Engineering Applications of Artificial Intelligence 51 (2016) 191–201

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Recognizing emotions in text using ensemble of classifiers

Isidoros Perikos*, Ioannis Hatzilygeroudis

Department of Computer Engineering & Informatics, University of Patras, 26504 Patras, Greece

ARTICLE INFO

Available online 5 February 2016

Keywords:
Sentiment analysis
Emotion recognition
Text mining
Classifiers ensembles
Affective computing
Machine learning

ABSTRACT

Emotions constitute a key factor in human nature and behavior. The most common way for people to express their opinions, thoughts and communicate with each other is via written text. In this paper, we present a sentiment analysis system for automatic recognition of emotions in text, using an ensemble of classifiers. The designed ensemble classifier schema is based on the notion of combining knowledge-based and statistical machine learning classification methods aiming to benefit from their merits and minimize their drawbacks. The ensemble schema is based on three classifiers; two are statistical (a Naïve Bayes and a Maximum Entropy learner) and the third one is a knowledge-based tool performing deep analysis of the natural language sentences. The knowledge-based tool analyzes the sentence's text structure and dependencies and implements a keyword-based approach, where the emotional state of a sentence is derived from the emotional affinity of the sentence's emotional parts. The ensemble classifier schema has been extensively evaluated on various forms of text such as, news headlines, articles and social media posts. The experimental results indicate quite satisfactory performance regarding the ability to recognize emotion presence in text and also to identify the polarity of the emotions.

1. Introduction

Human cognition and emotions are innate and very meaningful aspects of human nature. Research in Artificial Intelligence area tries to explore and get a better understanding of the mechanism underlying behavior aiming to give computer systems and applications the ability to recognize aspects of human nature, like emotions. Emotions constitute a key factor of human intelligence, which provides indicative characteristics of human behavior, colors the way of human communication and can play an important role in human computer interaction. The role of emotions was initially investigated by Picard, who introduced the concept of affective computing (Picard, 1997), indicating the importance of emotions in human computer interaction and drawing a direction for interdisciplinary research from areas, such as computer science, cognitive science and psychology. The aim of affective computing is to enable computers to recognize the emotional status and behavior of a human and bridge the gap between the emotional human and the computer by developing systems and applications that can analyze, recognize and adapt to the user's emotional states (Calvo and D'Mello, 2010).

Human emotions can be expressed through various media, such as speech, facial expressions, gestures and textual data. The most common way for people to communicate with other and with computer systems is via written text, which is the main communication mean and the backbone of the web and of social media. Over the last years the advent of the Web and the raising of social media have changed completely the way of human communication as they provide new means that connect people all over the globe with information, news and events in real time. Also, they have changed completely the role of the users; they have transformed them from simple passive information seekers and consumers to active producers (Kanavos et al., 2014). Every day, a vast amount of articles and text messages are posted in various sites, blogs, news portals, e-shops, social networks and forums. The vast amount of web textual content necessitates automated methods to analyze and extract knowledge from it (Anusha and Sandhya, 2015; Shaheen et al., 2014).

Analyzing web content and peoples' textual messages with the aim to specify their emotional status is a very interesting and challenging topic in the microblogging area (De Choudhury et al., 2012). The massive and continuous stream of textual data in the web can reflect the writers' feelings, opinions and thoughts about various phenomena ranging from political events across the globe to consumer products. It can convey people's emotional status and substantial information about their beliefs and attitudes (Qiu et al, 2012). The analysis of the textual data is necessary for deeper

* Corresponding author.
E-mail addresses: perikos@ceid.upatras.gr (I. Perikos), ihatz@ceid.upatras.gr (I. Hatzilygeroudis).

http://dx.doi.org/10.1016/j.engappai.2016.01.012
0952-1976/

understanding a person's emotional status and behavior and in this line can provide very indicative factors regarding public attitude towards different events and topics and also can describe the emotional status of a community, a city or even a country. From a person-centric scope, analyzing the text messages of a specific person can provide very indicative factors of the person's emotional situation, his/her behavior and also provide deeper clues for determining his/her personality (Qiu et al., 2012). Furthermore, regarding news, articles and people comments, from a topic-center perspective, the analysis of the people comments on a specific topic can provide very meaningful information about the public stance, feelings and attitude towards various topics and events. In this line, emotion models can be employed to understand how people feel about a given entity such as a movie, a topic or a live event (Wang and Pal, 2015).

However, the development of systems and applications for automatically analyzing natural language with the aim to understand its sentimental content is a very hard process. Several studies have shown that analyzing and recognizing emotion in text documents is considered to be a very complex, NLP-complete problem and the interpretation varies depending on the context and the world knowledge (Shanahan et al., 2006). Also, it is pointed that sentiment analysis and emotion recognition approaches should move towards content, concept, and context-based analysis of natural language text and also support time efficient analysis techniques suitable for the special needs of the analysis of the vast web content and the big social data (Cambria et al., 2013). This work is a contribution towards this direction.

