Identifying Controversial Issues and Their Sub-topics in News Articles

Yoonjung Choi, Yuchul Jung, and Sung-Hyon Myaeng

Korea Advanced Institute of Science and Technology
355 Gwahak-ro Yuseong-gu Daejeon, 305-701, South Korea
{choiyj35,enthusia77,myaeng}@kaist.ac.kr

Abstract. We tackle the problem of automatically detecting controversial issues and their subtopics from news articles. We define a *controversial issue* as a concept that invokes conflicting sentiments or views and a *subtopic* as a reason or factor that gives a particular sentiment or view to the issue. Conforming to the definitions, we propose a controversial issue detection method that considers the magnitude of sentiment information and the difference between the amounts of two different polarities. For subtopic identification, candidate phrases are generated and checked for containing five different features, some of which attempts to capture the relationship between the identified issue phrase and the candidate subtopic phrase. Through an experiment and analysis using the MPQA corpus consisting of news articles, we found that the proposed method is promising for both of the tasks although many additional research issues remain to be tapped in the future.

Keywords: Controversial issue detection, Subtopic identification, Sentiment analysis.

1 Introduction

People tend to be naturally interested in popular stories such as controversial issues and social events that invoke sentiment, which are usually found in news articles. Distinguished from opinions that have been a target for analysis in the natural language processing community[1], sentiment is more broadly defined to include one's judgment or evaluation, affective state, or intended emotional communication [Wikipedia].

Sentiment analysis, which determines the sentimental attitude of a speaker or writer, is known to be important for governments, companies, and individuals since governments should listen to public opinions to improve their services, companies need to scrutinize reviews to advance their products, and individuals want to know others' thinking or feeling about interesting subjects such as products, movies, and

[1] The term *sentiment analysis* has been used for the meaning of *opinion analysis* in the literature. In this paper, we define sentiment analysis to be broader and include opinion analysis.

H. Chen et al. (Eds.): PAISI 2010, LNCS 6122, pp. 140–153, 2010.

issues [3,4]. As such, many researchers have studied sentiment analysis and opinion analysis using such data as product reviews, blogs, and news articles, obtaining reasonable performances for the tasks of identifying subjective sentences or documents, determining their polarity values, and finding the holder of the sentiment or opinion found in a sentence [3,5,10].

While the past research has focused on the three tasks mentioned above, where sentiment is the focus, the main thrust of this paper is to identify controversial issues (or topics) in news articles and their reasons or subtopics for the controversy conveyed in the issues. That is, we focus on detecting controversial issues and their subtopics by means of sentiment information. We assume that a controversial issue receives sentiment of various sorts (e.g. positive vs. negative feelings, pros vs. cons, or rightness vs. wrongness in their judgments). A related subtopic often times invokes sentiment or serves as the reason why people feel or express particular sentiment. In short, our goal is to identify main topics/issues, which are controversial, and their subtopics by means of sentiment information conveyed in text segments.

Figure 1 shows an example using the topic "*Afghanistan War,*" one of the most controversial issues that appeared in newspapers for a while. Given sentiment-embedding sentences related to the topic, such as "*Obama supports the Afghanistan war*" and "*The Afghanistan war is perilous because of weapons of mass destruction*," we can identify subtopics such as *"Obama"* and *"weapons of mass destruction"*. By showing the subtopics on a time line as in the Figure 1, we can provide a sentiment-time summary of the subtopics for the issue of *Afghanistan war*.

The notion of topics in relation to sentiment analysis has been explored in the past. Several studies exploited topic relevance to analyze the sentiment [4, 6]. However, there was little explicit effort to find topics because topics were usually given for sentiment analysis. Even if topics were extracted as in [2, 4], they were different from the issues we deal with in this paper. A recent study [1] proposed a topic-sentiment mixture model to extract subtopics and sentiments, but the scope is limited in that they used product reviews. Compared to news articles, product reviews contain explicit opinions with obvious clues (e.g., "like", "good") and related subtopics that are usually pre-defined features given by the manufacturers (e.g. "battery"). Our work is unique in that we attempt to detect sentiment-invoking issues using news articles and then their related subtopics that serve as reasons or focal points for different views.

