Ranking News Events by Influence Decay and Information Fusion for Media and Users

Liang Kong, Shan Jiang, Rui Yan, Shize Xu, Yan Zhang*
School of Electronics Engineering and Computer Science, Peking University
Beijing 100871, China
{kongliang, jsh, r.yan, xsz}@pku.edu.cn zhy@cis.pku.edu.cn

ABSTRACT

In many cases, people would like to read the news with great importance on the Internet. However, what users can grasp covers a very small part compared with the huge amount of news which never stops increasing. In this paper, we try to find what users are most likely to be interested in. We notice that media focus plays an essential role in distinguishing news topics and user attention is also an important factor. Therefore, we first propose five strategies which only exploit media focus to decide news influence impact. Then we provide three strategies to combine user attention with media focus. Meanwhile, we also take four types of interaction between user attention and media focus into consideration. To the best of our knowledge, this is the first work to establish different models for computing influence decay of news topics. Experiments show that better influence scores will be achieved by a decay algorithm based on Ebbinghaus forgetting curve and information fusion by considering interactions between user attention and media focus.

Categories and Subject Descriptors

H.3.5 [**INFORMATION STORAGE AND RETRIEVAL**]: On-line Information Services—*Web-based services*

Keywords

News Ranking, Influence Decay, Media-user Interaction

1. INTRODUCTION

Currently news flood which comes from different websites spreads throughout the web. News agencies update their websites every few minutes and plenty of the latest contents are generated unceasingly. Without proper arrangement, readers can be easily trapped in news ocean, and wonder where to access. On some websites, the mess-up information could be gathered together, turned into news event, news issue, news topic or news special manually by editors, and shown to readers. *Yahoo! News Topics*[1] and *CNN Special Coverage* & *Hot Topics*[2], for instance, can deliver all above-mentioned service. With the development of the techniques of Topic Detection and Tracking (TDT)[1], some web service can gather news information from different websites and structure it into news topics which are constructed online automatically and updated temporally, such as *Google News*[3].

The huge amount of news articles available online reflects the users' need for a plurality of information and opinions. TDT makes it feasible for users to know "what's happening" and "what's new". Nevertheless, it is still impossible for users to read them all. So ranking news events which can provide the most valuable and influential news events is a valuable research subject. However, a new generated article barely has links pointing to it. So traditional link analysis techniques can not work well. What's more, the features which can be extracted from a news article are limited. So ranking news events is also a challenging and novel work.

According to the aging theory[4] for news event life-cycle modeling, the influence of a news event is to decay over time. The biological aging rate is not a constant ratio. In the same way, the decay value of a news event influence should not be a fixed parameter. The rate of decay is actually affected by current and past situations of a news event. In this paper, we establish different models to compute influence decay and discuss their performance.

Influence of a news event is the relative importance of an event over time, which is mainly determined by media focus and user attention. Media focus indicates the media report frequency during recent period. User attention indicates the user acceptability level. According to previous work [15], there is inconsistency between these two factors.

In this paper, we will discuss the problem of ranking news events. We propose several schemes to calculate influence of a news event based on the quantity of media reports and user comments. Besides, we discuss the interaction between the media focus and user attention. In order to demonstrate the effectiveness of different schemes, we conduct various experiments on real datasets and evaluate different schemes.

2. INFLUENCE DECAY WITH MEDIA FOCUS

A news topic usually has many related reports and the quantity of reports is an important factor, which indicates media focus. By analyzing features of chronological reports on a timeline, we can see the news topic's evolution tendency.

*Corresponding author

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.

[1]`http://news.yahoo.com/topics`
[2]`http://edition.cnn.com/SPECIALS`
[3]`http://news.google.com`

2.1 General Schemes and Constant Decay Scheme

To rank news topics, we should design a function to map the information of media to an influence value. Generally, this function should meet the following conditions:

- $0 \leq f(x) \leq 1$.
- $f(x)$ is a strictly increasing function of x.
- $f(\infty) = 1$ and $f(0) = 0$.

