Unsupervised Approach for Identifying Users' Political Orientations

Youssef Meguebli[1], Mouna Kacimi[2], Bich-Liên Doan[1], and Fabrice Popineau[1]

[1] SUPELEC Systems Sciences (E3S), Gif sur Yvette, France
{youssef.meguebli,bich-lien.doan,fabrice.popineau}@supelec.fr
[2] Free University of Bozen-Bolzano, Italy
mouna.kacimi@unibz.it

Abstract. Opinions, in news media platforms, provide a world wide access to what people think about daily life topics. Thus, exploiting such a source of information to identify the trends can be very useful in many scenarios, such as political parties who are interested in monitoring their impact. In this paper, we present an unsupervised technique to classify users based on their political orientations. Our approach is based on two main concepts: (1) the selection of the aspects and the sentiments users have expressed in their opinions, and (2) the creation of knowledge base from Wikipedia to automatically classify users according to their political orientations. We have tested our approach on two datasets crawled from CNN and Aljazeera. The results show that our approach achieves high quality results.

Keywords: Political leaning, Political Opinion Mining, Sentiment analysis.

1 Introduction

Political views are freely and explicitly expressed through opinions in news media platforms. These opinions represent an interesting sample about political trends and orientations of users. Extracting such type of knowledge would allow news portal publishers to have an idea about the orientation of their commenters, the main issues related to each orientation, and the possible political persuasions and ideological viewpoints for all topics. The opinions expressed by users are not restrained by journalism values such as fairness or balance, and do not go through a formal editorial process. Moreover, the number of opinions about a given topic might continuously increase. The unstructured and the dynamic nature of opinions, provided in news media platforms, call for effective and efficient techniques for identifying political trends.

Several approaches have been proposed to classify political positions from texts. One line of work focused on using SVM with optimization of text feature selection [7][14] [18] [6], as well as complementing with sentiment analysis [4] [13] [10] [2]. Another line of work used word frequencies, Bayesian statistical models, and topic models [8] [11] [9] [12] [16]. Most of these approaches use

M. de Rijke et al. (Eds.): ECIR 2014, LNCS 8416, pp. 507–512, 2014.

supervised techniques which can be expensive as they require training. Moreover, they mainly use semi-structured data to classify users. Examples include data extracted from twitter and microblogs which is characterized by short fragments (tweets, short messages), where each fragment covers a known and a unique aspect. More specifically, approaches based on twitter samples use hashtags of controversial topics such as *'USElection'* or *'Arabspring'* and a set of stakeholders such as actors or politicians, to classify the stakeholders opinions into pro or con categories for the respective topics [1]. Comments published in microblogs are frequently short and do not contain more than one aspect whereas, in news media platforms, users publish long opinions covering more than one aspect.

In this paper, we propose an unsupervised technique for defining the political orientation of users based on their opinions in news media platforms. To the best of our knowledge, we are the first to propose an unsupervised approach on such unstructured and dynamic data. Our contribution is twofold (1) we generate user profile based on the aspects he has discussed in his opinions and their sentiments and (2) we construct a knowledge base of political orientations, using Wikipedia, to automatically classify users based on their profiles. We have conducted extensive experiments with US and Egypt user groups crawled from CNN and Aljazeera. The experiments showed that our approach provides high quality results to classify US users into Republican/Democrat leanings and Egypt users into secular/Islamist leanings.

2 Generating User Profile

To define the political profile of a given user U, we collect the opinions he has expressed, in a given media platform, during a period of time T. Then, we analyze the opinions and extract from them all the aspects that user U has discussed. For each aspect, we define the sentiment expressed by the user U. For example, a user can discuss the aspect of *abortion rights* and be *negative* about it. As a result, the user U is described by a set of aspects $\{a_1, ...a_n\}$ and their related sentiments $\{s_1, ...s_n\}$. To this end, we proceed in three main steps.

Step1. Extraction of Opinionated Sentences. We first identify the sentences[1] expressed in all the opinions of user U. For each sentence, we extract all its contained terms and we assign to each term a sentiment that can be positive, negative, or neutral using the lexicon provided by Ding et.al., [3]. We count the number of positive and negative terms in each sentence. If the number of positive terms is higher than the number of negative terms, we classify the sentence as positive, otherwise it is classified as negative.

