Topic Models Incorporating Statistical Word Senses

Guoyu Tang[1,2,3], Yunqing Xia[1,2,3], Jun Sun[4],
Min Zhang[5], and Thomas Fang Zheng[1,2,3]

[1] Center for Speech and Language Technologies, Division of Technical Innovation and Development, Tsinghua National Laboratory for Information Science and Technology
[2] Center for Speech and Language Technologies, Research Institute of Information Technology
[3] Department of Computer Science and Technology, Tsinghua University, Beijing, China
[4] Institute for Infocomm Research, A-STAR, Singapore
[5] Soochow University, China
sweetyuer@gmail.com, {yqxia,fzheng}@tsinghua.edu.cn,
sunj@i2r.a-star.edu.sg, mzhang@suda.edu.cn

Abstract. LDA considers a surface word to be identical across all documents and measures the contribution of a surface word to each topic. However, a surface word may present different signatures in different contexts, i.e. polysemous words can be used with different senses in different contexts. Intuitively, disambiguating word senses for topic models can enhance their discriminative capabilities. In this work, we propose a joint model to automatically induce document topics and word senses simultaneously. Instead of using some pre-defined word sense resources, we capture the word sense information via a latent variable and directly induce them in a fully unsupervised manner from the corpora. Experimental results show that the proposed joint model outperforms the classic LDA and a standalone sense-based LDA model significantly in document clustering.

Keywords: topic modeling, word sense induction, document representation, document clustering.

1 Introduction

Latent Dirichlet Allocation (LDA) was developed as a powerful unsupervised algorithm in analyzing topic distribution for a document collection [1].

The classic LDA model relies on the co-occurrences of surface words to capture their semantic relations. In reality, a surface word is likely to be highly associated to more than one topic and presents different word senses in different topics. LDA considers the surface word to be identical in both contexts and leverages on its co-occurrences with other words in the context to differentiate those two topics. Ideally, if a model is able to differentiate word senses in different contexts, the sense disambiguated words can contribute more probability masses to the

A. Gelbukh (Ed.): CICLing 2014, Part I, LNCS 8403, pp. 151–162, 2014.

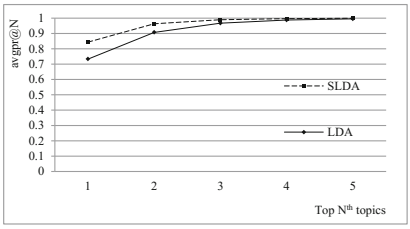


Fig. 1. Averaged per word (sense) topic distribution on the top-5 topics where the cumulative curve presents *avpgr* over the top-k topics

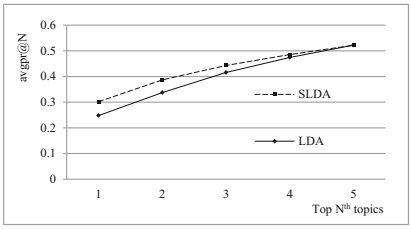


Fig. 2. Averaged per document topic distribution on the top-5 topics. where the cumulative curve presents the *avpgr* over the top-k topics.

corresponding topic than the surface words alone. In other words, word senses if applicable can serve as additional features for topic models and enhance the discriminative capability of topic models.

The intuition can be further verified with an empirical study on the average probability mass over the top N topics ($argpr@N$). We perform the surface word LDA model and the sense-based LDA model (SLDA) model (see details in Section 3.1) on Reuters20 data and compare the probability mass over the top N topics (N=5 in our analysis). We study two pairs of conditional probabilities: (1) The first pair is $p(z|w)$ and $p(z|s)$, which refer to the topic distribution given a surface word and a word sense respectively. The quantities are averaged over all word (sense) types in the data set. (2) The second pair is $p_w(z|d)$ in LDA and $p_s(z|d)$ in SLDA models, which refer to the topic distribution given a surface word based document and a word sense based document respectively. The quantities are averaged over all documents in the data set. The first pair is presented in Fig. 1 and the second is in Fig. 2, where the sense based model is referred as SLDA. All figures are drawn based on the experiments in Section 4.2. From Fig. 1, we can find that SLDA is above LDA in the cumulative curves. This suggests that word sense is a more indicative signature to describe the topic preferences for documents than surface word. From Fig. 2, we find that SLDA concentrates a document more on the top topics and provides sharper posterior topic estimation than LDA. This indicates that SLDA offers more confidence on the posterior estimation by means of the indicative word senses. Details of this analysis can be further found in Section 4.2.

