Research Article

Feng Jian*, Wang Yajiao, and Ding Yuanyuan

Microblog topic evolution computing based on LDA algorithm

https://doi.org/10.1515/phys-2018-0067
Received Apr 20, 2018; accepted May 29, 2018

Abstract: Research on topic evolution of Microblog is an effective way to analyze network public opinions. This paper proposes a method for mining changing of Microblog topics with time, and realizes topic evolution through topic extraction and topic relevance calculation. Firstly, latent Dirichlet allocation (LDA) model is used to automatically extract topics from different time slices; secondly, a similarity calculation algorithm is designed to calculate relevance of topic content through normalization of similarities among characteristic words and co-occurrence relations, to get evolutionary relationship among sub-topics of different time slices; thirdly, using probability distribution of blog article-topic to calculate topic intensity in each time slice, and then gets evolutionary relationship of topic intensity over time. Experiments show that the proposed topic evolution analysis model can effectively detect the evolution of topic content and intensity of real blogs.

Keywords: LDA, topic evolution, semantic similarity, relevance, co-occurrence

PACS: 07.05.Kf, 89.20.Hh

1 Introduction

Due to the characteristics of timeliness, flexibility, integration and grassroots, Microblog became the main source and an important distribution center of public opinions. In order to maintain healthy development of society, public opinions formed by Microblog should be monitored.

Network public opinions are sublimation of Microblog topics, and their evolution often depends on content and intensity changes. Topic content evolution refers to the changes of topic content over time, and topic intensity evolution shows the changes of attentiveness [1]. How to effectively track the development of Microblog topics becomes the key to evolution analysis of network public opinions.

Sina Microblog and Tencent Microblog are main research objects in China, and Twitter is the counterpart out of China. The traditional research method to describe topics is usually based on vector space model (VSM). In recent years, latent Dirichlet allocation (LDA) [2] has become the mainstream model of research because it can refine topics from network events more accurately, and was introduced to the study of topic evolution. However, due to the characteristics of dynamic interaction, inheritance and continuity of Microblog, existed researches did not fully consider semantic information of the corpus, resulting in insufficient tracking accuracy. Therefore, the field has still a great exploration space, requiring innovative information mining methods to enhance accuracy of topic mining and descript topic evolution dynamically. This paper studies the problem of topic evolution for Sina Microblog based on LDA, establishes topic evolution analysis model, and analyzes evolution of topic content and intensity with time.

The content of the paper is organized as follows. LDA model and its related researches in topic evolution are introduced in Section 2. In Section 3, we analyze the relevant characteristics of topic content and intensity evolution, and give corresponding calculation methods. New topic evolution model is descripted in Section 4. Section 5 presents experiments we have performed using the proposed model . Final conclusions and suggestions for future work are discussed in Section 6.

2 Related works

2.1 LDA model

LDA is a typical statistical topic model. Its basic idea is assuming implicit semantic structure of a document con-

***Corresponding Author: Feng Jian:** College of Computer Science & Technology, Xi'an University of Science and Technology, Xi'an, 710054, China; Email: fengjian@xust.edu.cn
Wang Yajiao: College of Computer Science & Technology, Xi'an University of Science and Technology, Xi'an, 710054, China; Email: wangyiyx0322@163.com
Ding Yuanyuan: Beijing Huade Hydraulic Industry Group, Beijing 100176, China; Email: yyding91@163.com

sisted by a set of interrelated topics, and topics are composed of a set of words; assuming that words are generated by the probability distribution of topics, each topic is represented by the words and their probability distribution on the topic; the document is a random finite mixing of probabilistic distribution of potential topics. Each document is sampled according to Dirichlet distribution to produce proportion of topics in the document, combined with the probability distribution of topic-words to generate every word, so that high dimensional lexical space of the document is reduced to low dimensional topic space to extract topics. Table 1 shows the main parameters used for LDA model generation.

