ISSN 2581 – 3781

AJCSE
ASIAN JOURNAL OF COMPUTER SCIENCE ENGINEERING

Available Online at www.ajcse.info
Asian Journal of Computer Science Engineering 2018;3(4):17-21

RESEARCH ARTICLE

Analysis of Supervised and Unsupervised Learning Classifiers for Online Sentiment Analysis

B. Usharani

Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

Received on: 31-08-2018; Revised on: 30-09-2018; Accepted on: 05-10-2018

ABSTRACT

Sentiment analysis is also known as opinion mining which it extracts opinions to learn about public point of view. In general, people prefer to take advice from others not only to get sensible products but also to invest in a wise way. Nowadays, the popular sources of personal opinion are blogs. The consumer review sites are helpful to know about the features and services available to a specific product. This sentiment analysis is also helpful to the manufactures to improve their business by knowing the public opinion. The online sources gather public opinion content in the form of blogs, forums, social media, review websites, etc. The sentiment analysis automatically mines the huge content that is available online. This paper gives a complete study of various supervised and unsupervised classifiers.

Key words: Machine learning, online reviews, sentiment analysis, supervised, unsupervised

INTRODUCTION

Present days, people want to know what they are getting into before they make purchase decisions. People are seeking out and curious what other people personal experiences to help them to decide what product to choose and what is right for them. For example, before spending a vacation, people want to know from other families about the local attractions and amenities that are providing by the hotels. People are searching for advice before purchasing anything online, and they feel that these reviews give them confidence, sense of security and credibility. For manufacturers the consumer reviews are a crucial source to evaluate the business decisions.

Customers do trust and engage with online reviews, and these reviews are very influential to make purchase decisions. These days the online reviews might be the best marketing tool, and these online reviews push a customer from consideration to purchase. The online review informs others about the customer's experience and helps other people to make decisions and let the business management know about the customers experience about a product. Traditional marketing and advertising are an excellent way to create awareness about the product, but it went down these days, people want to find the business only through online reviews.

The online reviews impact purchase, the customer decision on what to buy comes from the reviewers who share their experience and these reviews pulls the customers and informs them about their decision-making. The manufacturers manage the online reputation by marketing their products effectively based on the information what the customers receive from online reviews will help them to deal with the business. The negative reviews give an opportunity to do service to the customer, responses, and actions taken to correct the issue will show a potential customer that the manufacturer's true care about the feedback they receive.

The feedback is taken from the customers form various ways such as blogs, forums, social media (Facebook, LinkedIn, and Twitter), and review websites. After taking the feedback the automatic filtering of and analyzing of each word was calculated using the sentiment analysis approach. From this analysis, the words are categorized as positive, negative, or neutral. The detailed flow of the process is shown in diagrammatic form in Figure 1.

RELATED WORK

Tsytsarau and Palpanas[1] illustrates the sentiment analysis definition, problem, categorization, and development with the help of tables and graphs.

