Journal of Intelligent & Fuzzy Systems xx (20xx) x–xx
DOI:10.3233/JIFS-179479
IOS Press

Big data metrics: Time sensitivity analysis of multimedia news

Lin Gan[a], Shijun Li[a], Zhan Shu[b] and Wei Yu[a,*]
[a]*School of Computer Science, Wuhan University, Wuhan, The Netherlands*
[b]*Department of Computer Science, Binghamton University, Binghamton, USA*

Abstract. With the daily release of the huge amount of new information, the web information such as video, image, audio and text information is growing dramatically. Most of this vast information changes over time. It may become ineffective, obsolete and worthless, and even affect user's understanding of web information, degrade their experience. Therefore, analyzing this temporal-sensitive information in multimedia is a vital issue. To this end, this paper analyses the temporal sensitivity of multimedia web information to find what kind of information is temporal-sensitive, how this information is responsive to time, and how to evaluate the temporal sensitivity of information. We start with four types of features, that is time, content, user behavior and related multimedia news, then set up a triple model to depict news. By establishing the energy transfer relationship between news and related and similar news, time, and user behavior, we measure the energy, and use the change ratio of energy as the temporal sensitivity of news. The data set is a multimedia news corpus, including video, image and text news. In the experiment, we take the users' comments as the validation set. The result basically matches the validation set, and it shows our metric is reasonable.

Keywords: Big data analysis, temporal metric, temporal sensitivity, multimedia news

1. Introduction

Considerable information is being released every day. It makes the web information growing larger. These hypertext and multimedia information can fall into three categories based on the validity period of content, instantaneous information, short-term information and long-term information.

Instantaneous information means the information is effective in a short time. If it delays slightly, its content will be worthless and finally only of value as reference or history, such as video news, broadcast notifications, major social dynamics, stock validity and sudden events. Short-term information refers to the information works for a period, such as public policies, regulations, measures. Long-term information is essentially time-independent and remains useful for a long time, such as concept definition, historical events, physical laws, etc.

Among them, the value and effectiveness of both instantaneous information and short-term information vary over time. They are called temporal-sensitive information. And, information temporal sensitivity refers to some certain properties of information, i.e. value and effectiveness, changing over time. For example (1) News: the value of news varies with time; this kind of information is without a deadline. (2) "Call for papers" of an academic conference: it usually along with some important deadlines, after the deadline, the value of this information would suddenly decrease. (3) Big events: information about big events, will periodically grow and fade, as a reference to some related big events that users concerned very much. Temporal-sensitive

*Corresponding author. Wei Yu, School of Computer Science, Wuhan University, 430072 HB, Wuhan, The Netherlands. E-mail: yuwei@whu.edu.cn.

information may be ineffective, obsolete or worthless to affect user's understanding of web information, degrade their experience and waste their time without proper process. Therefore, analyzing this big temporal-sensitive information is an important issue.

Temporal-sensitive information has been studied for a few years, such as time-sensitive queries in information retrieval, asking time-related questions in question answer, ranking pages with time relevance in time-sensitive search engine. However, few people research into what time-sensitive information is, how this information responsive to time and how to evaluate their temporal sensitivity. To the end, this paper will concentrate on analyzing and evaluating temporal sensitivity of web information.

This research is built on a multimedia news corpus of Sina Web Portal, including video, image and text news. And we take each news as a unit rather than a page. Since pages are the carrier of news, what people interested in is the information on the page, a page without any information is nonsense. The second reason is, one page may carry multiple pieces of information. This is out of our consideration. This method starts with four types of features, time, content and user behaviour and related pages. Then we set up a triple model to depict news. Furthermore, we construct the transfer relationship of energy between news and related content, time, and user behaviour. Finally, this paper measures the energy of news and uses the change ratio of energy as the temporal sensitivity of news.

Thus, the contribution is threefold:

– We proposed a triple model to measure temporal sensitivity of web page.
– Four types of features are adopted to characterize the temporal sensitivity of news. They are (1) Time: include publish time and system time (2) User behaviour: users visiting and comment dynamics. (3) content: topic and temporal information in the news. (4) related news pages: similar topic news pages and linked-in news pages.
– When measuring the temporal sensitivity, we regard the web page as life with energy and use energy changing ratio to represent the temporal sensitivity of each page.