As a way to detect candidates for controversial issues, we first compute the weight of each noun phrase and verb phrase utilizing query generation methods [9]. We then determine whether the selected candidates are controversial or not by observing how many sentiment clues are included within pertinent sentences. If a candidate is not controversial enough, we choose another one until finding the proper one. For identification of related topics, phrases in the sentences near the identified issue are examined to determine whether each of them belongs to the sub-topic category or not. Features used for this classification function include collocation information between the candidate phrase and the sentiment clues and the degree to which the candidate is correlated with the issue.

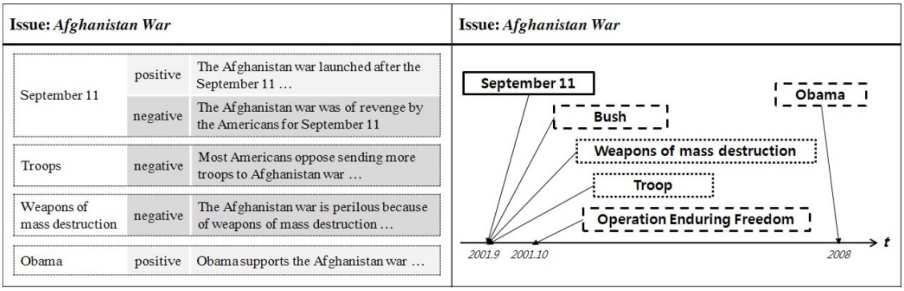

Fig. 1. A summary of the sentiment-generating subtopics for an issue "*Afghanistan War*"

2 Definition

We begin with some definitions of key concepts to be used in our models and analysis. While some of them are our own, other are borrowed from the existing literature and extended.

Controversial Issue: It is operationally defined to be a piece of text that lends itself to a query for a search engine and invokes conflicting sentiment or views. The criteria ensure its topical coherency and controversy aspects, respectively. While it can be any arbitrarily long text, we confine it to a noun or verb phrase in this paper. For example, "*Afghanistan war*" would be a controversial issue.

Subtopic: A subtopic is defined to be an entity or concept that is meaningfully associated and subordinated with a controversial issue. It plays the role of making the associated issue controversial and thus serves as a reason or a factor for a particular view or sentiment about the issue. Given a controversial issue, it is natural that related subtopics change over time. In this paper, the unit of a subtopic is confined to be a noun phrase. For example, "*troops*" can be a subtopic associated with "*Afghanistan war*". Note that a person who expresses an opinion on (e.g. *should send more*) or takes an action (e.g. *support*) for a topic can be a subtopic in that opinions or actions of different polarity (e.g. *should reduce* or *disapprove*) may exist.

Sentiment: It refers to an affective state invoked by a piece of text that describes an issue or topic involving real-world entities or abstract concepts such as events and policies. Such an affective state can be associated with evaluation or judgment, approval/disapproval, and feelings about an issue or a topic. Readers of a news article may feel some sentiment even from factual statements based on their own evaluations. Sentiment has three polarities: positive, neutral, and negative. For example, not only do subjective opinions about "*Afghanistan war*" contain sentiment but also a report on a death of a soldier and civilian generates sentiment.

Sentiment Clue: A sentiment clue is a word or phrase that causes positive or negative sentiment in a sentence [4]. Sentiment clues may be domain-dependent or independent. Some domain-independent clues are found in a lexicon such as SentiWordNet [7]. For example, "*sad*" or "*death*" can be sentiment clues for the topic "*troops*" under "*Afghanistan war*". However, the sentiment from a word like "*long*" or "*increase*" may depend on the domain, depending on what is described.

Sentiment Holder: It refers to the source of sentiment [8], which can be explicit or implicit, depending on whether a sentiment-revealing sentence contains the entity who expresses the sentiment. When a factual statement invokes sentiment, the author is assumed to be the implicit sentiment holder. For example, "*Obama*" in "*Obama supports Afghanistan war*" is the sentiment holder since the sentiment clue "*support*" generates positive sentiment. In this paper, a sentiment holder can be a subtopic as explained above.

