IJICTR International Journal of Information & Communication Technology Research

Volume 9 - Number 2 - Spring 2017(19-25)

A Probabilistic Topic Model based on an Arbitrary-Length Co-occurrence Window

Marziea Rahimi
School of Computer and IT Engineering
Shahrood University of Technology
Shahrood, Iran
marziea.rahimi@shahroodut.ac.ir

Morteza Zahedi
School of Computer and IT Engineering
Shahrood University of Technology
Shahrood, Iran
zahedi@shahroodut.ac.ir

Hoda Mashayekhi
School of Computer and IT Engineering
Shahrood University of Technology
Shahrood, Iran
hmashayekhi@shahroodut.ac.ir

Received: November 12, 2016 - Accepted: February 26, 2017

Abstract— **Probabilistic topic models have been very popular in automatic text analysis since their introduction. These models work based on word co-occurrence, but are not very flexible with respect to the context in which co-occurrence is considered. Many probabilistic topic models do not allow for taking local or spatial data into account. In this paper, we introduce a probabilistic topic model that benefits from an arbitrary-length co-occurrence window and encodes local word dependencies for extracting topics. We assume a multinomial distribution with Dirichlet prior over the window positions to let the words in every position have a chance to influence topic assignments. In the proposed model, topics being shown by word pairs have a more meaningful presentation. The model is applied on a dataset of 2000 documents. The proposed model produces interesting meaningful topics and reduces the problem of sparseness.**

Keywords- probabilistic topic modeling; co-occurrence; context window; Gibbs sampling; generative models.

I. INTRODUCTION

Nowadays we are faced with a vast amount of digitalized information. As the amount continues to grow, it becomes more and more difficult to find what we are looking for. It will be way more facile, if we could look for our needed information by exploring based on thematic data instead of raw data. Probabilistic topic modeling introduces methods which can extract thematic structure of documents. The basic idea of these methods is that a document is a mixture of latent topics and each topic is a distribution over words.

Suppose we have M documents $\{d_1, d_2, \ldots, d_m, \ldots d_M\}$ where each document d_m consists of N_m words $\{w_{m1}, w_{m2}, \ldots, w_{mn}, \ldots, w_{mN_m}\}$ and such that there are K topics and N unique words $v = \{v_1, v_2, \ldots, v_N\}$. The topic assigned to each word w_{mn} is denoted by z_{mn}. Based on this view we can approach the problem of extracting topics of a corpus as follows: each topic is a distribution over words where the words are exchangeable, i.e., each document is a bag of words. Documents are also exchangeable. Each word in each document is

extracted from the distribution of its assigned topic. For each document there is a distribution over topics which shows how the topics have been mixed to produce the document. Then there are two parameters in model; distribution of words in topics φ and distribution of topics in documents θ.

Set φ comprises K multinomial distributions φ^k over N words where φ_v^k is $p(w_{mn} = v \mid z_{mn} = k, \varphi)$ and the probability of topics θ is a set of M multinomial distributions θ^m over K topics where for each document d_m, $\theta_k^m = p(z_{mn} = k \mid \theta^m)$. We want to estimate φ such that gives the words of the documents high probabilities and it can be done by maximizing the corpus probability. Latent Dirichlet allocation (LDA) [1] is a generative probabilistic model for such estimation.

LDA has been very popular since its introduction and has been used in many application areas to help exploring large data and to reveal latent relationships in data. Luo and Zhang [2] have used LDA for image quality assessment. In their approach each distorted image is a document and distortion-aware features are modeled as words. Savoy [3] has used LDA for authorship attribution which is useful where it is needed to determine who has wrote a text when authorship of the text is in dispute. The main idea is that a topic model can capture the differences between writing styles. Razavi and Inkpen [4] use topic modeling to produce a multi-resolution view of the text. The basic idea of their research is that different number of topics reveals different aspects of texts.

