2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology

Stakeholder Mining and Its Application to News Comparison

Tatsuya Ogawa
Graduate School of Informatics
Kyoto University
Kyoto, Japan
Email: ogawa@db.soc.i.kyoto-u.ac.jp

Qiang Ma
Graduate School of Informatics
Kyoto University
Kyoto, Japan
Email: qiang@i.kyoto-u.ac.jp

Masatoshi Yoshikawa
Graduate School of Informatics
Kyoto University
Kyoto, Japan
Email: yoshikawa@i.kyoto-u.ac.jp

***Abstract*—In this paper, we propose a novel stakeholder mining mechanism for analyzing bias in news articles by comparing descriptions of stakeholders. Our mechanism is based on the presumption that interests often induce bias of news agencies. As we use the term, a "stakeholder" is a participant in an event described in a news article who should have some relationships with other participants in the article. Our approach attempts to elucidate bias of articles from three aspects: stakeholders, interests of stakeholders, and the descriptive polarity of each stakeholder. Mining of stakeholders and their interests is achieved by analysis of sentence structure and the use of RelationshipWordNet, a lexical resource that we developed. For analyzing polarities of stakeholder descriptions, we propose an opinion mining method based on the lexical resource SentiWordNet. We also describe an application system we developed for news comparison based on the mining mechanism. This paper presents a user study to validate the proposed methods.**

***Keywords*-Stakeholder Mining; RelationshipWordNet; Relationship Structure; Bias Analysis**

I. Introduction

Due to intentions of news agencies and their sponsors, in a sense, news is never free from bias. Bias often causes readers to misunderstand the facts of actual events and even the whole story. Although a large number of studies [1]–[4] have been made on analyzing bias by means of comparing related news articles, conventional methods present related articles and ask users to compare them. To the best of our knowledge, models and criteria for bias analysis have not yet been well studied.

In this paper, we focus on the relation between media bias and interests, on the presumption that the latter induce the former. Under this presumption, it is possible to make bias overt by analyzing the descriptions on stakeholders and their relationships appearing in news articles.

As a bias analysis method, we propose a novel stakeholder mining mechanism to extract stakeholders referred to in news articles and relationships among them.

Based on the stakeholder mining mechanism, we propose a system for comparing news articles (see Figure 1). The system enables us to elucidate bias from three descriptive aspects: 1) stakeholders in articles, 2) interests of stakeholders, and 3) the descriptive polarity of each stakeholder.

We define a "stakeholder" as a participant in an event described in a news article and who should have some relationships to another participant or participants. Stakeholders described in a news article specify the objects to which the article refers. When two articles are compared, some stakeholders are referred to by only one article. In other words, the scope of events that news articles deal with is biased because one article may describe the stakeholder while another does not.

An interest state is represented by a pair of values, one being a positive relationship indicating a degree of corresponding interests and the other being a negative degree indicating that of conflicting interests. Comparisons of these values and the existences or nonexistence of interests also show a bias.

Descriptive polarity is represented by two numerical scores indicating positivity and negativity of a descriptive tendency. The polarity makes it possible to deal quantitatively with the viewpoints of news articles.

A sentence referring to the relationship among stakeholders includes words that express their interests because stakeholders share interests. This suggests that 1) words that express interests indicate states of the interests and 2) the structure of a sentence specifies the stakeholders in each interest. Therefore, we constructed RelationshipWordNet, a lexical resource in which each word assigns scores indicating a state of relationship, and Relationship Structure, a sentence structure appropriate for extracting relationships. Relationship structure is also used to descriptive polarity analysis because the structure is helpful to identify stakeholders that a description expresses.

II. Stakeholder Mining Mechanism

A. RelationshipWordNet and Relationship Structure

To detect the participants sharing interests in an event, we build a lexical resource named RelationshipWordNet and construct a sentence structure named relationship structure.

1) RelationshipWordNet: We construct RelationshipWordNet from WordNet [5] to specify words expressing interests. WordNet is a lexical resource that groups terms into sets of cognitive synonyms (synsets), each expressing a distinct concept. RelationshipWordNet assigns all synsets

