Project for class NLP: Aspect-based sentiment analysis

Dina Sarajlić[1] and **Sanja Stojanoska**[2] and **Vanda Antolović**[3]
Faculty of Computer and Information Science, University of Ljubljana
Email: [1]ds0267@student.uni-lj.si, [2]ss3151@student.uni-lj.si, [3]va2351@student.uni-lj.si

Abstract

1 Introduction

Sentiment analysis on entity level is an important research in the field of Natural language processing (NLP). Sentiment analysis has become popular topic in the past few years, at the beginning used for determining polarity of a given document or text, but it has expanded since. As the Web content got enriched with many product and service reviews, tweets and comments there was an increased need for fine-grained sentiment analysis, also called Aspect-based sentiment analysis (ABSA), to get better insight of the user opinion. ABSA determines the polarity of each aspect, identifies sentiment's orientation to positive, negative or neutral.

This paper proposes ABSA on news articles. The goal is to build a hybrid approach of lexical structure and machine learning to determine the polarity of each given entity in a particular news text.

2 Related work

In general there are two main approaches for sentiment classification; knowledge based approach and machine learning approach.

Knowledge based approach uses predefined lexicons of opinion words labeled as: positive, negative or neutral as well as with polarity percentage. In this case the sentiment is determined by comparing the text of interest with the pre-defined entry in the lexicon. Machine learning approach involves training a sentiment classifier. Many related researches use: Naive Bayes (NB), Support Vector Machines (SVM) and Maximum Entropy.

The (Sweeney and Padmanabhan, 2017) research aims to investigate how entities and their descriptors can be used to identify the sentiment of tweets in relation to the entities or entities if more than one entity exists. The important novelty here is that they treat the tweets differently which have only one entity and the ones with more. In addition, many-entity tweets are analyzed in a way that particular descriptor words are extracted as features and their sentiment is identified using SentiWordNet lexicon while one-entity tweets are processed using the Word2Vec algorithm. Overall accuracy of this system is 71%.

(Ding et al., 2018) designed entity-level sentiment analysis tool SentiSW based on four modules: preprocessing, feature vectorization, sentiment classification and entity recognition. Preprocessing step aims to reduce noise words and uses stemming techniques. Vectorization module transfers bag of words (BOW) into vectors with the help of TF-IDF and Doc2Vec. Sentiment classification classifies comments into neutral, positive and negative. Last module takes only subjective sentiment sentences and recognizes the entity as 'Person' or 'Project' towards which the sentiment refers to. Their goal is to determine emotion on each entity written in a GitHub issue comment. The designed system outputs a tuple of (sentiment, entity) if the comment is subjective or 'neutral statement' if the text is objective. Sentiment classification evaluation achieved 77.19% accuracy.

(Biyani et al., 2015) addresses entity-specific sentiment classification of comments written on Yahoo News. It is formulated as two-stage binary classification. First, filtering the relevant entities. Second, classification of relevant entities as positive or negative. The approach follows three phases: context extraction, feature generation, sentiment classification. Context extraction connects each entity with its context in the given text. In case a sentence does not contain entities, its context belongs to all other entities. On the other hand, if a sentence contains more than one

entity, only phrases are taken as context related with each of the entities. Feature generation uses knowledge based approach to find interesting features. They noticed that entity of type person is more likely to be polar, compared to an entity of non-person type, which is a useful fact for the first-step classification. Moreover, they use algorithm for calculating sentiment score and create features like SentiPos and SentiNeg. Sentiment classification is done by using machine learning algorithm, more specifically Logistic Regression for the neutral-polar classification and Naive Bayes for positive-negative classification. Model evaluation resulted into 67% as F1 measure for the neutral-polar classification and 70% F1 score for the positive-negative.

3 Initial ideas

The goal of this project is to perform accurate sentiment analysis of given entities in a particular news article by using proposed workflow of (Biyani et al., 2015) as a baseline. The dataset is given with annotated entities and co-reference resolutions. Our plan is to structure the project into three stages: preprocessing, features extraction and sentiment classification.

Preprocessing each article by reducing the non-informative words and mapping each entity with its context by using dependency parsing provided by (cla). ReLDIanno text annotation tool from (cla) is a tool that provides four different annotation processes: tagging, lemmatizing, named-entity recognition and dependency parsing. Second phase will extract features from the context checking whether the entity is a person or a subject, determining its polarity and similar lexicon-based rules. Sentiment classification can be done in two stages: first deciding on the polarity and next positive-negative by using standard Machine Learning models proposed by similar research papers.

For the evaluation part, one idea is to compare results with results obtained from a software platform which has support for sentiment analysis. As for now, the idea is to use RapidMiner with defining the process with custom attributes, filters and removal of unnecessary data. For the results comparison, we will use standard measures like recall and precision, as well as their combination known as F1 measure. Data will be visualised with the option for creating association and presented as a dependency tree.

References

Reldianno – text annotation service for processing slovenian, croatian and serbian. https://www.clarin.si/info/k-centre/web-services-documentation/. Accessed: 2020-03-30.

Prakhar Biyani, Cornelia Caragea, and Narayan Bhamidipati. 2015. Entity-specific sentiment classification of yahoo news comments.

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*, page 7–13. Association for Computing Machinery.

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