Multi-lingual Opinion Analysis Applied to World News: A Case Study

David Kirk Evans
devans@nii.ac.jp
National Institute of Informatics
Tokyo, Japan

Noriko Kando
noriko.kando@nii.ac.jp
National Institute of Informatics
Tokyo, 101-8430, Japan

Abstract

In this paper, I present an analysis of one of the topics from the NTCIR-6 Opinion Analaysis Pilot Task, a multi-lingual workshop using data from Chinese, Japanese, and English documents, and develop a novel method for summarizing the results of opinion analysis of a topic in multiple languages. I present an analysis of words correlated to opinionated sentences specific to the analyzed topic as compared to general text, and an analysis of the types of opinions found and opinion holders across languages for the given topic. I also present plans for combining opinion analysis with summarization.

1 Introduction

Opinion and sentiment analysis has been receiving a lot of attention in the natural language processing research community recently. With the broad range of information sources available on the web, and rapid increase in the uptake of social community-oriented websites that foster user-generated content there has been further interest by both commercial and governmental parties in trying to automatically analyze and monitor the tide of prevalent attitudes on the web. As a result, interest in automatically detecting language in which an opinion is expressed, the polarity of the expression, targets, and opinion holders has been receiving more attention in the research community. Applications include tracking response to and opinions about commercial products, governmental policies, tracking blog entries for potential political scandals and so on.

In the Sixth NTCIR Workshop to be held in Tokyo, May 2007, a new pilot task for Opinion Analysis has been introduced. The pilot task has tracks in three languages: Chinese, English, and Japanese. In this paper, I examine the manual annotation results for one topic in the pilot task data-set, present a rough comparison of the Japanese, Chinese, and English data, and present an outline of how I plan to use opinion analysis for multilingual news summarization.

Analysis Task	Values	Req'd?
Opinionated Sentences	YES, NO	Yes
Opinion Holders	String, multiple	Yes
Relevant Sentences	YES, NO	No
Opinionated Polarities	POS, NEG, NEU	No
Application-oriented	Use annotations	No

Table 1: Opinion Analysis task descriptions

1.1 NTCIR6 Opinion Analysis Task

The NTCIR-6 Opinion Analysis Pilot Task extends previous work in opinion analysis to a multilingual corpus. The initial task focuses on a simplified sentence-level binary opinionated or not opinionated classification as opposed to more complicated contextual formulations, but we feel that starting with a simpler task will allow for wider participation from groups that may not have existing experience in opinion analysis.

The Opinion Analysis task has four subtasks, two of which are mandatory and two of which are optional. Table 1 summarizes the tasks, which are all being performed for all three languages. The two mandatory tasks are to decide whether each sentence expresses an opinion or not. For the Chinese data, all potential opinion holders are annotated whether the sentence in which the entity occurs is an opinionated sentence or not. In Japanese and English, opinion holders are only annotated for sentences that express an opinion, however, the opinion holder for a sentence can occur anywhere in the document. The annotators performed a kind of reference resolution by marking the opinion holder for the sentence, and if the opinion holder is an anaphoric reference noting the target of the anaphora. The opinionated sentences judgement is a binary decision, but in the case of opinion holders we allow for multiple opinion holders to be recorded for each sentence in the case that multiple opinions are expressed.

The two optional tasks are to decide the polarity of the opinionated sentences, and whether the sentences are relevant to the set topic or not. Each set contains documents that were found to be relevant to a particular topic, such as the one shown in Figure 1.

Language	Topics	Documents	Sentences
Chinese	28	2103	27741
English	28	439	8370
Japanese	26	422	12525

Table 2: General information about NTCIR6 Opinion Analysis Data Set

For those participating in the relevance subtask each sentence should be judged as either relevant (Y) or non-relevant (N) to the topic. Polarity is determined for each opinionated sentence, and for sentences where more than one opinion is expressed the annotators were instructed to determine the polarity of the most main opinion expressed. In addition, the polarity is to be determined with respect to the set topic description if the sentence is relevant to the topic, and based on the attitude of the opinion if the sentence is not relevant to the topic.

The Application-Oriented task is separate from the Opinion Analysis task, and allows for groups that would like to build natural language processing applications on top of opinion analysis systems. We provide participants in this track with the human-annotated gold-standard results which they can use to enrich existing applications with opinion information, build prototype systems using the data to judge feasibility of some task, evaluate systems that use opinion analysis by comparing performance using human annotations to automatic systems, and so on. The Application-Oriented track allows for a wide range of exploratory research over opinion analysis data.

2 Topic Analysis

This paper focuses on an analysis of topic 010, "History Textbook Controversies, World War II". The description, background, and relevant documents fields of the topic are shown in Figure 1.

Table 3 shows some general information about Topic 010. The Chinese data generally has more documents annotated than the English and Japanese portions of the data. For both Japanese and English, for topics with more than twenty documents, only twenty documents were selected for annotation.

