The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Twitter Summarization Based on Social Network and Sparse Reconstruction *

Ruifang He, Xingyi Duan
School of Computer Science and Technology, Tianjin University, Tianjin 300350, China
Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin 300350, China
{rfhe,xingyiduan}@tju.edu.cn

Abstract

With the rapid growth of microblogging services, such as Twitter, a vast of short and noisy messages are produced by millions of users, which makes people difficult to quickly grasp essential information of their interested topics. In this paper, we study extractive topic-oriented Twitter summarization as a solution to address this problem. Traditional summarization methods only consider text information, which is insufficient in social media situation. Existing Twitter summarization techniques rarely explore relations between tweets explicitly, ignoring that information can spread along the social network. Inspired by social theories that expression consistence and expression contagion are observed in social network, we propose a novel approach for Twitter summarization in short and noisy situation by integrating Social Network and Sparse Reconstruction (SNSR). We explore whether social relations can help Twitter summarization, modeling relations between tweets described as the social regularization and integrating it into the group sparse optimization framework. It conducts a sparse reconstruction process by selecting tweets that can best reconstruct the original tweets in a specific topic, with considering coverage and sparsity. We simultaneously design the diversity regularization to remove redundancy. In particular, we present a mathematical optimization formulation and develop an efficient algorithm to solve it. Due to the lack of public corpus, we construct the gold standard twitter summary datasets for 12 different topics. Experimental results on this datasets show the effectiveness of our framework for handling the large scale short and noisy messages in social media.

Introduction

Twitter has become one of the most popular social network platforms, through which amounts of users can freely produce content (called tweets) on their interested topics. However, the rapid growth of tweets makes it difficult for people to quickly grasp essential information. Twitter summarization aims to generate a succinct summary delivering the core information from a sheer volume of tweets in a given topic. It can be used to help ordinary people fastly acquire information and aid agencies monitor crisis progress so as to assist recovery and provide disaster relief.

Despite document summarization has been researched for many years, it is still a knotty problem due to the large scale short, noisy and informal nature of messages in social media, such as tweets. Existing Twitter summarization approaches usually regard tweets as sentences, and adopt traditional summarization methods (Inouye and Kalita 2011), such as SumBasic (Vanderwende et al. 2007), Centroid (Radev, Blair-Goldensohn, and Zhang 2001), LexRank (Erkan and Radev 2004) and TextRank (Mihalcea and Tarau 2004) to validate the relevant performance on microblogging posts. However, it is not clear whether adding the complexity of methods will improve the performance of Twitter summarization. Some other researches (Chang et al. 2016; Liu et al. 2012) explore to utilize static social features except for textual content, such as number of replies, number of retweets, number of likes, author popularity (i.e. number of followers for a given tweet's author) and temporal signals.

All the above methods ignore the fact that Twitter data is networked. There exist some researches (Chang et al. 2013; Duan et al. 2012) exploiting social network information. These approaches mainly consider network information from the user-level perspective, assuming that high authority users are more likely to post salient tweets. However, tweets are also potentially networked through user connections. Different from traditional methods, which obtain associated tweet information through measuring similarity between tweets purely based on content information, the networked tweets may contain more semantic clues than purely text-based methods. So we need to explore a new method for modeling the tweet-level networked information.

The social theories indicate the reciprocal influence of networked information. People themselves are more likely to keep the same sentiment (Hu et al. 2013), preference (Wang et al. 2015) on a specific topic in a short period, and this phenomenon is called **expression consistency**. Moreover, relationship between people is established through a series of interactions and feedbacks. The influence is subtle and can make a great impact on ones' preference, speaking manner or even expression content. Thus people gradually have similar viewpoints about a topic with their friends and show them with almost the similar tone and words, which is regarded as **expression contagion**. Inspired by these two

*This work is supported by NSFC (61472277) and partly supported by National Program on Key Basic Research Project (973 Program, 2013CB329301) and NSFC (61772361).

social theories, we explore how to utilize them for Twitter summarization.

Recently, sparse reconstruction based summarization methods have been proposed (He et al. 2012; Liu, Yu, and Deng 2016; Yao, Wan, and Xiao 2015), and show a significant performance on traditional corpus DUC/TAC. It is also because that social information can be seamlessly combined into the sparse reconstruction based method. In this paper, we propose to integrate social network into a unified optimization framework for Twitter summarization from the perspective of sparse reconstruction, through modeling the tweet-level networked information. It assumes that a good summary can best reconstruct the original corpus, and better address the coverage, sparsity and diversity of summary in social media. Our contributions are summarized as follows:

- From the statistical perspective, we verify the existence of two social theories in twitter data and formally define the problem of Twitter summarization to enable the utilization of social network;
- Model the tweet-level networked information as a social regularization through integrating social network into the sparse reconstruction-based method;
- Design the group sparsity regularization for Twitter summarization to keep salient tweets from the corpus-level, and the diversity regularization to avoid the more serious redundancy bought by social network;
- Construct 12 gold standard topic-oriented tweet datasets by asking 24 volunteers to manually select the most informative tweets, all in 48 expert summaries;
- We empirically evaluate the proposed SNSR framework on this datasets, elaborate the effectiveness of social network, and validate the new designed sparsity and diversity schemes.

