

Optimized Event Storyline Generation based on Mixture-Event-Aspect Model

Lifu Huang

Shenzhen Key Lab for Cloud Computing
Technology and Applications,
Peking University Shenzhen Graduate School,
Shenzhen, Guangdong 518055, P.R.China
warrior.fu@gmail.com

Lian'en Huang*

Shenzhen Key Lab for Cloud Computing
Technology and Applications,
Peking University Shenzhen Graduate School,
Shenzhen, Guangdong 518055, P.R.China
hle@net.pku.edu.cn

Abstract

Recently, much research focuses on event storyline generation, which aims to produce a concise, global and temporal event summary from a collection of articles. Generally, each event contains multiple sub-events and the storyline should be composed by the component summaries of all the sub-events. However, different sub-events have different part-whole relationship with the major event, which is important to correspond to users' interests but seldom considered in previous work. To distinguish different types of sub-events, we propose a mixture-event-aspect model which models different sub-events into local and global aspects. Combining these local/global aspects with summarization requirements together, we utilize an optimization method to generate the component summaries along the timeline. We develop experimental systems on 6 distinctively different datasets. Evaluation and comparison results indicate the effectiveness of our proposed method.

1 Introduction

With the rapid growth of the World Wide Web, information explosion has become an important issue to modern people. Those who search for information from the Internet often get lost and confused by the overwhelmingly large collection of web documents. So how to get a concise and global picture for a given event subject is an urgent problem to be solved. Although many document understanding systems have been proposed, such as

multi-document summarization systems, to generate a compressed summary by extracting the major information from the collection of documents, they ignored the dynamic development information of an event. Intuitively, each event is long-running and contains multiple sub-events, including related events. Users are likely to prefer a summary of all occurrences of all the sub-events along the timeline of the event. This motivates us to study the task of generating event storyline from a collection of web documents related to an event subject.

The research of event storyline summarization is popular in recent years. Its task is to summarize a collection of web documents by extracting representative information based on all the sub-events and generate a global summary. Generally, generating such a global storyline is quite interesting for the following main reasons: (1) It can help people catch the whole incident based on an overall temporal structured summary for a given subject, and understand the cause, climax, development process and result of an event. (2) It can also make people know what other events are related, or the effect of this incident to subsequent events, which can present the evolution of an event along a timeline.

Though several methods of generating event storyline have been proposed recently, there are still some problems unresolved. As event storyline summarization is a process to generate component summaries based on the multiple sub-events, which is different from traditional summarization focusing on only one subject, so how to exactly extract all the sub-events is the first challenge. Moreover, users tend to bias to the sub-events which have global

*Corresponding author

consistency with the given event subject, so the sub-events should not be considered equally when generating the component summaries. It is also a great challenge to generate a qualified summary based on the different types of sub-events. The component summaries should be correlative across different dates based on the global collection (Yan et al., 2011a).

Mei and Zhai (Mei and Zhai, 2005) proposed to use theme or topic to model different sub-events, which is to some extent similar to our method. To be different, in this paper we introduce “local/global” property to distinguish different part-whole relationship between the sub-events and the major event, which have not been considered before in storyline generation or summarization, to improve the quality of the storyline. The local/global property corresponds to the elements of an event, such as the place, characters and other body information. These information reflects the relationship between the sub-events and the major event. Some sub-events have distinctive body information and little relevance with each other. They generally occur for a local period, which we name as “local-sub-events”. While other sub-events often share common properties with each other and have close relationship with the major event and we call them as “global-sub-events”. Here we give some examples to illustrate the difference. For the event “Connecticut school shooting” which occurred on Dec.14 2012, its sub-events such as “Obama’s speech for this massacre” or “Gun control Act” have little word co-occurrences and distinctive event body information to each other, while the process and result of this tragedy can be regarded as global-sub-events which have a lot of word co-occurrences and share common properties with the major event.

Inspired by these, to detect different types of sub-events based on word co-occurrences between sub-events and the major event, we propose a mixture-event-aspect (MEA) model to formalize different types of sub-events into local/global aspects, which are implicated with clusters of sentences. Then combining the local/global aspects with summarization requirements together, we utilized an optimization approach to get the optimal component summaries along the timeline. We evaluate our method on 6 distinctively different datasets. Performance compar-

isons among different system-generated storylines demonstrate the necessity to distinguish different types of sub-events and also indicates the effectiveness of the proposed mixture-event-aspect model.

