Towards an Analysis of Opinions in News Editorials: How positive was the year?

Bal Krishna Bal

Madan Puraskar Pustakalaya, Lalitpur, PatanDhoka, Nepal
and IRIT, 118 route de Narbonne, 31062 Toulouse, France

bal@mpp.org.np

1   Introduction

Editorials represent opinion articles written by the publisher, editors or columnists of newspapers. From this perspective, editorials are ideal sources for outlining views on events and analyzing how they are perceived, e.g. whether they are positive or negative or the kind of feelings or analysis they involve (uncertainty, worries, etc.). The proposed work aims at identifying the linguistic criteria and at creating a model for the aforementioned purpose, which makes use of an adequate set of semantic tags that we have defined to annotate texts from which a detailed analysis can be carried out.

Currently, the work is in its preliminary stage, primarily focussed on defining the tags required for the semantic-pragmatic analysis. In parallel with the definition of tags and text annotation, we explore ways to construct synthesis of opinions on a given event from various editorials. One of the challenges is to organize the positive and the negative views, and the associated arguments and their strength. We noted that opinions in editorials are not so apparent, which makes synthesis construction a challenging task.

2   Opinion Mining as a problem

Although Opinion Mining has emerged only quite recently as a subdiscipline under computational linguistics, a considerable amount of work has already been done in this direction, which include but not limit to (Hearst,1992), (Hatzivassiloglou and Wiebe,2000), (Wilson et al., 2005), (Esuli and Sebastiani, 2006) and (Read et al., 2007). These works range from a variety of task domains like mining the product reviews available on the web, sentiment classification of documents, opinion mining and summarization to

much more. Irrespective of the nature of different specific tasks, Opinion Mining generally encompasses the following generic problems: (1) Determining the subjectivity or identifying the subjective and objective expressions in texts, where subjective expressions are opinions whereas objective ones denote facts; (2) Determining the orientation or polarity of the subjective expressions. A first analysis may be just identifying whether an expression is positive or negative, but in general other dimensions need to be integrated since opinions are rarely so vivid; (3) Determining the strength of the orientation of the subjective expressions.

Our problem of mining editorials falls under the larger task domain of sentiment classification of documents and this essentially involves the problems 1 and 2 above. Problem 3 also may be partially applicable depending upon the type of analysis we would like to have as an end result.

3   Linguistic framework and the distinction between facts and opinions

Since editorials are usually a mix of facts and opinions, there is a clear need to make a distinction between them. Opinions often express an attitude towards something. This can be a judgment, a view or a conclusion or even an opinion about opinion(s). Different approaches have been suggested to distinguish facts from opinions. Generally, facts are characteristic for the presence of certain verbs like "declare" and different tense and number forms of the verb "be" etc. Moreover, statements interpreted as facts are generally accompanied by some reliable authority providing the evidence of the claim, e.g.:

Fact: Both the two dates announced for the constituent assembly (CA) elections came and went without the vote taking place.
Reliable authority: *Election Commission for CA elections 2007.*
Fact: We have fewer people getting killed every day.
Reliable authority *Nepal Police Department of Crime and Investigation. (December 2007)*

Opinions, on the other hand, are characterized by the evaluative expressions of various sorts such as the following (Dunworth, 2008):
- Presence of evaluative adverbs and adjectives in sentences - "ugly" and "disgusting".
- Expressions denoting doubt and probability - "may be", "possibly", "probably", "perhaps", "may", "could" etc.

```
<expression expression_type="Fact" opinion_type="Null" opinion_orientation="Null" force="Null" opin-
ion_date="Null" authority="Yes">
This was the year Nepal declared itself a Federal Democratic Republic
</expression>
<expression    expression_type="Opinion"    opinion_type="Assumptive"    opinion_orientation="Positive"
force="Low" opinion_date="2007-12-31" authority="Null">,
but the decision needs to be endorsed which will probably happen in 2008.
</expression> </input>
```

- Presence of epistemic expressions - "I think", "I believe", "I feel", "In my opinion" etc.

It is obvious that the distinction between the two is not always straightforward. Facts could well be opinions in disguise and, in such cases, the intention of the author as well as the reliability of information needs to be verified. In order to make a finer distinction between facts and opinions and within opinions themselves, opinions are proposed for gradation as shown below:

Opinion type	Global definition
Hypothesis statements	Explains an observation.
Theory statements	Widely believed explanation
Assumptive statements	Improvable predictions.
Value statements	Claims based on personal beliefs.
Exaggerated statements	Intended to sway readers.
Attitude statements	Based on implied belief system.

4   Results

For the purpose of annotation, the following semantic tag set has been developed, subject to further extension or modification in the future. The current annotation scheme can be represented as a list of five parameters and their possible values as shown below:

Parameter	Possible values
expression_type	Fact, Opinion, Undefined
opinion_orientation	Positive, Negative, Neutral
opinion_date	Date of the editorial publication
force	Low, Average, High
authority	Yes, Average, No

Here, 'force' refers to the strength of the orientation of the opinion. Any textual expression subject to annotation would go within <expression> </expression>. A non-applicable parameter for a particular type of expression would receive a "Null" value. The annotation of text fragments has been guided by the presence of evaluative expressions and other criteria as explained in section III.

For annotation purposes, we have collected the editorials from two online newspapers (http://ekantipur.com,http://kantipuronline.com/ktmpost.php) of different dates of the year 2007, amounting to a total of 16 text files and approximately 320 sentences with an average of 20 sentences per editorial. Two annotators having a fairly good understanding of the English

language have been involved in the annotation work. The annotators have been assigned the same texts to see how semantic annotations can differ among annotators. Results have shown that the difficulties in the manual annotation exist at two levels, the first one in determining an expression as being a fact or an opinion and the other one in grading the opinion as one of the types, i.e., assumptive or value etc. Wherever the annotators have confusions about providing one particular value, they have been advised to provide multiple values separated by commas. A sample of the annotated text in XML format is given in Fig. 1. above.

5   Conclusion

The work that we have described is currently in the phase of manually annotating samples of editorials being based on the semantic tag set and methodology discussed in the document. Further, we plan to develop a computational model which would suggest methods to automate the process of analyzing and synthesizing opinions from editorials. The manually annotated texts and collected editorials would serve as training data and test data respectively for validating the proposed computational model.

References

[1] Hearst M. A., Direction-based text interpretation as an information access refinement, Text-based intelligent systems: current research and practice in information extraction and retrieval, 1992, Erlbaum.

[2] Hatzivassiloglou V. and Wiebe J. M., Effects of adjective orientation and gradability on sentence subjectivity, Proc. 18th Conference on Computational Linguistics, 2000, Saarbrücken.

[3] Wilson T. et al., Recognizing contextual polarity in phrase-level sentiment analysis, Proc. HLT/ EMNLP 2005, Vancouver.

[4] Esuli A. and Sebastiani F. , Determining term subjectivity and term orientation for opinion mining, Proc. EACL-06, Trento.

[5] Read J.et al., Annotating Expressions of Appraisal in English, Proc. ACL 2007 Linguistic Annotation Workshop, Prague.

[6] Dunworth K., UniEnglish reading: distinguishing facts from opinions, World Wide Web, 2008, http://unienglish.curtin.edu.au/local/docs/RW_facts_opinions.pdf