The paper is structured as follows. The related work is reviewed in Section 2. Section 3 describes our methodology for temporal sensitivity metric of news. We report our experiments in Section 4. Finally, Section 5 concludes the paper.

2. Related work

Web pages temporal analysis is about information quality assessment in temporal web, which is related to research work of temporal web, web information quality, web usage and News tracking.

2.1. Temporal web

Recently, Temporal Web turns to be the focus of scholars' concern. Mohammad Sadegh Zahedi et al. [1] introduced temporal property of queries into blog retrieval. They compute the temporal sensitivity of each query in different time intervals and incorporate with relevancy to rank blog posts. Vanessa Pena-Araya et al. [2] proposed a spatio-temporal context-aware representation of news events using social media information. by including social, temporal, and spatial information in the event representation, they are enabling the analysis of historical world news from a social and geopolitical perspective. Akira Yoshihara et al. [3] leveraged temporal effects of past news events to build a recurrent deep neural network, and proposed an approach to predict stock market trend based on this deep learning method. Fernando S. Peregrino et al. [4] made a temporal analysis of how terms evolve in time in both newspaper articles and twitter messages. Although the temporal Web has begun to be concerned and have a certain degree research on some aspects, it has no the same work and definition on Web Temporal sensitivity measurement of news.

2.2. Temporal patterns

Temporal pattern discovery is also widely studied in Temporal Web. Maria CarlaCalzarossa et al. [5] analysed the temporal patterns of content changes of news websites, they modelled and illustrated their dynamics with the trigonometric polynomials and ARMA components. Ida Mele et al. [6] track the news publishing patterns and rank the news streams according to timeliness. They use the average delay of events published by the news stream to define the timeliness of news stream, with the features of the delay to the first news and relative position among news sequence of the same event.Other related work includes exploring temporal aspects of Web behaviors [7] and utilizing the evolution patterns of pages and links to benefit Web based applications [8].

2.3. User behaviours

In recent years, there has been an increase in the number of studies on web usage mining [7–10]. The main motivation behind these studies is to get a better understanding of the reactions and of the users' navigation and then to help them in their interaction with the Web site. Steven Van Canneyt et al. [9] presented a novel method with temporal behaviours to model and predict the popularity of online news. The combination of publication time and date with content and meta-data significantly improved the popularity predictions.Filannino, Michele and Nenadic, Goran [10] tested the effect of user behaviors such as comments, retweets and likes to the news credibility.

2.4. Page quality assessment and timeliness measures

Existing studies have shown that, although some research results reveal the time distribution of network resources on a certain degree [5, 6], few scholars conduct research on temporal sensitivity metric of news page. Currently, the work on Web quality assessment and timeliness measure mainly include:

In our previous research [11, 12], a novel way of Web quality assessment and social annotations have been utilized to improve Web search performance. Damir Korencic et al. [13] study the topic coherence of news and provide a topic coherence measurement based on topic documents rather than topic words.

Timeliness is one of the crucial elements of quality assessment of Web information. The information should be updated with the changing of its resources' value; otherwise it will be out of date. Gam D.Nguyen et al. [14] quantified freshness in terms of the age metric in a communication problem. Rahime Belen Saglam and Tugba Taskaya Temizel [15] proposed a framework to assess the timeliness of health website contents in accordance with current literature disregarding the update time shown on the websites.

In fact, the above evaluation criteria just for the general quality and information timeliness of the webpage or website, and there have no work on the modeling and measurement of Temporal sensitivity of Webpage.

2.5. News corpus

News article is a formal and credible source of web information, a lot of researches are built on this corpus. Young JoonPark et al. [16] try to understand the sentence polarity of news articles on player performances to improve player evaluation in question-answering system in sports. Songqiao Han et al. [17] extract business event from online Chinese news with integrating patterns, machine learning models and word embedding technology in deep learning. Desmond Bala Bisandu et al. [18] use N-grams and improved sqrt-cosine similarity measure to cluster news articles. Myojung Chung [19] examines how social media metrics affect online news evaluations. Mohan Timilsina et al. [20] expand the conventional research citation graph based on news to improve prediction of the academic impact. As for numerous studies carried on news articles, we choose it as our corpus.

3. Methodology

Web information exhibits dynamics as it grows and fades over time. In some time, Web information is effective, while a few days later, it may out of date. In other words, the same sentence may have different value in different time. Being based on this character, we do a web page-level temporal analysis on web information to discover the sensitivity of web information over time. What's more, we assume one web page only carry one piece of news in this paper.

