PAPER Special Section on Data Engineering and Information Management

Predicting Political Orientation of News Articles Based on User Behavior Analysis in Social Network

Jun-Gil KIM†, *Nonmember and* **Kyung-Soon LEE**†a), *Member*

SUMMARY News articles usually represent a biased viewpoint on contentious issues, potentially causing social problems. To mitigate this media bias, we propose a novel framework for predicting orientation of a news article by analyzing social user behaviors in Twitter. Highly active users tend to have consistent behavior patterns in social network by retweeting behavior among users with the same viewpoints for contentious issues. The bias ratio of highly active users is measured to predict orientation of users. Then political orientation of a news article is predicted based on the bias ratio of users, mutual retweeting and opinion analysis of tweet documents. The analysis of user behavior shows that users with the value of 1 in bias ratio are 88.82%. It indicates that most of users have distinctive orientation. Our prediction method based on orientation of users achieved 88.6% performance in accuracy. Experimental results show significant improvements over the SVM classification. These results show that proposed detection method is effective in social network.

key words: *social network, user behavior analysis, political orientation, bias ratio, mutual retweet*

1. Introduction

Social media has become a great medium for communicating, socializing and engaging in arguments and debates [1]. The revolution in social data is the shift in human communication patterns towards increased individual information and opinion sharing on the social network service (SNS) such as Twitter and Facebook [2]. Twitter allows people to be a one-man medium sending their opinions simultaneously to their followers [3].

The social media has played a huge role in our society. However, media bias can occur in several reasons: whenever newspaper provides out of context facts to support a particular party or when journalists add personal opinion to the news report. News articles frequently represent a biased viewpoint on contentious issues. This potentially causes social problems [5], [6]. Providing predicted viewpoints of news articles help social users see issues from different perspectives.

We have observed following communication behavior between users in social media streams such as Twitter:

- When a social issue occurs, the number of tweets is rapidly increased since twitter users post tweets and retweets with or without opinion on the issue news article [7], [8]. The retweeting behavior without any opinion can be seen as sympathy for the opinion of the original tweet.
- Highly active users have rich relationships to other users by retweeting behavior to express their sympathy.
- Users' sentiments in tweets depends on their political orientation and the viewpoint of the news article

Based on the observation, our assumptions are as follows: (1) Highly active users may have consistent behavior patterns in social network. (2) Rarely active users' orientation may be estimated based on the relationships to the highly active users (3) The orientation of a news article depends on that of users which post a tweet to the news article.

In this paper, we propose a method for predicting orientation of news articles by measuring the orientation of users who posted tweets regarding to the issue. The motivation of our research is to provide social users with different perspectives on contentious issues.

We have conducted experiments to validate the proposed method by gradually constructing social user network with about 600,000 users and 14,000,000 tweets for 280 contentious news articles on 25 issues in Korean Twitter. Experimental results showed the significant improvement over the baseline SVM classification.

The rest of the paper is organized as follows: Section 2 presents related work. Section 3 describes a prediction method of users' political orientation based on user behavior by constructing social network and shows user behavior analysis. Section 4 describes a prediction method of orientation of news articles and shows the experimental results. We conclude in Sect. 5.

2. Related Work

Our work is related to previous work on classification of users, blogs, and news articles, and social network analysis.

Users' sides are classified based on conventional text classification method, sentiment analysis, and co-citation analysis using explicit references to other users for informal online political discourse [9]. Three semi-supervised learning methods are used for classification of political leaning of news articles and users from user votes with a few labeled articles and users [10].

On classification of political blogs, the link patterns are used to determine whether the blog is liberal or conservative [11]. Subjectivity analysis in blogs to improve political

Manuscript received July 4, 2013.
Manuscript revised October 29, 2013.
†The authors are with the Department of Computer Science and Engineering, CAIIT, Chonbuk National University, Korea.
a) E-mail: selfsolee@chonbuk.ac.kr (Corresponding author)
DOI: 10.1587/transinf.E97.D.685

leaning classification has been exploited [12]. Bootstrap algorithm is applied for classification of links and blogs and blog rankings from a set of seed blogs [13]. Graphical information on news articles are provided regarding user's political orientation by exploiting subjective analysis [14].

On classification of news articles, the commenters' sentiment pattern-based method is used for political view identification [6]. The commenters posted positive comments on articles that they agreed on, and posted negative comments on articles that they disagreed on. The method uses the commenters selected for the study and discovered predictive commenters, showing a high degree of regularity in their sentiment patterns. The sentiment expressed by the predictive commenters strongly indicates the political orientation of the article. The pattern is used to determine the orientations of the news articles, which depends on the patterns of commenters' analyzed by humans [15].

