2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
Banff Center, Banff, Canada, October 5-8, 2017

Mining Phase Evolution for Hot Topics: A Case Study from Multiple Social Media Platforms

Ruoran Liu[12], Qiudan Li[1], Can Wang[12], Lei Wang[1],
Daniel Dajun Zeng[123]
[1]The State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing 100190, China
[2]University of Chinese Academy of Sciences, Beijing, China
[3] University of Arizona
Tucson, Arizona, USA
{liuruoran2016,qiudan.li,wangcan2015,l.wang}@ia.ac.cn, zeng@eller.arizona.edu

Hongyuan Ma[4]
[4]CNCERT/CC, Beijing, China
mahongyuan@foxmail.com

***Abstract*—Monitoring the evolution phases of real-time event including occurrence, development, climax, decline and ending is crucial for management department to intuitively and comprehensively understand the event and then make better decisions. However, there have been very few studies on performing phase evolution analysis of event using the number of posts at the specific time unit. The challenge of this problem is how to identify temporal pattern and mine topic of different phases automatically. In this paper, we propose a unified phase evolution mining model, it firstly identifies the temporal patterns of phases based on k-means and empirical rules, then, burst detection algorithm is adopted to discover peak interval of all phases, finally, we use a summarization technique TextRank to extract keywords from contents to summarize the topics in each phase. In addition, we perform experiments on two real-world datasets collected from different social media platform to understand the event evolution in a more comprehensive way. Experimental results show the characteristics of event evolution on different social media platforms and verify the efficacy of the proposed model.**

Keywords—phase evolution; k-means; burst detection; textrank; social media

I. INTRODUCTION

With the rapid development of Web 2.0, social media such as Weibo, Twitter, News has become an increasingly popular information source for up-to-date topics about what is happening in the world. According to the report published by Weibo, a popular social media in China, as of September, 2016, it had 132 million daily active users, indicating a rise of 32% compared with 2015[1]. The massive information related to a hot topic results in cognitive load on management department to keep track of an event. Monitoring the evolution phases of a real-time event including occurrence, development, climax, decline and ending is crucial for management department to intuitively and comprehensively understand the event and then make better decisions. Therefore, it's urgent to put forward an event phase evolution mining model to identify event phases and analyze corresponding topics automatically.

However, most of the existing research mainly focus on topic evolution analysis of event. [1][2]used unsupervised clustering techniques to generate storylines from unorganized documents. [3]proposed a Bayesian model to generate storylines from Twitter data. [4]utilized user input queries to sketch the real-time storyline of the event. Few of these works utilized the timestamps of documents but only paid attention to textual analysis. In general, the number of documents published at the specified time interval related to an event reflects the degree of people's attention to this event. Therefore, it's necessary to take the temporal information into consideration so that we can sketch the evolution phases of an event by recognizing all temporal patterns including occurrence, development, climax, decline and ending. There have been some studies on finding a specified pattern in a temporal sequence[5], which aims to discover high quality rules efficiently and predict the occurrence of future events by representing temporal patterns in a meaningful and extendable way. Considering the volume, variety and velocity of information, the challenge of mining phase evolution is how to identify temporal pattern and mine topics of different phases automatically.

In this paper, we aim to fill in this research gap and focus on mining phase evolution by integrating temporal pattern learning and topic analysis. The number of posts published on social media reflects the degree of people's attention to a topic, so we define occurrence, development, climax, decline, ending pattern firstly by utilizing local maximum points and local minimum points in temporal sequence consisting of the number of posts published in a time unit. As k-means is an efficient cluster algorithm[10], we use it to conduct cluster analysis on temporal subsequence, adjacent points are clustered into the same cluster

[1] http://data.weibo.com/report/reportDetail?id=346

by minimizing the sum of within-cluster distance, then the extreme point of each cluster is picked out to recognize all event temporal phases. Furthermore, burst detection algorithm proposed in [11] is used to discover all peak interval in the mined phases. Finally, TextRank[12], an unsupervised summarization algorithm, which uses a graph-based ranking model to score each word in the document is adopted to generate the summarization of each event phase.

