Advances in Environmental Biology, 8(13) August 2014, Pages: 53-68

AENSI Journals

Advances in Environmental Biology

ISSN-1995-0756 EISSN-1998-1066

Journal home page: http://www.aensiweb.com/AEB/

A Comparative Study on Sentiment Analysis

[1]Alireza Yousefpour, [1]Roliana Ibrahim, [3]Haza Nuzly Abdull Hamed, [4]Mohammad Sadegh Hajmohammadi

Faculty of Computing, Universiti Teknologi Malaysia (UTM), 81310 Skudai, Johor, Malaysia

ARTICLE INFO

Article history:
R Received 25 April 2014
Received in revised form
8 May 2014
Accepted 20 July 2014
Available online 18 August 2014

Keywords:
Sentiment Analysis,
Feature Selection,
Sentiment Classification

ABSTRACT

With the genesis of the internet and the world wide web, we have seen an unprecedented growth of data and information on the web as well as a huge growth in digital or textual opinions, sentiments and attitudes as remarked upon in reviews. Special attention needs to be given to the processing and understating information used by information retrieval methods and natural language processing methods. One of the main problems in this scope is sentiment analysis whereby a review is classified into two classes, i.e. positive (thumbs up or favorable) or negative (thumbs down or unfavorable) opinions. This paper discusses related issues such as feature selection methods and different sentiment classifications as well as the main approaches currently being taken to solve this problem.

ToCite ThisArticle: A. Yousefpour, R. Ibrahim, H.N.A. Hamed, M.S. Hajmohammadi, A comparative study in sentiment analysis. ***Adv. Environ. Biol., 8(13),*** *53-68, 2014*

INTRODUCTION

Nowadays, both individuals and organizations need to make use people's public opinions and sentiments with regard to decision-making about their products. With the advent of Web 2.0, together with widespread internet and social media (such as social networks, reviews, comments, twitter and forum discussions on the web), focus can be given in order to collect public opinions, since there is such a rich amount of general information available. Thus, processing and extracting information and opinions on the web and then analyzing them is a powerful task. Sentiment analysis applications are now extracted to roughly every possible area, such as: services, financial services, political elections and customer products.

Sentiment analysis or opinion mining is a field of computational study that processes opinions, attitudes, sentiments, emotions, and appraisals of people concerning namely: products, movies, entities, events, issues, topics and their respective features. There are different names for this area of study, such as: sentiment mining, review mining, text mining, opinion extraction, subjectivity analysis, emotion analysis, etc. We know them primarily by the terms sentiment analysis or opinion mining. In fact, sentiment analysis basically means analyzing opinions in order to detect whether they are of a positive or negative opinion. The aim of sentiment analysis and opinion mining is to identify attitudes and emotions of a person about certain subjects including movies or products, etc. In fact, opinion mining is a process which draws out subjective information from a text corpus or reviews; while sentiment analysis is the evaluation of the information that is extracted. More recent researches studies into sentiment analysis have presented different techniques by which to extract and analyze sentiments.

One of the key issues is the identification of a semantic relationship among subject and sentiment-related expressions, which may quite different relevance on their relationships. Sentiment expressions such as adjectives, adverbs, nouns, sentiment verbs and transfer verbs have been used for sentiment analysis. The use of features including: natural language processing, POS tagging, as well as the Markov-Model-Based tagger model has essentially assisted in the demystification of some multi-meaning expressions such as "like". This can denote a verb, adjective or preposition and syntactic parsing for identifying relationships among an expression and subject term respectively [33].

Corresponding Author: Roliana Ibrahim, Computing of Faculty, Universiti Teknologi Malaysia, 81310 UTM, Skudai, Johor, Malaysia.
E-mail: roliana@utm.my

Background of The Problem:

However, the sentiment analysis may be the first research to become manifested by Nasukawa [33]and Dave [10], when the concept of opinion mining first appeared. At that time, there was already some extant research relating to sentiments and opinions. Nasukawa [33] took this a step further by describing a sentiment analysis approach to specific negative and positive opinions concerning particular subjects from a document, instead of classifying all of the documents. They focused on identifying the semantic relevancies among both expressions and subjects that can improve the accuracy of the analysis. However, there were some works already in existence by Pang [37] and Turney [52] which proposed sentiment classification tasks. Pang and Turney introduced two different approaches in their articles. Pang [37] exploited a supervised approach while Turney [52] proposed an unsupervised approach. In a supervised learning approach, a classifier is first trained based on a large feature set of labeled data; this classifier is then used to identify and classify unlabeled test data into two classes (or more) of positive and negative sentiments. Some researchers used a set of several features to improve the classification accuracy [28, 35, 63-64, 68-69]. In addition, other researchers proposed a shallow parsing by which to select appropriate feature sets [54]. Other feature selection techniques were used to achieve better performance such as Information Gain by Ye and Keogh [60]. Further, some works also tried to combine several approaches with hybrid classifiersPrabowo and Thelwall [39]. Most of the existing studies define sentiment classification as a supervised classification problem and try to train a classifier from a large amount of labeled data [37, 39, 66]. The main disadvantage of supervised methods is the difficulty involved in preparing and annotating a large amount of training data.

Parallel to this, some of the researches studies have been undertaken based on unsupervised learning methods [18, 20, 48, 52]. Hence, they used sentiment lexicon to identify and classify documents to polarity sentiment. This was attained through calculating the term “sentiment orientation” using a dictionary or part-of-speech patterns by exploiting a search engine to calculate the association of words with a known polarity seed set. These works are known as unsupervised learning methods and are strongly dependent on sentiment lexicons.

Labeled data and sentiment lexicon are considered to be key resources as data training in sentiment classification task, and we can find them on the web in some domain. However, they lack training data in other domains as the data of each domain are special and different. Gaining unlabeled data is easy and is freely available on different blogs, while obtaining labeled data is expensive because it is usually done by human beings. To solve this problem, it is necessary to have a powerful learning algorithm which is capable of determining and including a great deal of a lot of unlabeled data based on the training of small labeled data. As a result, one of the problems that arises is how can use low-labeled data be used for training while not losing accuracy. This challenge leads to an interesting research area namely semi-supervised learning. The goal of semi-supervised learning is to enable higher accuracy of sentiment classification via the use of both labeled and unlabeled data, instead of using only labeled data in supervised learning.

There are several classifiers in semi-supervised learning. For example, [14-15, 46] proposed a graph-based semi-supervised learning algorithm to rate and label unlabeled documents via calculation of the similarity rate between two documents. Also, [35] calculated the statistical relationships existing between target variables i.e. subjectivity, sentiment polarity, and the will to influence.

A significant amount of research studies have been conducted in the area of sentiment analysis via a single classifier however, some of the works have used multi-classifiers and their combinations (known as classifier ensemble in sentiment analysis) [11, 47, 64]. Prabowo and Thelwall [39] proposed a hybrid classification via the use of the following: combination rule-based classifiers, SVM algorithms and statistics based classifiers, e.g. document frequency, mutual information, Chi-square, and Log likelihood ratio used a rule set built by using 240 sentiment lexicons, consisting of 120 negative words and 120 positive words. Xia [55] proposed an ensemble framework that included two different feature sets, namely: part-of-speech and word-relation based sets of features. Further, they used three algorithms in base-level, i.e. naive bayes, maximum entropy, and SVM to classify each of the feature sets. In addition, they utilized three kinds of methods for the combination of classifiers, namely: weighted and fixed combinations and meta-classifiers. Su, Zhang *et al*., 2013, introduced a classifier ensemble that including five popular classifiers namely Naive Bayes (NB), Centroid-Based classification (CB), K-nearest neighbor algorithms (KNN), Support Vector Machine (SVM), and Maximum Entropy model (ME) respectively as base-level algorithms. They also employed used stacking generalization in high-level or meta-level areas.