In this paper, we present an ensemble classifier system for sentiment analysis of textual data. The ensemble schema seek to effectively integrate different types of learners and classification methods aiming to overcome the drawbacks of each one and benefit from each ones advantages and in this line, improve the overall performance of the sentiment classification. The system is based on three main learners, two statistical learners and a knowledge based classifier tool, ensembled on a majority voting approach. The statistical learners are a naïve Bayes learner and maximum entropy leaner, which are trained on ISEAR (International Survey on Emotion Antecedents and Reaction) (Scherer and Wallbott, 1994) and Affective Text (Strapparava and Mihalcea, 2007) datasets. The knowledge-based tool analyzes sentence's structure using tools such as Stanford parser (de Marneffe et al., 2006) to specify word dependencies and uses WordNet Affect (Strapparava and Valitutti, 2004), lexical resources to spot words known to convey emotions. Then, it specifies each emotional word's strength and determines the sentence's emotional status based on the sentence's dependency graph in an approach where the overall sentence emotional state is derived by the emotional affinity of the sentence's emotional parts. The ensemble classifier system performs sentiment analysis on sentence level and so, a new text is initially split in sentences and each sentence is forwarded to the ensemble classifier schema, where features are extracted, represented as bag-of-words, and then handled by the statistical classifiers. The ensemble classifier determines whether the sentence is emotional or neutral and, in case it is emotional, determines the underlying emotional polarity.

The rest of the paper is structured as follows. Section 2 presents background topics on textual emotion recognition and ensemble classifiers. Section 3 presents related work. Section 4 presents the ensemble classifier system, describes its architecture and analyzes its functionality. Section 5 presents the evaluation study conducted and analyzes its performance results. Finally, Section 6 concludes the work presented in this paper and draws directions for future work.

2. Background topics

2.1. Emotion models

What is and what defines an emotion is a philosophical question that remains open for more than a century. In general, emotion is considered to be a strong feeling deriving from one's circumstances, mood or relationships with others (Oxford Dictionary, 2008). The way that emotions are represented is a basic aspect of an emotion recognition system (Reisenzein et al., 2013). The most popular models for representing emotions are the categorical and the dimensional models. The categorical model assumes that there is a finite number of basic and discrete emotions, where each one is serving a particular purpose. On the other hand, the dimensional model follows a different way and represents emotions in a dimensional approach. In this approach, dimensional model assumes that an emotional space is created and each emotion lies in this space.

A very popular and widely used categorical model is the Ekman emotion model (Ekman, 1999), which specifies six basic human emotions: anger, disgust, fear, happiness, sadness, surprise. These emotions are characterized as universal, as they are expressed in the same way across different cultures and eras. Ekman's emotion model has been used in several research studies and in various systems that are used to recognize emotional state from textual data and facial expressions. Another model that is also adopted in many studies on human emotion recognition is the Ortony–Clore–Collins (OOC) emotional model (Ortony et al., 1988). OOC model specifies 22 emotion categories based on human emotional reactions to various situations and it is mainly designed to model human emotions in general. Also it has been established as the standard model for emotion synthesis and is mainly utilized in systems that reason about emotions or incorporate emotions in artificial characters. Parrot's model (Parrott, 2001), constitutes of a group of six basic emotions, which are: love, joy, surprise, anger, sadness and fear, and also created a tree structure of emotions consisting of three levels. The first level of this classification model consists of the aforementioned six basic emotions and each level refines the granularity of the previous level, making abstract emotions more concrete. Parrot's model identifies over 100 emotions, conceptualized in a tree structured list and is considered to be the most nuanced classification of emotions.

Plutchik's model of emotions (Plutchik, 2001) is a dimensional model which offers an integrative theory based on evolutionary principles and defines eight basic bipolar emotions. These eight emotions are organized into four bipolar sets: joy vs. sadness, anger vs. fear, trust vs. disgust, and surprise vs. anticipation. Each emotion can be further divided into three degrees, for example, serenity is a lesser degree of joy and ecstasy is a more intense degree of joy. Also, the eight basic emotions can be combined to form feelings. For example, joy and trust can be combined to form love. Russell (1980) proposed the circumplex model of emotions, where emotions are represented in a two-dimensional circular space. The one dimension of the space is used to represent the emotion's polarity and the other dimension the emotion's activation. The polarity dimension characterizes an emotion as positive or negative, whereas the activation characterizes an emotion as activated or deactivated.

In our work, the ensemble classifier system utilizes the Ekman's basic emotions and the two dimensional model of Russell, characterizing basic emotions in terms of polarity as either positive or negative. The Ekman emotion model was adopted since it is the basic model for recognition of emotional content not only in facial, but also in textual data (see Related Works section), and also is can be scaffold by available lexical resources. Also, Russell's scale is used in order to quantitatively describe emotions and, in this scale, each emotion can be placed on the two dimensional plane with polarity and activation as the horizontal and vertical axes.

2.2. Sentiment analysis methods and resources

Sentiment analysis and classification techniques can be divided into two main kinds of approaches, the knowledge-based approaches and the machine learning approaches, which are widely used for the automatic recognition of emotions in text (Chaffar and Inkpen, 2011). In general, machine learning approaches utilize machine learning algorithms to train learners using textual features of articles and make predictions regarding the sentimental content of new articles. In most cases, the classification learners are trained on an annotated corpus and rely on text features such as the bag-of-words text representation, according to which a document is represented as a binary or frequency-based feature vector of the tokens it contains, regardless of their position in the text. On the other hand, the knowledge based approaches utilize sentiment lexicons, which are collections of known and precompiled sentiment terms (Medhat et al., 2014). In general, they depend on analyzing text sentences, finding sentimental seed words and then search the dictionary of their synonyms and antonyms.