In summary, the major contributions of this work are: 1) This work is a first step towards utilizing temporal information, the number of posts, and topic information to mine evolution phases of hot events, a unified phase evolution mining model is proposed to identify the occurrence, development, climax, decline, ending phases and analyze the topics of each phase automatically. The model provides users an effective way to understand the event evolution process and make timely decision. 2) We design a quantitative method and empirically evaluate the performance of the proposed model on two real-world datasets collected from different social media platforms including Weibo and News, to understand the event evolution in a more comprehensive way. Experimental results show the characteristics of event evolution on different social media platforms. The results also show the potential of automatically mining evolution phases to alleviate decision specialists' workload.

The rest of this paper is organized as follows: In section II, we review related works. The components of the proposed model are introduced in section III. The experiments are discussed in section IV. Finally, we summarize our work and put forward future research directions.

II. LITERATURE REVIEW

A. Storyline Construction of Event

Some studies have been conducted to generate event storylines. [13] used self-organizing maps to identify events and assigned weight to each event according to the similarity between the event and given topics. Term weighting schemes were adopted to generate storyline summaries, but this model didn't take temporal information into consideration. [14] proposed a pictorial storyline generation method which aimed to sketch the event storyline with textual and pictorial information. [4] utilized user input queries to sketch the real-time storyline of the event. [3] introduced a Bayesian model to generate storylines from Social media Twitter. Different from these works, we introduce a unified phase evolution mining model for analyzing hot topics automatically which partitions the temporal sequence into occurrence, development, climax, decline and ending phases with cluster analysis and temporal pattern learning, then utilizes burst detection algorithm to discover peak intervals in the temporal sequence, finally, uses TextRank to help users better understand the topics in each phase.

B. Time Series Data Mining

There are extensive studies on time series data mining. [5] proposed a novel algorithms to discover rules in time series and predict the occurrence of future events. [6] put forward an admissible pruning strategy to accelerate density based time series clustering.[7] explored a dynamic time warping averaging algorithm to classify time series faster and more accurate. [8] made efforts to discover temporal pattern by using only some local patterns and deliberately ignoring the rest of the data. [9] emphasized the importance of giving a meaningful definition of the specified pattern before searching. Inspired by these works, so we define the development, climax and decline patterns with local maximum and minimum points in the temporal sequence based on the event evolution rules in the real world.

C. Text Summarization

Text summarization focuses on selecting the representative sentences or words to describe the semantics of documents[12][15]. [16] proposed a model to generate opinion summary for entities in Twitter by considering topic, opinion and insight as a whole. [17] focused on generating an opinion summary from a collection of microblogs related to a hot topic from the perspective of overall sentiment. In this paper, we use TextRank[12] to obtain topic descriptions in each event phase, which computes the words' significance with a graph-based ranking model.

III. EVENT PHASE EVOLUTION MINING MODEL

A. Problem Definition and Framework

Given a temporal sequence $P=\{P_1,\cdots,P_n\}$, where P_i is the number of posts published in the i^{th} time unit, the Event Phase Evolution Mining aims to find the beginning and ending of occurrence, development, climax, decline and ending phases respectively, and analyze the topic of each phase.

Fig. 1 shows the framework of the proposed model. It consists of temporal sequence clustering, peak interval detection and topic analysis. In the temporal sequence clustering module, local maximum points and local minimum points in temporal sequence are identified firstly, then, candidate phase intervals are mined based on clustering the identified local maximum points and minimum points, thirdly, empirical rules are used to mine the temporal patterns of each phase. In the peak interval detection module, the classical burst detection algorithm[11] is adopted to discover peak intervals of all phases, which helps gain insights into fine-grained attention. In the topic analysis module, the topics of each phase are summarized by TextRank. The details are described in the following sections.