Related Work

Our proposed method belongs to the extractive and unsupervised style. Therefore, we mainly review the relevant researches.

Multi-Document Summarization. Lots of traditional methods extract the result summary from top sentences with the highest scores, through assigning salient scores to sentences of the original document. The computation strategies of salience include: (1) Feature based methods, including Centroid and SumBasic, consider the frequency and the position of word to measure the sentence weight; (2) Graph based methods are the PageRank like algorithms, such as LexRank and TextRank built by random walk on sentence or word graph. However, these methods face the redundancy problem. Some researches propose to use cluster based strategies to keep the diversity of summary to avoid the redundancy (Cai et al. 2010; Wang et al. 2011; Shen, Li, and Ding 2010; Wang et al. 2009; Gao et al. 2012; Litvak et al. 2015). They mainly use topic modeling, cluster algorithms or matrix factorization to produce the more coverage summary. Recently, the appearance of sparse reconstruction based summarization methods, originally proposed by (He et al. 2012), brings us new possibility to resolve the classical challenges existed in summarization, including coverage, salience, and diversity. Further improved researches are contained in (Yao, Wan, and Xiao 2015; Liu, Yu, and Deng 2016). However, the large scale short and noisy texts in social media make these methods unsuitable for twitter.

Twitter Summarization. The prosperity of social media impels people to explore the adaptation of traditional summarization methods (Inouye and Kalita 2011) on twitter, including Hybrid TF-IDF model and phrase reinforcement algorithm to find the most commonly used phrase as summary (Sharifi, Hutton, and Kalita 2010; Nichols, Mahmud, and Drews 2012). All these methods only consider text information. However, social media platform can provide us much more rich information other than texts in Twitter. (Duan et al. 2012; Liu et al. 2012) extended the PageRank algorithm through incorporating social properties. (Alsaedi, Burnap, and Rana 2016) proposed three methods for event summarization, by using the temporal information and retweet information. (Chang et al. 2013; 2016) regarded Twitter summarization as a supervised classification task through mining rich social features, such as temporal signal and user influence. These approaches mainly use the static social information or user-level network information, and don't further explore tweet-level networked relations which may contain much more potential semantic clues.

Social Network Propagation. Social theories and social network analysis provide us useful insights to combine topology and content in Twitter summarization. Social network propagation also known as social influence or network influence has been researched in several domains, such as sentiment analysis (Hu et al. 2013), topic identification (Wang et al. 2015), topic detection (Bi, Tian, and Sismanis 2014), and network inference (He et al. 2015). From these researches, we know that sentiment and topic can spread along the network. In this paper, we will further explore how expression content, which is the carrier of sentiment and topic, can spread along the network and influence Twitter summarization. Social regularization considering social network propagation can be seamlessly integrated into sparse reconstruction based summarization methods. Therefore, we further study the coverage, salience and diversity challenges of summarization in social media from the sparse reconstruction perspective.

Problem Statement

Assume that the tweets corpus in a specific topic are represented as a weighted term frequency inverse tweet frequency matrix, denoted as $S = [t_1, t_2, \ldots, t_n] \in R^{m \times n}$, where m is the size of vocabulary and n is the total number of tweets. Each column t_i of S stands for a single tweet vector. $U \in R^{d \times n}$ denotes the user-tweet matrix, where $U_{ij} = 1$ means that the j*th* tweet is posted by the i*th* user. We construct the user-user matrix $F \in R^{d \times d}$ according to the following relationship, and $F_{ij} = 1$ indicates that the i*th* user is related to the j*th* user.

From the notation above, we formally describe Twitter summarization in short and noisy social media texts:

Table 1: Statistics of the Data sets

	Osama	Joplin	Mavs	Oslo
Date	0501	0522	0612	0722
# of Tweets	4780	2896	3859	4571
# of Users	1309	1082	1780	1026
Max Degree of Users	69	68	76	77
Min Degree of Users	1	1	1	1
Max Tweets Number of Users	42	93	92	56
Min Tweets Number of Users	2	1	1	2
Ave. Tweets per User	3.65	2.68	2.18	4.46
P-value(Consistency)	4.78e-125	2.1e-98	9.08e-211	2.62e-131
P-value(Contagion)	1.82e-33	6.6e-09	8.09e-08	4.98e-19

Given a topic oriented Twitter corpus C with their content S and social context including user-tweet matrix U and user-user matrix F, we aim to learn the reconstruction coefficient matrix W to automatically produce a summary.

Data and Observations

Due to the lack of public Twitter summarization corpus, in this section, we first introduce how to collect the data, and the construction scheme of ground truth corpus is listed in experiment section. Then, we explore whether the social theories can bring some motivating insights for Twitter summarization.