To enhance the knowledge of sentiment analysis systems and their efficiency in recognizing emotions and sentiments in text, several lexical resources have been developed. One of the first general-purpose computerized text analysis resources is the General Inquirer (Stone et al., 1966) which was developed by IBM and has 11,788 words labeled with 182 categories of word tags, including positive and negative polarity and can provide a binomial classification (either positive or negative) of sentiment bearing words. The General Inquirer relies on Harvard psychological dictionaries that were correlated with states, motives, social and cultural roles and also various aspects of general distress. The Affective Norms for English Words (ANEW) lexicon (Bradley and Lang, 1999) provides a set of normative emotional ratings for a large number of words in the English language and this set of verbal materials have been rated in terms of pleasure, arousal and dominance. It rates the words on the dimensions of pleasure (from pleasant to unpleasant), arousal (from calm to excited) and dominance (from control to out-of-control), and so, it can assist methods to take the valence dimension and classify the pleasant terms as positive and the unpleasant terms as negative to perform a polarity analysis.

A popular and wide used lexicon is the WordNet Affect (Strapparava and Valitutti, 2004) which is based on the WordNet database and extends it by adding subsets of synsets suitable for the representation of affective concepts. These synsets added are annotated and associated with one or more affective labels. On WordNet database is also based the SentiWordNet 3.0 (Baccianella et al., 2010), which associates each synset with three numerical scores, where each one indicates the degree that the synset is objective, positive or negative. SentiWordNet 3.0 includes around 200,000 entries and uses a semi-supervised method to assign each word with positive, negative and objective scores. In our approach, the Wordnet Affect lexicon is utilized by the knowledge based tool in order to assist the spotting of words known to convey emotions and also the specification of their emotional content.

2.3. Ensembles of classifiers

The combination of classifiers is an effective method for improving the performance of a classification system (Li et al., 2007). The aim of the ensemble is to benefit from the learners advantages and to minimize its ones drawbacks. The design and development of effective classifier ensembles requires that the used learner units have some level of diversity. In the literature, ensemble classifiers have applied successfully in various subdomains of text mining, such as named entity recognition, word sense disambiguation and text classification (Xia et al., 2011). In general, classifier ensemble methods rely on a set of classifiers and combine them in order to make a classification decision. Classic machine learning methods train by using a simple classification method on the domain's data, while classifier ensembles train by using multiple different classifiers.

There are many reasons for designing, developing and using classifier ensembles as indicated by Dietterich (2000). From a statistical scope, by constructing an ensemble schema out of trained classifiers, the algorithm can average their votes and reduce the risk of choosing the wrong or underperforming classifier on new data. Even when different classifiers are trained and report a good performance, when just one is chosen, it may not yield the best generalization performance in unseen data. From a computational perspective, many learning algorithms work by performing some form of local search and it is very possible to get stuck at a local optimum. So, an ensemble constructed by running the local search from many different starting points may provide a better approximation to the true unknown function than any of the individual classifiers. Finally, from a representational scope, in some cases the decision boundaries that separate data from different classes may be too complex and an appropriate combination of classifiers can make it possible to cope with this issue. In this line, given the characteristics of the textual data, the utilization of ensemble classifier methods seems to be a suitable and interesting approach and the work presented in this paper is a contribution towards examining this direction.

3. Related work

In recent years, recognizing emotions and analyzing human behavior have attracted the attention of researchers in computer science, natural language processing and sentiment analysis. In the literature, there is a huge research interest and many studies on the design of methods and the development of systems for the sentiments analysis of text. A detailed and complete overview of approaches can be found in Medhat et al. (2014), Liu and Zhang (2012), Vinodhini and Chandrasekaran (2012). Several works study the way human express emotions and try to specify emotions in news, web blogs, forums and social media (Thelwall et al., 2012; Cambria et al., 2013; Liu, 2015).

A wide range of works and approaches to recognize emotions utilize a machine learning approach, which relies on training machine learning algorithms to solve the emotion recognition as a regular text classification problem utilizing syntactic and/or linguistic features. A first work in the field was presented in Alm et al. (2005), which explores machine learning methods for automatic classification of sentences in children fairy tales. The authors developed a corpus consisting of fairy tales sentences, which were manually annotated with emotional information and explore sentence's classification according to the Ekman's emotion categories with satisfactory results.

The work presented in Neviarouskaya et al. (2007) aims to recognize Ekman's six basic emotions in online blog posts. The authors analyze the posts using Machinese Syntax parser, spot emoticons and keywords that may appear in the posts and use a rule-based approach to determine sentence emotional content. The system developed reports approximately 70% agreement with human annotators in recognition of emotional content of sentences.

Moreover, in Brilis et al. (2012) machine learning approaches to classify song lyrics into mood categories using are examined. The song lyrics undergo pre-processing steps such as stemming, stop words and punctuation marks removal. Then the lyrics are classified into mood categories using a bag-of-words approach, where each word is accompanied by its frequency in the song and its TF-IDF (term frequency-indirect document frequency) score. Machine learning classifiers are trained and utilized and authors indicate that Random Forest algorithm reported the best results with approximately 71.5% accuracy on stemmed dataset and 93.7% on unstemmed.

In the work presented in Danisman and Alpkocak (2008), authors present an approach based on vector space model to classify emotional text. They use the ISEAR (International Survey on Emotion Antecedents and Reaction) and the SemEval datasets and the classification, which focuses on the classification of emotions and valence in text is made based on vector space model on a total of 801 news headlines provided by the Affective Task of SemEval 2007. Authors report that the vector space model classification model can give better performance results than other classifiers including Concept Net, Naïve Bayes, and support vector machines.

In Ho and Cao (2012), authors present a hidden markov logic approach to specify the most probable emotion of a given text. Authors report an *F*-score of 35% on the ISEAR dataset. The low accuracy was mainly due to the fact that authors ignored semantic and syntactic features of the analysis of the sentence, something that made it non-context sensitive.

