

An Opinion Analysis System Using Domain-Specific Lexical Knowledge

Youngho Kim, Yuchul Jung, and Sung-Hyon Myaeng

Information and Communications University,
119, Moonji-ro, Yuseong-gu, Daejeon, 305-714, South Korea
{yhkim, enthusia77, myaeng}@icu.ac.kr

Abstract. In this paper, we describe an opinion analysis system using domain-specific lexical knowledge in Korean economic news. We tested our hypothesis that such domain-specific knowledge helps enhancing the performance of statistically based approaches and obtained a promising result.

1 Introduction

Gathering opinions about a specific subject is important in many areas such as governments to improve their services [1]. Since people often express their thoughts on articles, newspaper is one of the good resources where opinions of various sorts are found. Especially, in the economy domain, opinions about economic events flood into news. Sentiment information (e.g., negative or positive tendency) can be a reflection of people's opinions for a specific subject. Our research is motivated by this and centered around sentiment analysis of news articles on economy. More precisely, we attempt to build a system that determines the polarity of sentiment and its strength.

Previously, researchers have used a statistical learning method and semantically oriented *seed* terms as clues. Pang et al [2] adopted supervised machine learners to predict a document's semantic polarity. Ku et al [3] attempted to develop opinion extraction, summarization and tracking systems. Hatzivassiloglou [6] have attempted to predict semantic orientation of adjectives by analyzing adjective pairs. Turney [7] have bootstrapped from a seed set¹. Kamps et al [8] have focused on the use of lexical relations defined in WordNet. Esuli [5] proposed semi-supervised learning method started from expanding Turney's seed set. Bradley [4] tried psychological studies which have found measurable associations between words and human emotions.

Following the same line of thought, we take a novel approach that considers domain-specific lexical knowledge to complement generic methods entirely relying on statistical learning. In particular, we built an opinion analysis (OA) system for Korean economic news. Instead of constructing a costly deep knowledge base (KB) in the domain, we collected a set of domain-specific terms that indicate sentiments of an article in the domain. We postulate that domain knowledge is critical because a priori

¹ Seed set contains 7 positive words and 7 negative words as good, nice, excellent, positive, fortunate, correct, superior in positive set and bad, nasty, poor, negative, unfortunate, wrong, inferior in negative set.

knowledge on semantic orientation of domain-specific terms complements statistical learning methods. The KB contains sentiment information of economy terms such as “감자” (a reduction of capitals) and general sentiment clues such as “폭락” (falling).

In order to test our hypothesis that domain-specific lexicon should improve the performance of sentiment analysis, we evaluated our system using Korean economic news. The system consists of a KB that contains information about domain specific terms as well as domain independent terms that express sentiments, agent architecture for crawling news articles and determines their sentiments using the knowledge base, and an information retrieval (IR) system that enables end-users to capture the sentiment tendency of an interesting topic. In order to simulate an operational environment, for which this system was envisioned to begin with, about 170,000 news documents were collected initially. We ran experiments using topic-specific 200 articles of the collection to test our system’s effectiveness.

2 Constructing a Knowledge Base

In our system, the KB is a core according to our hypothesis. The lack of machine-readable knowledge in the economy domain drove us to gather sentiment knowledge manually. We employed 5 annotators, who were majoring in economics. They annotated semantic orientation and its strength for a given term, based on its perceived role in determining sentiments in the domain. They were instructed to make judgments in view of the Korean economy. For example, “부채 계정 (liability account)” is considered strongly negative since it has a negative effect on the Korean economy.

Our assumption is that some economy terms have polarity values and degrees of strengths. Several well-known dictionaries² in economy domain were the main source. In addition, non-economy terms (context terms) that influence the economy were collected. For example, “북한 핵 실험” (North Korea’s nuclear test) in news severely affects the economy as witnessed by North Korean’s announcement of their success in underground nuclear test last year, which shook South Korea’s stock market.

We also gathered words that frequently carry sentiment information in news because generally the sentiment often depends on an occurrence of *seed* words such as “비난” (denounce). These domain-independent words have some clues as to the polarity of text containing them. Our annotators extracted those terms from a randomly selected 300 sample articles from the collection of news articles collected during a month. All the collected terms were annotated, resulting in 13,564 dictionary terms, 620 context terms, and 176 general words. The annotation process was as follows. First, the annotator selected a term’s polarity: positive, negative, or neutral. Next, the annotator decided its strength if the term is polarized. If a term’s polarity is not obvious, the annotator was allowed to choose “neutral.” We designed the scale of 1 to 5 for polarity values (1 and 2 are weak, 3 is mild, 4 and 5 are strong). Naturally, we found many cases with disagreement among the annotators. For examples, “1월 효과” (January Effect) was judged as two positives, two negatives, and one neutral. For the cases with disagreements, we decided a term’s semantic orientation by majority

² Mae-il Economic terms dictionary, Economic dictionary by the bank of Korea, Dong-a Economic terms dictionary are freely open on the Internet.