Data

We use the public Twitter data collected by University of Illinois[1] as the raw data. According to the hashtags, we extract twelve popular topics happening in May, June and July 2012, including politics, science and technology, sports, natural disasters, terrorist attacks and entertainment gossips. Each topic can have multiple hashtags, such as "#osama" and "#osamabinladen". Then we search the tweets which contain any of these hashtags or any of the keywords obtained by getting rid of "#" from hashtags. Through observing the topic trends of tweet number over time, there are mainly emergence and hot event. To consider expression consistency and expression contagion in a short time interval, we further collect tweets within five days after the emergencies occurred (e.g. Oslo terrorist attack) and tweets within five days before and after the hot event occurred (e.g. Harrypotter). After obtaining the topic-oriented data, we filter some tweets beforehand if they satisfy one of the following conditions:

- Appear more than one time (only remain one of them);
- The number of words is less than 3 other than hashtags, keywords, mentions, URL and stop words;
- The user of a certain tweet is independent of others;

Due to the limited space, the statistics of partial topics are shown in Table 1.

Observations of Social Theories for Twitter Summarization

Social theories, such as consistency (P.Abelson 1983) and contagion (Shalizi and Thomas 2011; Harrigan, Achananuparp, and Lim 2012), have been proved to be useful for social media mining (Harrigan, Achananuparp, and Lim 2012). The analysis indicates that the members of a social network often exhibit correlated behavior, sentiment and topic can be diffused through network. Consistency means that social behaviours conducted by the same person keep consistent in a short period of time. Contagion means that friends can influence each other. In this subsection, we investigate expression consistency and expression contagion under a given topic for Twitter summarization. In our work, we redefine and explore the consistency and contagion as:

- **Expression consistency**: Whether the tweets posted by the same user are more consistent than two randomly selected tweets?
- **Expression contagion**: Whether the two tweets posted by friends are more similar than the two randomly selected tweets?

To verify the two questions, we measure the distance between two tweets as $D_{ij} = \|t_i - t_j\|_2$, where t_i denotes the vector of the ith tweet. The more similar the two tweets, the more D_{ij} tends to 0. For the first question, we construct two vectors named as $cons_c$ and $cons_r$ with equal number of elements. Each element of the first vector is obtained by calculating the distance of two tweets posted by the same user, and each element of the second vector is obtained by calculating the distance of two randomly selected tweets. Then we conduct the two-sample t-test on the two vectors $cons_c$ and $cons_r$. The null hypothesis is that there is no difference between the two vectors, $H_0 : cons_c = cons_r$. The alternative hypothesis is that the distance between two tweets posted by the same user is less than that of those randomly selected tweets, $H_1 : cons_c < cons_r$.

Similarly, to ask the second question, we construct two vectors named as $cont_c$ and $cont_r$ with equal number of elements. Each element of the first vector is obtained by calculating the distance of two tweets posted by friends, and each element of the second vector is obtained by calculating two randomly selected tweets. We also conduct the two-sample t-test on the two vectors $cont_c$ and $cont_r$. The null hypothesis $H_0 : cont_c = cont_r$, shows that there is no difference between two tweets posted by friends and those randomly selected tweets. The alternative hypothesis $H_1 : cont_c < cont_r$, shows that the distance between two tweets posted by friends is less than those randomly selected tweets. For all the topics, the consistency null hypothesis and the contagion null hypothesis are rejected respectively at significance level $\alpha = 0.01$ with p-values presented in the last two rows of Table 1.

This observation provides strong evidence for the existence of expression consistency and expression contagion. In the next section, we elaborate how to exploit these social theories for Twitter summarization.

Our Approach

The large scale short and noisy texts in social media bring more serious data sparseness, and content redundancy due to social network propagation. We investigate the new challenges from Twitter summarization through the perspective of sparse reconstruction.

[1] https://wiki.illinois.edu/wiki/display/forward/Dataset-UDI-TwitterCrawl-Aug2012

SR - Sparse Reconstruction

Coverage Through treating Twitter summarization as an issue of sparse reconstruction, original corpus should be reconstructed by summary tweets. Given the original corpus C, we can formally describe the reconstruction process as:

$$\min_{W} \frac{1}{2}\|S - SW\| \tag{1}$$

Where $W = [W_{*1}, W_{*2}, \ldots, W_{*n}] \in R^{n\times n}$ is the reconstruction coefficient matrix, and each column $W_{*j} = [W_{1j}, W_{2j}, \ldots W_{nj}]$ is a vector of coefficients used for representing tweet t_j, and each element W_{ij} of W_{*j} denotes the proportion of tweet t_i in reconstructing tweet t_j. To consider the original tweets can be regarded as the non-negative linear combination of summary tweets, we add constraint $W \geq 0$. To avoid tweet being reconstructed by itself, saying the reconstruction matrix $W \approx I$, we add the constraint $diag(W) = 0$. The formula Eq.(1) can be transformed as:

$$\begin{aligned} &\min_{W} \frac{1}{2}\|S - SW\| \\ &s.t. \quad W \geq 0,\ diag(W) = 0 \end{aligned} \tag{2}$$

Sparsity - Group Lasso Due to that we only select a few tweets as a summary to reconstruct the original corpus, and thus not all the tweets have an impact on reconstructing one certain tweet. Lots of coefficients of each column should tend to be zero. Inspired by sparse coding (Ye and Liu 2012), we regard each tweet as a group. The problem turns into selecting a subset of these groups to reduce the reconstruction loss. We add $l_{2,1}$ norm constraint on W, so the entire objective function is transformed as:

$$\begin{aligned} &\min_{W} \frac{1}{2}\|S - SW\| + \lambda \|W\|_{2,1} \\ &s.t. \quad W \geq 0,\ diag(W) = 0 \end{aligned} \tag{3}$$