A number of approaches to recognize emotions utilize the information included in affect dictionaries either exclusively or in addition to other features. In Osherenko and André (2007), a statistical approach for affect sensing in textual data is presented. Authors address the textual affect recognition task for spontaneous utterances based on affective qualities of words and classify emotions by using word counts. In their experiments, they consider affective dictionaries and general purpose lexical resources and indicate that affective annotations seem to provide a good means to reduce the number of features for emotional classification task.

In Chaumartin (2007), author examines a knowledge based approach for recognizing emotions in text and also present a tool he developed. The tool utilizes versions of SentiWordNet lexical resource, a subset of WordNet-Affect and also manual added words and detects contrasts between positive and negative words that shift emotion valence. It also uses Stanford parser to find the head word in a sentence that is considered to have the major importance and also to detect contrasts between positive and negative words that shift valence. In the experiments, the tool reports approximately 89% accuracy on headlines sentences and authors argue that working with linguistic techniques and a broad-coverage lexicon can be a viable approach to sentiment analysis of headlines.

Authors in the work presented in Ptaszynski et al. (2013) perform textual emotion analysis of Japanese narratives. In their research, they address the problem of person related affect recognition and they extract emotion subject from a sentence based on analysis of anaphoric expressions at first. Then, the affect analysis procedure estimates what kind of emotional state each character is in, for each part of the narrative. They use the ML-Ask (eMotive eLement and Expression Analysis system), a keyword based language dependent system for automatic affect annotation on utterances in Japanese, and are able to extract emotion types, including "joy", "fondness", "relief", "fear", "sadness", or "anger" from narrates with a performance of 0.60 Also, their approach is able to specify whether a sentence is emotional or not with approximately 90% accuracy.

Over the recent years, ensemble classification approaches are examined and their performances are studied on various types of textual data. In the work presented in da Silva et al. (2014), authors explore tweet sentiment analysis using classifier ensembles. A classifier ensemble is formed using the base machine learning classifiers: random forest, support vector machines, multinomial naïve Bayes and logistic regression. In their study, authors experimented with a variety of tweet datasets and report that the classifier ensemble can improve classification accuracy. Also, they have compared strategies for the representation of tweets, like bag-of-words and feature hashing, and indicate that bag-of-words representation can achieve better accuracy.

In the work presented in Xia et al. (2011), authors study an ensemble classifier for sentiment classification and use an ensemble schema combining three algorithms: naïve Bayes, maximum entropy and a support vector machine, to recognize polarity (positive or negative) in text. The classifiers utilize part-of-speech based feature sets and word-relation based feature sets and authors indicate that the ensemble of classification algorithms on the same feature set perform robustly better than individual classifiers.

In Wang et al. (2014), authors experimented with the performance of an ensemble classifier consisted of five base learners, that is naïve Bayes, maximum entropy, decision tree, k-nearest neighbor and support vector machine combined using random subspace method. Results indicate that the ensemble classifier substantially improve the performance of the individual base learners and reports better results than using solely the base learners and so, in this line, authors suggest that ensemble learning methods have the potential and can be used as a very viable approach for sentiment classification.

The ensemble classifier approaches in the literature mainly rely on sole machine learning classifiers. However, the machine learning approaches in general ignore semantic and syntactic features of the analysis of the sentences, something that made them non-context sensitive. On the other hand, the classification methods based only on keywords can suffer from the ambiguity in the keyword definitions in the sense that a word can have different meanings according to its usage and context and also the incapability of recognizing emotions within sentences that do not contain emotional keywords (Shaheen et al., 2014). So, based on the above, an ensemble classifier approach that would combine both machine learning and knowledge-based approaches could be of great interest. In addition, our work presented in this paper is, to the best of our knowledge, one of the first approaches in the sentiment analysis domain to examine this direction.

4. Emotion recognition system

In this section, the ensemble classifier developed to analyze natural language and recognize the emotional content of text, is presented and its functionality is illustrated. The analysis of the natural language is conducted at sentence level, so a given document is split in sentences. Many documents and articles may contain various emotional states, even about the same entities. So, systems and approaches that want to have a more fine-grained view of the different sentiments expressed in a document regarding entities or the writer's feelings, must deal with sentence level (Feldman, 2013).

The ensemble is based on three main classifiers. Two classifiers follow a statistical approach and one follows a knowledge-based one. More specifically, a naïve Bayes and a maximum entropy learners are trained to recognize sentiments in textual data. The learners are trained using the International Survey on Emotion Antecedents and Reaction (ISEAR) and the Affective text datasets. The knowledge-based tool performs a deep analysis of the natural language structure, specifies word dependencies and determines the way words are connected in order to specify words known to convey emotional content. The sentence's structure is analyzed using tools, such as Stanford parser (de Marneffe et al., 2006), and lexical resources, such as WordNet Affect (Strapparava and Valitutti, 2004), are utilized to spot words known to convey emotion. Then it specifies each emotional word's strength and determines the sentence's emotional status based on the sentence's dependency graph. The architecture of the ensemble classifier is illustrated in Fig. 1.

The ensemble schema connects the three classifiers on a majority voting approach. The ensemble classifier, given a new text document, initially splits it in sentences and each sentence is analyzed and its features are extracted. The ensemble classifier determines if the sentence is emotional or neutral, and in case it is emotional, determines the underling emotional content polarity.