The LDA model works based on word co-occurrence within a whole document but according to many pieces of research, a whole document is not always a suitable context for extracting co-occurrence statistics. With the context of a whole document, the model cannot consider any local information. Some effort has been made for incorporating such information into the LDA model. These models will be discussed in the next section. Many of these models use only the previous word for encoding local dependencies. We can consider it as if these models use a co-occurrence of length 2 which cannot provide enough evidence to derive a robust model in many applications. We will introduce a model that can use an arbitrary length co-occurrence window. We provide the model with a multinomial distribution over the positions of the co-occurrence window. Every word in the window has the chance to influence topics.

The proposed model assumes that each word in a document is determined by both its topic and a preceding word in the co-occurrence window. The preceding word is determined by the new multinomial distribution which is incorporated in the model. We show that this model reduces the number of zero occurrences compared to the base model [1] discussed in Section 3. We also evaluate the proposed model using a dataset of 2000 documents of Associated Press (AP), showing that it is a better model of the dataset in comparison to LDA and BTM and produces more meaningful topics.

The rest of the paper is organized as follows. In Section 2, we review the related works that consider local and spatial information for extracting topics. In section 3, LDA and Bigram Topic Model (BTM) are described in more details. In Section 4, we describe the proposed model and how collapsed Gibbs sampling is used to estimating model parameters. Section 5 contains the description about experiments and results.

II. RELATED WORKS

Probabilistic topic models have been improved in many directions. Some effort has been made to relax the basic model assumptions such as document exchangeability [5] and word exchangeability [6]. A supervised probabilistic topic model has been introduced in [7]. Several studies have been done for finding faster and less complex algorithms for extracting model parameters such as in [8, 9] and several for incorporating prior knowledge into the basic model such as in [10]. In this paper we are interested in studies that have tried to relax exchangeability in a document and incorporated local and spatial information into the model such as in [6, 11].

For this aim, some researchers have tried to incorporate word order into generative topic models. Wallach [6] has incorporated bigram language model into a generative probabilistic topic model, in which each word is dependent to its previous word in addition to its topic. This model is called Bigram Topic Model (BTM).

Barbieri et al. [11] have suggested a very similar model called Token-Bigram model along with two other models. One of them assumes the dependency between each word's topic and its previous word's topic, called Topic-Bigram. The other one assumes the dependency between each word and its previous word's topic called Token-Bitopic. These three models have been used in a recommendation system and all of them did a better job than the basic LDA.

Words can be divided into two categories: function words, which serve syntactic functions and content words, which provide semantics. Based on this idea, Griffiths et al. [12] have introduced a generative probabilistic topic model which can distinguish between function and content words without any prior knowledge of either syntax or semantic . This model allows a word to be generated from either a topic model or a hidden Markov model (HMM) reflecting syntactic classes.

Grifith, et al. [13] introduce a model called LDA-Collocation in which each word can be extracted from a topic or its previous word with which forms a collocation. The choice between these two options is handled by a Bernoulli distribution over the options. Wang, et al. [14] introduce a generalization over LDA-Collocation in which, the mentioned choice is dependent on the topic of the subjected word. This gives the model the ability to consider the context of the word when choosing if the word forms a

collocation with its previous word. Jameel, et al. [15] use similar settings in a supervised topic model. Yang, et al. [16] also use similar idea in combination with a topic hierarchy to capture the hierarchical nature of topics in a text.

.

All of these models incorporate local information into the basic LDA model. In this section we are going to look at those efforts from the perspective of word co-occurrence. As we mentioned before, LDA works based on word co-occurrence in the whole document and assumes no dependency among words or among assigned topics. On the other hand, the other discussed models assume that words or topics are dependent on only the previous word or topic. We consider the Bigram Topic Model (BTM) [6] shown in Fig. 2 as a base for such models. In this model, each word is assumed to be dependent on its previous word. We can consider this as a co-occurrence window of size 2. It is a well-known hypothesis in automatic text analysis that when we are trying to capture semantic relationships in a text by calculating word co-occurrences, a window length of a whole document is not a suitable context length. Semantic relationships have diverse relation to the distance between the considered words [17, 18], but a very short window is not suitable either especially when the ordering is maintained, because sparseness will grow unbearably and there will not be enough evidence for generating robust results. In this paper, we introduce a model that provides the desired flexibility to decide on the length of the co-occurrence window.