Where

$$\|W\|_{2,1} = \sum_{i=1}^{n} \|W(i,:)\|_2 \tag{4}$$

$$and \quad \|W(i,:)\|_2 = \sqrt{\sum_{j=1}^{n} \|W_{ij}\|^2} \tag{5}$$

SN - Model Networked Tweet-level Information

To reduce the reconstruction error and make a rectification during the process of reconstruction, we exploit social theories to model the networked tweet-level information as a social regularization. This means two correlated tweets which are originally close should keep close during the reconstruction. Essentially, we need to build a **graph Lasso** (Ye and Liu 2012).

In order to utilize the social relations for Twitter summarization, we model the two mentioned social theories to construct tweet-tweet correlation graph through transforming the user-tweet relations and social relations into tweet-tweet correlation relations. Given user-tweet matrix U and user-user matrix F, tweet-tweet correlation matrix for **expression consistency** T_{cons} is defined as $T_{cons} = U^T \times U$, where $T_{cons} = 1$ denotes two tweets are posted by the same user. For **expression contagion**, T_{cont} is defined as $T_{cont} = U^T \times F \times U$, where $T_{cont} = 1$ denotes two tweets are posted by friends. Then we obtain the tweet-tweet correlation matrix, which can be either T_{cons}, T_{cont} or the combination $T = T_{cons} + dT_{cont}$, where d is a balance parameter between the two relations. In this paper, we can simply set $d = 1$ to construct a relation matrix. $T_{ij} = 1$ denotes two tweets have a correlated connection, otherwise $T_{ij} = 0$. We define the reconstruction matrix of S as $\hat{S} = SW$, so graph Lasso penalty term, namely social regularization is formulated as:

$$\begin{aligned} \Omega_{graph} &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} T_{ij}\|\hat{S}_{*i} - \hat{S}_{*j}\| \\ &= \sum_{i=1}^{m} \hat{S}_{i*}\|D - T\|\hat{S}_{i*}^T \\ &= tr(SWLW^TS^T) \end{aligned} \tag{6}$$

Where $tr(\cdot)$ denotes trace of a matrix, $L = D - T$ is the Laplacian matrix, $D \in R^{n\times n}$ is a diagonal matrix with $D_{ii} = \sum_{j=1}^{n} T_{ij}$, and each diagonal element denotes the degree of a tweet in matrix T.

Finally, we incorporate the social regularization Eq.(6) into Eq.(3) as:

$$\begin{aligned} &\min_{W} \frac{1}{2}\|S - SW\| + \alpha tr(SWLW^TS^T) \\ &\quad + \lambda \|W\|_{2,1} \\ &s.t. \quad W \geq 0,\ diag(W) = 0 \end{aligned} \tag{7}$$

where α is the parameter of social regularization.

Diversity-Model the Redundancy Information

Redundancy removal has always been the focus of summarization researches. Social studies show (Harrigan, Achananuparp, and Lim 2012) that reciprocal ties and certain triadic structures substantially increase social contagion, yet which also brings the more inherent redundancy and the lack of novelty of messages in a certain social network. Therefore, this kind of redundancy challenge will be more serious than that of traditional summarization.

Sparse reconstruction based methods tend to select tweets that cover the whole corpus, yet there is no explicit tendency to select tweets containing different aspects of a topic. (Liu, Yu, and Deng 2016) introduced a correlation term to control the diversity, and their optimization process is very complex. (Yao, Wan, and Xiao 2015) introduced a dissimilarity matrix, which greatly reduces the optimization complexity. However, the computational method of this matrix is not suitable for tweets due to the nature of the large scale short and noisy texts in social media, since it measures the encoding cost for each word with sentence length or vocabulary size. This method makes each dissimilarity value pretty large, and leads each element of W closing to zero, so it is confused to identify the salience of a tweet.

Inspired by the dissimilarity matrix, we introduce a relatively simple but effective cosine similarity matrix ∇, and each element $\nabla_{ij} \in [0, 1]$ denotes the cosine similarity between tweet t_i and tweet t_j. In the process of sparse reconstruction, we add constraint $diag(W) = 0$ to avoid tweets reconstructing themselves. Based on this knowledge, we have reason to avoid tweets being reconstructed by those tweets pretty similar to them. Considering the example below:

- **Tweet1:** the mood was solemn at the garden of reflection in lower makefield following the death of osama bin laden. video: http://fb.me/tof3pqok
- **Tweet2:** the mood was solemn at the garden of reflection in lower makefield following osama bin laden's death. video: http://bit.ly/l9tvdw