3.1. Framework of temporal sensitivity measurement

The sensitivity of web information to time is not completely objective, it is decided by the temporal effectiveness of web information itself and the demand of users for the information. In order to better model the temporal sensitivity of news, four aspects, the content of value, time and user behaviours, have been taken into consideration to construct the temporal sensitivity measurement framework. Where (1) the content includes topic of news, temporal information in the news and the news with the similar topic. (2) The time indicates publishing time and system time. (3) The user behaviour refers to users' comment and join. (4) related news pages is including similar news with the same topic and the linked-in news page.

Which is given as a triple :

$$TS = f(IV, T, P)$$

where TS is the temporal sensitivity value of news page, IV is the information volume of news page, referring to the content (topic and temporal information) of news and similar news pages. T is short for Timeliness of news with age metric, which means how long the news page has been published. it relates to the time features. and P represents the popularity of news page, including user behaviours and linked-in news pages.

In this framework, information volume IV includes the information volume of temporal word and that of topic in the news page, they depict the temporal effectiveness of news page itself; popularity P which can be characterized by users' visiting, comment and supporting and link relationship among news pages reflect the demand of users for news information; and the time Tindicates the temporal property in information temporal sensitivity.

3.2. Features and representation

Time-line T: In this paper, we assume a discrete representation of time-line, denoted by T, and let the day be an atomic time interval called chronon. Then, the time-line T in this paper can be represent by a non-empty finite set of discrete time points $T = \{t_1, t_2, ..., t_i, ...\}$, where t_i is a discrete time point with a day granularity.

Based on this time-line T, we describe the temporal representation of pages. Set the publishing time of a given page d as $t_0(d)$, $t_0(d) \in T$. define a time minus function $f(t_0(d), t)$ on time-line T as $f(t_0(d), t) = t - t_0(d)$. If $f(t_0(d), t) = t - t_0(d) = 0$means the page d is published on the same day as t, then $t = t_0(d)$. And if $f(t_0(d), t) = n(n > 0)$indicates the page d is published n days before t, then $t = t_n(d)$.

In the temporal sensitivity framework, the information volume aspect is characterized by two features: information volume of temporal word and information volume of topic in the news page:

Information volume of temporal word (IV_{tw}): Given a news page d, we do a word divide process on it, and then it can be split into a word sequence $d = (w_1, w_2, ..., w_{nw})$, nd is the number of words in news page d. Find out all the temporal words and record their locations to re-display the news page as $d_{tw} = (< tw_1, l_1 >, < tw_2, l_2 >, ..., < tw_{ntw}, l_{ntw} >)$, where $tw_i$is i-th temporal word in news page d, l_i is the location of tw_i, it means that, if tw_i is the i'-th word in d (i.e. $tw_i = tw_{i'}$), then $l_i = i'$, $ntw$is the number of temporal words in news page d.

Then the density of temporal words in news page d is $\frac{n_{tw}}{n}$, and the center location of the temporal word is $\mu = \arg\min \frac{\sum_{0<i\leqslant n_{tw}} \|l_i-\mu\|^2}{n_{tw}}$, so the information volume of temporal word is $IV_{tw} = \frac{n_{tw}}{\mu \times nw} = \frac{n_{tw}}{nw \times \arg\min \frac{\sum_{0<i\leqslant n_{tw}} \|l_i-\mu\|^2}{n_{tw}}}$.

Information volume of topic (IV_{topic}): For every news page d in D, we describe its topic as a word set $S_{topic}(d) = \{w_1, w_2, ..., w_m\}$, where w_i is a representative word, m is the number of these representative words the topic contains. Note that each d just only has one topic set $S_{topic}(d)$, and no two words in this topic set are the same, and the number of topic words m varies in terms of the news d. now we have N news page in D, so the topic of each news can be denoted by $S_{topic}(d_i) = \{w_1^i, w_2^i, ..., w_{m(i)}^i\}$. If $S(d) \subseteq D$ is the set of all the news pages that have similar topic with news page d, and every news page d' in S is published later than news page d, which means $t_0(d') > t_0(d)$, $d' \in S(d)$, then the information volume of topic of news page d can be given as $IV_{topic} = \sum_{d' \in s(d)} sim(d, d')f(d')$. Where $f(d')$is the weighted function of news page $d' \in S(d)$, and $sim(d, d')$ is the similarity between news page d and news page d' (similarly hereinafter).