In this paper, we will not only verify that word sense features provide topic models with more confidence in the posterior estimation, but also propose appropriate approaches to verify that the reinforced confidence is meaningful and helpful to improve the quality of the induced topics. The major contributions in this paper thus can be compiled in two perspectives. First, we incorporate the word sense information in the LDA generative story and construct a joint model to infer word senses for words and topics for documents simultaneously. Rather than applying the word sense information as an external feature or isolate the word sense induction as a pre-processing step [2] the proposed model

is more generic by incorporating the word sense feature as a latent variable in the graphical model. Second, our model is completely unsupervised and is able to work with external resources minimized. Previous researches[3,4] attempted to introduce word senses from WordNet to topic models. However, their models rely on the external knowledge source, i.e. WordNet, to construct a pool of word senses for a given word. Alternatively, we induce word senses automatically from the corpora. This is especially advantageous for resource poor languages that are short of available pre-defined word senses as well as domain specific documents that may contain terms beyond the general resources.

Specifically, we employ Hierarchical Dirichlet Processes (HDP) [5] as a non-parametric prior for word sense induction, because HDP can prevent us from explicitly bounding the number of word senses for each word. Two models are proposed in this paper: Standalone SLDA (SA-SLDA) considers word sense induction and document representation as standalone modules; Collaborative SLDA (CO-SLDA) takes the topics of senses from SLDA as the pseudo feedback for Word Sense Induction (WSI) and iteratively infers both topics and word senses.

The reminder of this paper is organized as follows: we first present some background for this work in Section 2. After that, we describe the approaches to incorporate statistical word senses for the LDA topic models. Experimental results and discussions are presented in Section 4. We conclude this paper in Section 5.

2 Related Work

2.1 Semantic Interpretation of Documents

In Vector Space Model (VSM) [6] , it is assumed that terms are independent of each other and the semantic relations between terms are ignored.Recently, models are proposed to represent documents in a semantic concept space using lexical ontologies, i.e. WordNet or Wikipedia [7,8]. However, the lexical ontologies are difficult to be constructed and their coverage can be limited. In contrast, topic models are used as an alternative for discovering latent semantic space in corpora based on the per topic word distribution. LDA [1] as a classic topic model identifies topics of documents by evaluating word co-occurrences. Some work attempt to integrate word semantics from lexical resources into topic models [3,4]. Alternatively, our models are fully unsupervised and do not rely on any external semantic resources, which will be extremely applicable for resource poor languages and domains.

2.2 Word Sense Disambiguation and Induction

Word sense disambiguation, which identifies the correct word sense from a set of pre-defined sense candidates, has been proved to benefit various NLP tasks [9]. However, manually-compiled large lexical resources such as WordNet are

often required.Instead, Word Sense Induction (WSI) can learn word senses from corpora in an unsuper-vised manner. With respect to the Bayesian approaches, Brody and Lapata [10] used an extended LDA model to induce word senses which provide the state-of-the-art performance in SemEval-2007 evaluation[11] . Yao and Durme [12] used Hierarchical Dirichlet Process (HDP) [5] to induce word senses and empirically verified its advantage over LDA. WSI is also applied in other tasks like information retrieval [2], where word senses for query words are induced. To the best of our knowledge, no work has been reported to exploit WSI in document topic modeling as we do in this paper.

3 Topic Models Incorporating Statistical Word Senses

As shown in Fig. 3, the classic LDA assigns each word in the document a topic and considers the surface words as the basic granularity for a document. Alternatively, our model emits a sense for each surface word and assigns each sense a topic. Therefore, the basic granularity for our model is the word sense (Fig. 3). To address this motivation, we introduce an additional latent variable of word sense to LDA and induce it from the observed surface words.

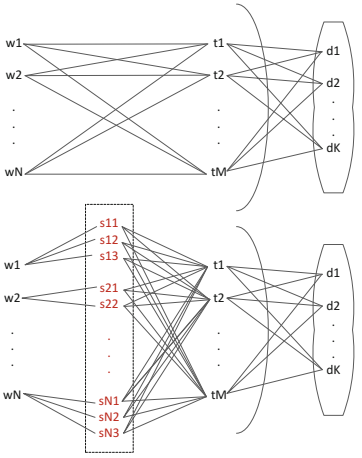


Fig. 3. Illustration of the classic LDA model (above) and the word sense extended LDA models (below). The values in the dot rectangle are assigned to the latent variable (i.e., word sense).

