Topic Evolution Modeling in Social Media Short Texts Based on Recurrent Semantic Dependent CRP

Yuhao Zhang[1,2] Wenji Mao[1,2] Daniel Zeng[1,2]
[1]State Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, Chinese Academy of Sciences
[2]School of Computer and Control Engineering, University of Chinese Academy of Sciences
{zhangyuhao2012, wenji.mao, dajun.zeng}@ia.ac.cn

***Abstract*—Social media has become an important platform for people to express opinions, share information and communicate with others. Detecting and tracking topics from social media can help people grasp essential information and facilitate many security-related applications. As social media texts are usually short, traditional topic evolution models built based on LDA or HDP often suffer from the data sparsity problem. Recently proposed topic evolution models are more suitable for short texts, but they need to manually specify topic number which is fixed during different time period. To address these issues, in this paper, we propose a nonparametric topic evolution model for social media short texts. We first propose the recurrent semantic dependent Chinese restaurant process (rsdCRP), which is a nonparametric process incorporating word embeddings to capture semantic similarity information. Then we combine rsdCRP with word co-occurrence modeling and build our short-text oriented topic evolution model sdTEM. We carry out experimental studies on twitter dataset. The results demonstrate the effectiveness of our method to monitor social media topic evolution compared to the baseline methods.**

Keywords—Social Media Analytics; Topic Modeling; Text Mining

I. INTRODUCTION

Social media is growing fast in recent years, which makes it an important platform for communication, opinion expression and information seeking and sharing. Social media texts are usually short, such as tweets and status messages. Detecting and tracking topics from social media short texts can help people understand key contents and their evolution in social media data, which facilitates many security-related applications, such as government decision-making, emergency response, crisis management and so on.

Topic model is one of the most important tool for topic detection and tracking. Conventional topic models like LDA [1] and nonparametric topic model HDP [2] have gained great success in extracting topics from documents. These models represent each document as a mixtures of topics and each topic as a distribution over words. Many topic evolution models are built based on these models. For example, TOT [3], DTM [4], Online LDA [5] and TTM [6] are built based on LDA. HDP-ISM [7] and EvoHDP [8] are built based on HDP. However, these topic evolution models usually not perform very well on social media short texts due to the data sparsity problem.

Recently, some models have been proposed which are suitable for tracking topics in short texts. TM-LDA [9] can model topics and topic transition in social media data. Twitter-TTM [10] is proposed by combining short text topic model Twitter-LDA [11] and the traditional topic evolution model TTM [6]. Online BTM [12] models the word co-occurrence patterns directly with an online approach. However, the above models all need to specify the topic number manually, and the topic number could not change during different time periods, which is inconsistent with the real world topic evolution process.

So in this paper, we propose a nonparametric topic evolution model (sdTEM) for social media short texts. To overcome the data sparsity problem of short texts, we use word embeddings [13, 14] to enrich the semantics and use the word co-occurrence modeling to capture the co-occurrence patterns directly. Word embeddings are distributed representations of words, which could project words with similar semantic into the same area in the vector space. So it is an effective tool for capturing the semantic information. We use word embeddings to measure the semantic similarity between words and topics and propose the recurrent semantic dependent Chinese restaurant process (rsdCRP). This process samples the topic for each word using additional semantic information including the semantics of word itself and the semantics of topics in current and previous time epochs. Combining the rsdCRP and word co-occurrence modeling together, we propose the nonparametric topic evolution model sdTEM for social media short texts.

We summarize the contributions of our work as follows:

(1). We propose the recurrent semantic dependent Chinese restaurant process, a nonparametric statistical process which could capture semantic similarity information using word embeddings;

(2). We combine the recurrent semantic dependent Chinese restaurant process with the word co-occurrence modeling and build the nonparametric topic evolution model sdTEM for social media short texts;

(3). We conduct experimental analysis on real world dataset. The experimental results demonstrate the effectiveness of our method compared with the baseline methods.

II. RELATED WORK

A. Topic Detection and Tracking

Topic detection and tracking (TDT) is one of the most important tasks in data mining and text analysis and has drawn a lot of research attention. Early works often adopt the clustering algorithms. Then the topic model gains great success in text analysis and becomes an important tool for TDT. Many topic evolution models are built based on conventional topic models like LDA and HDP.