Step2. Generation of Candidate Aspects. We take all the opinionated sentences extracted from the previous step, and we rank their contained terms using $tf * idf$ scoring function. In our work, tf represents the term frequency in the set of opinionated sentences of user U, and idf represents the inverted

[1] Using OpenNLP `http://opennlp.sourceforge.net/`

document frequency in the set of opinionated sentences of all users. The idea is to select highly scored unigrams as a base for generating candidate aspects. From these unigrams, we generate bi-grams, then we take the bi-grams as input and we build a set of n-grams by concatenating bi-grams that share an overlapping word. At each step we take the topk n-grams based on the score of their composed unigrams[2]. We check the redundancy of the generated candidates, using Jaccard similarity [15]. If two n-grams have a similarity higher than a defined threshold, we would discard one of them. In our work, we have set the maximum length of the n-grams to 5 since there were no meaningful n-grams of a higher length.

Step3. Selection of Promising Aspects. Generating n-grams that have high $tf * idf$ scores is not enough for identifying the aspects discussed in users' opinions. It is important for the words in the generated n-grams to be strongly associated within a sentence in the original text to avoid covering incorrect information. To capture this association, we use *pointwise mutual information* [17] (PMI) of words in n-grams.Formally, suppose $m_i = w_1...w_n$ is a generated n-grams. We define the $Score_n$ as follows:

$$S_{PMI}(w_1...w_n) = \frac{1}{n}\sum_{i=1}^{n} pmi_{local}(w_i) \tag{1}$$

where $pmi_{local}(w_i)$ is a local pointwise mutual information function defined as:

$$pmi_{local}(w_i) = \frac{1}{2C}\sum_{j=i=C}^{i+C} pmi'(w_i, w_j), i \neq j \tag{2}$$

where C is a contextual window size. The $pmi_{local}(w_i)$ measures the average strength of association of a word w_i with all its C neighboring words (on the left and on the right). For example, in *gun control law* phrase, assuming $C = 1$, for *gun* we would obtain the average PMI score of *gun* with *control* and for *control* we would obtain the average PMI of *control* with *gun* and *control* with *law*. When this is done for each $w_i \in m$, this would give a good estimate of how strongly associated the words are in m. We used a modified PMI scoring [5] referred to as *pmi* where the *pmi* between two words, w_i and w_j, is defined as:

$$pmi'(w_i, w_j) = \log_2 \frac{p(w_i, w_j) \cdot c(w_i, w_j)}{p(w_i) \cdot p(w_j)} \tag{3}$$

where $c(w_i, w_j)$ is the frequency of two words co-occurring in a sentence from the original text within the context window of C (in any direction) and $p(w_i, w_j)$ is the corresponding joint probability. The co-occurrence frequency, $c(w_i, w_j)$, which is not part of the original PMI formula, is integrated into our PMI scoring to reward frequently occurring words from the original text. By adding $c(w_i, w_j)$ into the PMI scoring, we ensure that low frequency words do not dominate and moderately associated words with high co-occurrences have relatively high scores.

[2] In this work we have set k=500.

3 Defining Users' Political Orientations

An unsupervised technique of identifying the political orientation of users based on their profiles calls for the use of a knowledge base. To this end, we have created a knowledge base of political orientations from Wikipedia. For a given political orientation, we start from a Wikipedia seed page. We extract from text part of the page all outgoing links that point to other Wikipedia articles. Then, we select the anchor text of these links as aspects related to the political orientation. Each aspect occurs in a sentence that have a sentiment orientation. Thus, in a similar way to user profile, we identify each political orientation to be described by a set of aspects $\{a_1, ...a_m\}$ and their related sentiments $\{s_1, ...s_m\}$. Table 1 shows an example of Liberal and Conservative orientation and their relevant aspects extracted from Wikipedia. To cover more aspects and enrich the knowledge base, we also include the Wikipedia pages pointed by the seed page of the political orientation. For example, a seed page reports that liberals are in favor of universal health care. We take the Wikipedia page of universal health care and add its aspects to the favorite list of Liberals.