Obviously the above two tweets are similar to each other, it will lead that both of the reconstruction coefficients W_{12} and W_{21} close to 1. Therefore, this fact raises the salience of two tweets throughout the corpus, and brings more redundancy. Through the preliminary experiments, we can discover lots of similar pairs presented in the final summary without handling the diversity. To better avoid the "similar" reconstruction phenomena, we design ∇ as:

$$\nabla_{ij} = \begin{cases} 1 & \text{if} \quad \nabla_{ij} \geq \theta, \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where θ is the threshold used to distinguish the similar pairs and normal pairs. Then we formally introduce the diversity regularization term,

$$tr(\nabla^T W) = \sum_{i=1}^{n} \sum_{j=1}^{n} \nabla_{ij} W_{ij}$$

into Eq.(7), so the objective function is transformed as:

$$\begin{aligned} \min_{W} \ & \frac{1}{2}\|S - SW\| + \alpha tr(SWLW^T S^T) \\ & + \gamma tr(\nabla^T W) + \lambda \|W\|_{2,1} \\ s.t. \quad & W \geq 0, \ diag(W) = 0 \end{aligned} \tag{9}$$

where γ is the parameter of diversity regularization term.

By solving Eq.(9), the ranking score of tweet t_i is calculated as:

$$Score(t_i) = \|W(i, :)\|_2$$

We select tweets according to this ranking score to form the final summary.

Optimization Algorithm for SNSR

Inspired by (Ji and Ye. 2009; Nesterov and Nesterov 2004; Hu et al. 2013), we derive an efficient algorithm to solve the optimization problem in Eq.(9). Objective function can be equivalently expressed as:

$$\begin{aligned} \min_{W} \ f(W) = \ & \frac{1}{2}\|S - SW\| + \alpha tr(SWLW^T S^T) \\ & + \gamma tr(\nabla^T W) \\ s.t. \qquad & W \geq 0, \ diag(W) = 0, \ \|W\|_{2,1} \leq z \end{aligned} \tag{10}$$

where $z \geq 0$ is the radius of the $\ell_{2,1}-$ball, and there is a one-to-one correspondence between λ and z.

Algorithm 1 An Efficient Optimization Algorithm for SNSR

Input: $S, U, F, \nabla, W_0, \alpha, \gamma, \lambda, \theta, \epsilon$
Output: W

1: Initialize $\mu_0 = 0, \mu_1 = 1, W_1 = W_0, lr = 0.1$
2: $A = U^T \times U + U^T \times F \times U, L = D - A, \nabla = \nabla \geq \theta$
3: **for** iter = 0,1,2,... **do**
4: $\quad V = W_1 + \frac{\mu_0 - 1}{\mu_1}(W_1 - W_0)$
5: $\quad \frac{\partial f(W)}{\partial W} = S^T SW_1 - S^T S + \gamma\nabla + \alpha S^T SW_1 L$
6: $\quad$ **loop**
7: $\quad\quad U = V - \frac{1}{lr}\frac{\partial f(W)}{\partial W}$
8: $\quad\quad$ **for** $each\ row\ U_{i*}\ of\ U$ **do**
9: $\quad\quad\quad W_{i*} = S_{\lambda/lr}(U_{i*})$
10: $\quad\quad$ **end for**
11: $\quad\quad W = W - diag(W), W = max(W, 0)$
12: $\quad\quad$ **if** $f(W) \leq G_{lr,V}(W)$ **then**
13: $\quad\quad\quad break$
14: $\quad\quad$ **end if**
15: $\quad\quad lr = 2 * lr$
16: $\quad$ **end loop**
17: $\quad Set\ funVal(iter) = f(W) + \lambda\|W\|_{2,1}$
18: $\quad$ **if** $|funVal(iter) - funVal(iter - 1)| \leq \epsilon$ **then**
19: $\quad\quad break$
20: $\quad$ **end if**
21: $\quad W_0 = W_1$
22: $\quad W_1 = W$
23: $\quad \mu_0 = \mu_1$
24: $\quad \mu_1 = \frac{(1 + \sqrt{1 + 4\mu_1^2})}{2}$
25: **end for**

We omit details of our mathematical derivations due to the limited space. Interested reader may reference (Hu et al. 2013) and SLEP package[2] (Sparse Learning with Efficient Projects). The entire algorithm is described in Algorithm 1, where

$$\begin{aligned} G_{lr,V}(W) = f(V) + tr(\left(\frac{\partial f(V)}{\partial V}\right)^T (W - V)) \\ + \frac{lr}{2}\|W - V\|_F^2 \end{aligned} \tag{11}$$

and the shrinkage operator in Line 9 is defined as:

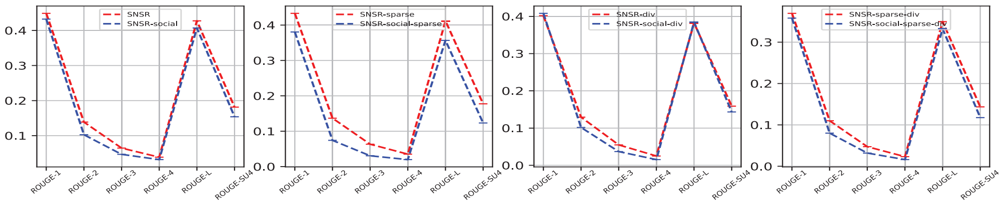
$$S_{\lambda/lr} = max(1 - \frac{\lambda}{lr\|U_{i*}\|_2}, 0)U_{i*} \tag{12}$$