In this paper, we adopt a sigmoid function which can meet above conditions well. Analogous to that used in [4], it is defined as follows:

$$f(x) = \begin{cases} \dfrac{\delta x}{1+\delta x}, & x > 0; \\ 0, & otherwise. \end{cases}$$

where δ is a smoothing parameter.

Starting from the above function, two simple schemes will be presented. The first one is "the frequency scheme" (FS), which maps the frequency of a news topic reported during a time unit to an influence value. For a topic ε, r_d^ε is the report-frequency of ε during the dth day, and

$$I_d^\varepsilon = f(r_d^\varepsilon),$$

where I_d^ε represents the influence value of the topic ε.

Obviously, consecutive new reports will increase the influence of the topic. So we propose "the accumulative scheme" (AS) and use s_d^ε to represent the accumulative-report-support:

$$I_d^\varepsilon = f(s_d^\varepsilon) = f(\sum_{i=1}^{d} r_i^\varepsilon).$$

Users usually focus on the latest news topics, so we must consider decay function although the report-support r_d^ε may still increase. In [15], decay function is simply a constant decay factor β_c, which is determined by experiments with training data. We borrow this factor and call this scheme "the constant decay scheme" (CDS):

$$s_d^\varepsilon = \begin{cases} \sum_{i=1}^{d}(r_i^\varepsilon - \beta_c) = (s_{d-1}^\varepsilon - \beta_c) + r_d^\varepsilon, & s_d^\varepsilon > 0; \\ 0, & s_d^\varepsilon < 0. \end{cases}$$

2.2 Dynamic Decay Scheme Based on Current Situation

CDS regards a news topic as a life form and its influence decays linearly. However, we think the decaying tendency is related to the topic's current situation. For a news topic ε, when the first report is published, we say ε is born and the original report-support $r_1^\varepsilon = 1$. After d day's accumulation, the topic has a sum of all report-supports $\sum_{i=1}^{d} r_i^\varepsilon$. We define $\bar{g}_d^\varepsilon$ as the average growth rate of the report-support sum from the topic's birthday to the dth day:

$$r_1^\varepsilon(1+\bar{g}_d^\varepsilon)^d = \sum_{i=1}^{d} r_i^\varepsilon. \tag{1}$$

As mentioned above, the accumulative-report-support s_d^ε can be divided into decaying part and developing part:

$$s_d^\varepsilon = s_{d-1}^\varepsilon \times decay(d) + r_d^\varepsilon, \tag{2}$$

where $decay(d)$ is a decay function represented as:

$$decay(d) = (1-\beta_d^\varepsilon)^\lambda. \tag{3}$$

λ is a constant determined by experiments. β_d^ε is the decay rate and is positively correlated to $\bar{g}_d^\varepsilon$. To map $\bar{g}_d^\varepsilon$ to β_d^ε, a sigmoid function is employed. From Equation (1), we have

$$\beta_d^\varepsilon = \frac{\bar{g}_d^\varepsilon}{1+\bar{g}_d^\varepsilon} = 1 - \frac{1}{\sqrt[d]{\sum_{i=1}^{d} r_i^\varepsilon}}. \tag{4}$$

Substituting Equation (3), (4) into Equation (2), we have

$$s_d^\varepsilon = s_{d-1}^\varepsilon \times \frac{1}{(\sqrt[d]{\sum_{i=1}^{d} r_i^\varepsilon})^\lambda} + r_d^\varepsilon, \tag{5}$$

and $s_0^\varepsilon = 0$. We name this scheme as "the dynamic decay scheme based on current situation" (DDSCS).