(i.e., counting voting results and selecting the most one) and the term's strength by the average of the selected polarity voters' strength values. When two or less annotators viewed a term as either positive or negative, we set it to be a neutral term. Therefore, we gathered 4043 dictionary terms, 531 context terms, and 176 general words which are either a single word or a phrase consisting up to three words. Table 2 shows the final results: the number of positive and negative terms with different strengths.

Table 1. Statistics of all annotation results

Annotator	Positive			Negative			Neutral
	Weak	Mild	Strong	Weak	Mild	Strong	
A	672	186	159	388	276	262	12,417
B	803	403	407	266	284	390	11,807
C	1,112	317	295	368	213	300	11,666
D	1,008	240	255	455	290	316	11,796
E	977	304	244	328	219	182	12,106

Table 2. Gold-standard

	Positive	Negative
Weak	2,739	1,322
Moderate	119	209
Strong	151	210

3 System Architecture

Our system determines a news article's sentiment information for end-users who want to read an opinionated news article on a specific topic, sometime with an interest on a particular sentiment. Due to an enormous number of daily news data, swiftness is an essential in our system. So, firstly, an agent³ collects news documents from news wires. Next, another agent detects sentiment terms in KB and reacts to determine the sentiment of news in the background of the scoring formula, paragraph segmentation and morphological analysis⁴. After that, the news documents sentiment information are tagged are transferred to IR system [10]. As a result, news articles with sentiment information are viewed by IR system. Also, a document's sentiment value, v is determined in Figure 1. For each term (tc) in a document (D), the formula averages all the detected terms by each term's frequency ($f(tc)$) and each term's sentiment value ($s(tc)$). We added a bias for the most frequent term (tm) since frequent term is an important sentiment factor. (In experiment, β is 0.4, α is 0.6). Also, we promoted one step for sentiment words in the first paragraph as like "weakly positive" to "mildly positive", which reflects the writing style that topics are arranged at the top.

³ AgletTM agent (<http://www.trl.ibm.com/aglets/>)

⁴ Morphological analyzer developed from ETRI (Korea Electronics and Telecommunications Research Institute) extracts noun portion from raw text and puts the part-of-speech tag on it.

$$v = \alpha \cdot \frac{\sum_{tc \in D} s(tc) \cdot f(tc)}{\sum_{tc \in D} f(tc)} + \beta \cdot tm, \text{ where } \alpha + \beta = 1$$

Fig. 1. Document Sentiment Value Determination Function

4 Experimental Results

Our collection covers 21 different newswire sources from July of 2005 to November of 2006. Among them, we tested the documents retrieved within top-200 based on two queries, “정부 부동산 정책” (government’s real estate policy) and “정몽준 현대 자동차 회장 구속” (arrest of Jung Mong Jun, President of Hyundai automobile company). Our goal of the experiments was testing two hypotheses: 1) our approach can complement the statistically based approach to OA; 2) our KB containing domain-specific lexical knowledge is useful in determining the polarity and the strength of news article. We assumed an information seeking scenario where a user wanting to get a feeling about the most sensational news in economy enters a query like the above query and reads the retrieved news articles. First, we measured the accuracy of the polarity determination function using three classes: negative, positive, and neutral. Second, we examined the sentiment strength determination function in terms of accuracy and the Mean-Squared Error (MSE) that can capture the problem of the huge distinction between the true strength and the predicted strength. If t_i is the true strength of document i , and p_i is the predicted strength of document i ,

$$MSE = \frac{1}{n} \cdot \sum_i^n (t_i - p_i)^2$$

where n is the number of documents in the test set. We obtained the results of strengths at the 3-step level (i.e., weak, mild, and strong) and 5-step level (i.e., 1~5).

For the gold-standard, we decided the polarities and their strength of the test articles by majority as used in Section 3. To verify our system’s competitiveness over statistical learning methods, we produced results of SVM and Naïve Bayes-based classifiers as the baseline. Those classifiers using uni-gram feature were trained by another 100 retrieved documents with the same queries (i.e., we searched another 100 samples and annotated them in the same manner as in the test set). Pang et al [2] report that SVM has the best performance in English text. In spite of the belief that adjective is more effective contributor than other parts-of-speech in classifying English documents based on their sentiment values [5, 6], noun is a much more critical feature in Korean text (Table 4). We used Joachim’s SVM^{light} package [9].

