Aspect based Sentiment Analysis using Machine Learning

Umair Cheema
Harvard Extension School
umair.cheema@ieee.org

Athar Pasha
Harvard Extension School
athar.n.pasha@gmail.com

ABSTRACT
In this paper, we describe some rudimentary approaches to combine Machine Learning and Natural Language Processing (NLP) for conducting Aspect Based Sentiment Analysis of customer product reviews. Aspect Based Sentiment Analysis is often formulated as a Multilabel or Multidimensional classification and is inherently more challenging than simple binary or multiclass classification problems. We have evaluated performance of Random Forest and Support Vector Machine (SVM) algorithms to detect aspects and sentiment polarities in customer reviews using traditional Bag of Words, n-gram and POS tag based features as well as using Vector Space Model based domain specific Word Embeddings created using Word2Vec and Doc2Vec.

Keywords
Sentiment Analysis, Multilabel classification, Word Embeddings

1. INTRODUCTION
Traditionally Customers used to consult their friends and family members before purchasing a product but over last few years potential customers of a product read online reviews before buying a product. These online reviews often consist of a couple of sentences accompanied with a star based rating. As a customer, reading a handful of reviews and getting a feeling about different pros and cons of a product and how it stands against similar products available in the market is relatively a straightforward task. As a business however the task becomes quite challenging as businesses not only want to understand the overall sentiment of their consumers regarding their product but also need additional information as to what product features are exceeding the expectations of the users and what features need improvement. The overall task of converting the often-textual reviews into useful actionable insight becomes more challenging due to increased volume and velocity of these reviews and unstructured nature of the review data. This task of extracting information about sentiment of the product's users is called Sentiment Analysis (SA).

Sentiment Analysis (SA), which is also known as Opinion Mining (OM) is formally defined as the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities such as products, services, organizations, individuals, issues, events, topics and their attributes[1]. It is the computational study of people's opinions attitudes and emotions toward an entity[2]. SA identifies the sentiment expressed in a text then and analyzes it. Its target is to find opinions, identify the sentiments they express and then classify their polarity.

2. Aspect based Sentiment Analysis
Sentiment Analysis is typically conducted at three main levels:

1. *Document-level SA* classifies the overall sentiment of the whole document into positive, negative, neutral classes or classes with discrete numbers showing strength of positive or negative sentiment.

2. *Sentence-level SA* conducts a sentence wide classification of sentiment polarity.

3. *Aspect-level SA* aims to classify the sentiment with respect to the specific aspects of entities. It directly looks at the opinion and its target instead of just looking at document, paragraph, sentence or phrase level sentiment.

Sentiment Analysis/Opinion mining is a fast growing research area and there are numerous research publications. A survey paper done by researchers at Ain Shams university [2] contained a comprehensive list of algorithms used for Sentiment Analysis/Opinion Mining. Some of the common Machine Learning algorithms used in different research papers are Naive Bayes Classifier, Bayesian Network, Support Vector Machine (SVM), Neural Network and Decision Trees.

Many researchers have explored tools and techniques offered by NLP to extract features to be fed to different Machine Learning algorithms. In [3], the researchers used sentence level subjectivity/objectivity classification, co-reference resolution, dependency parsing and SentiWordNet in combination with Support Vector Machines (SVM) for conducting Aspect Based Sentiment Analysis. In [4] Word Vectors and Convolutional Neural Networks (CNN) for aspect based sentiment analysis.

In [5] researchers used Deep Memory Networks to capture significance of context words for aspect level sentiment classification. The authors found the accuracy comparable to SVM based classifiers.

Some researchers have even tried unsupervised learning for Aspect level Sentiment Analysis as shown in [6].

Word-Aspect Association and Sentiment Lexicons were used in [7] for detecting aspects and their relative polarities in Customer Reviews. The authors also employed unlabeled Reviews Corpora such as Amazon laptop reviews corpus and Yelp restaurant reviews corpus and achieved an F1-score of 88.58. The feature vectors used in this particular case were ngrams, POS tags and lexicon features etc. We have used Yelp restaurant reviews corpora in our experiments but to generate domain specific Word Embeddings and Paragraph Vectors.