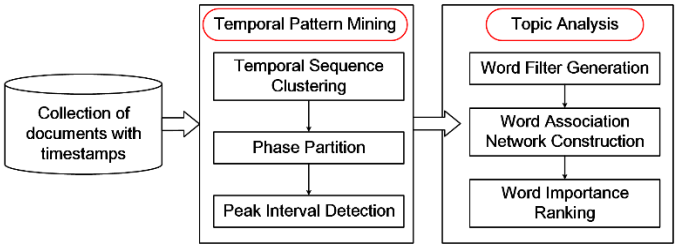

Fig. 1 A framework for phase evolution mining model

B. Temporal Pattern Mining of Different Phases

1) Temporal sequence clustering

The temporal sequence clustering module is designed to conduct cluster analysis on the temporal sequence. The variables in the temporal sequence $P=\{P_1,\cdots,P_n\}$ reflect the level of people's concern about a specific hot topic. To get the range of people's concern extent, we construct two subsequences $PMAX=\{P_{max1},\cdots,P_{maxT}\}$ and $PMIN=\{P_{min1},\cdots,P_{minT}\}$ which

consist of P's local maximal points and local minimal points respectively. Further, the adjacent time points in PMAX and PMIN are classified into the same cluster by using k-means. Specifically, this module classifies the adjacent time points in subsequence PMAX into the same cluster by minimizing objective function (1), where k is the number of clusters and μ_i is the center of the i^{th} cluster. Further, the maximum of each cluster is returned which reflects the degree of attention during the cluster. Therefore, a new subsequence $PMAX^{'}=\{P^{'}_{max1},\cdots,P^{'}_{maxk}\}(maxk=\sqrt{maxT})$ is returned, which will be used to recognize the occurrence, development, climax, decline and ending pattern in the temporal sequence with some empirical rules that we will discuss in the next section. The operations on the subsequence PMIN are similar and we obtain a new subsequence $PMIN^{'}=\{P^{'}_{min1},\cdots,P^{'}_{mink}\}(mink=\sqrt{minT})$.

$$d=\sum_{i=1}^{k}\sum_{P_{maxj}\in S_i}(P_{maxj}-\mu_i)^2 \quad (1)$$

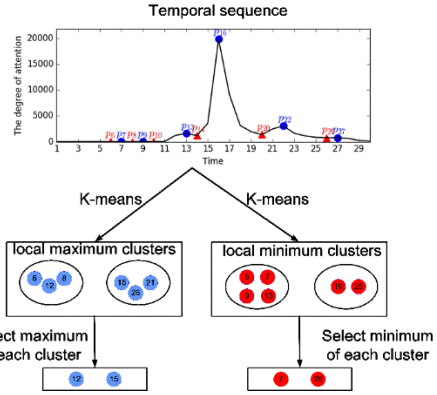

Fig. 2 A visual explanation of cluster analysis on a temporal sequence

The outline of temporal sequence clustering on the subsequence PMAX is described in Table I. And a visual explanation of this procedure is shown in the Fig. 2.

TABLE I THE OUTLINE OF TEMPORAL SEQUENCE CLUSTERING

1	Set the number of clusters k to $\sqrt{maxT}$ heuristically.
2	Choose k points from temporal sequence PMAX randomly to initialize the center of each cluster.
3	Assign each time point to the nearest cluster center.
4	Update the k cluster centers using the current memberships of each cluster.
5	If the value of objective function (1) don't change, go to 6; otherwise, go to 3.
6	Return the maximum in each cluster.

2) Phase partition rules

After obtaining the subsequences $PMAX^{'}$ and $PMIN^{'}$, temporal pattern learning aims to find occurrence, development, climax, decline and ending pattern in the subsequences. Before introducing how to recognize all the patterns, the definition for development, climax, decline pattern is given as follow:

a) Development pattern: In the development stage, the number of posts increases gradually. Fig. 3 is an example of development pattern. Given the $P^{'}_{maxi}$ obtained in the temporal sequence clustering module, the beginning of development pattern P_{devis} and the ending P_{devie} are determined by (2).

