined as follows:

$$\text{perplexity}(M^{\text{enc}}) = \exp \left\{ - \frac{\sum_{d=1}^M \sum_{v=1}^V n_m^{(t)} \log(\sum_{k=1}^K \beta_k \cdot v_{\theta_m k})}{\sum_{d=1}^M N_d} \right\}$$

OLT model is the extended model which is based on improved L-LDA and modified OLT model. When on-line style is not been considered and the number of topic associated with label is one, OLT model is translated to L-LDA model. If no account of label and dynamic is taken, OLT model will become the optimized OLDA model. In the first group of experiments, we compared On-line Topic Model (OLDA), L-LDA, LDA, with our presented on-line label topic model (OLT). In our experiments, the time slice and sliding windows are set to 1 day and 3 days respectively. The hyper-parameter α_0 is set to 0.001, a small initial value of α is employed to construct a sparse topic distribution over documents. 0.01 is adopted for β_0 . The average perplexity trained on Xinhua News and Enron corpus are presented in Fig. (2).

OLDA and OLT models all explore new ways of incorporating metadata of the time stamp information into their models. However, OLT model optimizes OLDA model through employing the set of labels attached with document to offer a better prior for corpus. One can see as shown in these two graphs that making use of the label information significantly improves the predictive log-likelihood.

In addition, the documents clustering of the standard LDA model can be seen as the clustering operation based on topics settled by the user. Since LDA model does not make use of the label or time information for labeled corpus ordered by time, which restricts generalization capability of the LDA model. Therefore, it is evidently couldn't exceed OLDA and OLT models in efficiency.

So, in most cases, the effective use of documents' labels or time information can optimize the performance of document clustering. During the process of document clustering, L-LDA and OLT models can both make use of the label to promote the generation of document's topics. So, the effect of document clustering is obviously better than LDA model. Furthermore, OLT model considers that each document may employ topics that are classified in the set of document's

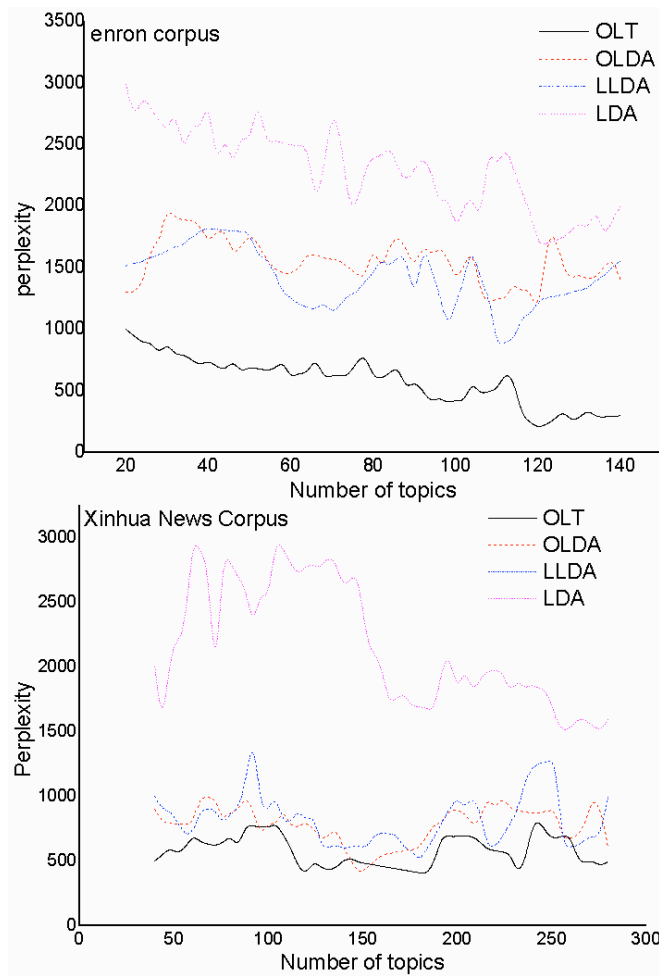


Fig. (2). Comparisons of perplexity.

labels and employs time information to improve the effect of document clustering with respect to L-LDA model. As shown in Fig. (2), the ability of document clustering of OLT model always outperforms L-LDA model.

5.2. Algorithmic Efficiency

The comparison of the average iteration time is shown in Fig. (3). Because there is a linear relationship between the complexity of the model and the number of topics, the average iteration time increases with the number of topics.

OLT and OLDA are able to detect topics without making use of the overall data. These models are constructed by settling sliding windows fraction of documents. However, L-LDA needs the entire memory information for more treatment; the average iteration time of L-LDA is remarkably higher than those of OLT and OLDA models.

Furthermore, OLT model adopts the dynamic maintenance of sliding windows before and after an update to control excessive growth, and only samples topics of labels of document. Most importantly, OLT model employs multi-threading technology to generate the latent label of topics. After several enhancements of OLDA model, as is shown, the running time of OLT model has been the shortest.

