2017 IEEE International Conference on Big Knowledge

Modeling Topic Evolution in Social Media Short Texts

Yuhao Zhang[1,2] Wenji Mao[1,2] Junjie Lin[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, linjunjie2013}@ia.ac.cn

***Abstract*—Social media short texts like tweets and instant messages provide a lot of valuable information about the hot topics and public opinion. Detecting and tracking topics from these online contents can help people grasp the essential information and its evolution and facilitate many applications. Topic evolution models built based on LDA need to set the topic number manually, which could not change during different time periods and could not be adjusted based on the contents. The nonparametric topic evolution models do not perform very well on short texts due to the data sparsity problem. So in this paper, we propose a nonparametric topic evolution model for short texts. The model uses the recurrent Chinese restaurant process as the prior distribution of topic proportions. Combining it with word co-occurrence modeling, we construct a topic evolution model which is suitable for social media short texts. We carry out experimental studies on twitter dataset. The results show that our method outperforms the baseline methods and could monitor the topic evolution in social media short texts effectively.**

Keywords—Topic Modeling; Text Mining; Social Media Analytics

I. Introduction

With the rapid growth of social media in recent years, it has become an important platform for information seeking and sharing. Social media texts like microblogs and status messages are usually short. Considering the large volume of these texts available online, modeling the topic evolution in social media short texts has become increasingly important for text stream analysis. It can help people understand the key contents and facilitate many applications related to semantic analysis like summarization [1-3], recommendation [4-6], information diffusion [7, 8] and so on.

Topic model is a widely used technique for topic detection and tracking. Lots of topic evolution models are built based on classic topic models. For example, TOT [9], DTM [10] and Online LDA [11] are built based on LDA [12]. TOT [9] could capture the changes of topic over time by associating each topic with a continuous distribution over timestamps. DTM [10] is built by using state space models on the natural parameters of the multinomial distributions that represent the topics. Online LDA [11] is an online version of LDA which could be updated incrementally. However, these topic evolution models constructed based on LDA need to manually set the topic number, which is not easy to determine when facing a new corpus. Besides, the predefined topic number could not change during different time epochs, which means these models could not adjust the topic number to fit the contents during the topic evolution process.

On the other hand, some topic evolution models adopt the nonparametric technique which could help determine the topic number in different time epochs automatically. These models are usually constructed by extending HDP [13]. For example, HDP-ISM [14] could incrementally derive semantic clusters in a text stream. EvoHDP [15] could model the topic evolution using multiple correlated time-varying text corpora. However, these models are not designed specifically for short texts. So they usually do not perform very well on social media short texts due to the data sparsity.

Recently, short text modeling has drawn a lot of attention. Several methods have been proposed for short texts topic detection with different strategies. Twitter-LDA [16] is built by restricting one tweet only containing one topic besides the background topic. SATM [17] and PTM [18] are built by adopting the self-aggregation strategy. BTM [19] is built by modeling the word co-occurrence directly. But these methods only focus on topic discovering without modeling the topic evolution. Online BTM [20] is proposed as an online version of BTM [19], but this model also need a manually defined topic number which is fixed during different time epochs.

So in this paper, we propose a nonparametric dynamic topic modeling method for short texts. We use the recurrent Chinese restaurant process [21] as the nonparametric prior for topic evolution and combine it with the word co-occurrence modeling technique, which is simple and effective for short texts modeling. Then we propose a collapse Gibbs sampling method for parameter inference of our model. Finally, we conduct experiment on twitter dataset and prove the effectiveness of our model.

II. Proposed Method

In this section, we will introduce the dynamic topic modeling method we proposed. We first demonstrate the generation process of our model and then present the parameter inference method for the model.

A. Topic Evolution Model for Short Texts

To overcome the data sparsity problem of short texts, we adopt the technique of word co-occurrence modeling [20]. The basic idea is that we first transform the original short texts into biterms, which are unordered word-pairs co-occurring in the

978-1-5386-3120-1/17 $31.00 © 2017 IEEE
DOI 10.1109/ICBK.2017.39

same context. As each piece of short text only contains a small number of words, any two words within the same text count as a biterm. So a text containing n words can be transformed into C_n^2 biterms. As two words within one biterm come from the same context, they usually represent the same topic. So we can assume that words within one biterm belong to same topic. Then we build our model based on these biterms and model the word co-occurrence directly.

To model the topic evolution, we partition the biterms into different time epochs. We establish the connection of topic distribution between different time epochs with a temporal nonparametric Bayesian process and use it as topic evolution prior to determine the topic of each biterm. Then we generate the words within one biterm according to the word distribution of the corresponding topic. The topic number of different time epochs are determined automatically during this process.

