DEGREE PROJECT IN TECHNOLOGY,
FIRST CYCLE, 15 CREDITS
STOCKHOLM, SWEDEN 2018

Sentiment classification on Amazon reviews using machine learning approaches

SEPIDEH PAKNEJAD

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

KTH Computer Science
and Communication

Sentiment klassificering på Amazon recensioner med hjälp av maskininlärningstekniker

Sepideh Paknejad

Degree Project in Computer Science, DD142X
Supervisor: Richard Glassey
Examiner: Örjan Ekeberg
CSC, KTH June 2018

Abstract

As online marketplaces have been popular during the past decades, the online sellers and merchants ask their purchasers to share their opinions about the products they have bought. As a result, millions of reviews are being generated daily which makes it difficult for a potential consumer to make a good decision on whether to buy the product. Analyzing this enormous amount of opinions is also hard and time consuming for product manufacturers. This thesis considers the problem of classifying reviews by their overall semantic (positive or negative). To conduct the study two different supervised machine learning techniques, SVM and Naïve Bayes, has been attempted on beauty products from Amazon. Their accuracies have then been compared. The results showed that the SVM approach outperforms the Naïve Bayes approach when the data set is bigger. However, both algorithms reached promising accuracies of at least 80%.

Sammanfattning

Eftersom marknadsplatser online har varit populära under de senaste decennierna, så har online-säljare och inköpsmän ställt kunderna frågor om deras åsikter gällande varorna de har köpt. Som ett resultat genereras miljontals recensioner dagligen vilket gör det svårt för en potentiell konsument att fatta ett bra beslut om de ska köpa produkten eller inte. Att analysera den enorma mängden åsikter är också svårt och tidskrävande för produktproducenter. Denna avhandling tar upp problemet med att klassificera recensioner med deras övergripande semantiska (positiva eller negativa). För att genomföra studien har två olika övervakade maskininlärningstekniker, SVM och Naïve Bayes, testats på recensioner av skönhetsprodukter från Amazon. Deras noggrannhet har sedan jämförts. Resultaten visade att SVM-tillvägagångssättet överträffar Naïve Bayes-tillvägagångssättet när datasetet är större. Båda algoritmerna nådde emellertid lovande noggrannheter på minst 80%.

Contents

1 Introduction **4**
1.1 Problem Statement 4
1.2 Outline of the report 5

2 Background **6**
2.1 Sentiment classification and analysis 6
2.2 Sentiment classification using Machine learning methods 6
2.2.1 SVM 8
2.2.2 Naïve Bayes 8
2.2.3 Feature extraction 8
2.3 Sentiment classification using Lexicon based methods 9
2.4 Related work 10

3 Method **12**
3.1 Programming environments 12
3.2 The data set 12
3.2.1 Data preparation 13
3.3 Machine learning classifiers 13

4 Result **15**
4.1 First experiment 15
4.2 Second experiment 15

5 Discussion **17**
5.1 Limitations and further study 17

6 Conclusion **19**

References **20**

1 Introduction

As online marketplaces have been popular during the past decades, the online sellers and merchants ask their purchasers to share their opinions about the products they have bought. Everyday millions of reviews are generated all over the Internet about different products, services and places. This has made the Internet the most important source of getting ideas and opinions about a product or a service.

However, as the number of reviews available for a product grows, it is becoming more difficult for a potential consumer to make a good decision on whether to buy the product. Different opinions about the same product on one hand and ambiguous reviews on the other hand makes customers more confused to get the right decision. Here the need for analyzing this contents seems crucial for all e-commerce businesses.

Sentiment analysis and classification is a computational study which attempts to address this problem by extracting subjective information from the given texts in natural language, such as opinions and sentiments. Different approaches have used to tackle this problem from natural language processing, text analysis, computational linguistics, and biometrics. In recent years, Machine learning methods have got popular in the semantic and review analysis for their simplicity and accuracy.

Amazon is one of the e-commerce giants that people are using every day for online purchases where they can read thousands of reviews dropped by other customers about their desired products. These reviews provide valuable opinions about a product such as its property, quality and recommendations which helps the purchasers to understand almost every detail of a product. This is not only beneficial for consumers but also helps sellers who are manufacturing their own products to understand the consumers and their needs better.

This project is considering the sentiment classification problem for online reviews using supervised approaches to determine the overall semantic of customer reviews by classifying them into positive and negative sentiment. The data used in this study is a set of beauty product reviews from Amazon that is collected from Snap dataset (Leskovec & Sosic, 2016).