Table 1. The structure of the orientation knowledge base

Orientation	**Some Aspects Extracted from Wikipedia**	
Liberals	**favor**	universal health care, strict gun control, diplomacy, stem cell research, same-sex marriage, abortion rights
	against	increased military spending The Ten Commandments display in public buildings
Conservatives	**favor**	small government, low taxes, limited regulation free enterprise, school prayer, capital punishment
	against	same-sex marriage, abortion rights, multiculturalism

To identify the political orientation of user U, we compute the similarity between its profile and the description of all the political orientations that exist in the knowledge base. The most similar description is assigned to the user as its political orientation.

4 Experimental Results

We have crawled 2 datasets from CNN and Aljazeera English news portals. From CNN, we have extracted 11,322 users, and their 684,058 opinions about 15,365 news articles. From Aljazeera, we have extracted 539, and their users 24,826 about 2,773 news articles. For each user, we have extracted all his opinions that concern politics from *October 2009* to *September 2013*. We have run our experiments on 500 users: 290 from US (CNN) and 210 from Egypt (Aljazeera). We have shown the list of opinions of each user and asked human assessors, who were students not involved in this project, to analyze the users opinions and classify them into the following categories: *Democrat/Republican* for US users and *Secular/Islamist* for Egypt users. The result of the human assessment is the ground truth for our evaluation.

We applied our approach on the same 500 users selected before. The outcome of the classification was then compared to our golden standard. To measure the effectiveness of our approach, we have computed the accuracy which represents the fraction of users that were correctly classified. We have compared different variations of our approach. The first one uses the *top*100 unigrams, based on $tf * idf$ scoring function, to classify the user. The second approach uses n-grams of length between 1 and 5. The *top*100 n-grams, based on $tf * idf$ scoring function, are selected to classify the user. In the third approach, the *top*100 n-grams, based on PMI, are selected to classify the user.

Table 2. Accuracy of User classification

	US		**Egypt**	
	Democrats	**Republicans**	**Islamists**	**Seculars**
Unigrams (tf*idf)	50%	16,66%	72,72%	25%
N-grams (tf*idf)	67,60%	50%	70,70%	51,56%
N-grams (PMI)	**95,07%**	**79,41%**	**85,85%**	**84,37%**

The results are shown in Table 2. We can see the impact of the different steps of our approach on the accuracy of our technique. Using only unigrams generates incomplete information about the aspects discussed by users and thus provides very inaccurate results. We can see that using n-grams improves the results in most cases, however they still have a low accuracy. Using PMI to select the aspects of opinions is the best providing an accuracy that goes up to $95,07\%$.

5 Conclusion and Future Work

We have proposed a new technique for defining the political orientation of users based on their opinions around news articles. The proposed approach is promising as it provides means for dealing with unstructured source of information. Moreover, it is completely unsupervised which makes it flexible to be applied on any kind of dynamic knowledge such as opinions. As future work, we plan to extend the knowledge base to other types of orientations in other domains and propose a general approach for extracting the main aspects of daily life topics and their main trends.

References

1. Awadallah, R., Ramanath, M., Weikum, G.: Harmony and dissonance: organizing the people's voices on political controversies. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, WSDM 2012, pp. 523–532. ACM, New York (2012)