We use the recurrent Chinese restaurant process (RCRP) as the nonparametric prior for topic evolution which satisfies the requirements of our model. The recurrent Chinese restaurant process is a temporal extension to Chinese restaurant process (CRP), one of the most widely used non-parametric technique for topic modeling. The CRP is an equivalent view of Dirichlet process. It works according to the following metaphor: suppose a sequence of customers entering a restaurant with infinite tables. When a new customer enter, she either sits at an occupied table with probability proportional to the number of previous customers sitting by the table, or at an unoccupied table with probability proportional to α. Besides, in CRP the probability of a particular seating configuration does not depend on 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 by table k and m denotes the total number of customers. Let K denotes the number of tables where $m_k > 0$ and $\bar{k}$ denotes new table. Let α be the parameter of CRP. The probability of customer i sitting at each table is represented as follows:

$$p(z_i = k \mid z_{-i}, \alpha) = \sum_{k=1}^{K} \frac{m_k}{m-1+\alpha} \delta(z_i, k) + \frac{\alpha}{m-1+\alpha} \delta(z_i, \bar{k}) \quad (1)$$

The CRP could partition data into a proper number of clusters automatically. RCRP further extends it by considering the time series information and could be used to model the topic evolution process. In the RCRP scenario, we partition the streaming data into different time epochs. For a data point in one time epoch, the cluster it belongs to is not only depend on the cluster distribution information in current time epoch, but also the information in history time epochs. This will help maintain the consistency of clusters in the evolution process. Furthermore, the history information in different time epochs has different influence to cluster chosen in current time epoch. The earlier the information is, the fewer influence it has.

So let $m_{k,t}$ be the number of customers in time epoch t sitting by table k and $m'_{k,t}$ be the prior weight of table k in time epoch t. Let λ denotes the decay factor and Δ denotes the width of time epochs we use. $m'_{k,t}$ can be defined as follows:

$$m'_{k,t} = \sum_{\delta=1}^{\Delta} e^{-\delta/\lambda} m_{k,t-\delta} \quad (2)$$

So in time epoch t, customer i sits at an occupied table k with probability proportional to $m_{k,t} + m'_{k,t}$. The recurrent Chinese restaurant process can be represented as follows:

$$\begin{aligned} p(z_{i,t} = k \mid z_{-i,t}, z_{t-\Delta:t-1}, \alpha, \lambda, \Delta) \\ \propto (m_{k,t} + m'_{k,t})\delta(z_i, k) + \alpha\delta(z_i, \bar{k}) \end{aligned} \quad (3)$$

in which $z_{-i,t}$ denotes the cluster assignment of all data in time epoch t except $z_{i,t}$ and $z_{t-\Delta:t-1}$ denotes the cluster assignment of all data from time epoch $t-\Delta$ to time epoch $t-1$.

In our approach, we use RCRP as the prior to determine the topic of each biterm. Then we generate the words within one biterm according to the word distribution of that topic. Let $b_{i,t}$ denotes the ith biterm in time epoch t. Let $w_{i,1,t}$ 和 $w_{i,2,t}$ denote the two words within biterm $b_{i,t}$. Let $\phi_{t,z_{i,t}}$ denotes the word distribution of topic $z_{i,t}$ in time epoch t. The generative process of biterm $b_{i,t}$ is as follows:

$$z_{i,t} \sim RCRP(z_{-i,t}, z_{t-\Delta:t-1}, \alpha, \lambda, \Delta) \quad (4)$$

$$w_{i,1,t}, w_{i,2,t} \sim Multinomial(\phi_{t,z_{i,t}}) \quad (5)$$

in which the word distribution $\phi_{t,z_{i,t}}$ is generated based on Dirichlet distribution as follows:

$$\phi_{t,z_{i,t}} \sim Dirichlet(\beta) \quad (6)$$

We apply the above generative process to each biterm in each time epoch. With the parameter inference method we proposed, we can get the word distribution of each topic in all time epochs.

B. Parameters Inference

We adopt the collapsed Gibbs sampling method to estimate the parameters of our model. Gibbs sampling is a widely used approximate inference method based on numerical sampling. The Gibbs sampling procedure replaces the value of one variable by sampling a value based on the distribution of that variable conditioned on the remaining variables. Then we repeat the procedure by cycling through the variables and estimating their values iteratively. The collapsed Gibbs sampling can reduce the variable number by taking advantage of the conjugate distribution and integrating out some variables.