The rest of the paper is organized as follows. We briefly review the related work in section 2. In section 3 we present the details of optimized event storyline generation based on mixture-event-aspect model. Experiments and results are discussed in Section 4. Finally we draw a conclusion of this study in Section 5.

2 Related Work

Our work is related to several lines of research in the literature including multi-document summarization (MDS), topic detection and tracking (TDT), temporal text mining (TXM) and temporal news summarization (TNS).

Multi-document summarization is a process to generate a summary by reducing documents in size while retaining the main information. To date, different features and ranking strategies have been studied. Radev et al. (Radev et al., 2004) proposed to implement MEAD as a centroid-based summarizer by combining several predefined features to score the sentence. LexPageRank (Erkan and Radev, 2004) is the representative work which is based on PageRank (Page et al., 1999) algorithm. Some methods have been proposed to extend the conventional graph-based models recently including multi-layer graph incorporated with different relationship (Wan, 2008), ToPageRank based on the topic information (Pei et al., 2012) and DivRank (Mei et al., 2010) balancing the prestige and diversity.

Topic detection and tracking (TDT) aims to group news articles based on the topics discussed in them, detect some novel and previously unreported events and track future events related to the topics (Wang et al., 2012). Kumaran and Allan (Kumaran and Allan, 2004) showed how performance on new event detection could be improved by the use of text classification techniques as well as by using named entities in a new way. Makkonen et al. (Makkonen et al., 2004) proposed a method that incorporated simple semantics into TDT by splitting the term space into groups of terms. Krause et al. Wang et al. (Wang et al., 2007) and Wang et al. (Wang et al.,

2009) worked on topic tracking from multiple news streams. Their methods extracted meaningful topics from multi-source news collections and tracked different topics as they evolved from one to another along the timeline.

Our work is also related to temporal text mining and temporal news summarization. The task of temporal news summarization is to generate news summaries along the timeline from massive data. Chieu et al. (Chieu and Lee, 2004) built a system that extracted events relevant to a query from a collection of related documents and placed such events along a timeline. Yan et al. (Yan et al., 2011b) designed an evolutionary timeline summarization approach to construct a timeline of a topic by optimizing the relevance, coverage, coherence, and diversity. Lin et al. (Lin et al., 2012) explored the problem of generating storylines from microblogs for user input queries. They first proposed a language model with dynamic pseudo relevance feedback to obtain relevant tweets and then generated storylines via graph optimization.

3 Approach Details

In this section, we first propose a mixture-event-aspect model to detect local/global sub-events based on part-whole relationship with the major event and then present a new method to estimate the bursty of each aspect on a certain date. Afterwards we utilize an optimization method based on local/global aspects to extract the qualified summary.

3.1 Mixture-Event-Aspect Model

The key challenge to our storyline generation task is to detect and distinguish different types of sub-events contained in the article collection. In the collection, each sentence is assigned with a certain date and sentences that are assigned with the same date are grouped into the same sub-collection. Considering the consistency of content between the sub-events and the major event, we model different sub-events into two types: local-sub-event and global-sub-event, and introduce local/global aspects correspondingly. Generally, local aspects which correspond to local-sub-events have distinctive words distribution from each other and sustain for a local context while the global aspects corresponding

to global-sub-events have coincident words distribution with the major event. To capture specific words, Titov and McDonald (Titov and McDonald, 2008) proposed a multi-grain topic model, relying on word co-occurrences within short paragraphs and Li et al. (Li et al., 2010) proposed a entity-aspect model based on word co-occurrences within single sentences. Inspired by these ideas, we rely on word co-occurrences within local period context to detect mixed local and global aspects implicated in the whole collection. We name this model as “Mixture-Event-Aspect (MEA)” model which can simultaneously detect local/global aspects and cluster sentences and words into different aspects.

3.1.1 Model Description

Our mixture-event-aspect (MEA) model can be extended from both the Probabilistic Latent Semantic Analysis (PLSA) (Hofmann, 1999) and Latent Dirichlet Allocation (LDA) (Blei et al., 2003). We model two distinct types of aspects: global aspects and local aspects, based on their relationship with the major event. The distribution of global aspects is fixed for the collection while the distribution of local aspects is fixed to a local period of sub-collections. That means a sentence is sampled either from the mixture of the global aspects or from the local aspects specific for the local context. Here we take the event “Connecticut school shooting” as an example. For the sentence “On Sunday, President Obama came to Connecticut to give a lecture, expressing his sorrow for ... and calling for an end to such incidents”, the words such as “Obama”, “lecture”, “express” are only occurred for the local period of two days and have no co-occurrence with other neighboring period sentences, so we sample the sentence as a local aspect sentence. But for the sentence “All schools in Newtown, the northeastern U.S. state of Connecticut were in lockdown after a shooting was reported at a local elementary school”, the words such as “Connecticut”, “shooting”, “elementary” have high co-occurrence frequency in the whole collection, so we sample the sentence as a global aspect sentence.