1.1 Problem Statement

Sentiment classification aims to determine the overall intention of a written text which can be of admiration or criticism type. This can be achieved by using machine learning algorithms such as Naïve Bayes and Support Vector Machine. So, the problem that is going to be investigated in the project is as follow:

Which machine learning approach performs better in terms of accuracy on the Amazon beauty products reviews?

1.2 Outline of the report

The rest of the thesis is structured as follow:
Section 2, the Background, consists of essential definitions and theory to understand the other sections of this thesis. It also introduces related work done in this area of research. This is followed by the Methods, in section 3, where the procedure of the study has been described. The results from the experiments are gathered in section 4 and discussed in section 5. Finally, section 6 concludes the study.

2 Background

2.1 Sentiment classification and analysis

Electronic commerce is becoming increasingly popular due to the fact that e-commerce websites allow purchasers to leave reviews on different products. Millions of reviews are being generated everyday by costumers which makes it difficult for product manufacturers to keep track of customer opinions of their products. Thus, it is important to classify such large and complex data in order to derive useful information from a large set of data. Classification methods are the way to tackle such problems. Classification is the process of categorizing data into groups or classes based on common traits (Pandey et al. 2016; Rain 2013). A common concern for organizations is the ability to automate the classification process when big datasets are being used (Liu et. al 2014).

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) problem which means identifying and extracting subjective information of text sources. The purpose of sentiment classification is to analyze the written reviews of users and classify them into positive or negative opinions, so the system does not need to completely understand the semantics of each phrase or document (Liu 2015; Pang et. al 2002; Turney & Littman 2003).

This however is not done by just labeling words as positive or negative. There are some challenges involved. Classifying words and phrases with prior positive or negative polarity will not always work. For example, the word "amazing" has a prior positive polarity, but if it comes with a negation word like "not", the context can completely change (Singla et. al 2013).

As Ye et. al (2009) state the word "unpredictable" camera has a negative meaning to that camera while "unpredictable" experience is considered as positive for tourists.

Sentiment classification has been attempted in different fields such as movie reviews, travel destination reviews and product reviews (Liu et al. 2007; Pang et al. 2009; Ye et al. 2009). Lexicon based methods and machine learning methods are two main approaches that are usually used for sentiment classification.

2.2 Sentiment classification using Machine learning methods

There is a large number of papers that have been published in the field of machine learning. One of the most used approaches for sentiment classification is machine learning algorithms. This section attempts to cover some of them.

One of the first definitions of machine learning that has been provided by Tom Mitchell (1997) in his book Machine Learning is as follow:
"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

Machine learning aims to develop an algorithm in order to optimize the performance of the system by using example data. The solution that machine learning provides for sentiment analysis involves two main steps. The first step is to "learn" the model from the training data and the second step is to classify the unseen data with the help of the trained model (Khairnar & Kinikar 2013). Machine learning algorithms can be classified in different categories:
a. supervised learning
b. semi-supervised learning
c. unsupervised learning

a. In **Supervised Learning** the process where the algorithm is learning from the training data can be seen as a teacher supervising the learning process of its students (Brownlee 2016). The supervisor is somehow teaching the algorithm what conclusions it should come up with as an output. So, both input and the desired output data are provided. It is also required that the training data is already labeled. If the classifier gets more labeled data, the output will be more precise. The goal of this approach is that the algorithm can correctly predict the output for new input data. If the output were widely different from the expected result, the supervisor can guide the algorithm back to the right path.

There are however some challenges involved when working with supervised. The supervised learning works fine as long as the labelled data is provided. This means that if the machine faces unseen data, it will either give wrong class label after classification or remove it because it has not "learnt" how to label it (Cunningham et al. 2008).

b. **Unsupervised learning** in difference with supervised learning is trained on unlabeled data with no corresponding output. The algorithm should find out the underlying structure of the data set on its own. This means that it has to discover similar patterns in the data to determine the output without having the right answers. One of the most important methods in unsupervised learning problems is clustering. Clustering is simply the method of identifying similar groups of data in the data set (Kaushik 2016).