In our model, the word distribution of each topic is multinomial distribution and its prior is Dirichlet distribution. The two distributions are conjugate. So we can integrate out $\phi_{t,z_{i,t}}$ and only sample the topic $z_{i,t}$ for each biterm.

We first initialize the topic of each biterm randomly. According to the Bayes formula, $z_{i,t}$ can be represented as follows:

$$\begin{aligned} P(z_{i,t} \mid b_{i,t}, b_{-i,t}, z_{-i,t}, z_{t-\Delta:t-1}, \alpha, \beta, \lambda, \Delta) \propto \\ P(b_{i,t} \mid b_{-i,t}, z_{i,t}, z_{-i,t}, \beta) \times P(z_{i,t} \mid z_{-i,t}, z_{t-\Delta:t-1}, \alpha, \lambda, \Delta) \end{aligned} \quad (7)$$

In the above formula, $P(z_{i,t} \mid z_{-i,t}, z_{t-\Delta:t-1}, \alpha, \lambda, \Delta)$ is represented as formula (3), while $P(b_{i,t} \mid b_{-i,t}, z_{i,t}, z_{-i,t}, \beta)$ can be calculated as follows:

$$P(b_{i,t} \mid b_{-i,t}, z_{i,t}, z_{-i,t}, \beta) = \frac{(n_{w_{i,1,t}|z_{i,t}} + \beta)(n_{w_{i,2,t}|z_{i,t}} + \beta)}{(\sum_w n_{w|z_{i,t}} + W\beta)^2} \quad (8)$$

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

Based on formula (3), (7) and (8), we can sample the topic $z_{i,t}$ for biterm $b_{i,t}$ as follows:

$$P(z_{i,t} = k \mid b_{i,t}, b_{-i,t}, z_{-i,t}, z_{t-\Delta:t-1}, \alpha, \beta, \lambda, \Delta) \propto \begin{cases} (m_{k,t} + m'_{k,t}) \dfrac{(n_{w_{i,1,t}|z_{i,t}} + \beta)(n_{w_{i,2,t}|z_{i,t}} + \beta)}{(\sum_w n_{w|z_{i,t}} + W\beta)^2} & (\textit{existing topic}) \\ \\ \alpha \dfrac{(n_{w_{i,1,t}|z_{i,t}} + \beta)(n_{w_{i,2,t}|z_{i,t}} + \beta)}{(\sum_w n_{w|z_{i,t}} + W\beta)^2} & (\textit{new topic}) \end{cases} \quad (9)$$

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

$$\phi_{t,w|z_{i,t}} = \frac{n_{w|z_{i,t}} + \beta}{\sum_w n_{w|z_{i,t}} + W\beta} \quad (10)$$

The parameter inference algorithm of our model is shown in Table I.

TABLE I. PARAMETER INFERENCE ALGORITHM

Input: Biterm set B_t for each time epoch, $t \in [1, T]$

Hyperparamters $\alpha, \beta, \lambda, \Delta$

Output: Word distribution $\phi_{t,k}$ of each topic in each time epoch

Algorithm:

For time epoch t=1 to T

  Initialize the topic of each biterm in B_t randomly

  For *iteration*=1 to N_{iter}

    For $b_{i,t}$ in B_t

      Sample topic $z_{i,t}$ according to formula (9)

      Update parameter $n_{w_{i,1,t}|z_{i,t}}$ and $n_{w_{i,2,t}|z_{i,t}}$

    End-For

  End-For

End-For

Calculate $\phi_{t,k}$ according to formula (10)

III. EXPERIMENTS

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

A. Experimental Setup

Dataset: Twitter is one of the most popular social media site in the world. As tweets are short and can reflect the topics people interest in and opinions people hold, we use twitter dataset to evaluate the topic evolution models. We sample tweets from the twitter7 dataset [22] to construct our experimental dataset. The twitter7 dataset is a large dataset which contains 467 million tweets collected from June 1, 2009 to December 31, 2009. We use the tweets in July and only remain English tweets for our expriment. We convert the letters of all tweets into lower case. After that we remove the stop words and restrict the vocabulary to 10000 words based on the word frequency. We keep the tweets with more than one word and sample 10000 tweets each day for three weeks. We use these sampled tweets as training dataset and use the remaining tweets as reference dataset for the topic coherence evaluation task. In our experiment, we split the training data into days for topic evolution evaluation.

