CUSTOMER SATISFACTION MEASUREMENT TOOL BY ANALYSING TURKISH PRODUCT REVIEWS

Zeynep Orhan,
Süleyman Fatih Giriş,
Migena Ceyhan
Fatih University
zorhan@fatih.edu.tr,
fatih.giris@fatih.edu.tr,
migena.ceyhan@fatih.edu.tr

Elton Domnori
Epoka University
edomnori@epoka.edu.al

ABSTRACT

Sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The aim of this study is classifying the text in social media with context analysis. Unlike the past, nowadays people are interacting via social networks. These analyses have potential of wide spread applicability. We examined the customer reviews of the products and determined the polarity of the reviews by considering the language features of the customers. The ultimate goal of this study is to extract the features of the humans' attitudes, polarity etc. towards any kind of topic by considering the language used in the product reviews. This study automatically classifies positive and negative comments using ML methods with high accuracy. The promising results indicate very high accuracy.

1 Introduction

Sentiment analysis refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Text categorizations according to affective relevance, opinion exploration for market analysis are examples of applications which use these techniques.

Emotion detection, subjectivity analysis, opinion mining, referred also as sentiment analysis, attitude and appraisal analysis or review mining [1], are some of the tasks developed in the field of Natural Language Processing (NLP). All of the above mentioned research fields are considered as part of the wider area of research in artificial intelligence (AI) of Affective Computing [2].

Sentiment Analysis' main target is to detect the expression of sentiments in text and classify them according to their polarity into separate categories (negative or positive), according to their semantic orientation.

The attitude of a person may be his or her judgment or evaluation, affective state, or the intended emotional communication. Opinion Mining is a process of automatic extraction of knowledge from the opinion of others about some particular topic or problem. The idea of Opinion Mining and Sentiment Analysis is to process a set of search results for a given item, generating a list of product attributes, as quality, features etc. and aggregating opinion.

Pang and Lee define this problem as the binary classification task of labeling an opinionated document as expressing either an overall positive or an overall negative sentiment. Expressions of sentiment are related to expressions of emotions in text. Emotions in themselves are complex phenomena, which haven't yet been given a generally accepted definition. There exist many definitions from different points of view, but none of them is fully sufficient [1]. Sentiment is a personal belief or judgment that is not founded on proof or certainty [3]. Although these affect-related phenomena have traditionally been studied in depth by disciplines such as Philosophy or Psychology, due to advances in computing and the growing role of technology in everyday life, the past decades have shown an increasing interest in building software systems that automatically process affect. In order for systems to benefit from the knowledge acquired in social sciences, interdisciplinary methods have been proposed, which use the existing theoretical models as basis for engineering computational ones.

People are expressing themselves freely in social network and millions of messages appear daily in popular web-sites for micro-blogging such as Twitter, Tumblr, Facebook, therefore they are the best places to start the analysis of people's feelings. Many analyses done based on data provided from internet have potential of widespread applicability to various fields ranging from the detection of psychological disorders to profit and loss analysis of the market for various productions.

2 Experimental Setup

It is easy for human beings to detect the feelings of the interlocutor from the person's mimics, voice tonality, etc. Nevertheless, in textual data it is harder to detect this kind of information and the meaning is derived from just a list of words. As a consequence, it is really hard to teach a computer how to understand the mood of a writer of a given text. Therefore, automatic detection of attitudes, polarity, etc. of data has gained an important attention recently.

The outline of the processes followed in this study is shown in Figure 1. Each step is interpreted further in this section.

