A Sentiment Analysis Plugin for Wordpress

by Hamid Hooshmandi and Dhuha Al-Amiri

Abstract

This report presents our experimental work of creating a sentiment analysis Wordpress plugin that classifies user reviews, in a hypothetical camera website. We adopt two of Machine learning classifiers; SVM, Support Vector Machine, and NB, Naive Bayes in order to implement our task. The results demonstrate that the performance of multinomial naive bayes (NB) and support vector machine(SVM) differs when unseen testing data is used for evaluation, resulting in a better accuracy of 79.01% for SVM and 72% for NB. Whereas their accuracies are similar at about 93% when training data used as testing data .

Introduction

In the recent years, consumers have mainly relied on E-commerce. Therefore, the demand for knowing about people's preference for some products than the others has significantly raised. This can be seen in commerce websites where consumers type their reviews and others benefit from them. It is also useful for companies who own these products in order to enhance their merchandise. Due to the fact that people’s review on products’ quality represents a reference for both customers and organizations, sentiment analysis is the method that has been utilized to automatically analyse people’s comments and classify them into positive, negative or neutral. Sentiment analysis is “a language processing task that uses a computational approach [6] ” to discover viewpoints about a topic or object in a text and classify them into different categories [6].

In this project, Machine Learning and text classification techniques were used to create a sentiment analysis plug-in for a hypothetical Wordpress website selling different camera models. The website allows consumers to type their comments on different camera

models. The website administrator then is able to see if each of these reviews has positive or negative sentiment and with what probability.

Currently there is only one sentiment analysis plugin for WordPress, and it is not designed to accurately classified text in a particular domain. In general many of the "one size fits all" tools are not appropriate and accurate for specific content type; therefore to have an accurate classifier for a specific domain, having relevant training data is very important [4]. Our proposed method uses a web crawler to collect and scrap relevant data and then uses this data to train domain specific classifier. The method can be easily adopted for creating different classifiers for many other classification tasks, everything from sentiment analysis of restaurant, movie or books reviews to task of spam detection. Our approach is based on Machine learning. Two classifiers has been created; one using Multinomial Naive Bayes and the other using Support Vector Machine (SVM) to classify the polarity of each product.

Related Work

In [7], Waila et al. examine the performance of Machine learning using Naive Bayes and SVM algorithms with SO-PMI- IR algorithm, which is a method of unsupervised learning, for sentiment analysis. They use three data sets of movie reviews. Two of the data sets are labeled from Cornell; one of the dataset has 1000 positive and 100 negative reviews while the the other dataset has an equal number of 700 positive and negative reviews. The third dataset has 1000 reviews labeled manually and collected from IMDB. Naive Bayes algorithm is implemented using Java with Eclipse IDE and the datasets provided to NB as k-folds, where k is chosen to be 3, 5, and 10. “3-folds application of test data” represent how the training and test data are divided. Two of the datasets are the training data and the the third is the test data. It is has been shown that for implementing the 3-folds in the first dataset, the accuracy of NB is about 83.8%, SVM is 78.15%, and the SO-PMI- IR is almost 84%. This indicates that the performance of NB is better than SVM in terms of the accuracy in the all the three datasets, but it is close to SO-PMI- IR algorithm.

The work of Neethu and Ragasree in [3], analyses twitter posts to determine people's opinions about electronic products using Machine learning. The datasets collected automatically using Twitter API and labeled manually as positive or negative. Due to the fact that tweet posts have misspelling and slang words, sentiment analysis is implemented based on sentence level. The work includes a new feature vector presented to classify each post as positive or negative. To illustrate how feature vector is utilized, specific features of twitter are determined and included into feature vector. Then these features are excluded from the tweets. The process is repeated on the remaining text. It is founded that the accuracy of feature vector using Naive Bayes, SVM and other machine learning algorithms are similar. Therefor, this feature vector is recommended to be used for sentiment analysis of electronic products.

Filipa and Jao in [1] evaluate three algorithms of Machine Learning; SVM, RIPPER and Sentence Level classifier. The datasets are different collection of movies(IMDb), books, DVDs and music (Amazon). This study focuses on using adjective- word phrases as a feature selection, instead of opinion words using unigrams, that are used in other works. It is concluded that for movie reviews that are analysed using pair adjective-word, SVM performs the best among other classifiers.