Baseline Methods: We compare our model with two representative methods on twitter dataset, which are DTM and Online LDA. In our model, we set the parameter as α=10, β=0.1, λ=0.5 and Δ=1. For DTM, we use the default setting. For Online LDA, we set the parameters as α=50/K and β=0.1, in which K is the topic number. Both DTM and Online LDA could not determine the topic number automatically. To see the performance change accompanied by the topic number, we choose three topic numbers 60, 90 and 120 respectively for evaluation.

B. Evaluation of Topic Quality

Topic models are usually evaluated by measures like perplexity or marginal likelihood on held-out test. But recent studies [23, 24] show these measures could not reflect the interpretability of topics very well. So to evaluate the topic quality, we use topic coherence to reflect the interpretability of topics and topic distinctiveness to reflect the distinctiveness between topics.

As our model is a non-parametric method, after the sampling process, small clusters containing only a few biterms may exist. These clusters contain little information and are usually noises. So in our experiment, we omit the clusters with less than 30 biterms.

Topic Coherence

Recent studies [25, 26] show topic coherence could reflect the topic interpretability very well. And it is more consistent to human judgements compared to traditional measures like perplexity. In our experiment, we use LCP [25] as the topic coherence measure.

TABLE II. TOPIC COHERENCE RESULTS OF TOPIC EVOLUTION MODELS

Method		DTM				Online LDA				Our Model
Topic Number		60	90	120	Average	60	90	120	Average	–
LCP	*L*=5	-79.83	-76.42	-82.14	-79.46	-76.27	-74.05	-78.36	-76.23	**-69.24**
	L=10	-381.13	-373.84	-385.26	-380.08	-374.81	-368.19	-379.73	-374.24	**-349.36**
	L=15	-914.26	-898.64	-923.71	-912.20	-901.35	-887.92	-910.27	-899.85	**-854.71**

The LCP can be calculated as follows:

$$LCP(t) = \sum_{j=2}^{L} \sum_{i=1}^{j-1} \log \frac{P(w_i, w_j)}{P(w_i)} \tag{11}$$

The word probabilities and the co-occurrence probabilities are computed on the large-scale external dataset referred as reference dataset. Higher LCP score indicates better topic quality. For each topic, we use the top *L* words to calculate the LCP score. Here we set *L* to 5, 10 and 15 respectively. We use the average LCP score of all topics during whole period as the model evaluation metric.

Table II shows the model evaluation scores based on LCP. We can see the performance of DTM and Online LDA varies with the predefined topic number. As it is difficult to set the topic number of a new corpora accurately, it is hard to achieve best performance of these models. Since DTM and Online LDA are not designed specifically for short texts, our model performs best among these models.

Topic Distinctiveness

To measure the distinctiveness of the generated topics, we define the corresponding metric $D(P\|Q)$ based on the Kullback-Leibler divergence (KLD) since it could measure the divergence of the topic distributions. Let *P* and *Q* denote two topic distributions. $D_{KL}(P\|Q)$ denotes the Kullback-Leibler divergence between *P* and *Q*. The $D_{KL}(P\|Q)$ can be calculated as follows:

$$D_{KL}(P \| Q) = -\int P(w) \ln(\frac{P(w)}{Q(w)}) dw \tag{12}$$

As the Kullback-Leibler divergence is non-symmetric, we define $D(P\|Q)$ as follows:

$$D(P \| Q) = \frac{1}{2}[(D_{KL}(P \| Q) + D_{KL}(Q \| P)] \tag{13}$$

We first calculate the topic distinctiveness scores between any pair of topics in one time epoch. We get the average score in one time epoch and then calculate the average score of the whole time period. We use this score as the topic distinctiveness metric for topic models.

Table III shows the topic distinctiveness results. We can see that the topic distinctiveness scores of DTM and Online LDA vary with the predefined topic numbers, which illustrates the predefined topic number could affect the topic distinctiveness of topic evolution models. Our model outperforms DTM and Online LDA, which means the topics generated by our model are more distinctive than baseline methods.

TABLE III. TOPIC DISTINCTIVENESS RESULTS OF TOPIC EVOLUTION MODELS

Method	Topic Number	Topic Distinctiveness
DTM	60	3.91
	90	3.61
	120	3.38
	Average	3.63
Online LDA	60	4.02
	90	3.68
	120	3.76
	Average	3.82
Our Model	—	**5.39**

Topic Illustration

We illustrate the topic about food generated by our method as well as DTM (K=90) and Online LDA (K=90) which perform well in our expriment. We show the top words of this topic in Table IV. The words not so relevent to this topic are marked bold. We can see the words our model generated are more relevent to this topic compared to baseline methods.