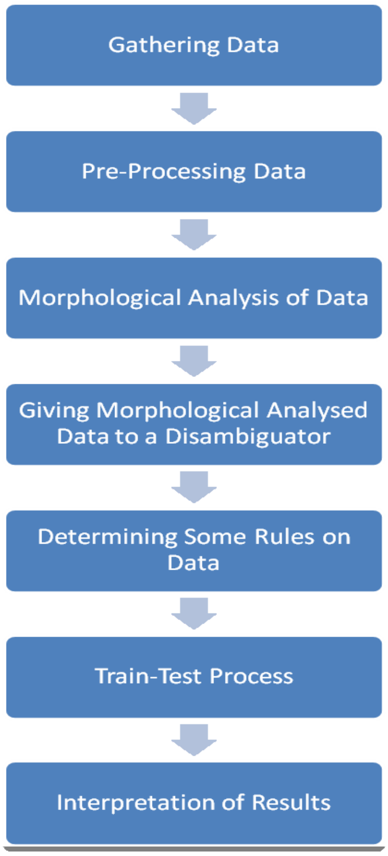

Figure-1. The outline of the processes

2.1. Gathering Data

One of the most important steps of this study is gathering data. Its aim is to classify the people's opinions as positive or negative. For this reason data from two different web sites of which we know for sure the polarity were gathered, according to the assumption that the positive content was extracted from the product reviews that brands allowed to display in their web sites [4], while the negative ones were obtained from a complaints site [5].

First, data was downloaded by means of the tool SiteSucker [6], which organizes them in a hierarchical way, from roots to leaves. The output of this process is files in html format. Next, comments were extracted from the raw text of the downloaded files by means of some php scripts. The latter were then saved in new files of text format. For ethical issues all personal information, such as user names were replaced by their identity numbers, which became the file names, while brand and product names were hidden. Figure 2 shows the content of one comment file.

> Ben xxx telefonu xxx vasıtasıyle aldım. Ürünün cam dokunmatik yüzeyi soyuldu. Defelarca xxx vasıtasıyle tamir için göndersemde ilgilenilmedi. Eğer böyle devam ederse bir daha xxx ürünlerini almayacağım. Aradığımda sanki ilk defa telefon kullanıyormuşum gibi kullanıcı hatası diyorlar. Bu sorunumun çözülmesini talep ediyorum.

Figure-2. The raw data format of one comment

All files are organized as positive or negative and grouped into relevant folders. A total of 2497 positive and 392 negative comments are collected.
To balance the inequality of positive and negative data, the positive data was divided into groups of 392 for the training and testing process, and their average results were calculated and displayed in this study.

2.2. Pre-Processing Data

The extracted data was further modified; sentences were separated according to punctuation marks, even though they are not very correctly used in social media comments. The raw text data was prepared as input for the morphological analyzer, which included each sentence between <S><S> and <\S><\S> tags, and each word was displayed in a separate line, as shown in Table 1a. Data in this format was then supplied to morphological analyzer and morphological disambiguator tools [7].

2.3. Morphological Analysis of Data

Morphological analysis is a simple creative method of forced association of attributes which simply divides the words into root and suffixes in possible formats. There could be more than one option or type for one word. The example output of morphological analysis is shown in Table 1b. Generally, the morphological analysis result yields multiple and ambiguous results about each word, which brings the necessity of morphological disambiguation to select the correct one [8].

2.4. Disambiguating Morphological Analyzed Data

Disambiguation is a natural language processing application that tries to determine the intended meaning of a word or phrase by examining the linguistic context [8]. The output after the disambiguation process is displayed in Table 1c.

2.5. Determining some Rules on Data

After the disambiguation process, the roots of the words are extracted, so that they can be counted by the system. However, some more preprocessing is done beforehand to differentiate the words when they are used in their positive or negative forms, or when their meaning is strengthened by means of quantity adverbs.

Words Indicating Negativity: In Turkish, there are some words which convert the meaning of the previous words or the ones associated to them to its opposite. The most widely used negation words included in this study are *değil*, meaning *not*, effecting mainly adjectives, verbs, adverbs and nouns. Another group of words is *ne ... ne ...*, referring to *neither ... nor ...* in English, effecting the words coming after each of the clauses preceded by *ne*.

Suffixes Indicating Negativity: The most used suffixes for negation in Turkish are *–me* and *-mez*. They are used mainly in verbs or nouns derived from verbs to negate their meaning.

Words Indicating Quantity: In this study, some of the quantity indicating words in Turkish which reduce or strengthen the meaning of the associated words like *çok*, referring to *very* or *a lot*, *az* meaning *a little* and *en* used to form the superlative of adjectives are the most used. Considering their quantity, the affected words' frequencies are arranged accordingly.