$$\frac{P^{'}_{maxi}-P_{devie}}{P^{'}_{maxi}}\geq threshold_{dev},\frac{P_{devie}-P_{devis}}{P_{devie}}\geq threshold_{dev} \quad (2)$$

$$\text{s.t. devis, devie} < \text{maxi}$$

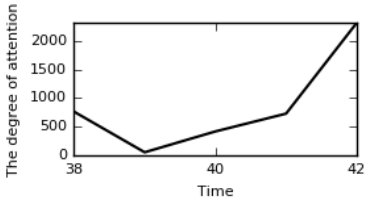

Fig. 3 The development pattern

b) Climax pattern: In the climax stage, the number of posts increases rapidly and the $P^{'}_{maxi}$ is covered with the i^{th} climax phase. Fig. 4 is an example of climax pattern. Given $P^{'}_{maxi}$ and P_{devie}, the beginning of climax pattern $P_{climaxis}$ and the ending $P_{climaxie}$ are determined by (3).

$$P_{climaxis}=P_{devie}$$
$$\frac{P^{'}_{maxi}-P_{climaxie}}{P^{'}_{maxi}}\geq threshold_{climax} \quad (3)$$
$$\text{s.t. climaxie}>\text{maxi}$$

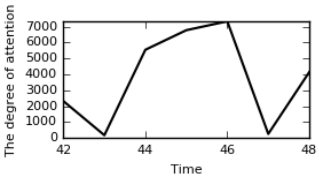

Fig. 4 The climax pattern

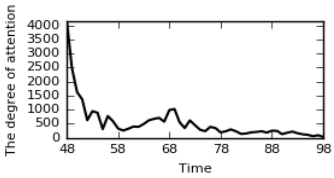

Fig. 5 The decline pattern

c) Decline pattern: In the decline stage, the number of posts starts to decrease and tends to be smooth. A visual explanation is shown in Fig. 5. Given $P_{climaxie}$ and $P^{'}_{mini}$, the beginning of the i^{th} decline phase P_{endis} and the ending P_{endie} are determined by (4).

$$P_{declineis}=P_{climaxie},\ P_{declineie}=P^{'}_{mini} \quad (4)$$
$$\text{s.t. declineie}>\text{declineis}$$

After recognizing the development, climax, decline patterns in the temporal sequence P, the beginning of the occurrence phase is the first point of P and the ending equals to the beginning of the first development phase. The beginning of the ending phase is the ending of the last decline phase and the ending is the last point of P.

3) Peak interval detection

The peak interval detection module is designed to discover the peak interval of each phase so that we gain insights into fine-grained attention. Take the development phase as an example, let $Pdev=\{P_{devs},\cdots,P_{deve}\}$ be the temporal subsequence in development phase, and $Sdev=\{S_{devs},\cdots,S_{deve}\}$ be its corresponding state sequence, where $S_i=1, 0$ means that the i^{th} time unit is in a positive state or a negative state respectively. And a peak interval consists of a continuous time units in the positive state. We adopt the burst detection technique proposed in [11] to determine the state of each time unit by minimizing the cost function defined as (5), where $\tau(S_i, S_{i+1})$ is the transfer cost, $f(P_i)$ is the distribution rule of Pdev, P_m is the maximum of Pdev, q_0 and q_1 is the probability of publishing a post in the negative state and positive state respectively. This problem can be solved by adapting the standard forward dynamic programming algorithm. Perform the similar operation on other phases in order to obtain all peak intervals.

$$C(Sdev|Pdev)=\sum_{i=devs}^{deve-1}\tau(S_i,S_{i+1})+\sum_{i=devs}^{deve}-\ln f(P_i)$$
$$\tau(S_i,S_{i+1})=\begin{cases}\ln(deve-devs+1) & \text{if } S_{i+1}>S_i\\ 0 & \text{otherwise}\end{cases} \quad (5)$$
$$f(P_i)=[\binom{P_m}{P_i}q_{S_i}^{P_i}(1-q_{S_i})^{P_m-P_i}]$$
$$q_0=\frac{P_{devs}}{P_m},\ q_1=\frac{2*P_{devs}}{P_m}$$

C. Topic Analysis of Different Phases

The topic analysis module aims to extract the keywords from the posts of each phase to describe the topics, and it consists of word filter generation, word association network construction and word importance ranking. The collection of posts in a phase is represented as $d=\{doc_1,\cdots,doc_D\}$, where D is the number of posts. The details of this module are described as follow:

1) Word filter generation: Since stop words refer to the most common words in a language and have no ability to describe the topic of each phase, they are filtered out before summarization. The rest of the words are extracted to construct word association network.