For sentiment classification in an unsupervised manner it is usually the sentiment words and phrases that are used. This means that the classification of a review is predicted based on the average semantic orientation of the phrases in that review (Turney 2002). This is obvious since the dominating factor for sentiment classification is often the sentiment words (Berk 2016). This technique has been used in Turney's study (2002).

c. Finally, **Semi-supervised learning** which has the benefit of both supervised and unsupervised learning, refers to problems in which a smaller amount of data is labelled, and the rest of the training data set is unlabeled. This is useful for when collecting data can be cheap but labelling it can be time consuming and expensive. This approach is highly favorable both in theory and practice because of the fact that having lots of unlabeled data during the training process tends to improve the accuracy of the final model while building it requires much less time and cost (Zhu 2005). In Dasgupta and Vincent Ng. (2009) a semi-supervised learning was experimented where they used 2000 documents as unlabeled data and 50 randomly labeled documents.

2.2.1 SVM

Support vector machines (SVM) are supervised learning method that can be used for solving sentiment classification problems (Cristianini & Shawe-Taylor 2000). This technique is based on a decision plane where labeled training data is placed and then algorithm gives an optimal hyperplane which splits the data into different groups or classes. As seen in figure 1. the best hyperplane is the one that separates the classes with the largest margin. This is achieved by choosing a hyperplane so that its distance from the nearest data on each class is maximized (Berk 2016).

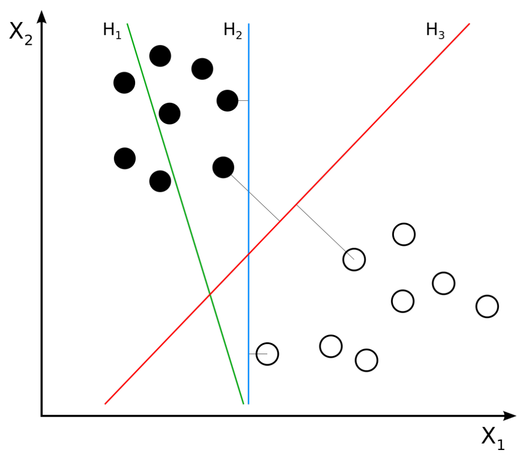

Figure 1: "H1 does not separate the classes. H2does, but only with a small margin. H3separates them with the maximum margin." (Zack Weinberg 2012)

2.2.2 Naïve Bayes

Naïve Bayes is another machine learning technique that is known for being powerful despite its simplicity. This classifier is based on Bayes theorem and relies on the assumption that the features (which are usually words in text classification) are mutually independent. In spite of the fact that this assumption is not true (because in some cases the order of the words is important), Naïve Bayes classifiers have proved to perform surprisingly well (Rish 2001). The first step that should be carried out before applying the Naïve Bayes model on text classification problems is feature extraction.

2.2.3 Feature extraction

Since machine learning algorithms work only with fixed-length vector of numbers rather than raw text, the input (in this case text data) need to be parsed. The method for transforming the texts into features is called the Bag of words model of text, which is a commonly used method of feature extraction. The approach works by creating different bags of words that occur in the training data set where each word is associated with a unique number. This number shows the occurrence of each word in the document. A simple illustration of the Bag of words model can be seen in figure 2. The model is called a bag of words because the position of the words is in the document discarded (Raschka 2014).

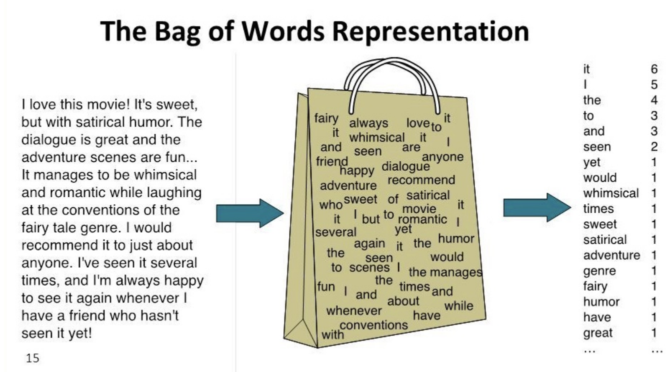

Figure 2: A simple illustration of the Bag of words model. Figure: (Prabhat Kumar Sahu, 2017)

2.3 Sentiment classification using Lexicon based methods

Lexicon based method is another unsupervised approach, which relies on word and phrase annotation. To compute a sentiment score for each text, this method uses a dictionary of sentiment words and phrases (Taboada et al. 2011). In lexicon-based methods the simplest approach for determining the sentiment of a review document where is to use a count-based approach. If we have a text and a lexical resource containing the positive and negative annotation of words and phrases, we can assign the polarity of the review. This means that if the number of positive words is more than the negative ones the polarity of the review is positive. If there are more negative sentiment words than positive sentiment words, the overall sentiment of the text is then negative (Mukherjee 2017).