Words and suffixes indicating negativity yield a new word, which is counted separately from the original word, i.e., in the sentences "Çayı çok *sıcak* severim" and "Bugün hava *sıcak* ***değil***", even though the word *sıcak*, meaning *hot*, appears twice in these sentences, they are counted as if they were two different words.

Similarly, quantity indicating words multiply or reduce the frequency of the effected words, assuming that the same word is used two or more times depending on the quantity indicating word. So, the word *sıcak* is counted differently in the previous examples.

Table-1. Sample data ready to pre-process by the morphological analyzer and Morphological disambiguator and their respective results

<S>	<S>	<S>		<S> <S>
Ben	Ben	Ben	+Noun+Prop+A3sg+Pnon	Ben ben+Pron+Pers+A1sg
xxx	Ben	ben	+Noun+A3sg+Pnon+Nom	xxx xxx+Noun+A3sg+Pnon
telefonu	Ben	ben	+Pron+Pers+A1sg+Pnon+Nom	telefonu telefon+Noun+A3sg
xxx	xxx	xxx	+Noun+A3sg+Pnon+Nom	xxx xxx+Noun+A3sg+Pnon+Nom
vasıtasıyle	telefonu	telefon	+Noun+A3sg+P3sg	vasıtasıyle
aldım	telefonu	telefon	+Noun+A3sg+Pnon	vasıtası+Noun+A3sg+Pnon+Ins
</S>	xxx	xxx	+Noun+A3sg+Pnon+Dat	aldım al+Verb+Pos+Past+A1sg
	xxx	xxx	+Noun+A3sg+Pnon+Nom	</S> </S>
	vasıtasıyle		vasıtası +Noun+A3sg+Pnon+Ins	
	vasıtasıyle		vasıtasıy +Noun+A3sg+Pnon+Ins	
	vasıtasıyle		vasıtasıyl +Noun+A3sg+Pnon+Dat	
	vasıtasıyle		vasıtasıyle +Noun+A3sg+Pnon+Nom	
	aldım	al	+Adj^DB+Verb+Zero+Past	
	aldım	al	+Verb+Pos+Past	
	</S>	</S>		
a.Data in <S><\S> tags	b. Morphological analyzer output			c. Morphological Disambiguator output

2.6. Train-Test Process

The features used in the experiments were extracted by the content analysis of the textual data. The roots of the words obtained from the morphological disambiguation process served to determine the features of the training data.

Three methods were applied. The first method uses the binary information of the existence or nonexistence of a word in the prepared set of features for each subject (Table 2). The second method provided also information about the frequency of the usage of each word, while the third one added to the frequency also information about the weight of each word. When frequency was used, normalization was also applied by dividing the calculated frequency by the text size of the comment.

Table-2. Training data including all distinct words as columns and binary values indicating word existence in comments

Id	W_1	W_2	W_3	…	W_n	Class (N:Neg P:Pos)
1	0	1	0	…	1	N
2	1	0	1	…	1	P
3	1	0	0	…	0	P

2.7. Interpretation of Results

Classification is achieved using ML (Machine Learning) methods provided by a data mining tool named Weka. Weka provides a collection of many well-known ML algorithms for data mining tasks. It is a very powerful tool for data pre-processing, clustering and classification processes. It provides results of train and test data across many data mining algorithms [9]. In this study for the classification of text, BayesNet (BN), a decision tree algorithm J48 and VotedPerceptron (VP) algorithms were used.

BayesNet is an algorithm which uses various search algorithms and quality measures to improve Bayes Network learning. It is a base class for a Bayes Network classifier and it provides data structures (network structure, conditional probability distributions, etc.) and facilities common to Bayes Network learning algorithms like K2 and B [10].

The method called *J48* is a class for generating a pruned or unpruned C4.5 desicion tree. C4.5 and its predecessor, ID3, use formulas based on information theory to evaluate the *goodness* of a test; in particular, they choose the test that extracts the maximum amount of information from a set of cases, given the constraint that only one attribute [11].

