A Fast and Effective Framework for Lifelong Topic Model with Self-Learning Knowledge

Kang Xu *, Feng Liu, Tianxing Wu, Sheng Bi, and Guilin Qi

School of Computer Science and Engineering, Southeast University, Nanjing, China
{kxu,fengliu,wutianxing,bisheng,gqi}@seu.edu.cn

Abstract. To discover semantically coherent topics from topic models, knowledge-based topic models have been proposed to incorporate prior knowledge into topic models. Moreover, some researchers propose lifelong topic models (LTM) to mine prior knowledge from topics generated from multi-domain corpus without human intervene. LTM incorporates the learned knowledge from multi-domain corpus into topic models by introducing the Generalized Polya Urn (GPU) model into Gibbs sampling. However, GPU model is nonexchangeable so that topic inference for LTM is computationally expensive. Meanwhile, variational inference is an alternative approach to Gibbs sampling and tend to be faster than Gibbs sampling. Moreover, variational inference can also be flexible for inferring topic models with knowledge, i.e., regularized topic model. In this paper, we propose a fast and effective framework for lifelong topic model, called Regularized Lifelong Topic Model with Self-learning Knowledge (RLTM-SK), with lexical knowledge automatically learnt from the previous topic extraction, then design a variational inference method to estimate the posterior distributions of hidden variables for RLTM-SK. We compare our method with 5 state-of-the-art baselines on a dataset of product reviews from 50 domains. Results show that the performance of our method is comparable to LTM and other knowledge-based topic models. Moreover, our model is consistently faster than the best baseline method, LTM.

Keywords: Variational Inference, Lifelong Topic Model, Knowledge-based Topic Model

1 Introduction

Topic models, such as pLSA (probabilistic Latent Semantic Analysis) [13] and LDA (Latent Dirichlet Allocation) [4], are popular content analysis techniques. Topic models are purely data-driven where topics are generated based on high-order word co-occurrence. However, some researchers find that the produced topics may not conform to human judgements [16]. One key problem is that the objective functions of topic models (e.g., LDA), which are fully based on

* Corresponding Author

implicit word co-occurrence patterns [18], often do not correlate well with human judgements [5].

To alleviate the above problem, some researchers try to leverage knowledge-based models which incorporate external knowledge to guide topic modeling. The existing knowledge-based topic models mainly incorporate lexical knowledge. For example, Dirichlet Forest-Latent Dirichlet Allocation (DF-LDA) [1] utilizes domain-specific lexical knowledge in the form of must-links to discover topics in accordance with domain knowledge. Thereinto, Must-link (u,v) represents two words u,v have a similar probability within any topics. A Must-link can be a word-pair with semantic association (e.g., "large" and "big"), or a word-pair occur in the same phrases (e.g., "battery" and "life") [1]. Chen et al. [8–11] expand domain-specific lexical knowledge to multi-domain knowledge. In their work, words in the same semantic-set share the same word sense. Each semantic-set is viewed as a must-link. However, all the knowledge based on lexicon are static and general, a majority of knowledge is not flexible for topic modeling in a specific domain, while domain-related knowledge is not covered in the lexicon. Chen et. al [6,7,17] further propose a method to mine Must-link knowledge automatically and incorporate these knowledge into topic models. LTM incorporates the learned knowledge from multi-domain corpus into topic models by introducing the Generalized Polya Urn (GPU) model into Gibbs sampling. There exist some disadvantages of Gibbs sampling approaches for LTM : 1) words in the same documents and topics influence each other during topic inference, hence it is difficult to conduct inference in parallel for large-scale corpus; 2) Gibbs sampling is computationally expensive. Based on Gibbs sampling method, only GPU model can be utilized for incorporating knowledge in LTM. However, GPU model is nonexchangeable so that topic inference for LTM is more expensive than Gibbs sampling. An optimal alternative to Gibb sampling is variational inference [3]. Variational methods use optimization to find a distribution over the latent variables that is close to the expected posterior of interest [20]. For variational methods, each iteration of variational inference is difficult and it requires the computation of complicated functions [4]; however, it only needs dozens of iterations to converge. Moreover, variational methods are flexible for incorporating Must-link knowledge in a fast and effective way, i.e, by appending regularized terms into lower bound function of topic models.

