Aspect-based Sentiment-Analysis. Final report

Diego Ihara (dihara2@uic.edu), Niharika Dharmapuri (ndharm3@uic.edu)

Abstract

Sentiment analysis is the process of analyzing people's opinions, appraisals, emotions and attitudes towards an entity and its various aspects. As a final project, we have been given sets of reviews along with given aspect terms. The objective was to build a model that identifies if the associated sentiment is positive, negative or neutral. In order to achieve this result, we applied techniques such as Naïve-Bayes, Support Vector Machines (SVM) and Neural Networks. In this report, we present and explain our methodology and obtained results.

1. Introduction

Opinions are important as they are key influencers of our behaviour. There are various techniques for identifying sentiment expressed by a review. However, much research is still needed in order to correctly identify sentiments involving a specific aspect term. We refer the reader to [1] for more details.

In this project, we were given two datasets: one about laptop reviews and another about restaurant reviews. The objective was to built two classifiers that give best accuracy for each data set.

In Section 2, we explain our initial approach without considering the aspect term. In Section 3 we detail the steps we followed in order to incorporate the aspect term into our model. We briefly mention other methods we tried but without success in Section 4. Finally, conclusions can be found in Section 5.

2. A naïve approach

Our initial approach was to train our model considering each complete review, without giving the aspect term any consideration.

2.1. Data pre-processing

For the pre-processing, we replaced all [COMMA] markers for the proper punctuation sign. Next, we removed stop-words to reduce the number of words in the bag of words format and improve accuracy. Next, we used stemming technique to reduce a word to their root. This reduces the number of words in bag of words and thereby speeds up the computations.

2.2. Feature representation

2.2.1. N-grams

Splitting reviews into n-grams lemmas for example grouping words in sets of two and three. Our final approach has given the best result using ngram range of 1 to 3.

2.2.2. Bag of words

Converting reviews into a list of lemmas . In the bag of words representation, we select unique words from the dataset and treat each of the word as a feature. If the word is present in a review, we give it 1 and 0 if it is not present. We tried both dense and sparse vector representations where appropiate.

2.2.3. TF-IDF

TF-IDF transformation: computing how important a lemma is based on number of times it appears in the dataset.

2.3. k-fold Cross-Validation

k-Fold Cross-Validation with k = 10 was used for dividing the data into training and testing. Each experiment was performed using 9 parts as training and performing the test on the remaining part. This was repeated 10 times. All results presented are the average of each metric on all ten runs.

2.4. Classification methods tried

2.4.1. Naïve Bayes

The first classifier we used is the Multinomial Naive Bayes from Scikit [2] or multinomial NB model, which is a probabilistic model that gives probability of the document given belonging to a class. More details in [3]. We achieve a decent accuracy. For the positive class, we got a precision of 0.69, recall of 0.92 and F1-score of 0.79 for dataset 1 and for dataset 2 we got precision of 0.63, recall of 0.96 and F-score of 0.77. For the negative class we got precision of 0.71, recall of 0.76 and F1-score of 0.73 for dataset 1 and for dataset 2 we got precision of 0.60, recall of 0.17 and F1-score of 0.27.

2.4.2. Random Forest

For the dataset1, Random Forest [4] was giving a test accuracy for is 0.76. It gave precision value of 0.80, recall value of 0.82 and F1-score of 0.81. For the dataset 2, we achieved overall accuracy of 0.68 with precision value of 0.70, recall value of 0.97 and F1-score of 0.81. These results can be seen in Table 1

Data set 1 Accuracy: 0.76				
Class	**Precision**	**Recall**	**f1-score**	**Support**
-1	0.70	0.89	0.78	72
0	0.88	0.46	0.60	48
1	0.80	0.82	0.81	100
avg/total	0.78	0.76	0.75	220
Data set 2 Accuracy: 0.68				
Class	**Precision**	**Recall**	**f1-score**	**Support**
-1	0.56	0.34	0.42	70
0	0.73	0.15	0.25	73
1	0.70	0.97	0.81	217
avg/total	0.68	0.68	0.62	360

Table 1: Random Forest Test

2.4.3. Convolutional Neural Network (CNN)

For the dataset1, CNN was giving a test accuracy for is 0.75. It gave precision value of 0.81, recall value of 0.86 and F1-score of 0.83. For the dataset 2, we achieved overall accuracy of 0.63 with precision value of 0.75., recall value of 0.81 and F1-score of 0.78. Results are presented in Table 3.

