

Twitter summarization with social-temporal context

Ruifang He¹ · Yang Liu² · Guangchuan Yu¹ ·
Jiliang Tang³ · Qinghua Hu¹ · Jianwu Dang¹

Received: 26 January 2015 / Revised: 8 December 2015 / Accepted: 16 March 2016
© Springer Science+Business Media New York 2016

Abstract Twitter is one of the most popular social media platforms for online users to create and share information. Tweets are short, informal, and large-scale, which makes it difficult for online users to find reliable and useful information, arising the problem of Twitter summarization. On the one hand, tweets are short and highly unstructured, which makes traditional document summarization methods difficult to handle Twitter data. On the other hand, Twitter provides rich social-temporal context beyond texts, bringing about new opportunities. In this paper, we investigate how to exploit social-temporal context for Twitter summarization. In particular, we provide a methodology to model temporal context globally and locally, and propose a novel unsupervised summarization framework with social-temporal context for Twitter data. To assess the proposed framework, we manually label a real-world Twitter dataset. Experimental results from the dataset demonstrate the importance of social-temporal context in Twitter summarization.

Keywords Twitter summarization · Social media · Time point detection · Wavelet denoising · Social-temporal context

✉ Ruifang He
he.ruifang@gmail.com

Qinghua Hu
huqinghua@tju.edu.cn

¹ Tianjin Key Laboratory of Cognitive Computing and Application, School of Computer Science and Technology, Tianjin University, Tianjin 300072, China

² School of Computer Science and Technology, Peking University, Peking 100871, China

³ School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, 85281, Tempe, AZ, USA

1 Introduction

Twitter has become a popular microblogging platform for online users to gather and share information. However, tweets are fragmented, informal, and large-scale, which makes it difficult for users to efficiently acquire core information of a topic. Hence Twitter summarization is proposed to mitigate the information overload problem and has attracted increasing attention in recent years [3, 5].

Traditional document summarization is extensively studied and has been successfully applied in various domains such as search engines [1, 10] and Q&A systems [22]. However, traditional document summarization becomes difficult when handling Twitter data. First, tweets are very short, and often do not provide sufficient statistical information for robust similarity measures [6]. Second, traditional documents usually have structure information, which provides important cues for summarization such as the positions of sentences in documents or paragraphs [10, 16], while a tweet is highly unstructured and usually only contains one sentence [3]. The unique properties of Twitter data not only determine that it is difficult to apply traditional document summarization but also manifest a new research challenge for summarization.

Meanwhile, Twitter often provides additional information other than texts as demonstrated in Figure 1. Users in Twitter can follow each other and form a user-user following network. Users, and their following relations provide rich social context for tweets, which is potentially useful for summarization. For example, tweets from users with high reputations

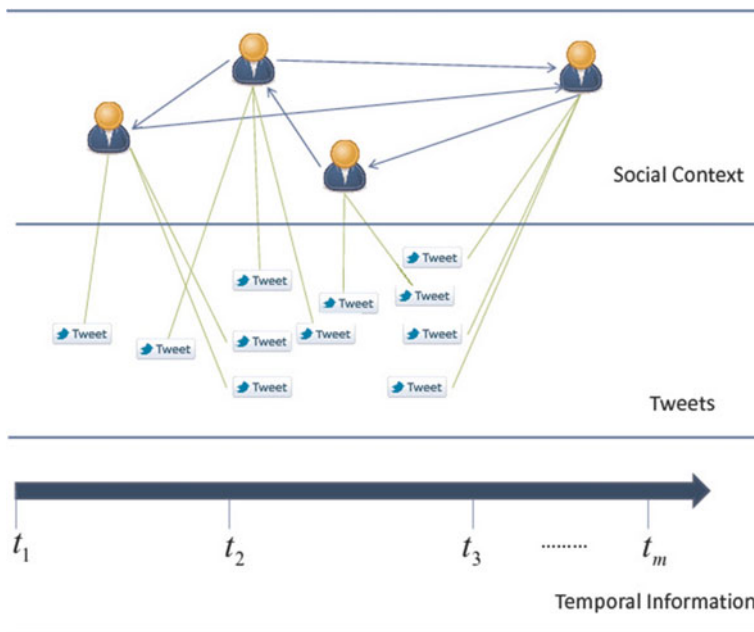


Figure 1 Twitter Data with Social-Temporal Information

are likely to be reliable and important. Tweets related to a topic are created sequentially but they may not be uniformly distributed. There are some time periods with very few tweets such as the period from t_1 to t_2 in Figure 1, while there are other time periods where tweets are bursty such as the period from t_2 to t_3 . These temporal changes might provide important cues for summarization. For example, the tweet burst means that there are sudden concentrations of the topic in these periods, and tweets in bursts should be more informative. The availability of social-temporal context provides unprecedented opportunities and can help advanced research on summarization for Twitter data. There is recent work exploiting social context for Twitter summarization [3, 5], however, social-temporal contexts are seldom touched.

The unique properties of tweets pose both challenges and opportunities for Twitter summarization. In this paper, we investigate how to exploit social-temporal context for Twitter summarization. In particular, we study: (1) how to model social-temporal context mathematically; (2) how to make use of social-temporal context for summarization with Twitter data. In our attempt to answer these questions, we propose a novel summarization framework T2ST for Twitter data. Our contributions are summarized below.

- Provide a methodology to model temporal context from both local and global perspectives mathematically;
- Propose a novel summarization framework T2ST for Twitter data which incorporates both social and temporal context; and
- Build a manually labeled Twitter dataset to evaluate the proposed framework T2ST.

The rest of paper is organized as follows: Section 2 introduces the proposed framework including our solutions about how to capture social context and how to model temporal context, and how to make use of them in Twitter summarization. Section 3 introduces the experiments with details about data set construction, metrics, and experimental results and discussions. Section 4 briefly reviews the related work. Finally, we conclude the paper with future work in Section 5.

2 The proposed framework for twitter summarization

In this paper, we aim to exploit social-temporal context for Twitter summarization. Assume that $\mathcal{T} = \{tw_1, tw_2, \dots, tw_N\}$ is the set of N tweets and $\mathcal{D} = \{d_1, d_2, \dots, d_M\}$ is the dictionary with M words. Before going to the details, we first introduce LexRank [6] as a basic summarization method of the proposed framework. LexRank is a graph-based ranking algorithm performing well in the field of traditional automatic summarization. It first segments documents into sentences, then constructs a graph with sentences as nodes and their similarities as the weight of edges, and finally performs a random walk on the graph. When a random walk on graph arrives at its stationary state, the probability of a node being visited is its importance. The whole process is repeated until a global stable state is achieved, and all nodes in the graph are ranked based on their final rank scores. When applying it to Twitter data, each tweet is considered as a node and their similarities as the weights of edges. The detailed algorithm is shown in Algorithm 1.

Algorithm 1 A Basic Algorithm for Twitter Summarization

- 1: Based on the vector space model, compute the tweet vector $D_i = \{w_1, w_2, \dots, w_M\}$ for each tweet tw_i where w_i is the TF-IDF like weight of d_i ;
- 2: Calculate the cosine similarity $\text{sim}(D_i, D_j)$ between tweets tw_i and tw_j ;
- 3: Obtain the transition probability matrix \mathbf{M} based on the similarity:

$$\mathbf{M}_{ij} = \begin{cases} \frac{\text{sim}(D_i, D_j)}{\sum_{j' \text{ adjacent to } i} \text{sim}(D_i, D_{j'})} & \text{if } \sum_{j'} \text{sim}(D_i, D_{j'}) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where j' refers to all the nodes adjacent to i ;

- 4: Perform a random walk on the tweet graph,

$$s_i = (1 - \alpha)/N + \alpha \sum_{j \neq i} s_j * \mathbf{M}_{ji} \quad (2)$$

, where α is a jumping factor, and s_i (initialized to $1/N$) is the rank score for each tweet;

Here, we just use the standard Cosine measure [6] to compute the similarity values. The weight w_i associated with term d_i is calculated with the $tf_{d_i} * idf_{d_i}$ formula, where tf_{d_i} is the frequency of term d_i in the tweet and idf_{d_i} is the inverse tweet frequency of term d_i , i.e. $1 + \log(N/n_{d_i})$, N is the total number of tweets and n_{d_i} is the number of tweets containing term d_i . Then $\text{sim}(D_i, D_j)$ is computed according to the normalized inner product of the corresponding term vectors. We also empirically set jumping factor parameter in the random walk as $\alpha = 0.85$.