III. PRELIMINARIES

LDA can be considered as a basic model for probabilistic topic models which do not consider local word relationships. BTM can be considered as a representative for models that consider local word relationships assuming a direct dependency between each word and its previous word in each document. Thus the proposed model is compared to these two models and therefore both need to be discussed in more detail.

A. LDA

LDA assumes the following generative process for generating each document d_m in corpus D :

- Choose $\theta^m \sim Dir(\alpha)$
- For each word in position n in document d_m
  - Draw a topic z_{mn} from θ^m
  - Draw a word w_{mn} from the distribution over words for the topic z_{mn} , i.e., $p(w_{mn} \mid z_{mn})$

LDA can be represented by the graphical model shown in Fig. 1. As shown in the figure, each document d_m is a mixture of latent topics represented as θ^m and the mixture weights follow a Dirichlet distribution with the hyperparameter α , i.e., $\theta^m \sim Dire(\alpha)$. Each topic is a distribution over exchangeable words.

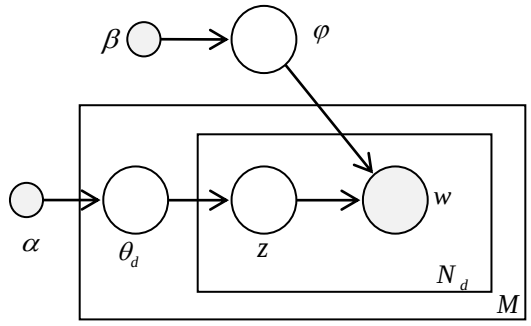

Fig. 1. Graphical model of LDA

B. BTM

As mentioned before in BTM each word is dependent to its previous word in addition to its topic. This means each word is sampled from a probability distribution conditioned on the chosen topic and also the previous word. A graphical representation of the model is shown in Fig. 2.

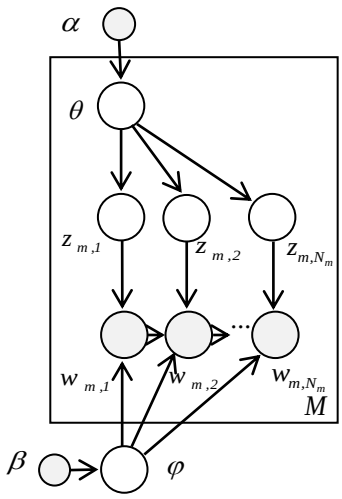

Fig. 2. Graphical model for BTM which incorporates bigram language model into LDA

As shown in the figure, words in each document are not exchangeable. The documents are still exchangeable. BTM assumes the following generative process for each document:

- Choose $\theta^m \sim Dir(\alpha)$
- For each word in position n in document d_m
  - Draw a topic z_{mn} from θ^m
  - Draw a word w_{mn} from the distribution over words for the context defined by the topic z_{mn}

and previous word w_{mn-1} i.e. $p(w_{mn} \mid w_{mn-1}, z_{mn})$

As one can see considering the generative process, each word is directly dependent on its preceding word.

IV. PROPOSED MODEL

In this section, we describe a model which has more flexibility in using the co-occurrence context. Recall that in the defined terminology of topic modeling, θ and φ denoted the distributions of topics and words in a topic respectively. Here, we consider an additional multinomial distribution with parameter π in the document level which lets the model to choose a word from a window of length L.

Assume a window of length L before each word w_{mn} of each document d_m. Words in every position of this window have the chance to influence topic assignment for w_{mn}. The value of a word w_{mn} is determined based on both its topic and the word residing in position t_{mn} of the preceding window. The value of t_{mn} is chosen based on π. In other words, now each topic is not a single multinomial distribution over words, but rather consists of a set of N distributions over words. The word in position t_{mn} of the preceding window decides on which distribution of topic z_{mn}, word w_{mn} will be chosen from. For more clarification, the window of word w_{mn} is shown in Fig. 3 below. The window corresponding to word w_{mn} is shown in this figure. If t_{mn} is 1 then w_{mn} is selected from the word distribution of topic z_{mn} which corresponds to the word in position $n-1$ i.e. w_{mn-1} and so on.