3 The Method

Our proposed method consists of two parts. The first one is for issue detection with which controversial issues are identified from news articles. We measure the "controversiness" of a phrase by its topical importance and sentiment gap it incurs. The other part is for extraction of subtopics that are related to the detected issue

3.1 Controversial Issue Detection

According to the definition in Section 2, a controversial issue would be often found in the form of a search query that retrieves a set of documents containing positive or negative sentiment towards the query topic. Therefore, the first step for identifying a controversial issue is to guess a potential query. We do this by adopting a known-item query generation method in [9] and extending it to generate controversial issues. A known-item query can be generated from the document known to be relevant based on its probabilistic model. Given a document, the algorithm selects a query length *l* and repeats *l* times the process of selecting a term based on its generation probability from the document model. Our algorithm follows the same flow:

1. Initialize issue term set: $issue\_terms = \{\}$
2. Repeat *l* times: /* *l* is an empirical value */
   i. Select a term t_i with probability $p(t_i \mid \theta_{it})$
   ii. Add t_i to the issue term

but uses a different method for the probability calculation to suit our need and extends it further to consider phrases. The issue term model can be expressed as in Equation (1) where θ_{it} is the model of the issue term. The probability is estimated based on the mixture of a topic model and sentiment model because a term in an issue may contain sentiment in it (e.g. "*war*").

$$p(t_i \mid \theta_{it}) = \lambda \cdot p(t_i \mid \theta_{topic}) + (1-\lambda) \cdot p(t_i \mid \theta_{senti}) \qquad (1)$$

where λ is parameter whose range is $0 \le \lambda \le 1$.

The topic model can be estimated as:

$$p(t_i \mid \theta_{topic}) = \frac{n(t_i, D)}{\sum_t n(t, D)} \tag{2}$$

where *D* is the set of news documents, and $n(t_i, D)$ is the number of occurrences of t_i in *D*. The sentiment model is based on calculation of sentiment score of each term:

$$scr(t) = MAX[scr(t \mid POS), scr(t \mid NEG)] \tag{3}$$

$$p(t_i \mid \theta_{senti}) = \frac{scr(t_i)}{\sum_t scr(t)} \tag{4}$$

where $scr(t \mid POS)$ and $scr(t \mid NEG)$ are a positive sentiment score and a negative score of the term *t*, respectively, which are provided by SentiWordNet[2].

In order to handle a phrase *ph*, we consider the issue terms contained in it and calculate its score, average probability:

$$w(t_i) = \begin{cases} p(t_i \mid \theta_{it}) & if\ t_i \in issue\_term, \\ 0 & otherwise \end{cases} \tag{5}$$

$$score(ph) = \frac{\sum_{t_i \in ph} w(t_i)}{\mid ph \mid} \tag{6}$$

where $\mid ph \mid$ is the number of terms of the phrase.

While the algorithm selects phrases with one or more issue terms, there is no guarantee that they are controversial enough. An additional step is to check the degree of controversy using the contextual information. We first compute the score for positive and negative sentiment for a phrase (Equation (7)) and then determine if it is sufficiently controversial not only by the sum of the magnitude of positive and negative sentiments but also the difference between them (Equation (8)). A topic may be of great importance for people but may not be controversial if its sentiment has one polarity (e.g. "*swine flu*" is only negative).

$$scr_{POS} = \sum_{t \in \theta_{ph}} scr(t \mid POS),\ scr_{NEG} = \sum_{t \in \theta_{ph}} scr(t \mid NEG) \tag{7}$$

$$controversial(ph) = \begin{cases} 1 & if\ scr_{POS} + scr_{NEG} \geq \delta, \mid scr_{POS} - scr_{NEG} \mid \leq \gamma \\ 0 & otherwise \end{cases} \tag{8}$$

where $\theta_{ph} = \{sentence \mid ph \in sentence\}$, and δ and γ are empirical values.

3.2 Subtopic Extraction

As we defined in Section 2, every noun phrase in the issue document (i.e., a news article containing one of the detected issues) is eligible as subtopic candidates. After

[2] SentiWordNet, `http://sentiwordnet.isti.cnr.it`

extracting all atomic noun phrases using a parser[3] from each document containing at least one detected controversial issue, we generate a statistical classifier to select a subtopic based on collocation information. A subtopic would co-occur not only with a detected issue but also with sentiment clues, which we use as classification features. In order to incorporate the features, we opt for the linear regression model that works well even with a small amount of training data. The five types of features we consider are explained below: two basic features and three statistical features.