2) Word association network construction: Let $G=(V,E)$ be a text graph with the set of vertices V and set of edges E. Each word w_i is regarded as a vertex V_i of the text graph G. And two vertices are connected if their corresponding words co-occur within a window of maximum N words.

3) Word importance ranking: Starting from arbitrary values assigned to each node in the text graph G we use (6) to compute the importance of a word within a document, where |V| is the number of words in the document, $In(V_i)$ is the set of vertices that point to V_i, $Out(V_i)$ is the set of vertices that V_i point to, $S(V_i)$ is the score of vertex V_i standing for the importance of word w_i, and α is a damping factor that can be set between 0 and 1. The computation iterates until convergence, thus we obtain the importance of each word. Rank the words according to their scores and the top W words are selected as the keywords of posts[12].

$$S(V_i)=(1-\alpha)+\alpha*\sum_{j\in In(V_i)}S(V_j)/\left|Out(V_j)\right| \quad (6)$$

IV. Experimental Results and Discussion

A. Dataset and Parameter Settings

To validate the performance of the proposed unified phase evolution mining model, we conduct experiments on two real-world datasets from Weibo and News. Both of datasets are about Rio Olympics and published from August 1,2016 to August 31, 2016.

1) The Weibo dataset contains 10,166 posts published by mainstream media and official certification users.

2) The News dataset consists of 319,785 pieces of news published by mainstream media such as Xinhua and Ifeng.

In the experiments on both datasets, we set $threshold_{dev}=0.5$, $threshold_{climax}=0.5$ in temporal pattern mining module. The window size used in word association network construction equals to 5, then top 5 words are extracted from each document, finally, 10 words with the highest frequency are picked out in topic analysis module.

B. Method for Comparison

We compare the proposed algorithm with the following baselines to evaluate the its performance:

- *Uniform partition(UP)*: We divide the temporal sequence into five phases evenly. That is to say we regard days from August 1 to August 6 as occurrence phase, days from August 7 to August 12 as development phase and so on.
- *Extended burst detection(EBD)*: Since there are no existing phase partition models, we extend the classical burst detection algorithm[11] to partition the temporal sequence. First, the burst detection algorithm is adopted to obtain all the burst interval. Then, the beginning of the first interval and the ending of the last interval are regarded as the beginning and ending of the climax phase. Last, the remaining temporal sequence is uniformly partitioned into occurrence, development, decline, ending phases respectively.

C. Gold Standard Generation

Based on the method in [18], we examine the phase partition results by human judgements. Formally, let $G=\{G_1,\cdots,G_N\}(G_i\in\{occ, dev, cli, dec, end\})$ be the phase partition results given by human, where occ, dev, cli, dec and end are the abbreviation of occurrence, development, climax, decline and ending phase respectively, N is the length of temporal sequence. For example, G_i=dev if the i^{th} time unit belongs to development phase. In our experiments, we invited three graduate students to label the temporal sequence according to the trend of the number of posts and the definition of temporal patterns we introduce in section III. The phase was determined only when two students agreed that a time unit was belonged to the specific phase. After discussion, the average interrater agreement as measured by Fleiss's kappa[19] is 0.75 and the results are shown in Table II. For brevity, we represent the results with the beginning and ending of each phase.

TABLE II PHASE PARTITION RESULTS GIVEN BY HUMAN JUDGEMENT

	occurrence	development	climax	decline	ending
Weibo	1 ~ 4	5 ~ 5	6 ~ 21	22 ~ 28	29 ~ 31
News	1 ~ 2	3 ~ 6	7 ~ 19	20 ~ 29	30 ~ 31

D. Evaluation Metrics

We put forward two quantitative methods to evaluate our model in the view of phase partition and topic analysis.