Users support and comment: For each news page d, there is a comment number time series $cs(d) = \{c_1, c_2, ..., c_n\}$and a support number time series $ss(d) = \{s_1, s_2, ..., s_n\}$, where c_i and s_i are the comment number and support number at time point t_i respectively, n that varies as the news changes is the span of the news d on the time-line T, that is the news document d having been published for n days. And $t_i \in T$. So there is a one-to-one correspondence relationship between elements in $cs(d)$ and elements in $ss(d)$. Let $D = \{d_1, d_2, ..., d_N\}$ be a set of web page news. N is the number of web page news. the comment time series and support time series of news document d_i can be donated by $cs(d_i) = \{c_{1i}, c_{2i}, \cdots, c_{n(i)i}\}$ and $ss(d) = \{s_1^i, s_2^i, ..., s_{n(i)}^i\}$ respectively.

Link network: We describe news pages as nodes and link relationship as edge to construct the link network in this paper, denoted by $G = (D, E)$, where G is the link network, D is the set of news pages nodes and E is the set of link relationships. If $d, d' \in D$ are in a link relationship and d links to d', then we can have the link relationship $d \rightarrow d' \in E$.

3.3. Energy function

The features we introduced above are deemed as factors that can increase or decrease the energy of every news page. Before compute the energy function of all the news pages, we have to define the energy that each feature contributes to every news page.

Define the publish time of news page d as $t_0, t_1 \in T$. For page d, the energy that each feature contributes to news page d can be defined as follows:

- Assume that the energy of information volume of temporal word follows the exponential decay. At time point t_k, the energy that the temporal words in news page d contribute is $e_{IV_{tword}}(t_k) = e_{IV_{tword}}(t_0) \times e^{-K \times (t_k - t_0)}$, where K is the parameter of temporal words decay, and $e_{IV_{tword}}(t_0) = IV_{tword}$.
- The increase of information volume of topic, especially for the information overlaying, which is the information volume increasing of the totally the same topic of news page, can reduce the energy of news page, for this reason, we define the energy that the topic of news page d contributes as $e_{IV_{tword}}(t_k) = -\sum_{d' \in S(d)} sim(d, d') f(d') Eng_d(t_{k-1})$, where $Eng_d(t_k - 1)$ is the total energy of news page d at time point t_{k-1}. And when $t_k = t_0$, $e_{IV_{topic}}(t_k) = e_{IV_{topic}}(t_0) = 0$.
- We define the initial energy that the comment number and support number of news page d at time point ti contribute to news page d as $e^i_{join}(t_i) = \alpha c_i + \beta s_i$. Assume all the comments and supports are all exponential decay, in terms of this assumption, at time point t_k, the energy that comments and supports created at time point ti contribute is $e^i_{join}(t_k) = \alpha c_i e^{Kc(t_k - t_i)} + \beta s_i e^{Ks(t_k - t_i)}$, then sum all the energy that comments and supports created from the time point t0 when news page d published to time point t_k contribute to news page d, there is $e_{join}(t_k) = \sum_{i=0}^{k} e^i_{join}(t_k) = \sum_{i=0}^{k} (\alpha c_i e^{Kc(t_k - t_i)} + \beta s_i e^{Ks(t_k - t_i)})$, where α and β are the weighted parameter that show the importance of comments and supports respectively, Kc is the decaying parameter of comments and Ks is the decaying parameter of supports, and c_i and s_i are the comment number and support number generated at time point t_i separately, note that $c_0 = s_0 = 0$, so $e_{join}(t_k) = 0$. (need to be modified)
- Consider the directed link-out network of news page d, we represent it as $G_{out}(d) = (D_{out}(d), E_{out}(d))$, where $D_{out}(d) = \{d' | d \to d'\}$is the set of all the news pages that page d links out to, and $E_{out}(d)$ is the set of all the directed linked-out edges. Similar to the definition of the energy of information volume of topic, all the news pages in $D_{out}(d)$ can reduce the energy of news page d. Accordingly, we can have the energy contributed by all the link-out news pages in $D_{out}(d)$: $e_{linkout}(t_k) = -\sum_{d' \in D_{out}(d)} relate(d, d') f(d') Eng_{d'}(t_{k-1})$, where $relate(d, d')$ is the relatedness between news page d and news page d'(similarly hereinafter). And when $t_k = t_0$, $e_{linkout}(t_k) = e_{linkout}(t_0) = 0$.