Sentiment Analysis:

Sentiment analysis is also known as opinion mining; it is a field of computational study that processes opinions, attitudes, sentiments, emotions, and appraisals of people concerning products, movies, entities, events, issues, topics and related features. There are different names for this area of study, such as: sentiment mining, review mining, text mining, opinion extraction, subjectivity analysis, and emotion analysis. We know them by names of sentiment analysis or opinion mining. In fact, sentiment analysis is basically a process of analyzing opinions in order to detect positive or negative opinion.

Applications of Sentiment Analysis:

Liu [29] proposed a sentimental model for forecasting of sales performance by the use of weblogs. In fact, weblogs provide a wide variety of information that can be helpful in both decision-making and for ascertaining the general public's opinions and sentiments. Asur [5] demonstrate that social media, in the form of twitter, can be used to forecast real-world results. They are used to predict box-office incomes for movies. As a result, they have shown that sentiment analysis can improve predictions of the power of social media. Yano [59] designed and evaluated a model that evaluates comment volume and contexts of political weblogs. In fact, their goal is to be able to forecast the rate of comment volume on new weblogs. Tumasjan [51]used this method to predict an election outcome in German federal election by using twitter. They wished to identify state that twitter can be a valid source by which to verify decisions and political sentiments. Mohammad [31] have shown that emotional differences between the sexes influence the use of the following sentiments in many types of emails. In fact, they can be used to quantify the following sentiments in the mail by extracting sentiment lexicon and comparing emotions of hate and love found in self-murder emails by sentiment analysis. Bollen [6] investigated the forecasting of stock markets by analysis of textual content of large-scale twitter mood. Pestian [38] posited that sentiment analysis was used to study movements in reference to suicide notes. Padmaja [36] stated that using techniques of opinion mining, linguistic analysis and machine learning methods helped when studying the measurement of peoples' belief. Kechaou [21] present a novel method for clustering of video news and media text based on sentiment analysis. Understanding and recognition of the needs of cancer patients helped to provided answers for other patients in using sentiment analysis in [34].

Different levels in sentiment analysis:

There are three levels in sentiment analysis, specifically: document level, sentence level and entity and aspect level.

Document Level:

At document level, classification of all opinion documents that can have negative or positive sentiments is called document-level sentiment classification. This includes, for example, prediction of negative and positive opinions on product reviews under all comments and sentiments that are written or spoken by an opinion holder. Hence, it cannot be used for comparing and evaluating between different entities. At this level, it is assumed that the document represents an entity or a product.

Sentence level:

In the sentence level, classification is performed by examining each sentence to evaluate whether the sentence consists of a neutral, negative or positive sentiment. Classification on a sentence level is done in two steps. The first step consists of sentences being divided into objective or subjective categories; thus, subjective sentences being classified to a positive or negative sentiment. As a result, sentences can take two forms, either objective or subjective. Hence, subjectivity does not relate to sentiment, but an objective sentence can imply decision. However, document and sentence levels can be used to determine opinion, but they cannot detect definitely what people either like or don't like. For example, [30] used a graph model that at the end links up to several sentence level sentiments. In a study by Nakagawa *et al*., the expanded pars tree of a sentence by conditional random field model was used for identifying positive and negative sentiments.

Aspect level:

Aspect-base or feature-base levels are based on ideas and opinions, composed of either negative or positive sentiments. They result in an opinion target that helps us to better notice of sentimental analysis problems. Many customers think that they can make better decisions when selecting and buying products based on reading the experiences of others in reviews on the web. Although customers can still make efficient decisions by studying product reviews, due to the increasing amount of product reviews appearing every moment on the web, reading and deciding entirely by means of relevant reviews is very difficult for customers. As a result, in order to solve the above problem, a good solution can be found by summarizing reviews. There are several approaches to summarizing evaluations. Liu [27] named this as aspect-base sentiment analysis to determine an accurate summary from reviews. There are two steps in this context as shown below:

Aspect (feature or entity) Extraction: feature-base sentiment analysis uses natural language processing methods to automatically extract aspects in documents. *Aspect Orientation:* predicting the orientation of each aspect in a sentence must classify to an orientation, i.e. positive sentiment, negative sentiment or neutral. For example, Godbole [14] proposed a sentiment lexicon based on WordNet approach and associated sentiment. They extracted sentiment lexicon as an aspect by which to express the sentiment of a relevant sentence. By assuming that there are syntactic relationships between sentiment word and entities, Qiu [40] introduce a propagation method by which to extract these relationships and express their sentiments. Table 1 shows some pervious research in a different level.

Table 1: Selected Previous Studies in Sentiment analysis

Paper	*Dataset*	*Level*
(Pang, Lee *et al*. 2002)	Movie Review	Document Level
(Turney 2002)	Movie Review, Travel Review, Bank Review, Automobile Reviews	Document Level
(Prabowo and Thelwall 2009)	Movie Review, Product Reviews	Document Level
(Taboada, Brooke *et al*. 2011)	Movie Review, Camera Reviews	Document Level
(McDonald, Hannan *et al*. 2007)	Product Review (Car, Fitness, mp3)	Sentence Level
(Nakagawa, Inui *et al*. 2010)	Japanese blogs, Movie Review, Product Review	Sentence Level
(Yi, Nasukawa *et al*. 2003)	Product Review, Music Review	Aspect Level
(Hu and Liu 2004)	Product Review	Aspect Level
(Qiu, Liu *et al*. 2011)	Product Review	Aspect Level

Sentiment Lexicon:

Sentiment words, also known as “opinion words” or “opinion-bearing words” are used in many researches for sentiment classification. These words show both positive and negative sentiments. For example, words such as excellent, good, amazing and cool are positive sentiment words; while words with negative sentiment include weak, faint, bad and terrible. A set of these words is called an opinion lexicon or a sentiment lexicon [27]. Although sentiment lexicons can be important and necessary for analysis, using them can sometimes present problems. Some of these problems are expressed as below:

a. Some words may have different meanings when used in different ways. For example, the word “suck” in the sentence, “X vacuum cleaner sucks well” has a positive sentiment, whereas the sentence “Y camera sucks” is negative.

b. There are many sentences that can show sentiment without a sentiment lexicon. For example, the following sentence has a negative sentiment, “Z washer used a lot of water for washing the car”. This results in an ironic sentence with or without the sentiment lexicon.

c. They are common in political conversation, but do not seem to appear in product reviews, e.g. “what a great car! It stopped working in two days”.

d. In some instances, question sentences and conditional sentences may not show any sentiment, but still include a sentiment lexicon. For example, “which one of Nikon’s camera is good?”

There are three common approaches for making a sentiment lexicon, namely: a manual construction method, a dictionary-based method and corpus-base method. The manual making of a sentiment lexicon is very tedious and time-consuming, so it is rarely used alone. However, it is usually used with other methods for improving and incrementing the accuracy of results. The following two methods are discussed in sub-sections as shown below.

Dictionary-based:

Dictionary-based methods are centered on bootstrapping by a small seed set of sentiment words and rely on available existing online lexical resources such as WordNet or thesaurus. The approach includes two main steps. At the first step, a small set of sentiment words is collected manually with known polarity sentiment. The next step involves trying to increase and grow an “upping” set with their antonyms and synonyms in the online dictionary. The new words are appended to the pervious set. This process is repeated until new words not previously added to the collection [22] used this approach. The dictionary-based methods and sentiment words extracted from them have a major problem [28]. This approach has been unable to detect sentiment words within an area of specific orientation. In fact, some words may have two different sentiments. For example, the word “quit” in the following sentence has quit different sentiments, for example: “The speaker in my iPhone quit” and “my car quit working”. The word “quit” in the first sentence is usually negative, while the word “quit” in the second sentence denotes a positive sentiment. However, this problem is quite common, but corpus-base methods can solve the above problem.