We design several models to implement this purpose as follows:

– Standalone SLDA (SA-SLDA): We isolate the Word Sense Induction (WSI) process as a standalone step. With the induced word senses in hand, we perform the word sense based LDA. .

– Collaborative SLDA (CO-SLDA): We identify the generative story as two iteratively interchangeable steps. Given an observed topic, we generate the word sense from the topic. Given an observed word sense, we generate the topic for each word sense, where the word sense is a point estimate from the mode of the distribution.

3.1 Standalone SLDA Model (SA-SLDA)

In the SA-SLDA model, WSI and documet representation (DR) are considered as standalone modules, where DR takes the output (i.e., word senses) of WSI as input(see Fig.5).

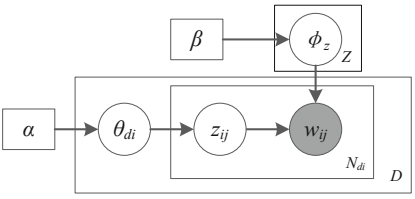


Fig. 4. Illustration of the standard LDA model

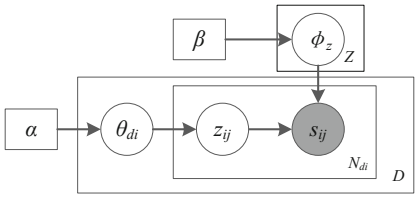


Fig. 5. Illustration of the SA-SLDA model

We follow [12] to employ Hierarchical Dirichlet Processes (HDP) for word sense induction.In this paper, we perform HDP on each word. We define a word on which the WSI algorithm is performed as a target word. We also define the words in the context of a target word as context words of the target word. After WSI, we simply take the mode sense in the distribution as the sense of the target word.

As shown in Fig. 5, we replace surface word with word sense in the gray plate. Given D documents and W word types, the formal procedure with Z topics of document representation in SA-SLDA is given as follows:

1. For each topic z:
   (a) choose $\phi_z \sim Dir(\beta)$.
2. For each document d_i:
   (a) choose $\theta_{d_i} \sim Dir(\alpha)$.
   (b) for each word w_{ij} in document d_i:
      i. choose topic $z_{ij} \sim Mult(\theta_{d_i})$.
      ii. choose sense $s_{ij} \sim Mult(\phi_{z_{ij}})$.

where d_i refers to i-th document in the corpus; w_{ij} refers to j-th word in document d_i; z_{ij} refers to the topic that word w_{ij} is assigned; s_{ij} refers to the sense that word w_{ij} is assigned from WSI; α, β are hyper-parameters of the model; $\phi_{z_{ij}}$ and θ_{d_i} are per topic sense distribution and per document topic distribution respectively which are drawn from Dirichlet distributions.

We use Collapse Gibbs Sampling to do inference for SA-SLDA [13]. Compared with LDA, we replace the surface words with the induced word senses. Therefore, the topic inference is similar to the classic LDA, where the condition probability $P(z_{ij} = z|\boldsymbol{z_{-ij}}, \boldsymbol{s})$ is evaluated by

$$P(z_{ij} = z|\boldsymbol{z_{-ij}}, \boldsymbol{s}) \propto \frac{n^{d_i}_{-ij,z} + \alpha}{n^{d_i}_{-ij} + Z\alpha} \times \frac{n^{s}_{-ij,z} + \beta}{n_{-ij,z} + S\beta} \tag{1}$$

In Eq.1,$n^{d_i}_{-ij,z}$ is the number of words that are assigned topic z in document d_i $n^{s}_{-ij,z}$ is the number of senses with sense s that are assigned topic z , $n^{d_i}_{-ij}$ is the total number of words in document d_i; $n_{-ij,z}$ is the total number of words assigned topic z; S is the number of senses for the data set.$-ij$ in all the above variables refers to excluding the count for word w_{ij}. Further details are similar to the classic LDA [13].

