Sentiment-Oriented Summarisation of Peer Reviews

Sunghwan Mac Kim and Rafael A. Calvo

School of Electrical and Information Engineering, University of Sydney
{sunghwan.kim,rafael.calvo}@sydney.edu.au

Abstract. It is common that students peer-review other students' writing, and these reviews are useful information to instructors, both on the particulars of the essay being reviewed, the feedback provided and the overall progress of the class. This paper describes a novel approach to summarising feedback in academic essay writing. We present a summarisation method for identifying and extracting representative opinion sentences from each feedback. Sentiment score-based techniques are employed and SentiWordNet is used as a linguistic lexical resource for sentiment summarisation. We evaluate our approach with the reviews written by a group of 50 engineering students.

Keywords: sentiment summarisation, peer review, student feedback.

1   Introduction

Online reviews are becoming increasingly significant on the Internet. Users create and rate significant amounts of content with useful information for academic and commercial purposes. Within the fields trying to automatically make sense of this content, sentiment (or opinion) summarisation has become a growing research topic.

Most existing research has been based on movie or product reviews. However, sentiment summarisation would be useful in other situations. For example, sentiment summaries can be used in students' feedback to interpret the rationale behind an evaluation. This valuable information can, for example, help a university lecturer obtain a more precise understanding of the[1] feedback for his or her lecture. Peer feedback of writing tasks is helpful to discover positive and negative features linked to the quality of the document [1].

In this study, we concentrate on a specific domain, which is peer review on student essay. Providing feedback to students is crucial to the learning process during the process of writing. Automatically computer generated summary on peer review can effectively enhance student's learning in academic writing. Furthermore, the highlighted feedback encourages the students to engage in the next writing and to make more meaningful changes on their work. To our best knowledge, there is not any research on summarising sentiments or opinions under the context of peer review. We suggest summarisation in terms of sentiment in order to provide more meaningful feedback to students.

---

2   Sentiment-Oriented Summarisation (SOS)

We define three methods to extract a set of candidate 'sentiment' sentences. Each sentence is ranked based on the scores below, and the highest ranked sentences are used to produce the summary. In our experiment we selected the top six sentences often considered enough to summarize a document [2]. If there are less than six sentences in a document, the entire content of document is included as a summary.

Three sentiment scores are produced for each sentence. The first, Intensity Sentiment Score (ISS), attempts to extract opinion sentences that contain as much sentiment information as possible. The sentiment information of each word in a sentence is calculated with the linear combination of TF-IDF (Term Frequency-Inverse Document Frequency) weight, Part-Of-Speech (POS) and Polarity. POS and Polarity values come from SentiWordNet 3.0 [3]. We define the Representativeness Sentiment Score (RSS) function, which computes the representativeness of a sentence by finding keywords that capture the main content in a document. RSS is a measurement of how well a summary captures the principal information content of the original document. We adopt and improve Luhn's approach [4]. RSS is a linear combination of occurrence frequency and TF-IDF weight. The Coverage Sentiment Score (CSS) function calculates the relative significance degree of a sentence with respect to the remaining sentences in a document by means of SentiWord Correlation Factor (SCF). The SCF is computed by iteratively measuring the relevance of each word in a sentence to words in the whole document.

3   Evaluation and Discussion

This study used peer-reviews collected from 50 undergraduate students in ELEC3610 (E-Business Analysis and Design). First students were required to write a first draft of a Project Proposal, then they acted as peer-reviewers to evaluate each a draft as a part of course assessment. Both the draft and the peer-review were part of the assessment and were supported through iWrite [5].

We use the Jensen-Shannon (JS) divergence, a symmetric measure of the similarity of two pairs of distributions as evaluation metric. The JS divergence without smoothing is considered a good measure for comparing input text and summary content [6]. The smaller the value of divergence, the better the summary. We compared the SOS measure based on the three scores and two baseline measures: the *First-6* and *Last-6* summarisers, which are constructed by extracting the first and the last six sentences in the peer review, respectively [7]. The baseline summaries are based on the assumption that students usually start writing their feedback with overall sentiment or conclude them with good summaries. Figure 1 shows the JS divergence for the different peer-reviews and their summaries. Overall, *SOS* outperforms *First-6* and *Last-6*. A notable aspect observed in the Figure 1 is that *First-6* and *Last-6* perform better than or equally to *SOS* in some reviews. In these cases, the reviews usually have the short length of a document, which is approximately equal to the size of a summary. Hence, the length of a review has a considerable influence on summarisation quality. In addition, the presented result shows an interesting fact that students tend to write overall sentiment sentences at the end rather than at the beginning of the feedback. It means that they conclude with opinionated sentences.
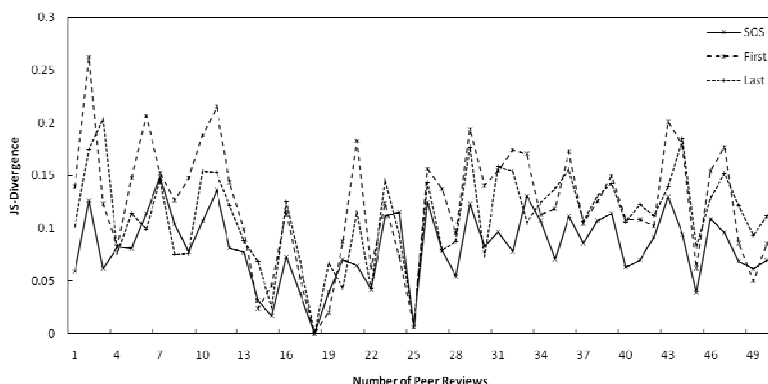
Fig. 1. Jensen-Shannon divergence between peer-review and its summary

4 Conclusions

Our study evaluated an approach for summarization of reviews using sentiment analysis. The method showed to be more accurate than the baseline using heuristics described in the literature. Such techniques can be useful to understand the impact of systems designed to teach academic writing [5] and other automated feedback systems.

References

1. Cho, K., Schunn, C.D., Charney, D.: Commenting on Writing: Typology and Perceived Helpfulness of Comments from Novice Peer Reviewers and Subject Matter Experts. Written Communication 23, 260–294 (2006)
2. Mihalcea, R., Hassan, S.: Using the Essence of Texts to Improve Document Classification. In: Proceedings of Recent Advances in Natural Language Processing (2005)
3. Baccianella, S., Esuli, A., Sebastiani, F.: SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In: Proceedings of the 7th Conference on Language Resources and Evaluation LREC 2010, pp. 2200–2204 (2010)
4. Luhn, H.P.: The Automatic Creation of Literature Abstracts. IBM Journal of Research and Development 2, 159–165 (1958)
5. Calvo, R.A., O'Rourke, S.T., Jones, J., Yacef, K., Reimann, P.: Collaborative Writing Support Tools on the Cloud. IEEE Transactions on Learning Technologies 4, 88–97 (2011)
6. Louis, A., Nenkova, A.: Automatic Summary Evaluation without Human Models. In: Proceedings of the Text Analysing Conference (2008)
7. Pang, B., Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In: 42nd Annual Meeting of the Association for Computational Linguistics, pp. 271–278 (2004)