Research on social network services such as Twitter and Facebook is currently being processed very actively. There is a research on studying the topological and geographical properties of Twitter's social network to understand the user intentions and community structure [16]; there are also researches on extracting influential figures from user groups on Twitter [17], [18], and a research on how users would mainly be influenced by urgent news, friends' posts, and interesting posts. The models are suggested to classify the starting point of a discussion on Twitter to extract issues [19].

3. Predicting Political Orientation of Users Based on Social Network Analysis

We propose a novel framework for predicting political orientation of social users and contentious news articles based on social behavior in Twitter. Political orientation of a user and a news article is categorized into three classes: *liberal*, *conservative* or *moderate*. First, social user network is constructed based on user interaction for issue news and gradually expanded when issue news are added. The highly active users are extracted for prediction with high confidence from the social network. Next, political orientation of a user is predicted by analyzing retweeting behaviors on social networks and opinion analysis of tweet documents. Finally, political orientation of a news article is predicted based on the consistent user behavior pattern and opinion analysis of tweet documents.

3.1 Construction of Social User Network

In order to build an initial user network, politicians and political words are used as a seed information. The political orientation of politicians is explicitly classified depending on the party they belong to, based on orientations of congressmen [20]. On social network, the nodes are from general users who post political views on their tweets by retrieving twitter test collection with political words as a query, the edges between users are connected when users do retweeting behaviors to the other user's tweets.

Table 1 Seed information for the initial social user network.

Politicians		Political word list (136 words)
Conservative Party	Liberal Party	congressman, political party, intimidation People's Revolutionary Party, suspicion, political party, election, president, ...
423	418	

Table 2 Expanded social user network.

	Date	# of users	# of tweets	# of retweets
Initial network	2012.9.1 - 9.15	337,737	5,191,643	3,618,415
Expanded network	2012.9.16 - 10.10	265,575	9,807,818	6,968,346
Final social network	2012.9.1 - 10.10	603,312	14,999,461	10,586,761

Table 1 shows the seed information of social user network built based on politicians' party information and political keywords. The social network is gradually expanded by adding issue news articles and the related users to the initial social network.

In order to expand the social network, issue news articles are selected. To analyze people's political behavior, we focus on hot issues on social networks. In order to select issue news on Twitter, all tweets are gathered which contain news URL from 16th September 2012 to 10th October 2012 (25 days). And the top 10 news articles with the most frequently mentioned URLs are selected as hot issues for each day. Then tweets and users who post the tweets for news articles are added to the social network.

Table 2 shows the number of twitter users and retweets. Social network are gradually expanded by adding issue news articles and users information on the initial network.

Through these expansions, social network is solidly built depending on temporal flow by expanding information of users and tweet documents on the network.

3.2 User Behaviors on the Social Network

In order to predict political views of users' for issue news articles, user behavior should be analyzed according to their behaviors such as tweeting with/without opinion and retweeting with/without opinion in Twitter.

When users have the same viewpoint with the news article, they tend to post a tweet with only news URL to spread out the news and arouse sympathy. Besides when users agree to other's viewpoint, they tend to do 'retweet' behavior to show sympathy.

Based on our observation, the following tweet relations are defined according to user behavior on social network as shown in Fig. 1.

- **Tweeting relation:** when a user post a tweet of news URL without any opinion, the user can be considered as agreeing to the political view of the news article (*user A*).
- **Opinionated tweeting relation:** when a user post a tweet of news URL with some opinion, the user's con-

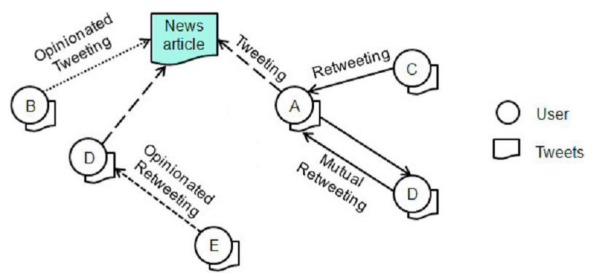

Fig. 1 Tweet relations according to user behavior on social network.

sent or objection to the political view of the news article depends on the opinion (*user B*).

- **Retweeting relation:** when user C does retweet user A's tweet, it can be seen their opinions are the same (*user C and A*).
- **Opinionated retweeting relation:** when user E does retweet user D's tweet with some opinion, the orientation of user E depends on the opinion (*user E and D*).
- **Mutual retweeting relation:** when user D and A retweet each other's tweets, it can be seen they have a very close relationship and hold the same political view (*user A and D*).

The retweeting and mutual retweeting relations are directly used for building user association network without opinion analysis. The edge of retweeting relation is weighted by the frequency of retweeting behaviors.