3.2 Collaborative SLDA (CO-SLDA)

Alternatively, we induce word senses and the document topics simultaneously in a joint model (see Fig.6). We are interested in whether the topic assigned to a word has a positive feedback on WSI, which then can be used to refine the topic distribution. Inspired by this motivation, we propose a Collaborative SLDA model which takes the topics of senses from SLDA as the pseudo feedback for WSI and iteratively infers both topics and word senses. Specifically, we achieve a point estimate for the target word in WSI and feed this estimated sense to DR.

In this model, a three-level HDP algorithm is used to capture the relationship between word senses and topics of a target word w (see Fig. 6). In the

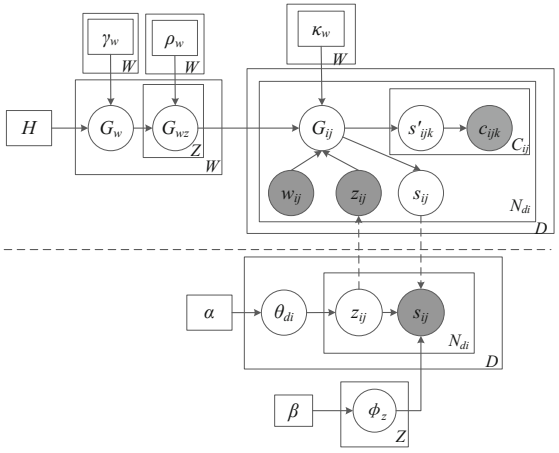


Fig. 6. Illustration of the CO-SLDA model

three-level HDP, for each word type w , we choose for each topic a probability measure G_{wz} which is drawn from Dirichlet Process $DP(\rho_w, G_w)$. For each word w_{ij} in document d_i , given topic $z_{ij} = z$, we use G_{wz} as the base probability measure for the context of w_{ij} and draw its own G_{ij} from Dirichlet process $G_{ij} \sim DP(\kappa_{wz}, G_{wz})$. This means that word w may have different sense distributions in different topics. For each context v_{ij} of the target word w, the sense s_{ijk} for each word c_{ijk} in v_{ij} has a nonparametric prior G_{ij}. H is a Dirichlet distribution with hyper-parameter ϵ. The context word distribution η_s given a sense s is generated from H:$\eta_s \sim H$. Hyper-parameters γ_w, ρ_w and κ_w are the concentration parameters for DP, controlling the variability of the distributions of G_w,G_{wz} and G_{ij} respectively.

We show the graphical presentation for CO-SLDA in Fig. 6. C_{ij} refers to the number of words in the context window v_{ij} for word w_{ij} in document d_i . The above dotted line shows the WSI process while the below shows the DR process. Given observed topics $\{z_{ij}\}$, word senses $\{s_{ij}\}$ are inferred in WSI. Given observed senses $\{s_{ij}\}$, topics $\{z_{ij}\}$ are inferred in DR. The two processes are interchangeably performed. We provide the dashed arrows in Fig. 6 to connect $\{s_{ij}\}$ and $\{z_{ij}\}$ that will change from hidden to observed during the alternation of two processes.

The word sense induction process is as follows:

1. For each word type w:
   (a) choose $G_w \sim DP(\gamma_w, H)$.
   (b) For each topic z:
      i. choose $G_{wz} \sim DP(\rho_w, G_w)$.
2. For each document d_i:
   (a) For each context v_{ij} of word w_{ij}:
      i. choose $G_{ij} \sim DP(\kappa_{wz}, G_{wz})$.
      ii. For each context word c_{ijk} of target word w_{ij}:
         A. choose $s_{ijk} \sim G_{ij}$.
         B. choose $c_{ijk} \sim Mult(\eta_{s_{ijk}})$.
      iii. set $s_{ij} = \arg\max_s P(s|G_{ij})$.

The document representation process is just like SA-SLDA.

For inference, we interchangeably infer two groups of hidden variables in CO-SLDA,

1. Given that the topic for each word sense z_{ij} is observed, we infer the sense distribution G_{ij} in the context window around a target word. This is achieved through the same scheme as [5]. Then we estimate s_{ij} for the target word as sense with the highest probability in G_{ij}.
2. Given that the word sense s_{ij} is observed, we infer the topic z_{ij} for each word sense. This can be achieved using the same inference scheme as SA-SLDA.

As the iteratively process can refine both topics and word senses based on each other's prediction, intuitively, the CO-SLDA model should be advantageous over the SA-SLDA model, which only provides single round estimation of the variables.