2. Conover, M.D., Gonçalves, B., Ratkiewicz, J., Flammini, A., Menczer, F.: Predicting the political alignment of twitter users. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (passat), and 2011 IEEE Third International Conference on Social Computing (socialcom), pp. 192–199. IEEE (2011)
3. Ding, X., Liu, B., Yu, P.S.: A holistic lexicon-based approach to opinion mining. In: WSDM, pp. 231–240 (2008)
4. Durant, K.T., Smith, M.D.: Mining sentiment classification from political web logs. In: Proceedings of Workshop on Web Mining and Web Usage Analysis of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (WebKDD 2006), Philadelphia, PA (2006)
5. Ganesan, K., Zhai, C., Viegas, E.: Micropinion generation: an unsupervised approach to generating ultra-concise summaries of opinions. In: Proceedings of the 21st International Conference on World Wide Web, WWW 2012, New York, NY, USA, pp. 869–878 (2012)
6. Hirst, G., Riabinin, Y., Graham, J.: Party status as a confound in the automatic classification of political speech by ideology. In: Proceedings of JADT 2010 (2010)
7. Jiang, M., Argamon, S.: Political leaning categorization by exploring subjectivities in political blogs. In: DMIN, Citeseer, pp. 647–653 (2008)
8. Laver, M., Benoit, K., College, T.: Extracting policy positions from political texts using words as data. American Political Science Review, 311–331 (2003)
9. Lin, F., Cohen, W.W.: The multirank bootstrap algorithm: Self-supervised political blog classification and ranking using semi-supervised link classification. In: ICWSM (2008)
10. Malouf, R., Mullen, T.: Graph-based user classification for informal online political discourse. In: Proceedings of the 1st Workshop on Information Credibility on the Web (2007)
11. Martin, L.W., Vanberg, G.: A robust transformation procedure for interpreting political text. In: SPM-PMSAPSA, vol. 16, pp. 93–100 (2008)
12. Monroe, B.L., Colaresi, M.P., Quinn, K.M.: Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. In: SPM-PMSAPSA, vol. 16, pp. 372–403 (2008)
13. Mullen, T., Malouf, R.: A preliminary investigation into sentiment analysis of informal political discourse. In: AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, pp. 159–162 (2006)
14. Oh, A.H., Lee, H.-J., Kim, Y.-M.: User evaluation of a system for classifying and displaying political viewpoints of weblogs. In: ICWSM (2009)
15. Real, R., Vargas, J.M.: The probabilistic basis of jaccard's index of similarity, vol. 45, pp. 380–385. Oxford University Press (1996)
16. Slapin, J.B., Proksch, S.-O.: A scaling model for estimating time-series party positions from texts, vol. 52, pp. 705–722. Wiley Online Library (2008)
17. Terra, E., Clarke, C.L.A.: Frequency estimates for statistical word similarity measures. In: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, NAACL 2003, vol. 1, pp. 165–172. Association for Computational Linguistics, Stroudsburg (2003)
18. Yu, B., Kaufmann, S., Diermeier, D.: Classifying party affiliation from political speech, vol. 5, pp. 33–48. Taylor & Francis (2008)

User Perception of Information Credibility of News on Twitter

Shafiza Mohd Shariff, Xiuzhen Zhang, and Mark Sanderson

School of Computer Science and IT, RMIT University, Australia
{shafiza.mohdshariff,xiuzhen.zhang,mark.sanderson}@rmit.edu.au

Abstract. In this paper, we examine user perception of credibility for news-related tweets. We conduct a user study on a crowd-sourcing platform to judge the credibility of such tweets. By analysing user judgments and comments, we find that eight features, including some that can not be automatically identified from tweets, are perceived by users as important for judging information credibility. Moreover, distinct features like link in tweet, display name and user belief consistently lead users to judge tweets as credible. We also find that users can not consistently judge or even misjudge the credibility for some tweets on politics news.

1 Introduction

As of May 2013, an average of 58 million tweets are posted per day on Twitter.[1] Currently Twitter not only acts as a social medium, it is also becoming a news media source. Twitter citizens not only share news headlines from newswires, but also report real time events before they reach the press [5]. News on Twitter comes from a wide variety of sources: some from well known news organisations and government departments, while most from members of the public. Consequently twitterers often need to judge the credibility of tweets. Morris et al. (2012) discovered that twitterers have poor judgement on the truth of information on Twitter. Features such as the number of retweets, information on users who post tweets and their relationship network (number of followers and followees) help little in determining the level of information credibility on Twitter.

Spammers exploit the anonymity feature of Twitter to propagate their messages, retweeting them to increase their popularity rating [10]. In a Twitter dataset analysed by Gupta and Kumaraguru (2012), nearly half of the tweets about an event were found to be spam. In the work by Castillo et al. (2011), it was discovered that the credibility of information on Twitter is determined mainly by four types of features: message-based, content-based, user-based, and propagation-based. In most existing work, the features may need to be compiled by crawling the Twitter space and extracting the link relationship between twitterers. The purpose of these features are for automatic prediction and may not necessarily be users' perception of important signals for credibility.