One of the two basic features is whether or not the title section in a document includes a given candidate; a candidate is likely to be a subtopic if it appears on the title section in a document. The other is where in the sentence the candidate appears as it is likely to be a subtopic if it appears on the subject or object position in a sentence. In the case of "*Afghanistan war*", for example, words like "*Bush*", and "*Troop*" would appear on the title of a news article. In addition, they would occupy the position of the subject or object in a sentence.

In addition, three statistical features are used in our method: contextual similarity measured with issue likelihood and subtopic likelihood distribution models, subtopic likelihood computed through a sentiment model obtained for the issue, and direct correlation between an issue and a subtopic in terms of their co-occurrence in the same sentences. The first statistical feature considers the extent to which an issue and a candidate subtopic emerge from the same context whereas the second measures how strongly a subtopic is related with the sentiment in the article. The third feature measures how close the subtopic is to the issue.

We calculate contextual similarity between an issue and a noun phrase (NP) subtopic candidate through KL-divergence that calculates the difference between two probability distributions. The NP and issue models are constructed out of the word frequencies of the sentences containing the term, after removing stop words.

$$KL(\theta_{NP} \,||\, \theta_{Issue}) = \sum\nolimits_{t_i \in \theta_{Issue \cup NP}} p(t_i \mid \theta_{NP}) \cdot \log \frac{p(t_i \mid \theta_{NP})}{p(t_i \mid \theta_{Issue})} \tag{9}$$

where $\theta_{NP} = \{sentence \mid NP \in sentence\}$, $\theta_{Issue} = \{sentence \mid Issue \in sentence\}$, and $\theta_{Issue \cup NP} = \{sentence \mid (Issue \in sentence) \cup (NP \in sentence)\}$. Furthermore, the probabilities are estimated as follows:

$$p(t_i \mid \theta_{NP}) = \frac{n(t_i, \theta_{NP})}{|\,\theta_{NP}\,|} \tag{10}$$

where $|\,\theta_{NP}\,|$ is the number of sentences, and $n(t_i, \theta_{NP})$ is the number of sentences which contains a term t_i in θ_{NP}. $p(t_i \mid \theta_{Issue})$ can be estimated in a similar way. KL divergence is 0 if and only if two models (i.e., NP and Issue model) are the same. In this paper, we take its inverse and normalize it to make the similarity value range between 0 and 1.

[3] We used Stanford Parser in
`http://nlp.stanford.edu/software/lex-parser.shtml`

To compute the relatedness of a subtopic candidate with the sentiment expressed for the issue, we estimate the probability of a candidate given an issue as follows:

$$P_{senti}(NP \mid Issue) = \frac{\sum_{t_i \in \theta_{NP \cap Issue}} scr(t_i)}{\sum_{t \in \theta_{Issue}} scr(t)} \qquad (11)$$

where $\theta_{Issue \cap NP} = \{sentence \mid (Issue \in sentence) \cap (NP \in sentence)\}$.

While the contextual similarity (the first statistical feature) measures the extent to which an issue and a subtopic share the same context, they may occur in different sentences. In order to give a higher weight to a subtopic that occurs together with an issue in the same sentence, we measure their sentential correlation by counting the number of times they appear in the same sentence as follows:

$$Cor(S_{Issue}, S_{NP}) = \frac{\mid S_{Issue} \cap S_{NP} \mid}{\sqrt{\mid S_{Issue} \mid} \cdot \sqrt{\mid S_{NP} \mid}} \qquad (12)$$

where $|.|$ is the number of sentences in a set.

4 Experiment

4.1 Experimental Setup

We used the MPQA corpus that contains news articles from 187 different foreign and U.S. news sources, ranging June 2001 to May 2002, some of which are classified into 10 different topics [12]. The articles were collected based on information retrieval system search results and manual analysis. Since they contain sentiment information, the 10 topics can be regarded as controversial issues. It contains 355 documents with topic information and 8,955 sentences. Table 1 lists the topics and the number of documents in each topic.