Kim [22] used a method that assigns a score ratio of positive and negative to each of the sentiment words. Because of the difficulty in scoring words, some of the words may seem to have differing degrees of both negative and positive sentiments; while some words may be seen as having a stronger positive or negative sentiment than other words. Their algorithm first starts with a manually-seen set of sentiment words (i.e. positive and negative words); other seed set words are then added by synonyms and antonyms of positive and negative seed words. Hence, their system first extracts word sentiment as adjectives, verbs and nouns sentiment by POS tagger and WordNet. Support Vector Machine (SVM) method is used in this research.

The main problem in dictionary-base is the inability to detect sentiment words with area specific orientation. A word may have a different meaning in different sentences. For example, the word “large” in the

Advances in Environmental Biology, 8(13) August 2014, Pages: 53-68

sentence, “this computer has a large screen” is positive-oriented; whereas, if it is used in the sentence, “the screen of the laptop is large”, this is usually negative-oriented. This is the problem of domain dependency in sentiment classification; therefore this problem cannot be solved even in the corpus-based methods.

Corpus-base:

The corpus-base methods use a seed set of sentiment words with polarity sentiment and then try to expand this by helping syntactic patterns or co-occurrence patterns to find new sentiment words in a large corpus. Hatzivassiloglou [19] proposed one of the basic methods in this approach. Their methodologies start with a seed set of adjectives, and then go on to use linguistic constraints. There are sets or rules on connectives governing the addition of new words and their orientation. The authors used a graph-base method of reading a large corpus to extract the learning lexicon. Sometimes, people declared their opinions with a pair of conjoined adjectives using conjunctions such as AND, OR, BUT, etc. Their algorithm tries to determine conjoined adjectives from a large corpus that can be either in the opposite or same orientation. For example, considering to conjunction AND in the sentence, “this car is beautiful and spacious”. If “beautiful” is a positive sentiment; so too is the word “spacious” positive. The parser creates a graph whereas nodes are words and links between two nodes in the same or opposite orientation. They applied a parser based on clustering to cluster adjectives in two classes of positive and negative sentiment.

Feature Selection And Extraction Methods:

There are further reviews in document-level which have many features. In fact, many features express a high dimensional space. The main task of feature selection and feature extraction is dimension reduction in feature space while losing the minimum of accuracy. There are several factors relating to reduction of dimension in the term-document matrix of feature space. It causes irrelevant features to be removed; the following results have been achieved, namely: many have efficient categories, analysis of sentiment after reduction will be easier, results may be visualized, and items of low dimension will have better perception.

There are two main approaches required to achieve the appropriate size of dimension for classifier:

a. Feature selection:identifying and extracting features that lead to class separability. Using the univariate approach for rating and multivariate approach for optimization result in a criterion function that may result in a better chip in classification.

b. Feature extraction: this is a reduction of high-dimensional space to lower feature space via a linear transformation or nonlinear transformation. This transformation takes three forms, namely: supervised learning, semi-supervised learning and unsupervised learning methods.

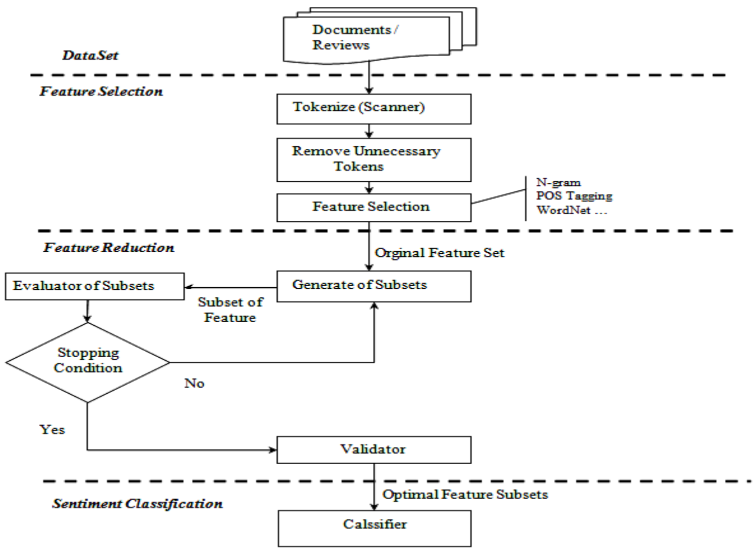

Fig. 1: The steps of feature selection and reduction

Feature Selection:

At the first step, all tokens are extracted from the document-level. At the second step, slight changing of preprocessing with handling emoticons and punctuations is made [3]. The list of emoticons and punctuations are derived from the Wikipedia list. This is then hand-tagged into five classes (neutral, positive, extremely positive, negative and extremely negative).

Stop Words Removal:

Stop words in reviews are seen as a negative role. They can appear in training sets of both positive and negative; as a result, they increase ambiguity in sentiment classification since stop words do not have any sentiment information [3]. Stop words removal in the preprocessing step, including less than 2 words in length, as well as words such as: she, he, at, about, at, the, etc. The list of English stop words can be found in the website: norm.al/2009/04/14/list-of-english-stop-words/

N-gram Method:

N-gram is an ordered set of words. The advantage of using N-gram rather than a single word is, because of the dependencies existing between some words and also the importance of individual phrases. Three forms of N-gram are currently popular and include: unigram, bigram and trigram. Agarwal [3] used a unigram model to compare a feature based model and tree kernel-based model for Twitter data.

Unigram, bigram and a combination of these two have been used to extract features of movie reviews by Pang [37]. Authors experiment by extracting features with three classifiers, i.e. Support Vector Machine (SVM), Naive Bayes (NB) and Maximum Entropy (ME). Also Gamon [13] used lemma unigram, bigram and trigram respectively. There are researches showing that 'N' is greater than 3. For example, Cui *et al*., used high 'N' (up to 6) to extract features from online product reviews. There are several research works that have used N-gram [28, 34-35, 63-64, 68-69].

Part-Of-Speech:

Part-Of-Speech (POS) is used to determine a more useful sentiment classification. However, adjectives are key factors in forecasting sentiment analysis, but Benamara *et al*.showed that the use of adverbs and adjectives words has higher accuracy than merely using adjectives alone. As a result, using POS can reduce feature space, resulting in improved classifier performance via the deletion of less useful words. Table 2 shows some of the Penn Tree POS tags [27].

Table 2: Some POS tags

Tag	Description		Tag	Description
JJ	Adjective		RB	Adverb
JJR	Adjective, Comparative		RBR	Adverb, Comparative
NN	Noun, singular, or mass		SYM	Symbol
NNS	Noun, Plural		VB	Verb
PRP	Personal Pronoun		VBN	Verb, past participle

Feature Extraction:

Dash [9] expressed the finding that there are four basic steps in any feature extraction method. They are, namely: the *generation* that produces a candidate feature subset; while next step is that of *evaluation.* In this step, produced subsets with their relevancy value are evaluated by the evaluation function or classifier and s*topping criteria* that occur in two states. If a subset is optimal, it can be said to be at the end of the process; whilst if it is in use, the generation process calls again for creating a following subset of features. The last stage to emerge from the feature selection process is *validation.* The validation step investigates how to verify a selected feature subset so that it may be a valid subset as required.

Generation:

In the generation stage, a subset of feature is created from a feature set. In each iteration, a subset is created and evaluated by the next stage until the gain of optimal subsets. There are different techniques by which to create subsets, i.e. creating feature subsets completely, randomly and in a heuristic fashion, whereas producing subsets are different from one method to another.

- *Complete generation:* complete or exhaustive generation is a combination of all the feature subsets is tested. If 'n' is number of features, number of subsets are $O(2^n)$. Order of search space is large. Even if we are able to find the best or optimal feature subset, it is generally too expensive and sometimes not very practical for commercial use.
- *Heuristic generation:* the other method to product subset is via a heuristic method, i.e. search algorithms of forward and backward selection that try to find the best subsets. The algorithm adds features one by one to the candidate feature subset until the evaluation function return the target relevancy value. There is one problem for such heuristic approach, in that there may exist a high order combination of relevant feature subsets because some relevant feature of subsets may have been omitted {f1, f2}.
- *Random generation:* In a random state, the feature subset is created randomly without having any criterion or algorithm. As a result, number of optimal feature subsets can be identified by the number of user trying to create subsets.