TOT [3] is a LDA-style topic model which could capture the changes of topic over time by associating each topic with a continuous distribution over timestamps. DTM [4] is built by using Gaussian time series on the natural parameters of the multinomial topics and logistic normal topic proportion models. Online LDA [5] works in an online fashion and could identify the topic changes. TTM [6] can be used for tracking trend in each topic.

The above LDA-based topic evolution models all need to specify the topic number manually, which is difficult to determine accurately and the topic number could not change during different time periods. So some non-parametric topic evolution models based on HDP are proposed. HDP-ISM [7] could incrementally derive semantic clusters in text stream. EvoHDP [8] could detect and track topics from multiple correlated time-varying text corpora.

However, these traditional topic evolution models are not designed specifically for short texts. So they often suffer from the data sparsity problem when used on short texts.

Some recently proposed topic evolution models are more suitable for short texts. TM-LDA [9] could be used for modeling the topic transitions in social media data. Twitter-TTM [10] is designed based on Twitter-LDA [11], a user based topic model for short texts. Online BTM [12] is proposed based on BTM [12], which is a successful short text topic model built by modeling the word co-occurrence patterns directly. However, these models also need to specify the topic number manually. So in this paper, we propose a non-parametric topic evolution model which is suitable for short texts.

B. Short Text Modeling

As short texts are usually sparse and noisy, using conventional topic models directly on short texts often not perform very well. Since short texts are widely used in the Web now days, short text modeling has drawn a lot of attention.

Traditional methods [15] adopt the “aggregation strategies” by combining short texts together as pseudo documents and then use conventional topic model for topic analysis. These methods could improve the performance of the conventional topic models in short texts. However, these methods do not really model the short texts and their performance vary with datasets and aggregation strategies.

Several strategies have been proposed for modeling short texts. One way is to restrict the number of topics in each document. For example, the Twitter-LDA [11] is built by assuming that each tweet only contains one topic besides the background topic.

Another way is to model the word co-occurrence. Using non-negative matrix factorization on global word co-occurrence matrix could detect topics in short texts effectively [16]. BTM [12] further extends it to a more principle approach by modeling the generative process of word co-occurrence patterns in corpus. It gets good performance in both short and normal texts.

Self-aggregation strategy is also proposed for short texts modeling. Both SATM [17] and PTM [18] are built by automatically self-aggregating of short texts during the topic modeling process.

All the above models are used for static topic mining. Works on topic evolution for short texts are relatively small. In this paper, we propose a topic evolution model for short texts by adopting the strategy of modeling the word co-occurrence. We combine this strategy with the rsdCRP we proposed to build the topic evolution model.

III. PROPOSED METHOD

In this section, we will introduce the nonparametric topic evolution model for short texts we proposed. We first give a brief introduction about the word embedding. Then we introduce the recurrent semantic dependent Chinese restaurant process we proposed, which incorporates semantic information with word embeddings. Finally, we present our short text topic evolution model sdTEM as well as the inference process for the method.

A. Word Embeddings

In short texts, words with high semantic relation may not co-occur frequently in same context. But word embeddings as distributed representations of words could capture the semantic information effectively and put semantically related words into same area in the vector space. So the word embedding could be seen as an effective semantic supplement to the original texts. We could use word embeddings pre-trained on large corpus to get accurate distributed representations and incorporate them into our model to enrich the semantic information.

Word2Vec [13] is one of the most popular methods for generating word embeddings. It learns vector representations for words using a shallow neural network architecture. Word2Vec has two model architectures. One is the skip-gram model, and the other one is continuous bag-of-words model. The skip-gram model uses the current word to predict the context words in the surrounding window. On the contrary, the continuous bag-of-words model predicts the current word using the context words in surrounding window. However, Word2Vec only utilizes the local context. The GloVe [14] further improves the word vector representation by adding global word co-occurrence information of corpora into the model. In this paper, we use the GloVe pre-trained word embedding vectors and incorporate them into our topic evolution model.

B. Recurrent Semantic Dependent Chinese Restaurant Process

Traditional clustering techniques need to manually specify the number of clusters. But the non-parametric technique could adapt the number of clusters automatically based on the data itself. So we adopt the non-parametric technique in our model to help determine the topic number automatically.