4.1. Feature representation

The way that a document is analyzed and how it is represented is important for the performance of a machine learning approach. In this work, for the representation of a natural language text, we use the bag-of-words (BOW) representation technique. BOW is used more often because of its simplicity for the classification process. It is widely used in text mining applications in combination with removal of stop-words and stemming of useful words. In this approach, a document is considered to be an unordered collection of words, whereas the position of words in the document bears no importance. In the system, a new sentence initially is tokenized and is broken up into words and then each word is lemmatized and its base form is specified. Also, the sentence's stop words are filtered out and the sentence features are forwarded to the base learners.

4.2. Ensemble classifier voting

The ensemble classifier adapts a majority voting approach to make a classification decision based on the outputs of each base classifier. Each classifier has a vote that, for each text sentence, is a class determined by the classifier. The majority voting approach is considered to be the simplest and most intuitive method for combining classifier outputs (Kuncheva, 2004). In general, the majority vote counts the votes for each class over the input classifiers and selects the majority class. The idea of selecting a number of classifiers to make up an ensemble instead of using all classifiers has been dealt with in different ways. Theoretically, if the base classifiers selected can make independent errors, it is proven that the majority vote is suitable and can outperform the best classifier (Orrite et al., 2008).

4.3. Naïve Bayes classifier

Naïve Bayes is a simple model for classification and can achieve good performance in text categorization. It is based on Bayes theorem and is a probability based classification approach that assumes that documented words are generated through a probability mechanism. In general, the lexical units of a corpus are labeled with a particular category or category set and are processed computationally. During this processing, each document is treated as a bag-of-words, so the document is assumed to have no internal structure, and no relationships between the words exist. A universal feature of Naïve Bayes classification is the conditional independence assumption. Naïve Bayes assumes that words are mutually independent and so, each individual word is assumed to be an indication of the assigned emotion. The Bayesian formula calculates the probability of a defined class, based on document's features and is calculated as:

$$P(cs) = \frac{P(c)P(s|c)}{P(s)}$$

where $P(c)$ is the probability that a sentence belongs to category c, $P(s)$ is the probability of sentence s occurrence, $P(s|c)$ is the probability that the sentence s belongs to category c and $P(c|s)$ is the probability that given the sentence s it belongs to category c. The term $P(s|c)$ can be computed taking into consideration the conditional probabilities of occurrences of sentence's words given the category c, as follows:

$$P(sc) = \prod_{1 \le k \le n} P(s_k(c)$$

where $P(s_k|c)$ represents the probability that term (word) s_k occurs given the category c and n represents the length of sentence s.

4.4. Maximum Entropy classifier

The Maximum Entropy classifiers are feature based models that prefer the most uniform models that satisfy a given constraint. Labeled data in training phase are used to derive the constraints for the model that characterize the class. In contrast to Naïve Bayes, the Maximum Entropy classifier does not make independence assumption for its features. So, it is possible to add features to a Maximum Entropy classifier like word unigrams, bigrams and N-grams in general, without worrying about the overlapping of the features. Maximum Entropy classifiers can achieve very difficult classification tasks and indicate good performance in various natural language processing tasks such as sentence segmentation, language modeling and named entity recognition (Nigam et al., 1999). MaxEnt classifier can also be used when we cannot assume the conditional independence of the features, something that is particularly true in text mining and sentiment analysis problems, where features such as words are not independent. The Max Entropy classifier requires more time to be trained compared to Naïve Bayes, mainly due to the optimization problem that needs to be solved in order to estimate the parameters of the model. The two statistical, machine learning approaches (Naïve Bayes, MaxEnt) are fed with large training corpus of sentimental annotated texts in order to be trained aiming not only to learn the emotional status and strength of emotional words, as in knowledge-based

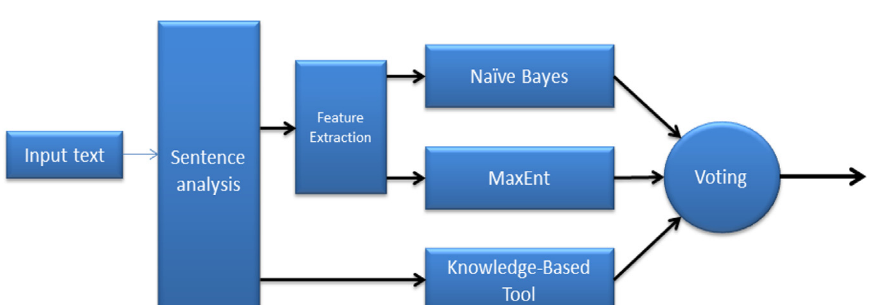

Fig. 1. An overview of the ensemble classifier architecture.

approaches that utilize lexicons, but also to take into account strengths of other arbitrary words, word co-occurrences, word frequencies and their combinations (Cambria et al., 2013).

The development of the learners was made in Python. For the training of the naïve Bayes classifier, the Python NLTK toolkit was utilized and for the training of the Maximum Entropy classifier the Python's TextBlob (Loria, 2014) module was used. The training of both classifiers was made based on the ISEAR and Affective Text datasets, which were enriched with additional sentences mainly from twitter posts and news articles. More specifically, the training data was enriched with neutral sentences and also with sentences that denote surprise. This was made to assure that all the emotional categories equally appear in the training dataset. That was necessary, given that the ISEAR dataset does not include sentences of the surprise emotional category.