Rather, denote the link-in network of news page d as $G_{in}(d) = (D_{in}(d), E_{in}(d))$, where $D_{in}(d) = \{d' | d \to d'\}$is the set of all the news pages that links in to news page d, and $E_{in}(d)$ is the set of all the directed linked-in edges, all the news pages in Din can enhance energy of news page d. as for this consideration, we can have the energy contributed by all the link-in news pages of page d: $e_{linkin}(t_k) = -\sum_{d' \in D_{in}(d)} relate(d, d') f(d') Eng_{d'}(t_{k-1})$. The same as the initial energy value in link-out network, when $t_k = t_0$, $e_{linkin}(t_k) = e_{linkin}(t_0) = 0$.

Eventually, the energy contribution of link network of news page d is:

$$
\begin{aligned}
= \quad & e_{link}(t_k) \\
= \theta & \sum_{d' \in D_{in}(d)}^{\theta e_{linkin}(t_k) + \delta e_{linkout}(t_k)} related\,(d, d') f(d') f(d') Eng_{d'}(t_{k-1}) \\
-\delta & \sum_{d' \in D_{out}(d)} related\,(d, d') f(d') f(d') Eng_{d'}(t_{k-1})
\end{aligned}
\tag{1}
$$

where θ and δ are both the weighted parameters. And when $t_k = t_0$, $e_{link}(t_k) = e_{link}(t_0) = 0$.

To synthesize above energy definition, the final energy contribute by all features to news page d is

$$
Eng_d(t_k) = e_{IV_{tw}}(t_k) + e_{IV_{topic}}(t_k) + e_{join}(t_k) + e_{link}(t_k) \tag{2}
$$

If $t_k = t_0 = 0$, the initial energy value of news page d is

$$
Eng_d(t_0) = e_{IV_{tw}}(t_0) + e_{IV_{topic}}(t_0) + e_{join}(t_0) + e_{link}(t_0) = IV_{tw} \tag{3}
$$

3.4. Temporal sensitivity measurement

As information changes over time, the temporal sensitivity of a page at different time point is also not the same. This paper take the slope of pages' energy as the temporal sensitivity, for instance, at time point t_k, if the energy of page d is $eng(t_k, d)$, then the temporal sensitivity of the page at this point is $\frac{\partial Eng(t_k, d)}{\partial t_k}$.

In this paper, we deem news web pages as a kind of living creature. Each page is of energy, which changes as some factors, such as time goes by and user's concern. The value of this energy of every news web page indicates how old the news web page is, and the slope of the energy shows the temporal sensitivity of news pages.

3.5. Clustering – compute $S(d)$and $sim\,(d, d')$

In the energy function, the energy the topic of news page d contributes relies on the $S(d)$and $sim\,(d, d')$. so in this section, we discuss the how to compute this two items.

3.5.1. Our algorithm to compute $S(d)$ is

1. Build the Vector Spacial Model of news text, and use a word vector $V_{topic}(d)$ to represent every news page. Instead of use Weighted TF/IDF to quantify weights of words in a news text, we compute the cross entropy of each word.
2. Compare news page d against the previous news pages d'in memory. Based on a time window T_{window}, find news pages d'in the time window. Compute the similarity $sim\,(d, d')$between news page $d$and every news page d'in the time window.
3. If$sim\,(d, d')$ is bigger than the threshold δ, then put news page d'into the set $S(d')$and record the $sim\,(d, d')$.

3.5.2. Computing $sim\,(d, d')$

The similarity $sim\,(d, d')$ between news page d and news page d' is calculated as the following way: do numeralization on the topic vectors v_{td} and $v_{td'}$, then take the Euclidean distance between this two vectors as the similarity $sim(d, d')$ between this two pages. We denote it by the equation that.