Evaluation of Subsets:

In the evaluation stage, produced subsets are evaluated by evaluating the function or classifier so as to determine if they are two types, namely, filter and wrapper where this will computes with some relevancy values. Such value is then compared with the previously-known best value.

Similarity with the generation step, we are able to categorize different feature selection methods according to the way the evaluation is carried out. There currently exist 5 different evaluation methods [9].

Table 3: A cooperation of evaluation functions

Approach	Examples	Generality	Time	Accuracy	Type
Distance	Euclidean Distance Measure	yes	low	-	Filter
Information	Entropy, Information Gain, etc.	yes	low	-	Filter
Dependency	Correlation Coefficient	yes	low	-	Filter
Consistency	Min-Features Bias	yes	moderate	-	Filter
Classifier error-rate	The Classifier Themselves	no	high	very high	Wrapper

Table 3 shows a comparison of different evaluation functions regardless of the kind of procedure used. The '-' in the accuracy column means that nothing about the accuracy of the corresponding evaluator cannot be concluded. Following is a brief discussion of each of these types of evaluation functions:

- In *distance measure*, we are computing the physical distance. Features that can support instances/records of the class in order to stay together are selected. The key concept is the assumption that instances of the same class must be closer than those in different class.
- *Information measure* is selects a feature subset that can yield the maximal information gain.
- *Dependency measure*, this measures the correlation between a feature and a class label. If feature A has a higher relation to the class than feature B, then we select feature A. It measures how closely a feature is related to the outcome of the class label. A slight variation of the definition can be used to measure the degree of redundancy between features. For example, if feature A is heavily dependent upon feature B, then feature A is redundant. Since correlation is only a measure of relationship, we need some kind of physical measure in order to define such a relationship.
- *Consistency measure*, two instances are considered in-consistence if such a situation occurs while having all matching feature values, except its class. Select only if there is no such case. It makes use of the Min-feature bias where Find is a minimally sized subset that satisfies the acceptable inconsistency rate (i.e. defined by the user). This bias may lead to problems when one feature alone guarantees no inconsistency. IC value is unique for all instances, since you can never find two people with the same IC number (i.e. two feature values that are the same).
- *Classifier error rate,* in this approach, feature selection has lost its generality, but gained accuracy towards the classification task. It was very costly from a computationally aspect.

In Table 3, some literature studies categorized the first four as a *filter approach* and the final one as a *wrapper approach*. In a study by Saeys [43], the feature extraction methods were classified into three classes:

- Filter techniques
- Wrapper techniques
- Embedded techniques

Filter techniques:

Filter methods are independent of inductive algorithms. Filter methods select the best of features based on some intrinsic properties criteria e.g. using their Euclidean Distance measure (that is, choosing features to stay with the same proximity by instance of the same class). It is important to assume that samples of the same class should be nearest to those in other classes. In fact, filter techniques selected a related feature that has both high-scoring and removal low-scoring features. The best of the feature subsets are then sent to the classifier [43].

Wrapper techniques:

Wrapper methods are inductive algorithms such as an evaluator. These methods select the best of features subsets by generation and evaluation of different subsets in space of states. The selection and evaluation of a specific feature subset is achieved from classifiers by training and testing algorithms. Hence, wrapped algorithms search the space of all subsets of features as classification methods. Heuristic techniques can help to search for optimal subsets [43].

Advances in Environmental Biology, 8(13) August 2014, Pages: 53-68

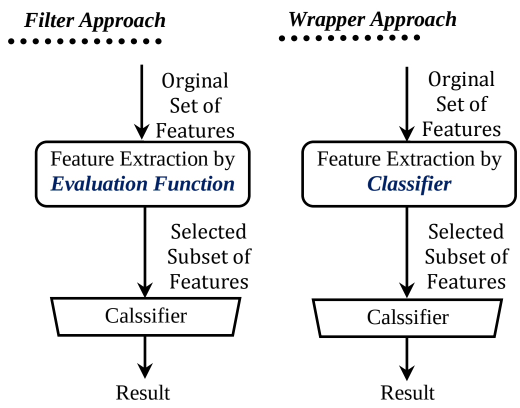

Fig. 2: Comparison of two approaches based on the type of feature extraction functions

Figure 2, shows two different types of feature extraction. Although filter methods have several benefits such as: fast, simple, independent processing of classifiers and capable of being easily scaled to very high-dimensional; they have a common problem in that relinquish classifier and the majority of suggested methods are univariate so that it may reduce the accuracy of classification. Multivariate filter methods can result in a reduction of the above problem, i.e. ignoring feature affiliations. In fact, they connect affiliations of features to the same degree [43]. The benefit of wrapper methods is the semantic relationship among model selections and subsets of search features. Further, they interrelate with the classification algorithms. Nevertheless, they are a higher over-fitting risk and have a significant complexity of computation and cost.

Embedded techniques:

Embedded methods look for an optimal subset of features via search in hypotheses and space of feature subset. In fact, this method creates the classifier construction. Embedded methods are a special form of existing classifiers with learning algorithms. The benefit of this method is far less complexity than wrapper methods; and, at same time, there is interplay with classifier and dependence upon feature. For example, [16, 67] used the weight vector of each feature in SVM as a linear classifier. The weights express a relationship of multivariate features in the result, causing cancellation of features with light weight.

Table 4 shows a taxonomy and category of feature extraction methods. Saeys [43], for each feature extraction expressed a set of characters which can help to select a suitable and better method of attaining the goals via listing the advantages and disadvantages of each method.

Table 4: A category on pros and cons of feature extraction methods

Type Search		Positive	Negative	Example
Filter methods	Univariate	. Quick . Gradable . No dependence to classifier	. Relinquish dependence to feature . Relinquish interplay with classifier	. Information Gain (IG) . x^2 – CHI . *t* - test
	Multivariate	. Dependence to feature . No dependence to classifier . Better time complexity than wrapper	. Slower than univariate methods . Less gradable than univariate methods . Relinquish interplay with classifier	. Correlation-based feature selection (CFS) . Markov blanket filter (MBF) . Fast correlation-based feature selection (FCBF)
Wrapper methods	Deterministic	. Simple . Dependence to feature . Interplay with classifier . Slower than Randomize	. High risk to over-fitting . More entrapment to local optimum than Randomize . Classifier dependent selection	. Sequential forward selection (SFS) . Sequential backward elimination (SBE) . Beam search
	Randomize	. Dependence to feature . Less entrapment to local optimum . Interplay with classifier	. Classifier dependent selection . More risk of over-fitting than deterministic	. Simulated Annealing . Randomized hill climbing . Genetic algorithms . Estimation of distribution algorithms
Embedded methods		. Dependence to feature . Interplay with classifier . Better time complexity than wrapper	. Classifier dependent selection	. Decision trees . Weighted naive Bayes . Feature selection using the weight vector of SVM

Advances in Environmental Biology, 8(13) August 2014, Pages: 53-68

Methods of Feature Selection:

There are several methods of feature selection and reduction that are gaining in popularity. The following survey uses six methods, namely: Document Frequency (DF), Term Frequency-Inverse Document Frequency (TF-IDF), Mutual Information (MI), GHI-square statistic (CHI), Information Gain (IG), and Principal Component Analysis (PCA). All of these Methods use score in term extraction and select the size of the predefined set of characteristics.

Document Frequency:

In the Document Frequency (DF) method, features are ordered by document frequency for each feature in the whole document [24, 39]. This method is the simplest measure for feature reduction and has a linear time complexity to scale a large dataset.