After the algorithm converges, we will do some necessary post-processing. We first extract the tweets based on their rank scores; meanwhile we use the MMR (Maximal Marginal Relevance) [2] algorithm to removal the redundancy. The basic idea of MMR is that if one tweet is selected into the summary, we will rerank the other tweets according to the dissimilarity with the tweet summary. The more dissimilar, the more probable one tweet is selected into the tweet summary set. In comparison with traditional documents, Twitter data is short and highly unstructured and we would like to use social-temporal context to mitigation these challenges in this paper. In the following subsections, we will provide our solutions to exploit social-temporal context, which leads to the proposed framework T2ST.

2.1 Capturing social context

Twitter has a strong social characteristic considering the asymmetry of following and followed among users, which implies that tweets posted by users with different authorities have a potentially logical relation and thus have a different salience. A Twitter user can follow each other and actively choose the friends. Such a relationship is not reciprocal [3]. A Twitter user does not need to follow back if the user is followed by another. However, the number of its followers still affect whether this user can be selected as a friends of others. Being a Twitter follower means that the user will receive all tweets from other users that he/she follows. A Twitter follower could broadcast a tweet to its own followers by retweeting or replying to the tweet. Through this way, Twitter followers propagate the influences of a celebrity. Thus the users' social statuses contribute to determining a tweet's importance.

Due to the availability of social context for tweets, the importance of tweet content should be strongly related to the authority of the users who post the tweet. We use $AS(D_i)$ to denote the author's authority of the i -th tweet tw_i , which is defined as

$$AS(D_i) = \frac{fol(D_i)}{fri(D_i)} * RT(D_i) \quad (1)$$

where $fol(D_i)$ and $fri(D_i)$ are the numbers of followers and followees of the author of tw_i . The rationale under this definition is based on much observations and the assumption. (1) If a user has many friends, but few followers, it may be an advertiser with a high probability. (2) If a user has many followers but few friends, then the person likely receives approval and recognition and might be an authority with a high probability. Not following many people arbitrarily shows that the post author might be publishing valuable and original information. (3) If the i -th tweet tw_i is retweeted many times, its author's influence could be enlarged, where $RT(D_i)$ indicates the retweet number. Therefore, we can combine these factors to get a measure about social context of a user in a hybrid division and multiplication way.

Consider $AS(D_i)$ can be larger than 1, we divide it by maximum value MAX_A of the users' influences, mapping it into the range of $[0,1]$ as

$$A(D_i) = \begin{cases} \frac{AS(D_i)}{MAX_A} & \text{if the author of } D_i \text{ exists} \\ AS_{AVE} & \text{otherwise} \end{cases} \quad (2)$$

The reason why we use MAX_A is that it can normalize the influence of social context into the range of $[0,1]$, which can conveniently be integrated into the transition probability matrix since cosine similarity score lies in $[0,1]$. If the author of D_i does not exist, we will take the average score AS_{AVE} of all authors for D_i . A tweet is more likely to be an informative tweet if it is posted or retweeted by authoritative users. Therefore we use social context to modify the transition probability by considering the authorities of users as

$$\mathbf{M}_{ij} = \begin{cases} \frac{\text{sim}(D_i, D_j) * A(D_j)}{\sum_{j'} \text{sim}(D_i, D_{j'}) * A(D_{j'})} & \sum_{j'} \text{sim}(D_i, D_{j'}) * A(D_{j'}) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

2.2 Capturing temporal context

Before modeling, we first build heat signals to investigate the temporal characteristics of tweets related to a given topic. We use tweet publishing speed to judge at which moment the topic is lively discussed in Twitter. We define update rate of tweets v_{t_i} for a given topic K at time point t_i as,

$$v_{t_i} = \frac{N_{t_i}}{\Delta_t} * \frac{1}{N} \quad (4)$$

where N_{t_i} denotes the number of tweets containing K within the time period $[t_i, t_i + \Delta_t]$, N is the number of all tweets, and Δ_t is a time window. Finally we get the heat signals $f(t)$ as $f(t) = \{v_{t_1}, v_{t_2}, \dots, v_{t_{N_t}}\}$ where N_t is the number of total time points.

Figure 2 demonstrates the heat signals built from tweets with the topic "Obama" during the first half of 2011, and the time window Δ_t is set to one hour. We observe that the heat

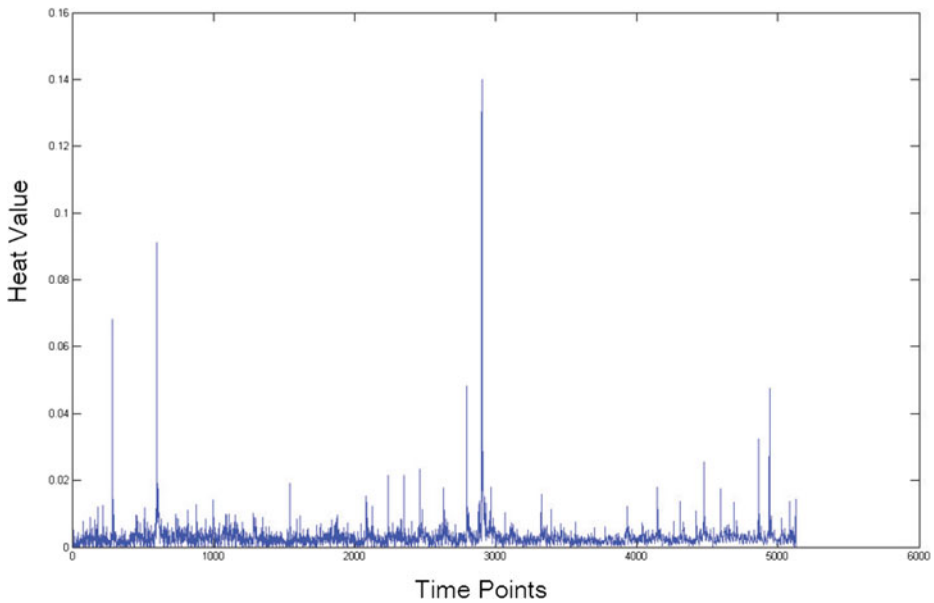


Figure 2 Heat signals built with the topic “Obama”

signals are not uniformly distributed. We further examine the highest heat value and find that it appears on May 1, 2011, which is the day when Obama announced that U.S. Army has killed Osama bin Laden. We also find that other peak points are in line with significant events related to “Obama” that occur in reality.

The update rate of a topic strongly reflects its popularity at that time, and the peak heat signals correspond to the concentrations of the topic. Globally, there are some key time points with the concentrations of the topic, and locally the heat signal of a tweet reflects its importance under the time point. Tweets are likely to be informative if they are posted in the key time points with high heat signals. The above analysis paves a way for us to exploit temporal context from both global and local perspectives. In particular, to capture global temporal context, we detect the key time points for a given topic, while for each key time point, we use heat signals in a day to indicate the importance of tweets. More details will be given next.

2.2.1 Global key time point detection

In our model, we assume that a good Twitter summary should consist of time points that can best reconstruct the heat signals of a topic. From the geometric interpretation, we should select time points that span the intrinsic subspace of the heat signals about a topic so that it is able to cover the core signals of a topic. Next we will give details about the algorithm.

Through analyzing the heat signals from a real-world dataset, we find that most peaks in the curve are redundant and only some of them essentially indicate the curve trend. Therefore it is unnecessary to consider each time point for Twitter summarization. Although the heat signals can substantially reflect the importance of time points, there may still exist noise in the signals due to the following reasons.