The Chinese restaurant process (CRP) is one of the most widely used non-parametric technique for static topic modeling, which could partition data into a proper number of clusters automatically. To model the text streams, a framework called recurrent Chinese restaurant process (RCRP) [19] has been proposed. In the RCRP scenario, the streaming data is divided into different time epochs. The number of clusters at each epoch is unbounded and the cluster could remain, die out or emerge over time. However, the RCRP treats each arriving data equally without considering the information the data carrying or its relationship with other data in the epoch.

So we adopt the concept of the distance dependency [20] and propose the recurrent semantic dependent Chinese restaurant process. In this process, arriving data and clusters are all represented by embedding vectors. So we can calculate the similarity between the arriving data and clusters. By this means, the data is more likely to be distributed into the right cluster. And the embedding vectors carrying the semantic information of clusters could be transmitted through different epochs.

Now we first give a brief introduction of the CRP and then introduce our recurrent semantic dependent Chinese restaurant process. The CRP is a distribution on partitions of integers. It works according to the following metaphor: suppose a sequence of customers enter a restaurant with infinite tables, each customer sits at an occupied table with probability proportional to the number of previous customers sitting by the table, and at an unoccupied table with probability proportional to α. In CRP, the probability of a particular seating configuration is not influenced by the customers' arriving order.

Let z_i be the index of the table where customer i sits and z_{-i} be all the other customers' seats. Let m_k denotes the number of customers sitting at table k and K denotes the number of tables where $m_k > 0$. Let α be the parameter of CRP. The probability of customer i sitting at each table is represented as follows:

$$p(z_i = k \mid z_{-i}, \alpha) = \begin{cases} \dfrac{m_k}{\sum_{j=1}^{K} m_j + \alpha} & k \le K \\ \dfrac{\alpha}{\sum_{j=1}^{K} m_j + \alpha} & k = K+1 \end{cases}$$

To incorporate the semantic information into the CRP, we assume each customer is represented by a embedding vector carrying its semantic information. And each table is represented by a embedding vector which is the sum of the vectors of customs sitting by the table. So a customer sits at an occupied table with probability proportional to the similarity between the customer and the table, and at an unoccupied table with probability proportional to α. The process can be represented as follows:

$$p(z_i = k \mid z_{-i}, \alpha) = \begin{cases} \dfrac{f(s_{i,k})}{\sum_{j=1}^{K} f(s_{i,j}) + \alpha} & k \le K \\ \dfrac{\alpha}{\sum_{j=1}^{K} f(s_{i,j}) + \alpha} & k = K+1 \end{cases}$$

in which $s_{i,k}$ represents the similarity between customer i and table k. Let e_i be the embedding vector of customer i and S_k be the set of customers sitting at table k. The $s_{i,k}$ can be calculated as follows:

$$s_{i,k} = e_i \bullet \sum_{j \in S_k} e_j$$

And f is a function defined as follows:

$$f(a) = \begin{cases} a & a > 0 \\ 0 & a \le 0 \end{cases}$$

For data stream, we partition it into different time epochs. To ensure the consistency of clusters in different epochs, choosing a cluster for a customer should consider the information in current epoch as well as in previous epochs. So in time epoch t, customer i sits at an occupied table k with probability proportional to $f(s_{i,k,t}) + f(s'_{i,k,t})$, in which $f(s_{i,k,t})$ is calculated based on the information in current epoch and $f(s'_{i,k,t})$ based on the information in previous epochs.

Information in early epochs should have less influence to the customer, so we define $f(s'_{i,k,t})$ as follows:

$$f(s'_{i,k,t}) = \sum_{\delta=1}^{\Delta} f(s_{i,k,t-\delta}) \times e^{-\delta/\lambda}$$

in which λ is the decay factor and Δ is the width of time epochs we use.