4.5. Knowledge-based classification tool

The knowledge-based approach and the tool developed, in contrast to the statistical approaches, tries to analyze and extract knowledge from each sentence in order to specify its sentimental status (Perikos and Hatzilygeroudis, 2013). The architecture of the tool is depicted in Fig. 2. The tool performs sentiment analysis at a sentence level. It uses Tree Tagger (Schmid, 1994), a part-of-speech tagger, to specify each word's grammatical role in the sentence and its base form (lemma), and the Stanford parser (de Marneffe et al., 2006) to analyze sentence's structure and create the dependency tree based on the words' relationships. The Named Entity Recognizer (NER) tool (Finkel et al., 2005) is utilized to detect proper names and named entities that appear in the sentence aiming to assist the sentence analysis and the specification of the way that emotional parts are associated with sentence's entities, such as persons. Words known to convey emotions are spotted using the lexical resources of the knowledge base (KB) and each emotional word detected is further analyzed by the tool and its relations and the way it interacts with the sentence's words are determined. Based on the words' relationships, identifies specific types of emotional word's interactions with quantification words, in order to specify its emotional strength. Finally, the emotion extractor unit specifies the sentence's overall emotional status based on the sentence emotional parts.

In this way, the tool for a given natural language sentence proceeds as follows:

1. *Uses tree tagger to specify the words' lemmas and grammatical roles.*
2. *Uses Stanford parser to analyze sentence structure and get the dependencies and the dependency tree.*
3. *Uses NER to recognize named entities and persons.*
4. *For each word uses the knowledge base to determine whether it is emotional or not. If it is,*
   4.1 *Analyzes its relationships,*
   4.2 *Checks if a modification relationship with quantification words exists, analyzes it and determine emotion strength.*
   4.3 *Analyzes the dependency tree, recognizes sentence pattern/structure and based on it, determines the sentence's emotional content.*

The tool's knowledge base (KB) is used to store information regarding emotional words known to convey emotions. It relies on WordNet Affect lexical resource, a widely used extension of the Wordnet which was also extended by adding some more emotional words and their grammatical role. In addition, the knowledge base stores quantification words, which constitute a special type of words that can quantify and appraise the content of emotional words. Some examples of such words are presented in Table 1. The knowledge base assists the determination of the emotional content of a sentence based on the analyses performed by Tree tagger and Stanford parser and the linguistic knowledge it holds, as presented above.

The system, given a natural language sentence, initially uses Tree Tagger, a well-known statistical morphosyntactic part of speech tagger and lemmatizer, to specify for each word its base form (lemma) and its part of speech tag, identifying its grammatical role in the sentence. Part of speech tagging is a fundamental process in a NLP system and gives a first level analysis of words' roles in the sentence.

Then, Stanford parser, which is a very popular morphosyntactic analysis tool, is used to perform a deeper analysis of the sentence's structure. Stanford parser analyzes the structure of a sentence, specifies the relationships between the sentence's words and determines the corresponding dependencies. The dependency tree represents the grammatical relations between the sentence's words in a tree based approach. Those relationships are presented as triplets consisting of the name of the relation, the governor and the dependent respectively. Dependencies indicate the way that words are connected and interact with each other. When the sentence morphosyntactic analysis is completed and the dependency tree is created, special parts of the dependency tree and specific words are further analyzed. The dependency tree is analyzed and the relationships and types of interactions/connections between the sentence words are examined. After that, the system specifies the existence of named entities in the sentence.

The determination of the named entities is conducted with the utilization of Stanford Named Entity Recognizer tool (Finkel et al.,

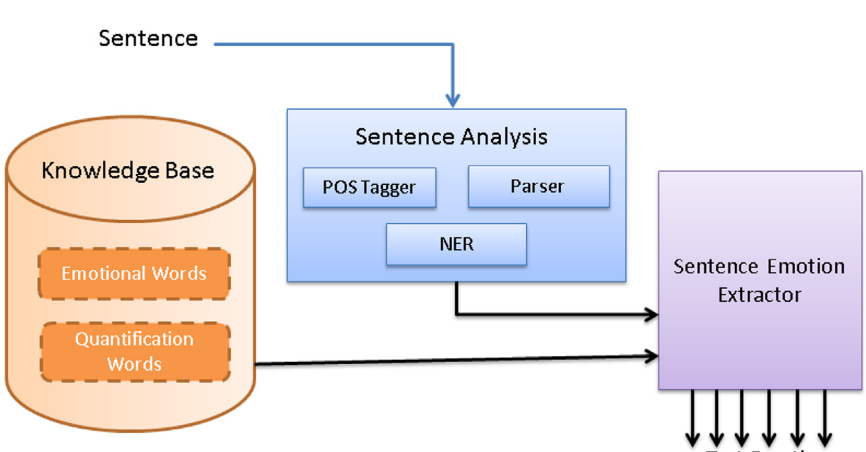

Fig. 2. Overview of the knowledge-based classification tool architecture.

2005). The tool can label sequences of words in a sentence, which are the names of things, such as person and company names, with the proper category label. Examples of named entities are, a person (e.g. John), a country (e.g. Greece), a city (e.g. Athens) etc. The specification of the named entities of the natural language sentence can assist in the analysis of the sentence structure and the way that emotional parts are connected with entities, such as persons.

As an example, consider the sentence "She kissed her aunt with great happiness". In Fig. 3, the sentence parse tree and the dependencies as specified by the Stanford parser are presented.

The nodes of the dependency tree are the sentence's words and the edges specify the existing relationships between the words. For each word, its grammatical role in the sentence and the way it interacts with other words are specified. The interaction between two words is denoted by the existence of an edge and the exact type of interaction is denoted by the edge's name. For example, nsubj (she, kissed) is a nominal subject relationship between the two words, defining that the word 'she' is the subject of the word 'kissed'.

4.5.1. Formulation of emotional units

After a given natural language sentence is analyzed, the system proceeds in spotting emotional words and formulating emotional units. To do so, the system utilizes lexical resources to spot words known to convey emotional context. The knowledge base, as mentioned above, stores information about (a) emotional words and (b) quantification words.