TABLE IV. Food TOPIC Illustration

Method	Top Words
Our Model	cream ice cake salad bread peanut
DTM (*K*=90)	cream ice chocolate **yum** bread **dinner**
Online LDA (*K*=90)	ice cream salad chocolate chicken **great**

We also show the top words of the topic about IT companies and social media sites in three time epochs, which is an example of the topic evolution. We can see the word distribution evolves over time in the topic evolution process.

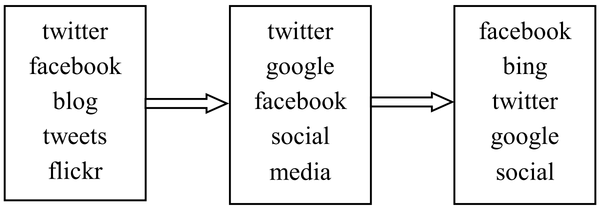

Fig 1. Topic Evolution Example

IV. Conclusions

As short texts are widespread in the Web, discovering and tracking topics of these contents become a very important task, which could not only help people grasp essential information but also facilitate many applications like social media mining, information retrieval and so on. Most existing topic evolution models either suffer from the data sparsity problem or need to specify the topic number manually. So in this paper, we propose a nonparametric topic evolution model for short texts by combining the recurrent Chinese restaurant process with word co-occurrence modeling technique. We carry out empirical studies on twitter dataset. The experimental results prove the effectiveness of our method.

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] R. Yan, X. Wan, J. Otterbacher, L. Kong, X. Li, and Y. Zhang, “Evolutionary Timeline Summarization: A Balanced Optimization Framework via Iterative Substitution,” in *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, 2011, pp. 745–754.

[2] O. Alonso and K. Shiells, “Timelines as Summaries of Popular Scheduled Events,” in *Proceedings of the 22Nd International Conference on World Wide Web Companion*, Republic and Canton of Geneva, Switzerland, 2013, pp. 1037–1044.

[3] L. Shou, Z. Wang, K. Chen, and G. Chen, “Sumblr: Continuous Summarization of Evolving Tweet Streams,” in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, 2013, pp. 533–542.

[4] B. Hu and M. Ester, “Spatial Topic Modeling in Online Social Media for Location Recommendation,” in *Proceedings of the 7th ACM Conference on Recommender Systems*, New York, NY, USA, 2013, pp. 25–32.

[5] X. Liu, “Modeling Users’ Dynamic Preference for Personalized Recommendation,” in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[6] L. Gao, J. Wu, C. Zhou, and Y. Hu, “Collaborative Dynamic Sparse Topic Regression with User Profile Evolution for Item Recommendation,” in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[7] C. X. Lin, Q. Mei, J. Han, Y. Jiang, and M. Danilevsky, “The Joint Inference of Topic Diffusion and Evolution in Social Communities,” in *2011 IEEE 11th International Conference on Data Mining*, 2011, pp. 378–387.

[8] C. Xide, L. Qiaozhu, M. Yunliang, J. Jiawei, and H. S. Qi, “Inferring the Diffusion and Evolution of Topics in Social Communities,” in *Proc. of 2011 ACM SIGKDD Workshop on Social Network Mining and Analysis*, 2011.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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*, New York, NY, USA, 2010, pp. 1079–1088.

[16] W. X. Zhao *et al.*, “Comparing Twitter and Traditional Media Using Topic Models,” in *Advances in Information Retrieval*, P. Clough, C. Foley, C. Gurrin, G. J. F. Jones, W. Kraaij, H. Lee, and V. Mudoch, Eds. Springer Berlin Heidelberg, 2011, pp. 338–349.

[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] X. Yan, J. Guo, Y. Lan, and X. Cheng, “A Biterm Topic Model for Short Texts,” in *Proceedings of the 22Nd International Conference on World Wide Web*, Republic and Canton of Geneva, Switzerland, 2013, pp. 1445–1456.

[20] 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.

[21] 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.

[22] J. Yang and J. Leskovec, “Patterns of temporal variation in online media,” in *Proceedings of the fourth ACM international conference on Web search and data mining*, New York, NY, USA, 2011, pp. 177–186.

[23] J. Chang, S. Gerrish, C. Wang, J. L. Boyd-graber, and D. M. Blei, “Reading Tea Leaves: How Humans Interpret Topic Models,” in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, Eds. Curran Associates, Inc., 2009, pp. 288–296.

[24] D. M. Blei, “Probabilistic topic models,” *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012.

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

[26] J. H. Lau, T. Baldwin, and D. Newman, “On Collocations and Topic Models,” *ACM Trans. Speech Lang. Process.*, vol. 10, no. 3, p. 10:1–10:14, 2013.