To detect aspects, we first divide words into two types: aspect words and background words. Background words are commonly used in the whole event corpus while aspect words are clearly associat-

$$p(v_{c,s} = v_h, \eta_{c,s} = loc, z_{c,s} = a | v', z', y, w) \propto \frac{n_{h^*}^c + \lambda}{n_{(\cdot)}^c + S_p^* \lambda} \cdot \frac{n_{loc}^{c,v_h} + \alpha_{loc}^{mix}}{n_{(\cdot)}^{c,v_h} + \sum_{r' \in gl, loc} \alpha_r^{mix}} \cdot \frac{n_{loc,a}^{c,v_h} + \alpha^{loc}}{n_{loc}^{c,v_h} + A_{loc} \alpha^{loc}} \cdot \frac{\prod_{l=1}^L \prod_{i=0}^{E(l)-1} (C_{(l)}^a + i + \beta)}{\prod_{i=0}^{E(\cdot)-1} (C_{(\cdot)}^a + i + L\beta)}$$

where $n_{h^*}^c = \#\{s_i | v_{c,s_i} = v_{i-S_p+h^*+1}\}$, denotes the number of times a sentence s_i in collection c is assigned to its h^{*th} window and for each sentence s_i , we have $h = i - S_p + 1 + h^*$. S_p^* is the number of windows that contain sentence s . $n_{(\cdot)}^c$ denotes the number of sentences in collection c . S_p represents the number of dates a window covered. n_{gl}^{c,v_h} and n_{loc}^{c,v_h} are the number of sentences in window v_h that are assigned to global or local aspects. $n_{(\cdot)}^{c,v_h}$ is the number of sentences assigned to window v_h . $n_{gl,a}^c$ is the number of sentences in all global aspects that are assigned to aspect a and $n_{loc,a}^{c,v_h}$ is the number of local aspect sentences in the window v_h are assigned to aspect a . A_{gl} is the number of global aspects in collection C while A_{loc} is the number of local aspects. $E(l)$ represents the number of times of word l occurs in the current sentence and is assigned to be an aspect word, while $E(\cdot)$ is the total number of words in the current sentence that are assigned to be an aspect word.

Then we compute the assignments of the considered word. We sample the hidden variables $y_{c,s,n}$ for each word with the following rules:

$$p(y_{c,s,n} = 1 | z, y') \propto \frac{C_{(1)}^\pi + \gamma}{C_{(\cdot)}^\pi + 2 \cdot \gamma} \cdot \frac{C_{(w_{c,s,n})}^B + \beta}{C_{(\cdot)}^B + L \cdot \beta}$$

$$p(y_{c,s,n} = 2 | z, y') \propto \frac{C_{(2)}^\pi + \gamma}{C_{(\cdot)}^\pi + 2 \cdot \gamma} \cdot \frac{C_{(w_{c,s,n})}^a + \beta}{C_{(\cdot)}^a + L \cdot \beta}$$

where $C_{(1)}^\pi$ and $C_{(2)}^\pi$ are the numbers of words assigned to be background words and aspect words. $C_{(\cdot)}^\pi$ is the total number of words. $C_{(\cdot)}^B$ is the total number of background words and $C_{(\cdot)}^a$ is the number of words assigned to aspect a .

By using the samples from Gibbs sampling, we can effectively make the following estimations:

$$\varphi_w^B = \frac{C_w^B + \beta}{C_w^B + L \cdot \beta}, \quad \varphi_w^a = \frac{C_w^a + \beta}{C_w^a + L \cdot \beta}, \quad \psi_{h^*} = \frac{n_{h^*}^c + \lambda}{n_{(\cdot)}^c + S_p^* \lambda}$$

$$\pi_y = \frac{C_y^\pi + \gamma}{C_{(\cdot)}^\pi + 2 \cdot \gamma}, \quad \rho_\eta^{c,v} = \frac{n_\eta^{c,v} + \alpha_\eta^{mix}}{n_{c,v} + \sum_{r \in (gl, loc)} \alpha_r^{mix}}$$

$$\theta_a^{gl} = \frac{n_{gl,a}^c + \alpha^{gl}}{n_{gl}^c + A_{gl} \alpha^{gl}}, \quad \theta_a^{loc} = \frac{n_{loc,a}^{c,v} + \alpha^{loc}}{n_{loc}^{c,v} + A_{loc} \alpha^{loc}}$$

The hyper-parameters like $\alpha, \beta, \gamma, \lambda$ can be estimated using standard methods introduced in (Minka, 2000).