Table 1: Notation correspondence

Parameter	Descriptions
D	*Total number of documents*
K	*Number of hidden topics*
N_d	*Number of words in the dth document*
$w_{d,n}$	*The nth word in document d*
$z_{d,n}$	*The topic associated with* $w_{d,n}$
θ_d	*The multinomial distribution of topics specific to the document d*
φ_z	*The multinomial distribution of words specific to the topic z*
α	*Hyperparameter for the multinomial* θ_d
β	*Hyperparameter for the multinomial* φ_z

The process of simulating topics generation in LDA is as follows:

(1) For each blog article $d \in D$, according to $\theta_d \sim \mathrm{Dir}(\alpha)$, gets multinomial distribution parameter θ_d of topics on document d;
(2) For each topic $z \in K$, according to $\varphi_k \sim \mathrm{Dir}(\beta)$, gets multinomial distribution parameter φ_k of words on topic z;
(3) For the nth word $w_{d,n}$ in document d:
   ① According to the multinomial distribution $z_{d,n} \sim \mathrm{Mult}(\theta_d)$, gets the topic $z_{d,n}$.
   ② According to the multinomial distribution $w_{d,n} \sim \mathrm{Mult}(\varphi_k)$, gets the word $w_{d,n}$.

In LDA, a topic is represented by a set of semantically related words and the probability that words appear on the topic. Namely, $z = \{(w_1|z), \cdots, (w_V, p(w_V|z))\}$, where $\mathrm{p}(w_V|z)$ indicates probability that word w_V appears where topic z has been observed.

2.2 Tracking topic evolution based on LDA

LDA model assumes that documents are exchangeable, that is, it ignores time information of documents. For revealing dynamics and development of topics over time, researchers introduced time information to LDA to extend the model [3–6].

There are three kinds of topic evolution methods based on LDA. The first kind considers the words in the document to be influenced by time, so it combines the time information of the document into LDA. The representative model is called topic over time (TOT) [3], which can use distribution of the topic at different time to get evolution of topic intensity, but cannot get evolution of topic content; the second kind first uses LDA to get topics in whole document set, and then checks the distribution of topics in discrete time to measure evolution. This kind of method also cannot get the evolution of topic content, and the evolution of topic intensity depends on time granularity; the third kind of method separates the document into different time slice in accordance with time information, and then deals with the collection of documents on each time slice in turn. It can simultaneously achieve the evolution of topic content and intensity. The typical model includes dynamic topic model (DTM) [4], which uses the state space to record the change of topic content and distribution intensity; continuous time dynamic topic model (CTDMT) [5] adopts Brownian motion model to model topic evolution in continuous time; multiscale topic tomography (MTT) [6] studies the topic evolution of multi-time granularity. The work of this paper belongs to the third kind of methods.

2.3 Topic similarity calculation

The relevance of reports in adjacent time periods should be determined when evolution of topic content is analyzed. If a new report is related to the priori report, it can be used to track the development of the transcendental report. If it is not relevant, the report can be judged as a new event. Relevance judgment is also called correlation detection, which is the original problem of topic tracking. In text mining methods, the similarity between two topics is generally calculated according to characteristic words in the text. According to difference of text representation models, the topic similarity calculation method can be summed up as Jaccard coefficient and cosine similarity based on word pocket model, KL (Kullback-Leibler) distance method [7] and words co-occurrence method based on language model.

The similarity algorithm used in conjunction with LDA is usually KL algorithm, that is, the similarity between two topics is seen as the distance of vector space formed by words of two topics, as follows:

$$KL(z_1, z_2) = D(P(w|z_1)\|P(w|z_2)) = \sum_{w \in W} P(w|z_1) log \frac{P(w|z_1)}{P(w|z_2)} \quad (1)$$

Where $D(P(w|z_1)\|P(w|z_2))$ is the distance of two vectors. The algorithm measures the difference of probability distribution of two sets of characteristic words in same semantic space. In order to improve the accuracy of this kind of algorithms, Chu proposed an improved similarity calculation method which combines classical cosine similarity algorithm and Jensen-Shannon divergence (JSD) [8], the latter is an improved KL algorithm [9], hereinafter referred to JSD_COS algorithm:

$$JSD(z_1, z_2) = \frac{1}{2}(D(z_1\|m) + D(z_2\|m)), \quad m = \frac{1}{2}(z_1 + z_2) \quad (2)$$

$$JSD\_COS(z_1, z_2) = \lambda COS(z_1, z_2) + (1 - \lambda) JSD(z_1, z_2) + \lambda \quad (3)$$

Where $COS(z_1, z_2)$ is cosine similarity, which represents cosine of the angle of two vectors composed by characteristic words from two topics, and λ is a adjustment coefficient.

