Aspect-based sentiment analysis

Miha Bizjak, Anže Gregorc, Rok Grmek
University of Ljubljana
Faculty for computer and information science
Večna pot 113, SI-1000 Ljubljana
mb—@student.uni-lj.si, ag—@student.uni-lj.si, rg6954@student.uni-lj.si

Abstract

TODO: ...

1 Introduction

Online news, forums and social media are a place for everyone to read and write articles and posts across various domains. People can also leave comments and giving their opinion and express their feelings about the topics. That leads to a huge amount of text content. That is probably why natural language analysis is currently a hot topic around the world. We wanted to extract useful information out of large amount of text data. Since we have no time for reading all the words that are written nowadays, we hope to build a good computer program to do that for us. In this project we chose to do aspect-based sentiment analysis. Our task is to get the subjective information from text material that refer to a entity with the use of natural language processing and other methods. An entity is considered as a person, organization or a location and can be represented multiple times in one document or a sentence and there could be more entities in one document.

2 Related work

The main challenge of entity-based analysis is how to find words that describe the entity and identify if contributes to positive or negative sentiment to a given entity. A lot of related work tried to predict sentiment of the whole document. But in many cases a text can describe polarity of a more entities. That is why we suggest that sentiment analysis is done on entity level. Since our task is more specific we focused more on methods that identify entities.

In the paper (Ding et al., 2018) they developed an entity-based sentiment analysis SentiSW and tested it on issue comments from GitHub. SentiSW can classify issue comments into *<sentiment, entity>* tuples. They evaluate the entity recognition by manually annotation and it achieves 75% accuracy. The main pipeline of this tool is preprocess (words removal, words replacing, stem), feature vectorize (TF-IDF, Doc2vec), classifier (random forest, bagging and other supervised machine learning methods) and entity recognition (rule-based method).

The use of word embeddings provide powerful methods for semantic understanding without the need of creating large amounts of annotated test data. The paper (Sweeney and Padmanabhan, 2017) enhanced the word embeddings approach with the deployment of a sentiment lexicon-based technique to appoint a total score that indicates the polarity of opinion in relation to a particular entity. They associate a given entity with the words describing it and extracting the associated sentiment to try to infer if the text is positive or negative in relation to the entity.

3 Initial ideas and work plan

- Which measures are useful for comparing our results to the ground truth? We have already implemented average class deviation measure, but a simple majority class classifier will probably easily outperform our methods in terms of average deviation (because of the class distribution).

- Any other related work?

- Implement some methods from related work (if they are directly applicable).

- Derive and implement our own methods.

- Are deep learning approaches applicable on a data set like this one?

4 References

References

Jin Ding, Hailong Sun, Xu Wang, and Xudong Liu. 2018. Entity-level sentiment analysis of issue comments. In *Proceedings of the 3rd International Workshop on Emotion Awareness in Software Engineering*, pages 7–13.

Colm Sweeney and Deepak Padmanabhan. 2017. Multi-entity sentiment analysis using entity-level feature extraction and word embeddings approach. In *RANLP*, pages 733–740.