Another work is by Qiang et al. [5] for mining reviews from travel blogs where Naive Bayes, SVM and N-gram model are conducted. The tourist reviews are collected from Yahoo.com for the most seven common places visited in the USA and Europe. each post is rated from 1 to 5 star. 1 and 2 start represents the negative reviews, 4 and 5 star are considered as positive opinions. The result indicated that all the three algorithms can exceed 80% of accuracy. However, SVM and N-gram model have better performance than Naive Bayes. It is also noticed that the distinction of the accuracy among the three classifiers differ based on the size of the training data. If the training data is small that ranges from 40 to 100 reviews, the difference is significantly increases. Whereas the larger training data shows no enormous difference.

Hanhoon et al in [2] conduct a research on finding “a new senti-lexicon for sentiment analysis of restaurant reviews.” 70,000 restaurant reviews are collected from restaurant search sites and the star scores are involved in the reveiws. In addition, they propose an improved Naive Bayes algorithm due to the issue of “classifying a review document as a positive sentiment or negative sentiment”. The issue is that the accuracy of positive sentiment seems higher than the negative sentiment by 10%. In this case, when calculating the average of the accuracy of the two categories (positive and negative), this causes the average to be decreased. As a result, Naive Bayes is improved in this research to prevent this issue. The result of the proposed work with using of unigrams and bigrams as the feature confirms the improved gap between the accuracy of positive and accuracy of negative sentiment which reduced to 3.6%, comparing to the original Naive Bayes which poses a gap of 28.5% between the accuracy of positive and negative sentiment.

Methodology

1.Collecting Training Data

To collect domain specific training data, a customized crawler was used. Specific Amazon urls related to camera reviews were manually selected and added to the list of urls that has to be crawled by the crawler. The crawler scrapes 1 star reviews and labels them as negative, and scrapes 5 star reviews and labels them as positive. The number of star is self-reported by the poster of the review. All the reviews are saved to a local database. Based on querying the database, testing and training data are created. In total, 1023 labeled reviews was collected. 555 reviews being positive, and 468 being negative.

2.Training the Classifiers

Weka (Waikato Environment for Knowledge Analysis) was used for training the classifier. Two classifiers were created; one using Multinomial Naive Bayes, and the other is using Support Vector Machines. Different tokenization setups were used. The tokenization setup

is shown in Table 1. The accuracy of NB and SVM classifiers are both tested on the training data with 10 fold cross validation as shown in Table 2 and Table 3.

Tokenization #	IDF	TF	Lowercase	Document length Normalization	Stemer	tokenizer	Stop words
1	yes	yes	no	no	Lovins Stemer	Alphabetic	no
2	no	no	no	no	no setemer	Alphabetic	no
3	no	no	yes	no	no setemer	Alphabetic	no
4	yes	yes	yes	no	Lovins Stemer	Alphabetic	yes
5	no	no	yes	no	Lovins Stemer	Alphabetic	yes
6	no	yes	yes	yes	no setemer	Alphabetic	yes

Table 1. Tokenization setups

Tokenization #	1	2	3	4	5	6
accuracy	91.21	92.09	92.87	90.13	92.38	93.75
precision	0.92	0.91	0.93	0.91	0.91	0.93
recall	0.92	0.94	0.94	0.91	0.95	0.96
F_measure	0.92	0.93	0.93	0.91	0.93	0.94

Table 2. NB with 10 fold cross validation, using different tokenization setups

#	1	2	3	4	5	6
accuracy	92.28	91.5	93.17	90.82	91.89	93.56
precision	0.92	0.92	0.92	0.92	0.93	0.95
recall	0.94	0.93	0.95	0.92	0.92	0.93
F_measure	0.93	0.92	0.94	0.92	0.92	0.94

Table 3. SVM with 10 fold cross validation, linear u*v, using different tokenization setups

3. Creating Wordpress plugin

A plugin for wordpress was created to use the model files created in step 2.

When the Sentiment Analysis admin page is viewed by the administrator, the plugin queries the wordpress database for comments related to a particular product. The plugin run these comments through the classifier file on the web server, and the prediction results is read by the plugin and displayed on the screen. Figure1.