3. Dataset

The datasets used for this project were made available by International Workshop on Semantic Evaluation 2016 (SemEval2016) and Yelp dataset challenge. We used 325 labeled customer reviews on restaurants for training and 90 reviews for testing from SemEval datasets [8]. We used around 9,90,627 reviews on 21,892 restaurants from the Yelp dataset [9].

The restaurant reviews in SemEval dataset are labeled with Entity, Attribute pairs along with their sentiment polarities. The possible combinations of the entities and attributes are given in table 1.

Table 1. SemEval Restaurant Reviews Entity Attribute Combinations [10]

	GENERAL	PRICES	QUALITY	STYLE& OPTIONS	MISCELLANEOUS
RESTAURANT	✓	✓	✗	✗	✓
FOOD	✗	✓	✓	✓	✗
DRINKS	✗	✓	✓	✓	✗
AMBIENCE	✓	✗	✗	✗	✗
SERVICE	✓	✗	✗	✗	✗
LOCATION	✓	✗	✗	✗	✗

We encoded the entity aspect pairs and polarities into target variable for classification such that there were 12 labels with 2 polarities (Positive, Negative) each. This transformation formed a Multidimensional classification scheme.

The reviews text was cleaned to remove punctuation marks, newline characters, stop words removal. We then conducted a round of exploratory data analysis to look at the word frequencies for both SemEval and Yelp datasets given in Figure 1 and Figure 2 respectively.

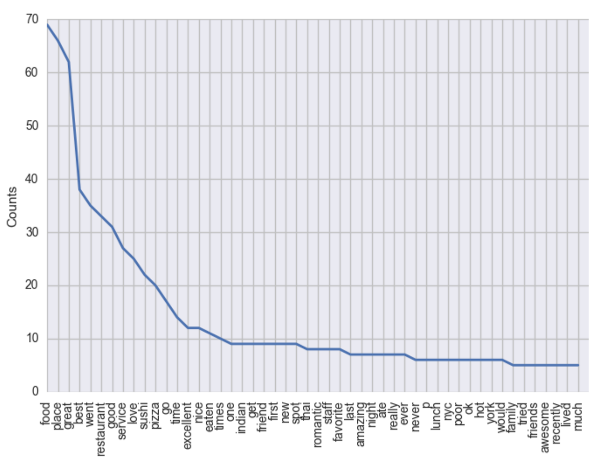

Figure 1: Word Frequency for SemEval dataset

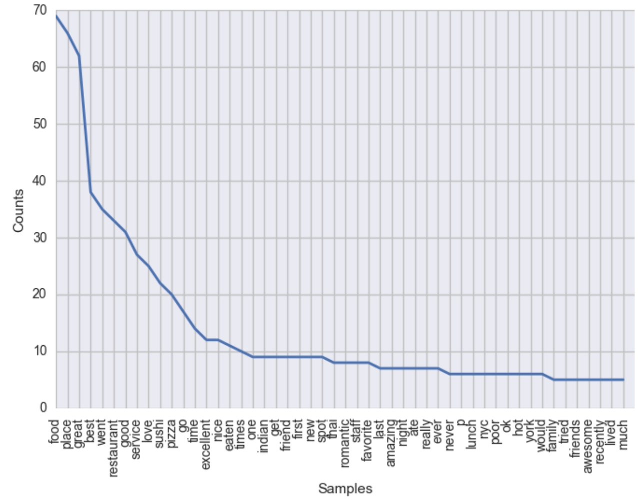

Figure 2: Word Frequency for Yelp dataset

One thing that stood out by observing these diagrams was that both restaurant review datasets had *food* and *place* as the most frequently used words followed by positive sentiment words such as *great* and *best*. This was an indication that there were more positive reviews than negative and mostly about the food. Other aspects such as Drink prices and service were not mentioned as frequently as food and place. We therefore checked the aspect sentiment class distribution for our training dataset to ascertain the balance between positive and negative examples for each aspect. The distribution is shown in Figure 3.