- The diversity of topics - it is possible that users inadvertently mention the topic while discussing personal things.
- Asynchronous publishing - users are likely to publish tweets online after the change or the progress actually happens.
- The presence of redundant information - a large number of meaningless tweets may be mixed with useful tweets, which may make the heat signals inaccurate to reflect the true popular trends.

To detect key time points, it is necessary to eliminate noise existing in the heat signals. The wavelet analysis provides precise measurements regarding when and how the frequency of the signals change over time, which helps to extract valid information from signals. We use $f(t)$ to denote the heat signals of a given topic and next we will introduce our signal denoising algorithm based on wavelet analysis.

The wavelet analysis generally uses a quickly vanishing oscillating function, i.e., mother wavelet, to match the signal $f(t)$. Wavelet transform [4] is the core method of the wavelet analysis. It scales and translates the mother wavelet $\psi(t)$ to form a set of linearly independent basis functions, which are called a wavelet family and defined as,

$$\{\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), a > 0, b \in R\} \quad (5)$$

where a is the scaling factor, and b refers to the translating factor. Due to its good performance in the literature [4], we select DWT (discrete wavelet transform) for the heat signal reconstruction in this paper.

Let the scaling factor $a = a_0^m$ ($m \in Z, a_0 \neq 1$) and the translating factor $b = nb_0a_0^m$ ($n \in Z$). The discrete mother wavelet and the discrete wavelet transform function of the heat signals $f(t)$ can be represented as,

$$\psi_{m,n}(t) = |a_0|^{-m/2} \psi(a_0^{-m}t - nb_0) \quad (6)$$

$$W_f(m, n) = \langle f(t), \psi_{m,n} \rangle = \int_R f(t) \psi_{m,n}(t) dt \quad (7)$$

where m is the resolution parameter and n is the shift parameter of wavelet. Using the orthonormal basis ($a_0 = 2, b_0 = 1$), we can apply the transform multi-resolution analysis [13] and wavelet denoising [9] to the original heat signal of a topic.

Since the valid information and the noise information have different frequency characteristics, we can effectively remove noisy signals by selecting appropriate thresholds at each resolution. This process is called “wavelet denoising”, and we will use this to eliminate noisy signals in the heat signals of the topic. Wavelet denoising performs the following steps:

- Use DWT on the origin heat signals $f(t)$ and calculate the wavelet coefficients $W_f(m, n)$;
- Process coefficients at each resolution m with thresholds, coefficients are sparse and only time points with strong signals are kept;
- Reconstruct the signals with the processed wavelet coefficients, and the reconstruction formula is represented as $\tilde{f}(t) = \sum_m \sum_n W_f(m, n) \psi_{m,n} = \sum_m \sum_n \langle f, \psi_{m,n} \rangle \psi_{m,n}$;

The threshold is important for wavelet denoising. If the threshold value is too high, some valid information in the signals may be filtered out as noises; while a low threshold will keep some noises in the reconstructed signals. We choose a soft-threshold function [15] based on HeurSure threshold to mitigate this problem.

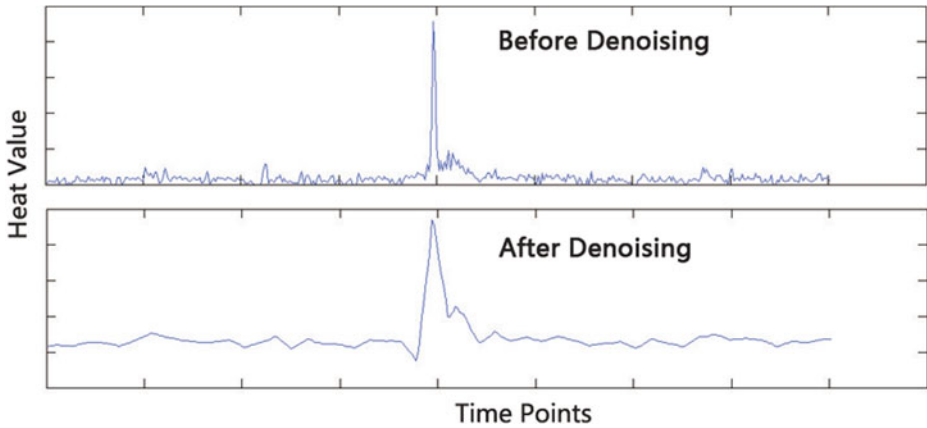


Figure 3 Results of Wavelet Denoising

Figure 3 shows a segment of the signals built with the topic “Obama” before and after wavelet denoising. The wavelet denoising smooths the waveform and eliminates part of the redundancy in the heat signals. Therefore, the maximums of the signals can be selected and ranked to be the key time points of a topic. The time points we select are based on a time window of a date. If there are multiple peaks appearing in the same day, we only use the highest one to compete with other dates.

2.2.2 Modeling local temporal context

The time window for key time point detection is one day, and the heat signals for tweets within the day may be different. In each key time point, to capture the local temporal context, we discriminate the temporal influence of a tweet according to the time window “hour”. We also define the transition probability based on the heat signals $H(D_i)$ for the tweet tw_i , which is used to measure the influence of dynamic context.

$$H(D_i) = \frac{v_{p(D_i)}}{MAX_H} \quad (8)$$

where $p(D_i)$ represents the time of the tweet t_i posted, and v is update rate defined in (4). MAX_H is the maximum value of update rate during a day, which is used to map the heat signals to the range of [0,1]. The reason why we normalize $H(D_i)$ with MAX_H is similar with the situation of capturing the social context.

After this, the transition probability matrix with local temporal context is shown in (9).

$$\mathbf{M}_{ij} = \begin{cases} \frac{\text{sim}(D_i, D_j) * H(D_j)}{\sum_{j'} \text{sim}(D_i, D_{j'}) * H(D_{j'})} & \sum_{j'} \text{sim}(D_i, D_{j'}) * H(D_{j'}) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where the weights of edges are changed based on the heat signals, and nodes with a higher heat receive a higher salience score. In (9), we not only consider the content importance of each tweet among the entire dataset, but also the local influence of topic heat signals when a tweet is published.

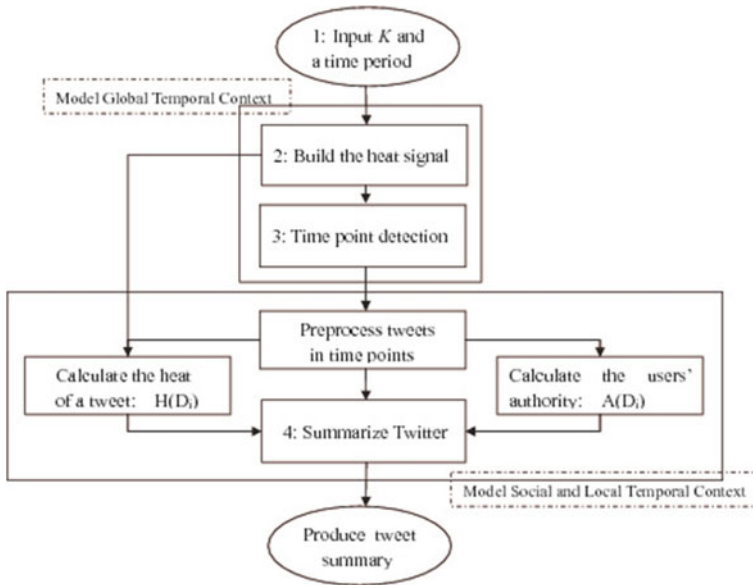


Figure 4 The Proposed Framework for Twitter Summarization T2ST

2.3 Our framework with social-temporal context

With our solutions to exploit social-temporal context, the proposed framework is shown in Figure 4. In general, the input is a time sequence with tweets about a topic K and the output is Twitter summarization. According to Figure 4, details of the proposed framework are shown in Algorithm 2. We first build the heat signals, then detect key time points and finally extract tweets from each key time point using Algorithm 1. About how to integrate the social and local temporal context collectively, we propose two methods according to the different intuitions. (1) Consider the two kinds of context influence with each other, we take the similar mechanism with the single context. The transition probability matrix of Algorithm 1 is simply defined as their multiplication combination shown in (10).