Table 4 lists some opinion holders from topic 010 from each language. For both English and Japanese for the top opinion holders I manually combined counts for some of the holders, performing a kind of named entity conflation. I did not perform this for Chinese, due to unfamiliarity with the language. There are a total of 299 opinion holders in English, 764 in Japanese, and 425 in Chinese with on average of 1.5 mentions per entity in English, 2.7 in Japanese, and 2.5 in Chinese.

In all three languages the largest source of opinions is unsurprisingly the document author, but following that it is interesting that the Japanese documents feature more Japanese opinion holders, the Chinese documents feature more Chinese holders, and the English data is split between Japanese and Chinese figures. Since there is little overlap between the top opinion holders in the different languages, it is difficult to draw any comparative conclusions about the polarity data. The English authors tend to be the most negative, with a large number of positive opinions from the Chinese authors. A further step is to identify the target of the expressed opinion, but for this pilot task that was not performed, although there are plans to identify the target in a future run of the opinion analysis task. In both Japanese and English, Korean and Chinese government officials predictably present mostly negative and neutral opinions.

Count	Holder	P/Neg/Neu
60	(English) Author	1/17/32
21	South Korean Gov.	0/10/11
20	Zhu Bangzao	0/2/18
14	History Textbook Group	2/7/5
7	South Korean Legislators	0/1/6
4	Japanese Ministry of Education	0/2/2
127	(Japanese) Author	28/54/42
37	Korea / Korean Government	1/18/19
30	Hata Ikuhiko (Prof.)	11/15/4
26	Takamori Akinori	13/5/8
23	Tanaka Toshiaki (Prof.)	3/4/16
11	China / Chinese Government	1/6/4
41	(Chinese) He	20/14/7
24	Koizumi Junichiro	11/9/4
13	Ministry of Foreign Affairs	8/4/1
11	Hsieh, Chi-ta	1/3/7
9	Chu, Te-Lan	1/1/7
8	Lo, Fu-chen	4/4/0

Table 4: Top Opinion Holders for each language

3 Opinion Summary Terms

I am interested in discovering terms that are interesting or important for the opinionated text in the document set. Based on work in corpus linguistics for comparing corpora (see [5] for a nice overview) and research in technical terminology identification, I decided to use the look at word-based mutual information and log liklihood scores to sort words from the opinionated sentences in the topic compared to the non-opinionated sentences across all topics.

The equation $MI_{w,A} = log_2\left(\frac{|A_w|}{|A|} \times \frac{|A|+|B|}{|A_w|+|B_w|}\right)$ computes the mutual information [2] score for a word, which gives higher scores for words that are more strongly associated with corpus A than corpus B. In

Find reports on the controversial history textbook about the Second World War approved by the Japanese Ministry of Education. The Japanese Ministry of Education approved a controversial high school history textbook that allegedly glosses over Japan's atrocities during World War Two such as the Nanjing Massacre, the use of millions of Asia women as "comfort women" and the history of the annexations and colonization before the war. It was condemned by other Asian nations and Japan was asked to revise this textbook. Reports on the fact that the Japanese Ministry of Education approved the history textbook or its content are relevant. Reports on reflections or reactions to this issue around the world are partially relevant. Content on victims, "comfort women", or Nanjing Massacre or other wars and colonization are irrelevant. Reports on the reflections and reactions of the Japanese government and people are also irrelevant.

Figure 1: Topic title, description, and relevance fields for set 010

Language	Docs	Sents	POS	NEU	NEG	Relevant
Chinese	41	1641	198 (12%)	199 (12%)	528 (32%)	966 (58.9%)
English	20	1829	8 (0.4%)	57 (3.1%)	224 (12.2%)	359 (19.6%)
Japanese	20	2358	149 (6.3%)	148 (6.3%)	319 (13.5%)	1269 (53.8%)

Table 3: General statistics over Topic 010

my implementation, I only considered words that occurred more than five times between the corpora. The mutual information score is thought to overemphasize rare terms. In addition, I implemented the log liklihood statistic [4] as: $G^2 = 2(alog(a) + blog(b) + clog(c) + dlog(d) - (a+b)log(a+b) - (a+c)log(a+c) - (b+d)log(b+d) - (c+d)log(c+d) + (a+b+c+d)log(a+b+c+d))$

The log liklihood statistic measures how surprising an event is even when there are very low occurrences of the event, as is often the case with words that appear infrequently.

Table 5 shows a list of the top ten words with high log likelihood and mutual information scores for words from opinionated sentences for topic 010 when compared to all non-opinionated words. I did not perform any stemming or morphological analysis on the text, so both "textbook" and "textbooks" show up separately. The two different measures identify very different types of words, with mutual information terms being more rare terms, but which still seem related with terms central to the controversy. "Invaders", "denigration", and "tragedies" possibly relate to the historic account events that is controversial, while "blurs" and "biased" are criticisms that were heavily leveled against the new textbook. The other words express some of the activism popular in Korea and China protesting the textbook.