3.2 Bursty Period Detection

We borrow the definition of ‘‘bursty’’ from (Lappas et al., 2009) to measure the popularity of the event on a certain date. Intuitively, each aspect have different bursties on different dates. In this section, we try to obtain the temporal aspect sequences of an event based on the bursty periods of all the aspects. During its bursty period, one aspect should (1) be more popular than other aspects (2) be continuously more popular than other time. Following these intuitions, we design a method to measure the bursty of each aspect and get the bursty period.

Let A_k be the k^{th} aspect obtained from the mixture-event-aspect model, we estimate the bursty of A_k at a certain date t as follows.

$$bursty(A_{k,t}) = p(t|A_k) = \frac{p(A_k|t) \cdot p(t)}{\sum_{t'} p(A_k|t') p(t')}$$

where $p(A_k|t)$ is measured by the number of sentences assigned to aspect A_k in date t divided by the total number of sentences in date t . $p(t)$ is estimated by the total number of sentences in aspect A_k divided by the overall number of sentences in the collection C .

After getting the bursty of aspect A_k at each date, we can find the most popular date and expand on both sides to obtain the whole burst period in which the bursties are higher than the neighboring aspects and continuous higher than other dates.

3.3 Optimization-based Storyline Generation

With the methods discussed in previous sections, we can get the local/global aspect sequence. Each aspect contains numbers of sentences and we are aiming to select the most representative ones to compose the final storyline. Considering users’ bias and the length requirement, different aspects should have different proportions in the last storyline. For global aspects which correspond more to users’ interest, they should share a larger proportion in the final storyline than local aspects. Thus, we use an optimization method to determine if a sentence is selected to be a summary sentence or to be discarded based on the multiple local/global aspects and finally get the optimal storyline. We formalize this problem as selecting a subset of sentences S from the aspect A_k

to minimize the information loss.

$$\arg \min \sum_{A_k \in C, S \in A_k} \sum_{z \in A_k - S, s \in S} O(z, s)$$

where $O(z, s)$ is the cost function which measures the cost of representing sentence z with sentence s . Generally, this is an NP-hard problem (Cheung et al., 2009) but we can use POPSTAR, an implementation of an approximate solution proposed by Resende and Werneck (Resende and Werneck, 2004). To model different costs between global or local aspects and determine the proportions of different aspects in the final storyline, we utilize a function $\zeta(s)$. When sentences z and s are local aspect sentences, $\zeta(s) = \chi$, or, $\zeta(s) = 1 - \chi$. Formally, we incorporate two kinds of decreasing/increasing logistic functions, $\ell_1(x) = 1/(1 + e^x)$ and $\ell_2(x) = e^x/(1 + e^x)$, to define the cost function as

$$O(z, s) = \zeta(s) \cdot \ell_1(S(s)) \cdot \ell_2(S(z)) \cdot D_{KL}(s, z)$$

where $S(s)$ and $S(z)$ are the ranking scores of sentences s and z among the aspect A_k with LexPageRank algorithm. $D_{KL}(s, z)$ is used to measure the similarity between sentence s and z with Kullback-Leibler divergence here.

With this optimization method, we get the representative sentences of each aspect for the given event subject. Combining all the representative sentences together based on the aspect sequence, we finally generate the storyline.

4 Experiments and Evaluation

4.1 Datasets

To evaluate our framework for event storyline generation, we conduct our experiments on the datasets amounting to 12418 articles for 6 event subjects from 6 famous news websites, which provide date edited by professional editors. Each article consists of three parts, title, publish-time and news content. Table 1 and Table 2 give the brief description of the 12418 articles. To generate reference summary, we invite 12 undergraduate students with good English ability to read the sentences, and for each event subject we ask two students to label human storylines.