Experiments

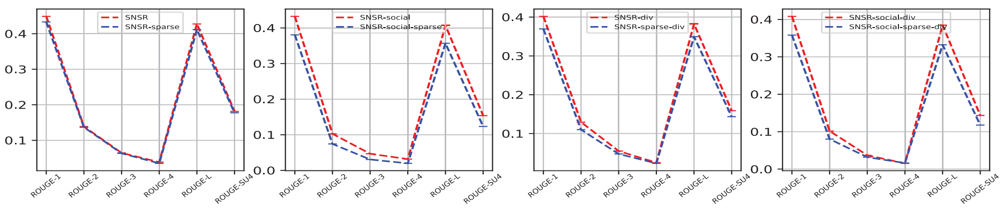
Ground Truth and Evaluation Metric

In order to evaluate our approach, we construct the ground truth (expert summaries) **C**orpus for **T**witter **S**ummarization (CTS). For each of the twelve topics, we ask four volunteers

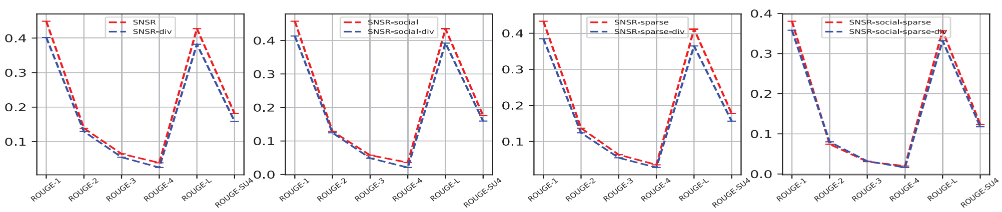
[2] http://www.yelab.net/software/SLEP/



(a) Social influence. From left to right are SNSR and SNSR-social, SNSR-sparse and SNSR-social-sparse, SNSR-div and SNSR-social-div, SNSR-sparse-div and SNSR-social-sparse-div



(b) Sparse influence. From left to right are SNSR and SNSR-sparse, SNSR-social and SNSR-social-sparse, SNSR-div and SNSR-sparse-div, SNSR-social-div and SNSR-context-sparse-div



(c) Diversity influence. From left to right are SNSR and SNSR-div, SNSR-social and SNSR-social-div, SNSR-sparse and SNSR-sparse-div, SNSR-social-sparse and SNSR-social-sparse-div

Figure 1: The influences of social regularization, sparse regularization and diversity regularization

to selects 25 tweets as a summary respectively, altogether 48 expert summaries. Then we ask three other volunteers to score all summaries based on coverage, diversity and writing quality of tweets in range $[1, 5]$. If only 0-6 tweets are satisfactory, then this summary is scored as 1, 12 as 2, 18 as 3, 24 as 4, and if all the tweets are good, we score it 5. The higher score, the more possible it is a better summary. We remain the summaries whose scores are greater or equal to 3, and require those low-quality summaries to be modified until they are eligible.

We use ROUGE as our evaluation metric (Lin 2004), which measures the overlapping N-grams between expert summaries and the model summary. In our experiment, we report the F-measures of ROUGE-1, 2, and ROUGE-SU4.

Performance Evaluation

Since our model belongs to the extractive and unsupervised style, we only compare with the relevant systems, including text and sparse reconstruction based methods. The upper bound of human summary and the baselines shown in Table 2 are as follows:

Expert denotes the average mutual assessment of expert summaries. **Random** selects tweets randomly; **Centroid** (Radev, Blair-Goldensohn, and Zhang 2001) ranks tweets by calculating the similarity between tweets and pseudo-center tweet; **LexRank** (Erkan and Radev 2004) ranks tweets by the PageRank like algorithm; **LSA** (Gong and Liu. 2001) exploits SVD (singular value decomposition) to decompose the TF-IDF matrix, and then selects the highest ranked tweets from each right singular vector; **NNMF** (Park et al. 2007) performs non-negative matrix factorization on the TF-IDF matrix, and chooses tweet with the maximum probability in each cluster; And two other methods based on sparse reconstruction are namely **DSDR** (He et al. 2012), and **MDS-Sparse** (Liu, Yu, and Deng 2016). The **SNSR-div**, **SNSR-sparse**, **SNSR-social** are the degradation models of **SNSR**, "-" denotes deleting the corresponding diversity, group sparse and social regularizations from our model **SNSR**. The "-" setting rules of models in Figure 1 are similar.

Through the overall comparisons seen in Table 2, we have the following observations:

- Our model outperforms all the baselines, and is below the upper bound. Yet it is worthy to be noted that all the methods make a high performance, especially for ROUGE-1. This phenomenon can be explained that our task is topic-oriented and we collect tweets according to the hashtags, thus tweets tend to have the coherent content. It also may be due to that we conduct an effective preprocessing;
- Among all the comparison experiments, the methods us-

Table 2: Performance on the Twitter data

System	ROUGE-1	ROUGE-2	ROUGE-SU4
Expert	0.47814	0.16337	0.20389
Random	0.41701	0.09439	0.14231
Centroid	0.38190	0.12384	0.15668
LexRank	0.42046	0.13273	0.17366
LSA	0.43474	0.13023	0.16625
NNMF	0.43784	0.13321	0.17433
DSDR	0.43236	0.12946	0.16521
MDS-Sparse	0.42240	0.10060	0.14666
SNSR-div	0.40191	0.12940	0.15894
SNSR-sparse	0.43327	0.13692	0.17749
SNSR-social	0.43236	0.10271	0.15379
SNSR	**0.44887**	**0.13882**	**0.18147**

ing matrix factorization, especially NNMF shows a comparable performance. The probable reasons come from two aspects: (1) NNMF can also be regarded as a reconstruction method, furthermore, it is similar to (Li et al. 2017) that exploits aspect term vectors to reconstruct the original term space; (2) It solves coverage and diversity challenges to some degree by mining sub-topics.