978-0-7695-4191-4/10 $26.00 © 2010 IEEE
DOI 10.1109/WI-IAT.2010.156

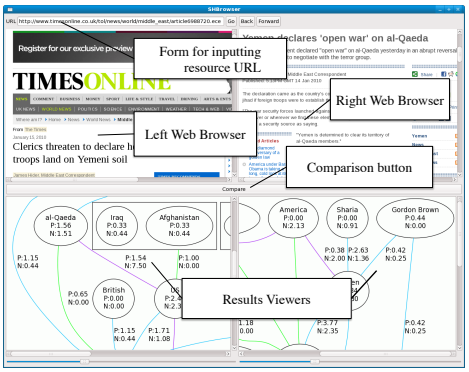

Figure 1. Screenshot of News Comparison System

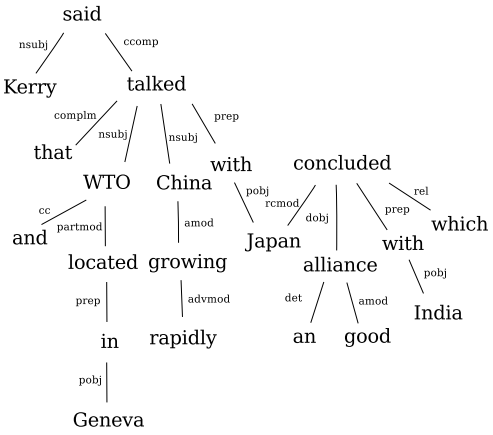

Figure 2. Example of relationship structure for the sentence "Kerry said that the WTO, located in Geneva, and the rapidly growing China talked with Japan, which concluded a good alliance with India."

with three numerical scores: "positivity", "negativity", and "objectivity". A synset that express a better (resp. worse) relationship has a larger positive (resp. negative) score. The "objective" score indicates the degree of objectivity the corresponding synset has.

RelationshipWordNet is constructed by modifying the construction method of SentiWordNet [6]. Three scores of the former indicate the state of relationships, while those in the latter express the degree of sentiments. Therefore, we need to assign different scores to synsets for RelationshipWordNet. To obtain the scores, we change the seed sets of SentiWordNet. First, We collect synsets that each includes any word in 2193 words of The Longman Definition Vocabulary. Then, we manually classify the synsets into the three kinds of seed sets L_p, L_n and L_o by labeling each synset "positive", "negative", or "objective" according to the state of relationship the synset expresses in case it is used for describing the relationship of stakeholders. As a part of seed sets, L_p includes "depend" and "comfort". L_n includes "attack" and 'blame". L_o includes "say" and "go".

Hereinafter, we use the expression *relation word* to refer to a term included in a synset whose "positive" or "negative" score is more than 0.

2) Relationship Structure: We define relationship structure, a sentence structure suitable for identifying stakeholders who share interests. An example of relationship structure is shown in Figure 2.

Relationship structure is constructed on the basis of Stanford Dependencies [7]. Stanford Dependencies provide grammatical relations in the form: *type* (*governor*, *dependent*). We can obtain a tree structure of a sentence by identifying *governor* as a parent node and *dependent* as a child node.

It is important to identify positional relations among relation words, verbs, and stakeholder candidates in a sentence structure to specify the relationships among stakeholders. Tree structure operations are necessary due to the need to identify the positional relations on the basis of the sentence structure. Therefore, to construct a relationship structure, we operate a tree structure corresponding to the grammatical relations shown in Table I.

Table I
OPERATIONS FOR GRAMMATICAL RELATIONS

Type	Operation
conj	Omit this relation from the structure and the parents of the *governor* are added to the parents of *dependent*. Furthermore, if both *governor* and *dependent* are verbs, the children of *governor* except *dependent* is added to those of *dependent*.
appos	
rcmod	Interchange *governor* and *dependent*.
cop	

B. Details of Stakeholder Mining Mechanism

1) Extraction of Interests: An interest is represented as $(\{s_1, s_2\}, P, N)$ where s_1 and s_2 are stakeholders that have an interest and P and N are numerical scores ranging from 0.0 to 1.0, in which P indicates the degree of coincidence between the interests and N indicates the degree of conflict between them.

The interests extraction procedure is executed for each relation word w_R in a sentence. Here, we define a "stakeholder candidate" as named entities such as country, person and organization. Furthermore, in relationship structure of a sentence, we define a "root" as a node that does not have any parent nodes, a "leaf" as a node that does not have any child nodes, $V(w_R)$ as a set of verbs for which the distance between w_R and the verb is the shortest on each path from w_R to a root in ascendants of w_R, and $N_e(v)$ as a set of stakeholder candidates for which the distance between a verb v and the stakeholder candidate is the shortest on each path from v to a leaf in descendants of v. In the interests