4.2 Evaluation Metrics

We use the ROUGE¹ (Lin and Hovy, 2003) (Recall Oriented Understudy for Gisting Evaluation) toolkit

¹<http://www.isi.edu/licensed-sw/see/rouge/>

Table 1: News sources of the 6 datasets

News Sources	Number of Articles
CNN	2357
Fox News	1936
New York Times	2178
ABC	2113
Washington Post	1405
Xinhua	2429

Table 2: Event subjects of the datasets

Event Subjects	Number of Articles
Connecticut school shooting	1792
The earthquake in Tokyo, Japan	2046
The U.S. presidential election	2573
Sandy hurricane attacked America	1827
American curiosity rover landed on Mars	1651
The 30th London Olympic Games	2529

to evaluate our framework, which has been widely applied for summarization evaluation. It evaluates the quality of a summary by counting the overlapping units between the candidate summary and reference summaries. There are many kinds of ROUGE metrics to measure the system-generated summarization such as ROUGE-N, ROUGE-L, ROUGE-W, and ROUGE-U, of which the most important one is ROUGE-N with 3 sub-metrics: precision, recall, and F-score.

$$ROUGE - N = \frac{\sum_{S \in RS} \sum_{N-gram \in S} Count_{match}(N - gram)}{\sum_{S \in RS} \sum_{N-gram \in S} Count(N - gram)}$$

where RS represents the reference summaries. $N-gram \in RS$ in the metrics denotes the N-grams in reference summaries. $Count_{match}(N - gram)$ is the maximum number of N-grams co-occurring in the candidate summary and in the set of reference summaries. $Count(N - gram)$ is the number of N-grams in the reference summaries.

The ROUGE toolkit can report separate scores for 1, 2, 3, and 4-gram. In the experimental results we report three ROUGE F-measure scores: ROUGE-1, ROUGE-2, ROUGE-W metrics. The higher the ROUGE scores, the better the summary is.

4.3 Algorithms for Comparison

Given a collection of news articles, we first decompose them into sentences, and then assign each sentence with a certain date, afterwards stop-words removing and words stemming are performed. We choose the following algorithms as baseline systems. Specifically, baseline 2 and 3 are summarization

systems which are similar to our storyline generation system. Then we choose baseline 4, 5 to evaluate the effectiveness of the proposed method. It must be said that all the systems are required to generate the same number of summary words with the human reference. We conduct the same preprocessing for all algorithms for fairness.

- **Random** : The method selects sentences randomly from the sentence collection.

- **LexPageRank (LexRank)**: This method applies the graph-based multi-document summarization algorithm which first constructs a sentence connectivity graph based on the cosine similarity and then chooses top-ranked sentences with PageRank.

- **Chieu** : This method was proposed by Chieu (Chieu and Lee, 2004), utilizing interest and burstiness to rank sentences, and choosing the top-ranked query related sentences to construct the timeline.

- **LDA+LexPageRank (LDALR)** : This method first applies standard LDA to detect latent topics from the collection and clusters sentences to multiple aspects, then utilizes PageRank to generate the most representative component summaries from all the aspects.

- **MEA+LexPageRank (MEALR)** : This method applies the proposed mixture-event-aspect model to cluster sentences into multiple aspects and then utilizes PageRank to generate the most representative component summaries from all the aspects.

- **MEA+Optimization (MEAOp)** : This method extracts local/global aspects with the proposed mixture-event-aspect model, and then utilizes the optimization method to get the qualified summary.

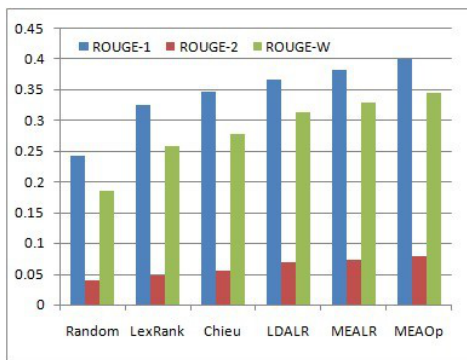


Figure 3: Overall performance for comparison

Table 3: Results of different systems on 6 subjects

Systems	Subject1			Subject2		
	R-1	R-2	R-W	R-1	R-2	R-W
Random	0.234	0.037	0.188	0.242	0.039	0.192
LexRank	0.317	0.045	0.257	0.326	0.051	0.262
Chieu	0.332	0.056	0.277	0.351	0.055	0.283
LDALR	0.356	0.069	0.297	0.369	0.066	0.327
MEALR	0.369	0.072	0.313	0.381	0.076	0.348
MEAOp	0.381	0.075	0.331	0.398	0.081	0.364