In this paper, we propose a new framework, Regularized Lifelong Topic Model with Self-learning Knowledge (RLTM-SK), which can be divided into two main steps: knowledge mining with topic models and knowledge-constrained topic modeling. Our process for mining high-quality topics from multi-domain documents involves first mining frequent co-occurred word-pairs (knowledge) based on topics mined from multi-domain corpus in advance, and then incorporating these knowledge into topic models for the next round of topic extraction. For lifelong topic model with self-learning knowledge, we will repeat the two steps alternately until convergence.

There exist three main contributions of RLTM-SK: 1) We propose a fast and effective framework for RLTM-SK, which contains the method for automatically mining Must-link knowledge and the method of incorporating Must-links into topic modeling, via appending a regularized term into lower bound function of LDA model. 2) We design a gradient descent approach for parameter estimation of RLTM-SK. 3) We implement experiments on a product review dataset from 50 domains to evaluate the effectiveness and efficiency of topic extraction in RLTM-SK.

2 Related Work

Topic Models, e.g. pLSA [13] and LDA [4], modeled semantic relations among words in an unsupervised way. The primitive topic models do not introduce any prior knowledge or other external resources, and topic models produce topics with uncontrolled quality. Nowadays, some researchers tried to leverage the domain knowledge of words to promote topic modeling. All the knowledge based topic models incorporated lexical (word-level) knowledge into topic models. The DF-LDA topic model [1] used tree-based priors to encode domain-specific expert knowledge on topic models in the form of must-links and cannot-links. Thereinto, a must-link indicates that two words must be assigned to the same topic, while a cannot-link states that two words should not be in the same topic. In [19], a factor graph framework was proposed to incorporate prior knowledge into topic models, where prior knowledge is modeled as sparse constraints (must-links and cannot-links) to speed up model training. In [2,15], they all incorporated domain knowledge into topic modeling in the form of first-order logic.

Recently, Chen et al. [6–11] proposed a series of research works that incorporated prior lexical knowledge from multi-domains into topic models. Thereinto, the first related work is MDK-LDA (LDA with Multi-Domain Knowledge) [11], a framework that exploited prior knowledge (must-links) from the past domains in topic models to bias topic assignment in the new domains. MC-LDA (LDA with M-set and C-set) [10] is the extension of MDK-LDA which used must-links and cannot-link prior knowledge. In GK-LDA (General Knowledge based LDA) [9], the model not only incorporated the general prior knowledge, but also handled incorrect knowledge without user input. Further they proposed AKL (Automated Knowledge LDA) [8] and LTM (Lifelong Topic Model) [7] that learned prior knowledge automatically from multiple domains to produce more coherent topics. All the knowledge-based topic models are based on Gibbs sampling, which is computationally intensive and cannot be scaled to large dataset [3]. Based on Gibbs sampling, LTM used GPU model for incorporating lexical knowledge. However, GPU model is nonexchangeable, the inference for LTM can be more computationally expensive than Gibbs sampling due to the non-exchangeability of words [6].

3 Lifelong Topic Model with Self-learning Knowledge

To extract topical words that satisfy our desired requirements, we propose a framework that can be divided into two main steps: knowledge-mining with topic models and knowledge-constrained topic modeling. Our process for transforming documents into high-quality topics involves first mining frequent co-occurred word-pairs (knowledge), and then incorporating these knowledge into topic models for the next round of topic mining. For lifelong topic model with self-learning knowledge, we will repeat the two steps alternately until convergence. In this section, we firstly make a brief review of LDA; then describe the process of knowledge mining and utilization; finally introduce the regularized topic model with self-learning knowledge.