Table II
STRUCTURE TYPES FOR INTERESTS EXTRACTION

Structure type	A set of stakeholder candidate pairs
N_sVN_p	$\{(s_i, s_j) \mid s_i \in N_s, s_j \in N_p\}$
N_sVN_n	$\{(s_i, s_j) \mid s_i \in N_s, s_j \in N_s, s_i \neq s_j\}$
N_nVN_p	$\{(s_i, s_j) \mid s_i \in N_p, s_j \in N_p, s_i \neq s_j\}$

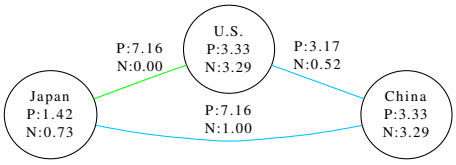

Figure 3. Example of Relations Graph

extraction procedure, interests are extracted by referring to the structure types defined in Table II. A structure type matches a part of the relationship structure of a sentence. That part forms a tree structure. In a structure type, V matches the root verb of a part of a relationship structure, the symbol to the right of V means whether stakeholder candidates exist in the subject, and the symbol to the left of V means whether stakeholder candidates exist in the predicate. $N_s(v)$ indicates stakeholder candidates in $N_e(v)$ exist in the subject. $N_p(v)$ indicates stakeholder candidates in $N_e(v)$ exist in the predicate. N_n means there are no stakeholder candidates in the subject or predicate where N_n is located. The interests extraction procedure for each w_R in a sentence is described below.

Step 1 If any stakeholder candidates are on a path from w_R to a verb v in $V(w_R)$, w_R is considered as a modification to the nearest stakeholder candidate to w_R on that path, and does not represent interests. If w_R satisfies the condition, the procedure for w_R stops.

Step 2 For each v in $V(w_R)$, a set of stakeholder candidate pairs is extracted according to the expression in Table II where the structure type matches the tree structure of which v is the root. If no structure type matches the tree structure, a set of candidate pairs is not extracted. Each pair of stakeholder candidates shares interests. We consider those candidates as stakeholders. If no interests are extracted for any v, the procedure for w_R stops.

Step 3 If an extracted interest is obtained for the first time, initialize P and N values of the interest to 0. Positive and negative scores of w_R are respectively added to P and N values of each interest. At this time, if the sentence includes a negation word, that is, the grammatical relations of the sentence include *neg*(*governor*, $v \in V(w_R)$) or *neg*(*governor*, w_R), w_R is handled in the way its positive and negative scores are interchanged because w_R is used to represent the opposite meaning.

2) Descriptive polarities of stakeholders: Descriptive polarity denotes the opinion of a news agency on a stakeholder. The polarity is represented by two scores, each indicating a degree of positive or negative description of a stakeholder. Descriptive polarities enable us to deal quantitatively with the viewpoints of news articles to stakeholders.

We use relationship structure and a sentiment dictionary to analyze descriptive polarity. Relationship structure is used to identify stakeholders that a description expresses, and the sentiment dictionary SentiWordNet is used to analyze polarities. Hereinafter, we use the expression *sentiment word* to refer to a term included in a synset whose "positive" or "negative" scores in SentiWordNet is more than 0.

The procedure of descriptive polarity analysis is executed for each sentiment word w_S in a sentence. Here, in relationship structure of a sentence, we define $V(w_S)$ as a set of verbs for which the distance between w_S and the verb is the shortest on each path from w_S to a root in ascendants of w_S, and $S(v)$ as a set of stakeholders for which the distance between a verb v and the stakeholder is the shortest on each path from v to a leaf in v's descendants. The procedure for each w_S in a sentence is described below.

Step 1 If any stakeholder is on a path from w_S to a verb in $V(w_S)$, w_R is considered as a modification to the nearest stakeholder to w_S on that path. Therefore, positive and negative scores of w_S are respectively added to the P and N values of the stakeholder ' s descriptive polarity.

Step 2 If no stakeholder is on each path from w_S to a verb in $V(w_S)$, w_S is considered to modify all the stakeholders in $S(v)$, $v \in V(w_S)$. Therefore, positive and negative scores of w_S are respectively added to the P and N values of the stakeholder ' s descriptive polarity in $S(v)$.

In these two steps, if the sentence has negation words the scores of w_S are dealt with in the same way as that in the stakeholder extraction process described above.

3) Construction of Relations Graph: We obtain the interests described in an article as a whole by summing up all the scores of the same interest in each sentence. Similarly, we obtain the descriptive polarities by summing up all the scores of the same stakeholder in each sentence.