$w_{m(n-L)}$	$\cdots$	$w_{m(n-2)}$	$w_{m(n-1)}$	w_{mn}

Fig. 3. Window of word w_{mn}

The model assumes the following generative process where α, β, γ are hyper parameters and φ, θ and π are corresponding to topic distributions, topic proportions and position proportions respectively.

- Choose $\varphi \sim Dir(\beta)$
- For each document d_m
  - Choose $\theta^m \sim Dir(\alpha)$
  - Choose $\pi^m \sim Dir(\gamma)$
  - For each word in position n in document d_m
    - Draw $z_{mn} \sim multinomial(\theta^m)$
    - Draw $t_{mn} \sim multinomial(\pi^m)$
    - Draw $w_{mn} \sim multinomial(\varphi^{z_{mn}, w_{m(n-t_{mn})}})$ i.e. $p(w_{mn} \mid z_{mn}, w_{m(n-t_{mn})}, \varphi)$

According to the proposed model, a word can be dependent on any word in its preceding window and it is $multinomial(\pi^m)$ that chooses which window word it is dependent on. This can be considered as if the co-occurrence is calculating on a context of arbitrary length L and each window is overlapped with $n-1$ elements of the previous window. The proposed model specification is shown in Fig. 4 below.

$$
\begin{aligned}
w_{mn} \mid z_{mn}, w_{m(n-t_{mn})}, \varphi_{z_{mn}, w_{m(n-t_{mn})}} &\sim multinomial(\varphi_{z_{mn}, w_{m(n-t_{mn})}}) \\
\varphi &\sim Dirichlet(\beta) \\
z_{mn} \mid \theta^m &\sim multinomial(\theta^m) \\
\theta &\sim Dirichlet(\alpha) \\
t_{mn} \mid \pi^m &\sim multinomial(\pi^m) \\
\pi &\sim Dirichlet(\gamma)
\end{aligned}
$$

Fig. 4. Proposed model specification

A. Estimating the Model Parameters

The model parameters are estimated by Gibbs sampling. To apply Gibbs sampling we need $p(z_{xy}, t_{xy} \mid z_{-xy}, t_{-xy}, w)$ where z_{xy} is an instance of z in position y of document x and $-xy$ means all the other positions except the position xy and according to Bayes rule:

$$p\left(z_{xy},t_{xy} \mid z_{-xy},t_{-xy},w\right)=\frac{p\left(z_{xy},t_{xy},z_{-xy},t_{-xy},w\right)}{p\left(z_{-xy},t_{-xy},w\right)}. \quad (2)$$

We first calculate the numerator and then eliminate all things that are related to the denominator as follows where $f(t)$ maps each position to its corresponding word in each window.

$$\begin{aligned}&p\left(z_{xy},t_{xy},z_{-xy},t_{-xy},w\right)=p\left(z,t,w\right)=\\&\iiint p\left(z,t,w,\theta,\pi,\varphi\right)d\theta d\pi d\varphi=\\&\iiint p(\theta)\,p(z\mid\theta)\,p(\pi)\,p(t\mid\pi)\\&p(\varphi)\,p(w\mid z,f(t),\varphi)d\theta d\pi d\varphi=\\&\int p(\theta)p(z\mid\theta)d\theta\times\\&\int p(\pi)\,p(t\mid\pi)d\pi\times\int p(\varphi)\,p(w\mid z,f(t),\varphi)d\varphi.\end{aligned} \quad (3)$$

Based on the conjugacy and by cancellation we reach to the following result:

$$p\left(z_{xy},t_{xy} \mid z_{-xy},t_{-xy},w\right)\propto\frac{n^{d_x}_{-xy,z_{xy}}+\alpha}{n^{d_x}_{-xy,(.)}+K\alpha}\times\frac{n^{d_x}_{-xy,t_{xy}}+\gamma}{n^{d_x}_{-xy,(.)}+L\gamma}\times\frac{n^{z_{xy},w_{x(y-t_{xy})}}_{-xy,w_{xy}}+\beta}{n^{z_{xy},w_{x(y-t_{xy})}}_{-xy,(.)}+N\beta}. \quad (4)$$