- In comparison with three degradation models, both social regularization, diversity regularization are useful for SNSR, and group sparsity regularization is also effective in obtaining salient tweets patterns from corpus-level rather than cluster-level.
- Through observing the last four rows in Table 2, we can discover that social regularization has an obviously effect on ROUGE-2 and ROUGE-SU4, and diversity regularization dose well on ROUGE-1.

In summary, our SNSR method achieves the better performance. It suggests that integrating social network information into the proposed sparse reconstruction framework helps improve Twitter summarization. Mining the group sparsity patterns of salient tweets and designing the diversity regularization in terms of redundancy brought by social network are also effective.

Effect of Different Regularization

To further evaluate the effectiveness of (1) **Social regularization**, we conduct four groups of comparison experiments, seen in Figure 1(a). For each group, two models are presented. In addition, we conduct similar evaluations on (2) **sparse regularization** and (3) **diversity regularization**, seen in Figure 1(b) and Figure 1(c) respectively. By analyzing experiment results, we have the following observations:

- The experiment performance will drop down without any of these three terms. It demonstrates the effectiveness of adding social regularization, group sparse regularization and diversity regularization.
- For each regularization, we compute the average growth percentage under ROUGE-1, ROUGE2 and ROUGE-SU4 in comparison with all of the corresponding degradation models. Seen from Table 3, social regularization has the greatest influence for the entire model, secondly sparse regularization and then diversity regularization term. Especially for ROUGE-2 and ROUGE-SU4, adding social regularization outperforms the degradation models of 46.01% and 23.69% respectively, which demonstrates that adding social regularization tends to select tweets close to expert summaries.

Table 3: Average improvement of different regularization in SNSR over the degradation models

Regularization	ROUGE-1	ROUGE-2	ROUGE-SU4
social	4.84%	46.01%	23.69%
sparse	9.98%	21.03%	14.87%
diversity	10.27%	6.34%	12.51%

- It is noted that diversity regularization makes a better performance in ROUGE-1 than the other two terms, which demonstrates that redundancy removal indeed decreases duplicated words and increases words coverage in comparison with expert summaries.

For social regularization, we just simply set $d = 1$, and have not discussed the different influences of consistency relation and contagion relation due to the limited space. They will be further explored in the future.

Parameters Settings and Tunings

In our experiment, there are mainly four parameters to be analyzed: social regularization α, diversity regularization γ, group sparse regularization λ and diversity threshold parameter θ. We tune four parameters greedily through setting step size, such as α in range $[0, 1]$, setting step size as 0.01.

Through preliminary experiments, we set $\alpha = 0.03$, $\lambda = 1$, and $\gamma = 1$. And we set the similarity threshold $\theta = 0.1$, which is consistent with the observation that the similarity between sentences is mostly distributed in the area of $[0, 0.1]$. Through this, we try to avoid the "similar" reconstruction phenomena.

Conclusion

The large scale short and noisy texts in social media bring new challenges for summarization, which make traditional document summarization methods unsuitable for twitter. In this paper, we study Twitter summarization from the sparse reconstruction perspective, and propose a SNSR framework. Social network makes tweets be correlated by user relations and also bring more serious coherent redundancy. Therefore, we model the networked tweet-level information by social relations as a social regularization, and integrate it into the sparse optimization framework. Meanwhile, we also design the diversity regularization to avoid the redundancy, especially due to the inherent redundancy brought by social network and the consistent property of topic-oriented corpus. And we use group sparse regularization to extract the better tweets patterns to form summary from the corpus-level, mining the salient tweets as well as keeping the diverse tweets. Apart from this, we construct the CTS corpus. Experimental results on this corpus show that our model achieves the better performance and the proposed framework is effective.

References

Alsaedi, N.; Burnap, P.; and Rana, O. 2016. Automatic summarization of real world events using twitter. *The Tenth International AAAI Conference on Web and Social Media*.

Bi, B.; Tian, Y.; and Sismanis, Y. 2014. Scalable topic-specific influence analysis on microblogs. *ACM International Conference on Web Search and Data Mining*.

Cai, X.; Li, W.; Ouyang, Y.; and Yan, H. 2010. Simultaneous ranking and clustering of sentences: A reinforcement approach to multi-document summarization. *COLING*.

Chang, Y.; Wang, X.; Mei, Q.; and Liu, Y. 2013. Towards twitter context summarization with user influence models. *WSDM*.