VotedPerceptron algorithm is implemented by Freund and Schapire. It globally replaces all missing values, and transforms nominal attributes into binary ones [12].

The accuracy of the classification of texts by using each of the methods when applied to the abovementioned ML techniques are grouped in Table 5, while their graphical visualization is shown in Figure 3.

Table-3. The accuracy percentages obtained from different algorithms

Methods	Binary	Frequency	Weighted
BayesNet	94.9	92.4	93.2
J48	86.3	84.5	84.4
Voted Perceptron	93.6	91.2	90.2

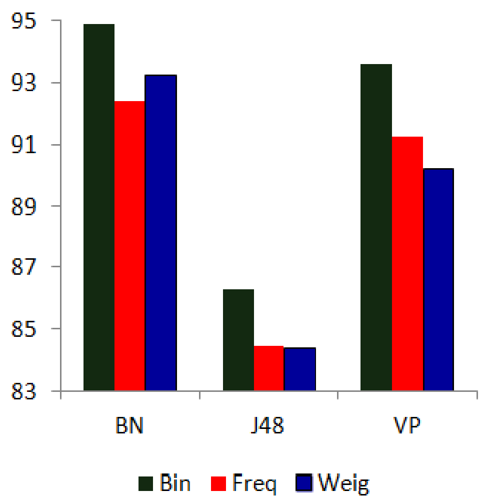

Figure-3. Graphical view of the accuracy percentages obtained from different methods

3. Conclusion

We examined the customer reviews of the products and tried to determine the polarity of the reviews by considering the language features of the customers.

Negative and positive reviews are collected from two different web sites and these are automatically classified by using ML methods with high accuracy.

These results support that language usage is an important issue about obtaining valuable information related with the opinions, attitudes, and emotions of the people on social media.

The methodology of this system uses all distinct words of the text as the feature according to their existence, frequency or weighted frequency. These features are given to different algorithms provided by Weka library, such as BayesNet, a decision tree algorithm J48 and VotedPerceptron algorithm.

The ultimate goal of this study is to extract the features of the humans' attitudes and polarity towards any kind of topics by considering the language used in the product reviews.

These analyses have potential of wide spread applicability especially to achieve profit and loss analysis of the market for various goals.

The results are very promising and to the extent of our knowledge as being one of the initiators of Turkish sentiment analysis and opinion mining.

4. References

[1] PANG, B., AND LEE, L., "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval, Vol 2, No 1-2, pp 1-135, 2008.

[2] PICARD, R. W., "Affective computing: challenges." International Journal of Human-Computer Studies 59.1 (2003): 55-64.

[3] BALAHUR, A., HERMIDA, J.M., AND MONTOYO, A. Detecting implicit expressions of emotion in text: A comparative analysis. Decision Support Systems, 2012,53.4:742-753.

[4] Samsung Reviews Website,http://reviews.tr.samsung.com/7463-tr_tr/allreviews.htm, Last accessed:December, 2013

[5] Turkish Complement Website, www.sikayetbank.com, Last accessed: December, 2013

[6] Web Page Downloader,http://www.sitesucker.us/home.html, Last accessed: December, 2013

[7] Turkish Language Resources,http://www.denizyuret.com/2006/11/turkish-resources.html, Erişim Tarihi: Haziran 2014.

[8] OFLAZER. K, 1993. Two-level description of Turkish morphology. In Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics, April. A full version appears in Literary and Linguistic Computing, Vol.9 No. 2, 1994.

[9] Witten, I. H., AND FRANK, E., Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, 2 edition, 2005.

[10] BEN-GAL I., Bayesian Networks, in Ruggeri FALTIN F., F. & KENETT R., Encyclopedia of Statistics in Quality & Reliability, Wiley & Sons, 2007.

[11] QUINLAN J. R., *C4. 5: programs for machine learning*. Vol. 1. Morgan Kaufmann, 1993.

[12] FREUND Y., AND SCHAPIRE R. E.: Large margin classification using the perceptron algorithm. In: 11th Annual Conference on

Computational Learning Theory, New York, NY, 209-217, 1998.