Where $n^{d_x}_{-xy,z_{xy}}$ is the number of times a word has been assigned to topic z_{xy} in document d_m ignoring the current position (xy) in the document. Term $n^{d_x}_{-xy,t_{xy}}$ denotes the number of times a word in position t_{xy} of the window has been selected in document d_m. Term $n^{z_{xy},w_{x(y-t_{xy})}}_{-xy,w_{xy}}$ denotes the number of times word w_{xy} has been assigned to topic z_{xy} where word $w_{x(y-t_{xy})}$ has been appeared somewhere in its window throughout the dataset. In these formulae, $(.)$ refers to all values of the corresponding variables. After enough iterations, we can calculate the values of parameters

$$\theta^{d_m}_k=\frac{n^{d_m}_k+\alpha}{n^{d_m}_{(.)}+K\alpha},\qquad \pi^{d_m}_l=\frac{n^{d_m}_l+\gamma}{n^{d_m}_{(.)}+L\gamma},$$
$$\varphi^{k,v_t}_v=\frac{n^{k,v_t}_v+\beta}{n^{k,v_t}_{(.)}+N\beta}. \quad (5)$$

Where $n^{d_m}_k$ is the number of times a word has been assigned to topic k in document d_m. Term $n^{d_m}_l$ denotes the number of times a word in position l of the window has been selected in document d_m. Term n^{k,v_t}_v denotes the number of times word v has been assigned to topic k where word v_t has been appeared somewhere in its window throughout the dataset.

V. Experimental Results

We applied the proposed model on a dataset consisting of 2000 documents. The dataset was constructed by removing stopwords, numbers and signs from the Associated Press (AP) dataset provided by [1]. All the words that occurred only once in the corpus were also removed. 1248 documents are randomly selected or training and the reset are used for test. Model specifications are shown in Table 1.

As mentioned before, the size of a co-occurrence window has to be long enough to avoid sparseness and obtain accurate statistics. On the other hand, if it is too long it will lose the ability to capture local word relationships. Fig. 5, which shows the perplexity of our model as a function of window length, confirms this statement. We used the settings in Table 1 for the experiment.

Table 1. Experimental settings

Number of topics (K)	20
Window length (L)	10
α	$50/K$
β	0.01
γ	$1/(50L+K)$
Number of iterations	1000
Burn-in period	500
lag	100

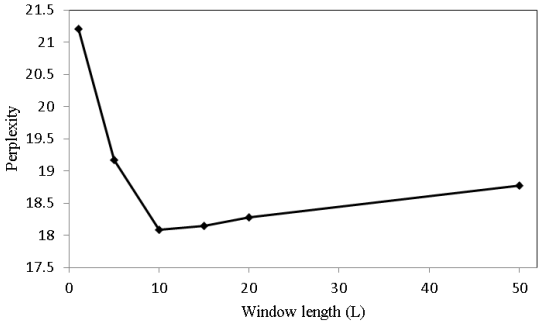

Fig. 5. Perplexity of the model as a function of window length according to the settings reported in Table 1**.**

Table 2. Zero-occurrences decrease by incrementing co-occurrence window.

Size of the co-occurrence window	Number of nonzero elements of φ
$L=1$	1567457
$L=5$	1950822
$L=10$	2446655
$L=20$	2524235

When the window size is 1, the proposed model is equivalent to BTM shown in Fig. 2. The results of our experiments, reported in Table 2, show that by increasing the window size, the number of zero-occurrences decreases. Zero-occurrences are the number of possible elements of φ which have not

been occurred in the iterations. Although increasing the window size in our model does not affect the number of possible states, i.e., $N \times N \times K$. It means that by increasing the window size sparseness decreases, which theoretically improves the model robustness.