Systems	Subject3			Subject4		
	R-1	R-2	R-W	R-1	R-2	R-W
Random	0.258	0.042	0.191	0.234	0.036	0.179
LexRank	0.339	0.049	0.272	0.309	0.043	0.242
Chieu	0.364	0.059	0.296	0.329	0.053	0.264
LDALR	0.383	0.071	0.331	0.347	0.065	0.308
MEALR	0.396	0.076	0.3512	0.368	0.069	0.312
MEAOp	0.419	0.082	0.371	0.376	0.071	0.323

Systems	Subject5			Subject6		
	R-1	R-2	R-W	R-1	R-2	R-W
Random	0.222	0.034	0.166	0.264	0.045	0.195
LexRank	0.309	0.042	0.237	0.349	0.054	0.276
Chieu	0.319	0.049	0.258	0.371	0.062	0.293
LDALR	0.342	0.062	0.291	0.392	0.073	0.325
MEALR	0.372	0.068	0.299	0.406	0.079	0.349
MEAOp	0.384	0.070	0.309	0.427	0.087	0.368

4.4 Overall Performance Comparison

We experiment with all the baselines and our framework on the 6 datasets. We take the average F-score performance in terms of 3 ROUGE-F scores: ROUGE-1, ROUGE-2 and ROUGE-SU4. The overall results are shown in Figure 3 and details are listed in Tables 3.

Figure 3 and Table 3 show the performance of these systems on the same datasets. The local/global optimization balance parameter $\chi = 0.5$. From Figure 3 and Table 3 we have following observations:

- Generally, the Random gets the worst performance;

- The LexRank system outperforms Random algorithm. This is due to the fact that LexRank ranks all the sentences based on eigenvector centrality and the global relationship between sentences, which tends to select the most informative sentences as the summary.

- The results of Chieu (Chieu and Lee, 2004) system are better than those of LexRank. This may be mainly for the reason that Chieu used the date dimension to filter away uninteresting sentences by paraphrasing and defined two different ranking measures: interest and burstiness, to select top-ranked informative sentences.

- The LDALR system outperforms the Chieu sys-

tem. This may be for the fact that Chieu’s method is actually based on flat clustering-based summarization, which is not as effective as LDA topic model to extract latent sub-events.

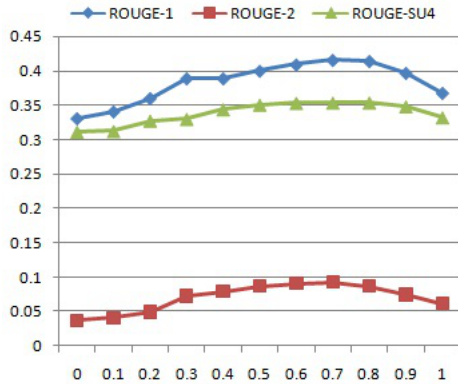


Figure 4: Examine the performance of the balance parameter χ

- The MEALR system outperforms the LDALR system. This may be mainly for the reason that MEALR utilizes the mixture-event-aspect model to detect the more salient sub-events based on the sub-whole relationship, which seems to satisfy users’ bias to different sub-events.

- The MEALR system which utilizes our method outperforms all the baselines, indicating the effectiveness of detecting different types of sub-events with mixture-event-aspect model and the necessity to distinguish different proportions of the component summaries based on local/global aspects.

4.5 Parameter Tuning

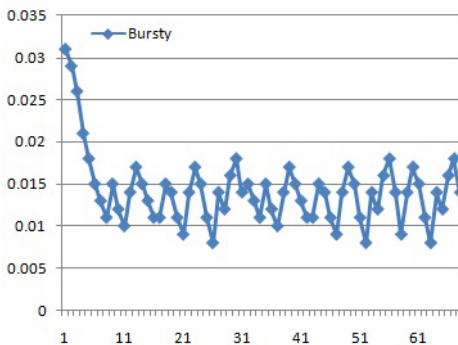


Figure 5: Aspect sequence of event “Connecticut school shooting” (X-axis is the number of days after Dec. 14, 2012. Y-axis is $bursty(A_{k,t})$)

In this section, we compare the performance of the parameters. The hyper-parameters such as α , β , γ , λ can be estimated using standard methods introduced by Minka (Minka, 2000). So we mainly examine the local/global optimization balance parameter χ . We try to evaluate the influence of this parameter on the three kinds of ROUGE measure results respectively. Figure 4 shows the performance of the balance parameters χ . It is obvious that when the balance parameter χ is set to 0.7 this method performs best.