Looking at the class distribution it became clear that the classes were highly imbalanced as there were more positive reviews than negative reviews for all aspects.

4. Method

The first task was to select an evaluation metric for this multidimensional classification problem. We used F1-micro metric [11] for aspect detection. The accuracy for polarity detection is based on a simple ratio of correctly identified polarities to the total predictions across all aspects.

$$F_1^{\mu} = 2\sum_{i=1}^{N} TP_i / [\sum_{i=1}^{N} FP_i + \sum_{i=1}^{N} FN_i + 2\sum_{i=1}^{N} TP_i]$$

We then trained RandomForest and Support Vector Machine classifiers using Bag of Words, Bag of n-grams and POS tag based features. We used POS tags of the review appended to the actual tokens before feeding to vectorization stage. This improved the 71%accuracy from base classification model to 73%. Our next step was to prepare domain specific Word Embeddings using Yelp reviews. Word Embeddings are dense low dimensional representation of words. Word Embeddings convert words from a vocabulary into vectors of real numbers. They have been used for sentiment analysis in the past such as in [12].

We tried tweaking several parameters for Word2Vec and Paragraph Vector based embeddings. Word2Vec is described in [13] and Paragraph Vectors are introduced in [14].

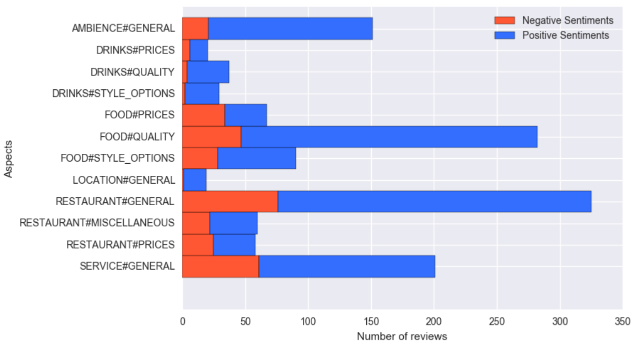

Figure 3:Sentiment Class distribution in SemEval dataset

We tried generating domain specific Word Embeddings using both Continuous Bag of Words (CBOW) and Skip-gram approaches. Skip-gram approach of generating Word Vectors when fed into a SVM classifier with Radial Basis Function (RBF) kernel improved the accuracy for both aspect detection as well as sentiment polarity detection.

Our paragraph vector (Doc2Vec) approach did not give promising results. We tried changing the word window size , switching between Distributed Bag of Words (DBOW) and Distributed Memory (DM) paragraph vectors and also changing the size of the output vector features. We used a subset of around 40,000 reviews from the Yelp dataset for this part.

The best Word Embedding model was CBOW based model with 500 features but with an additional step of collocation detection using phrases before feeding into Word2Vec.

This model was then used to vectorize most frequently used words in the training dataset. To visualize these frequent words we used t-Distributed Stochastic Neighbor Embedding (t-SNE) to reduce the dimensions of the vectorized data into two components. The resulting visualization is shown in Figure 4.

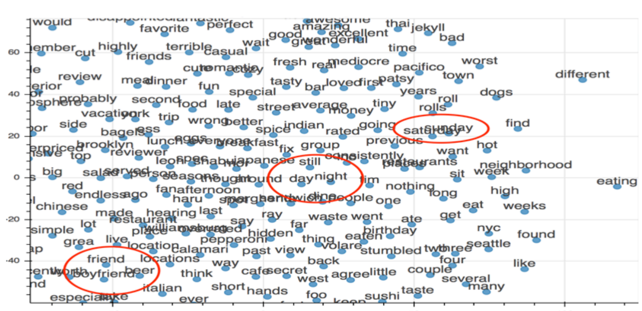

Figure 4: Vector Space Model shown with t-SNE

The visualization shows that the Word2Vec model captured implicit semantic similarity between words such as days of the week and day and night etc.

We also evaluated at how well this domain specific Word Embedding model can identify a word such as 'Sushi'. The result is shown in Table 2.