The rsdCRP we proposed is represented as follows:

$$p(z_{i,t} = k \mid z_{-i,t}, z_{t-\Delta:t-1}, \alpha, \lambda, \Delta, E) \propto \begin{cases} f(s_{i,k,t}) + f(s'_{i,k,t}) & \textit{existing cluster} \\ \alpha & \textit{new cluster} \end{cases}$$

where $z_{-i,t}$ denotes the cluster assignment of all data in epoch t except $z_{i,t}$. $z_{t-\Delta:t-1}$ denotes the cluster assignment of all data from epoch $t-\Delta$ to epoch $t-1$. E denotes the set of embedding vectors.

C. sdTEM Model

To reduce the limitation of data sparsity, we adopt the idea of word co-occurrence modeling and model the word co-occurrence directly. By this means, we can use the aggregated word co-occurrence patterns in documents at each epoch instead of patterns at document level, which could help solve the data sparsity problem.

We turn the original texts into biterms. Each biterm is an unordered word-pair co-occurring in the same context. So a document containing n words will be converted into C_n^2 biterms. As words within the same biterm come from the same context, they usually belong to same topic. So we assume words in one biterm share the same topic. We use the sum of word embeddings of the two words within a biterm as the embedding vector for the biterm.

In our approach, we partition the text data into different epochs. The original texts are turned into biterms for all epochs. In each epoch, we use rsdCRP as the prior to determine the topic of each biterm. The rsdCRP could use the semantic information based on the biterm embedding vectors and the topic cluster embedding vectors. Besides, rsdCRP could take advantage of the topic semantic information of previous epochs. As rsdCRP is a non-parametric process, it could adjust the topic number in each epoch automatically to fit the data.

In sdTEM, for each biterm, we first sample its topic using rsdCRP, then we sample the words in this biterm based on the corresponding word distribution of the determined topic. The generative process of the sdTEM is as follows:

For each time epoch t

For each biterm $b_{i,t}$

a) Draw topic $z_{i,t} \sim$ rsdCRP(α,λ,Δ)

b) If $z_{i,t}$ is a new topic in epoch t,

— draw topic word distribution $\phi_{t,z_{i,t}} \sim Dirichlet(\beta)$

c) For word $w_{i,1,t}$ 和 $w_{i,2,t}$ in $b_{i,t}$:

i. — draw $w_{i,1,t} \sim Multinomial(\phi_{t,z_{i,t}})$

ii. — draw $w_{i,2,t} \sim Multinomial(\phi_{t,z_{i,t}})$

D. Parameters Inference

We use collapsed Gibbs sampling to estimate the parameters of our model. Gibbs sampling is a widely used inference method based on Markov chain Monte Carlo algorithm. It cycles through the variables and estimates them iteratively. The method samples one variable each time based on the distribution of that variable conditioned on the remaining variables. The collapsed Gibbs sampling takes advantage of the conjugate distribution by integrating out some variables. It can simplify the sampling procedure.

In our model, we only need to sample the topic $z_{i,t}$ for each biterm $b_{i,t}$. The topic-word distributions $\phi_{t,z_{i,t}}$ are integrating out during the sampling process and could be calculated based on the sampling results. To perform the Gibbs sampling, we first initialize the variables randomly. By applying the chain rule on the joint probability of the whole dataset, we get the sampling formula for $z_{i,t}$ as follows:

$$P(z_{i,t}=k \mid b_{i,t}, z_{-i,t}, z_{t-\Delta:t-1}, \alpha, \beta, \lambda, \Delta, E) \propto$$
$$\begin{cases} (f(s_{i,k,t}) + f(s'_{i,k,t}))\dfrac{(n_{w_{i,1,t}|z_{i,t}}+\beta)(n_{w_{i,2,t}|z_{i,t}}+\beta)}{(\sum_w n_{w|z_{i,t}}+W\beta)^2} & (existing\ topic) \\ \\ \alpha\dfrac{(n_{w_{i,1,t}|z_{i,t}}+\beta)(n_{w_{i,2,t}|z_{i,t}}+\beta)}{(\sum_w n_{w|z_{i,t}}+W\beta)^2} & (new\ topic) \end{cases}$$

where W denotes the size of the vocabulary and $n_{w|z_{i,t}}$ denotes the number of word w assigned to topic $z_{i,t}$ in time epoch t.