Address of correspondence:
B. Usharani
E-mail: ushareddy.vja@gmail.com

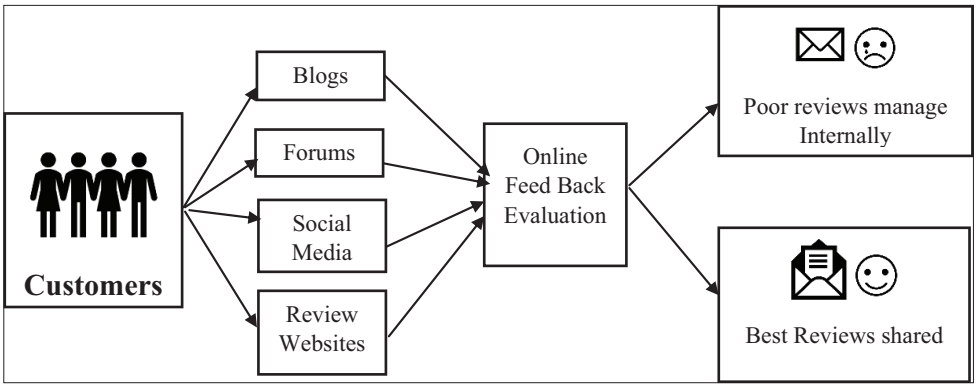

Figure 1: Online feed back flow

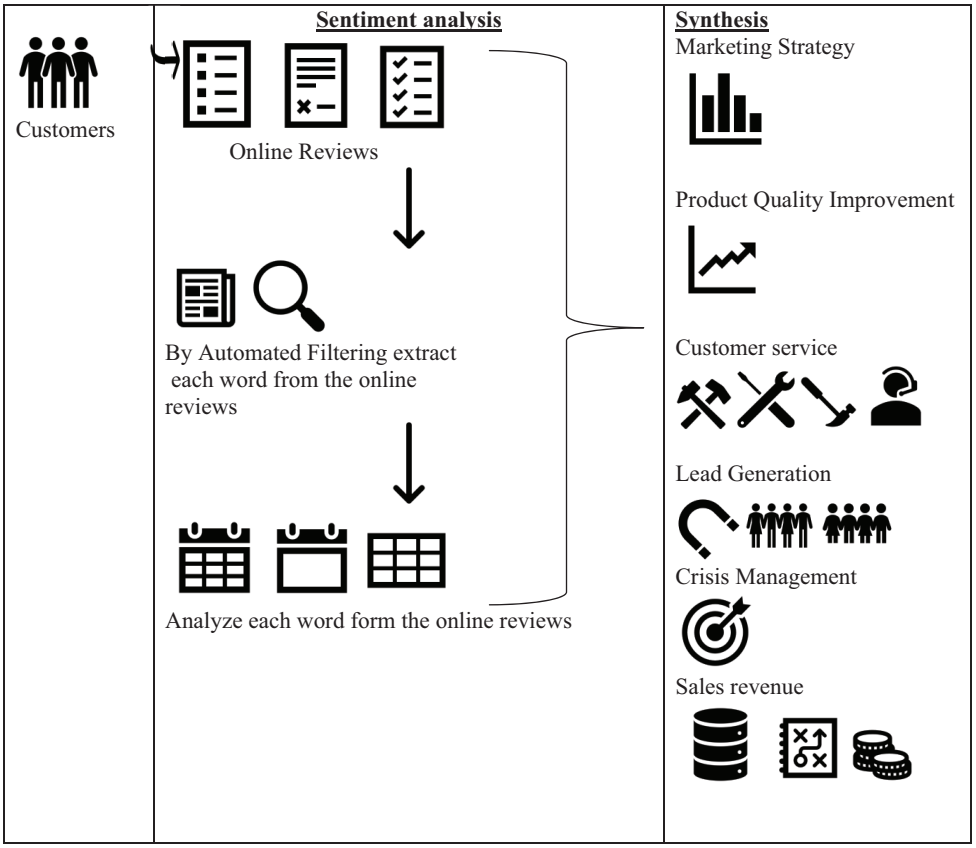

Figure 2: Sentiment analysis flow

The surveys[2,3] discussed the problem of sentiment analysis from the application point of view and focused on the applications and challenges in sentiment analysis and mentioned the techniques to solve each problem in sentiment analysis.

The surveys[4-6] illustrate the trends in sentiment analysis.

Li and Wu[7] takes the sequence of words and calculated the sentiment score of each keyword based on the dictionary and its privative and modifier near it.

Balahur[8] replaced the sentiment word and modifiers by sentiment labels such as positive, negative, high positive, and high negative and then applied support vector machine sequential minimal optimization to classify data.

Nakagawa *et al*.[9] used a probabilistic model of the information gathered from the dependency tree to determine the sentiment of the sentence.

Moilanen and Pulman[10] presented a combination model for three classes, i.e., positive, negative, and neutral for phrase level and sentence level sentiment analysis.