Term Frequency-Inverse Document Frequency:

Term Frequency-Inverse Document Frequency (TF-IDF) is a numerical static in that it uses a score as a weighting factor for important features in a corpus. In other words, it is the number of documents that contains the desired feature [64]. TF-IDF value is calculated based on the number of times a feature appears in a target document and corpus. It is defined to be:

$$TF - IDF(f) = TF(f) * \log\left(\frac{N}{DF(f)}\right) \quad (1)$$

Where *f* is a feature, *TF(.)* or term frequency denotes the number of features that appear in a target document, *N* is the number of all documents in a corpus, and *DF(.)*, or document frequency, refers to the number of documents that contain this feature.

Chi-square statistic:

The Chi-square (CHI) is a normalized value. This value calculates the degree of the relationship between the feature and the category [64]. The feature-goodness criterion is defined as follows:

$$CHI(f, c_i) = \frac{N \times (AB - CD)}{(A+D)(B+C)(A+C)(B+D)} \quad (2)$$

And, $CHI_{max}(f) = \max_i (CHI(f, c_i)) \quad , \quad i = 1, .., M$

Where, *f* is feature, *c* is category, *N* is the number of all documents in the corpus, *M* is the number of the category (e.g. M=2 for polarity classification), *A* is the number of times *f* and *c* co-situate, *B* is the number of time neither *c* nor *f* situate, *C* is the number of times *f* situates without *c*, and *D* is the number of times *c* situates without *f*.

Mutual Information:

Mutual Information (MI) is a common measure used in statistical language modeling of feature community and relevant applications [58]. It is defined as follows:

$$MI(f, c_i) = \log\left(\frac{A \times N}{(A+D)(A+C)}\right) \quad (3)$$

And, $MI_{max}(f) = \max_i (MI(f, c_i)) \quad , \quad i = 1, .., M$

Where the meaning of *M*, *N*, *A*, *B*, *C*, and *D* share the same denotation in CHI.

Information Gain:

Information gain (IG) refers to the number of bits of information that gained forecasting of classification by the presence or absence of a feature in a document. Yang [58] presented a greater generality that can be seen as follows:

$$IG(f) = -\sum_{i=1}^{M} P_r(c_i) \times log\, P_r(c_i) + \quad (4)$$
$$P_r(f) \times \sum_{i=1}^{M} P_r(c_i|f) \times log\, P_r(c_i|f) +$$
$$P_r(\bar{f}) \times \sum_{i=1}^{M} P_r(c_i|\bar{f}) \times log\, P_r(c_i|\bar{f})$$

Where *f* is a feature, $P_r(f)$ indicates the probability situates feature, $P_r(\bar{f})$ means that probability does not situate a feature and $P_r(c_i)$ denotes the situate of class c_i.

Advances in Environmental Biology, 8(13) August 2014, Pages: 53-68

Principal Component Analysis:

Principal Component Analysis (PCA) is a computational method that uses orthogonal transformation to transform a set of associated features into a set of values of linear non-associated features. The number of features in PCA is less than or equal to the number of original features. This transformation function can definitively vary features which have higher values of selected variables. PCA is relative to the relevant scaling of original features. If the features are normally spread, then the principal components can be independent [25, 44]. It will be calculated as follows:

$$\overline{x}_k = \frac{1}{N}\sum_{j=1}^{N} x_{jk} \quad , \quad k = 1,2,..,M \tag{5}$$

Where x is a feature value, $\overline{x}$ denotes the mean (or average) of feature values, N is the number of all documents, and M is the number of all features. Variance (S^2_{kk}) is calculated by:

$$S_{kk}^2 = \frac{1}{N}\sum_{j=1}^{N}(x_{jk} - \overline{x}_k)^2 \quad , \qquad k = 1,2,..,M \tag{6}$$

And the covariance is

$$S_{ik} = \frac{1}{N}\sum_{j=1}^{N}\left(x_{ji} - \overline{x}_i\right) \times \left(x_{jk} - \overline{x}_k\right) \quad , \qquad i = 1,2,..,N \quad k = 1,2,..,M \tag{7}$$

Sentiment Classification:

Sentiment Classification methods can be categorized into three groups, namely: supervised learning, semi-supervised learning and unsupervised learning.

Supervised Learning:

Supervised learningmethods usually use two datasets, specifically: a training dataset and a testing dataset. Form the first instance, classifiers try to learn by samples or training dataset and then predict a set of testing. In sentiment classification, classes are divided into two, namely: positive class and negative class. Some of the reviews appearing on the web have rating levels of 1 to 5 stars. Hence, the stars can be defined in a form as follows: the first and second stars are negative class, the third star is neutral class and the fourth and fifth stars are positive class. Since sentiment classification is similar to the text classification, accordingly, it can be applied as an available method of supervised learning of text classification. A study by [37] applied three methods of machine learning for sentiment classification. *Naive Bayes* (NB), *Support Vector Machine* (SVM), *Maximum Entropy* (ME or MaxEnt) are frequently used in sentiment classification. They classify datasets of movie reviews in order to determine positive or negative opinions. In addition, the using of unigram as features has performed well with SVM and NB.

Semi-supervised Learning:

In Natural Language Processing (NLP), one of the problems confronting sentiment analysis method is poor labeling of data. In semi-supervised learning, there are fewer numbers of labeled data and, hence, it is necessary to use unlabeled data. In fact, semi-supervised learning methods should train to use both labeled and unlabeled data. Thus, we should apply active learning to unlabeled data or reviews as training data by which to identify opinion polarity [68]. The authors have presented a semi-supervised algorithm named Active Deep Networks (ADN) to be used in conjunction with active learning. ADN made by the algorithm of Restricted Boltzmann Machine (RBM) and unsupervised algorithms use a great deal unlabeled data and little labeled data. ADN is capable of selecting appropriate training labeled data and training in an active deep structure method at the same time.

Unsupervised learning for polarity sentiment is difficult due to the fact that there are a significant number of ambiguous sentiments in reviews [8]. At the first instance, they attempt to solve this problem, by extracting unambiguous reviews and then using a novel synthesis of transductive, active learning and ensemble learning so as to identify and classify ambiguous reviews.

Most present research has assumed the balance among samples of positive and negative in labeled and unlabeled data may not actually be true. Li [26] proposed a solution to this problem. They surveyed a more common case approach towards imbalances by creating different random sub-spaces in a dynamic fashion. In fact, they offered a novel semi-supervised learning algorithm based on the dynamic generation of random sub-space in the iteration process. Arcilla [4] introduced a novel lexical resource builder for semi-supervised learning. In this process, they run their lexical builder online product reviews which automatically extract word pairs versus opinionated words. Zhu [69] improved the abilities of the SVM process by use of a semi-supervised method based on analysis of Chinese micro-blog data using an iteration method. They improved performance through combinations of different features and weighted factors on a text source of objective and subjective unrelated experiments.

Examples using previous lexical knowledge in relation to unlabeled and labeled data in recent research are shown. For example, [46] introduced a semi-supervised forecasting method by connecting sentiment analysis lexicon and documents based on bi-partite graph representations of both labeled and unlabeled data. [35]presented a solution for the real-world multi-dimensional sentiment analysis problem via calculation of the statistical relationship existing among target variables when a customer expresses his opinion about a specific topic. Hence, there are three distinct target variables, specifically: sentiment polarity, subjectivity and the will to influence. Due to the aforementioned problems, they have used a multi-dimensional Bayesian Network classifier that can connect the assorted target variables to the same category for using their potential relationship.

Unsupervised Learning:

Words and phrases of sentiment are the main indexes in sentiment analysis and classifications. Hence, some researches have been done based on unsupervised learning methods. The following survey presents some samples of unsupervised learning.