3.1 Brief review of LDA

LDA assumes that a document is a mixture of topics where a topic is a multinomial distribution over words in the vocabulary. The generative process is as follows:

1. For topic index $k \in \{1, ..., K\}$
   i. Choose a word distribution $\beta_k \sim \mathrm{Dir}(\eta)$
2. For document $d \in \{1, ..., D\}$
   i. Choose a topic distribution $\theta_d \sim \mathrm{Dir}(\alpha)$
   ii. For $n \in \{1, ..., N_d\}$ word
      a. Choose a topic assignment $z_{d,n} \sim \mathrm{Multi}(\theta_d)$
      b. Choose a word $w_{d,n} \sim \mathrm{Multi}(\beta_{z_{d,n}})$

In this process, $Mult()$ is a multinomial distribution, and $Dir()$ is a Dirichlet distribution which is a prior distribution of $Mult()$, α and η are hyperparameters. The total probability of LDA is as Eq. 1.

$$p(\theta, \beta, z, w|\alpha, \eta) = \prod_{d=1}^{D} P(\theta_d; \alpha) \prod_{n=1}^{N} P(z_{d,n}|\theta_d) P(w_{d,n}|z_{d,n}, \beta) \prod_{k=1}^{K} P(\beta_k; \eta) \quad (1)$$

The mean-field variational distribution q for LDA breaks the relevance between words and documents, the detailed q is shown in Eq. 2. Based on Eq. 2, we can get lower bound on the likelihood $\mathbf{L}$ as Eq. 3, the object is to maximize $\mathbf{L}$ with respect to λ, γ and ϕ, where λ, γ and ϕ are utilized for estimating the objective posteriors.

$$q(\theta, \beta, z) = \prod_{d=1}^{D} q(\theta_d|\gamma_d) q(z_{d,n}|\phi_{d,n}) \prod_{k=1}^{K} q(\beta_k|\lambda_k) \quad (2)$$

$$\begin{aligned} \mathbf{L} &= E_{q(\theta,\beta,z)} \log \left[p(\theta|\alpha) \cdot p(\beta|\eta) \cdot p(z|\theta) \cdot p(w_n|z, \beta)\right] \\ &- E_{q(\theta,\beta,z)} \log \left[q(\theta|\gamma) \cdot q(\beta|\lambda) \cdot q(z|\phi)\right] \end{aligned} \quad (3)$$

3.2 Knowledge Mining and Utilization

The key object of LTM-SK is to extract Must-link knowledge from topics mined in the previous topic modeling. These knowledge contains word-pairs with semantic association (e.g., “large” and “big”), or word-pairs occur in the same phrases (e.g., “battery” and “life”). By incorporating these self-learning knowledge into topic modeling, we can mine more coherent topics.

Multi-Domain Knowledge Mining Given a set of documents $D = [D_1, ..., D_n]$ from n domains, LTM-SK mine knowledge from these documents with 3 main steps :

(1) Topic models are utilized to produce a set of topics S. In the initial phrase of topic modeling, LDA (Variational Inference) is utilized to mine topics, in the latter pharse of topic modeling, when knowledge has been learnt, we use Topic Model with Self-learning Knowledge to generate topics. (2) Topics are mined from multi-domain corpus, so many topics are not semantically related. If we mine knowledge from all the topics, much noisy knowledge will be introduced. To reduce noise in knowledge mining, we only mine knowledge in semantically similar topics from multi-domain documents, hence K-means clustering algorithm [12] is used to cluster topics. Thereinto, each topic in S corresponds to a word distribution, in our work, we choose the word distribution as the feature of the topic for clustering. (3) Only topics in the same clusters are utilized for mining knowledge together. In this step, we aim to mine Must-links from topics S, i.e., word-pairs co-occur multiple times in topics from multi-domain corpus. Here, each topic is represented as a list of words ranked with top-T probabilities in the topic-word distribution. To generate Knowledge Base (KB), i.e., a set of Must-links, from topics in different clusters, we use frequent itemset mining (FIM) [14] to mine Must-links. FIM is stated as follows: Given a set of transactions (topics) X, where each transaction (topic) $x_i \in X$ is a set of items (words). The goal of FIM is to discover every itemset (a set of items) that stratifies user-specified frequency threshold (i.e., minimum support), which the minimum times of an itemset must occur in X. Such itemsets are $frequentitemsets$, i.e., KB we need to learn. These $frequentitemsets$ are frequently co-occurred words in our work. To guarantee the quality of knowledge, we only use $frequentitemsets$ with 2-length, i.e., Must-links in our context. Must-links can be word-pairs with semantic association (e.g., "large" and "big"), or word-pairs occur in the same phrases (e.g., "battery" and "life")