Because LDA and its improved models have some shortcomings, such as high frequency words tends to be selected for topic distribution, the above methods will lead to a problem that semantics of generated topics is not clear, for lacking semantic analysis.

In the other hand, some scholars use co-occurrence relationship between characteristic words to express their semantic relations [10], so as to calculate relevance between words and topics. The basic idea is if a few words often appear together in the same text, then they express the semantic information of the text to a certain extent. For example, in a news report, if "Haiti" and "earthquake" appear together repeatedly, it can be inferred that this is a report on the earthquake in Haiti. The existing co-occurrence analysis model is mainly used to extend VSM, and with no research on extending LDA.

3 Analyses of topic evolution

In this paper, we use the following strategies to study the evolution of Microblog topics: according to the release time of blog articles, we will scatter the blogs into different time slices, and then use LDA to extract topics in each time slice. The evolution of topics is mainly based on changes of topic similarity and intensity in different time slice.

3.1 Topic content evolution

The evolution of topic content shows the difference of characteristic words sequence in different time slices. This difference is mainly manifested in semantic relevance. From the analysis of Section 2.3, it can be seen that the topic relevance calculation based on LDA often based on characteristic words matching without considering semantic relevance of the topic, and the co-occurrence relationship of characteristic words can express semantic relations among words, but considering only the co-occurrence relation between the characteristic words is not enough. Therefore, this paper measures the relevance of topic content in different time slices from two aspects: the proportion of the same characteristic words included in the two topics and co-occurrence frequency of the characteristic words. The former is calculated by Jaccard coefficient, while the latter is measured by co-occurrence probability of characteristic words.

Suppose that z_1 and z_2 are two topics of the adjacent time slices respectively, and C and D are the characteristic word sets of z_1 and z_2, respectively.

(1) The proportion of the same characteristic words

The proportion of the same characteristic words in z_1 and z_2 is calculated by Jaccard coefficient, that is, the matching degree of the characteristic words. The Equation 4 is:

$$Jaccard(z_1, z_2) = \frac{|C \bigcap D|}{|C \bigcup D|} \quad (4)$$

(2) The co-occurrence frequency

The term co-occurrence here refers to the situation in which two different characteristic words appear simultaneously in topics of two different time slices. The higher the co-occurrence rate among words, the more likely semantics of topics may be similar, and the more likely the topics may be related. The co-occurrence frequency of characteristic words is the synchronized appearance frequency of the characteristic word pairs in z_1 and z_2:

$$WC(z_1, z_2) = \sum_i \sum_j \frac{\|Segment(w_{z_{1i}}, w_{z_{2j}})\|}{\|Segment\|} \quad (5)$$

Where $\|$Segment$\|$ represents the total number of blog articles in two adjacent time slices where z_1 and z_2 are located; and $\|$Segment$(w_{z_{1i}}, w_{z_{2j}})\|$ is the number of blogs including the characteristic words $w_{z_{1i}}$ and $w_{z_{2j}}$ simultaneously in both time slices. The algorithm hereinafter is referred to as WC algorithm.

(3) The relevance degree of z_1 and z_2

Based on the above calculations, the method to calculate the relevance degree of z_1 and z_2 are given, namely JW algorithm:

$$JW(z_1, z_2) = \gamma Jaccard(z_1, z_2) + (1-\gamma)WC(z_1, z_2) \quad (6)$$

Where Jaccard(z_1, z_2) calculates the similarity of the characteristic words, and WC(z_1, z_2) reinforces its semantic similarity. γ is the weighting coefficient, which reflects the contribution of two different similarities to the overall similarity.

In order to determine the relevance of two topics by value of JW, experience threshold is needed. If JW value is greater than the threshold, it is considered there is a correlation between the two topics on the adjacent time slices, so content evolution happens.

3.2 Topic intensity evolution

Generally, the more the number of blogs to discuss a topic, the higher the hotness of the topic is. In each time slice, the average value of the topic distribution probability on each blog is calculated to determine the mean hotness of a topic. The Equation 7 is as follows:

$$\overline{\theta_z^t} = \frac{\sum\limits_{d \in D_t} \theta_{dz}^t}{D_t} \quad (7)$$

Where t is a time slice, z is a topic, D_t is total number of blogs in time slice t, and θ_{dz}^t is the probability of blog d belonging to z in t. The topic intensity is represented by the average of θ, namely the proportion of the topic in the time slice, so the evolution of the topic intensity can be obtained by combining the topic intensity of all time slices.