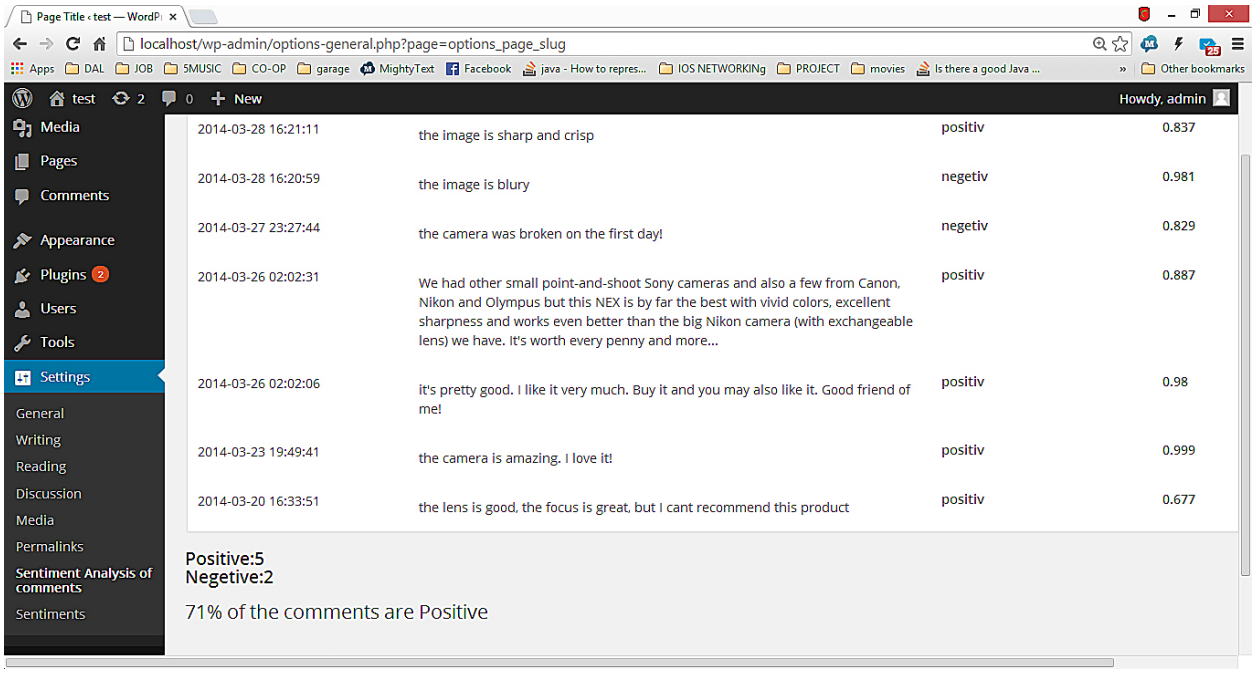
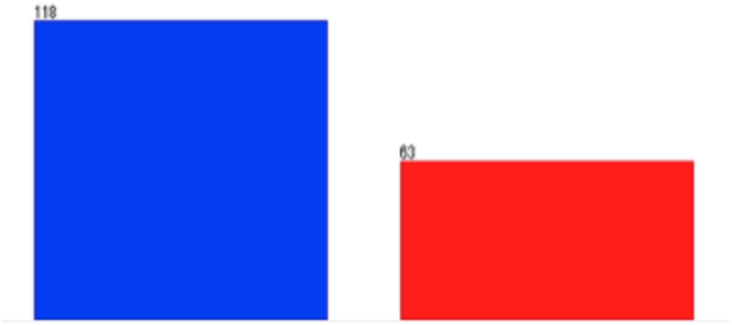


Figure 1. Sentiment Analysis Wordpress Plugin

4. Evaluation of Classifiers

Aside from evaluating the classifier on training data, the two classifiers were also evaluated using testing data. The testing data consists of 181 unseen camera reviews collected from Amazon website. 1, 2 stars reviews labeled as negative, and 4, 5 stars

reviews labeled as positive. The total reviews are 118 positive and 63 negative reviews. Figure 2. Both classifiers then were reevaluated on these test data using setup #6 as tokenization method. When using unseen test data, the SVM classifier performs better with 79.01%. accuracy (Table 7.) compare to 72% accuracy of NB classifier. See Table 4



test data positive (blue) and negative(red) reviews Figure2

Correctly Classified Instances	**131**	**72.38%**
Incorrectly Classified Instances	**50**	**27.62%**
Total Number of Instances	**181**	

Table 4. NB, accuracy using test data, setup #6 for tokenization

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.619	**0.079**	**0.936**	**0.619**	**0.745**	**positive**
	0.921	**0.381**	**0.563**	**0.921**	**0.699**	**negetive**
Weighted Avg	**0.724**	**0.184**	**0.806**	**0.724**	**0.729**	

