Evaluation of Techniques Used for Sentiment Analysis

MILESTONE

Yongjie Kang

A20349674

Problem Overview

Sentiment analysis aims to extract affective states and subjective information from texts [1] and is a well-known NLP problem. The big raise of social media such as Facebook and Twitter brings tons of text data which can be used for analyzing sentiment. There are many state-of-the-art algorithms available, for example, logistic regression, support vector machine and neural network, and they do outperform on text categorization than human beings when the size of data is large. However, among those algorithms, which one is outperformed than the others in sentiment analysis? The goal of this project is to make comparison between different algorithms and evaluate their performance, and discuss the availability of combining the advantage of each algorithm to obtain better results.

Data

The data we are using are collected from IMDB[1] and can be downloaded directly from Stanford research datasets[2]. The data contain 12500 positive reviews and 12500 negative reviews. Due to the IMDB comment system, we are able to know people's opinions from the number of stars they give to the movie and this feature largely simplify the evaluation process.

Method

The algorithms include logistic regression, support vector machine, neural network and naïve Bayes classifier. For each algorithm, we are using grid search provided by sklearn[3] to train the model with different parameters and select the model with the highest accuracy. The comparison contains both time consuming and accuracy.

Experiments

1. Preprocessing

We are using NLTK packages to remove stopwords, however, in sentiment analysis, words such as "not", "don't'", "rarely" are informative to determine a review as negative reviews, we need to build a subset by using stopwords set provided by NLTK and only contains useless words.

1.www.imdb.com
2.http://ai.stanford.edu/~amaas/data/sentiment/
3.http://scikit-learn.org/stable/

2. Generate Models

We use model generator created by sklearn package to create models and use them to fit the data. Here are sample outputs of logistic regression and SVM. As we can see, in general, logistic regression performed better. Note that the time consuming of training SVM was much longer than SVM.

	Time	Category	Precision	Recall	F1-score
Logistic regression	10.025s	neg	0.89	0.85	0.87
		pos	0.85	0.89	0.87
SVM	609.37s	neg	0.89	0.45	0.60
		pos	0.63	0.94	0.76

Related Works

A broad overview of existing related works, the range varies from phrase-level sentiment analysis to document-level sentiment analysis. On document level sentiment analysis, the authors of (Bo Pang, 2008)demonstrated different machine learning techniques and the background knowledge of each algorithm. It's a general view of all state-of-the-art approaches dealing with sentiment and opinion analysis. It inspired me to make comparison between multiple algorithms. Regarding feature engineering, in (Efthymios Kouloumpis, 2011), they considered features such as n-gram features, part-of-speech features and lexicon features to train the models. Although they made a conclusion that the part-of-speech features are not useful in microblogging domain, but considering them in movie reviews domain could be an interesting trial. Pre-processing has an insignificant role in opinion mining task for the specific domain of reviews (Fernando Leandro dos Santos, 2004). In (Ryan Coughlin, 2017) Labeled Latent Dirichlet Allocation was applied to recognize the topic words (keywords). In this way, those most import words would have higher weights. the authors of (Changhua Yang, 2007) applied SVM and CRF classifiers on web blog opinion classification, they first categorized sentiment on sentence level and took the whole document into account by combining the results of sentence level analysis. They also concluded that considering the last sentence of a document would generate the best classifier.

Tasks

It's an individual project and I'll write program and paper from sketch.

References

[1] Pang, B., and Lee, L. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2):1–135.
[2] E. Haddi, X. Liu, and Y. Shi, ''The role of text pre-processing in sentiment analysis,'' Procedia Comput. Sci., vol. 17, pp. 26–32, Sep. 2014.
[3] ZHAO JIANQIANG and GUI XIAOLIN, "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis" IEEE., vol. 5, Feb. 2017.
[4] Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg!. Icwsm, 11(538-541), 164.

[5] Yang, Changhua, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. "Emotion classification using web blog corpora." In Web Intelligence, IEEE/WIC/ACM International Conference on, pp. 275-278. IEEE, 2007.