4 Evaluation

In the experiments, we first evaluate the latent topics in document clustering task and then analyze the averaged per sense topic distribution and averaged per document topic distribution of the proposed models.

4.1 Document Clustering

In this section, we apply the proposed models on the document clustering task and evaluate the performance against the baselines of LDA and K-means algorithms.

4.1.1 Setup

Data Set: Three data sets are used in our experiments.

1. **TDT4**:Following [14], we use the English documents from TDT2002 and TDT2003, i.e., TDT41 and TDT42. .
2. **Reuters**: Documents are extracted from Reuters-21578[15] with the most frequent 20 categories, i.e., Reuters20.

In the experiments, only nouns and verbs are used as target words for word sense induction and topic inference. We use per sentence as the context window for each target word. TreeTagger[16] is used to for Part-of-speech labeling.

Evaluation Metrics: In the experiments, we intend to evaluate the proposed topic models in document clustering task. Each topic in the test dataset is considered as a cluster and each document is clustered into the topic with the highest probability. We adopt the evaluation criteria proposed by [17]. The calculation starts from maximum F-measure of each cluster. The general F-measure of a system is the micro-average of all the F-measures of the system-generated clusters.

4.1.2 Experiment 1.1: Different Word Sense Induction Approaches

In this experiment, we aim to investigate how well the different word sense induction approaches contribute to the task of document clustering. We compare the performance of two different Bayesian models, i.e. LDA vs. HDP, in our SA-SLDA model.

System Parameters: As we isolate the WSI process from the document representation process in SA-SLDA, we present the parameters accordingly. (1) In the WSI step, for HDP, we set the hyper-parameters γ_w , ρ_w , ϵ for every word type to be $\gamma_w \sim Gamma(1, 0.001)$, $\rho_w \sim Gamma(0.01, 0.028)$,$\epsilon = 0.1$; for LDA, we set $\alpha = 0.2$,$\beta = 0.1$ and set the sense numbers for all words to be 4. (2) In the Document representation step, we set $\alpha = 1.5$ and $\beta = 0.1$. All hyper-parameters are tuned in the TDT42 dataset. The number of topics is set to be equal to the number of clusters in each dataset.

In all experiments, we let the Gibbs sampler burn in for 2000 iterations and subsequently take samples 20 iterations apart for another 200 iterations.

Table 1. Results of SA-SLDA with different WSI approaches (i.e., LDA and HDP)

Method	TDT41	TDT42	Reuters20
SA-SLDA(LDA)	0.787	0.842	0.490
SA-SLDA(HDP)	0.792	0.870	0.512

Experimental results are presented in Table 1.

Discussions: From Table 1, we can find that WSI with HDP outperforms WSI with LDA in all data sets when integrated into the SA-SLDA model. This is because LDA is a parametric model which requires user's explicit setting of the parameters. Alternatively, HDP, as a non-parametric model, can automatically infer the number of senses for each word. This provides reasonable interpretation for word sense modeling and additional flexibility for document representation. This advantage of HDP also provides our series of SLDA models with better interpretation. As a result, we employ HDP as a non-parametric prior for all proposed models.

4.1.3 Experiment 1.2: Different Extended LDA Models

In this experiment, we aim to verify the effectiveness of the proposed models in document clustering. Other than the proposed models, i.e., SA-SLDA and CO-SLDA , we also present K-means and LDA as our baselines. Specifically, we implement the Bisecting K-Means algorithm [17] which computes the cosine similarity between documents based on the TF-IDF features.

System Parameters In the WSI step we set the hyper-parameters γ_w, ρ_w , ϵ for every word type to be $\gamma_w \sim Gamma(8, 0.1)$, $\rho_w \sim Gamma(5, 1)$, $\kappa_w \sim Gamma(0.1, 0.028)$, $\epsilon = 0.1$; (2) in the DR step, we set $\alpha = 1.5$ and $\beta = 0.1$. In LDA, we set $\alpha = 1.5$, $\beta = 0.1$. The number of topics is set to be equal to the number of clusters in each dataset. In K-Means, we set K to be equal to the number of clusters in each dataset.

Experimental results are presented in Table 2.

Discussions: From Table 2, we can find that: First, SLDAs outperform the two baselines in all data sets. This indicates that using word senses other than surface words improves the document clustering results, which is due to the fact that SLDAs are facilitated with more fine-grained features of word sense induced from the context.