3.6. Computing $relate\,(d, d')$

$relate\,(d, d')$ represents the link relationship strength between news page d and news page d'. In this paper, we simply set the $relate\,(d, d')$ as follows:

$$relate\,(d, d') = \begin{cases} 0, \textit{There is no link relationship} \\ \textit{between page } d \textit{ and page } d', \\ 1,\ d' \to d,\ if\ d' \in D_{in}(d) \\ or\ d \to d',\ if\ d' \in D_{out}(d) \end{cases} \quad (4)$$

Table 1
Channels and number of news

Channel name	Text news	Video news	Total news
Entertainment	17,244	11,069	28,313
World	11,517	2,678	14,195
Society	9,651	4,812	14,463
Finance	32,447	151	32,598
Technology	7,704	98	7,802
Stock Market	17,695	0	17,695
Army	5,002	0	5,002
Health	0	2	2
Domestic	25,322	4,340	29,662
Sports	35,045	14,868	49,913
USA Stock	487	43	530
Others	4,257	2	4,259
Total	166,371	38,063	204,434

where $D_{in}(d)$ is the set of all the news pages that links in to news page $_{\mathrm{d}}$, in contrast, the $D_{out}(d)$is the set of all the news pages that links out from news page $_{\mathrm{d}}$.

In this paper, we deem news web pages as a kind of living creature. Each news is of energy, which changes as some factors, such as time goes by and user's concern. The value of this energy of every piece of news indicates how old the news web page is, and the slope of the energy shows the temporal sensitivity of news pages.

4. Experiment

4.1. Data set

At the beginning of experiments, we give a brief description of the data set. All these news come from the news portal Sina, one of the top four Chinese portals. we crawled about 204 thousand news for a period of more than 4 months (published from August 4 to December 23, 2013). The data Set has two styles, text news and video news. They fall into 12 types, statistics showed in Table 1. And we tracked the dynamics of these news from 2013-08-05 to 2013-12-27. Besides, some related news links, comments and supports of every piece of news have also been crawled. Where 63.3% percent of news have related news links and 63%percent of news have comment and join information.

4.2. News content

Temporal Words: We identity and collect all the temporal words in the news article with the public system HeidelTime [21]. It is one of the best performed systems in TempEval-2 and TempEval-3

Table 2
Statistic of news label

	Text news	Video news
Number of news	135,096	36778
Percentage of news	81.2%	96.6%

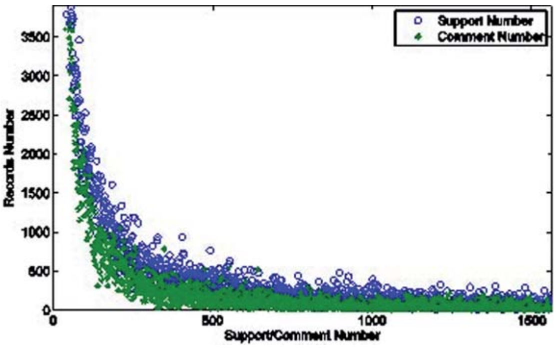

Fig. 1. Distributions of support number and comment number.

tasks. What's the most important, it is a multilingual public tool including Chinese.

Topic Words: Since most of news have labels on the news page, and they are create manually that much accurate than generating by topic models, we take these labels as the topic words of news. the statistics is as Table 2.

4.3. User behaviors

In this part, we have done the correlation analysis between the comment number and support number. By looking at the values, note that support number are about 10 times the number of comments,

Before analysis the correlation between the comment number and support number, it is essential to display the data of these two features. However, the range of the two features are all too wide that comment number ranges from 0 to 2.15×105 and Join number ranges from 0 to 3.7×105. It is very hard to clearly show in one figure. So, we just display the local data which located at the most densest place.

In Fig. 1 the Records Number means that how many records are with the Comment Number or Support number at the given support number or comment number. Note that Fig. 1 shows that support number and Comment number of all news in our data set are both follow the exponential distribution, and it indicates that there is likely a strong correlation between the two features.

Table 3
Clustering results comparison

Method	Run times	Precision
TF/IDF	2160s	0.88
Entropy	2170s	0.93

4.4. Clustering

We have adopted the traditional single-pass clustering method with the energy function. The Time window in the clustering is a fixed instant. And the threshold we set is 0.9.

In Table 3, we can see that our method are prior to the TF/IDF representation method in precision, as we adopt the cross entropy to compute the vector in VSM. While in run time, no obvious difference. Besides, The two methods both perform good in run time as it drops the pages that are in the memory so long.