3.3 Extraction of Highly Active Users

In this paper, a highly active user is defined as one with a high degree of retweeting or retweeted relation to or by other users. On the other hand, a rarely active user is defined as one with a low degree of relations among social users. Highly active users may have consistent behavior patterns on social network. Their orientations can be used for rarely active users.

In order to select highly active users, HITS (Hyperlink-Induced Topic Search) algorithm [21] is modified for our purpose to estimate user activities on social network.

In our framework, user activity is measured based on user behavior such as tweeting and retweeting frequency with or without opinion to the news articles and to the others' tweets. In HITS graph, users are represented as nodes and retweet relations between users are represented as edges. Highly active users tend to be connected by lots of edges to others on social network.

The modified part for our HITS is the initial authority and hub score, and edge weight. The weight of edge, w_{ij} represents the frequency of retweet from user u_i to u_j, and w_{ji} is from user u_j to u_i. The authority and hub scores are calculated iteratively as following:

$$AuthScore^{(k+1)}(u) = \sum_{j\in(j,i)} w_{ji} \times HubScore^{k}(u_j) \qquad (1)$$

$$HubScore^{(k+1)}(u) = \sum_{j\in(i,j)} w_{ij} \times AuthScore^{k}(u_j) \qquad (2)$$

The initial authority score is set to the ratio of the retweeted tweets by other users' tweets, and the initial hub score is set to the ratio of the retweeting tweets to other users' tweets. While the general HITS has the same initial authority and hub scores, our AuthScore^0(u) and HubScore^0(u) are calculated as follows:

$$AuthScore^{0}(u) = \frac{\text{\# of tweets of user u } \textbf{\textit{retweeted}} \text{ by other users}}{\text{\# of tweets of user u}} \qquad (3)$$

$$HubScore^{0}(u) = \frac{\text{\# of tweets of user u } \textbf{\textit{retweeting}} \text{ by other users}}{\text{\# of tweets of user u}} \qquad (4)$$

Finally, the activity scores of users are calculated by combining authority and hub scores:

$$ActiveScore(u) = \sqrt{AuthScore(u) \times HubScore(u)} \qquad (5)$$

Highly active users are selected according to *ActivityScore*. In our model, a good hub represents a user that *retweets* to many other users, and a good authority represents a user that is *retweeted* by many different hubs.

Users with activity score higher than a threshold θ (set to 0.1e-10 by training in our experiment) is considered as highly active users and the rest of the users are considered as rarely active users. Out of 603,312 users, 130,419 are considered highly active users and 472,893 are considered as rarely active users.

According to the user activity scores, highly active users tend to have consistent viewpoints compared to rarely active users.

3.4 Predicting Political Orientation of Users

To predict political orientation of highly active users, the *bias ratio* to Conservative and Liberal side is measured for each user by relationships on the social user network in descending order of the activity scores. For the rarely active users, opinions in their tweets are analyzed based on support vector machine (SVM) learning [22].

3.4.1 On Highly Active Users

When a user's orientation is obvious, it can be used to analyze other user's orientation. This results in extending users network with the same viewpoints. The users are distinguished on how heavily related they are with either liberal or conservative group. If one is heavily involved with users in the conservative group and not as much involved with users in the liberal group, this user can be considered as a user with a conservative orientation. Likewise, a user can be liberal oriented when the user is deeply involved with the liberal and less than the conservative. However, there can be a moderate category, if the user's involvement with either party is similar. Therefore, to distinguish those whose orientations are distinct, the ***bias ratio of a user*** is measured based on retweeting and retweeted relationships on social

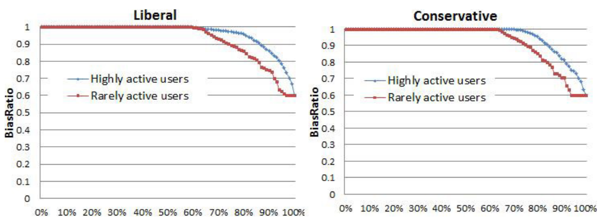

Fig. 2 Distribution of *BiasRatio* on users.

network as follows:

$$BiasRatio(u, L) = \frac{Liberal(u)}{Liberal(u) + Conservative(u)} \quad (6)$$

$$BiasRatio(u, C) = \frac{Conservative(u)}{Liberal(u) + Conservative(u)} \quad (7)$$

where

$$Liberal(u) = RTing(u, L) + RTed(u, L) \quad (8)$$

$$Conservative(u) = RTing(u, C) + RTed(u, C) \quad (9)$$

where *RTting(u, X)* represents the frequency of retweets from a user u to any users who belong to X group, and *RTed(u, X)* represents the frequency of retweets from any users who belong to X group to a user u. C and L describe Conservative and Liberal, respectively. When the two fractions are added, it should be equal to 1.