I ran the same statistical measures over the Japanese text, although slightly different steps were taken for data processing. Opinionated and non-opinionated sentences were extracted as in English, then the sentences were analyzed with the ChaSen [1] morphological analysis system. I removed punctuation and kept each analysis entry as a separate "word" for counting. A more sophisticated approach would filter out syntactical markers such as particles and make informed use

Log Likelihood	Mutual Information
textbook	invaders
history	denigration
Japanese	blurs
textbooks	biased
facts	Stage
Japan	Rally
Asian	Netizens
draft	Cyber
descriptions	militarists
distorted	tragedies

Table 5: Top 10 English opinionated words for topic 010 for Log Likelihood and Mutual Information statistics.

Log Likelihood	Mutual Information
Textbook	faithful
History	rash, thoughtless
Official approval	strange
revision	unfair? (ikou)
(negation)	cultural progress
Korea	permission
1	deception
description	Tokyo City U.
Tsukuru	Yamazumi
glorification	fear, misgivings

Table 6: Top 10 Japanese opinionated words for topic 010 for Log Likelihood and Mutual Information statistics.

of the verb conjugation information, but I was interested to see if the non-content bearing tokens would follow similar usage characteristics in opinionated and non-opinionated text.

Table 6 lists the top ten opinion words for the Japanese text for the mutual information and log likelihood scores. The log likelihood terms are quite similar to the terms from English, with the exception of the regional descriptor "Asia" which isn't necessary to orient readers. The fifth entry, (negation) is the Japanese conjugation of the negative form (nai). Other morpho-syntactic tokens that I expected might show up (case markers such as wo, ha, ga, ha, etc.) had similar usage across the corpora, but opinionated text made more usage of negation, although I did not track what is being negated. "Tsukuru" is interesting because it is from the "New Textbook Creation Committee", often written as Tsukuru-kai with furigana. For Mutual Information, entry four, "ikou" is in hiragana, so I can only guess that it might be "unfair". In the 8th and 9th entries, Yamazumi refers to Yamazumi Masaki, a professor from Tokyo City University that figures prominently in the debate (but appears as an opinion holder only four times.)

I was unable to generate a table of words from the Chinese text due primarily to unfamiliarity with Chinese language processing tools and time constraints.

3.1 Application to Summarization

Recent summarization evaluations such as the Document Understanding Conference [3] and 2006 Multilingual Summarization Evaluation generally have focused on full sentence extraction and revision via paraphrasing or syntactic simplification. While there have been recent trends towards query and opinion focused summarization [6], the multi-lingual corpus used in this pilot task offers a new opportunity to look at multilingual opinion-focused summarization.

An immediately interesting question to ask about opinionated multi-lingual document sets is "how do these documents differ across languages?" The terms presented in Section 3 are a first approach at answering that question. The same methods used in this paper to identify interesting terms can be applied to new document sets given the existence of a system that can automatically label a sentence as opinionated or not opinionated. These systems exist for Chinese, English, and Japanese, and we plan to continue evaluation to improve performance in future NTCIR Opinion Analysis tasks.

A summary can then generated for an opinionated document set by identifying opinion terms for each language, presenting them as a concise summary of the document set, which can be supplemented with sentence-extraction based summaries in the language of the users' preference. A benefit of using simple extracted terms from each language to characterize the document set is that a short list of terms can be quickly automatically translated, and make for an interesting contrast with the terms from the other languages. While many researchers worry about presenting machine translation output of full sentences to users for use in summaries, a list of terms is more robust to translation errors and more likely to be adopted by end users who would not accept simple syntactic mistakes in a full sentence. Combined with a brief list of the major opinion holders and in the future targets of opinions and polarity, multi-lingual summaries can mix both free text summaries and more constrained types of information that are more easily translated across languages.

4 Conclusion

In this paper I introduced the NTCIR-6 Opinion Analysis Task, gave a brief overview of the Chinese, English, and Japanese documents in the corpus, and presented an analysis of one of the topics. I presented opinionated terms found using statistical techniques in both English and Japanese. I also presented an approach to multilingual summarization using term translation, traditional sentence-extraction based methods, and opinion holder identification that can be used to contrast differences between the three languages.

References

[1] M. Asahara and Y. Matsumoto. Extended models and tools for high-performance part-of-speech tagger. In *Proceedings of COLING 2000*, July 2000.

[2] K. W. Church and P. Hanks. Word association norms, mutual information, and lexicography. In *Proceedings of the 27th. Annual Meeting of the Association for Computational Linguistics*, pages 76–83, Vancouver, B.C., 1989. Association for Computational Linguistics.

[3] H. Dang. Overview of duc 2005. In *Document Understanding Conference (DUC)*, Boston, MA, May 2005.

[4] T. Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74, 1994.

[5] A. Kilgarriff. Comparing corpora. *International Journal of Corpus Linguistics*, 6(1):97–133(37), 2001.

[6] Y. Seki, K. Eguchi, N. Kando, and M. Aono. Opinion-focused summarization and its analysis at duc 2006. In *Document Understanding Conference (DUC)*, Brooklyn, New York, June 2006.