$$\mathbf{M}_{ij} = \begin{cases} \frac{\text{sim}(D_i, D_j) * A(D_j) * H(D_j)}{\sum_{j'} \text{sim}(D_i, D_{j'}) * A(D_{j'}) * H(D_{j'})} & \sum_{j'} \text{sim}(D_i, D_{j'}) * A(D_{j'}) * H(D_{j'}) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

(2) From the viewpoint of set theory, we can extract the tweets with the high ranking scores based on a certain summary length ratio of the method with social context to the method with temporal context. The final summary tweet set **T2ST-Set** (such as abbreviation, please refer in Section 3.2) is defined as (11), where TSS_{set} , TST_{set} is top tweet set from TSS and TST based on the summary length parameter λ and $1 - \lambda$ respectively. Here, $0 \leq \lambda \leq 1$.

$$\mathbf{T2ST-Set} = \lambda * TSS_{set} + (1 - \lambda) * TST_{set} \quad (11)$$

Algorithm 2 The Proposed Twitter Summarization Framework with Social-Temporal Context (T2ST).

- 1: Given a topic K and a time period $[t_{start}, t_{end}]$, search all the tweets relevant to the topic published during the given period;
 - 2: Build the heat signals $f(t)$ according to the time when tweets are posted;
 - 3: Detect key time points with the time window of one day based on wavelet denoising;
 - 4: **for** Each key time point **do**
 - 5: Extract tweets to satisfy the summary length through two methods separately: (1) based on Algorithm 1 with the transition probability matrix integrating social and local temporal context shown in (12); (2) based on (13).
 - 6: **end for**
-

3 Experiments

In this section, we conduct experiments to assess the proposed framework on a real-world Twitter data. After introducing the dataset, and the evaluation metric, we investigate the following questions from the experiments.

- Can exploiting social-temporal context improve the performance of Twitter summarization? and
- what are the effects of social-temporal context on the proposed framework?

Finally we conduct further experiments to understand the working of temporal context.

3.1 Data settings

3.1.1 Data preparation

To the best of our knowledge, there is no corpus available to evaluate Twitter summarization with social-temporal context. Thus we manually label a real-world Twitter dataset by inviting seven college students as volunteers.

We choose a real Twitter dataset collected by University of Illinois¹ as the raw data and further construct the ground truth data. There are 400 million tweets crawled from 3 million users, which were published between January 2011 to August 2011. Besides the tweet content, this corpus also contains the Twitter metadata such as publishing time and forwarding times. Based on the raw data, we construct Twitter summarization corpus in the selected time points. In this paper, we only care about the topics from economic and politic fields, and summarize them based on the clues of person, product and company. For the popular and public topic, the user authority has some common characteristic. Thus we select three topics “Obama”, “iPad”, and “Microsoft”, and search all the tweets related to these topics during the period. The statistics of this dataset are shown in Table 1.

For the golden standard key time points, each volunteer is required to pick up the most important 20 time points from each topic under the premise of carefully reading topic relevant news. We arrange three volunteers to make ground truth data for each topic, which is used to evaluate key time point detection.

¹<https://wiki.engr.illinois.edu/display/forward/Dataset-UDI-TwitterCrawl-Aug2012>

Table 1 Data for time point detection

Topics	Size	# of Tweets
Obama	20.7MB	221364
iPad	13.6MB	143887
Microsoft	15.3MB	172664

Secondly, we only select four key time points to further construct expert tweet summary since it is very time-consuming for all key time points given the informal and fragmented tweets. The statistics of this dataset is shown in Table 2. The volunteers are asked to extract expert summaries with the above data, which are used to evaluate the content selection of system summaries. We give the tweet collections, topics and the posting time information to the volunteers. Since the different users have the different understandings about a topic, we ask volunteers to select the expert summaries and take the average ROUGE scores to obtain the evaluation about expert summary quality. For each selected time point, three volunteers are required to produce three tweet expert summaries, and each expert summary has about 10 tweets. Altogether there are 12 tweet expert summaries for four time points.

In order to measure the inter-annotator agreement, we measure the content quality of expert summary tweets by calculating the average ROUGE value among them [12], which are shown in Table 3. We don't do anything about language quality evaluation. After all, this expert summary is extractive but not abstractive, and the quality of expert summary can be acceptable in comparison with the level of ROUGE scores in DUC or TAC evaluation, .

3.1.2 Evaluation metric

We use the common ROUGE [12] metric for Twitter summarization, which is an n-gram recall based statistic and widely used in summarization evaluation. It measures summary content quality by counting overlapping units such as the N-gram, word sequences and word pairs between the system summary and the reference summary. ROUGE-N is defined in (12).

$$\text{ROUGE-N} = \frac{\sum_{m \in MS} \sum_{u \in m} \text{match}(u)}{\sum_{m \in MS} \sum_{u \in m} \text{count}(u)} \quad (12)$$

where MS is the set of manual summaries, u is the set of n-gram in a particular manual summary. $\text{match}(u)$ represents the number of n-grams that appear in both automatic and manual summary, and $\text{count}(u)$ is the number of n-grams in manual summary. In this paper, we use ROUGE-1 and ROUGE-2 to evaluate the summarization performance.

In order to obtain better summary results, some preprocessing needs to be conducted, including case conversion (all letters will be converted to lowercase), stopping words

Table 2 Data for twitter summarization

Keyword	Time point	Size	# of tweets
Obama	2011/01/25 (Obama-1)	133KB	1328
Obama	2011/05/19 (Obama-2)	113KB	1108
iPad	2011/03/11 (iPad)	114KB	1308
Microsoft	2011/05/10 (Microsoft)	149KB	1758

Table 3 Evaluation on expert summaries

Data Set	ROUGE-1	ROUGE-2
Obama-1	0.32423	0.06329
Obama-2	0.35131	0.09576
iPad	0.44568	0.25695
Microsoft	0.39981	0.11179
Average	0.38026	0.13195

removal, stemming (same with the paper [6]), removing all non-alphanumeric characters and “RT” strings.

3.2 Performance comparisons for twitter summarization

To answer above questions, we choose the following baseline methods for comparison.

- **Random**: selects summary tweets for a given topic randomly;
- **SumBasic**: SumBasic [23] uses simple word probabilities with an update function to compute the best k posts. It was chosen because it depends solely on the frequency of words in the original text and is conceptually very simple;
- **TS**: this baseline method is based on LexRank, which is shown in Algorithm 1;
- **TSS**: this baseline method is TS with social context whose transition probability matrix is defined as (3);
- **TST**: this baseline method incorporates TS with temporal information and its transition probability matrix is defined as (9);
- **T2ST-M**: this is the proposed framework which exploits social and temporal context based on TS in a multiplication way shown in (10);
- **T2ST-Set**: this is the proposed framework which exploits social and temporal context based on TSS and TST in a set theory viewpoint shown in (11);
- **TS+DSocial**: combines the social context based on Duan’s work [5] with TS method, and only integrates his social context shown in (13), where $d = 0.5$; $F(u_i)$ is the static influence of the i -th user, and U is the set of all user publishing the current tweets. We take the simplified $F(u_i)$, which is the follower number of u_i . $flw(u_i)$ is the follower set of u_i . $|frd(u_j)|$ is the followee number of user u_j , that is how many people u_j follows .

$$Score^{(r+1)}(u_i) = (1 - d) \frac{F(u_i)}{\sum_{u \in U} F(u)} + d \sum_{u_j \in flw[u_i]} \frac{1}{|frd[u_j]|} Score^{(r)}(u_j) \quad (13)$$

- **SumBasicS**: integrates SumBasic with our social context, and the original tweet score is multiplied by (2);
- **SumBasicT**: integrates SumBasic with our temporal context, and the original tweet score is multiplied by (8);
- **SumBasicST-M**: SumBasic together with our social and temporal context, and tweet score is multiplied by (2) and Eq. (8).
- **SumBasicST-Set**: SumBasic integrating with our social and temporal context through the method similar to **T2ST-Set** shown in (11).