Figure 2 shows the distribution of BiasRatio on users in the social network. The 80% of highly active users shows strict BiasRatio with value 1. The 60% of rarely active users shows BiasRatio with value 1. The more accurate analysis can be carried out by using highly active users as a basis.

The prediction algorithm for user's orientation is based on *MutualRetweet* and *BiasRatio* as shown in Fig. 3. The *MutualRetweet* represents the retweet frequency between users in the same group. The MutualRetweet is calcuated by the minimum frequency to reflect the common activity between two users. The parameters are set to α=3, β=0.7 and γ=1.0 empirically. If a user has some degree of mutual retweet, α, on Liberal group and there is no relationship by mutual retweet on Conservative group, then the user can be classified into Liberal side. Or, the *BiasRatio(u, L)* of a user is greater than the threshold β, the user can be categorized into Liberal side.

If a user has distinctive orientation, which means *BiasRatio* is higher than the threshold γ, the user will be added to the orientation group in social network. It is possible to expand the relationship of users who have not a previous relationship. It makes incremental prediction.

3.4.2 On Rarely Active Users

On the social network, it is difficult to predict orientation of rarely active users since there is no enough relationships among users. To categorize these types of users, sentiments of tweet documents of users are analyzed by using support vector machines (SVMs) classification which showed to give high performance on text classification problems.

From the initial user network, tweet documents depending on the orientation and sentiments of users are divided into four categories: *Liberal-Positive, Liberal-Negative, Conservative-Positive and Conservative-Negative* to learn SVMs. When a tweet only mentions a particular politician with positive or negative sentiment, the tweet document is selected for learning. The sentiment lexicon consists of 589 positive words and 782 negative words for Korean sentiment analysis. Although the lexicon contains more negative words than positive ones, the training collection has no influence by the size of sentiment words as shown in Table 3.

Table 3 shows Twitter training tweet collection automatically constructed for SVM classification on rarely active users. For example, the tweet document can be seen as on the liberal side when a tweet contains liberal politicians and a positive sentiment word such as "support". In contrast, the tweet document can be seen on the liberal side when a tweet contains conservative politicians and a negative sentiment word such as "can't believe".

The SVM classification with tweet contents is based on the traditional approach. If most of a user's tweet documents are classified into Liberal-Positive or Conservative-Negative categories, then political orientation of the user is predicted to be liberal side. The conservative side is vice versa.

```
Procedure  PredictUserOrientation(ActiveUserSet)
  ActiveUserSet: set of users extracted by HITS algorithm;
  C, L: empty set for Conservative and Liberal Side

for each user u in ActiveUserSet do
  // the degree of mutual retweet between user u and Liberal
  MutualRetweet(u, L) = ∑l∈L min(retweeting(u,l), retweeted(u,l))
  MutualRetweet(u, C) = ∑c∈C min(retweeting(u,c), retweeted(u,c))

if((MutualRetweet(u, L) ≥ α) and MutualRetweet(u, C) = 0)
    or (BiasRatio(u, L) ≥ β))
  Classify user u to "Liberal" side
  if( BiasRatio(u, L) ≥ γ ) then Add user u to L
else if((MutualRetweet(u,C) ≥ α) and MutualRetweet(u,L)=0)
      or (BiasRatio(u, C) ≥ β))
  Classify user u to "Conservative" side
  if( BiasRatio(u, C) ≥ γ ) then Add user u to C
else
  Classify user u to "Moderate" side
end
```

Fig. 3 Incremental prediction algorithm for orientation of users.

Table 3 Twitter training collection automatically constructed for SVM.

Orientation	Liberal side		Conservative side	
Categories	Liberal-Positive	Conservative-Negative	Conservative-Positive	Liberal-Negative
# of tweets	11,114	2,781	5,463	6,448

Table 4 Twitter test collection for analyzing orientations of users.

# of issue news	# of users	# of tweets	# of retweets	Avg. # of tweets per news article
280	22,875	136,223	132,304	486.5

3.5 Empirical Study on Predicting User's Orientation

3.5.1 Experimental Set-Up

We have evaluated the effectiveness of the proposed method in social user network on Twitter. All tweets, users and contentious news articles are extracted in Twitter from September 1 to October 10, 2012 by using Twitter API (all news articles and tweets are written in Korean).

One thousand users are randomly selected in the social network and judged by two human assessors by analyzing the user's tweets and retweet behavior. The answer set consist of 711 Liberal oriented, 274 Conservative oriented and 15 Moderate oriented users. Table 4 shows Twitter test collection for predicting user's orientation.

