Proceedings of the International Workshop on Intelligent Information Systems
IIS'2011, September 13-14, 2011, Chisinau, Republic of Moldova

Sentiments and Opinions Annotation in User-generated Texts

Victoria Bobicev

Abstract: In this work, we analyze sentiments and opinions expressed in user-written Web messages. The messages discuss health related topics: medications, treatment, illness and cure, etc. Recognition of sentiments and opinions is a challenging task for humans as well as an automated text analysis. The paper presents the annotation model, discusses characteristics of subjectivity annotations in health-related messages, and reports the results of the annotation agreement.

Keywords: Natural Language Processing, text annotation, inter-annotation agreement, opinion mining, sentiment analysis.

1 Introduction

In recent years, Text Data Mining (TDM) and Natural Language Processing (NLP) intensively studied sentiments and opinions in user-written Web texts (e.g., tweets, blogs, messages). Researchers analyzed sentiments and opinions that appear in consumer-written product reviews, financial blogs, political discussions [1], [2], [3].

The goal of this work is to study sentiments and opinions in health-related Web messages. We start with building a data set of annotated sentences. We present an opinion and sentiment annotation scheme and its application to tag sentences harvested from the Web messages. We report evaluation of manual annotation agreement.

2 Opinion and Sentiment Annotation

We are interested in the expressions of user *private state* which is not open to objective observation or verification. These personal views are revealed through thoughts, perceptions and other subjective expressions that can be found in text [6]. We assume that the private states can be revealed by emotional statements, *sentiments*, and subjective statements that may not imply emotions, *opinions*. In this work, statements are considered within the sentence bounds; thus, sentences are the units of our language analysis. We agree with [1] that opinion can be expressed about a fact of matter, and should not be treated as identical to sentimental expression. We further sub-categorize sentiments into *positive* and *negative*, opinions – into *positive*, *negative* and *neutral*. Sentences that do not bear opinions or sentiments are considered objective by default.

Annotation of subjectivity can be centered either on perception of a reader/annotator [5] or the author of a text [4]. Our model is author-centric. We requested that annotators do not impose their own sentiments and attitudes towards information in the text. Instead we suggested that an annotator imagined sentiments and attitudes that the author possibly had while writing.

For example, "I don't know if that makes sense, it seems to me that the new drug which stimulates red blood cell production would be a more logical approach, erythropoiten (sp?)" exposes the author's thoughts and ideas. It should be annotated as an opinion though without an emotional attitude. Another example, "Alas, I didn't record the program, but wish I had" expresses the author's regret and should be annotated as a negative opinion about the action (i.e., not recording the program).

Our annotation schema is based on the following assumptions: (a) annotation was performed on a sentence level; one sentence expressed only one assertion; this assumption held in a majority of cases; (b) only author's subjective comments were marked as such; if the author conveyed opinions or sentiments of others, we did not mark it as subjective as the author was not the holder of these opinions or sentiments; (c) we did not differentiate between the objects of comments; author's attitude towards a situation, an event, a person or an object were considered equally important.

3 Annotation process description

For the annotation we used the sci.med texts of *20 Newsgroups* (http://kdd.ics.uci.edu/databases/20newsgroups/20newsgroups.html). It is a benchmark data set of 20,000 messages, popular in applications of machine learning techniques, such as text classification and text clustering. There are 1000 sci.med messages.

To group messages by their content, we merged the messages with the same topic. A script automatically placed all messages with the same Subject line in the file with the same title. Thus, we obtained 365 files named "Arrhythmia", "arthritis and diabetes", "Athletes Heart", etc. Finally, 357 files were left for the annotation.

10 undergraduate and 10 master students were involved in the process. A master student had 30 files to annotate. The results of the annotation were examined; students with better annotations received more files. An undergraduate student had 10 files to annotate; again, only students with the satisfactory quality annotations were given more files. Finally, all the 357 files have been annotated by at least one annotator. 216 have been tagged by two annotators, and 21 have been tagged by three annotators. 120 files have been tagged by only one annotator. Thus we could compare the annotation for 237 files.

We have divided the results of comparison into 3 categories: subjective sentences: both annotators identified them as subjective, sentiment or opinion, and marked either the same polarity or neutral; weak subjective sentences: only one annotator identified them as subjective; non-subjective and uncertain sentences: sentences that both annotators did not mark as subjective and sentences marked with the opposite polarity.

4 Results and Discussion

6408 sentences were annotated in total. The majority – 4190 sentences – were considered nonsubjective*:* by both annotators. *Neutral opinion* was the most frequent subjective label, some persons asked questions and some replied in many cases expressing their own opinions. 85 sentences were marked *neutral opinion* by both annotators. In 655 cases, it was a weak subjectivity (i.e., identified by one annotator). The latter set contained ambiguous sentences, without clear indicators was the expressed statement author’s thought or just information taken from some sources. We report some examples: “Symptoms can be drastically enhanced by food but not inflammation”, “The low residue diet is appropriate for you if you still have obstructions”, “Then they may be able to crowd out garbage genes” *Negative sentiment* was another large set of the ambiguous annotation. Often *negative sentiment* was attributed to sentences that were interpreted as subjective only in the message context. For example, “I said that I PERSONALLY had other people order the EXACT SAME FOOD at TWO DIFFERENT TIMES from the SAME RESTAURANT” was marked *negative sentiment* in context of a very opinionated discussion. For the annotator, it was clear that the author of the text had been really angry, and the sentence did carry negative emotion even if it did not contain indicative words.

We have found that sarcasm was a strong factor for the polarity disagreement between annotators. “I’m forever in your debt” was marked as *positive sentiment* and *negative sentiment*, because it was positive as is but was used in a sarcastic answer to another message; one annotator took the whole context in consideration but another one did not. Perhaps, a more complex set of sentiment annotation tags can help to capture such sentiments.

5 Conclusion

In this paper, we have presented a study of sentiments and opinions in user-written Web messages. We focused on messages posted on health discussion boards. For the annotation we have designed an author-centric annotation model. The model shows how positive and negative sentiments and positive, negative and neutral opinions can be identified in text.

We applied the annotation model to the *sci.med* messages of *20 NewsGroups*. The results show that annotators better identify sentiments than opinions and stronger agree on what type of sentences do *not* belong to positive or negative subjective categories.

Our future plans are to continue the annotation; the final aim is to have all texts annotated by at least five persons.

References

[1] Kim, S.-M., and E. Hovy. Crystal: Analyzing predictive opinions on the web. *EMNLP-CoNLL* (2007), pp. 1056–1064.

[2] Blitzer, J., M. Dredze, and F. Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. *45th Annual Meeting of the ACL* (2007), pp. 440–447.

[3] Ferguson, P., O’Hare, N., Davy, M., Bermingham, A., Tattersall, S., Sheridan, P., Gurrin, C., Smeaton, A. Exploring the use of Paragraph-level Annotations for Sentiment Analysis of Financial Blogs. *WOMAS 2009*.

[4] Balahur, A. and R. Steinberger Rethinking Sentiment Analysis in the News: from Theory to Practice and back. *1st Workshop on Opinion Mining and Sentiment Analysis*, 2009.

[5] C. Strapparava, R. Mihalcea. Learning to Identify Emotions in Text, ACM-SAC 2008.

[6] Wiebe, J. Tracking point of view in narrative. *Computational Linguistics*, 20 (1994), pp. 233–287.

Victoria Bobicev
Technical University of Moldova, E-mail: victoria_bobicev@rol.md