1) Overlapping Ratio

The results given by the proposed model are denoted as $R=\{R_1,\cdots,R_N\}(R_i\in\{occ, dev, cli, dec, end\})$. The overlapping ratio is used as a quantitative evaluation measure, which is defined as (7).

$$ratio=\frac{\sum_{i=1}^{N}I(G_i,R_i)}{N}$$
$$I(G_i,R_i)=\begin{cases}1 & G_i=R_i\\ 0 & G_i\neq R_i\end{cases} \quad (7)$$

2) Inverse Phase Frequency

The topic of each phase is described with the keywords extracted from the posts published in the corresponding phase. Intuitively, the topics are different in each phase of event evolution. So we can compare the performance on event evolution mining by evaluating the diversity of topics. Inspired by this, we design a quantitative method to evaluate the phase partition results from the perspective of semantics. Let pf_i be the phase frequency of the word w_i, which means the number of phases that w_i appears in. And the inverse phase frequency of word w_i is denoted by $ipf_i = 5/pf_i$. Further, the inverse phase frequency of a model equals to the average of inverse phase frequency of all keywords mined by the model. The formal definition is given as (8), where W is the set of all keywords and |W| is the number of words in the set. The higher IPF represents better performance on event evolution mining.

$$IPF = \sum_{w_i \in W} \frac{ipf_i}{|W|} \tag{8}$$

E. Results and Analysis

The results of phase partition are shown in Table III and Table IV. Each phase is represented with its beginning and ending. A visual explanation of phase partition and peak interval detection results on Weibo and News datasets are shown in Fig. 5 and Fig. 6. The green, red, yellow, blue, purple part represent occurrence, development, climax, decline, end phase respectively. The red dashed rectangles are the peak intervals discovered by the model.

TABLE III PHASE PARTITION RESULTS ON WEIBO DATASET

	occurrence	development	climax	decline	ending
our model	1 ~ 4	5 ~ 6	7 ~ 19	20 ~ 28	29 ~ 31
UP	1~6	7~12	13~18	19~24	25~31
EBD	1~2	3~5	6~22	23~26	27~31

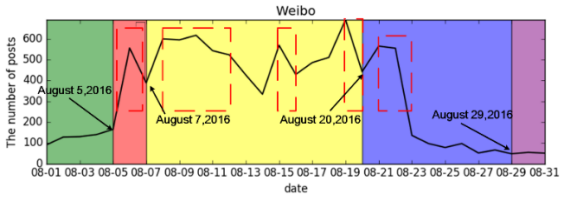

Fig. 5 The result of phase partition and peak interval detection on Weibo dataset.

TABLE IV PHASE PARTITION RESULTS ON NEWS DATASET

	occurrence	development	climax	decline	ending
our model	1 ~ 2	3 ~ 5	6 ~ 14	15 ~ 25	26 ~ 31
UP	1~6	7~12	13~18	19~24	25~31
EBD	1~3	4~6	7~8	9~19	20~31

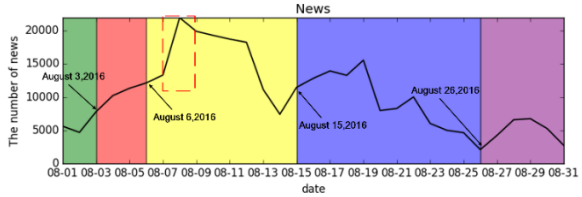

Fig. 6 The result of phase partition and peak interval detection on News dataset.

The overlapping ratio and inverse phase frequency of each dataset are displayed in Table V. From the view of overlapping ratio, the proposed model match the human judgement better than baseline methods, which proves that the event evolution mined by the model conforms to human cognition. On the other hand, the inverse phase frequency obtained by the proposed model is higher, which means that the event evolution mined by the model satisfies the event development process in the semantic level. In addition, we find that the proposed model performs better on the Weibo dataset than News. The main reason is that for both datasets we set $threshold_{climax}$ equals to 0.5 and don't take the different features of multiple platforms into consideration. In fact, we learn that the changing trends are sharper on Weibo from Fig. 5 and Fig. 6.