We have compared the proposed method with SVM classification based on tweet contents, SVM classification based on analysis of positive and the negative tweets:

- *SVM based on tweet contents*: Users' tweets are divided into Liberal or Conservative group according to the user's side from the initial network. Each tweet document is represented as a vector with tf-idf weighting. By SVM learning each category, a user is classified depending on how many tweets are categorized into one side.
- *SVM based on sentiment analysis for tweets*: all tweets are divided into four groups by sentiment analysis: Liberal-positive, Liberal-negative, Conservative-Positive and Conservative-negative group. Tweets are classified by SVM into Liberal-positive or Conservative-negative are considered as Liberal side. A user is classified according to how many tweets are categorized into one side.
- *Proposed method by social user network analysis*: According to the degree of user activity, highly active users are predicted by the relationships in the network, and rarely active users are classified by SVM learning with sentiment analysis of tweets.

3.5.2 Experimental Results and Analysis

The performance of each method is measured in accuracy. Experimental results are shown in Table 5. The proposed method achieves 10.1% and 26.2% improvements over SVM classifications by tweet documents with or without sentiment analysis. The result indicates that social users have consistent behavior patterns in social network for the contentious issues. When the parameter β is set to 0.75, the proposed method could correctly classify four users as Moderate side. However, the overall performance is the best

Table 5 Experimental results for political orientations of users.

Method	Lib.	Cons.	Mod.	Total	chg %
SVM by tweets	0.764	0.481	0	0.675	-
	543/711	132/274	0/15	675/1000	
SVM by sentiment analysis	0.790	0.661	0	0.743	+10.1%
	562/711	181/ 274	0/15	743/1000	
Proposed method	0.879	0.828	0	0.852	**+26.2%**
	625/711	227/274	0/15	852/1000	

Table 6 Result analysis according to the degree of user activity.

Highly Active Users				Rarely Active Users			
Lib.	Cons.	Mod.	Total	Lib.	Cons.	Mod.	Total
0.877	0.915	0	**0.874**	0.890	0.308	0	**0.727**
528/602	215/235	0/13	743/850	97/109	12/39	0/2	109/150

Table 7 Examples of predicting orientation for highly active users.

user	Case 1		Case 2			Orientation
	MutualRetweet		RTing+RTed		BiasRatio	
	Lib.	Cons.	Lib.	Cons.	(L or C)	
u1	0	41	7	182	0.963(C)	Cons.
u2	15	1	155	3	0.981(L)	Libs.
u3	24	18	35	87	0.713(C)	Mod.

when the β is set to 0.7 in training, which is shown in Table 5.

The results are analyzed according to the degree of user activity as shown in Table 6.

We have examined highly active users to see the effectiveness of the proposed algorithm. Table 7 shows the examples correctly predicted by each case of our algorithm. The user u1 is classified as Conservative side by the case 1 that there is no MutualRetweet among users in Liberal side and there are more than three times (α=3) in MutualRetweet among users in Conservative side. For the user u2, the user can not be applied to the case 1 since there are not zero MutualRetweet to one side. The user u2 is classified as Liberal by the case2 that the BiasRatio to Liberal side (0.981) is above the threshold (β=0.75). The user u3 is classified as Moderate by the default that BiasRatio to each side is below the threshold (β=0.75).

The prediction based on user behavior analysis in social network shows 87.4% for highly active users, and the classification based on SVM learning shows 72.7% for rarely active users in accuracy.

3.5.3 User Behavior Analysis in Social Network

The BiasRatio values of 444,052 users are measured, who have relationships to other users among all 603,312 users in the social network. The number of users without any relationships to other users is 159,260. The proportion of users with the value of 1 in BiasRatio is 88.82%. It means that the orientations of most of users are obvious.

In the result, lots of users have mutual relationships with the same orientation group. By the proposed method, 17,116 users are predicted by MutualRetweet behaviors

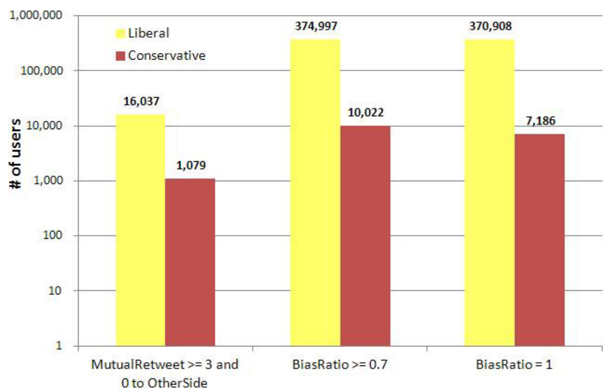

Fig. 4 Distribution of users for *MutualRetweet* and *BiasRatio*.

without interaction to the other side, and 374,997 users are predicted by BiasRatio with the value of 0.7 or above. Figure 4 shows the distribution of users for prediction algorithm. This analysis tells that the user behavior pattern is too distinct on social network. The desirable direction is social change is that news articles fairly deliver arguments in each stance on conflicting issues, and the level of user's bias ratio go down.