Based on the sampling results, we can calculate the topic word distribution $\phi_{t,w|z_{i,t}}$ in time epoch t as follows:

$$\phi_{t,w|z_{i,t}} = \frac{n_{w|z_{i,t}}+\beta}{\sum_w n_{w|z_{i,t}}+W\beta}$$

IV. Experiments

In this section, we evaluate the performance of our topic evolution model sdTEM and compare it with the baseline methods on twitter dataset.

A. Experimental Setup

Dataset: We sample tweets from the twitter7 dataset for our experiment. The twitter7 dataset contains hundreds of millions of tweets covering 7 month from June 1, 2009 to December 31, 2009. In our expriment, we use the tweets from July to examine our method. We first filter out the non-English tweets and remove stop words from the remaining ones. Then we convert the letters of all tweets into lower case and restrict the vocabulary to 10000 words based on the word frequency. We filter out tweets with only one word. After that, we sample 10000 tweets each day for two weeks. We use these tweets to train the topic evolution model and the remaining tweets will be used as reference dataset for the topic coherence evaluation task.

Baseline Methods: We compare our model with three baseline methods on twitter dataset. As the traditional topic evolution methods often suffer from the data sparsity, we only compare our method sdTEM with one typical traditional method DTM. The second baseline is Online BTM, which is a representative topic evolution method for short texts. The third one is rTEM, a variation of our method. It is built by combining the RCRP with the word co-occurrence modeling without considering the semantic information.

Parameter Setting: We use one day as one time epoch and divide the tweets into days. In our model sdTEM, we set the parameters as α=10, β=0.1 and λ=0.5. Since the topics of tweets in each epoch are most close to those in the previous epoch, without loss of generality, we set Δ to 1. For rTEM, we use the same setting as sdTEM. For DTM, we adopt the defaut setting. For Online BTM, we adopt the setting in the original work [12] as β=0.005 and α=50/K, in which K is the topic number. As DTM and Online BTM could not determine the topic number automatically, we choose three topic numbers 80, 100 and 120 for evaluation. By this means, we can see the relation between the model performance and predefined topic number.

B. Evaluation of Topic Quality

Traditionally, topic models could be evaluated by comparing the perplexity or marginal likelihood on a held-out dataset. However, recent studies show that these evaluation metrics could not reflect the interpretability of topics very well. Besides, our model optimizes the likelihood of word occurrences directly instead of the whole documents, which is different from traditional models like LDA.

TABLE I. TOPIC COHERENCE RESULTS OF TOPIC EVOLUTION MODELS

Method		DTM				Online BTM				rTEM	sdTEM
Topic Number		80	100	120	Average	80	100	120	Average	–	–
LCP	L=5	-74.12	-77.83	-82.14	-78.03	-64.18	-68.93	-71.23	-68.11	-68.64	**-63.91**
	L=10	-370.18	-385.72	-403.51	-386.74	-321.75	-339.78	-357.28	-339.60	-343.35	**-314.12**
	L=15	-906.42	-940.64	-978.26	-941.78	-791.72	-835.81	-884.14	-837.22	-847.35	**-778.05**

So in this paper, we use topic coherence and topic distinctiveness as the evaluation measures. Topic coherence could reflect the interpretability of topics and topic distinctiveness could reflect the distinctiveness between topics. The two measures could reflect the quality and diversity of the generated topics.

Our model is a non-parametric approach and few small cluster may exists after the sampling process. In our experiment, we omit topics with less than 30 biterms, since these topics contain too little information to form any meaningful topics in practice.

Topic Coherence

Topic coherence is an effective indicator of topic quality. It could reflect the topic interpretability and match well with human judgements of topic quality. Here we use LCP [21] as the topic coherence measure.

$$LCP(t) = \sum_{j=2}^{L}\sum_{i=1}^{j-1}\log\frac{P(w_i, w_j)}{P(w_i)}$$

Higher LCP score indicates better topic quality. The word probabilities and word co-occurrence probabilities should be computed on a large-scale external dataset which is often referred as reference dataset. Here we use remaining tweets of July as reference dataset to calculate these probabilities.

The LCP score is calculated based on the top L words of each topic. In our experiment, we set L to 5, 10 and 15 respectively. We use the average LCP score of the detected topics in the whole time period to evaluate topic evolution models.