TABLE V THE OVERLAPPING RATIO AND INVERSE PHASE FREQUENCY RESULTS

	overlapping ratio			IPF		
	our model	UP	EBD	our model	UP	EBD
Weibo	**90.3%**	51.6%	83.9%	**3.91**	3.79	**3.91**
News	**67.7%**	48.4%	29.0%	**3.22**	2.74	2.80

Based on the phase partition and peak interval detection results, we perform topic analysis on both datasets to understand the topic of each phase on different social media. The results on the two datasets are shown in Fig. 7. On the basis of the results, we can find that the posts which indicate the event happening emerge in the occurrence phase. For example, in the Olympic event, the posts containing words 'arrival' and 'Brazil' indicates that the game was going to start. In the development phase, a specific affair appears which attracts more people's attention and the number of the posts increases rapidly. For example, the opening ceremony made more people pay attention to the Olympic game. Further, the event is continually discussed in the climax phase and the degree of the public's attention reach its maximum. In the Olympic event, there were massive posts related to competition and players on the social media during the game. As the event unfolds, the discussion about the event starts to decrease in the decline phase. Specifically, the relevant posts started to decrease as most of the matches in the Olympic game were over. Finally, the information related to the event is dying out in the ending phase.

Meanwhile, according to the event phases and topics of each phase on Weibo and News platform, we can find that there are some different features between Weibo and news media platform:

1) Because of the convenience of publishing posts on Weibo, the changing trends are sharper on Weibo than news media platform. For example, the fastest grow rate is more than 200% on Weibo, while less than 100% on news media platform. This indicates that it's necessary to keep track of the event evolution trends on multiple social media platforms in order to respond to urgent social events timely.

2) The ending of climax phase on news media platform is earlier than Weibo. Most of the content in news dataset are about Olympic games and the amount of information is related to game schedule. And in the later period of Olympic games, the number of matches starts to reduce, so there is less related news. However, due to the recreational and informal characteristics of Weibo, the number of posts growth on Weibo is usually caused by some entertainment events or popular sports stars. For example, the keywords extracted from one peak interval of

climax phase on Weibo contains ‘Yang Sun’ and ‘Yuanhui Fu’, which are names of popular sports stars in China.

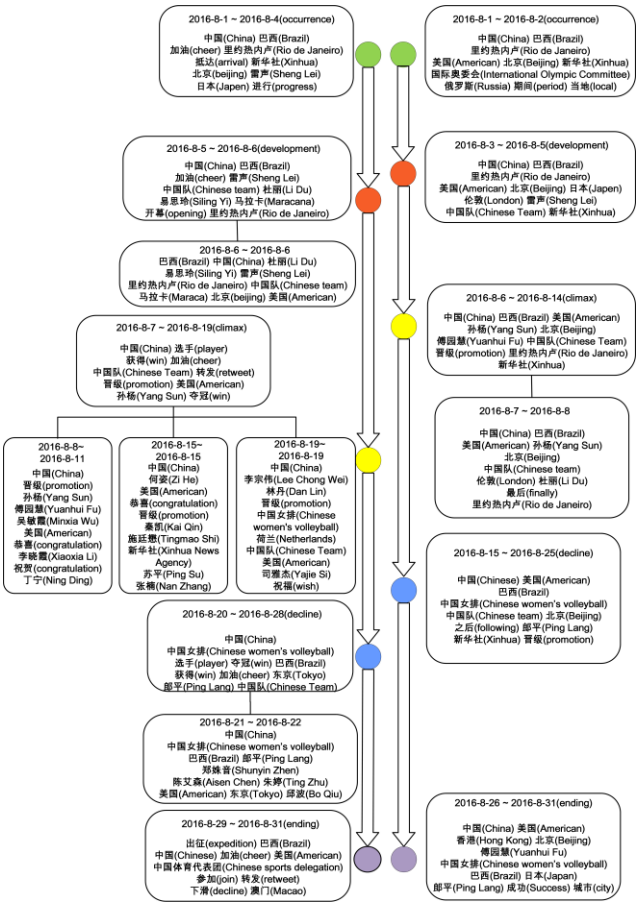

Fig. 7 The results of topic analysis on both social media platforms

The experimental results on two real-world datasets also show that the proposed model will be helpful for management department to keep track of social event. First, the management department can learn about the event evolution process easily based on the phase partition results given by this model. So they are able to evaluate the level of urgency and offer guidance to public opinion timely. Further, based on the topic analysis results, the management department can realize the specific topic in each phase. So they can judge the sentiments of the public and respond to the negative opinions.