4. Predicting Political Orientation of News Articles

In general, news articles are frequently biased and fail to fairly deliver conflicting arguments of an issue. When users perceive an issue passively through a single article, the thought of the users can be biased without understanding other viewpoints that drive social conflict. To mitigate media bias, we propose a method to predict political orientation of news articles by analyzing orientation of users in social user network. Then users can easily understand the issue from various perspectives. In our research, the users' behavior patterns can be used in the social network, which shows contemporary society by exhibiting the polarities of news articles, which lead polarities of users' viewpoints.

4.1 Observation on Relationships between Tweet and News

When a user posts a tweet just including a news article URL without any opinion, it is highly likely that the user agrees with the viewpoint of the news article. In this case, it is safe to say that orientation of the news article and orientation of the user are the same. However, if there are some opinions in a tweet, their viewpoints of the news article and the tweet may vary according to the sentiments of the tweet and user's political orientation. Through sentiment analysis of a tweet, if the sentiment of a tweet is positive (agreeing with the viewpoint of a news article), orientation of the news article is the same with the user of the tweet; if the sentiment of a tweet is negative (disagreeing with the viewpoint of a news article), orientation of the news article is the opposite of the user.

Based on our observation on the relationships between

```
Procedure    PredictNewsOrientation()
  AgreeTweets: users of tweets with the same viewpoint of news
  DisagreeTweets: users of tweets with the opposite viewpoint of news
  N: total tweet documents regarding the news article

// To sum all BiasRatio of users whose tweets agreeing
// with the viewpoint of the news article
LiberalSide = ∑u∈AgreeTweets BiasRatio(u, L) / N
ConservativeSide = ∑u∈AgreeTweets BiasRatio(u, C) / N

// To sum all BiasRatio of users whose tweets disagreeing
// with the viewpoint of the news article
AntiLiberalSide = ∑u∈DisagreeTweets BiasRatio(u, L) / N
AntiConservativeSide = ∑u∈DisagreeTweets BiasRatio(u, C) / N

totLiberalSide = LiberalSide + AntiConservativeSide
totConservativeSide = ConservativeSide + AntiLiberalSide

// Predict political orientation of the news article d
 if ((LiberalSide - ConservativeSide ≥ α) or
      (totLiberalSide - totConservativeSide ≥ β))
   Classify d to "Liberal viewpoint"
else if((ConservativeSide - LiberalSide ≥ α) or
        (totConservativeSide - totLiberalSide ≥ β))
   Classify d to "Conservative Viewpoint"
else
   Classify d to "Moderate Viewpoint"
end
```

Fig. 5 Prediction algorithm of orientation of news articles.

a tweet and a news article, tweets are divided into two groups: agreeing or disagreeing with the viewpoint of a news article.

- ***Tweets agreeing*** with the viewpoint of a news article (AgreeTweets) are selected as follows:
  1) tweets only including the news URL
  2) tweets with positive sentiment words
- ***Tweets disagreeing*** with the viewpoint of a news article (DisagreeTweets) are selected as follows:
  1) tweets including words expressing the article with negative sentiment words
  2) when the sentiment on the important person in the article contradicts the sentiment in the tweet

Here, the important person in a article is identified based on the seed politicians information used for the initial social network. The orientations of users whose tweets belong to AgreeTweets and DisagreeTweets are considered to predict political orientation of a news article.

4.2 Predicting Orientation of News Articles Based on Social User Behaviors

The political orientation of a news article is predicted based on orientations of users whose tweets with the same viewpoints or the opposite viewpoints to the news article. The prediction algorithm is shown in Fig. 5.

If there are lots of agreeing tweets whose users' orientations are outweighed to one side, it can be seen that the news article may have the same orientation with that of

Table 8 Twitter test collection for predicting orientation of news articles.

	# of news	Political Orientation			# of users	# of tweets	
		Lib.	*Cons.*	*Mod.*		*Lib.*	*Cons.*
train	116	74	35	5	10,646	39,721	18,533
test	164	118	42	6	15,908	50,186	27,783

users. It is assumed that when a user's tweet has negative sentiment to the news article, then viewpoint of the news article is opposite to that of the user. Therefore, the score of anti-liberal side is summed up bias ratio of users in liberal side who disagree the viewpoint of the news article. The final score of *totConservativeSide* is by adding *ConservativeSide* with *AntiLiberalSide* score. The vice versa is the same. The parameters used in the algorithm are set to α=0.5 and β=0.4 empirically.

When users' behavior has consistency in social network, the proposed method can be useful to predict the orientation of news articles.