It can also be a good criterion for deciding on the size of the co-occurrence window. One can see in Table 2 that from $L=10$ to $L=20$ the number of non-zero occurrences have not increased significantly and therefore we used $L=10$ in our experiments. This is another reason that

We ran 2 Markov chains for 1000 iterations and discarding the first 500 iterations we took one sample form the chain at a lag of 100 iterations. For all runs of the algorithm, we used $\beta=0.01$, $\alpha=50/K$ and $\gamma=1/(50*L+K)$.

Table 3. One of the produced topics: each topic is a set of several distributions over words.

Topic: 0			
Jesus:	Student:	Assault:	Authorities:
save apparently troubled boy family god death grandmother friends	veterans save teacher friends classroom students jesus jammed convicted excepted	victims weapon past stein fear decade related night forever	federal adult things brush spokesman border charge due largest lost

One randomly-selected topic produced by this model is shown in Table 3. As shown in Table 3, in the proposed model, each topic is not a single distribution over words. It consists of N distributions over words, where N is the size of the vocabulary. It means that the same word, according to its vicinity, may be assigned to deferent topics.

In Table 3, for example, the probability of the pair "Jesus" and "save" is 0.001 or the probability of "Jesus" and "god" is 0.0007. We selected the 7 most probable of such pairs. The results are shown in

Table 4 for some random topics.

Table 4 contains 3 topics generated by the proposed model and Table 5 contains similar topics generated by BTM. Consider topic 3 of the proposed model, which is about stock exchange in Wall Street. Topic 18 of BTM is also about stock exchange in Wall Street. The pairs in each topic are sorted in descending order according to their probability in the topic. The forth pair and the last two pairs of topic 3 of BTM are general or unrelated words but it is not the case for topic 3 of the proposed model. The last word of topic 3 of BTM is “assistant, attorney” which can be considered more related to topic 10 of our proposed model. It seems that the first topic in Table 5 is divided into more coherent topics in the proposed model; topic 3 and topic 10. Topic 18 of our proposed model and similar topics 2 and 19 of BTM are other examples of such behavior.

As one can see in Table 6, the presented topic, generated by LDA, is similar to topic 3 in

Table 4 generated by the proposed model. Apparently both topics are related to "stock exchange" but the topic generated by LDA, including words like rose and rate, is a general topic while the one generated by our model being shown by word pairs is more specific and meaningful.

The three models are also compared objectively, according to their perplexity. Fig. 6 shows the perplexities of the tree models as a function of the iterations of the Gibbs sampling process. As one can see in this figure, the lowest perplexity belongs to the proposed model.

Table 4. A representation of some random topics, generated by the proposed model, by their most probable pairs.

Topic 3	Topic 10	Topic 18
wall, street jones, average average, industrials exchange, market nyse, composite stock, exchange trading, stock	supreme, court grand, jury district, judge death, penalty northern, Ireland appeals, court law, enforcement	billion, billion stock, exchange composite, index bush, administration savings, loan real, estate dow, jones

Table 5. A representation of some random topics, generated by BTM, by their most probable pairs.

Topic 18	Topic 2	Topic 19
wall, street dow, jones composite, index coast, guard jones, average big, board assistant, attorney	mercantile, exchange share, index miles, south district, attorney cents, cents cents, lower cent, higher	savings, loan bank, board executive, officer north, america chairman, executive inches, snow inches, rain

Table 6. A topic generated by LDA.

Topic
Stock market index rose share average rate

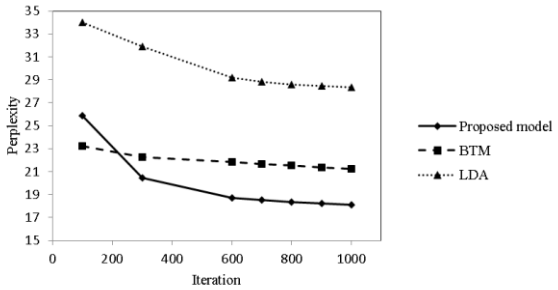

Fig. 6. Perplexity as a function of Gibbs sampling iterations

VI. CONCLUSION

Many current topic models, despite not being restricted to textual data, can only be applied to one-dimensional data. These models do not consider the local or spatial information of data and their performance is poor when it comes to short documents. They are all restricted to a very long co-occurrence window of a whole document or a window as short as two words. In this paper, we proposed a model which lifts this constraint and lets the designer decide on the suitable length of co-occurrence window based on dataset or application at hand. We derived the model parameters using Gibbs sampling and applied it on a dataset of 2000 documents. The evaluation results show that the model reduces the sparseness compared to the BTM that takes the local word dependencies into account. Also the proposed model produces more meaningful topics than LDA and BTM and is a better model of the corpus according to its perplexity.