2.4.4. Support Vector Machine (SVM)

For the dataset1, SVM [5] was giving a test accuracy for is 0.80. It gave precision value of 0.84, recall value of 0.90 and F1-score of 0.87. For the dataset 2, we achieved overall accuracy of 0.71 with precision value of 0.80, recall value of 0.82 and F1-score of 0.81. Metrics can be seen in Table **??**.

For all of the above classifiers we only did sentiment classification without using aspect terms. SVM was giving best accuracy so far. In the following models we tried to incorporate aspect terms for the classification.

Data set 1 Accuracy: 0.75				
Class	**Precision**	**Recall**	**f1-score**	**Support**
-1	0.75	0.83	0.79	72
0	0.62	0.44	0.51	48
1	0.81	0.86	0.83	100
avg/total	0.75	0.76	0.75	220
Data set 2 Accuracy: 0.63				
Class	**Precision**	**Recall**	**f1-score**	**Support**
-1	0.51	0.39	0.44	70
0	0.34	0.34	0.34	73
1	0.75	0.81	0.78	217
avg/total	0.62	0.63	0.62	360

Table 2: Neural Network Test

3. Incorporating Aspect Terms

In order to incorporate aspect terms into the model, we performed the following:

- POS Tagging.
- Clause separation.

For POS Tagging, we used [6] to better analyzed each component of the review [7]. We then performed chunking using a context free grammar we defined manually, were words acting as CC (coordinating conjuntion) in addition to parenthesis and commas were used to separate each review into clauses. Finally, only that clause that contained the given aspect term was considered both when training and testing.

After considering aspect terms, all accuracies dropped by at least five percent in all methods tried.

4. Other techniques tried

We also tried other approaches such as Convolutional Neural Networks and Long short-term memory Neural Networks. For these approaches, the accuracies we obtained were below 25 percent, therefore they were not reported but mentioned during presentations. More time was needed in order to properly configure and tune all parameters involved in this type of Neural

Data set 1 Accuracy: 0.80

Class	Precision	Recall	f1-score	Support
-1	0.73	0.92	0.81	72
0	0.91	0.44	0.59	48
1	0.84	0.90	0.87	100
avg/total	0.82	0.80	0.79	220

Data set 2 Accuracy: 0.71

Class	Precision	Recall	f1-score	Support
-1	0.54	0.56	0.55	70
0	0.60	0.53	0.57	73
1	0.80	0.82	0.81	217
avg/total	0.71	0.71	0.71	360

Table 3: SVM Test

Networks in order to produce good results as reported in the literature [8] and [9].

5. Conclusions

Aspect-Based Sentiment-Analysis is a very hard problem that many approaches attempt to solve with various degrees of success. As is prevalent in Machine Learning, the effectiveness of a given method greatly depends on the domain in which it is applied and also on the level of experience of the practitioner.

Sometimes, the latest and most popular may not work as expected (deep learning), but a known and perhaps not as exciting tool (SVM) produces better results when combined with careful pre-processing and good data representation.

[1] B. Liu, Sentiment analysis and opinion mining, Synthesis lectures on human language technologies 5 (2012) 1–167.

[2] SCIKIT, Multinomial NB, http://scikit-learn.org/stable/modules/ generated/sklearn.naive$_{b}$*ayes.MultinomialNB.html*, 05/06/2018.*A. McCallum*, *K. Nigam*, *et* *AAAI*−98*workshoponlearningfortextcategorization*, *volume*752, *Citeseer*, *pp*.41− −48.

[3] SCIKIT, Random Forest, http://scikit-learn.org/stable/modules/ generated/sklearn.ensemble.RandomForestClassifier.html, 05/05/2018.

[5] SCIKIT, Support Vector Machine, http://scikit-learn.org/stable/modules/svm.html, 05/04/2018.

[6] NLTK, Categorizing and Tagging, https://www.nltk.org/book/ch05.html, 05/06/2018.

[7] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, D. McClosky, The stanford corenlp natural language processing toolkit, in: Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, pp. 55–60.

[8] S. Pateria, P. Choubey, Aktski at semeval-2016 task 5: Aspect based sentiment analysis for consumer reviews, in: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pp. 318–324.

[9] T. T. Thet, J.-C. Na, C. S. Khoo, Aspect-based sentiment analysis of movie reviews on discussion boards, Journal of information science 36 (2010) 823–848.