In a sentence, Turney [52] presented a simple unsupervised learning method relating to recommended and not-recommended reviews. The authors used the Point-Wise Mutual Information (PMI) method to determine words that can be denoted as positive or negative by either a negative seed word ("poor") or a positive seed word ("excellent"); these words are called words of semantic orientation. He used POS patterns as "adjectives and adverbs" for search and extraction of words and phrases in reviews. Harb [18] improved Turney's work via Google's search engine by two sets of semantic oriented processes. They introduced an association of rules whereby one can discover more words and phrases. They achieved this by using their method to create a domain-oriented sentiment lexicon. However, they cannot connect words and documents bilaterally based on labeled seed words of "excellent" and "poor". Taboada [48] presented a lexicon-based approach to identify polarity sentiment words, i.e. positive and negative, to sentiment classification via dictionary-based. They introduced a Semantic Oriented CALculator (SO-CAL) to increase the word domain by combining negation words and reinforces. Usha [53] exploited a model called Combined Sentiment Topic (CST) in order to find sentiments and topics at the same time. They claimed that their model is portable compared to other domains. Further, they claimed that it is better than existing semi-supervised methods. They experimented with classifying polarity sentiment on documents in general domains. Xu [56] exploited an approach for identifying word triples of aspect-modifier-sentiment with shallow semantic information by using a semantic role labeling (SRL) tool. As a result, they used POS and SRL information with heuristic rules.

Table 5: A taxonomy on previous studies in Sentiment Classification

Authors	Methods	Features	Dataset	Type
(Pang, Lee *et al*. 2002)	SVM, Naive Bayes, Maximum Entropy	(Uni + Bi)grams, Adjective, Position of words	Movie Review	Supervised
(Prabowo and Thelwall 2009)	SVM, Rule-based Classifier	N-grams, POS tagging	Movie reviews, Product reviews	Supervised
(Zhang and Liu 2011)	SVM,Naïve Bayes	(Uni + Bi + Tri)grams	Restaurant review	Supervised
(Goldberg and Zhu 2006)	Graph-base	(Uni + Bi)grams, Adjective, Position of words	Movie Review	Semi-supervised
(Arcilla, Esquivel *et al*. 2013)	FeLex Builder	Lexicon Builder	Product Review	Semi-supervised
(Zhou, Chen *et al*. 2013).	Active Deep Networks (ADN), Boltzmann machines (RBM)	Unigram	Movie Review, Product Review	Semi-supervised
(Zhu, Xu *et al.* 2013)	SVM, Bootstrapping	Objective, Subjective	Chinese micro-blogs on Twitter	Semi-supervised
(Harb, Plantié *et al*. 2008)	Association Rule	Adjectives, Adverbs	Movie review	Unsupervised
(Taboada, Brooke *et al.* 2011)	Dictionary based approach	Adjectives, Nouns, Verbs, Adverbs, Intensifier, Negation	Movie review, Camera review	Unsupervised
(Hu, Tang *et al*. 2013)	General Inquirer (GI) - label, MPQA-label, K-Means, *ESSA*	Unigram	Stanford Twitter Sentiment, Obama-McCain Debate	Unsupervised

Classification Techniques:

Sentiment classification can be divided into two types of classification forms. First, binary sentiment classification also known as a polarity sentiment classification, includes both positive and negative classes. Finally, multi-class sentiment gives classification by using rating marks, for example: five-class, i.e. 1 to 5 stars or classes of, namely, {strong positive, positive, neutral, negative, strong negative} instead of a two-class system of negative and positive classes.

Advances in Environmental Biology, 8(13) August 2014, Pages: 53-68

Naive Bayes:

The Bayesian classification is a statistical method underlying a probabilistic model and supervised learning algorithms. Naive Bayes (NB) use a features vector matrix to determine a document that chooses which of the polarity classes (i.e. positive and negative classes) it belongs to by using a probability. It attaches a document to the relevant class having the highest probability [49-50, 58]. The probability is calculated as follows:

$$P(c|d) = \frac{P(c)\times P(d|c)}{P(d)} \tag{8}$$

Where *P(c)* is prior probability of category *c*, *P(d)* is prior probability of training dada *d*, *P(c|d)* is probability *c* given *d*, and *P(d|c)* is probability *d* given *c*.

Support Vector Machine:

Support Vector machine (SVM) is a most popular algorithm that can classify data as both linear and nonlinear, and map input data to high-dimensional feature spaces. In additional, classifiers' SVM supports regression, binary and multiclass classification respectively. For example, SVM classifier on binary classification tries to find a decision surface that can separate data into two classes and result in making a decision based on this support vector [49, 58, 64]. Following is an equation that should be minimized for optimization of SVM:

$$\vec{\alpha}^* = \arg\min \{ -\sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle \vec{x}_i, \vec{x}_j \rangle \} \tag{9}$$

Where, $\sum_{i=1}^{n} \alpha_i y_i = 0 \quad , \qquad 0 \le \alpha_i \le C$

The support vector can be linear or nonlinear. The nonlinear classification of SVM can be perfected if the kernel trick is used, and if kernel used Gaussian Radial Basis Function (RBF) then feature space will be a Hilbert space of infinite dimension. Hence, classifiers of maximum borders are well-regulated. Thus, infinite dimension does not destroy the results. Some of the kernels are defined as *a Gaussian Radial Basis Function (RBF)*, *polynomial (homogeneous)*, *polynomial (inhomogeneous)*, and *hyperbolic tangent*. RBF kernel is a popular function as SVM classification. SVM with RBF kernel is closely related to RBF neural networks, with the centers of the radial basis functions automatically chosen for SVM.

Maximum Entropy:

Maximum Entropy (ME) classifier is one of the machine learning methods for natural language processing application as it is implemented using a multinomial logit model as the classifier rule. ME is a kind of statically inference that can be used to estimate any probability distributions on the partial knowledge.

$$P(c|d) = \frac{1}{F(d)} exp\left(\sum_{i=1}^{N} W_{i,c} X_{i,c}(d,c)\right) \tag{10}$$

And $X_{i,c}(d,c') = \begin{cases} 1 & if\ n_i(d) > 0\ and\ c = c' \\ 0 & else \end{cases}$

Where $X_{i,c}$ is a function of feature/class for i feature and c class, $W_{i,c}$ is weight of features, and F(d) is a normalization function.

Artificial Neural network:

The artificial neural network has some benefits as regards adaptive learning, error tolerance, parallelism, and generalization. Neural networks can be divided into two types, namely: feed forward networks and feedback networks.

On their own, semantic oriented indexes do not have good performance but they return results rapidly. Further, a Machine Learning method, such as an artificial neural network extract has better classification accuracy, but requires a considerable amount of learning time. Chen [7] combined two approaches of Neural Networks (NN) and Information Retrieval (IR) techniques like the Semantic Oriented (SO) indexes, because there are benefits to using both of them together. The authors proposed four different kinds of semantic oriented indexes, specifically: Semantic Oriented Association (SO-A), Point-Wise Mutual Information (PMI), SO-PMI, and SO-LSA, as inputs of the multilayer feed forward perceptron neural network with back-propagation learning algorithm. Their approach shows that they improved classification accuracy and training time for movie and product reviews.

The deep architecture neural network is a one of the semi-supervised learning approaches that has achieved high performance in an object recognition task. Yanagimoto [57]used a combinatorial network of deep architecture neural network and Restricted Boltzmann Machine (RBM) as semi-supervised methods for raw features. They used T&G news as a corpus that included 62478 articles about stock prices. Some 100 articles have been labeled (positive, negative or neutral), and 71 articles have been used for evaluation and testing. They claimed that this could improve performance namely by reducing time complexity and increasing the accuracy of classification. A study by (Rustamov, Mustafayev *et al*. 2013), proposed three methods, specifically: Adaptive Neuro-Fuzzy Inferences System (ANFIS), Fuzzy Control System (FCA) and Hidden Markov Model (HMM) in sentence-level subjectivity. They used fuzzy inference for making an input vector of a multilayer perceptron neural network with back-propagation learning algorithm. Further, a research work by Sharma [45] proposed that artificial neural network with back-propagation learning algorithms is used for sentiment analysis in the document-level. For extracting the sentiment lexicon, they used 4 types of feature selections, i.e. Information Gain, Opinion Lexicon that includes opinion polarity words, General Inquire lexicon, and the Hatizassiloglou and McKeown method to train and test. Movie and hotel dataset review are used in this work.