However, using only sentiment words and phrases for sentiment classification is not enough. Sentiment lexicon for sentiment analysis is necessary but it is not sufficient. There are some issues involved with this method that Liu (2012) argues about:

1. positive or negative sentiment words may have different interpretation in different domains. For instance, the word "suck" usually have a negative sentiment, but it can also indicate a positive sentiment. For example, "This camera sucks," expresses a negative opinion but "This vacuum cleaner really sucks." is a positive opinion.

2. Sarcastic sentences are usually hard to deal with even if they contain sentiment words. For example, "What a great car! It stopped working in two days." This is a negative opinion even though it contains the word "good" which is a positive word.

3. it can happen that a phrase or opinion does not have a sentiment word, so it makes it hard for the machine to determine the compute a sentiment score for the opinion. "This washer uses a lot of water" has no sentiment words but it implies a negative opinion about the washer.

Hu and Liu, (2004), work is one of the first studies on this method. They proposed a lexicon-based algorithm for sentiment classification. Since they believed that a review usually contains some sentences with negative opinions and some sentences with positive opinions, they performed classification at the sentence level. For each sentence they identify if it is expressing a positive or negative opinion and then a final summary of the review is produced.

2.4 Related work

Due to the proliferation of online reviews, Sentiment analysis has gained much attention in recent years. Therefore, many studies have been devoted to this research area. In this section, some of the most related research works to this thesis are presented.

Joachims (1998) experimented SVM for text classification and showed that SVM performed well in all experiments with lower error levels than other classification methods.

Pang, Lee and Vaithyanathan (2002) tried supervised learning for classifying movie reviews into two classes, positive and negative with the help of SVM and Naïve Bayes and maximum entropy classification. In terms of accuracy all three techniques showed quite good results. In this study they tried various features and it turned out that the machine learning algorithms performed better when bag of words was used as features in those classifiers.

In a recent survey that was conducted by Ye et al. (2009), three supervised machine learning algorithms, Naïve Bayes, SVM and N-gram model have been attempted on online reviews about different travel destinations in the world. In this study, they found that in terms of accuracy, well trained machine learning algorithms performs very well for classification of travel destinations reviews. In addition, they have demonstrated that the SVM and N-gram model achieved better results than the Naïve Bayes method. However, the difference among the algorithms reduced significantly by increasing the number of training data set.

Chaovalit and Zhou (2005) compared the supervised machine learning algorithm with Semantic orientation which is an unsupervised approach to movie review and found that the supervised approach provided was more reliable than the unsupervised method.

According to many research works, Naïve Bayes, SVM are two most used approaches in sentiment classification problems (Joachims 1998; Pang et al. 2002; Ye et al. 2009). This thesis, therefore tries to apply supervised machine learning algorithms of Naïve Bayes and SVM to the beauty product reviews of Amazon website.

3 Method

This section presents the method of the study. The programming environments will be described in the first part. How and where the data was gathered as well as the data preparation approach will be discussed in the second part. In the last part, the procedure of machine learning classifiers will be explained.

3.1 Programming environments

Python is one of the most widely used programming language in machine learning and data science. Python has a huge set of libraries that can be used for solving various machine learning algorithms. The programming language used in this study is Python because of its wealth of libraries and ease of use. Scikit-learn is one of many libraries in Python that features a variety of supervised machine learning algorithms (Pedregosa et al. 2011). It provides different classification techniques such as SVM, Naïve Bayes. It also offers techniques for feature extraction.

3.2 The data set

The first step for conducting the research includes data collection for training and testing the classifiers. The data is collected from SNAP data set because Amazon does not have an API like Twitter to download reviews with. The format of the downloaded file was one-review-per-line in JSON. The file was converted to the Comma Separated Values (CSV) format, as it is more convenient for python to handle this type of files. The data set consists of 252000 reviews of different beauty products. Each review includes nine features as follow:

Feature	**Description**
reviewerId	Id of the user
"asin"	productId
reviewerName	name of the user
helpful	fraction of users who found the review helpful
reviewText	text of the review
overall	rating of the product
summary	review summary
unixReviewTime	time of the review in unix time
reviewTime	date of the review