2.3 Decay Scheme Based on Ebbinghaus Curve

From the equations in Section 2.2, we can have

$$s_d^\varepsilon = \sum_{i=1}^{d-1}(r_i^\varepsilon(\prod_{j=i+1}^{d}(\frac{1}{\sqrt[j]{\sum_{k=1}^{j} r_k^\varepsilon}})^\lambda)) + r_d^\varepsilon. \tag{6}$$

We define the sum decay function $sumofdecay(x)$, where x is the interval from $(d-x)$th day to dth day. $\prod_{j=i+1}^{d}(\frac{1}{\sqrt[j]{\sum_{k=1}^{j} r_k^\varepsilon}})^\lambda$ can be regarded as a kind of $sumofdecay(x)$ and

$$s_d^\varepsilon = (\sum_{i=1}^{d-1} r_i^\varepsilon sumofdecay(d-i)) + r_d^\varepsilon. \tag{7}$$

Since Equation (6) is too complex to compute, we improve accumulation by modifying the sum decay function. The function must meet the following conditions: i) $sumofdecay(x)$ is a strictly decreasing function of x; ii) $sumofdecay(+\infty) = 0$. We adopt a log function, which is defined as follows:

$$sumofdecay(x) = \alpha_e \ln(\frac{\beta_e}{x}),$$

where both α_e and β_e are fixed parameters.

The prototype of this function is used in the research field of memory, which is the famous Ebbinghaus forgetting curve[6]. As the reports have blown over, the influence will decline, just like memory loss. So we use the log function to fit the curve as the sum decay function and name this scheme as decay scheme based on Ebbinghaus curve (DSEC):

$$s_d^\varepsilon = (\sum_{i=1}^{d} \alpha_e r_i^\varepsilon \ln(\frac{\beta_e}{d-i})) + r_d^\varepsilon.$$

3. INFORMATION FUSION FOR MEDIA FOCUS AND USER ATTENTION

User attention indicates the user acceptability level. In this section, we adopt user comments to reflect users' attention and discuss the schemes of combination of media focus and user attention.

3.1 Linear Combination Scheme

Intuitively, frequent comments indicate strong influence. Calculating user attention based on schemes mentioned in Section 2 and linearly combining it with media focus, the influence I_d^ε can be computed as:

$$I_d^\varepsilon = \mu IU_d^\varepsilon + (1-\mu) IM_d^\varepsilon \tag{8}$$

where IM_d^ε represents the influence from media focus, and IU_d^ε represents the influence from user attention. $0 \leq \mu \leq 1$. This scheme is called linear combination scheme (LCS).

3.2 Fusion Schemes Based On Interaction between Media Focus and User Attention

In fact, there is interaction between media focus and user attention. According to previous work [15], inconsistency exists between media focus and user attention. The interaction is caused by the inconsistency and related to variation of the media focus and user attention. The interaction can be divided into four cases:

Thriving: both media focus and user attention increase.
Crashing: both media focus and user attention decrease.

Pushing: media focus increases but user attention decreases.
Fading: media focus decreases but user attention increases.

In the case of Thriving, a news event has become a social focus and the influence should be promoted. On the contrary, the influence will drop faster in the case of Crashing. The situation is more complicated in the case of Pushing or Fading and we have to make clear which tendency dominates, increasing or decreasing.

For a news event ε, let r_d^ε and c_d^ε denote the report and the comment frequency of it on the dth day respectively. Δr_d^ε and Δc_d^ε are the variation of the frequency:

$$\Delta r_d^\varepsilon = r_d^\varepsilon - r_{d-1}^\varepsilon, \qquad \Delta c_d^\varepsilon = c_d^\varepsilon - c_{d-1}^\varepsilon,$$

where $r_0^\varepsilon = 0$ and $c_0^\varepsilon = 0$.

The influence from media focus IM_d^ε and user attention IU_d^ε can be respectively represented as:

$$IM_d^\varepsilon = f(sm_d^\varepsilon), \qquad IU_d^\varepsilon = f(su_d^\varepsilon),$$

where sm_d^ε and su_d^ε are the accumulative support of media reports and user comments respectively. Let binary function $I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon)$ represent interaction of media focus and user attention. Based on above discussion, we can use the interaction to affect on the variety of the influence. So we improve the IM_d^ε and IU_d^ε, and get

$$IM_d^{\varepsilon'} = f(sm_d^\varepsilon \cdot I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon)), \quad IU_d^{\varepsilon'} = f(su_d^\varepsilon \cdot I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon)).$$

Finally we have

$$I_d^\varepsilon = \mu IU_d^{\varepsilon'} + (1-\mu) IM_d^{\varepsilon'} \tag{9}$$

$I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon)$ must meet the following conditions to fit the four cases of inconsistency:

- $I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon) > 1$ when $\Delta r_d^\varepsilon > 0, \Delta c_d^\varepsilon > 0$.
- $I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon) < 1$ when $\Delta r_d^\varepsilon < 0, \Delta c_d^\varepsilon < 0$.
- $I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon)$ is a strictly increasing function of $\Delta r_d^\varepsilon, \Delta c_d^\varepsilon$.
- $0 < I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon) < c$, where c is constant.