V. Conclusions

In this paper, we present a unified phase evolution mining model to analyze hot topics. It identifies the temporal patterns based on k-means and empirical rules and discover all peak intervals with burst detection algorithm. Further, it describes the topic of each phase with a summarization technique TextRank. Finally, we perform experiments on two real-world datasets collected from different social media platform. Experimental results verify the efficacy of the proposed model and show the characteristics of event evolution on different social media platform. In future work, we will test the proposed model on other hot topics such as national policies and social events, and represent the temporal patterns with a more meaningful way to improve the performance of our model.

Acknowledgment

This research is supported by the Key Research Program of the Chinese Academy of Sciences under Grant No. ZDRW-XH-2017-3; National Key Research and Development Program under Grant No. 2016YFC1200702 ;National Natural Science Foundation of China under Grant No. 71621002, 61671450, 61402123.

References

[1] Shahaf, Dafna, et al. "Information cartography: creating zoomable, large-scale maps of information." Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013.

[2] Yan, Rui, et al. "Evolutionary timeline summarization: a balanced optimization framework via iterative substitution." Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. ACM, 2011.

[3] Hua, Ting, et al. "Automatical Storyline Generation with Help from Twitter." Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016.

[4] Lin, Chen, et al. "Generating event storylines from microblogs." Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012.

[5] Shokoohi-Yekta, Mohammad, et al. "Discovery of meaningful rules in time series." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

[6] Begum, Nurjahan, et al. "A General Framework for Density Based Time Series Clustering Exploiting a Novel Admissible Pruning Strategy." arXiv preprint arXiv:1612.00637 (2016).

[7] Petitjean, François, et al. "Faster and more accurate classification of time series by exploiting a novel dynamic time warping averaging algorithm." Knowledge and Information Systems 47.1 (2016): 1-26.

[8] Zakaria, Jesin, et al. "Accelerating the discovery of unsupervised-shapelets." Data mining and knowledge discovery 30.1 (2016): 243-281.

[9] Keogh, Eamonn, Stefano Lonardi, and Bill'Yuan-chi Chiu. "Finding surprising patterns in a time series database in linear time and space." Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2002.

[10] Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A k-means clustering algorithm." Journal of the Royal Statistical Society. Series C (Applied Statistics) 28.1 (1979): 100-108.

[11] Kleinberg, Jon. "Bursty and hierarchical structure in streams." Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2002.

[12] Mihalcea, Rada, and Paul Tarau. "TextRank: Bringing order into texts." Association for Computational Linguistics, 2004.

[13] Lin, Fu-ren, and Chia-Hao Liang. "Storyline-based summarization for news topic retrospection." Decision Support Systems 45.3 (2008): 473-490.

[14] Wang, Dingding, Tao Li, and Mitsunori Ogihara. "Generating Pictorial Storylines Via Minimum-Weight Connected Dominating Set Approximation in Multi-View Graphs." AAAI. 2012.

[15] Yih, Wen-tau, et al. "Multi-Document Summarization by Maximizing Informative Content-Words." IJCAI. Vol. 7. 2007.

[16] Meng, Xinfan, et al. "Entity-centric topic-oriented opinion summarization in twitter." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012.

[17] Li, Qiudan, et al. "Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems." Knowledge-Based Systems 107 (2016): 289-300.

[18] Chang, Jonathan, et al. "Reading Tea Leaves: How Humans Interpret Topic Models." Advances in Neural Information Processing Systems 32(2009):288-296.

[19] Fleiss, Joseph L. "Measuring nominal scale agreement among many raters." Psychological bulletin 76.5 (1971): 3