To build a gold standard for evaluation of the subtopic extraction method, a portion of the corpus was manually analyzed and the subtopics for the issues were tagged. Five topics, AE, GB, HR, KP, and ZI, were chosen randomly because of the limited human resources. Since polarized sentences are marked in the MPQA corpus, we easily obtained 2,002 (36.84%) polarized sentences among 5,434 in the set of 204 documents in the chosen topics. Three annotators picked subtopics from the polarized sentences. When there are conflicts, they were resolved by majority voting. As a sentence may have more than one subtopic instance, a total of 2,723 subtopics (1.36 subtopics / sentence) were identified by one or more annotators. Among those, a total of 1,947 subtopics (71.6%) were agreed by two or three annotators. After collapsing the redundant subtopic instances (i.e. the same phrases), the average number of unique subtopics per issue was 17.8.

Table 1. The topic list and the number of documents about each topic in the MPQA corpus

Topic	# documents
President Bush's 2002 State of the Union Address – axis of evil (AE)	30
U.S. holding prisoners in Guantanamo Bay (GB)	55
Reaction to U.S. State Department report on human rights (HR)	30
Ratification of Kyoto Protocol (KP)	43
2002 president election in Zimbabwe (ZI)	46
Israeli settlements (IS)	25
Space missions of various countries – space station (SM)	30
Relations between Taiwan and China (TC)	28
Presidential coup in Venezuela (VE)	37
Economic collapse in Argentina (AR)	31

4.2 Experimental Results

Controversial Issue Detection. It is not easy to determine how appropriate a detected issue is. The goal of this part of the experiment is not to quantify any performance but to obtain a qualitative assessment of the proposed issue detection method. Issue phrases were detected based on the algorithm with two thresholds δ and γ set to 250.0 and 50.0, respectively (As we mentioned, two thresholds are empirical values).

Table 2 shows top-three phrases detected as controversial issues for the ten topic categories, together with the number of detected phrases over the thresholds. The underlined parts indicate they appear in the original topic category names in the corpus as in Table 1. Although the algorithm does not consider the location of phrases, those in the human generated topic titles were captured as controversial issues. Even in the case of IS (Israeli settlement), where none of the system-generated issues appear in the human-generated titles, our perusal of the related articles indicates that the main issue centres on "*occupied Palestinian territory*" for which "*Palestinian*" and "*Palestinian state*" seem appropriate. We feel that not only are the issues close to the topics but also the granularity of the detected issues is appropriate by and large.

Table 2. Sentiment issues extracted from MPQA corpus

Topic	Count	Detected sentiment issues		
AE	5	axis of evil	president Bush	Iran Iraq and North Korea
GB	4	prisoner	Guantanamo bay	Guantanamo detainees
HR	3	state department report	human rights	the human rights report
KP	4	Kyoto protocol	greenhouse	climate change
ZI	4	Mugabe	Zimbabwe election	presidential election
IS	3	Palestinian	Israel	recognize Palestinian state
SM	3	space station	Russian space officials	space activities
TC	4	Taiwan	China	Taiwan relations
VE	3	president Hugo Chavez	Venezuela	Chavez government
AR	4	Argentine government	economy	help Argentina

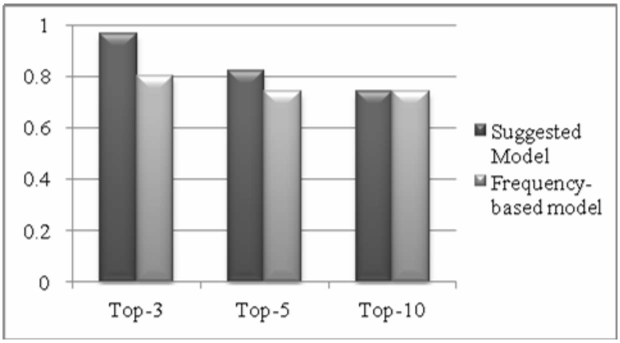

Fig. 2. Comparison of the controversial issue detection model with the frequency-based model

We adopted a user evaluation to quantify the performance. We chose top-ten phrases detected as controversial issues for each topic and provided these to three users with random-ordering. Then, users judged whether a given phrases is appropriate for a controversial issue or not. We considered phrases correct when they were agreed by two or three users. We compared the ranking generated by the proposed model with that of the frequency-based model which ranks the phrases in their frequency (Figure 2). We experimented in three cases: top-three phrases, top-five phrases, and top-ten phrases. The precision is calculated as follow:

$$precision = \frac{the\ number\ of\ correct\ phrases}{the\ total\ number\ of\ phrases} \tag{13}$$

In the top-ten phrases case, they showed the same result because the set of the phrases is identical and only rank lists are different. Although the difference between two models is small in the top-five phrases case (the precision of our suggested model is 0.83 while that of the frequency-based model is 0.74), our suggested model outperforms the frequency-based model, and the precision is nearly 1.0 in the top-three phrases case (the precision of our suggested model is 0.9667 while the frequency-based model is 0.8).