Logistics function models the "S-shaped" curve of growth of some set. We employ logistics function to separately calculate tendency of variation of media report and user attention. After that, we merge the tendencies. Since $I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon)$ should not excessively change the accumulative support, we restrict $I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon)$ within $[0.5, 1.5]$. We define

$$I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon) = 0.5 \times \left(\frac{\lambda_r e^{\Delta r_d^\varepsilon}}{1+\lambda_r e^{\Delta r_d^\varepsilon}} + \frac{\lambda_c e^{\Delta c_d^\varepsilon}}{1+\lambda_c e^{\Delta c_d^\varepsilon}}\right) + 0.5, \tag{10}$$

where λ_r and λ_c are two smoothing parameters. By substituting Equation (10) into Equation (9), the influence can be calculated and the method is named as linear combination scheme based on interaction (LCSI).

3.3 Media Priority Combination Scheme Based on Interaction

Since users read the news after they appear, there is always a delay from media report to users' comments. Based on this, it is reasonable to suppose that the media focus is prior in determining the influence of a news event. Thus we propose a media priority combination scheme.

Let us reexamine Equation (5), λ can affect rate of influence decay. As mentioned before, the inconsistency of media focus and user attention have four cases. When the quantities of media focus and user attention increase, λ should be adjusted to low to postpone decay. When both of them decrease, λ should be adjusted to high to promote decay.

The problem is how we can calculate λ. We define

$$\lambda = \phi(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon) \tag{11}$$

where $\phi(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon)$ is a function, related to variation of media focus and user attention. $\phi(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon)$ should have the opposite properties to $I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon)$. So we let

$$\phi(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon) = \beta - I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon) \tag{12}$$

where β is a constant and represents fixed basis rate of decay. Finally, we have

$$s_d^\varepsilon = s_{d-1}^\varepsilon \times \left(\frac{1}{(\sqrt[d]{\sum_{i=1}^d r_i^\varepsilon})^\lambda}\right)^{\beta - I(\Delta r_d^\varepsilon, \Delta c_d^\varepsilon)} + r_d^\varepsilon \tag{13}$$

This method is called media priority combination scheme based on interaction (MPCSI).

4. EXPERIMENTS AND DISCUSSIONS

4.1 Datasets

Experiments are performed on three datasets. The first dataset (SES) is constructed from Google News Archives. We crawl the reports using a query list which mentions mainstream news events during a period of time. Since it is difficult to get comments from search engine, we construct another dataset from web portals, namely WPS from news web portals (such as SOHU[4], Tencent[5], which are the most influential news websites in China). The third dataset (TS) also comes from the same resources as WPS, which publishes earlier than SES and WPS. A formal training/test division is necessary to justify the experiment results. So we use TS as our training set. Our method is language-independent since it only need to use the quantity of reports and comments.

Table 1 shows statistics of the datasets.

Table 1: Statistics of datasets

Dataset	Event	Report	Comment	Time Span
SES	133	1,361,776	0	Jan 31 - Mar 1, 2010
WPS	112	138,557	1,391,112	Jan 1 - Mar 31, 2010
TS	24	20,594	184,389	July 20 - Aug 31, 2009

4.2 Evaluation

Regarding some day's influential events as a query, we can compare events ranking to an information retrieval problem. Thus we can employ DCG[9] to evaluate different schemes. We have seven volunteers (no author involved) who have their own investigation method (For example, a volunteer may have his own investigation team). The grade for the topics is limited to three levels:

- **Weak Influence(1 Point)**: Readers never hear the topic, or think the topic is boring.
- **Common Influence(2 Points)**: Readers know the topic, but they do not care it.
- **Powerful Influence(3 Points)**: Readers are attracted by the topic.