4 Topic evolution system architecture

Microblog is a text stream with a timing relationship. On the basis of the time slice segmentation, this paper uses LDA to extract topics on each time slice, and uses the topic content evolution method and the topic intensity evolution method proposed in 3.1 and 3.2, respectively, to achieve tracking and evolution of topic content and intensity. System architecture is shown in Figure 1, and the specific steps are as follows:

(1) Gets and pre-processes blog sets, including eliminating duplicates, making word segmentation and removing stop words.
(2) According to time information on blogs, divides the blog sets into different subsets.
(3) Models blogs in each time slice by LDA, extracts subtopics of each time slice, and gets the probability distributions of the topic-word and document-topic, so as to extract topic collection from each time slice.
The specific process is: using Gibbs sampling algorithm to estimate the posterior distribution θ_d and φ_z of LDA [10]. Takes parameters of Dirichlet prior distribution $\alpha = 50/\mathrm{K}$, $\beta = 0.1$, iterates 1000 times to obtain document-topic distribution matrix and topic-word distribution matrix.
Characteristic words selection: for each blog, according to the probability distribution of words, extracts multiple words with high probability as characteristic words (through artificial excavation of the experimental corpus, when number of characteristic words is bigger than 8, the best specificity and coverage of the topic should be obtained. Therefore, in order not to lose generality, we extract the first 10 words as the characteristic word for each sub-topic).
(4) Analyzes and calculates the similarity among topics in each adjacent time slice according to Equation 6, so as to analyze the detailed evolution process of topic content in the whole time interval.
(5) According to Equation 7 to calculate the hotness of each topic under different time slices, and then gets the topic intensity evolution process.

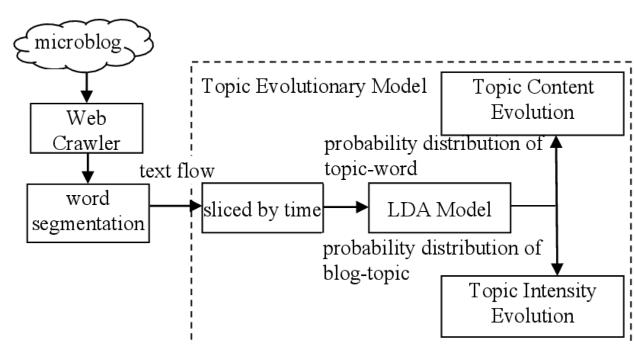

Figure 1: Topic evolution system architecture

5 Experimental analyses

In this paper, two sets of experiments were designed. The first set of experiments uses KL algorithm, JSD_COS algorithm, WC algorithm and JW algorithm, respectively, to calculate relevance of the topic, aiming at verifying the feasibility of JW algorithm; the second set of experiments evaluates the effect of topic content and intensity evolution, to verify the effectiveness of the overall scheme.

Experiments had been done on PC with Windows XP system, 2.93GHz CPU and 2GB memory. In experiments, Niuparser system developed by Northeastern University in China was used to carry on Chinese word segmentation.

5.1 Experimental data sets

There are no widely accepted corpus and annotation results for the study of topic evolution, so we chose Microblog data sets from an open source community aiming to do the social network analysis, namely Social Analysis [11]. These experimental blogs have been manually marked by the website. Each topic includes a large number of blogs and each blog clearly belongs to one topic. We used two data sets from August 2012 to November 2012: M-blog1 and M-blog2. The basic information is shown in Table 2.

The data sets M-blog1 and M-blog2 are divided into different time slices by month, and the number of blogs corresponding to each time slice is shown in Table 3.

Table 2: Basic information of data sets

Data sets	Number of blogs	Number of topics
$M-blog1$	88480	6
$M-blog2$	217308	29

Table 3: Distribution of blogs in time slices

Time slice	2012/08	2012/09	2012/10	2012/11
$Number\ of\ blogs\ (M{-}blog1)$	17991	33147	20315	17027
$Number\ of\ blogs\ (M{-}blog2)$	45106	62735	59268	50199

5.2 Experiment 1 – JW algorithm evaluation

5.2.1 Evaluation indexes

The Precision(P), Recall(R) and F1-Measure(F1) are used in the evaluation of the experiment. F1 is the comprehensive evaluation of the first two.

$$P = \frac{A}{A+B} \tag{8}$$

$$R = \frac{A}{A+C} \tag{9}$$

$$F1 = \frac{2 \times P \times R}{P+R} \tag{10}$$

Where A represents the extracted contents related to the topic; B represents the extracted contents that are not relevant to the topic; and C represents the contents related to the topic that has not been extracted. The number of blogs related to the topic in all blogs is A+C, and the number determined to be related with the topic is A+B.