However, our model includes some problems. Most of the detected phrases are noun phrases while verb phases would be equally useful as issues. Our analysis shows that the frequency-based algorithm prefers noun phrases because noun phrases are repeated more often than verb phrases in the same article. This is the reason why "*recognize Palestinian state*" for IS and "*help Argentina*" in AR were ranked low in the list. Taking this phenomenon into account is left for future work.

Subtopic Extraction. The subtopic extraction method was evaluated with a manually annotated gold-standard. Its performance was measured in terms of precision and recall, which counts how many of the extracted subtopics were found in the gold-standard and how many of the annotated subtopics were detected by the method, respectively.

The precision and recall results are shown in Table 3 for the five topic areas chosen for this part of the experiment. It was not possible to make comparisons with other methods, simply because we do not know any other previous approach to sub-topic identification for a given issue. The precision and recall values are not high in general. There are several weaknesses in the proposed algorithm. Most notably, its inability to handle anaphora that is found very often in news paper articles gave many incorrect subtopics. Since we considered all noun phrases in a sentence as candidates of subtopics, anaphoric phrases, such as "*these three counties*", "*such policy*", and "*his point of view*", also are considered as candidates. For example, for the incorrectly detected subtopic "*these three countries*", the correct answer in the gold standard is the list of actual country names such as "*Iran, Iraq, and North Korea*", i.e., the noun phrase which the anaphor refers to. Furthermore, semantically identical expressions that differ from each other at the surface-level are treated as different phrases. For example, "*Iran, Iraq and DPRK*" and "*Iran, Iraq, and North Korea*" are treated as different phrases. Another example is "*mass destruction weapons*" and "*weapons of mass destruction.*" Since the expressions in news articles are very diverse, we found there are many cases involved with this problem. Another weakness is associated with reliance on frequency in the features. Some of the important subtopics do not actually occur very often in news articles. Finally, incongruent candidates, such as "*all efforts*" and "*a few countries*", also are problem because we considered all noun phrases as candidates.

Table 3. Precision and recall of our proposed model

Topic	Precision	Recall	F1-score
AE	0.4170	0.7778	0.5429
GB	0.4000	0.6842	0.5049
HR	0.3167	0.7778	0.4501
KP	0.3167	0.8333	0.4590
ZI	0.4000	0.7500	0.5217
Total	**0.3700**	**0.7640**	**0.4986**