The system, for every word of the sentence, searches to see if it is stored as an emotional word in knowledge base. If the tool's knowledge base has recorded the word as emotional, it returns the emotional category the word belongs to. If the system does not find any same emotional word in the sentence, then it performs a deeper analysis of the sentence's words in terms of synonyms and antonyms. The assumption is that a word's synonym or an antonym may be recognized as an emotional word and thus an emotional content may underlie in the word.

The knowledge base also stores information about quantification words that quantify and modify the strength of emotional words through interacting with them. So, we developed a list of *quantification words,* such as: {very, some, all, hardly, less etc}. A special category of the quantification words are words that denote negation. The *negation words*, when appear in a sentence can flip the polarity of the words that interact with. Examples of negation words are: {none, no, not, never, nobody}. For each one, its modification impact on words that interacts with is specified. So, words like 'very' and 'great' have a strong positive impact on words that are related with, increasing their emotional content, while negation words flip the emotional content. In Table 1, example quantification words and their impact on emotional words are presented.

Table 1
Example quantification words and their impact.

Modification impact	Words
High	Very, great, huge, extensive
Average	Hardly, quite
Low	Little, less
Flip	No, not

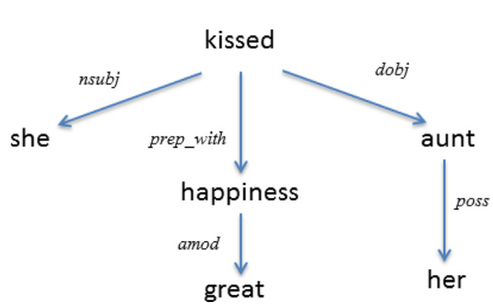

Fig. 3. The sentence's dependency tree.

Low value is set to 20, average to 50 and high to 100. So, the quantification word 'extremely' when modifies an emotional word has an emotion strength set to 100 (max emotional strength), whereas the quantification word 'quite' sets the emotion strength to 20. These values have been specified based on empirical and experimental studies.

After emotional words are recognized, the system performs a deeper analysis regarding their role in the sentence and the type of their relationships/connections with other words. More specifically, it tries to analyze special types of relationships that may appear with quantification words. These relationships are recognized as 'mod' dependencies by the Stanford parser. So, these dependencies, connecting emotional words with quantification words, are analyzed and a quantification word's impact defines the strength of the connected emotional word.

Finally, consider negation cases such as 'is not very furious', where the emotional word 'furious' is detected to denote 'anger'. The word 'very' has a high quantification to the emotion and the negation detected flips/reverse the quantification to 'low' and thus this sentence part emotional content is determined to be low (20) 'anger'.

4.5.2. Determining sentence emotional content

After the emotional words of the sentence are recognized and their strengths are determined, the system specifies the sentence's overall emotional content. To do so, it analyzes the sentence's structure. More specifically, it recognizes and analyses the basic pattern of the sentence consisting of the sentence's main verb, the object and the subject of it. So the pattern "Subject–Verb–Object" is extracted from the sentence based on the dependencies. This pattern is the backbone of the sentence structure and holds the core meaning of the sentence. Moreover, analyzing it can help the system in understanding the interactions of the sentence parts that is the way the emotional parts are connected. So, the system processes the sentence structure as follows:

1. *Analyze the sentence dependencies and extract the subject-verb-object pattern.*
2. *For each grammatical role of the pattern (e.g. object or verb or subject).*
   - 2.1 *Specify whether it is an emotional part,*
   - 2.2 *Analyze its relationships with emotional parts (if any),*
   - 2.3 *Specify its emotional content.*
3. *Combine emotional contents of the parts to specify the sentence overall emotions.*

4.6. Specify emotional polarity

According to Russell's two-dimensional model of affect (Russell, 1980), emotions can be presented in a dimensional space of two dimensions, where the one dimension represent the emotion's polarity and the other dimension the emotion's activation. The polarity dimension characterizes an emotion as positive or negative, whereas the activation characterizes an emotion as activated or deactivated. In this representation approach, four areas of emotions are created: activated-positive, activated-negative, deactivated-positive and deactivated-negative (Ptaszynski et al., 2009). In Fig. 4, the emotional space and the mapping of emotions are depicted.

The mapping enables the system to specify the polarity of a sentence based on its underlying emotional content. That is, in case a sentence is recognized by the system to convey emotions, its emotional content is specified and then, the polarity of the sentence is determined based on the mapping on Russell's space. The joy emotion is associated with positive polarity, while the emotions of anger, disgust, fear and sadness characterize a sentence as negative (Chaumartin, 2007). In this line, the surprise emotion can characterize a sentence either as positive, in cases it is accompanied with joy emotion (happy surpise), or as negative, in cases where it is associated with fear, anger, sadness and disgust emotions. The system adopts this emotional mapping and it relies on it in order to specify the emotional polarity of a sentence based on the emotional content of the sentence.

5. Evaluation study

An extended evaluation study was designed and conducted to analyze the system's performance. Initially, for the evaluation, different types of textual data were used to assess the system's performance and also provide a deeper insight of the system's performance on different textual data and sources.

5.1. Data collection

For evaluation purposes, we created a corpus of textual data from different sources and a human expert was used to manually annotate each one. The total number of the sentences of the created dataset was 750. They were selected from different sources, such as news headlines, news articles and Twitter posts. Articles and article headlines were harvested from news portals such as BBC, CNN and Euronews. A total number of 250 headline sentences and 250 sentences from article contents were selected for the formulation of the corpus. The sentences were selected not to be very lengthy and also not suffering from many ubiquities and anaphoric expressions. Also, for the corpus creation, users' posts from Twitter were selected. The posts were selected from different users and from various topics. A total of 250 various emotional posts were selected.

