CIS 434 Final Project Report: Airline Sentiment Classification

MSBA Miao Xi

Introduction

Social media analytics is very useful for all consumer-facing industries to measure their performance and help to craft future improvement. One typical implementation is in airline industry. Sentiment analysis could be the first step of determining the general consumer feedback and identifying low satisfaction area. In this case, a set of un-labeled tweets about US airline companies is given and the task is to classify them into negative and non-negative sentiments.

Methods

There are 3 common approaches toward this task. First is lexicon-based method, comparing words from a tweet with dictionaries with words expressing positive and negative sentiment. Then calculate a sentiment score for each tweet. The second method is supervised machine-learning. Use labeled tweets about US airlines from Kaggle https://www.kaggle.com/crowdflower/twitter-airline-sentiment , apply supervised machine learning algorithms such as SVM, neural network to construct models, then apply such model to the unlabeled tweets. Third method is an unsupervised machine learning approach. Applying LDA to the document term matrix (DTM) to extract underlying topic of the collection of tweets, then identify positive and negative topics, later to measure a tweet's sentiment as combination of negative and positive topics.

In this project, the first 2 approaches are implemented, the lexicon-based approach is more promising after manually inspect label output with original tweet. So following content would largely focus on this approach.

Lexicon-based Sentiment Classification

Dictionaries

Lexicon-based sentiment classification relies heavily on the coverage of the positive and negative dictionaries. In this project, dictionaries are acquired from http://ptrckprry.com/course/ssd/data/positive-words.txt and http://ptrckprry.com/course/ssd/data/negative-words.txt two links.

Class	Number of Words
Negative	4781
Positive	2005

The two dictionaries contain words and their frequent miss-spelling, as such miss-spellings are very common in social media settings.

The dictionaries mentioned above are not sufficient for the task in this project because they lack industry and situation specification. Later on, after construct the initial model and obtained initial prediction, more words would be added to the dictionaries in order to increase F1 measure, which will be sued throughout this project as the model evaluation metrics. The process of adding new words to the dictionary will be explained in later sections.

Data Preprocessing

Each tweet would be tokenized on word level only with special character removed and stop words removed.

Expansion of Negative Dictionary

The generic 'stopwords' function in R would remove interrogative pronoun such as 'how', 'why'. But after manually inspect the tweets, it is clear that interrogative pronouns are largely used in negative comments. Harsh and sarcastic questions are used to express negative sentiment towards the airlines. So in this project, interrogative pronouns are kept and added into negative dictionaries to reflect this phenomenon.

Also added into the negative dictionary are common curse words that are not previously included. Words such as 'piss', 'screw' and their other forms are commonly used negative tweets.

Some words normally convey neutral sentiments are also included in the negative dictionary because their correlation with negative experience in airline industry setting. Verb 'wait' could indicate that the passenger encountered a delay. 'Kick' could signal that the passenger was unwillingly asked to surrender their seat to airline employees. Noun 'tarmac' is also added, because there are multiple observations that this indicates the plane was unable to take off for a substantial amount of time after it left the terminal. Those words and their other forms are also included as negative words. What are also included are words describing time, when 'day', 'hour', 'minute' appear, they often signal a long wait time either for a delayed flight or a prolonged customer hotline.

Model Construction

After a more comprehensive negative dictionary is obtained, we would construct a model to measure the overall sentiment. Sentiment Score (S_s) for each tweet is computed as the number of positive words (N_p) minus the number of negative words (N_n) multiplied by a weight (w_n) on N_n.

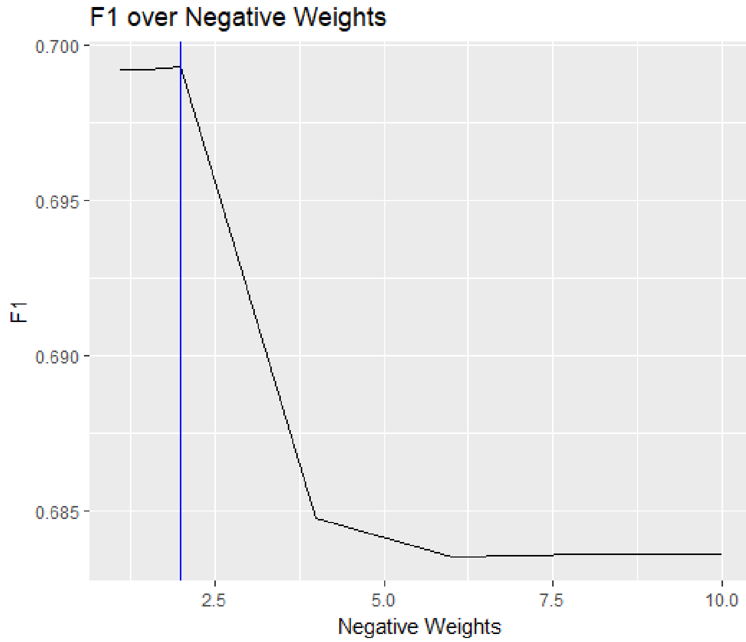
$$S_s = N_p - w_n \times N_n$$

Because the existence of negation words such as 'not', 'no'. The negative weight w_n should be larger than 0, in order to bring the combination of 'not good' to a negative sentiment. The tweets that contain neither words form both positive and negative dictionaries are seen as neutral, hence are classed with positive tweets together as a 'non-negative' class.

Model Tuning

Apply data preprocessing procedure and model on labeled US airline tweet data from Kaggle. When adjusting the weight w_n, the F1 measure of the model prediction would change accordingly. So finding an optimal weight to acquire the highest F1 measure is very important. The below graph shows the change of F1 measure over negative weight w_n. The optimal $w_n = 2$, with a F1 measure for non-negative class of 0.6957. The confusion matrix is shown as follow:

	negative	non-negative
negative	6569	979
non-negative	813	2084



Prediction for Unlabeled Data

	negative	non-negative
Count	4460	123