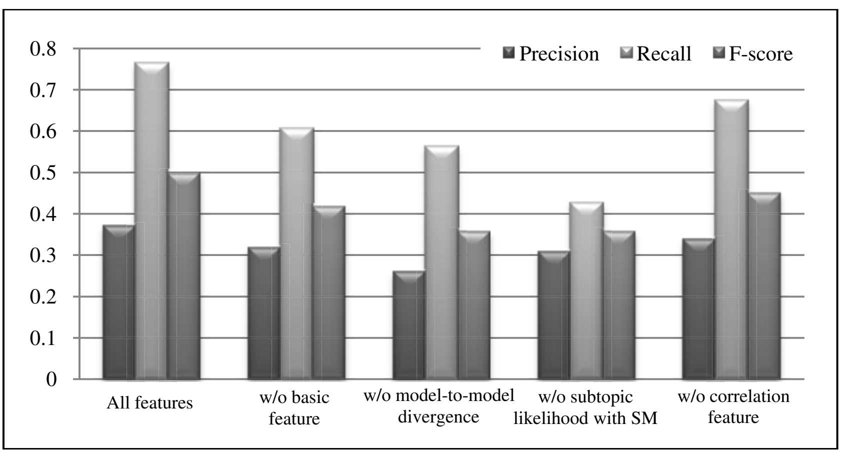

Fig. 3. The effect of ignoring one feature at a time

In order to see relative importance of the five features, we measure changes of classification performance caused by excluding one feature at a time. As in Figure 3, the subtopic likelihood with respect to the sentiment model (SM) and the model-to-model divergence turn out to be the most important since the performance was decreased most significantly. Without associating sentiment information, the subtopics were mostly terms with high frequency and named entities (e.g., "*Bush*" and "*U.S.*"). Without the contextual similarity measured by KL-divergence, many of the extracted subtopics are not directly related to the issue. Although the basic features are simple, their effect cannot be disregarded. The correlation feature contributes the least but also affect the performance in a non-trivial way.

Application. We envision that the proposed method can be used to build a system that detects a controversial issue (or alternatively received a user query for it) and organizes the subtopic detection result as in Figure 4, where the identified subtopics are shown on the left together with the polarity, the actual text, and the date it appears first. This will allow users to quickly understand what subtopics exist, what sentiment has been expressed, and what the examples are.

Axis of Evil			
A spokesman of the EU Commission	NEG	A spokesman of the EU Commission, at a press conference, expressed concerns over the Bush gaffe labeling North Korea, Iran, and Iraq as the axis of evil and said that the EUs high representatives do not agree to such policy.	2002.03.14
the Finnish and B elgian	NEG	On 17 February, the Finnish and Belgian Foreign Ministers also expressed opposition to the so-called axis of evil remark and the US plan to launch strikes against Iraq.	2002.03.15
a new threat	NEG	In relation to Bushs axis of evil remarks, the German Foreign Minister also said, Allies are not satellites, and the French Foreign Minister caustically criticized that the United States unilateral, simplistic worldview poses a new threat to the world.	2002.03.15
weapons of mass destruction	POS	A day after President Bushs threat to crush these countries and use all means to prevent them from developing weapons of mass destruction, Secretary Powell said before the Senate Foreign Relations Committee that characterizing these counties as the axis of evil does not mean that his government intends to invade them. Another goal is to prevent countries that support terrorism from threatening the United States …	2002.02.01

Fig. 4. An example result of a controversial issue and subtopics

5 Related Work

To the best of our knowledge, there has been no attempt to identify controversial issues and their related subtopics that serve as reasons for different sentiments from news papers. The closest work we know is found in Qiaozhu Mei et al. [1]. It proposes a topic-sentiment mixture model to extract topics and their sentiments from blog articles. They used a mixture of multinomials; background topic model, subtopic model, positive sentiment model, and negative sentiment model. To evaluate the topic extraction model, they used only two data sets which are constructed by submitting queries such as "*ipod*" and "*da vinci code*". Since a query is a product like a movie or a book and the blog articles are review data, its subtopics are related to its attribute. Aside from the fact that they only deal with unigrams where as we deal with phrase, their approach is not to detect issues and their subtopics but to find products and their attributes, which is much simpler problem.

Another is Ba-Quy Vyong et al. [11] work. They suggested Controversy Rank model in Wikipedia. They thought a dispute in an article is more controversial, so this model utilized the controversy level of disputes which can be derived from the articles' edit histories. The CR Models defined the article controversy score (an article is controversial when it has a lots disputes among less contributor) and the contributor controversy score (a contributor is controversial when he/she is engaged in a lots disputes in less articles). However, this model can only apply to Wikipedia because general news articles have no edit histories and no contributor except writer.

There was an attempt to extract a sentiment topic, which is similar to a controversial issue. Kerstin Denecke et al. [6] focused on discovering a main topic and identifying sentiment at sentence level from blogs. They detected a topic by applying the Latent Dirichlet Allocation algorithm, and identified its sentiment using SentiWordNet. Although their output is a topic and its sentiment at sentence-level, to be accurate, we cannot say the definition of a topic in this study is same as our definition of a controversial issue because they do not use the sentiment model when a main topic is detected.

Soo-Min Kim and Eduard Hovy [10] tried to extract an opinion topic utilizing FrameNet from news articles. If all sentences are simple and FrameNet covers lots of opinion words, this method would work very well because an opinion topic may be semantically related to an opinion word. However, the method cannot capture a sentiment topic if there are complicated semantic relations. To resolve this problem, it is necessary to analyze text in a complex manner. Besides, it can be applied to only sentence-level topic detection.