SENTIMENT ANALYSIS

The sentiment analysis is the process of identifying and categorizing opinions expressed as a form of text to determine the customer opinion toward a specific product, i.e. positive, negative, and neutral. Social media are an excellent source of

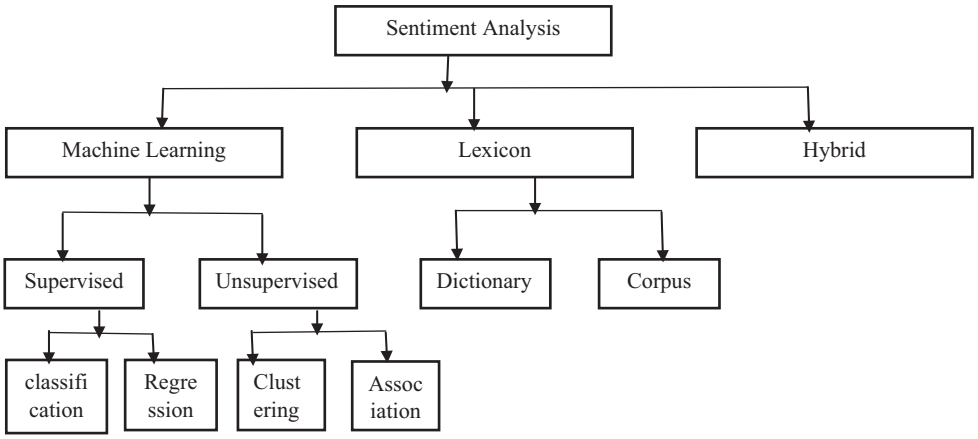

Figure 3: Sentiment classification diagram

Table 1: supervised versus unsupervised learning

Supervised	Unsupervised
Labeled data	No labeled data
Direct feedback	No direct feedback
Predict outcome/future	Find hidden structure in data
Training	
input → output	input → output
X2, X1	clusters, X2, X1
Algorithms: Naive Bayes Support vector machine Linear regression Artificial neural networks Decision Tree	Algorithms: Nearest neighbor Random forest Logistic regression Apriori Clustering Fuzzy
Applications: Biometrics Database marketing Handwriting recognition Information extraction Information retrieval Pattern recognition Speech recognition Spam detection	Applications: Data compression Data decompression Data Visuallization Project prioritization and selection Meteorology and oceanography Failure mode and effects analysis Creation of artwork Seismic facies analysis

information for sentiment analysis to provide perceptions to improve campaign success, determine marketing strategy, improve customer services, improve product messaging, and so on.

Most vendors take care that their sentiment analysis algorithm should be accurate otherwise by making decisions using incorrect sentiment analysis can be catastrophic.

Table 2: Confusion matrix

N-Measure	Actual	
	Positive	**Negative**
Predicted		
Positive	TP	FP
Negative	FN	TN

TP: True positive, FP: False positive, FN: False negative, TN: True negative

The sentiment analysis flow and the synthesis that was calculated from the sentiment analysis is shown in Figure 2.

Types of sentiment analysis

1. Manual processing: Human interpretation of the sentiment must be accurate.
2. Keyword processing: Assign positivity or negativity to individual words and calculates the overall percentage score to the post.
3. Natural language processing (NLP): Also called text analytics, computational linguistics. NLP is superior to keyword processing. NLP works by analyzing language for its meaning.

The information what the vendors get from sentiment analysis provides them to improve their marketing strategy. By sentiment analysis, the vendor can see the positive or negative discussions among their audience. By sentiment analysis, the vendors know the customer's opinions about their products or services, the quality, and features of their products. The products are not judged by their functionality, instead of how well it is presented on the online reviews.