Table I shows the average LCP score of all methods. We can see that DTM do not perform very well, since it is not designed specifically for short texts. Online BTM performs better than DTM, but its score varies with the predefined topic numbers. sdTEM performs better than rTEM, which illustrates that taking advantage of semantic information carried by word embeddings could help the topic evolution modeling. The sdTEM also performs better than the other two methods, which proves the effectiveness of our method.

Topic Distinctiveness

The topics generated by topic models should be distinctive from each other. The distinctiveness of topics extracted by a topic model could reflect the information diversity. If two topics are similar, they are considered carrying redundant information.

To measure the topic distinctiveness, we use Kullback-Leibler divergence (KLD) to measure the divergence of two topic distributions. As Kullback-Leibler divergence is non-symmetric, we define the metric of topic distinctiveness $D(P\|Q)$ as follows:

$$D(P \| Q) = \frac{1}{2}[(D_{KL}(P \| Q) + D_{KL}(Q \| P)]$$

in which P and Q represent two topic distributions. And $D_{KL}(P\|Q)$ denotes the Kullback-Leibler divergence between P and Q. To evaluate the topic model, we first calculate the topic distinctiveness scores between any pair of topics in one time epoch and get their average. Then we further average the scores of all time epochs. We use the final score as the evaluation metric of topic model.

The topic distinctiveness results are shown in Table II. We can see that the topic distinctiveness scores of DTM and Online BTM vary with the predefined topic numbers. The topic distinctiveness of sdTEM is larger than other methods. The reason may be that the sdTEM incorporates more semantic information in the modeling process, which makes it more likely to put semantic related words into same topic. So each generated topic is focused on one theme and the generated topics are more distinctive from each other compared to other methods.

TABLE II. TOPIC DISTINCTIVENESS RESULTS OF TOPIC EVOLUTION MODELS

Method	Topic Number	Topic Distinctiveness
DTM	80	4.06
	100	3.72
	120	3.43
	Average	3.74
Online BTM	80	6.27
	100	5.81
	120	5.29
	Average	5.79
rTEM	—	5.83
sdTEM	—	**6.29**

Topic Illustration

We illustrate the topics about Obamacare generated by our method sdTEM as well as rTEM and Online BTM (K=80) which perform well in the expriments. The results are shown in Table III.

TABLE III. Obamacare TOPIC Illustration

Method	Top Words
sdTEM	health care obama public plan bill president insurance lead option
rTEM	health obama care people **american** public support insurance lead president
Online BTM (K=80)	health care obama support president lead **lower** insurance americans plan

The words which are not so relevent to the obamacare topic are marked bold in Table III. We can see these methods all provide reasonable topics. But overall, sdTEM derives relatively more coherent topic than the baselines. This may due to the sdTEM incorporating semantic information into the modeling process which help increase the topic quality.

We also show an example of the topic evolution process in Figure 1, which is about the memorial to Michael Jackson. We can see the evolution of topic words distribution over time.

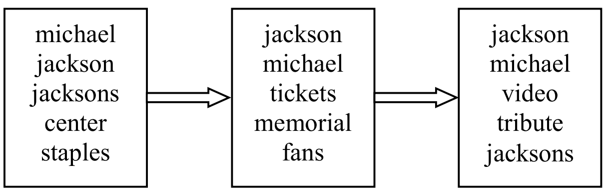

Fig 1. Topic Evolution Example

V. CONCLUSIONS

Social media has become an important platform for people to express opinions, share information and communicate with others. The textual information in social media is usually short. So detecting and tracking topics from social media short texts becomes a very important task, which could not only help people grasp essential information from social media but also facilitate many security related applications. Most topic evolution models are not designed specifically for short texts. So they often suffer from the data sparsity problem when facing short texts. Recently proposed topic evolution models for short texts need to specify topic number manually. So in this paper, we propose a nonparametric topic evolution model for short texts. We propose the recurrent semantic dependent Chinese restaurant process which could capture semantic similarity information using word embeddings. We combine it with co-occurrence modeling and build our model sdTEM. We carry out empirical studies on real-world twitter dataset. The experimental results show our method could generate coherent topics and monitor the topic evolution effectively compared to the baseline methods.