We obtained 74% accuracy which is 9% increase over the better baseline performance and 30% increase over the case where only the general KB was used. Since the terms in KB are only nouns, adjective was not encountered in the cases of KB. We observed, at least in Korean news, that domain specific terms play a more important role in determining the sentiment because the context of words provide additional information in capturing more precise sentiment. For example, general words such as “하락(decrease)” that may be considered as negative in many context have different meanings depending on its context. For example, “분양 가격 하락 (decrease of the

Table 4. Accuracy results of polarity

Features	Naïve Bayes	SVM	KB with domain knowledge	KB with only general knowledge
Adjective	51%	58%	N/A	N/A
Noun	59%	68%	74%	57%
Adjective + Noun	58%	68%	N/A	N/A

Table 5. Polarity determination results for different annotators

Methods	A	B	C	D	E
KB with domain knowledge	68%	69%	69%	74%	73%
KB with only general knowledge	58%	51%	62%	53%	50%

Table 6. Results of determination of polarity strengths

Methods	Accuracy		MSE	
	3 step	5 step	3 step	5 step
KB with domain knowledge	63%	33%	1.40	2.83
KB with only general knowledge	51%	21%	2.73	4.37

selling price of an apartment)” is positive, but “부동산 대출금리 하락 (decrease of the interest for real-estate loans)” is negative since it may cause land speculation. To resolve certain variation, we need domain knowledge.

Our approach is more effective across all the annotation results although there were quite a wide range of variations. The results in the table reflect the diversity compared to the gold-standard. However, our approach is shown to outperform the one with general term knowledge across all the cases, varying from 24% to more than 50%.

5 Conclusions

We proposed a knowledge-based OA system that does not rely on heavy natural language processing to figure out the semantics or pragmatics of text, which is not really feasible for the given task at this point. Our approach is practical since the knowledge base contains lexical information, i.e. terms and their polarity information in a specific domain like economy, which can be captured with relative ease. The machine-readable dictionary we developed contains not only domain specific terms, but also domain-independent opinion clue terms. We verified that our approach is effective

and promising in an experiment by showing that the proposed system determines the polarity and the strength more accurately than statistically based machine learning methods and the method with only general seed words. While the experiment was done with Korean news articles, the result is valuable since most popular statistical machine learning approaches and the approach with knowledge on general, domain-independent terms only have clear limitations that should be overcome. In the future, a more detailed analysis of the reasons why machine learning based approaches are inferior would be of great value.

Acknowledgements

This work was supported by KICOS through a grant by Ministry of Science & Technology (K20711000007-07A0100-00710) and 2nd Phase of Brain Korea 21 project sponsored by Ministry of Education and Human Resources Development, Korea.

References

1. Dave, K., Lawrence, S., Pennock, D.M.: Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In: Proceedings of the 12th International Conference on the World Wide Web, Budapest, Hungary, ACM Press, New York, US (2003)
2. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment Classification using Machine Learning Techniques. In: Proceedings of the 7th Conference on Empirical Methods in Natural Language Processing, Philadelphia, US, Association for Computational Linguistics, Morristown, US (2002)
3. Ku, L., Liang, Y., Chen, H.: Opinion Extraction, Summarization and Tracking in News and Blog Corpora. In: Proceedings of AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs, AAAI Technical Report SS-06-03, Stanford University, California, US (2006)
4. Bradley, M.M., Lang, P.J.: Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Technical Report C-1, The Center for research in Psychophysiology, University of Florida, Gainesville, Florida, US (1999)
5. Esuli, A., Sebastiani, F.: Determining the Semantic Orientation of Terms through Gloss Classification. In: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, ACM Press, New York, US (2005)
6. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, Madrid, Spain, Association for Computational Linguistics (1997)
7. Turney, P.D., Littman, M.L.: Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems* (2003)
8. Kamps, J., Marx, M., Mokken, R.J., Rijke, M.D.: Using WordNet to measure semantic orientation of adjectives. In: Proceedings of the 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal (2004)
9. Joachims, T.: Making large-scale SVM learning practical. In: Scholkopf, B., Burges, C., Smola, A. (eds.) *Advances in Kernel Methods – Support Vector Learning*, MIT Press, Cambridge (1999)
10. Hatcher, E., Gospodnetic, O.: *Lucene IN ACTION*. Manning Publications Co. (2004)