Sentiment analysis can be measured using three approaches. They are machine learning, lexicon based, and hybrid-based approaches. In the machine learning approach, the supervised learning model

can be easily trained, and the unsupervised model can be easily categorized the data. The lexicon-based approach can be easily calculate the sentiment scores for each word. The hybrid is a combination of both machine learning and lexicon-based approaches and measures the sentiment for noisy and less sensitive data. The classification is given in Figure 3.

The sentiment analysis can be divided into different categories as shown in Figure 3.

Sentiment analysis deals with identifying opinion patterns and presenting them in a way that is easy to understand. The outcome of the sentiment analysis can be in the form of categorizing opinions such as like or dislike about a product in an online review. Sentiment analysis has various sub-streams such as bias analysis and emotion detection.

MACHINE LEARNING APPROACH

The machine learning-based techniques are supervised, unsupervised, and semi-supervised. The supervised technique uses labeled data; the unsupervised data use clustering of data. Table 1 gives the difference between the supervised and the unsupervised approaches.

A machine learning algorithm that maps the input data into a category is called classifier.

Calculating the performance of a classifier

The performance of a classifier can be calculated with the help of the confusion matrix given in Table 2.

Some metrics are given below:

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{false positive}}$$

$$\text{Sensitivity (or) recall} = \frac{\text{True positive}}{\text{True positive} + \text{false negative}}$$

$$\text{F1} = 2*\frac{\text{Precision*Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Accuracy} = \frac{\text{True positive} + \text{true negative}}{\text{True positive} + \text{false positive} + \text{false negative} + \text{ture negative}}$$

$$\text{Specificity} = \frac{\text{True negative}}{\text{True negative} + \text{false positive}}$$

$$\text{Mean absolute error} = \frac{1}{N}\sum_{n=1}^{N}\left(a_n - b_n\right)$$

$$\text{Mean squared error} = \frac{1}{N}\sum_{n=1}^{N}\left(a_n - b_n\right)^2$$

$$\text{Mathews correlation coefficient} = \frac{TP*TN - FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Eg:

***n*-=165**	**Actual**	
	Positive	**Negative**
Predicted		
Positive	50	10
Negative	5	100

Precision = 0.91
Recall = 0.95
Accuracy = 0.90
F_1 = 0.92

CONCLUSION

This paper presented an overview of sentiment analysis, classification techniques on machine learning. Machine learning techniques such as supervised and unsupervised techniques have their own pros and cons. The social media play a major role to express public opinion. We will apply these both supervised and unsupervised techniques to various online review datasets to analyze sentiment.

REFERENCES

1. Tsytsarau M, Palpanas T. Survey on mining subjective data on the web. Data Min Knowl Discov 2012;24:478-514.
2. Liu B. Sentiment Analysis and Opinion Mining. San Rafael: Morgan and Claypool Publishers; 2012. p. 1-168.
3. Pang B, Lee L. Opinion mining and sentiment anlaysis. Found Trends Inf Retr 2008;2:1-135.
4. Cambria E, Schuller B, Xia Y, Havasi C. New avenues in opinion mining and sentiment analysis. IEEE Intell Syst 2013;28:15-21.
5. Feldman R. Techniques and applications for sentiment analysis. Commun ACM 2013;56:82-9.
6. Montoyo A, Barco PM, Balahur A. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. Decis Support Syst 2012;53:675-9.
7. Li N, Wu DD. Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decis Support Syst 2010;48:354-68.
8. Balahur A. Sentiment Analysis in Social Media Texts. Proceedings of the 4th Workshop on Computational

Approaches to Subjectivity, Sentiment and Social Media Analysis; 2013. p. 120-8.
9. Nakagawa T, Inui K, Kurohashi S. Dependency Tree Based Sentiment Classification Using CRFs with Hidden Varaibles, Human Language Technologies: The 2010 Annual Conference of the North America Chapter of ACL; 2010. p. 786-94.
10. Moilanen K, Pulman S. Sentiment Composition. The Oxford Computational Linguistics Group. Proceedings of RANLP; 2007. p. 1-5.