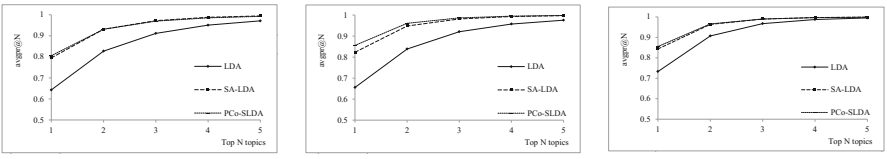
Table 2. Results of the proposed models and baselines

Method	TDT41	TDT42	Reuters20
K-Means	0.727	0.843	0.501
LDA	0.744	0.867	0.496
SA-SLDA	0.792	0.870	0.512
CO-SLDA	0.825	0.874	0.597

Second, CO-SLDA outperforms SA-SLDA in all data sets. This indicates that the joint inference process for topics of words and word senses provides positive impact to refine the results. Two reasons are worthy of noting: (1) In common sense, instances of the same word type in different topics may have different senses while instances in the same topic often refer to the same thing. Since CO-SLDA can jointly infer topics and word senses, instances of the same word in the same topic are more likely to be assigned the same sense while instances in different topics are likely to be assigned differently. As a result, word senses will be better identified. (2) Using topics as a pseudo feedback will facilitate the target words with topic-specific senses. For example, the word *election* only has one sense in general cases. However, in the TDT42 data set, topics are labeled in a more fine-grained perspective. For example, the following two sentences are labeled to be from two different topics as the countries of elections are different: z_1: *Ilyescu Wins Romanian Elections*, z_2: *Ghana Gets New Democratically Elected President*. With the joint inference of topic and sense, we can induce the word 'election' with two senses, i.e., *election#1* and *election#2*, related to the electing processes in Romania and Ghana respectively. By incorporating these topic-specific senses, *election* with context word *Romania* is identified as *election#1* and more likely to be assigned topic z_1 while *election* with context word *Ghana* is identified as *election#2* and more likely to be assigned z_2 .

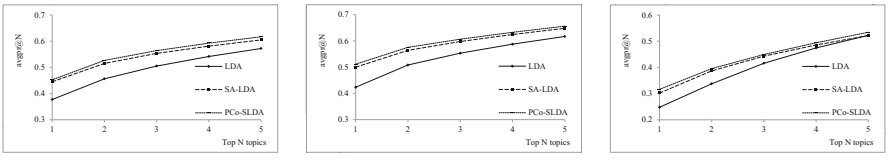
4.2 Distribution Analysis

In this study, we aim to analyze the averaged per sense topic distribution and averaged per document topic distribution of the proposed models.



(a) TDT41 (b) TDT42 (c) Reuters20

Fig. 7. Averaged per word (sense) topic distribution on the top-5 topics



(a) TDT41 (b) TDT42 (c) Reuters20

Fig. 8. Averaged per document topic distribution on the top-5 topics

In Fig. 7, we present the $argpr@N$ values of averaged per word topic distribution for all words in LDA and averaged per sense topic distribution for all senses in SLDAs on the three datasets. Specifically, for each word, the topics that a word w is associated are ranked according to the probability $p(z|w)$[1] . $argpr@N$ is calculated by averaging the probabilities $p(z|w)$ of all words on the top N topics. For each word sense, we calculate $argpr@N$ based on $p(z|s)$. We use the senses in the last iteration of SLDA models(e.g. 2200) and the topics inferred by these word senses. For each data set, we draw the cumulative curve over the top N topics. Furthermore, we measure $avgpr@N$ of document-topic distribution by averaging the probabilities $p(z|d)$ for all documents in LDA and SLDA models on three datasets are presented in Fig. 8.

Discussions: From Fig. 7, we can observed that: First, SLDAs are all above LDA in the cumulative curves. This indicates word senses provide better discriminative capabilities for topic models than surface words. Second, the cumulative curves of CO-SLDA are above SA-SLDA. This benefit comes from that fact that CO-SLDA induces topic-specific senses by using topics as a pseudo feedback. The topic-specific senses are more discriminative than common senses.