REFERENCES

[1] Blei, D.M., A.Y. Ng, and M.I. Jordan, *Latent dirichlet allocation.* Journal of machine Learning research, 2003. **3**(Jan): p. 993-1022.

[2] Luo, W. and T.B. Zhang. *Blind Image Quality Assessment Using Latent Dirichlet Allocation Model*. 2014. Trans Tech Publ.

[3] Savoy, J., *Authorship attribution based on a probabilistic topic model.* Information Processing & Management, 2013. **49**(1): p. 341-354 %@ 0306-4573.

[4] Razavi, A.H. and D. Inkpen. *Text Representation Using Multi-level Latent Dirichlet Allocation*. in *Canadian Conference on Artificial Intelligence*. 2014. Springer.

[5] Blei, D.M. and J.D. Lafferty. *Dynamic topic models*. 2006. ACM.

[6] Wallach, H.M. *Topic modeling: beyond bag-of-words*. 2006. ACM.

[7] McAuliffe, J.D. and D.M. Blei. *Supervised topic models*. 2008.

[8] Porteous, I., et al. *Fast collapsed gibbs sampling for latent dirichlet allocation*. 2008. ACM.

[9] Anandkumar, A., et al. *A spectral algorithm for latent dirichlet allocation*. 2012.

[10] Andrzejewski, D., X. Zhu, and M. Craven. *Incorporating domain knowledge into topic modeling via Dirichlet forest priors*. 2009. ACM.

[11] Barbieri, N., et al., *Probabilistic topic models for sequence data.* Machine learning, 2013. **93**(1): p. 5-29 %@ 0885-6125.

[12] Griffiths, T.L., et al. *Integrating Topics and Syntax*. 2004.

[13] Griffiths, T.L., M. Steyvers, and J.B. Tenenbaum, *Topics in semantic representation.* Psychological review, 2007. **114**(2): p. 211 %@ 1939-1471.

[14] Wang, X., A. McCallum, and X. Wei. *Topical n-grams: Phrase and topic discovery, with an application to information retrieval*. 2007. IEEE.

[15] Jameel, S., W. Lam, and L. Bing, *Supervised topic models with word order structure for document classification and retrieval learning.* Information Retrieval Journal, 2015. **18**(4): p. 283-330 %@ 1386-4564.

[16] Yang, G., et al., *A novel contextual topic model for multi-document summarization.* Expert Systems with Applications, 2015. **42**(3): p. 1340-1352 %@ 0957-4174.

[17] Momtazi, S., S. Khudanpur, and D. Klakow. *A comparative study of word co-occurrence for term clustering in language model-based sentence retrieval*. 2010. Association for Computational Linguistics.

[18] Evert, S., *The statistics of word cooccurrences: word pairs and collocations.* 2005.

Marziea Rahimi is a Ph.D. student at the Deprartment of Computer Engineering of Shahrood University of Technology. Her research interests revolve around statistical machine learning, text mining and topic modeling.

Morteza Zahedi is an assistant professor at the Department of Computer Engineering of Shahrood University of Technology. He received his Ph.D. from RWTH Aachen University. His research interests focus on statistical pattern recognition, data mining and machine vision.

Hoda Mashayekhi is an assistant professor at the Department of Computer Engineering of the Shahrood University of Technology. She received her Ph.D. from Sharif University of Technology in 2013. Her research interests include parallel and distributed computing, data mining, decision making, peer-to-peer (P2P) networks and semantic structures.

This Page intentionally left blank.