Building Domain-Specific Sentiment Lexicon by Sentiment Seed Expansion

Bin Liang

Soochow University

* Corresponding author email: 158966900@qq.com

Abstract: Sentiment words extraction is one of the most important subtask of sentiment analysis. However, most sentiment analysis tasks are based on emotion words extracted manually, which requires a lot of manual intervention. In order to overcome this problem, this paper presents a framework to automatically expand the domain-specific sentiment lexicon by sentiment seeds extraction from a large domain corpus of user comments in the automotive field. We use word2vec to learn word embedding and label a small amount of dataset as positive or negative data as training dataset. We extract domain-specific sentiment seeds embedding from training dataset by calculating the sentiment score of the words in training dataset. Afterwards, these sentiment seeds are expanded to cast a large-scale automotive-specific sentiment lexicon by synonyms and k-means algorithm, without any manual annotation. Experimental results show that our approach is able to obtain a large number of new domain-specific sentiment words in the automotive field, and our lexicon reveals better performance than universal sentiment lexicon on user comments sentiment classification.

Conclusions.
.
.
.
References
.
.
.
Acknowledgements
.
.
.