4.3 Empirical Study on Orientation of News Articles

4.3.1 Experimental Set-Up

The contentious news articles, users and their tweets are selected in the social network to evaluate the proposed method. Twitter test collection for predicting orientation of news articles consists of 280 news articles, 22,875 users and their tweets on 25 issues as shown in Table 8.

The orientation of each news article is judged by two human analysts by reading the news articles. Out of the 280 news articles, 192 are classified into Liberal side, 77 are classified into Conservative, and 11 are classified into Moderate side.

In order to prove the effectiveness of the proposed method, we have compared with SVM classification, sentiment pattern based classification and the proposed method without regarding the bias ratio:

- *SVM classification based on tweet documents and news articles*: Tweets of users are divided into Liberal or Conservative group according to the politician's side from the initial network. Each sentence of the news article and each tweet document is represented as a vector with conventional tf-idf term weighting. By SVM learning each side by using tweet vectors, each sentence of the news article is classified. Then the news article is classified depending on how many the sentences are categorized into one side.
- *Social annotation analysis approach*: The related work [6] is compared in social environment. The comments and the users who commented on the news article are all included in the experiment. The probability table for each user is built by training the orientations of the news articles and the users. The orientation of news articles is predicted based on political orientation from the sentiments expressed in the comments.
- *Proposed method without regarding user's bias ratio*: Prediction method based on the users in relation to the articles. The value of BiasRatio is set to binary: 0 or 1 according to the predicted side.
- *Proposed method*: Prediction method based on user's bias ratio for agreeing tweets and disagreeing tweets.

Table 9 Experimental results of predicting orientation of news articles.

Method	*Liberal*	*Conservative*	*Moderate*	Total
SVM based on tweets and news articles	0.786	0.571	0	0.696
	151/192	44/77	0/11	195/280
Social annotation analysis	0.875	0.753	0	0.807
	168/192	58/77	0/11	226/280
Proposed w/o bias ratio	0.953	0.766	0	0.864
	183/192	59/77	0/11	242/280
Proposed method	0.969	0.805	0	**0.886 +27.3%**
	186/192	62/77	0/11	**248/280**

Table 10 Examples of correctly predicted news articles.

issue	Orientation (by human)	# of users			
		AgreeTweets		DisagreeTweets	
		Lib.	*Cons.*	*Lib.*	*Cons.*
#2	*Lib.*	1336	75	85	8
#17	*Lib.*	274	12	14	152
	Cons.	33	353	133	21

Table 11 Examples of incorrectly predicted news articles.

issue	Orientation (by human)	# of users			
		AgreeTweets		DisagreeTweets	
		Lib.	*Cons.*	*Lib.*	*Cons.*
#10	*Lib.*	170	103	24	102
#21	*Lib.*	185	102	19	83
#25	*Cons.*	384	299	0	18

4.3.2 Experimental Results and Analysis

Experimental results of predicting orientation of news articles are shown in Table 9 in accuracy.

The proposed method shows 88.6% in accuracy. It achieves 27.3% and 9.8% performance improvements compared to the SVM classification based on tweet documents, and social annotation analysis approach, respectively. This result indicates that our approach based on user behavior in social network is effective.

We have analyzed the results. The correctly predicted examples are shown in Table 10. All the comparison methods correctly predicted the orientation of issue #2. For issue #17 news article, only our approach could correctly predict the orientation of news articles by considering users' orientations in each side of AgreeingTweets and DisagreeTweets.

The incorrectly predicted news examples are shown in Table 11. For the issue #10 on "Extend voting time", issue #21 on "Half tuition debate", and international issue #25, users belonging to Conservative and Liberal side have the same opinion. From error analysis, when users have the same opinion on issues, the prediction based on users' orientation lead wrong results.

From error analysis, when most users have the same opinion on international issues, their bias ratio does not work to predict news issues. This social user behavior will be desirable in debates on issues when the viewpoint of news article fairly deliver on each stance and social users suggest some directions on issue for reasonable society.

5. Conclusion

Predicting orientation of news articles based on social user behavior is effective in social network. Our contribution is to automatically estimate bias ratio of users by gradually constructing social network according to user behaviors, and to automatically predict the orientation of news articles based on users' orientations.

The prediction based on user behavior analysis in social network showed 85.2% performance in accuracy, which achieved 26.2% improvement over the SVM methods. The proportion of users with the value of 1 in BiasRatio is 88.82%. This result shows that social users have consistent patterns in bias ratio and mutual retweet behavior.

Based on orientations of users, the proposed method for political orientation of news articles showed 88.6% performance in accuracy, which achieved 27.3% and 9.8% improvements over SVM classification, and social annotation analysis approach, respectively. This result indicates that proposed method is useful for predicting the orientation of news articles when users' behavior has consistency in social network. Providing predicted news articles help users easily understand the issue from various perspectives.