We define DCG value at position r on d-th day as $DCG_d@r$:

$$DCG_d@r = \sum_{j=1}^{r} \frac{2^{r(j)-1}}{\log_2(j+1)},$$

where $r(j)$ is rating of the event ranked at the j-th position, according to the ground truth label. We define *Average DCG* of a scheme at r position during some period:

$$Average\ DCG@r = \frac{\sum_{d=Start\ Day}^{End\ Day} DCG_d@r}{number\ of\ days}.$$

4.3 Results

4.3.1 Results of Using Media Reports Only

We perform the first experiment using only media reports. Five schemes are employed in SES.

[4]http://www.sohu.com
[5]http://www.qq.com

We calculate the average DCG value of the five schemes and DSEC outperforms others. We also conduct t-test on the improvements in terms of Average DCG and the results indicate that the improvements of DSEC over others are statistically significant (p-value<0.05). Figure 1 shows the trends of $DCG_d@8$ of five schemes from February 1 to February 28, 2010. For most days, DSEC delivers DCG@8 values higher than other schemes.

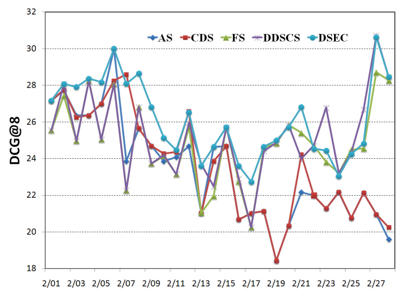


Figure 1: Trends of the $DCG_d@8$ of five schemes from February 1 to February 28, 2010

4.3.2 Results of Combination of Media Reports and User Comments

In this part, we conduct the experiment using both media reports and user comments. Owing to using user comments, the experiments are mainly performed in WPS. DSEC is used to compute the influence of media and user. All the parameters used will be discussed in Section 4.5.

Table 2 demonstrates the Average DCG values of LCS, LCSI, MPCSI and DSEC from 1-st position to 8-th position, respectively. We can see that LCSI performs better than other schemes. LCS and LCSI achieve much better performance than DSEC. The t-test result shows that the improvements are statistically significant (p-value<0.05). We further check the performance of DSEC and LCSI on the level of each time unit (one day in this paper) and find LCSI outperforms DSEC significantly most of the time. It illustrates that interaction between media focus and user attention is an effective feature to compute influence of events. However MPCSI dosen't perform better than DSEC. Maybe media focus is stressed too much in MPCSI.

Table 2: Average DCG values of LCSI, LCS, MPCSI and DSEC in WPS from March 1 to March 31, 2010. (@r represents Average DCG at r-th position)

Average DCG	DSEC	MPCSI	LCS	LCSI
@1	7.19	7.19	7.32	**7.84**
@2	10.83	10.80	10.88	**11.77**
@3	13.66	13.55	13.82	**14.45**
@4	15.67	15.41	16.00	**16.55**
@5	17.25	17.11	17.83	**18.57**
@6	18.71	18.75	19.61	**20.16**
@7	20.10	20.17	21.06	**21.64**
@8	21.19	21.41	22.31	**22.98**

4.4 Performance Evaluation

4.4.1 Performance of Using Media Reports Only

Owing to space constraints, it is impossible to illustrate the performance of five schemes in all the events. We choose typical styles of events and analyze the performance.

Figure 2(a) is the event about Health Care Reform. It is considered as a stable event because a lot of reports are published almost every day. Figure 2(b) shows a bursting event, which is about death of Haig. The reports erupt when the event happened. Figure 2(c) is the event about Ukraine's election, which is a wave event. There are two phases - one is that start to vote and the other one is that Yulia Tymoshenko drops Ukraine election challenge.