Specific-Domain Knowledge Utilization Must-links are mined from multi-domain topics, however, these Must-links are only applicable in a specific domain. For example, Must-link "battery" and "life" is useful for topic modeling in the domain of *Phone*, but it is inapplicable in the domain of *Book*, even it can be adverse for topic modeling in the domain of *Book*. Hence, to measure the correlation of Must-links in the current domain, we use Pointwise Mutual Information (PMI) to estimate the correctness of the Must-link towards the current domain, i.e., it measure the extent of two words , in a Must-link, co-occur in the current domain. The PMI of words w_1 and w_2 is $PMI(w_1, w_2) = log\frac{P(w_1,w_2)}{P(w_1)P(w_2)}$, where $P(w)$ denotes the probability of seeing word w in a document, and $P(w_1, w_2)$ denotes the probability of seeing both words co-occurring in a document. These probabilities are empirically estimated from the current domain D^t, where $\#D^t(w)$ is the number of documents in D^t that contains words w and $D^t(w_1, w_2)$ is the number of documents that contain both words w_1 and w_2. $\#D^t$ is the total number of documents in D^t.($P(w) = \frac{\#D^t(w)}{\#D^t}, P(w_1, w_2) = \frac{\#D^t(w_1,w_2)}{\#D^t}$) A high PMI value implies a true semantic correlation of words in the current domain.

3.3 Regularized Lifelong Topic Model with Self-learning Knowledge (RLTM-SK)

This model is an extension of LDA, the generative process of RLTM-SK is the same as LDA. Hence, as is shown in Eq. 3, the objective function of LDA is lower bound of the total probability of LDA model. Two words u,v in a Must-link have a similar probability within any topics, i.e., $\lambda_{k,u} \approx \lambda_{k,v}$ for each topic $k = 1, ..., T$. It means two words in the same Must-link have a high similarity over their topic distribution, where the similarity is measured by vector inner product between λ_u and λ_v. λ_u and λ_v are topic distributions of words u and v. The more likely two words exist in a Must-link, the more similar topics of two words are . As the aforementioned discussion, words in a Must-link will influence the topic assignment of the two words and then influence the topic distribution of a document.

$$
\begin{aligned}
&\mathbf{L_{RLTM-SK}} = \mathbf{L} + \mathbf{L_{KRT}} \\
&= \quad \sum_{d=1}^{D}(\log\Gamma(\sum_{k=1}^{T}\alpha) - \sum_{k=1}^{T}\log\Gamma(\alpha) + \sum_{k=1}^{T}(\alpha-1)[\Psi(\gamma_{dk}) - \Psi(\sum_{j=1}^{T}\gamma_{dj})]) \\
&+ \quad \sum_{k=1}^{T}(\log\Gamma(\sum_{v=1}^{V}\eta) - \sum_{v=1}^{V}\log\Gamma(\eta) + \sum_{v=1}^{V}(\eta-1)[\Psi(\lambda_{kv}) - \Psi(\sum_{j=1}^{V}\lambda_k j)]) \\
&+ \quad \sum_{d=1}^{D}(\sum_{n=1}^{N}\sum_{k=1}^{T}\phi_{dnk}\cdot[\Psi(\gamma_{dk}) - \Psi(\sum_{j=1}^{T}\gamma_{dj})]) \\
&+ \quad \sum_{d=1}^{D}(\sum_{n=1}^{N}\sum_{k=1}^{T}w_{d,n}\cdot\phi_{dnk}\cdot[\Psi(\lambda_{k,w_{dn}}) - \Psi(\sum_{j=1}^{V}\lambda_{kj})]) \\
&- \quad \sum_{d=1}^{D}((\log\Gamma(\sum_{j=1}^{T}\gamma_{dj}) - \sum_{k=1}^{T}\log\Gamma(\gamma_{dk}) + \sum_{k=1}^{T}(\gamma_{dk}-1)[\Psi(\gamma_{dk}) \\
&- \Psi(\sum_{j=1}^{T}\gamma_{dj})])) - \quad \sum_{k=1}^{T}((\log\Gamma(\sum_{v=1}^{V}\lambda_{kv}) - \sum_{v=1}^{V}\log\Gamma(\lambda_{kv}) + \sum_{v=1}^{V}(\lambda_{kv}-1) \\
&[\Psi(\lambda_{kv}) - \Psi(\sum_{j=1}^{V}\lambda_{kj})])) - \quad \sum_{d=1}^{D}(\sum_{n=1}^{N}\sum_{k=1}^{T}\phi_{dnk}\log(\phi_{dnk}))\{\mathbf{L}\} \\
&+ \quad \sum_{(u,v)\in KB} PMI(u,v) * log\sum_{k=1}^{K}(\lambda_{k,u} * \lambda_{k,v})\{\mathbf{L_{KRT}}\}
\end{aligned}
\tag{4}
$$