Table 4 Performance comparison under ROUGE-1

Method	ROUGE-1					
	Obama-1	Obama-2	iPad	Microsoft	AVE	IPR (%)
Random	0.27824	0.27846	0.32392	0.39231	0.31823	–
TS	0.32427	0.36382	0.33145	0.38462	0.35104	–
TSS	0.34937	0.38415	0.36347	0.41795	0.37874	+7.7790
TST	0.33264	0.38211	0.36158	0.39744	0.36844	+4.8500
T2ST-M	0.36611	0.41057	0.35405	0.39744	0.38204	+8.7200
T2ST-Set	0.36611	0.39024	0.38041	0.42821	0.39124	+11.338
TS+Dsocial	0.33682	0.33537	0.29944	0.39487	0.34163	–2.7820
SumBasic	0.3159	0.33333	0.33522	0.34615	0.33265	–
SumBasicS	0.38075	0.31098	0.3371	0.42308	0.36298	+9.1170
SumBasicT	0.29707	0.33943	0.35217	0.34872	0.33435	+0.5100
SumBasicST-M	0.36611	0.30894	0.3371	0.42821	0.36009	+8.2490
SumBasicST-Set	0.35983	0.28252	0.32957	0.40769	0.34490	+3.6830

For our method, we extract summary in a tweet level and the summary length is set to be the average number of words included in the expert summary tweets. We detect a key time point within the time window of one day and the number of selected time points is set to be 10. More details about the investigations of the summary length and the number of time points will be discussed in the following subsections. Note that the goal of this paper is to investigate the impact of social-temporal information and we choose LexRank as the basic algorithm. Each method has a different way to integrate social-temporal context, hence we do not compare with other more traditional document summarization methods [7, 24] except for SumBasic method. We would like to investigate social-temporal context with other more basic algorithms as our future work. Also we do not compare our unsupervised framework with the supervised framework in [3].

3.2.1 Overall performance

The whole experimental results are shown in Tables 4 and 5, where AVE indicates the average system performance on four topics and IPR (Improvement Rate) indicates the system performance improvement rate of comparison methods series in contrast to their basic method. For example, the methods from the 3rd line to 8th line belongs to the TS series, and the 9th to 13th is the SumBasic series; TS, SumBasic is the basic method respectively. Thus, $IPR = (a - b) / b$, where a , b is the performance of the comparison method and the baseline respectively.

It can be seen that Random method is the worst among the basic methods, and TS is a little better than SumBasic. Observed from the performance at each topic, AVE and IPR, most separately or collectively integrating social and temporal context improves the system performance. The performance on TS series methods is better than that of SumBasic series methods. The reasons are two-fold. One is that SumBasic itself performs worse than TS; the other is that how to use the social and temporal context depends on each specific method, and the same integrating context way could not have the same effect for all methods.

Table 5 Performance comparison under ROUGE-2

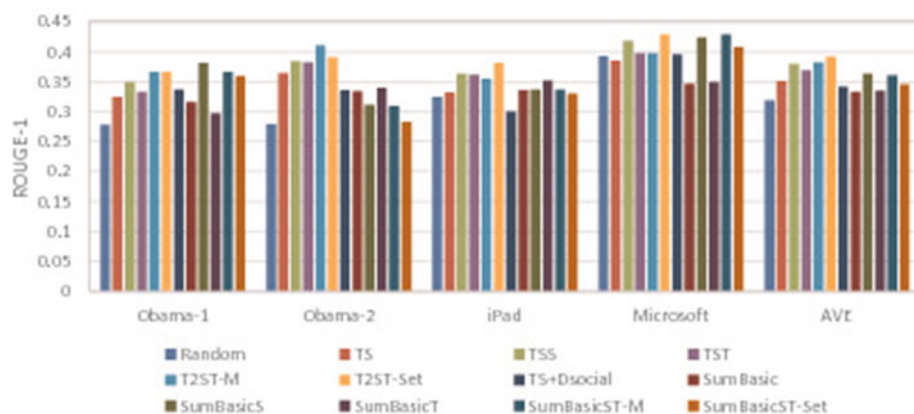
Method	ROUGE-2					
	Obama-1	Obama-2	iPad	Microsoft	AVE	IPR (%)
Random	0.04842	0.0409	0.04924	0.07752	0.05402	–
TS	0.06737	0.09816	0.10795	0.12403	0.09938	–
TSS	0.09474	0.11861	0.12121	0.14987	0.12111	+21.866
TST	0.06947	0.10429	0.13826	0.13437	0.11160	+12.296
T2ST-M	0.09474	0.11452	0.14205	0.14212	0.12336	+24.130
T2ST-Set	0.10316	0.12679	0.13447	0.14987	0.12857	+29.377
TS+Dsocial	0.09053	0.04908	0.05303	0.13178	0.08111	–18.387
SumBasic	0.07158	0.05521	0.11553	0.10336	0.08642	–
SumBasicS	0.11789	0.04294	0.12311	0.16021	0.11104	+ 28.486
SumBasicT	0.04632	0.05521	0.11364	0.10336	0.07963	–7.8540
SumBasicST-M	0.10737	0.04294	0.12311	0.15504	0.10712	+ 23.947
SumBasicST-Set	0.09895	0.03476	0.10038	0.12920	0.09082	+ 5.0940

To deeply understand results, we also demonstrate the visual comparisons in Figure 5 under ROUGE-1 and ROUGE-2, respectively. It shows that the proposed T2ST-M and T2ST-Set methods both perform best among all methods, and IPR is about 10 % under ROUGE-1, 25 % under ROUGE-2. Overall speaking, our proposed framework is effective. We will further analyze the impact of social and temporal context as follows.

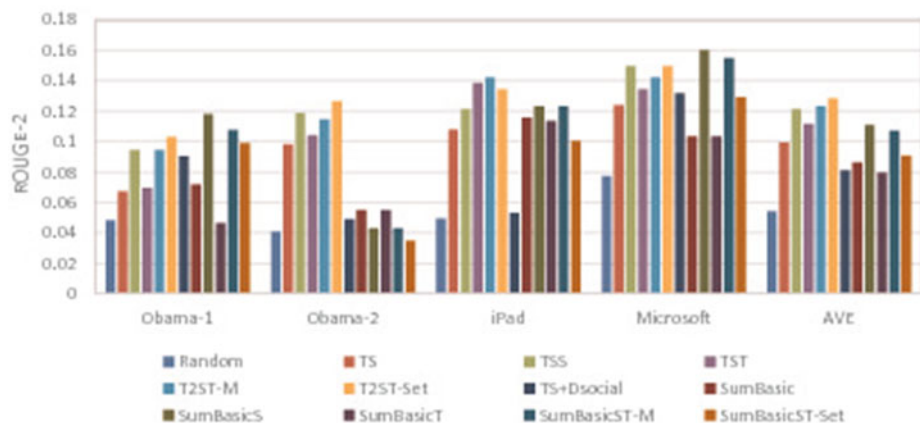
3.2.2 Impact of social context

The user authority $A(D_i)$ in (2) is introduced to exploit social context. Yet this popularity distribution is very unbalanced and authority features are highly skewed, e.g., the followers of a popular elite Twitter user could reach tens of millions, while most of the long tail Twitter users barely have more than 100 followers. Figure 6a shows an example on Obama-1 corpus. Due to the existence of absolute authority, we could not almost see the long tail Twitter user. Actually, it is not suitable for ordinary topic discussion. Therefore, we smooth the social context through degrading $AS(D_i)$ with very large value and normalize it by dividing the maximum $AS(D_i)$. Figure 6b is the smoothed and normalized social context on Obama-1 corpus. How much $AS(D_i)$ needed to degrade is up to the specific data set. This smoothing is important and very useful for extracting the summary tweet.