Table 1: Describes features of each product

The following is an example review in Json file:

{"reviewerID": "A4DIVP0NEQDA3",
"asin": "B000052WYD",
"reviewerName": "Amazon Customer",
"helpful": [1, 1],
"reviewText": "This product met my needs for Upper Lip shadow. I found it to be non oily, easy to smooth over skin and light on skin. You will need to use a press power for your skin tone after applying the concealer. Very please with the results.",
"overall": 5.0,
"summary": "Magic Stick for Upper Lip Shadow",
"unixReviewTime": 1386806400,
"reviewTime": "12 12, 2013"}

3.2.1 Data preparation

For preparing the desired data a simple code was written in python to remove the useless features. Many features were removed except the summary of the review, the text of the review itself, score and productId. The score that is generated by the reviewer includes a number of stars on scales of 1 to 5. Reviews that were rated with one or two stars were considered as negative and those with four or five stars were considered as positive. Reviews with three stars usually contain many mixed reviews and are difficult to be labeled into a positive or negative category.

In this study, two experiments have been conducted. In **the first experiment** the whole data set was used. Since the number of reviews were quite enough to get a reasonable result from the classifiers the reviews with three stars were omitted to avoid any complication while training the algorithms.

However, in **the second experiment** due to the small number of data the reviews with three stars were also considered as negative.
The same code was then used to label the data. The reviews that were considered as positive got a score of "1" and the remaining ones got a "0" score.

3.3 Machine learning classifiers

To carry out the experiments, each classifier algorithm needs to be trained before being tested. In order to train and use the classifiers, the data was divided into two data sets as training and testing data sets. As mentioned earlier, two experiments have been conducted in this research. In each experiment, the classifiers were trained and tested once on the reviews itself and once on the review summaries.

For **the first experiment** a corpus of 150000 data were collected as training data set and the remaining 48500 for testing the accuracy of the classifiers. The next step was to transform the review texts into numerical features before being fed to the algorithms. This was done by using the Bag of words model. The third step was to train the Naïve Bayes and SVM classifiers. The last step was to apply the trained classifiers on the test data to measure their performance by comparing the predicted labels with the actual labels that have not been given to the algorithms. Figure 3 shows an illustration of the wholr procedure.

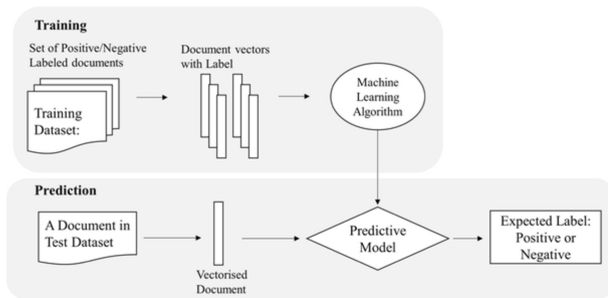

Figure 3: A basic illustration of the sentiment classification by supervised machine learning algorithms. Figure: (Choi & Lee, 2017)

For **the second experiment** a smaller number of data were chosen. The reviews were grouped by the productId and the first 10 products with most reviews, as shown in table 2, were chosen to apply the algorithms on them. 300 data were then collected from each product as training data set. The rest of the procedure was the same as the first experiment.

productId	**Nr of reviews**
B004OHQR1Q	431
B0043OYFKU	403
B0069FDR96	391
B000ZMBSPE	389
B00150LT40	329
B003V265QW	328
B006L1DNWY	321
B008U1Q4DI	310
B007BLN17K	305
B000142FVW	302

Table 2: First 10 products with most reviews

4 Result

This section presents the results of the study. The accuracy value shows the percentage of testing data set which were classified correctly by the model. The accuracy of two different machine learning algorithms on 2 sets of experiments are shown in tables 3, 4 and 5.

4.1 First experiment

The results from the first experiment are shown in Table 3 , where it shows the accuracy of Naïve Bayes and SVM on the whole data set both on the reviews and the summaries.

	Naïve Bayes	**SVM**
On reviews	90.16%	93.02%
On summaries	92.72%	93.20%

Table 3: The accuracy of the machine learning methods on the whole data set

Both algorithms have achieved accuracies over 90% for both cases, although SVM got better results. The Naïve Bayes approach has got better result when applied on the summaries.