5. Conclusion

This paper has presented a triple model with four sorts of features to depict the temporal sensitivity of news, and employ an energy based method to compute the value of temporal sensitivity degree at every time point. Currently, it is a prevalence that web contains ineffective, inconsistency and obsolete information. An important factor that results in this phenomenon is web information is temporal sensitive, that is information is changing over time. However, different kinds of information vary at different ratio. In order to uniformly quantify the value of temporal sensitivity of news, this paper has described the news as lives, and used change ratio of the energy of every life to represent the temporal sensitivity in terms of the time, popularity and information volume. This model can be applied to web page quality assessment and ranking, as well as to information retrieval to increase the quality of result lists of every query, further to improve the experience of users when surf or search on the Internet. The experiment demonstrates on Chinese Sina news. In the future, Multiple data sets such as weibo or twitter will be considered.

Acknowledgments

The authors acknowledge the support by the China Scholarship Council(CSC) and the Fundamental Research Funds for the Central Universities.

References

[1] M. Zahedi, et al., Time sensitive blog retrieval using temporal properties of queries, *Journal of Information Science* **43**(1) (2017), 103–121.

[2] V.P.N.A. Araya, et al., Gaining historical and international relations insights from social media: Spatio-temporal real-world news analysis using Twitter, *EPJ Data Sci* **6**(1) (2017), 25.

[3] A. Yoshihara, K. Seki and K. Uehara, Leveraging temporal properties of news events for stock market prediction, *Artif Intell Research* **5**(1) (2016), 103–110.

[4] F.S. Peregrino, D.T.A. S and F. Llopis, Temporal. Language Analysis in News Media and Social Networks, *Research in Computing Science* 144 (2017), 125–136.

[5] M.C. Calzarossa and D. Tessera, Modeling and predicting temporal patterns of web content changes, *Journal of Network and Computer Applications* **56** (2015), 115–123.

[6] I. Mele, S.A. Bahrainian and F. Crestani, Event mining and timeliness analysis from heterogeneous news streams, *Information Processing & Management* **56**(3) (2019), 969–993.

[7] S. Wu, A Traffic Motion Object Extraction. Algorithm, *International Journal of Bifurcation and Chaos* **25**(14) (2015), Article Number 1540039.

[8] S. Wu, M. Wang, Y. Zou, Research on internet information mining based on agent algorithm, *Future Generation Computer Systems* **86** (2018), 598–602.

[9] S. Wu, M. Wang and Y. Zou, Bidirectional cognitive computing method supported by cloud technology, *Cognitive Systems Research* **52** (2018), 615–621.

[10] S. Wu, Nonlinear information data mining based on time series for fractional differential operators, *Chaos* **29** (2019), 013114. https://doi.org/10.1063/1.5085430.

[11] W. Yu, S. Li and X. Tang, et al., An efficient top-k ranking method for service selection based on ε-ADMOPSO algorithm, *Neural Computing and Applications* **31**(1) (2019), 77–92.

[12] W. Yu and S. Li, Recommender systems based on multiple social networks correlation, *Future Generation Computer Systems* **87** (2018), 312–327.

[13] D. Korencic, S. Ristov and J. Snajder, Document-based topic coherence measures for news media text, *Expert Syst Appl* **114** (2018), 357–373.

[14] G.D. Nguyen, et al., Information freshness over a Markov channel: The effect of channel state information, *Ad Hoc Networks* **86** (2019) 63–71.

[15] R.B. Sağlam and T.T. Temizel, Automatic information timeliness assessment of diabetes web sites by evidence based medicine, *Computer Methods and Programs in Biomedicine* **117**(2) (2014), 104–113.

[16] Y. Park, et al., A deep learning-based sports player evaluation model based on game statistics and news articles, *Knowl-Based Syst* **138** (2017), 15–26.

[17] S. Han, X. Hao and H. Huang, An event-extraction approach for business analysis from online Chinese news, *Electronic Commerce Research and Applications* **28** (2018), 244–260.

[18] D.B. Bisandu, R. Prasad and M.M. Liman, Clustering news articles using efficient similarity measure and N-grams, *IJKEDM* **5**(4) (2018), 333–348.

[19] M. Chung, Not just numbers: The role of social media metrics in online news evaluations, *Computers in Human Behavior* **75** (2017), 949–957.

[20] M. Timilsina, et al., Social impact assessment of scientist from mainstream news and weblogs, *Social Netw Analys Mining* **7**(1) (2017), 48:1–48:15.

[21] J.S.O. Tgen and M. Gertz, Domain-Sensitive Temporal Tagging, *Computational Linguistics* **44**(2) (2018), 375–377.