5.2.2 Experimental results and analysis

In order to verify the effect of JW algorithm, both 4 algorithms including KL, JSD_COS, WC and JW were used to calculate similarities among topics on M-blog1 and M-blog2, and then R, P, and F1 were calculated according to artificial tagging. After comparing the experimental results, the value of γ in JW algorithm was taken by 0.6.

Figure 2-4 show that the R, P and F1 of JW in topic similarity calculation on different time slices on M-blog1 are better than KL, JSD_COS and WC. In order to further verify the stability of JW algorithm, to avoid the possible coincidence of the experimental results, the same experiment is

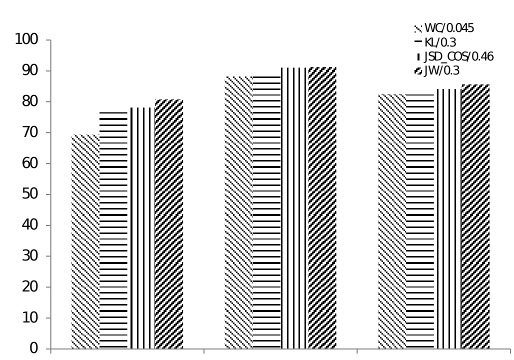

Figure 2: Results of relevance between August and September in M-blog1

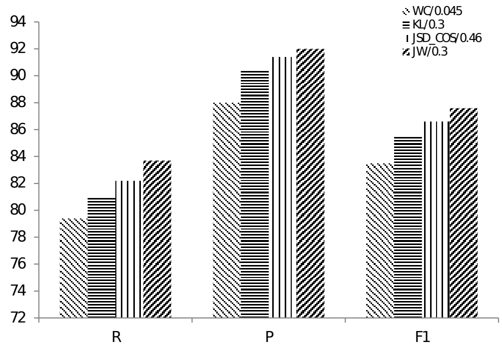

Figure 3: Results of relevance between September and October in M-blog1

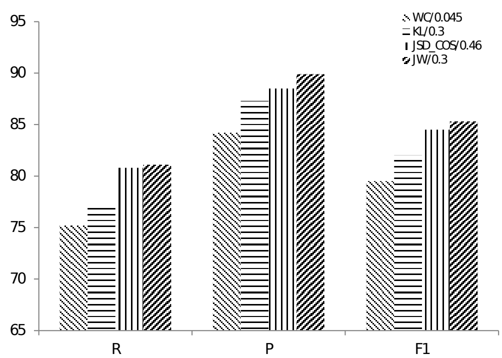

Figure 4: Results of relevance between October and November in M-blog1

carried out on M-blog2. The results are consistent and are not listed for length relation. The results show that JW algorithm has the best result for the relevance calculation and can track the topic content evolution better.

5.3 Experiment 2 – Topic evolution evaluations

5.3.1 Evaluation standard

There is no uniform criterion for evolution of topic content and intensity, and no effective method for quantifying the comparison. Generally, it is necessary to measure trend and intensity of real events by contrast with the artificial summary or the artificial description.

5.3.2 Experimental results and analysis

In order to verify whether the topic evolution method proposed in Section 4 can track the development of real events, measurement and judgment of multiple topics in two data sets are carried out. Due to the space relationship, only the "Liu Xiang" topic in M-blog2 is taken as an example, and its content and intensity changes in the time series are tracked and compared with the real event development process.

(1) Topic content evolution analysis

According to LDA, the sub-topics with relevance more than 0.3 and its corresponding characteristic words of Liu Xiang topic in different time slices are calculated and listed in Table 4.

Similarities more than 0.3 among sub-topics are related to Liu Xiang showed in Figure 5, calculated by JW algorithm. Among them, "Liu Xiang earned high advertising fees" sub-topic in September and "Cybernaut questioned Liu Xiang" sub-topic in August have slightly higher similarity than others; this is because the two sub-topics have more common characteristic words and co-occurrence words.