[1] http://www.statisticbrain.com/twitter-statistics/

M. de Rijke et al. (Eds.): ECIR 2014, LNCS 8416, pp. 513–518, 2014.

In this paper, we focus on studying the tweet-based features that the general public mostly use to determine the credibility level of newsworthy tweet messages. The research questions that we will cover in this current work are:

1. What features do users use to judge the credibility of tweets?
2. How do users use tweet features to make their credibility judgment?
3. Does the tweet topic have an effect on a user's credibility judgment?

Related Work: Information credibility on Twitter has attracted significant research recently [1–4, 6, 7], where most focuses on automated approaches predicting the credibility of topics [1] and events [2, 3] by engineering complex features based on data and meta-data for tweets as well as their social structures. Morris et al. [6] study user perception of information credibility on Twitter and show that users rely on tweet contents and other heuristics for credibility judgments. On the other hand, it is established that Twitter posts report real-time news overlapping with the reported news in newswire with the addition of minor and local news not reported by other sources [8]. Despite the real-time news post, Twitter users are more concerned with the credibility of tweets relating to breaking news, politics, and disaster events [6].

2 Methodology

We design a user study based on the CrowdFlower[2] crowd source platform to examine user perception of the credibility of tweets for news events. We select news event topics, and their relevant tweets for our study. We recruit crowd source evaluators to judge the credibility level of tweets, and leave comments on their judgments. Through their comments, we extract the features and apply predictive association rule analysis [9] to establish the associations between features and credibility levels.

In this research, credibility is defined as *"the quality of being believed or accepted as true, real, or honest"*. [3] The criteria to determine credible tweets [1] are that they must affirm a fact, be informative for the public, not be self opinionated, and not be a chat between friends. To ensure that relevant and truthful news is used in our dataset, we selected twenty news event-related topics (judged by the authors) based on major news recently reported on on-line newswire including BBC, Reuters, CNN, Guardian, and The New York Times. The news events occurred between 1 June 2013 and 15 October 2013. Table 1 describes the twenty topics we selected. Tweets were collected based on the search API of Twitter, using the news event topics shown in the left column of Table 1 as query terms. To ensure that we do not include redundant tweets, directly retweeted messages are excluded. In total, 400 credible tweets in English for twenty news events were presented to CrowdFlower evaluators to judge.

[2] https://crowdflower.com
[3] http://www.merriam-webster.com/dictionary/credibility

Table 1. Twenty news event topics

Topic	News event description
US government shutdown	US Government heads toward a shutdown
Iran-US relationship	Iranian President takes steps to thaw relations with the West
Sarin attack in Syria confirmed	United Nations confirms use of chemical weapons in Syria
Shipwrecked at Europe	Boat sinks in the Mediterranean, killing dozens
Egypt state of emergency	Egypt declares state of emergency
Train kills dozens in India	Train kills dozens of religious pilgrims in India
Navy Yard shooting	Gunman and 12 victims killed in Washington D.C. Navy Yard shooting
Earthquake in Pakistan	Magnitude 7.7 earthquake kills at least 327 in Pakistan
Terrorist attack mall	Somalian militants terrorize luxury mall
Military ousted president	President Morsi deposed by military after one year in office
NSA whistle blower	Edward Snowden: whistle-blower behind NSA surveillance revelations
UK new prince	The Duchess of Cambridge gives birth to a baby boy
Oil train derails	A train in Quebec derails and explodes
Colorado flood	Colorado flood 2013 tragedy
Australia's new prime minister	Australia's new Prime Minister Tony Abbott
Iraq suicide attacks	Suicide bomb attacks on Iraqi school, Shi'ite pilgrims, kill 29
Mexico storm disaster	Mexico storms death toll rises, crop lands damaged
Cyclone hits India	Many evacuated as Powerful Cyclone Hits India
Protest in Egypt	More than 50 people are killed as pro-Morsi protest
Riot in Moscow	Rioting erupts in Moscow after killing blamed on migrant

In the crowd source evaluation, the date, topic and topic description of each tweet are given to the evaluators to help them distinguish the credibility level of tweets. The credibility definition and criteria are also presented to the evaluators. To trap unreliable evaluators, gold questions are set up, which are credible news tweets mingled with not credible tweets containing opinions or social chats. For each of the twenty topics, two gold questions are randomly inserted into the credible tweets. Only evaluators that judge the gold questions correctly are considered reliable, and their judgments are accepted.