The only difference between TS and TSS is that TSS exploits the social context. Therefore by comparing the performance of TS and TSS, we can investigate the impact of social context for Twitter summarization. We note that TSS always obtains better performance than TS, which gains 8 % IPR under ROUGE-1 and 22 % IPR under ROUGE-2 in comparison with TS. In order to validate the generalization of this kind of social context, we do the experiments using SumBasicS, which gains 9 % IPR under ROUGE-1 and 28 % IPR under ROUGE-2 in contrast to SumBasic. It suggests that the user's authority and importance of tweets are highly correlated. Therefore such social context can provide an important cue for the discrimination of tweet content, and can significantly improve the performance of Twitter summarization.



(a) ROUGE-1



(b) ROUGE-2

Figure 5 Performance Comparison of Different Summarization Systems

In order to compare with the other social context, we do the experiment using TS+Dsocial, but its IPR decreases about 2.8 %. The reasons are three-fold. One main reason may be that this method had not done the smoothing processing like shown in Figure 6, which brings the unbalanced summary tweets since they are predominately from super authorities; The second is that the parameter d may be unreasonable; The third is that we could not fully implement their experiment, and the static user influence $F(u_i)$ in our reconstruction is simple. The special difficulty for not fully reconstructing this experiment includes two aspects: (1) we use the different data set, which might have the different detailed information; (2) for their topic-irrelevant static influence of a user, it is predicted by a linear SVM model. However, there is no specific preprocessing and parameter setting. The other reasons include that the random walk on graph formed by following relationships might be not work due to the very sparse user relationships and could not fully use the retweet information.

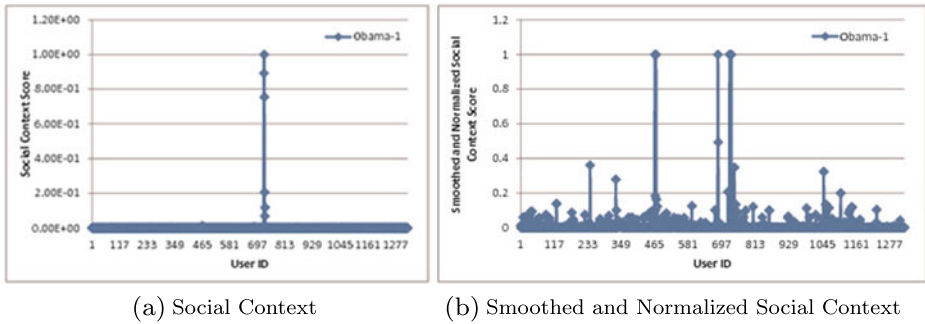


Figure 6 Curve of Social Context

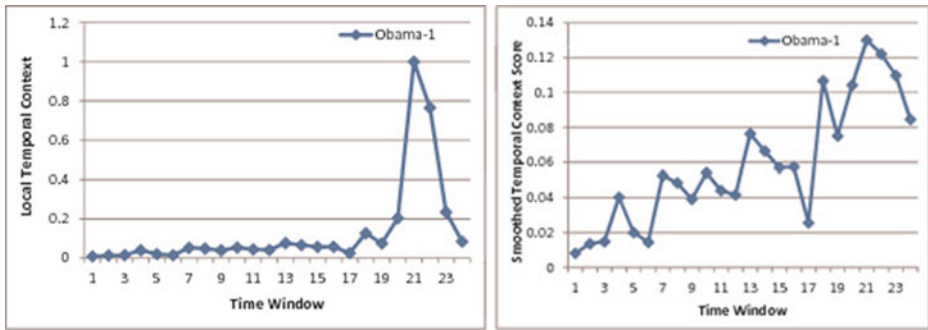
In our proposed method, $A(D_i)$ could not measure the follower and the friend quality, and only consider the static influence of the user. Meanwhile, the different users have different influences given a topic. We will further explore these dynamic factors in Twitter summarization.

3.2.3 Impact of local temporal context

Based on the global temporal context, we detect the key time points, and then extract summary from each key time point. The local heat signals of tweets $H(D_i)$ are introduced to exploit local temporal context. The time window is an hour. There are 24 local time points, which show the micro trend of one topic. Due to the regular life style and the topic burst in one day, such local temporal context shown in Figure 7a provides us a kind of cue for discriminating the importance of a tweet. However, the similar skewed phenomena also exists in local temporal context. When the number of posted tweets in a non-working time or a bursting time is very large, it will almost cover the function of other local time points. In order to avoid much extreme tweets entering into summary, we smooth the large peak area in a scaled way. Figure 7b is the smoothed and normalized local temporal context curve drawn from Obama-1 corpus. It is seen that we can observe the detailed temporal trend beyond the large peaks, which is useful for selecting the important tweets.

Compared to TS, TST exploits the local temporal context. Therefore via the performance comparison of TS and TST, we can study the effect of temporal context in Twitter summarization. We observe that TST often outperforms TS. It obtains 5 % IPR under ROUGE-1 and 12 % IPR under ROUGE-2 in comparison with the result of TS, respectively. Therefore such temporal context provides useful cues about the importance of a tweet on TST. The performance improvement suggests the importance of temporal context. However, the performance on SumBasic is not better. It only obtains 0.5 % IPR under ROUGE-1 and decreases 7.8 % under ROUGE-2. Meanwhile, the IPR of TST is not higher than that of TSS. The reasons are two-fold. One may be because that the local temporal signal is sparse in comparison with social context, which only has the 24 local time points. Thus the ranking score of many important tweet in content aspect is pulled down when meeting with the bad temporal context; the other is that this integrating method is specific on one method, and not suitable for the other method. In the future, we will explore the more generalized local temporal context.

In our current version, we only take into account the external information of a tweet, i.e., the publishing speed of tweets. The internal textual content information, such as the posting



(a) Temporal Context

(b) Smoothed Temporal Context

Figure 7 Curve of Temporal Context

speed of particular words and the accelerated speed, is not modeled in the heat signals, which might be exploited to further improve the performance.

3.2.4 Hybrid impact of collective social-local temporal context

We propose two methods to collectively integrate social-temporal context based on TS system, including T2ST-M and T2ST-Set. One is multiplying the effects in the basic TS algorithm. One is selecting the top tweets from both methods and concatenating them together. T2ST-M combines both social and temporal context into the transition matrix of random walk on graph in a multiplication way. T2ST-Set integrates social-temporal context in a set theory. On average, both of two methods improves the system performance and are effective on the TS series methods. T2ST-Set is more stable than T2ST-M and obtains the best performance at most of the time. The reasons are two-fold. One is that T2ST-Set combines the best top tweets from TSS and TST in a certain ratio, under the premise that both TSS and TST have a good performance. Consider that TSS has a larger IPR than TST on average, we let $\lambda : 1 - \lambda$ is about equal to 3 : 1, where summary length of tweets from TSS is about 75 % of the whole summary length, and the remaining is 25 %. The other is that if one of social or temporal context is relatively not good, the whole effect of T2ST-M through integrating in a multiplication is pulled down. Yet this situation does not bring too much impact for T2ST-Set.

Through observations, the performance improvement of T2ST-M and T2ST-Set indicates that social-temporal context may contain complementary information with each other. It is necessary to combine them collectively in an elegant way on the TS series methods. However, the performance on SumBasic is a little different. The whole IPR of SumBasicST-M and SumBasicST-Set is lower than that of T2ST-M and T2ST-Set. The performance of SumBasicST-Set is worse than that of SumBasicST-M. The reason is that SumBasicT does not improve too much or even becomes worse. Therefore, if SumBasicST-Set combines the summary tweets from the SumBasicT lower than the basic system, it must bring the worse performance. On the other hand, the performance of SumBasicST-M is also pulled down due to the sparse local temporal signal and one of the bad social or temporal context. Meanwhile, it shows that the fusion way of social-temporal context also depends on the specific baseline methods. If two systems have a huge performance gap, the fusion way in a set theory is not a good choice.