To guarantee the words in a Must-link share similar topics, the similarity of two words in a Must-link, $log\sum_{k=1}^{K}(\lambda_{k,u} * \lambda_{k,v})$, is introduced into the objective function $\mathbf{L}$ as regularized item $\mathbf{L_{KRT}}$. Since Must-links are learnt from multi-domain corpus, Must-links are not flexible in all the domains. In our work, we use $PMI(u,v)$ to measure the flexibility of a Must-link in the current domain. Hence, $PMI(u,v)$ is set as a weight of $log\sum_{k=1}^{K}(\lambda_{k,u}*\lambda_{k,v})$, $PMI(u,v)*log\sum_{k=1}^{K}(\lambda_{k,u}*\lambda_{k,v})$. The bigger the value of $PMI(u,v)$ is, the more similar of two words u and v in a Must-link are. The objective function of regularized lifelong topic model is shown in Eq. 4.

Inference for RLTM-SK

$$\gamma_{d,k} = \alpha + \sum_{n=1}^{N}\phi_{dnk} \tag{5}$$

$$\phi_{d,n,k} \propto exp([\Psi(\gamma_{dk}) + \Psi(\lambda_{k,w_{d,n}}) - \Psi(\sum_{j=1}^{V}\lambda_{kj})]) \tag{6}$$

$$\lambda_{k,v} = \eta + \sum_{d=1}^{D}\sum_{n=1}^{N} w_{dn}\cdot\phi_{dnk} \tag{7}$$

$$\begin{aligned}\frac{\mathrm{d}L}{\mathrm{d}\lambda_{k,v}} &= [\Psi'(\lambda_{k,v}) - \Psi'(\sum_{j=1}^{V}\lambda_{k,j})][\eta + \sum_{d=1}^{D}\sum_{n=1}^{N} w_{d,n}*\phi_{d,n,k} - \lambda_{k,v}] \\ &+ \sum_{(v,u)\in KB} PMI(v,u)*\frac{1}{\sum_{k=1}^{K}\lambda_{k,v}*\lambda_{k,u}}*\sum_{k=1}^{K}\lambda_{k,u}\end{aligned} \tag{8}$$

Through the adjusting of variational parameters based on LDA, the distributions can maximize Eq. 4. The inference of γ and ϕ are shown as Eq 5 and Eq. 6. The computation of $\lambda_{k,w}$ is divided into two conditions: if $w \notin KB$, the equation is shown as Eq. 7; otherwise, a gradient-based optimization method is adopted to get the optimized $\lambda_{k,w}$ as Eq 8. The whole procedure, named variational inference for RLTM-SK, is shown in Algorithm 1.

4 Experiment Results

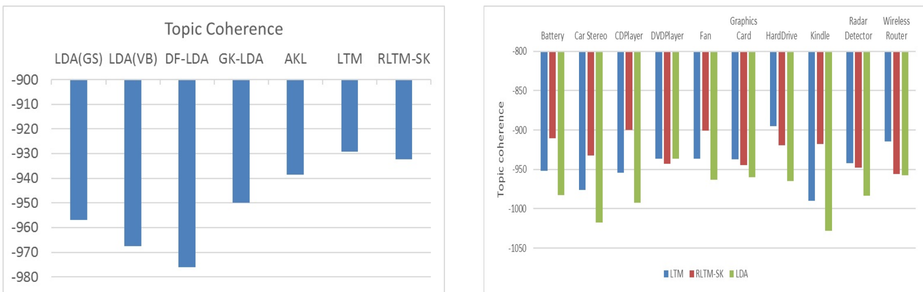
This paper evaluated the proposed RLTM-SK model and compares it with six state-of-the-art baselines: **LDA-GS** (Latent Dirichlet Allocation -Gibbs Sampling), **LDA-VB** (Latent Dirichlet Allocation -Variational Inference) [4], **DF-LDA** (Dirichlet Forest LDA) [1], **GK-LDA** (General Knowledge LDA) [9], **AKL** [8] (Automated Knowledge LDA), **LTM** [7](Lifelong Topic Model).