To further elicit the features the public uses to judge the tweet credibility, we also ask CrowdFlower evaluators to leave textual comments to explain their judgements. We manually examine the comments to ensure quality comments are used to analyse user perception. To this end, we remove nonsensical comments, such as those containing the word "none", numbers or words that are out of context for the topic.

Table 2. Distribution of credibility ratings for 400 tweets

Credibility level	**#comments**
Definitely credible	342 (85.5%)
Seems credible	2 (0.5%)
Not credible	35 (8.75%)
Can't decide	0 (0%)
No consensus rating	21 (5.25%)

Table 3. Features derived from user comments for credibility rating

Category	**Feature**	**#cmts**
Topic-based	Topic keyword - *e.g. Prince (UK new prince topic)*	315 (54%)
Message-based	Link in tweet - *URLs, URL shortener, image links*	95 (16.3%)
User-based	Display name - *Twitter ID e.g. BBCNews, Anonymous*	88 (15%)
User-based	User belief of the topic - *e.g. plausible, professional, it actually happened, facts, informative*	44 (7.5%)
Message-based	Credibility keyword - *e.g. Update, Breaking, Liveupdates*	26 (4.5%)
Message-based	Hashtag - *e.g. #Lampedusa, #Egypt*	8 (1.4%)
Message-based	Retweet - *Contains the letters 'RT' in the tweet messages*	6 (1%)
User-based	User mention - *e.g. @OMBPress, @cctvnewsafrica*	2 (0.3%)

3 Deriving and Analysing Features for Credibility

In our user study, evaluators were asked to judge the credibility level for each tweet as "Definitely credible", "Seems credible", "Not credible", or "Can't decide". At the conclusion of our user study, a total of 2,005 judgements by 98 evaluators for 400 tweets were collected, where five out of 400 tweets received six judgments and the rest received five judgments each. The consensus rule was used to assign credibility rating for tweets. If a tweet receives three out of five or four out of six votes for a credibility level, the message is assigned the corresponding credibility rating; otherwise no consensus credibility rating (recall that there are three credibility levels) can be reached for the tweet. Table 2 lists the distribution of credibility ratings for all tweets. Note that none of the tweets received the judgment of "Can't decide". Our results confirm that users generally trust the information disseminated on Twitter, which mirrors the findings in [1].

3.1 Analysing User Comments for Credible Tweets

We analyse user comments for 342 and two tweets received "Definitely credible" and "Seems credible" ratings to derive features users use for their credibility judgments. The comments collected from the user study consist of 558 valid comments from 22 evaluators, which describe features they feel important for their judgment of the truth and falseness for tweets. Following the categorisation in [1] and [3] we manually summarise the comments into three categories of eight features, as shown in Table 3.

Table 4. Top association rules

Association Rules	Accuracy
Link in Tweet=available 74 => Credible 72	97.7%
Hashtag=yes 8 => Credible 8	97.6%
Retweets=yes 6 => Credible 6	97.2%
Twitter display name=yes, User belief=yes 3 => Credible 3	96.2%
Twitter display name=yes 88 => Credible 81	91.0%
User belief=yes, Topic keyword=yes 36 => Credible 27	77.4%
User belief=yes 44 => Credible 33	76.7%

Note that "User belief of the topic" refers to user's prior belief on the relevant topic and is external to Twitter, while in [1] all features are derived based on Twitter. Table 3 shows that users perceive these features in general with significantly different weights, where Topic keyword is commonly used and User mention is rarely used. In contrast the carefully engineered tens of features in [1] are used collectively by machine learning models for predicting topic credibility.