Table 6 Effect of wavelet denoising in time point detection

Keyword	MAP		
	TPD-WD	TPD	IPR (%)
Obama	0.32200	0.38680	+6.4800
iPad	0.13840	0.1789	+4.0500
Microsoft	0.19600	0.22160	+2.5600

The detailed summary examples on Microsoft corpus are shown in the Appendix. It is seen that an expert seems to select the more formal and informative tweets to be reference summary tweets if there is no guidance with social and temporal context. That is to say the experts tend to discriminate the tweet quality based on the content itself and their cognition from one topic. Even though experts carefully build the reference summary tweets, there still exist some inconsistency since everyone can have the different understanding about one topic. It is very challenging to construct these short text expert summaries. We will explore more reasonable ways to build the more Twitter summarization corpus. With the above evidence, we have shown that (1) exploiting social-local temporal context can improve the performance of Twitter summarization; (2) normalized smoothing social and local temporal context is significant due to the skewed context distribution; and (3) the performance improvement is from both social context and local temporal context, which are complementary with each other.

3.3 Impact of denoising in key time point detection for twitter summarization

Key time point detection plays an important role in Twitter summarization, which helps to capture the global temporal context and have a macro summary content selection. In this subsection, we investigate the impact of wavelet denoising in key time point detection by comparing the performances with and without the wavelet denoising. We use TPD and TPD-WD to denote the key time point detection algorithms with and without wavelet denoising. In this experiment, each algorithm selects 10 time points for each dataset. To implement the wavelet denoising, db3 wavelet is selected as the wavelet function. And after observing the denoised signal, we set the decomposition level (the maximum resolution) $m = 5$. We use MAP (mean average precision) value to assess the performance for key time point detection. The time point detection can be considered as a kind of information retrieval problem. Therefore, we can use precision in (14) to evaluate the algorithm.

$$Precision = \frac{1}{R} \sum_{i=1}^R \frac{i}{Rank(i)} \quad (14)$$

where R indicates the number of points that appear in both detected results and manual results, i indicates the rank of a time point provided by manual labels, while $Rank(i)$ refers to the rank suggested by the detection algorithm. By calculating the MAP value, we can judge the accuracy of the detection performance. For example, if there are 4 from the 10 time points detected by the algorithm, and their rankings are 1, 3, 6, and 10. In accordance with (14), we can calculate the precision value as $\frac{1}{4} * (\frac{1}{1} + \frac{2}{3} + \frac{3}{6} + \frac{4}{10}) = 0.642$. The higher the precision, the better the performance is.

Table 7 Event comparison before denoising (BD) and after denoising (AD)

Time points	Related events	Ranking BD	Ranking AD
01/12	Remarks by the President on the Shootings in Arizona	2	4
01/25	State of the Union	–	2
04/08	No major events	9	–
04/13	Text of Obama's deficit-plan speech	8	–
04/20	Obama gets wake-up call on credit outlook after S&P kept U.S. AAA credit rating	–	8
04/27	Obama releases long-form birth certificate	4	10
04/28	No major events, still related to birth certificate	7	–
05/01	Obama announced that U.S. Army killed Osama Bin Laden	1	1
05/02	No major events still related to Bin Laden	3	–
05/19	Obama delivers a policy address on events in the Middle East	10	6
06/22	Obama says 33,000 troops will leave next year	–	7
07/06	Obama kicks off Twitter town with a tweet	6	5
07/22	Obama ends ban on openly gay military service	–	9
07/25	Obama fundraisers postponed amid debt limit talks	5	3

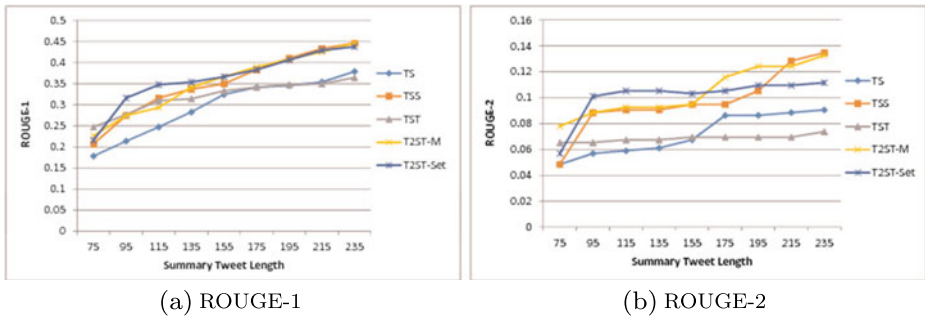


Figure 8 Performance Influence Curve of Summary Length on Obama-1

The MAP (mean average precision) values of TPD and TPD-WD are presented in Table 6. Note that in the table Improvement indicates the performance improvement of TPD from TPD-WD.

It can be seen that, the introduction of wavelet denoising can help improve the accuracy of key time point detection. To further illustrate the effect of wavelet denoising, Table 7 shows the detailed results of time point detection of heat signals with the topic “Obama” during the first half of 2011. We also asked a volunteer to write out the real events corresponding to the selected key time points. Since we have similar observations in other topics, we have not listed their experimental results.

As shown in the table, there is partially redundant information in the key time points selected before denoising. For example, on May 2, 2011, the discussion on Twitter is actually a continuation of the event that the U.S. Army shot Bin Laden. This can be understood as the noise caused by asynchronous publishing which we mentioned above, while this kind of redundancy is reduced after denoising. These results demonstrate the existence of irrelevant time points, and also suggest that denoising can improve the performance of key time point detection. Therefore, global temporal context is very useful to Twitter summarization, which helps to have a coarse content filtration.

3.4 Effect of tweet summary length

Constructing tweet expert summaries is a challenging task due to the short, free text style and cognition diversity among the different people. Therefore it is important to investigate the stability of the proposed framework with the changes of the length of summary. Figure 8 demonstrates how the system performance changes on Obama-1 with the increasing number of words L under ROUGE-1 and ROUGE-2, respectively. Here, we vary L from 75 to 235 with an increasing step of 20 words.

We make the following observations:

- with the increase of L , the performance first increases rapidly and increases gradually after a certain region;
- the turning point is basically where the system summary length is about equal to the average expert summary length, and the performances of TSS and T2ST-M increase a little faster under ROUGE-2 after the turning point, which further show that our social-temporal context is useful for extracting summary tweets.

To sum up, social-temporal context is effective on TS series methods for Twitter summarization. How to collectively integrate them still depends on specific data characteristics and the basic methods. Meanwhile, it is better to make the system and expert summary length match with each other in order to keep a good readability and system performance.

4 Related work

Twitter summarization plays an important role in helping users to efficiently acquire social media information, which has been attracting more and more attention in recent years [3, 5, 8, 20, 21, 26]. Most previous summarization studies have focused on the well-formatted news documents, as driven by DUC² and TAC³ evaluations. Existing methods can be classified from two aspects, (1) Unsupervised or supervised learning; (2) Extractive or abstractive style. Next we review the unsupervised and extractive summarization methods because our proposed framework belongs to this category.

4.1 Twitter summarization

Traditional summarization methods only consider text information and become difficult when handling Twitter data. Most tweets are relatively short and informal. Twitter summarization can not be considered as a combination of single document summarization. Automatic summarization algorithms targeted at micro-blog are continuations of the multi-document summarization (MDS) [8]. They normally treat a tweet as a sentence and then use the similar methods in MDS to extract summaries. Typical examples include SumBasic [17] and centroid algorithms [1] which produce a summary by simply computing the term frequency and inverse document frequency. SumBasic' underlying premise is that words that occur more frequently across documents have a higher probability of being selected for human created multi-document summaries than these that occur less frequently. The centroid algorithm measures a sentence with the pseudo-centroid topic of a document cluster. Another type of method is based on graph, such as LexRank [6] and TextRank [14]. These algorithms represent documents as weighted graphs to select salient ones for the final summary.