4.1 Datasets and methods for comparison

We used a large dataset containing 50 review collections from 50 product domains from Amazon.com as LTM [7], where each domain has 1000 reviews. The preprocess of the dataset was the same as LTM.

Algorithm 1 Mean-field variational inference for RLTM-SK

Require: $\alpha, \eta, K, iterNum$
Ensure: λ, γ, ϕ
  if $i < iterNum$ **then**
    for each topic k and vocabulary v **do**
      if $w_i \notin$ KB **then**
        Update $\lambda_{k,v}$ Using Eq 7
      end if
      if $w_i \in$ KB **then**
        Update $\lambda_{k,v}$ by gradient-based optimization with derivate in Eq 8
      end if
    end for
    for each document d **do**
      Update $\gamma_{d,k}$ Using Eq 5
      for each document d **do**
        Update $\phi_{d,n,k}$ Using Eq 6
      end for
    end for
  end if



(a) Average coherence score in 50 domains

(b) Coherence score in 10 domains

Fig. 1. (a) Average coherence score on the top 10 words in the 15 topics discovered on 50-domains product reviews. (b) Coherence score on the top 10 words in the 15 topics in the selected 10 domains.

Parameter Settings As the setting of LTM [7], we also set $\alpha = 1, \beta = 0.1$ and $K = 15$. For parameters of other baselines were set as their paper suggested. In all the baseline methods, Gibbs sampling was run for 2,000 iterations with 200 burn-in periods. For our variational inference method, the number of iterations was set as 200 (The variational inference method can converge faster than the Gibbs sampling method [20]). As LTM, the top 15 words of each topic were used to represent the topic for frequent itemset mining. For the number of clusters, $|S|$ is 10. The minimum support threshold was empirically set to max (5,0.4*$\#Trans$) where $\#Trans$ is the size of Transactions in each cluster (as LTM).

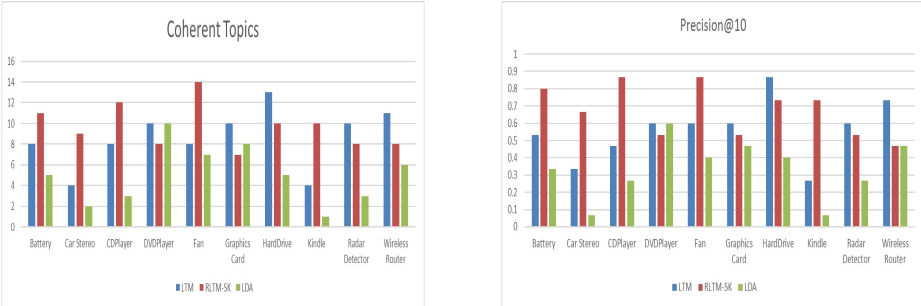
4.2 Topic Coherence

Another goal of RLTM-SK is to extract coherent topics from document collection and evaluate the effectiveness of topics captured by our models. In order to conduct quantitative evaluation of topic coherence, we used an automated metric proposed in [16], $C(t;V^{(t)}) = \sum_{m=2}^{M}\sum_{l=1}^{m-1} log\frac{D(v_m^{(t)},v_l^{(t)})+1}{D(v_l^{(t)})}$, where topic coherence, denoted as $D(v)$, is the document frequency of word v, $D(v,v^{'})$ is the co-document frequency of word v and $v^{'}$ and $V^{(k)} = (v_1^{(k)},...,v_T^{(k)})$ is a list of the T most probable words in topic k. The key idea of the coherence score is that if a word pair is related to the same topic, they will co-occur frequently in the corpus. In order to quantify the overall coherence of the discovered topics, the average coherence score, $\frac{1}{K}\sum^{k} C(z_k;V^{(z_k)})$, was utilized. The topic coherence is bigger, the topic quality is better. Here we compared RLTM-SK with six knowledge-based topic models: LDA(GS), LDA(VB), DF-LDA, GK-LDA, AKL and LTM. The result is shown in Fig 1(a). From the topic coherent results, the overall topic coherence score is close to the best baseline, LTM, and better than other baseline methods. We randomly selected 10 domains to compare the topic coherence score with LTM, which is shown in Fig. 1(b). In domain Battery, Car Stereo, CD player, Fan and Kindle, our model performed better than LTM in topic coherence; in other domains, LTM performed better than our model. Because the applicability of knowledge varied in different domains, topic coherence didn't perform better than LTM in all domains consistently. It is clear that the performance of our simple framework can be comparable with computationally expensive LTM.