4.6 Sample Output and Case Study

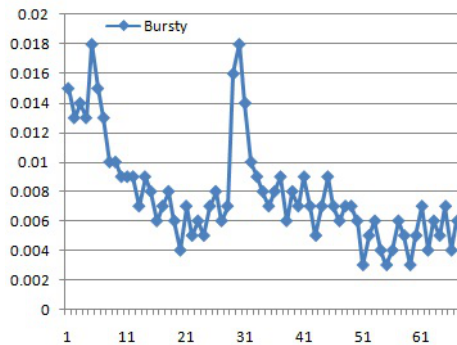


Figure 6: Bursties of sub-event “Gun control debate” (X-axis is the number of days after Dec. 14, 2012. Y-axis is $bursty(A_{k,t})$)

We take the event “Connecticut school shooting” as an example to show the usefulness of our method. Figure 5 shows the aspect sequence based on the bursty periods of all aspects. We select a sub-event “Gun control debate” and Figure 6 shows the bursties of this sub-event on the whole timeline. Table 4 shows part of the storylines for the event “Connecticut school shooting” generated by human and our method. Through observation, we find that the peak of the event “Connecticut school shooting” is around the date when it occurred, and the sub-event “Gun control debate” has two bursty periods around the two peaks. Compared with the human summary, our framework can extract the important sub-events contained in the collection, and satisfy users’ interest on different sub-events based on the part-whole relationship with the event subject.

From the sample output and the human storylines, we also get some observations. (1) The component summary of global aspect tend to share larger pro-

Table 4: Selected part of storyline generated by MEAOp and human

Storyline Generated by human:	
December 14, 2012	Global Aspect
Shooting massacre occurred in a primary school in Connecticut town, Sandy hoot Newton. 20-year-old Adam lanza broke into the primary school after shotting his mother, and in 10 minutes shot more than 100 times, killing twenty children and eight adult, including himself. The youngest death was a children in preschool students.	
December 15, 2012	Global Aspect
Photos of the teachers and students shoot in the Connecticut massacre are released as well as the shooter's. The shooter was very smart but lonely.	
December 16-17, 2012	Local Aspect
President Obama arrived in the locality of school shooting, mourned for the victims and made a speech.	
December 18-20, 2012	Local Aspect
American gun control bill was put on the agenda again.	
Storyline Generated by MEAOp:	
December 14, 2012	Global Aspect
Children and adults gunned down in Connecticut school massacre. 20 children, six adults and the shooter are dead after shooting at Sandy Hoot Elementary School in Newtown, Connecticut. Three law enforcement officials say Adam Lanza, 20, was the shooter, and that he died apparently by his own hand. Suspect's mother, Nancy Lanza, found dead in suspects home in Newtown, law enforcement source says. Ryan Lanza, older brother of Adam Lanza, questioned by police but not labeled a suspect.	
December 15-16, 2012	Global Aspect
Victims' names released Saturday; all of the slain children were either 6 or 7 years old. Understanding school lockdowns in regards to Connecticut shooting. Connecticut gunman recalled as intelligent but remote.	
December 17, 2012	Local Aspect
President obama leads interfaith preyer vigil in newtown connecticut. A tearful Obama says "we've endured too many of these tragedies".	
December 18-20, 2012	Local Aspect
Moderate dems join gun control debate call for commission on us violence gains. Gun debate gains traction as some lawmakers say its time to act.	

portion in the final storyline. This is mainly for the reason that when researching for an event subject, users bias more to the information about the global-sub-events that have closely connection and coincident properties with the major event based on the part-whole relationship. So it is really necessary to distinguish different sub-events with distinctive properties. (2) Our system performs better for the persistent event, such as "The U.S. presidential election". This may be for the fact that these events are usually long running and have more global-sub-events than local-sub-events.

5 Conclusion

In this work, we study the task of event storyline generation and present a novel method. We innovatively introduce the properties of different sub-events based on word co-occurrences to determine the part-whole relationship with the major event and develop a mixture-event-aspect (MEA) model to formalize different types of sub-events into local/global aspects. Based on these local/global aspects, we utilize an optimization method to get the optimal com-

ponent summaries along the aspect sequence. We conduct experiments with our method and various baselines on real web datasets. Through our experiments we notice that our method generates overall better storyline than other baselines. This indicates the effectiveness to detect different types of sub-events with the proposed mixture-event-aspect model and the necessity to distinguish different proportions of the component summaries based on local/global aspects.