Group Method of Data Handling:

Group Method of Data Handling (GMDH) is a self-organizing method for modeling complex systems and allows exploring the internal laws of the proper object area. The benefit of the GMDH approach is the opportunity make optimization models with a very small number of features while not knowing the dynamic among these features [65]. One of the reasons for using GMDH is a combination of high-dimensional data and relatively few training data [2].

GMDH tries to find a solution to the hierarchy with many simple models while keeping the best model by making repeated efforts to gain a suitable combination of function as models. The inter structure of each node or block is composed two forms, linear and nonlinear [41, 62] as follows:

$z = w_0 + w_1 f_1 + w_2 f_2 + w_3 f_1 f_2$

And nonlinear is, $$z = w_0 + w_1 f_1 + w_2 f_2 + w_3 f_1 f_2 + w_4 f_1^2 + w_5 f_2^2 \qquad (11)$$

Where $w_i$is weight vector or coefficient of polynomial as a vector, and x_1 and x_2 are inputs of each neuron. Output of network (y) calculates by one of below equations (Kondo and Pandya 2000):

$$y = z\ ,\ y = e^{-z^2},\ y = \frac{1}{1+e^z}\ ,\ y = a_0 + a_1 z + a_2 z^2 + \ \dots \qquad (12)$$

Ravisankar [41] exploited the GMDH network and five other machine learning methods in order to identify and predict companies that resort to financial statement fraud on a dataset of 202 Chinese companies with ten-fold cross-validation. Abdel-Aal [1] used GMDH-base to rank and select medical features for diagnosis classification. They used a hierarchical approach for selecting optimum predictors by complete ranking of features subsets based on their predictive quality by using GMDH type learning algorithms. This involved featuring ranking and dividing into groups by their order via a GMDH type.El-Alfy [12] continued the work of Abdel-Aal [1] by investigating spam detection massages from legal emails by GMDH-based networks. One supervised machine learning, i.e. Abductory Inductive Mechanism (AIM) and GMDH-type approach has tried to make an efficient high degree polynomial model in a repeating method without over-fitting the training dataset. They compared their proposed network (Abductive networks) with multilayer perceptron neural network and Naive Bayes classifier.

Classifier Ensemble:

Network ensembles are trained to find solutions for the same problem in a parallel independent [12]. There are several classifiers for sentiment classification that have both advantages and disadvantages. The aim of a classifier ensemble is combine these classifiers while still gathering their benefits, thus improving performance.

Xia [55] investigated the effectiveness of ensemble technique on feature sets and sentiment classification in a three step process. In the first step, they extracted sets of features by means of POS-base and word-relation-base. At the second step, they used three base-classifiers for each sets by the usage of SVM, maximum entropy, and naive bayes classifiers. In the last step, they combined these methods based on fixed combination, meta-classifier combination, and weighted combination as ensemble strategies. Tests using ensemble methods result in higher accuracy that shows an efficient way to improve classification performance via a combination of different feature sets and classifiers. Ekbal [11]proposed their approach over two stages. At the first stage, two statistical classifiers of support vector machine (SVM) and conditional random field (CRF) were used to select features based on multi-objective feature selection. In the last stage, these classifiers were combined as a classifier ensemble based on multi-objective-simulated annealing (MOSA), this combination is then able to select appropriate weights as voted in each classifier for named entity recognition. Su [47] proposed one of the classifier ensemble methods, stacking generalization five base-level classifiers (naive bayes, centroid-based, K-nearest neighbors, maximum entropy, and support vector machine) with different settings, compared with the majority voting. They used an opinion summary; to this end, they considered only the first two and last two sentences at most in a review. Also, they used three types of weighting measures of presence, term frequency

(TF), and TF-IDF for extracting features of book, hotel and notebook reviews respectively. They achieved higher performance via opinion summary.

Conclusion:

With the advent of the Web together with widespread internet and social media (such as social networks, reviews, comments, twitter and forum discussion on the web), we are currently seeing a huge growth of data and information on the web. This includes the growth of digital or textual opinions, sentiments and attitudes as remarked upon in reviews. Special attention needs to be given to the processes and understating of information by information retrieval methods and natural language processing methods. One of the main problems in this scope is sentiment analysis whereby a review is classified into two classes, i.e. positive or negative opinions.

The purpose of this article was to prepare a way for increased learning about sentiment analysis and existing techniques. We investigated different methods of feature selection as well as a variety of different algorithms of sentiment classification in sentiment analysis and compared the results in several studies.

REFERENCES

[1] Abdel-Aal, R., 2005. GMDH-based feature ranking and selection for improved classification of medical data. Journal of Biomedical Informatics., 38(6): 456-468.
[2] Abdel-Aal, R.E. and E.-S.M. El-Alfy, 2009. Constructing optimal educational tests using GMDH-based item ranking and selection. Neurocomputing, 72(4): 1184-1197.
[3] Agarwal, A., *et al*., 2011. Sentiment analysis of twitter data. Proceedings of the Workshop on Languages in Social Media, Association for Computational Linguistics.
[4] Arcilla, M., *et al*., 2013. FELEX BUILDER: A SEMI-SUPERVISED LEXICAL RESOURCE BUILDER FOR OPINION MINING IN PRODUCT REVIEWS. The Second International Conference on Digital Enterprise and Information Systems (DEIS2013).The Society of Digital Information and Wireless Communication.
[5] Asur, S. and B.A. Huberman, 2010. Predicting the future with social media. Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on, IEEE.
[6] Bollen, J., *et al*., 2011. Twitter mood predicts the stock market. Journal of Computational Science, 2(1): 1-8.
[7] Chen, L.-S., *et al*., 2011. A neural network based approach for sentiment classification in the blogosphere. Journal of Informetrics., 5(2): 313-322.
[8] Dasgupta, S. and V. Ng, 2009. Mine the easy, classify the hard: a semi-supervised approach to automatic sentiment classification. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: 2-Volume 2, Association for Computational Linguistics.
[9] Dash, M. and H. Liu, 1997. Feature selection for classification. Intelligent data analysis, 1(3): 131-156.
[10] Dave, K., *et al*., 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. Proceedings of the 12th international conference on World Wide Web, ACM.
[11] Ekbal, A. and S. Saha, 2013. Combining feature selection and classifier ensemble using a multiobjective simulated annealing approach: application to named entity recognition. Soft Computing, 17(1): 1-16.
[12] El-Alfy, E.-S.M. and R.E. Abdel-Aal, 2011. Using GMDH-based networks for improved spam detection and email feature analysis. Applied Soft Computing, 11(1): 477-488.
[13] Gamon, M., 2004. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. Proceedings of the 20th international conference on Computational Linguistics, Association for Computational Linguistics.
[14] Godbole, N., *et al*., 2007. Large-Scale Sentiment Analysis for News and Blogs. ICWSM 7.
[15] Goldberg, A.B. and X. Zhu, 2006. Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing, Association for Computational Linguistics.
[16] Guyon, I., *et al*., 2002. Gene selection for cancer classification using support vector machines. Machine learning., 46(1-3): 389-422.
[17] Hajmohammadi, M.S., R. Ibrahim, A. Selamat and A. Yousefpour, 2014. *Combination of Multi-view Multi-source Language Classifiers for Cross-Lingual Sentiment Classification*. In *Intelligent Information and Database Systems*. (pp. 21-30). Springer.
[18] Harb, A., *et al*., 2008. Web opinion mining: how to extract opinions from blogs? Proceedings of the 5th international conference on Soft computing as trans disciplinary science and technology, ACM.
[19] Hatzivassiloglou, V. and K.R. McKeown, 1997. Predicting the semantic orientation of adjectives. Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, Association for