For future work, we will study a prediction method considering the characteristics of issue news such as national issues or global issues to which all the people have the same viewpoints.

Acknowledgements

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A2044811).

References

[1] A.-M. Popescu, M. Pennacchiotti, and D. Parajpe, "Extracting events and event descriptions from Twitter," Proc. 20th International Conference on World Wide Web (WWW), 2011.
[2] Wikipedia, "Social Data Revolution," http://en.wikipedia.org/wiki/Social_data_revolution
[3] Korea Times, SNS emerges as influential political leverage, http://www.koreatimes.co.kr/www/news/nation/2011/12/113_101962.html, 2011.
[4] Pew Internet Center, Politics on Social Networking Sites, http://pewinternet.org/~/media/Files/Reports/2012/PIP_PoliticalLifeonSocialNetworkingSites.pdf
[5] S. Park, M.K. Lee, and J. Song, "Contrasting opposing views of news articles on contentious issues," Proc. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT), 2011.
[6] S. Park, M. Ko, J. Kim, Y. Liu, and J. Song, "The politics of comments: Predicting political orientation of news stories with commenters' sentiment patterns," Proc. ACM 2011 Conference on Computer Supported Cooperative Work (CSCW), 2011.
[7] Z. Yang, J. Guo, K. Cai, J. Tang, J. Li, L. Zhang, and Z. Su, "Understanding retweeting bahaviors in social networks," Proc. 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010.
[8] D. Boyd, S. Golder, and G. Lotan, "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter," Proc. 43rd Hawaii International Conference on System Sciences (HICSS), 2010.
[9] R. Malouf and T. Mullen, "Taking sides: user classification for informal online political discourse," Internet Research, 2008.
[10] D.X. Zhou, P. Resnick, and Q. Mei, "Classifying the political leaning of news articles and users from user votes," Proc. 5th International AAAI Conference on Weblogs and Social Media, 2011
[11] L.A. Adamic and N. Glance, "The political blogosphere and the 2004 U.S. election: Divided they blog," Proc. 3rd International Workshop on Link Discovery (LinkKDD), 2005.
[12] M. Jiang and S. Argamon, "Exploiting subjectivity analysis in blogs to improve political leaning classification," Proc. 31st annual international ACM SIGIR Conference on Research and Development in Information Retrieval, 2008.
[13] F. Lin and W. Cohen, "The MultiRank bootstrap algorithm: Semi-supervised political blog classification and ranking using semi-supervised link classification," Proc. International Conference on Weblogs and Social Media (ICWSM), 2008.
[14] M. Gamon, S. Basu, and D. Belenko, "BLEWS: Using blogs to provide context for news articles," Proc. International Conference on Weblogs and Social Media (ICWSM), 2008.
[15] S. Park, S. Kang, S. Chung, and J. Song, "NewsCube: Delivering multiple aspects of news to mitigate media bias," Proc. SIGCHI Conference on Human Factors in Computing Systems, 2009.
[16] A. Java, "Why we twitter: Understanding microblogging usage and communities," Proc. 9th WebKDD and 1st SNA-KDD Workshop on Web Mining and Social Network Analysis, 2007.
[17] R. Tinati, L. Carr, and W. Hall, "Identifying communicator roles in twitter," Proc. 21st international conference companion on World Wide Web, 2012.
[18] J. Weng and Q. He, "TwitterRank: finding topic-sensitive influential twitterers," Proc. 3rd ACM International Conference on Web Search and Data Mining, 2010.
[19] A.M. Popescu and M. Pennacchiotti, "Detecting controversial events from twitter," Proc. 19th ACM International Conference on Information and Knowledge Management, 2010.
[20] W. Kang, "Ideology and policy attitudes of 19th legislators," Legislative Research, vol.36, pp.5–38, 2012.
[21] J.M. Kleinberg, "Authoritative sources in a hyperlinked environment," J. ACM (JACM), vol.46, Issue 5, pp.604–632, 1999.
[22] B. Scholkopf, C. Burges, and A. Smola, "Advances in kernel methods: Support vector learning," MIT Press, Cambridge, MA, 1999.

Jun-Gil Kim is a researcher at Center for Advanced Image and Information Technology (CAIIT). He received the MS degree in the Department of Computer Science and Engineering, CAIIT, Chonbuk National University in 2013. His research interests include information retrieval and social information mining.

Kyung-Soon Lee is an associate professor in the Department of Computer Science and Engineering, CAIIT, Chonbuk National University since 2004. She received the Ph.D. degree in computer science from Korea Advanced Institute of Science and Technology (KAIST) in 2001. Her research interests include information retrieval, natural language understanding and social information mining in computer science.