Chang, Y.; Tang, J.; Yin, D.; Yamada, M.; and Liu, Y. 2016. Timeline summarization from social media with life cycle models. *IJCAI*.

Duan, Y.; Chen, Z.; Wei, F.; Zhou, M.; and Shum, H.-Y. 2012. Twitter topic summarization by ranking tweets using social influence and content quality. *COLING*.

Erkan, G., and Radev, D. R. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research 22 (2004) 457-479*.

Gao, D.; Li, W.; Ouyang, Y.; and Zhang, R. 2012. Lda-based topic formation and topic-sentence reinforcement for graph-based multi-document summarization. *AIRS*.

Gong, Y., and Liu., X. 2001. Generic text summarization using relevance measure and latent semantic analysis. *In Proc. of the 24th ACM SIGIR, 19-25. ACM.*

Harrigan, N.; Achananuparp, P.; and Lim, E.-P. 2012. Influentials, novelty, and social contagion: The viral power of average friends, close communities, and old news. *Social Networks, 34 (2012):470-480*.

He, Z.; Chen, C.; Bu, J.; Wang, C.; Zhang, L.; Cai, D.; and He, X. 2012. Document summarization based on data reconstruction. *AAAI*.

He, X.; Rekatsinas, T.; Foulds, J.; Getoor, L.; and Liu, Y. 2015. Hawkestopic: A joint model for network inference and topic modeling from text-based cascades. *ICML*.

Hu, X.; Tang, L.; Tang, J.; and Liu, H. 2013. Exploiting social relations for sentiment analysis in microblogging. *WSDM*.

Inouye, D., and Kalita, J. 2011. Comparing twitter summarization algorithms for multiple post summaries. *Socialcom*.

Ji, S., and Ye., J. 2009. An accelerated gradient method for trace norm minimization. *ICML*.

Li, P.; Wang, Z.; Lam, W.; Ren, Z.; and Bing, L. 2017. Salience estimation via variational auto-encoders for multi-document summarization. *AAAI*.

Lin, C.-Y. 2004. Rouge: A package for automatic evaluation of summaries. *In Proceedings of Workshop on Text Summarization Branches Out, Post-Conference Workshop of ACL 2004*.

Litvak, M.; Vanetik, N.; Liu, C.; Xiao, L.; and Savas, O. 2015. Improving summarization quality with topic modeling. *CIKM*.

Liu, X.; Li, Y.; Wei, F.; and Zhou, M. 2012. Graph-based multi-tweet summarization using social signals. *COLING*.

Liu, H.; Yu, H.; and Deng, Z. 2016. Multi-document summarization based on two-level sparse representation model. *AAAI*.

Mihalcea, R., and Tarau, P. 2004. Textrank: Bringing order into texts. *In Proceedings of the 2004 conference on Empirical Methods in Natural Language Processing*.

Nesterov, Y., and Nesterov, I. 2004. Introductory lectures on convex optimization. *A basic course*.

Nichols, J.; Mahmud, J.; and Drews, C. 2012. Summarizing sporting events using twitter. *IUI*.

P.Abelson, R. 1983. Whatever became of consistency theory? *Personality and Social Psychology Bulletin, Vol. 9 No. 1, March 1983 37-54*.

Park, S.; Lee, J.-H.; Kim, D.-H.; and Ahn, C.-M. 2007. Multi-document summarization based on cluster using non-negative matrix factorization. *In Proceedings of Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM)*.

Radev, D. R.; Blair-Goldensohn, S.; and Zhang, Z. 2001. Experiments in single and multidocument summarization using mead. *First Document Understanding Conference*.

Shalizi, C. R., and Thomas, A. C. 2011. Homophily and contagion are generically confounded in observational social network. *Sociological Methods & Research, 2011 , 40 (2) :211*.

Sharifi, B.; Hutton, M.; and Kalita, J. 2010. Summarizing microblogs automatically. *HLT-NAACL*.

Shen, C.; Li, T.; and Ding, C. H. Q. 2010. Integrating clustering and multi-document summarization by bi-mixture probabilistic latent semantic analysis (plsa) with sentence bases. *AAAI*.

Vanderwende, L.; Suzukia, H.; Brocketta, C.; and Nenkovab, A. 2007. Beyond sumbasic: Task-focused summarization with sentence simplification and lexical expansion. *Information Processing and Management 43(2007):1606-1618*.

Wang, D.; Zhu, S.; Li, T.; and Gong, Y. 2009. Multi-document summarization using sentence-based topic models. *ACL-IJCNLP 2009 Conference Short Papers* 297–300.

Wang, D.; Zhu, S.; Li, T.; Chi, Y.; and Gong, Y. 2011. Integrating document clustering and multidocument summarization. *KDD*.

Wang, X.; YingWang; Zuo, W.; and Cai, G. 2015. Exploring social context for topic identification in short and noisy texts. *AAAI*.

Yao, J.; Wan, X.; and Xiao, J. 2015. Compressive document summarization via sparse optimization. *IJCAI*.

Ye, J., and Liu, J. 2012. Sparse methods for biomedical data. *SIGKDD Explorations Newsletter of the Special Interest Group on Knowledge Discovery & Data Mining* 14(1):4–13.