Obviously, AS has poor performance and it will lead to a continuous high influence when a high frequency exists on some day. Similarly, CDS does not work well with a fixed parameter representing the decay. FS is based on report frequency and the influence varies rapidly with report frequency every day. However, when the frequency declines fast, the influence will disappear immediately. In fact, the influence should be maintained for a while. DDSCS improves the performance of FS, but the trend of rapid varying with frequency still exists. DSEC is well performed in all events. The change of influence seems to be more smooth. Besides it indicates the event evolving trends with report frequency.

4.4.2 Performance of Combination of Media Reports and User Comments

Figure 3 shows the performance of the three combination schemes of media report and user comment. To clearly show the chart, we normalize the number of reports and comments which are based on second Y-axis.The event talks about Haiti earthquake.

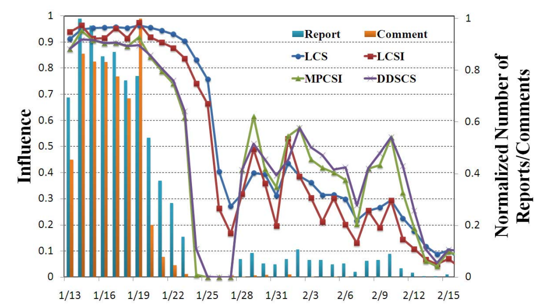


Figure 3: Fusion Schemes about Haiti Earthquake in WPS

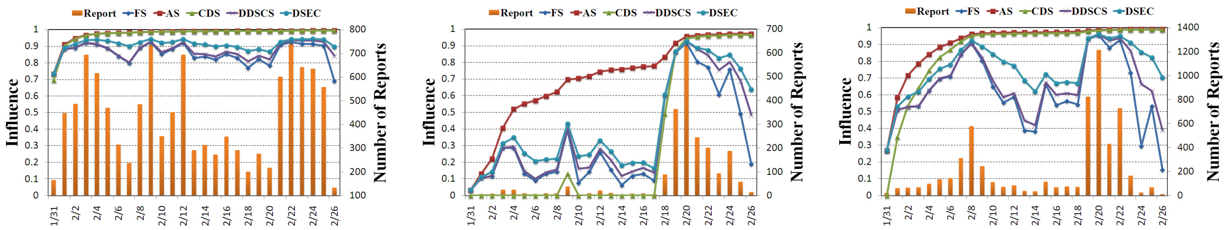
LCSI and MPCSI import interaction of media and user into LCS and DDSCS, respectively. Comparing LCSI with MPCSI, during Jan 25 to Jan 27, influence based on MPCSI nearly decreases to 0, but the influence based on LCSI is higher. Because MPCSI is reports prior and comments impact is weaker. But LCS and LCSI still keep high influence value when the number of reports approaches 0 but there are still a lot of comments. Generally, LCSI performs better than DDSCS and MPCSI.

4.5 Parameter Tuning

In this section, we discuss parameter tuning. α_e and β_e are set to be 0.3 and 7 respectively. δ, λ_c and λ_r don't affect the result of ranking significantly and proper values are used to conduct the experiments. μ, λ and β need to be adjusted based on practical application scheme. We repeatedly cycle through each parameter in TS, holding all others fixed while optimizing it. Finally we choose $\mu = 0.5$, $\lambda = 4$ and $\beta = 5$.

5. RELATED WORK

Topic detection and tracking (TDT) in news has been extensively studied in previous work. Pilot work includes [16] and [2]. Others follow their work on event detection [3],



(a) Decay on Obama Health Care Reform in SES (b) Decay on Haig Dead at 85 in SES (c) Decay on Ukraine Election 2010 in SES

Figure 2: Performance comparison

novelty detection [17], etc. Features such as named entities [18], date and place information [13] and term trajectory are analyzed [7]. Our work focuses on how to measure the news importance rather than how to find news events.

[4] proposes an aging theory to model the life cycle of sequential events, which incorporates a traditional single-pass clustering algorithm to detect and track events. The thought of aging theory is also employed in our work. In [8], He et al. rank the events by the frequencies of event reported in time units and the number of consecutive effective time units. However their approach need to know the “ascending-stage” and “steady-stages”, and it is inadequate to be applied online. In [5], a ranking algorithm is proposed to find the most authoritative news sources and identify the most interesting events in the different categories to which news article belongs. In [14], a three-step automatic online algorithm for news issue construction is proposed, and two simple rules are separately used for ranking news events. The first rule is time preference, and the second rule is quantity preference. In their next work [15], an automatic online news topic ranking algorithm is proposed based on inconsistency analysis between media focus and user attention. However, it only considers the constant decay in the aging model. Besides, it ignores the interaction between media focus and user attention.