Stakeholders and their interests can be represented as a graph. The vertices correspond to stakeholders and the edges correspond to their interests. The label of an edge denotes the interest. The label of a vertex denotes the descriptive polarity. We call such a graph a "relations graph" (e.g. Figure 3). Stakeholders who have positive relationships and have a meeting of interests against other stakeholders in an event are grouped by using a relations graph.

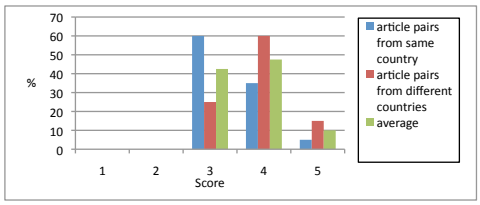

Figure 4. Results for Question 1

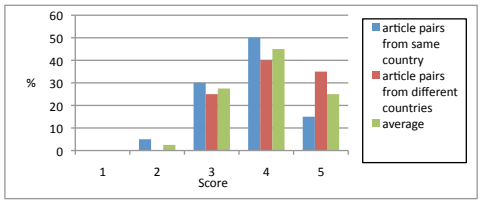

Figure 5. Results for Question 2

III. Experiments

We developed an application system for news comparison using a mining results (see Figure 1), and conducted a user study with the system to validate the bias analysis.

The subjects evaluated the method by comparing two articles using the mining results of each article. As a dataset, we selected ten pairs of articles from ten news topics. Five of them were reported by news agencies located in the same countries and the other five pairs were reported by news agencies located in different countries. As the subjects, four undergraduate students participated in the evaluation. They were asked to compare the news article pair for each topic while referring to the corresponding mining results and then fill out a questionnaire. The questions they were asked were:

1) Did the mining results help you to understand the bias of the two articles?
2) Were the mining results representative of the bias of the two articles?

The subjects gave point-score answers ranging from 1 to 5, with 5 being the best and 1 being the worst.

Figure 4 shows the results obtained for Question 1 in the form of score percentages corresponding to the same country pairs, different country pairs, and the average for both. Figure 5 shows the results obtained for Question 2. Table III shows the averages for both questions and the percentages of pairs whose score was 4 or 5.

The user study we conducted showed the effectiveness of using mining results to analyze bias in news articles. As shown in Table III, the percentages of the evaluations in which our stakeholder mining mechanism was found to be useful for bias analysis were 58.5% for Question 1 and 70.0% for Question 2. Obvious differences were found in comparing results obtained for article pairs prepared by news agencies in the same country and in different countries,

Table III
User Study Results

		score average	score of 4 or 5
Q1	same country pair	3.45	40.0%
	different country pair	3.90	75.0%
	average	3.67	58.5%
Q2	same country pair	3.75	65.0%
	different country pair	4.10	75.0%
	average	3.93	70.0%

i.e., bias appeared more clearly in the latter case. For both Questions 1 and 2, 75.0% of the scores were either 4 or 5. In addition, it can be said that bias is revealed even for article pairs prepared by news agencies in the same country because 65.0% of the scores for Question 2 were 4 or 5.

IV. Conclusion

We proposed a stakeholder mining mechanism and described how it can be applied to bias analysis. The experimental results obtained indicate that the stakeholder mining mechanism is helpful for discovering news articles' bias.

In future work we will attempt to identify identical entities that were referred to using different words. It will also be necessary to use lexical resources appropriately to the meanings of words according to context.

Acknowledgment

This research is partly supported by the research for the grant of Scientific Research (No.20700084, 20300036 and 20300042) made available by MEXT, Japan.

References

[1] S. Park, S. Kang, S. Chung, and J. Song, "Newscube: delivering multiple aspects of news to mitigate media bias," *Proc. of SIGCHI2009*, 2009, pp. 443–452.

[2] J. Liu and L. Birnbaum, "Localsavvy: aggregating local points of view about news issues," *Proc. of LOCWEB*, 2008, pp. 33–40.

[3] A. Nadamoto and K. Tanaka, "A comparative web browser (cwb) for browsing and comparing web pages," *Proc. of WWW2003*, 2003, pp. 727–735.

[4] Q. Ma and M. Yoshikawa, "Topic and viewpoint extraction for diversity and bias analysis of news contents," *Proc. of APWeb/WAIM2009*, 2009, pp. 150–161.

[5] G. A. Miller, "Wordnet: a lexical database for english," *Commununications of ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[6] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining," *Proc. of the 5th Conference on Language Resources and Evaluation*, 2006, pp. 417–422.

[7] M. C. De Marneffe and C. D. Manning, "The Stanford typed dependencies representation," *Proc. of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*, 2008, pp. 1–8.