Computational Linguistics.Hu, M. and B. Liu (2004). Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM.
[20] Hu, X., *et al*., 2013. Unsupervised sentiment analysis with emotional signals. Proceedings of the 22nd international conference on World Wide Web, International World Wide Web Conferences Steering Committee.
[21] Kechaou, Z., *et al*., 2013. A novel system for video news' sentiment analysis. Journal of Systems and Information Technology, 15(1): 24-44.
[22] Kim, S.-M. and E. Hovy, 2004. Determining the sentiment of opinions. Proceedings of the 20th international conference on Computational Linguistics, Association for Computational Linguistics.
[23] Kondo, T. and A.S. Pandya, 2000. GMDH-type neural networks with a feedback loop and their application to the identification of large-spatial air pollution patterns. SICE 2000. Proceedings of the 39th SICE annual conference. International session papers, IEEE.
[24] Lam, S.L. and D.L. Lee, 1999. Feature reduction for neural network based text categorization. Database Systems for Advanced Applications, 1999. Proceedings, 6th International Conference on, IEEE.
[25] Lee, Z.S., *et al*. 2006. "Automated web pages classification with integration of principal component analysis (PCA) and independent component analysis (ICA) as feature reduction."
[26] Li, S., *et al*., 2011. Semi-supervised learning for imbalanced sentiment classification. Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Three, AAAI Press.
[27] Liu, B., 2012. "Sentiment analysis and opinion mining." Synthesis Lectures on Human Language Technologies., 5(1): 1-167.
[28] Liu, B. and L. Zhang, 2012. A survey of opinion mining and sentiment analysis. Mining Text Data, Springer., pp: 415-463.
[29] Liu, Y., *et al*., 2007. ARSA: a sentiment-aware model for predicting sales performance using blogs. Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, ACM.
[30] McDonald, R., *et al*.,2007. Structured models for fine-to-coarse sentiment analysis. Annual Meeting-Association For Computational Linguistics.
[31] Mohammad, S.M. and T.W. Yang, 2011. Tracking sentiment in mail: how genders differ on emotional axes. Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (ACL-HLT 2011.
[32] Nakagawa, T., *et al*., 2010. Dependency tree-based sentiment classification using CRFs with hidden variables. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Association for Computational Linguistics.
[33] Nasukawa, T. and J. Yi, 2003. Sentiment analysis: Capturing favorability using natural language processing. Proceedings of the 2nd international conference on Knowledge capture, ACM.
[34] Ofek, N., *et al*., 2013. Improving Sentiment Analysis in an Online Cancer Survivor Community Using Dynamic Sentiment Lexicon. Social Intelligence and Technology (SOCIETY), 2013 International Conference on, IEEE.
[35] Ortigosa-Hernández, J., *et al*., 2012. "Approaching Sentiment Analysis by using semi-supervised learning of multi-dimensional classifiers." Neurocomputing, 92: 98-115.
[36] Padmaja, S. and S.S. Fatima, 2013. "Opinion Mining and Sentiment Analysis–An Assessment of Peoples' Belief: A Survey." International Journal.
[37] Pang, B., *et al*., 2002. Thumbs up?: sentiment classification using machine learning techniques. Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, Association for Computational Linguistics.
[38] Pestian, J.P., *et al*., 2012. "Sentiment analysis of suicide notes: A shared task." Biomedical Informatics Insights 5(Suppl. 1): 3.
[39] Prabowo, R. and M. Thelwall, 2009. "Sentiment analysis: A combined approach." Journal of Informetrics., 3(2): 143-157.
[40] Qiu, G., *et al*., 2011. "Opinion word expansion and target extraction through double propagation." Computational linguistics., 37(1): 9-27.
[41] Ravisankar, P., *et al*., 2011. "Detection of financial statement fraud and feature selection using data mining techniques." Decision Support Systems, 50(2): 491-500.
[42] Rustamov, S., *et al*., 2013. Sentiment analysis using Neuro-Fuzzy and Hidden Markov models of text. Southeastcon, 2013 Proceedings of IEEE, IEEE.
[43] Saeys, Y., *et al*., 2007. "A review of feature selection techniques in bioinformatics." Bioinformatics., 23(19): 2507-2517.

[44] Selamat, A. and S. Omatu, 2004. "Web page feature selection and classification using neural networks." Information Sciences, 158: 69-88.
[45] Sharma, A. and S. Dey, 2012. "A document-level sentiment analysis approach using artificial neural network and sentiment lexicons." ACM SIGAPP Applied Computing Review, 12(4): 67-75.
[46] Sindhwani, V. and P. Melville, 2008. Document-word co-regularization for semi-supervised sentiment analysis. Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on, IEEE.
[47] Su, Y., *et al*., 2013. Ensemble learning for sentiment classification. Chinese Lexical Semantics, Springer, pp: 84-93.
[48] Taboada, M., *et al*., 2011. "Lexicon-based methods for sentiment analysis." Computational linguistics., 37(2): 267-307.
[49] Tan, S. and J. Zhang, 2008. "An empirical study of sentiment analysis for chinese documents." Expert Systems with Applications., 34(4): 2622-2629.
[50] Tang, H., *et al*., 2009. "A survey on sentiment detection of reviews." Expert Systems with Applications 36(7): 10760-10773.
[51] Tumasjan, A., *et al*., 2010. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment." ICWSM 10: 178-185.
[52] Turney, P.D., 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th annual meeting on association for computational linguistics, Association for Computational Linguistics.
[53] Usha, M. and M. Indra Devi, 2013. Analysis of sentiments using unsupervised learning techniques. Information Communication and Embedded Systems (ICICES), 2013 International Conference on, IEEE.
[54] Whitelaw, C., *et al*., 2005. Using appraisal groups for sentiment analysis. Proceedings of the 14th ACM international conference on Information and knowledge management, ACM.
[55] Xia, R., *et al*., 2011. "Ensemble of feature sets and classification algorithms for sentiment classification." Information Sciences, 181(6): 1138-1152.
[56] Xu, P.P., *et al*., 2013. "An Unsupervised Sentiment Information Identification Approach." Applied Mechanics and Materials., 263: 3330-3334.
[57] Yanagimoto, H., *et al*., 2013. Word Classification for Sentiment Polarity Estimation Using Neural Network. Human Interface and the Management of Information. Information and Interaction Design, Springer, pp: 669-677.
[58] Yang, Y. and J.O. Pedersen, 1997. A comparative study on feature selection in text categorization. ICML.
[59] ano, T. and N.A. Smith, 2010. What's Worthy of Comment? Content and Comment Volume in Political Blogs. ICWSM.
[60] Ye, L. and E. Keogh, 2009. Time series shapelets: a new primitive for data mining. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM.
[61] Yi, J., *et al*., 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. Data Mining, 2003. ICDM 2003. Third IEEE International Conference on, IEEE.
[62] Yousefpour, A. and Z. Ahmadpour, 2011. "The prediction of air pollution by using Neuro-fuzzy GMDH." The Journal of Mathematics and Computer Science, 2(3): 488-494.
[63] Yousefpour, A., R. Ibrahim and H.N. Abdull Hamed, 2014. A Novel Feature Reduction Method in Sentiment Analysis. International Journal of Innovative Computing, 4(1).
[64] Yousefpour, A., R. Ibrahim, H.N.A. Hamed and M.S. Hajmohammadi, 2014. Feature Reduction Using Standard Deviation with Different Subsets Selection in Sentiment Analysis. In Intelligent Information and Database Systems. (pp. 33-41). Springer.
[65] Zaychenko, Y.P. and I. Zayets, 2003. "The fuzzy group method of data handling and its application to the tasks of the macroeconomic indexes forecasting." Systems Analysis Modelling Simulation., 43(10): 1321-1329.
[66] Zhang, L. and B. Liu, 2011. Identifying Noun Product Features that Imply Opinions. ACL (Short Papers).
[67] Zhang, X., *et al*., 2006. "Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data." BMC bioinformatics., 7(1): 197.
[68] Zhou, S., *et al*., 2013. "Active deep learning method for semi-supervised sentiment classification." Neurocomputing.
[69] Zhu, S., *et al*., 2013. Chinese Microblog Sentiment Analysis Based on Semi-supervised Learning. Semantic Web and Web Science, Springer, pp: 325-331.