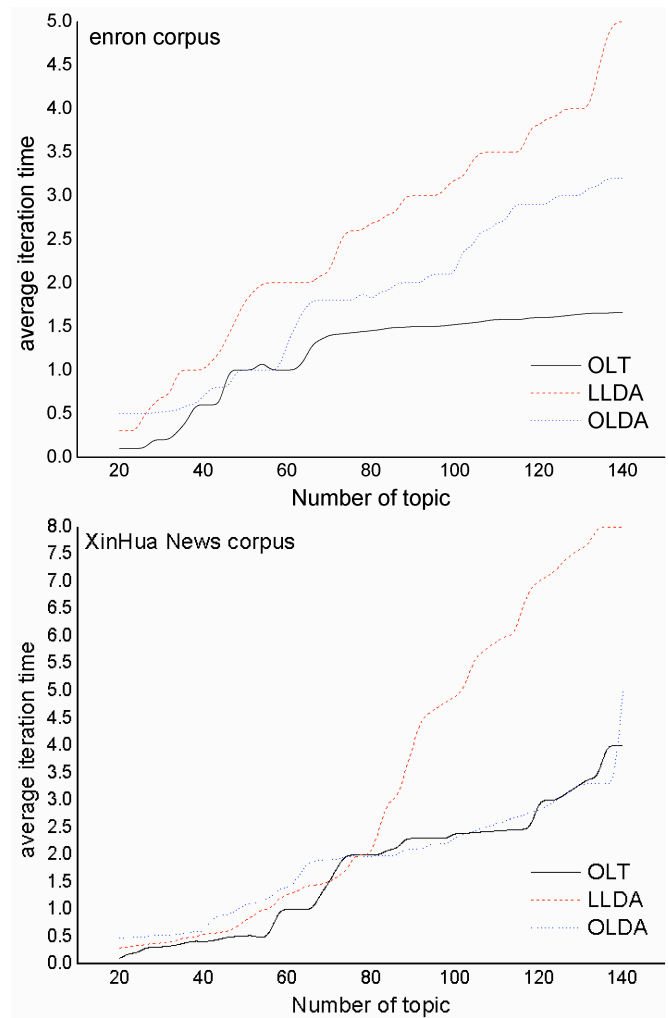


Fig. (3). Comparisons of running time.

CONCLUSION

An OLT model has been proposed in this paper, which provides simple probabilistic model to extract four levels of relationship, *i.e.*, documents, labels, topics, time slices and words. Firstly, this model prefers advanced constraints on potential topics which is in accordance with the human-assigned labels. And then, because this model could fit into the Bayesian framework, we further extended this model to incorporate time or sequence information based on optimized OLDA model. Empirical results prove that OLT model offers an obviously higher performance with respect to perplexity, operating efficiency and predictive ability of a label.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work is supported by science and technology project of Fujian provincial education department: No. JA13196; and the National Natural Science Foundation of China 6130313, 61402216.

REFERENCES

- [1] T. Hofmann, "Probabilistic latent semantic indexing", In: *Proceedings of the 22nd International Conference on Research and Development in Information Retrieval (SI-GIR'99)*, New York, NY, USA, 1999, pp. 50-57.
- [2] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation", *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003.
- [3] D. M. Blei, "Supervised Topic Models", In: *NIPS, Vancouver, B.C., Canada, 2007*, pp. 21.
- [4] H. Daume, "Markov random topic fields", In: *Proceedings of the ACL-IJCNLP, Suntec, Singapore, 2009*, pp. 293-296.
- [5] T. Iwata, T. Yamada, and N. Ueda, "Modeling social annotation data with content relevance using a topic model", In: *Advances in Neural Information Processing Systems*, Curran, Red Hook, NY, USA, 2009, pp. 835-843.
- [6] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis", *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391-407, 1999.
- [7] D. M. Blei, "Supervised Topic Models", In: *NIPS, Vancouver, B.C., Canada, 2007*, pp. 21.
- [8] Y. H. Chen, Y. Lin, W. Zuo, and K. Chen, "Time label topic model", *International Journal of Database Theory & Application*, vol. 8, no. 1, pp. 213-226, 2015.
- [9] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, "Labeled LDA: a supervised topic model for credit attribution in multi-label corpora", In: *EMNLP, Stroudsburg, PA, USA, 2009*, pp. 248-256.
- [10] Y. Zhou, S. Ji, and K. Xu, "Massive scientific paper mining: modeling, design and implementation", *Lecture Notes in Computer Science*, vol. 82, pp. 343-352, 2013.
- [11] D. Newman, A. Asuncion, P. Smyth, and M. Welling, "Distributed inference for latent Dirichlet allocation", In: *Neural Information Processing Systems, Vancouver, B.C., Canada, 2007*.
- [12] X. Wang, "Structured topic models: jointly modeling words and their accompanying modalities", PhD thesis, University of Massachusetts Amherst, 2009.
- [13] F. Yan, N. Xu, and Y. Qi, "Parallel inference for latent Dirichlet allocation on graphics processing units", In: *Advances in Neural Information Processing Systems*, Curran, Red Hook, NY, USA, 2009, pp. 2134-2142.
- [14] L. AlSumait, D. Barbará, and C. Domeniconi, "On-line LDA: adaptive topic models for mining text streams with applications to topic detection and tracking", In: *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM-08)*, 2008, pp. 3-12.
- [15] M. Hoffman, D. M. Blei, and F. Bach, "On-line learning for latent dirichlet allocation", In: *Advances in Neural Information Processing Systems 23 (NIPS-10)*, Vancouver, Canada, 2010, pp. 856-864.
- [16] X. Wang, "Structured topic models: jointly modeling words and their accompanying modalities", PhD thesis, University of Massachusetts Amherst, 2009.
- [17] X. Wang, and A. McCallum, "Topics over time: a non-Markov continuous time model of topical trends", In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, PA, USA, 2006, pp. 424-433.

Received: March 16, 2015

Revised: July 23, 2015

Accepted: July 23, 2015

© Chen et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.