3.2 Analysing Misjudged and Difficult-to-Judge Tweets

We analyse the 35 tweets with the "Not credible" rating in Table 2. These tweets are misjudged by evaluators, as all tweets in our study have been manually verified as credible. The politics news topics 'Iran and US relationship' and 'US Government shutdown' have the largest number of misjudged tweets. We observe that these tweets are often questions, which may be why users have misperception of their credibility; indeed they are titles for news articles from reliable news agencies with short url links. Although Link in tweet is an important feature for users to judge credible tweets (See Table 4), the language features of tweets also play important roles for user perception of credibility.

We also analyse the 21 difficult-to-judge tweets where users could not reach consensus ratings. 95.6% of these difficult tweets are breaking news (42.8%) and politics news (42.8%). We observe that these tweets mostly lack links to external sources, which result in that users can not consistently judge their credibility. This Link in tweet and tweet credibility association is also shown in Table 4.

3.3 Feature and Credibility Association Analysis

To uncover relationships between features and tweet credibility, we apply association rule mining to the 379 tweets in Table 2 with consensus ratings of Definitely credible, Seems credible and Not credible based on the features in Table 3. We use the WEKA Predictive Apriori package [9][4] to mine for the best 100 association rules of the form "feature set => credibility" with an accuracy threshold of 70%. Table 4 lists the top association rules, where numbers of comments supporting the left and right hand sides are shown. According to the table,

[4] http://www.cs.waikato.ac.nz/ml/weka/

for all top rules the right hand side is always the Credible rating – users tend to believe in the information conveyed in tweets yet can not reach the Not credible rating consistently. Moreover, Link in tweet, Display name and User belief are important features often leading users to the Credible rating for tweets. From Tables 3 and 4 it can be seen that the Topic keyword feature, the most important feature commented by evaluators, does not form a strong association rule; only when combined with User belief it gives high accuracy for predicting credible tweets. The Link in tweet feature is used as an indicator for credibility.

4 Conclusions

We have studied the user perception of credibility for news tweets on Twitter via a user study on the CrowdFlower platform. Through analysing user credibility judgements and comments, eight features have been identified, where display name, link in tweet and user belief in the tweet topic are most important. By feature and credibility association analysis, we find strong associations between features and tweet credibility. We further find that politics and breaking news are more difficult for users to consistently reach credibility rating.

Acknowledgments. This research is partially supported by Universiti Kuala Lumpur (UniKL), Majlis Amanah Rakyat (MARA), and by the QNRF project "Answering Real-time Questions from Arabic Social Media" NPRP grant number 6-1377-1-257.

References

1. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on Twitter. In: Proc. WWW, pp. 675–684 (2011)
2. Gupta, A., Kumaraguru, P.: Credibility ranking of tweets during high impact events. In: Proc. PSOSM, pp. 2–9 (2012)
3. Gupta, M., Zhao, P., Han, J.: Evaluating event credibility on twitter. In: Proc. SDM, pp. 153–164 (2012)
4. Kang, B., O'Donovan, J., Höllerer, T.: Modeling topic specific credibility on twitter. In: Proc. IUI, pp. 179–188 (2012)
5. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proc. WWW, pp. 591–600 (2010)
6. Morris, M., et al.: Tweeting is believing?: understanding microblog credibility perceptions. In: Proc. ACM CSCW, pp. 441–450 (2012)
7. O'Donovan, J., et al.: Credibility in context: An analysis of feature distributions in twitter. In: PASSAT and SocialCom, pp. 293–301 (2012)
8. Petrovic, S., et al.: Can twitter replace newswire for breaking news? In: Proc. AAAI (2013)
9. Scheffer, T.: Finding association rules that trade support optimally against confidence. In: Siebes, A., De Raedt, L. (eds.) PKDD 2001. LNCS (LNAI), vol. 2168, pp. 424–435. Springer, Heidelberg (2001)
10. Wang, A.: Don't follow me: Spam detection in twitter. In: Proc. SECRYPT, pp. 1–10 (2010)