From the Figure 5, topic content evolution process of Liu Xiang is summarized as follows:

In reality, content changes of Liu Xiang event in the time series are: in August, Liu Xiang fell in London Olympic Games, but he still adhered to hop to end point, finally has been questioned and accused by cybernaut, and he was in the whirlpool of public opinion. In September, Liu Xiang received a successful surgery, but cybernaut still discussed huge amount of advertising earned by him in last 8 years, and expressed their dissatisfaction. In October, there were nosy parkers spreading rumors about disagreement between Liu Xiang and his coach Sun Haiping, and Sun Haiping came forward against the rumor.

Table 4: Sub-topics of Liu Xiang in different time slices

Sub-topics	Time slice	Characteristic words
Liu Xiang wrestled	2012/08	Liu Xiang, results, London, Olympic Games, foot injury, preliminary contest, audience, stumbled, retire, hurdling
Cybernaut questioned		Olympic Games, Liu Xiang, London, hurdling, deliberate, acting, wrestling, earn money, diving, public opinion
Liu hopped to end point		Liu Xiang, London, condition of an injury, fall, end point, Olympic Games, regret, hop, kiss, retire
Zhao Benshan donated couplet		Liu Xiang, horizontal scroll, first scroll, Zhao Benshan, second scroll, swindle, cheat, Paralympic Games, rewards, rich man
Foot surgery success	2012/09	Liu Xiang, surgery, achilles tendon, success, hospital, Wellington, doctor, operating room, sober, narcosis
Liu earned high advertising fees		Liu Xiang, Forbes, celebrity, endorse, money, Chinese Yuan, advertisement, China, track and field, Olympic Games
Contradiction between Liu Xiang and Sun Haiping	2012/10	Liu Xiang, coach, Olympic Games, Sun Haiping, disagreement, increase, old wounds, intensity, training, recrudesce
Sun Haiping refuted the rumor		Liu Xiang, Sun Haiping, recover, training, relationship, National Games, contradiction, treat injury, next year, display talents
Nike applied for Liu Xiang trademark	2012/11	Liu Xiang, trademark, Nike, agent, register, popularity, right of personal name, commercialization, preemptive registration, reject

In November, a commercial dispute emerged about Nike wanting to use Liu Xiang as a brand. It can be seen from Figure 5 that the content change tracked in this paper is the real trajectory of Liu Xiang event development.

(2) Topic intensity evolution analysis

We calculate topic intensity of each time slice according to Equation 3. As shown in Figure 6, in August, Liu Xiang topic received wide attention, and topic intensity is very high; with the time of migration, the intensity gradually decreased. The actual situation is, in August 2012, event of Liu Xiang falling in the London Olympic Games caused uproar, cybernaut started a heated discussion about injury and hidden situations behind wrestling of Liu Xiang; in September, event of Liu Xiang falling continued to ferment, more were concerned about the subsequent recovery, retirement and other issues; until November, discuss about the event slowly subsided. The intensity curve of Figure 6 conforms to the actual hotness change of Liu Xiang topic.

We also carried out analysis on content and intensity evolution of "Diaoyu Islands", "Yan' an car accident" topic in M-blog1 and "Yang Ming Tan Bridge collapsed" topic in M-blog2, and their analysis results are all consistent with the development process.

The experimental results show that the JW algorithm has high accuracy and can effectively describe the relevance of the topic. The proposed topic evolution model can track the change of topic content and intensity.