From Fig. 8, we can observe that: First, the cumulative curves of SLDAs are all above LDA. This indicates that in SLDAs, documents concentrate on fewer topics which makes topics from sense-based topic models more discriminative. Second, the cumulative curves of CO-SLDA are above SA-SLDA. This suggests that the iteratively refined topics and words senses provide reinforcement of the posterior estimation of topics for documents.

5 Conclusion

In this paper, we propose to represent topics with distributions over word senses. In order to achieve this purpose in a fully unsupervised manner without relying on any external resources, we model the word sense as a latent variable and induce it from corpora via WSI. We design several models for this purpose. Distributions analysis of average sense-topic distribution and the average document-topic distribution shows a sharper distribution of topics in SLDAs which suggests that the proposed models provide more confidence on the posterior estimation. Empirical results verify that the word senses induced from corpora can facilitate the LDA model in document clustering. Specifically, we find the joint inference model (CO-SLDA) outperforms the standalone model (SA-SLDA) as the estimation of sense and topic can be collaboratively improved.

Acknowledgments. This work is supported by NSFC China (No. 61272233). We thank the reviewers for the valuable comments.

[1] $p(z|w)$ can be calculated with $p(z|w) \propto p(w|z)\Sigma p(z|d)p(d)$ where $p(w|z)$ and $p(z|d)$ are parameters of the model thus can be estimated while we estimate $p(d)$ to be the proportion of d's document length to the length of the entire document collection.

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022 (2003)
2. Navigli, R., Crisafulli, G.: Inducing word senses to improve web search result clustering. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP 2010, Stroudsburg, PA, USA, pp. 116–126. Association for Computational Linguistics (2010)
3. Boyd-Graber, J.L., Blei, D.M., Zhu, X.: A topic model for word sense disambiguation. In: EMNLP-CoNLL, pp. 1024–1033 (2007)
4. Guo, W., Diab, M.: Semantic topic models: Combining word distributional statistics and dictionary definitions. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, pp. 552–561. Association for Computational Linguistics, Stroudsburg (2011)
5. Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M.: Hierarchical dirichlet processes. Journal of the American Statistical Association 101 (2004)
6. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Commun. ACM 18(11), 613–620 (1975)
7. Hotho, A., Staab, S., Stumme, G.: Wordnet improves text document clustering. In: Proc. of the SIGIR 2003 Semantic Web Workshop, pp. 541–544 (2003)
8. Gabrilovich, E., Markovitch, S.: Computing semantic relatedness using wikipedia-based explicit semantic analysis. In: Proceedings of the 20th International Joint Conference on Artifical Intelligence, IJCAI 2007, San Francisco, CA, USA, pp. 1606–1611. Morgan Kaufmann Publishers Inc. (2007)
9. Tufiş, D., Koeva, S.: Ontology-supported text classification based on cross-lingual word sense disambiguation. In: Masulli, F., Mitra, S., Pasi, G. (eds.) WILF 2007. LNCS (LNAI), vol. 4578, pp. 447–455. Springer, Heidelberg (2007)
10. Brody, S., Lapata, M.: Bayesian word sense induction. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2009, pp. 103–111. Association for Computational Linguistics, Stroudsburg (2009)
11. Agirre, E., Soroa, A.: Semeval-2007 task 02: Evaluating word sense induction and discrimination systems. In: Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval 2007, pp. 7–12. Association for Computational Linguistics, Stroudsburg (2007)
12. Yao, X., Van Durme, B.: Nonparametric bayesian word sense induction. In: Proceedings of TextGraphs-6: Graph-based Methods for Natural Language Processing, pp. 10–14. Association for Computational Linguistics (2011)
13. Griffiths, T.L., Steyvers, M.: Finding scientific topics. PNAS 101(suppl. 1), 5228–5235 (2004)
14. Kong, J., Graff, D.: Tdt4 multilingual broadcast news speech corpus. Linguistic Data Consortium (2005), http://www. ldc. upenn. edu/Catalog/CatalogEntry. jsp
15. Lewis, D.D.: Reuters-21578 text categorization test collection, distribution 1.0 (1997), `http://www.research.att.com/~{}lewis/reuters21578.html`
16. Schmid, H.: Probabilistic part-of-speech tagging using decision trees. In: Proceedings of International Conference on New Methods in Language Processing, Manchester, UK, vol. 12, pp. 44–49 (1994)
17. Steinbach, M., Karypis, G., Kumar, V.: A comparison of document clustering techniques. In: KDD Workshop on Text Mining (2000)