4.2 Second experiment

Table 4 and Table 5 show the results from the second experiment where the classifiers were trained and tested on the first 10 products with the most reviews and similar to the first experiment once on the reviews and once on the summaries.

productId	**Naïve Bayes**	**SVM**
B004OHQR1Q	89.98%	89.43%
B0043OYFKU	89.98%	89.15%
B0069FDR96	89.98%	84.13%
B000ZMBSPE	89.98%	84.12%
B00150LT40	89.98%	89.41%
B003V265QW	89.98%	83.09%
B006L1DNWY	89.98%	86.55%
B008U1Q4DI	89.98%	87.96%
B007BLN17K	88.97%	83.02%
B000142FVW	89.98%	89.25%

Table 4: The accuracy of the machine learning methods on 10 products using the reviews itself

productId	Naïve Bayes	SVM
B004OHQR1Q	89.98%	88.73%
B0043OYFKU	90.11%	84.90%
B0069FDR96	90.02%	82.93%
B000ZMBSPE	90.04%	85.23%
B00150LT40	89.97%	86.50%
B003V265QW	89.91%	82.43%
B006L1DNWY	89.96%	85.03%
B008U1Q4DI	89.99%	87.32%
B007BLN17K	88.55%	81.54%
B000142FVW	89.98%	85.53%

Table 5: The accuracy of the machine learning methods on 10 products using the summaries

In the second set of experiments, where the number of reviews is much smaller than the first experiment, the Naïve Bayes method achieved better accuracy than the SVM in both cases. However, the accuracy is still quite good ranging from 80% to 90%.

5 Discussion

The main goal of this study was to determine which machine learning algorithm of SVM and Naïve Bayes methods performs better in the task of text classification. This was accomplished by using the Amazon beauty products as data set. The classifiers were evaluated by comparing their accuracies in different cases of experiments.

The overall accuracies of two machine learning algorithms in different experiments are shown in tables 3, 4 and 5. The results from the first set of experiments shown in table 3 indicated that the SVM approach got better accuracy than the Naïve Bayes in both cases where the algorithms applied on the reviews and when they have been applied on the summaries. The difference in accuracies between these approaches is however very small.

From this experiment it can be found out that well trained machine learning algorithms with enough data as training data set can perform very good classification. In terms of accuracies, SVM tends to do better than Naive Bayes, although the differences aren't very large, and the algorithms can reach more than 90% of classification correctly.

The results from the second set of experiments shown in Table 4 and 5 indicated that the Naïve Bayes approach had better accuracy than the SVM in both cases. In this experiment the data set was much smaller than the previous experiment, 300 reviews from 10 different products.

A possible explanation for this difference can be the size of the data sets. In the first experiment the size of the training data set is much bigger than the second experiment. This can lead to an assumption that the SVM model works better with more data. Fang and Zhan (2015) had a similar result in their experiment. In their paper they tackle the problem of sentiment analysis on Amazon reviews by using the Naïve Bayes, SVM and Random Forest model. They demonstrated that when the models get more training data the SVM model outperforms the other classifiers.

Another reason can be the reviews with 3 stars which usually are categorized as neutral. However, In the second experiment due to the small size of data set they were considered as negative. This could have affected the result of the second experiment. In future study it would be interesting to categorize them as positive to see whether it gives a different result.

5.1 Limitations and further study

For conducting the experiments in this thesis, no pre-processing has been done on the data set. Pre-processing is the process where the data is being cleaned and prepared before being fed to the algorithms. Online reviews usually contain many irrelevant and uninformative features which may not even have an impact on the orientation of them. This process involves removing many steps such as white space removal and stop words removal etc.

The results from all experiments implies that both approaches give higher accuracies when they are being applied on the summaries of the reviews. The possible explanation for this result might be the nature of the reviews. The reviews itself contain a large amount of words, which can lead to sparsity in bag of words features. As a result we see that the accuracies of the algorithms for all experiments are higher when applied on the summaries which are more informative and contain limited number of words.

In this study the text reviews were not pre-processed before being fed to the classifiers. However, Haddi et al. (2013) show in their study that pre-processing the data can significantly enhance the classifier's performance. Their experiment demonstrated that the accuracy of a classifier like SVM can be improved by using appropriate pre-processing methods. Further studies thus should investigate whether pre-processing the training data set would improve the results.

Another limitation is that the bag of words model does not care about the positioning of the words in a text which can have a negative effect on the semantic of a review. For instance, in the following example although the overall sentiment of the review is negative, a machine would classify the review as positive due to the number of positive words in contains (Pang et al. 2002).