Recent research takes into account the integration of other factors including social context and content quality [5], KL-divergence [26] and tweet term frequency [20]. Some studies organize twitter streams according to clues such as events [26] or participants [21]. The introduction of various factors improves the performance of summarization algorithms to some extent. However social-temporal context is seldom touched in literature.

4.2 Key time point detection

Key time points refer to points where there is a significant number of events (e.g., tweets) occurring at the point. Usually, key time point detection is a preprocessing step for timeline summarization. It is related to event detection and its underlying assumption is that some related words would show an increase in the usage when an event is happening. Most existing work detects events by grouping a number of keywords with similar burst appearance.

²<http://duc.nist.gov/>

³<http://www.nist.gov/tac>

We can broadly classify the existing time point detection into two categories -(1) Temporal expression based approaches and (2) Text feature based approaches.

4.2.1 Temporal expression based approaches

Since the important events usually accompany with temporal expressions, such as “15th of this month” or “last Sunday”. This kind of term can be used to infer the importance of that time point. Kessler et al. [11] presented a method based on ISO-TimeML to extract temporal expressions from texts and rank the dates with these features. This method is a supervised classification and highly dependent on text contents, while tweets are short and usually do not contain temporal expressions.

4.2.2 Text feature based approaches

Text feature based approaches generally analyze the distribution of certain text features and evaluate each time point by a statistical way. Nichols [18] proposed a method using Twitter to summarize sports events. In their approach, the publishing volume of tweets is represented as a waveform, and the spike of the waveform is considered to describe a sport event. Shamma [19] used a slope based method to simply find the peaks of the term frequency signals. Most of methods mentioned above could not control global redundancy. Recently, Weng and Lee [25] introduced the wavelet analysis into the field of event detection by calculating the wavelet entropy to cluster a set of words with similar patterns of burst based on a graph partitioning technique. For global grasping information, event detection focused on clustering key words into several groups along the time line [25], which is different from Twitter summarization. In this paper, we are the first to employ wavelet analysis in exploiting temporal-context for Twitter summarization.

5 Conclusion and future work

Traditional document summarization methods become difficult when handling short and highly unstructured Twitter data. However, Twitter data provides rich context information over texts. In this paper, we study the problem of Twitter summarization with social-temporal context. We provide ways to exploit social and temporal context, and propose a novel Twitter summarization framework which incorporates social, local and global temporal contexts. The key time points are detected through the macro temporal analysis based on wavelet denoising, and then the summary tweets are extracted from each key time point. Extensive experiments are conducted to evaluate the proposed framework on a real-world dataset with manual labels. The experimental results demonstrate the effectiveness of our proposed framework as well as the importance of social-temporal context for Twitter summarization.

Meanwhile, it shows that (1) global temporal context is very useful to Twitter summarization, which helps to have a coarse content filtration; (2) during local content selection, normalized smoothing social and local temporal context is significant due to the skewed context distribution. And T2ST-Set performs better than T2ST-M on the TS method series. The reasons to keep both of them in the proposed method are two-fold: (a) the fusion way through set theory (T2ST-Set) performs better has a certain premise, where two single systems need to both have a relatively good performance; (b) if we have the different baseline methods, where the two single systems have a huge performance gap, the fusion way with

set theory does not perform well. At this moment, the fusion way on behalf of T2ST-M in a multiplying effect is not a bad way. Through the experimental results, we hope to show that the fusion ways are method specific, but not absolutely.

There are several interesting directions for further investigation. First, in our current work, we build heat signals with a time window of one hour; and we will investigate how to build heat signals with multi-scale. Second, we would like to introduce trust to better capture social context for Twitter summarization.

Acknowledgments This work was supported in part by National Key Basic Research and Development Program of China (973 Program) under Grant 2013CB329304, 2013CB329301, National Natural Science Foundation of China (Grant No:61100123,61472277), Ministry of Education Fund of China for the Doctoral (Grant No:20110032120040) and Tianjin Younger Natural Science Foundation (Grant No:14JCQNJC00400).

References

1. Aker, A., Plaza, L., Lloret, E., Gaizauskas, R.: Multi-Document Summarization Techniques for Generating Image Descriptions: a Comparative Analysis. In: Multi-Source, Multilingual Information Extraction and Summarization (2013)
2. Carbonell, J., Goldstein, J.: The Use of MMR, diversity-based reranking for reordering documents and producing summaries. In: Proceedings of ACM SIGIR (1998)
3. Chang, Y., Wang, X., Mei, Q., Liu, Y.: Towards Twitter Context Summarization with User Influence Models. Proceedings of WSDM (2013)
4. Daubechies, I.: The wavelet transform, time-frequency localization and signal analysis IEEE Transactions on Information Theory (1990)
5. Duan, Y., Chen, Z., Wei, F., Zhou, M., Shum, H.Y.: Twitter Topic Summarization by Ranking Tweets Using Social Influence and Content Quality. In: Proceedings of COLING (2012)
6. Erkan, G., Radev, D.R.: Lexrank: Graph-based lexical centrality as salience in text summarization JAIR (2004)
7. Gong, Y., Liu, X.: Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis. In: SIGIR (2001)
8. Inouye, D., Kalita, J.K.: Comparing Twitter Summarization Algorithms for Multiple Post Summaries. In: Socialcom (2011)
9. Johnstone, I.M., Silverman, B.W.: Wavelet threshold estimators for data with correlated noise. Journal of the Royal Statistical Society: Series B (Statistical Methodology) (1997)
10. Jones, K.S.: Automatic summarising: The state of the art Information Processing & Management (2007)
11. Kessler, R., Tannier, X., Hagege, C., Moriceau, V., Bittar, A.: Finding Salient Dates for Building Thematic Timelines. In: ACL (2012)
12. Lin, C.Y.: Rouge: A package for automatic evaluation of summaries. In: Text Summarization Branches Out: Proceedings of the ACL-04 Workshop
13. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation TPAMI (1989)
14. Mihalcea, R., Tarau, P.: TextRank: Bringing Order into Texts. In: Proceedings of EMNLP (2004)
15. Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J.: Wavelet toolbox. The MathWorks Inc., Natick, MA (1996)
16. Nenkova, A., McKeown, K.: A Survey of Text Summarization Techniques. In: Mining Text Data. Springer (2012)
17. Nenkova, A., Vanderwende, L.: The impact of frequency on summarization. Microsoft Research. Technical Report MSR-TR-2005-101, Redmond, Washington (2005)
18. Nichols, J., Mahmud, J., Drews, C.: Summarizing Sporting Events Using Twitter. In: IUI (2012)
19. Shamma, D.A., Kennedy, L., Churchill, E.F.: Peaks and Persistence: Modeling the Shape of Microblog Conversations. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work (2011)
20. Sharifi, B., Hutton, M.A., Kalita, J.: Summarizing Microblogs Automatically. In: Proceedings of HLT-NAACL (2010)

21. Shen, C., Liu, F., Weng, F., Li, T.: A Participant-Based Approach for Event Summarization Using Twitter Streams. In: NAACL-HLT (2013)
22. Shi, Z., Melli, G., Wang, Y., Liu, Y., Gu, B., Kashani, M.M., Sarkar, A., Popowich, F.: Question Answering Summarization of Multiple Biomedical Documents. In: Advances in Artificial Intelligence (2007)
23. Vanderwende, L., Suzuki, H., Brockett, C., Nenkova, A.: Beyond SumBasic: Task-focused summarization with sentence simplification and lexical expansion. *Information Processing & Management* (2007)
24. Wang, D., Li, T., Zhu, S., Ding, C.: Multi-Document Summarization via Sentence-Level Semantic Analysis and Symmetric Matrix Factorization. In: SIGIR (2008)
25. Weng, J., Lee, B.S.: Event Detection in Twitter. In: Proceedings of ICWSM (2011)
26. Zubiaga, A., Spina, D., Amigó, E., Gonzalo, J.: Towards real-time summarization of scheduled events from twitter streams. In: Proceedings of the 23rd ACM conference on Hypertext and Social Media (HT'12)