6. CONCLUSION

In this paper, we construct an automatic news topic ranking framework. We not only take both media focus and user attention into consideration, but also bring in the interaction between them to our framework. Experimental results show that LCSI as a typical scheme using both reports and comments with interaction impact and DSEC as a typical scheme using reports only.

As a future work, we will consider some social network features to improve the ranking schemes, since such features have been widely used [10, 11, 12] in hot news detection.

7. ACKNOWLEDGMENT

We thank the anonymous reviewers for their valuable and constructive comments. This work is supported by NSFC with Grant No. 61073081 and 60933004.

8. REFERENCES

[1] J. Allan, J. Carbonell, G. Doddington, J. Yamron, and Y. Yang. Topic Detection and Tracking Pilot Study: Final Report. In *Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*, volume 1998, pages 194–218, 1998.

[2] J. Allan, R. Papka, and V. Lavrenko. On-line New Event Detection and Tracking. In *SIGIR’98*, pages 37–45, 1998.

[3] T. Brants, F. Chen, and A. Farahat. A System for New Event Detection. In *SIGIR’03*, pages 330–337, 2003.

[4] C. Chen, Y. Chen, Y. Sun, and M. Chen. Life Cycle Modeling of News Events using Aging Theory. In *ECML’03*, pages 47–59, 2003.

[5] G. Del Corso, A. Gullí, and F. Romani. Ranking A Stream of News. In *WWW’05*, pages 97–106, 2005.

[6] H. Ebbinghaus. *Memory: A Contribution to Experimental Psychology*. Teachers College, Columbia University, 1913.

[7] Q. He, K. Chang, and E. Lim. Analyzing Feature Trajectories for Event Detection. In *SIGIR’07*, pages 207–214, 2007.

[8] T. He, G. Qu, S. Li, X. Tu, Y. Zhang, and H. Ren. Semi-automatic Hot Event Detection. *ADMA’06*, pages 1008–1016, 2006.

[9] K. Jarvelin and J. Kekalainen. Cumulated Gain-based Evaluation of IR Techniques. *ACM TOIS*, 20(4):422–446, 2002.

[10] Y. Lee, H. Jung, W. Song, and J. Lee. Mining the blogosphere for top news stories identification. In *SIGIR’10*, pages 395–402, 2010.

[11] R. McCreadie, C. Macdonald, and I. Ounis. A Learned Approach for Ranking News in Real-Time Using the Blogosphere. In *SPIRE’11*, pages 104–116, 2011.

[12] X. Shuai, X. Liu, and J. Bollen. Improving news ranking by community tweets. *WWW 2012 - MSND’12 Workshop*, 2012.

[13] D. Smith. Detecting Events with Date and Place Information in Unstructured Text. In *JCDL’02*, pages 191–196, 2002.

[14] C. Wang, M. Zhang, S. Ma, and L. Ru. Automatic Online News Issue Construction in Web Environment. In *WWW’08*, pages 457–466, 2008.

[15] C. Wang, M. Zhang, L. Ru, and S. Ma. Automatic Online News Topic Ranking using Media Focus and User Attention based on Aging Theory. In *CIKM’08*, pages 1033–1042, 2008.

[16] Y. Yang, T. Pierce, and J. Carbonell. A Study of Retrospective and On-line Event Detection. In *SIGIR’98*, pages 28–36, 1998.

[17] Y. Yang, J. Zhang, J. Carbonell, and C. Jin. Topic-conditioned Novelty Detection. In *SIGKDD’02*, pages 688–693, 2002.

[18] K. Zhang, J. Zi, and L. Wu. New Event Detection based on Indexing-tree and Named Entity. In *SIGIR’07*, pages 215–222, 2007.