Table 2. Most similar words to the word 'sushi' identified by domain specific word embeddings model using Word2Vec

Word	Score
nigiri	0.63327
sashimi	0.62191
ayce_sushi	0.60983
raw_fish	0.53127
dim_sum	0.52201

5. Results

We have found domain specific Word Embeddings to be an invaluable tool for preparing features for Aspect Based Sentiment Analysis. A summary of F1-Scores and polarity accuracy scores are given in Table 3.

Table 3. Comparison of Methods

Method	F1-Score (Aspect)	Polarity Accuracy
RandomForest (Bag of Words)	0.710	0.84
RandomForest (Bag of Words + POS)	0.732	0.83
RandomForest (Word2Vec CBOW)	0.715	0.87
SVM(Word2Vec CBOW)	0.739	0.91
SVM(Word2Vec Skip gram)	0.752	0.92
SVM(Word2Vec Phrase detection + CBOW)	0.760	0.92
SVM (Doc2Vec)	0.7	0.83

The best results were obtained using SVM with RBF kernel applied on vectors generated from Word2Vec based embeddings model created using Yelp dataset. This Word2Vec based model was fed bigrams for collocation identifications.

6. Conclusion

Word2Vec and other word embedding models when trained using domain specific corpuses have been found to be invaluable to convert textual data into vector space model. The main advantage is that feature engineering required is significantly simple as opposed to complex dependency parsing based deep NLP techniques to get similar accuracies. One shortcoming that needs to be addressed is balancing the classes using any of the Multilabel oversampling techniques such as SMOTE before data is fed into SVMs.

7. Future Work

We would like to explore trying the same models but after balancing the classes using Synthetic Minority Oversampling Technique (SMOTE). We also intend to use Stanford University's Global Vectors for Word Representation (Glove) for building domain specific word vectors. Another possible area we would like to explore is using exploiting Convolutional Neural Networks with Word vectors as input features.

8. REFERENCES

[1] Bing Liu, 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. 1 Edition. Cambridge University Press*

[2] Mehdat W, Hassan A, Korashy H. *Sentiment Analysis algorithms and applications: A survey. Ain Shams Engineering Journal (2014)*

[3] Raisa V, Jayasree.M J. 2013. *Aspect Based Sentiment Analysis using Support Vector Machine Classifier. International Conference on Advances in Computing, Communications and Informatics (ICACCI)*

[4] Bo Wang and Min Liu. 2015. *Deep Learning For Aspect-Based Sentiment Analysis*. Stanford University report, https://cs224d.stanford.edu/reports/WangBo.pdf

[5] Tang, D., Qin, B., Liu, T.2016: *Aspect Level Sentiment Classification with Deep Memory Network. In: EMNLP*.

[6] Salud M, M.Teresa, Eugenio, L. Alfonso.2016: *Combining resources to improve unsupervised sentiment analysis at aspect level. Journal of Information Science 2016 Vol. 42(2)213-229*

[7] Svetlana, K., Xiaodan Zhu, Colin C, Saif M, 2014: *Detecting Aspects and Sentiment in Customer Reviews*. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 437-442*

[8] SemEval-2016 Task 5. (n.d.). Retrieved November 2016, http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools

[9] Yelp. Data Set Challenge Website Retrieved November, 2016, https://www.yelp.com.au/dataset_challenge

[10] SemEval 2016 Task 5, Aspect Based Sentiment Analysis (ABSA - 2016), Annotation Guidelines . http://alt.qcri.org/semeval2016/task5/data/uploads/absa2016_annotationguidelines.pdf

[11] Sheng G, Wen W, Chin-Hui L, Tat-Seng C , 2004. *A MFoM Learning Approach to Robust Multiclass Multi-Label Text Categorization. Proceedings of the 21st International Conference on Machine Learning, pp. 42*

[12] Pengfei L, Shafiq J, Helen M.,2015. Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing.*

[13] Tomas Mikolov et al 2013. *Efficient Estimation of word representations in vector space arXiv preprint arXiv:1301.3781*

[14] Quoc Le, Tomas Mikolov 2014. *Distributed Representations of Sentences and Documents. Proceedings of the 31st International Conference on Machine Learning, pp. 1188-1196*