"This film sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up."

This is the concept of "thwarted expectations" which is fairly common in reviews and has been recognized by Pang et. al (2002) and Turney (2002) who noted that "the whole is not necessarily the sum of the parts".

There is also another challenge that should be addressed in the problem of sentiment classification which is identification of negation and its effect on the semantic understanding of sentences. Future studies could fruitfully explore this issue further by developing approaches to tackle this issue.

In this study just two main machine learning algorithms have been examined. Future research could continue to explore another efficient sentiment classifiers like Decision tree etc.

6 Conclusion

This study has applied two different machine learning algorithms of SVM and Naïve Bayes on the Amazon beauty products reviews. The results from the study showed that in terms of accuracy the SVM approach achieves better results than the Naïve Bayes approach when the whole data set was used as training and testing data set. As the number of reviews decreased the Naïve Bayes method achieved better performance than the SVM method.

References

[1] Richard A Berk. *Statistical learning from a regression perspective.* Springer, 2016.

[2] Jason Brownlee. Supervised and unsupervised machine learning algorithms, Mar 2016.

[3] Pimwadee Chaovalit and Lina Zhou. Movie review mining: A comparison between supervised and unsupervised classification approaches. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*, pages 112c–112c. IEEE, 2005.

[4] Nello Cristianini and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods.* Cambridge university press, 2000.

[5] Pádraig Cunningham, Matthieu Cord, and Sarah Jane Delany. Supervised learning. In *Machine learning techniques for multimedia*, pages 21–49. Springer, 2008.

[6] Sajib Dasgupta and Vincent Ng. Topic-wise, sentiment-wise, or otherwise?: Identifying the hidden dimension for unsupervised text classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, pages 580–589. Association for Computational Linguistics, 2009.

[7] Xing Fang and Justin Zhan. Sentiment analysis using product review data. *Journal of Big Data*, 2(1):5, 2015.

[8] Emma Haddi, Xiaohui Liu, and Yong Shi. The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, 17:26–32, 2013.

[9] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.

[10] Thorsten Joachims. Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*, pages 137–142. Springer, 1998.

[11] SAURAV KAUSHIK. An introduction to clustering and different methods of clustering, Nov 2016.

[12] Jayashri Khairnar and Mayura Kinikar. Machine learning algorithms for opinion mining and sentiment classification. *International Journal of Scientific and Research Publications*, 3(6):1–6, 2013.

[13] Jure Leskovec and Rok Sosič. Snap: A general-purpose network analysis and graph-mining library. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(1):1, 2016.

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

[15] Bing Liu. *Sentiment analysis: Mining opinions, sentiments, and emotions.* Cambridge University Press, 2015.

[16] Bingwei Liu, Erik Blasch, Yu Chen, Dan Shen, and Genshe Chen. Scalable sentiment classification for big data analysis using naive bayes classifier. In *Big Data, 2013 IEEE International Conference on*, pages 99–104. IEEE, 2013.

[17] Jingjing Liu, Yunbo Cao, Chin-Yew Lin, Yalou Huang, and Ming Zhou. Low-quality product review detection in opinion summarization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 2007.

[18] Tom M Mitchell et al. Machine learning. wcb, 1997.

[19] Subhabrata Mukherjee. Sentiment analysis of reviews.

[20] Priyank Pandey, Manoj Kumar, and Prakhar Srivastava. Classification techniques for big data: A survey. In *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on*, pages 3625–3629. IEEE, 2016.

[21] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics, 2002.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[23] Callen Rain. Sentiment analysis in amazon reviews using probabilistic machine learning. *Swarthmore College*, 2013.

[24] Sebastian Raschka. Naive bayes and text classification i-introduction and theory. *arXiv preprint arXiv:1410.5329*, 2014.

[25] Irina Rish. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46. IBM, 2001.

[26] Zeenia Singla, Sukhchandan Randhawa, and Sushma Jain. Statistical and sentiment analysis of consumer product reviews. In *Computing, Communication and Networking Technologies (ICCCNT), 2017 8th International Conference on*, pages 1–6. IEEE, 2017.

[27] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.

[28] Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

[29] Peter D Turney and Michael L Littman. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21(4):315–346, 2003.

[30] Qiang Ye, Ziqiong Zhang, and Rob Law. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert systems with applications*, 36(3):6527–6535, 2009.

[31] Xiaojin Zhu. Semi-supervised learning literature survey. 2005.