4.3 Human Evaluation

As our objective is to discover more coherent topics, so we chose to evaluate the topics manually which is based on human judgement. Without enough knowledge, the annotation will not be credible. Following [16], we asked two human judges, who are familiar with common knowledge and skilled in looking up the test tweet dataset, to annotate the discovered topics manually. To ensure the annotation reliable, we labeled the generated topics by all the baseline models and our proposed model at learning iteration 10.

Here we only compared with our model with the best baseline and LDA model in human evaluation. Following [16], we asked the judges to label each topic as *coherent* or *incoherent* (a topic as *coherent* when at least half of top 15 words were related to the same semantic-coherent concept; others were *incoherent*). Then we chose *coherent* topics which were judged before and asked judges to label each word of the top 15 words among these *coherent* topics. When a word was in accordance with the main semantic-coherent concept that represents the topic, the word was annotated as *correct* and others were *incorrect*. Fig. 2(a) shows that , in 5 of 10 domains, RLTM-SK can discover more *coherent* topics than LTM and LTM can perform better in 4 domains. Fig. 2(b) shows that ,in 5 of 10 domains, RLTM-SK performed better than LTM in Precision. In summary,



(a) Number of Coherent Topics

(b) Precision of Coherent Topics

Fig. 2. (a) Proportion of *coherent* topics generated by each model (b) Average Precision @10 (p@10) of words in *coherent* topics

our model, RLTM-SK, can perform better than LTM on human evaluation in the randomly selected 10 domains.

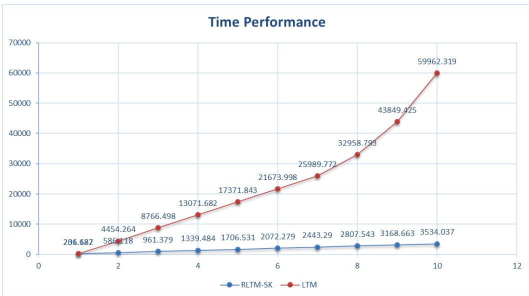


Fig. 3. The plot above demonstrates the speed of LTM and RLTM-SK.

4.4 Scalability

In the result of topic coherence and human evaluation, our model can achieve a competitive performance. Moreover, the significant advantages of our model are simplicity, efficiency and easily scalable. To understand the run-time complexity of our framework, which contains two main separate procedures, i.e., knowledge mining with topic models and knowledge-constrained topic modeling. The framework first involves topic modeling with self-learning knowledge by RLTM-SK model. However, there exists no prior knowledge initially, hence we use basic LDA to modeling topics. The second step is take the learnt topics as the input for knowledge mining, By jointly timing these two steps in our framework, we can empirically analyze the expected runtime of each iteration.

Fig. 3 shows the runtime of LTM, which is a previous lifelong topic model, and RLTM-SK over learning iterations. For our framework, with the growth of the learning iteration, the time cost did not increase significantly; On the contrary, LTM increases with a significant growth. When the iteration is 10, the run-time of LTM is far more than RLTM-SK. It shows that our framework can run lifelong topic model within a relatively short time, hence RLTM-SK can be much more easily scaled to large-scale corpus than LTM.