The work in [4] extracts a sentiment topic exploiting co-occurrence information with sentiment clues. It computes cosine similarity between a candidate (noun phrase) and sentiment clues to identify a sentiment topic. However, the goal is to generate domain specific sentiment clues for the sentiment classification by using topics as a vehicle, not to identify a sentiment topic of issues. Besides, the boundary of sentiment topics is confined to a sentence.

6 Conclusion and Future Work

This paper tackles two problems: controversial issue detection and their subtopic extraction. For issue detection, we identify noun or verb phrases as candidate issues using a mixture of topical and sentiment models. To compute the degree of controversy, we measure the amount of both positive and negative sentiment and the difference between them. For subtopic extraction, we generate noun phrases as candidates and calculate their feature scores for classification. We use two positional features and three statistical features. The statistical features are: contextual similarity between the issue and a subtopic candidate, relatedness of a subtopic to sentiment, and the degree of physical vicinity between the issue and the candidate phrases. The experimental result shows that the proposed method is reasonable as the first attempt in extracting controversial issues and their related subtopics.

Our qualitative analysis of the result generates a number of possible extensions to the current method. Besides adding more meaningful features for classification and improvements in probability estimates, which are always possible, we need to investigate on issue and subtopic boundary detection for the right level of granularity, beyond the current approach of using verb and noun phrases. Another important issue is to deal with paraphrases both for the purpose of identifying both issues and subtopics but also for evaluations.

Acknowledgments. This work was financially supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency" (NIPA-2010-C1090-1011-0008).

References

1. Mei, Q., Ling, X., Wondra, M., Su, H., Zhai, C.X.: Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs. In: The 16th International Conference on World Wide Web (WWW), pp. 171–180. ACM, Canada (2007)
2. Denecke, K., Taytsarau, M., Palpanas, T., Brosowski, M.: Topic-related Sentiment Analysis for Discovering Contradicting Opinions in Weblogs. Technical Report, University of Trento (2009)
3. Ku, L.-W., Liang, Y.-T., Chen, H.-H.: Opinion Extraction, Summarization and Tracking in News and Blog Corpora. In: American Association for Artificial Intelligence-Spring Symposium on Computational Approaches to Analyzing Weblogs (AAAI-CAAW), pp. 100–107 (2006)
4. Choi, Y., Kim, Y., Myaeng, S.-H.: Domain-specific Sentiment Analysis using Contextual Feature Generation. In: 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement (TSA), pp. 37–44. ACM, Hong Kong (2009)
5. Zhuang, L., Jing, F., Zhu, X.-Y.: Movie Review Mining and Summarization. In: 15th ACM International Conference on Information and Knowledge Management (CIKM), pp. 44–50. ACM, USA (2006)
6. Zhang, M., Ye, X.: A Generation Model to Unify Topic Relevance and Lexicon-based Sentiment for Opinion Retrieval. In: 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 411–418. ACM, Singapore (2008)
7. Esuli, A., Sebastiani, F.: SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In: 5th Conference on Language Resources and Evaluation (LREC), pp. 417–422 (2006)

8. Stoyanow, V., Cardie, C.: Annotating Topics of Opinions. In: 6th International Conference on Language Resources and Evaluation (LREC), pp. 3213–3217 (2007)
9. Azzopardi, L., de Rijke, M., Balog, K.: Building Simulated Queries for Known-Item Topic: An Analysis using Six European Languages. In: 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 455–462. ACM, Amsterdam (2007)
10. Kim, S.-M., Hovy, E.: Extracting Opinions, Opinion Holder, and Topics Expressed Online News Media Text. In: The Workshop on Sentiment and Subjectivity in Text, pp. 1–8. Association for Computational Linguistics, Sydney (2006)
11. Vuong, B.-Q., Lim, E.-P., Sum, A., Le, M.-T., Lauw, H.W., Chang, K.: On Ranking Controversies in Wikipedia: Models and Evaluation. In: 1st ACM International Conference on Web Search and Data Mining (WSDM), pp. 171–182. ACM, USA (2008)
12. MPQA corpus, `http://www.cs.pitt.edu/mpqa`