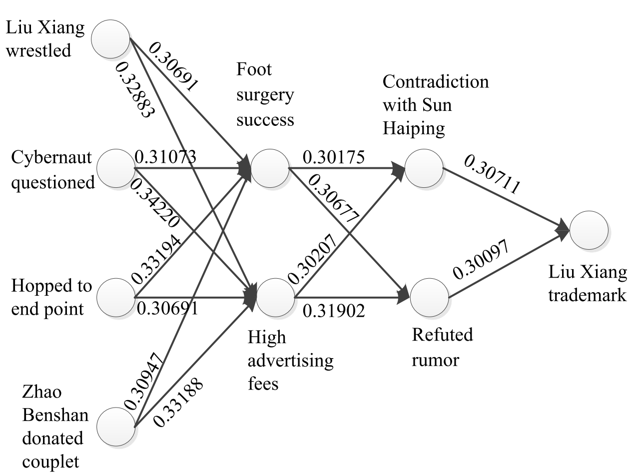

Figure 5: Content evolution and Similarities of topic Liu Xiang

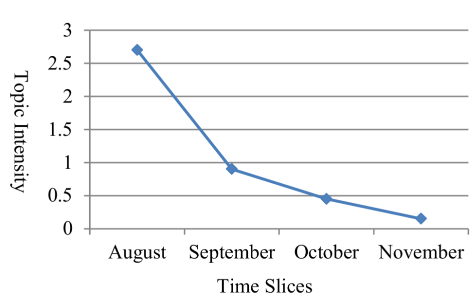

Figure 6: Topic intensity evolution of Liu Xiang

6 Conclusions

This paper describes a method of Microblog topic evolution based on LDA. The main contributions are as follows: ① A topic similarity calculation method based on Jaccard coefficient and co-occurrence frequency is proposed; ② A topic evolution model is presented through topic extraction from different time slices, and topic relevance and topic intensity calculation. Applying the proposed scheme to actual Microblog data sets shows that the evolution of topic content and intensity are in good agreement with the development of real events.

The paper only considered content and time information of blog articles, but did not consider author, reply, reference and other attribute information of blogs. How to integrate these into evolution of topics, that is making the evolution more directional and synergitic is our next direction of efforts. At the same time, evaluation for Microblog topic evolution has no unified evaluation criteria and no corresponding testing corpus. Whether it is topic content or topic intensity evolution, these can only be judged by people's subjective understanding of the topic. This is not comparable. Therefore, proposing criteria for evaluation is also an important problem to be solved.

Acknowledgement: This work is supported by Natural Science Foundation from Department of Education in Shaanxi Province (No. 15JK1468) and Shaanxi Provincial Natural Science Foundation Project (No. 2017JQ6053).

References

[1] Qin Y., Luo Y.Y., Zhao Y.Q., Zhang J., Research on relationship between tourism income and economic growth based on meta-analysis, Appl. Math. Nonl. Sci., 2018, 3, 105-114.

[2] Blei D.M., Ng A.Y., Jordan M.I., Latent Dirichlet Allocation, J. Mach. Learn. Res., 2003, 3, 993-1022.

[3] Wang X., Mccallum A., Topics over time: a non-Markov continuous-time model of topical trends, In: T. Eliassi-Rad(Ed.), Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(20-23 August 2006, Philadelphia, PA, USA), ACM, 2006, 424-433.

[4] Blei D.M., Lafferty J.D., Dynamic topic models, In: W. Cohen (Ed.), Proceedings of the 23rd International Conference on Machine Learning(25-29 June 2006, Pittsburgh, Pennsylvania, USA), ACM, 2006, 113-120.
[5] Wang C., Blei D., Heckerman D., Continuous Time Dynamic Topic Models, In: D. McAllester(Ed.), Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence(9-12 July 2008, Helsinki, Finland), AUAI Press Arlington, Virginia, United States, 2008, 579-586.
[6] Nallapati R.M., Cohen W., Ditmore S., Lafferty J., Ung K., Multiscale topic tomography, In: P. Berkhin(Ed.), Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(12-15 August 2007, San Jose, CA, USA), ACM, 2007, 520-529.
[7] Delpha C., Diallo D., Youssef A., Kullback-Leibler Divergence for fault estimation and isolation: Application to Gamma distributed data, Mech. Syst. Sign. Proc., 2017, 93, 118-135.
[8] Chu K.M., Li F., Fast algorithm for random walk centerlity, Comp. Appl. Softw., 2011, 28, 4-7.
[9] Wu L.F., Wang D., Zhang X Z, Liu S, Zhang L, Chen C.W., MLLDA: Multi-level LDA for modelling users on content curation social networks, Neurocomputing, 2017, 236, 73-81.
[10] Li Y.G., Zhou X.G., Sun Y., Zhang H.G., Design and Implementation of Weibo Sentiment Analysis Based on LDA and Dependency Parsing, Chin. Comm., 2016, 13, 91-105.
[11] Social Analysis. http://www.socialsis.org/data/dataset/dataset, 2016-03-15.