ACKNOWLEDGEMENTS

This work is supported in part by NSFC Grant #71621002, the Ministry of Science and Technology of China Major Grant #2016QY02D0205, and CAS Key Grant #ZDRW-XH-2017-3.

REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[2] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, “Hierarchical dirichlet processes,” *Journal of the american statistical association*, vol. 101, no. 476, 2006.

[3] 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*, New York, NY, USA, 2006, pp. 424–433.

[4] D. M. Blei and J. D. Lafferty, “Dynamic Topic Models,” in *Proceedings of the 23rd International Conference on Machine Learning*, New York, NY, USA, 2006, pp. 113–120.

[5] L. AlSumait, D. Barbara, and C. Domeniconi, “On-line LDA: Adaptive Topic Models for Mining Text Streams with Applications to Topic Detection and Tracking,” in *Eighth IEEE International Conference on Data Mining, 2008. ICDM ’08*, 2008, pp. 3–12.

[6] T. Iwata, S. Watanabe, T. Yamada, and N. Ueda, “Topic Tracking Model for Analyzing Consumer Purchase Behavior,” in *Proceedings of the 21st International Jont Conference on Artifical Intelligence*, San Francisco, CA, USA, 2009, pp. 1427–1432.

[7] Z. Gao *et al.*, “Tracking and Connecting Topics via Incremental Hierarchical Dirichlet Processes,” in *2011 IEEE 11th International Conference on Data Mining (ICDM)*, 2011, pp. 1056–1061.

[8] J. Zhang, Y. Song, C. Zhang, and S. Liu, “Evolutionary Hierarchical Dirichlet Processes for Multiple Correlated Time-varying Corpora,” in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010, pp. 1079–1088.

[9] Y. Wang, E. Agichtein, and M. Benzi, “TM-LDA: Efficient Online Modeling of Latent Topic Transitions in Social Media,” in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2012, pp. 123–131.

[10] K. Sasaki, T. Yoshikawa, and T. Furuhashi, “Online topic model for Twitter considering dynamics of user interests and topic trends,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1977–1985.

[11] W. X. Zhao *et al.*, “Comparing Twitter and Traditional Media Using Topic Models,” in *Advances in Information Retrieval*, 2011, pp. 338–349.

[12] X. Cheng, X. Yan, Y. Lan, and J. Guo, “BTM: Topic Modeling over Short Texts,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 2928–2941, 2014.

[13] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *Advances in Neural Information Processing Systems*, 2013, pp. 3111–3119.

[14] J. Pennington, R. Socher, and C. Manning, “Glove: Global Vectors for Word Representation,” presented at the Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1532–1543.

[15] L. Hong and B. D. Davison, “Empirical study of topic modeling in twitter,” in *Proceedings of the First Workshop on Social Media Analytics*, 2010, pp. 80–88.

[16] X. Yan, J. Guo, S. Liu, X. Cheng, and Y. Wang, “Learning Topics in Short Texts by Non-negative Matrix Factorization on Term Correlation Matrix,” in *Proceedings of the 2013 SIAM International Conference on Data Mining*, 0 vols., Society for Industrial and Applied Mathematics, 2013, pp. 749–757.

[17] X. Quan, C. Kit, Y. Ge, and S. J. Pan, “Short and Sparse Text Topic Modeling via Self-aggregation,” in *Proceedings of the 24th International Conference on Artificial Intelligence*, Buenos Aires, Argentina, 2015, pp. 2270–2276.

[18] Y. Zuo *et al.*, “Topic Modeling of Short Texts: A Pseudo-Document View,” in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2016, pp. 2105–2114.

[19] A. Ahmed and E. Xing, “Dynamic Non-Parametric Mixture Models and The Recurrent Chinese Restaurant Process : with Applications to Evolutionary Clustering,” in *Proceedings of The Eighth SIAM International Conference on Data Mining (SDM2008)*, 2008.

[20] D. M. Blei and P. I. Frazier, “Distance Dependent Chinese Restaurant Processes,” *J. Mach. Learn. Res.*, vol. 12, pp. 2461–2488, Nov. 2011.

[21] J. H. Lau, D. Newman, and T. Baldwin, “Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality,” 2014, pp. 530–539.