5 Conclusion and Future Work

In this paper, we propose a fast and effective framework for lifelong topic modeling. In our framework, we firstly mined knowledge automatically from topics extracted from multi-domain corpus and give different weights to knowledge so as to make knowledge adapting to a specific domain. Then, we used a variational inference method, which is fast and flexible for incorporating domain-adaptation knowledge, to infer parameters of topic models during lifelong topic modeling. Experimental results on 50-domains product reviews showed that our framework can consume much shorter time than LTM (the best baseline method), and it can perform as well as LTM in topic extraction. As future work, we plan to transfer this framework for jointly modeling sentiment and topic. Meanwhile, we can mine complex knowledge to improve topic modeling.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant No.61672153, the 863 Program under Grant No.2015AA015406 and the Fundamental Research Funds for the Central Universities and the Research Innovation Program for College Graduates of Jiangsu Province under Grant No. KYLX16_0295.

References

1. David Andrzejewski, Xiaojin Zhu, and Mark Craven. Incorporating domain knowledge into topic modeling via dirichlet forest priors. In *Proc. of ICML*, pages 25–32. ACM, 2009.
2. David Andrzejewski, Xiaojin Zhu, Mark Craven, and Benjamin Recht. A framework for incorporating general domain knowledge into latent dirichlet allocation using first-order logic. In *Proc. of IJCAI*, pages 1171–1192. AAAI, 2011.
3. David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *CoRR*, abs/1601.00670, 2016.
4. David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
5. Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L Boyd-graber, and David M Blei. Reading tea leaves: How humans interpret topic models. In *Proc. of NIPS*, pages 288–296. MIT Press, 2009.

6. Zhiyuan Chen. Lifelong machine learning for topic modeling and beyond. In *Proc. of NAACL-HLT*, pages 133–139, 2015.
7. Zhiyuan Chen and Bing Liu. Topic modeling using topics from many domains, lifelong learning and big data. In *Proc. of ICML*, pages 703–711. ACM, 2014.
8. Zhiyuan Chen, Arjun Mukherjee, and Bing Liu. Aspect extraction with automated prior knowledge learning. In *Proc. of ACL*, pages 347–358. ACL, 2014.
9. Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. Discovering coherent topics using general knowledge. In *Proc. of CIKM*, pages 209–218. ACM, 2013.
10. Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. Exploiting domain knowledge in aspect extraction. In *Proc. of EMNLP*, pages 1655–1667. ACL, 2013.
11. Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. Leveraging multi-domain prior knowledge in topic models. In *Proc. of IJCAI*, pages 2071–2077. AAAI, 2013.
12. JA Hartigan and MA Wong. Algorithm as 136: A k-means clustering algorithm. *Applied Statistics*, pages 100–108, 1979.
13. Thomas Hofmann. Probabilistic latent semantic indexing. In *Proc. of SIGIR*, pages 50–57. ACM, 1999.
14. Jacek Koronacki, Zbigniew W. Ras, Slawomir T. Wierzchon, and Janusz Kacprzyk, editors. *Advances in Machine Learning II, Dedicated to the Memory of Professor Ryszard S. Michalski*, volume 263 of *Studies in Computational Intelligence*. Springer, 2010.
15. Shike Mei, Jun Zhu, and Jerry Zhu. Robust regbayes: Selectively incorporating first-order logic domain knowledge into bayesian models. In *Proc. of ICML*, pages 253–261. ACM, 2014.
16. David Mimno, Hanna M Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. Optimizing semantic coherence in topic models. In *Proc. of EMNLP*, pages 262–272. ACL, 2011.
17. Shuai Wang, Zhiyuan Chen, and Bing Liu. Mining aspect-specific opinion using a holistic lifelong topic model. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*, pages 167–176, 2016.
18. Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. A biterm topic model for short texts. In *Proc. of WWW*, pages 1445–1456. Springer, 2013.
19. Yi Yang, Doug Downey, IL Evanston, Jordan Boyd-Graber, and Jordan Boyd Graber. Efficient methods for incorporating knowledge into topic models. In *Proc. of EMNLP*, pages 308–317. ACL, 2015.
20. Ke Zhai, Jordan L. Boyd-Graber, Nima Asadi, and Mohamad L. Alkhouja. Mr. LDA: a flexible large scale topic modeling package using variational inference in mapreduce. In *Proc. of WWW*, pages 879–888, 2012.