Acknowledgments.

We thank the anonymous reviewers for their valuable and constructive comments. This work is financially supported by NSFC under the grant NO.61272340 and 60933004.

References

- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.
- Jackie Chi Kit Cheung, Giuseppe Carenini, and Raymond T Ng. 2009. Optimization-based content se-

- lection for opinion summarization. In *Proceedings of the 2009 Workshop on Language Generation and Summarisation*, pages 7–14. ACL.
- Hai Leong Chieu and Yoong Keok Lee. 2004. Query based event extraction along a timeline. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 425–432. ACM.
- G. Erkan and D.R. Radev. 2004. Lexpagerank: Prestige in multi-document text summarization. In *Proceedings of EMNLP*, volume 4.
- Thomas Hofmann. 1999. Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 289–296. Morgan Kaufmann Publishers Inc.
- Giridhar Kumaran and James Allan. 2004. Text classification and named entities for new event detection. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 297–304. ACM.
- Theodoros Lappas, Benjamin Arai, Manolis Platakis, Dimitrios Kotsakos, and Dimitrios Gunopulos. 2009. On burstiness-aware search for document sequences. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 477–486. ACM.
- Peng Li, Jing Jiang, and Yinglin Wang. 2010. Generating templates of entity summaries with an entity-aspect model and pattern mining. In *Proceedings of the 48th annual meeting of the Association for Computational Linguistics*, pages 640–649. ACL.
- C.Y. Lin and E. Hovy. 2003. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 71–78. ACL.
- Chen Lin, Chun Lin, Jingxuan Li, Dingding Wang, Yang Chen, and Tao Li. 2012. Generating event storylines from microblogs. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 175–184. ACM.
- Juha Makkonen, Helena Ahonen-Myka, and Marko Salmenkivi. 2004. Simple semantics in topic detection and tracking. *Information Retrieval*, 7(3-4):347–368.
- Qiaozhu Mei and ChengXiang Zhai. 2005. Discovering evolutionary theme patterns from text: an exploration of temporal text mining. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 198–207. ACM.
- Q. Mei, J. Guo, and D. Radev. 2010. Divrank: the interplay of prestige and diversity in information networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1009–1018. ACM.
- Thomas Minka. 2000. Estimating a dirichlet distribution.
- Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The pagerank citation ranking: bringing order to the web.
- Yulong Pei, Wenpeng Yin, et al. 2012. Generic multi-document summarization using topic-oriented information. In *PRICAI 2012: Trends in Artificial Intelligence*, pages 435–446. Springer.
- D.R. Radev, H. Jing, M. Styś, and D. Tam. 2004. Centroid-based summarization of multiple documents. *Information Processing & Management*, 40(6):919–938.
- Mauricio GC Resende and Renato F Werneck. 2004. A hybrid heuristic for the p-median problem. *Journal of heuristics*, 10(1):59–88.
- Ivan Titov and Ryan McDonald. 2008. Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th international conference on World Wide Web*, pages 111–120. ACM.
- X. Wan. 2008. Document-based hits model for multi-document summarization. *PRICAI 2008: Trends in Artificial Intelligence*, pages 454–465.
- Xuanhui Wang, ChengXiang Zhai, Xiao Hu, and Richard Sproat. 2007. Mining correlated bursty topic patterns from coordinated text streams. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 784–793. ACM.
- Xiang Wang, Kai Zhang, Xiaoming Jin, and Dou Shen. 2009. Mining common topics from multiple asynchronous text streams. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 192–201. ACM.
- Dingding Wang, Tao Li, and Mitsunori Ogihara. 2012. Generating pictorial storylines via minimum-weight connected dominating set approximation in multi-view graphs. *Proceedings of AAAI 2012*.
- Rui Yan, Liang Kong, Congrui Huang, Xiaojun Wan, Xiaoming Li, and Yan Zhang. 2011a. Timeline generation through evolutionary trans-temporal summarization. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 433–443. ACL.
- Rui Yan, Xiaojun Wan, Jahna Otterbacher, Liang Kong, Xiaoming Li, and Yan Zhang. 2011b. 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*, pages 745–754. ACM.