Table 5. NB, F-measure using test data,

a	b	<-- classified
73	45	a = positive
5	58	b = negetive

Table 6. NB, confusion matrix, using test data

a	b	<-- classified
73	45	a = positive
5	58	b = negetive

Table 6. NB, confusion matrix, using test data

Correctly Classified Instances	143	79.01%
Incorrectly Classified Instances	38	20.99%
Total Number of Instances	181	

Table 7. SVM, accuracy using test data, setup #6 for tokenization

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.72	0.079	0.944	0.72	0.817	0.82	positive
	0.921	0.28	0.637	0.921	0.753	0.82	negetive
Weighted Avg	0.79	0.149	0.838	0.79	0.795	0.82	

Table 8. SVM, F-measure using test data

a	b	<-- classified
85	33	a = positive
5	58	b = negetive

Table 9. SVM confusion matrix, using test data

Findings

The best tokenization technique is set up #6 with no IDF Transform and stemmer, but using Term Frequency normalization, lowercase, document normalization, a list of stopwords and alphabetic tokenizer. SVM performs better on this particular text classification task even though the model creation time was longer. While using the training data as test data with 10 fold cross validation, accuracy of NB and SVM are almost the same with 93.75 and 93.56 respectively. However, when using unseen test data the SVM classifier performs better with 79.01% accuracy compare to 72% accuracy of NB classifier.

Conclusion

Wordpress is content management system, millions of websites are currently using it including many ecommerce websites. The only Sentiment Analysis plugin currently available for wordpress is only designed to classify a general text and does not perform well when it comes to domain specific text. we built a plugin that does Sentiment Analysis but our method uses a crawler to scrape relevant training data. with a few modifications this method can be used to create different classifiers to classify any content type, also the Wordpress plugin can be used on any Wordpress website.

We created two classifiers for classification of Camera reviews, one using SVM and one using NB. we found that SVM and Naive Bayes have similar accuracy when they were evaluated on training data with 10 fold cross validation. The accuracy was 93.75% for NB, and 93.56% for SVM. In contrast, when evaluated on unseen test data SVM had accuracy of 79.01% compare to NB that had an accuracy of 72%.

Reference

[1] Filipa P., and Jao M. “Opinions in User Reviews: An Evaluation of Sentiment Analysis Techniques.” Department of Information, Faculty of Science and Technology, University of Nova de lesboa, Portugal, 2013.

[2] Hanhoon K., Seong J. Y. and Dongil H. (2012, Apr.). “Senti-lexicon and improved Naive Bayes algorithms for sentiment analysis resturant reviews.” *Expert Systems with Applications.* [On-Line]. 39(5), pp. 6000-6010 Available: http://www.sciencedirect.com.ezproxy.library.dal.ca/science/article/pii/S0957417411016538, [Mar. 23, 2014].

[3] Neethu, M.S. and Rajasree, R., "Sentiment analysis in twitter using machine learning techniques," *Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference on* , vol., no., pp.1,5, 4-6 July 2013
doi: 10.1109/ICCCNT.2013.6726818
URL:
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6726818&isnumber=6726462

[4] Peter J. and Isabelle M., *Natural Language Processing for Online Applications.* John Benjamins publishing company, 2007, [E-Book]. Available: http://www.library.dal.ca/Files/How_do_I/pdf/IEEE_Citation_Style_Guide.pdf

[5] Qiang Y., Ziqiong Z. and Rob L. (2009, Apr.). “Sentiment Classification of online reviews to travel destinations by supervised machine learning approaches.” *Expert Systems with Applications.* [On-Line]. 36(3), pp. 6527-6535. Available: http://www.sciencedirect.com.ezproxy.library.dal.ca/science/article/pii/S0957417408005022, [Mar. 22, 2014].

[6] Singh, V.K.; Piryani, R.; Uddin, A.; Waila, P., "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification,"*Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on* , vol., no., pp.712,717, 22-23 March 2013
doi: 10.1109/iMac4s.2013.6526500
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6526500&isnumber=6526372

[7] Waila, P.; Marisha; Singh, V.K.; Singh, M.K., "Evaluating Machine Learning and Unsupervised Semantic Orientation approaches for sentiment analysis of textual reviews," *Computational Intelligence & Computing Research (ICCIC), 2012 IEEE International Conference on* , vol., no., pp.1,6, 18-20 Dec. 2012
doi: 10.1109/ICCIC.2012.6510235
URL:
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6510235&isnumber=6510164