After the corpus formulation, the sentences of the corpus were annotated by the human expert. The annotation stage was conducted via the system in a scenario, where each sentence was presented to the annotator and through the interface the annotator determined a couple of parameters for the sentence's emotional content. More specifically, during the annotation stage, for each sentence were determined by the human annotator (a) the existence and the (b) degree of each of the six basic emotions. The emotion level ranges from 0 to 100, where 0 is used to denote the absence of a specific emotion and 100 denotes that the specific emotion is very strong. Based on the expert's annotations, the sentence emotional polarity is specified, characterizing the sentence as positive, negative or neutral based. The annotations of the human expert are used as a 'gold standard' for the evaluation of the system.

5.2. Evaluation results

An evaluation study was conducted in order to assess the developed mechanism and provide an insight of its performance. The evaluation conducted in two parts. The first one evaluates the system performance in recognizing emotions present in natural language and the second part evaluates its performance in recognizing the emotional polarity of the emotional sentences. More specifically, in the first part of the evaluation study, the system was evaluated on characterizing a natural language sentence as either emotional, in case it conveys emotion(s) and generates feeling/s to the human reader, or neutral in case of emotional absence. The system classifies a sentence either in the emotional or in the neutral class. Evaluation was based, given the binary output, on the following metrics which are: accuracy, precision, sensitivity and specificity, defined as follows:

$$acc = \frac{TP+TN}{TP+FP+TN+FN}, prec = \frac{TP}{TP+FP}, sen = \frac{TP}{TP+FN}, spec = \frac{TN}{FP+TN}$$

where TP denotes the number of valid cases correctly classified, FP is the number of invalid cases that are misclassified, TN is the number of invalid cases correctly classified and FN is the number of valid cases that are misclassified. The performance results of the systems are illustrated in Table 2.

The results show a very good performance of the three classifiers and the ensemble classifier schema. The ensemble schema performs robustly better in all the experiments than the sole classifiers. This is due to the fact that the classification is performed with very good performance by each one of three main classifiers of the ensemble schema and in many sentences that one of the classifiers fails to make a correct prediction the final prediction is corrected by the remaining two. The analytic performance results of the classifiers and the ensemble schema for each type of the textual data are illustrated in Table 3.

The best performance of a sole classifier is achieved by the Naïve Bayes classifier, which accuracy and precision are better than those of the knowledge based tool and the MaxEnt in all the experiments. Its best performance is achieved in the news headlines, while its performance decreases in user post in twitter. The MaxEnt classifier performance is slightly lower than that of naïve

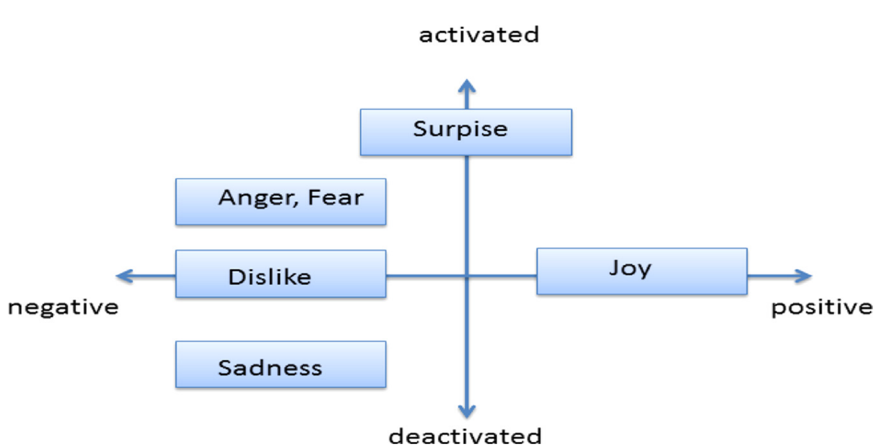

Fig. 4. Emotions mapping on Russell's space.

Bayes. It goes better in headline and article sentences, but its performance decreases in tweets almost in the same degree as naïve Bayes. The knowledge based tool performs very well in news headlines. This is mainly due to the fact that headlines are short and in most cases emotions are expressed with expressive emotional keywords and highly sentimental words. Also, headline and article sentences have well grammatical structure and proper syntax. A notable point of the knowledge based tool concerns its performance to recognize neutral sentences. Indeed, the high sensitivity denotes that the sentences that were neutral were recognized by the tool and classified correctly in the neutral category. Additionally, the low specificity is mainly a result of classifying some sentences that were emotional in the neutral category. In most of these cases, the emotion in the sentences was expressed without the existence of strong emotional words. On the other hand, the lower performance of the knowledge based tool may due to the test dataset, probably being in close line with the training sets.

Regarding the textual characteristics of the sentences and their sentimental analysis, classifiers report a better performance in headlines and lower in tweets. The reason is that headlines and article sentences are following grammatical rules and express emotions in a direct approach, using words that convey and generate emotion to the reader. On the other hand, tweets do not have a good grammatical syntax and in many cases express emotions in an indirect way, without strong emotional words.

As far as the second part is concerned, an evaluation of the mechanism's performance in determining the emotional polarity of an emotional sentence and its characterization as either positive or negative was conducted. To this end, the sentences that were characterized as emotional, and were indeed emotional, were selected. The performance of the classifiers is presented in Table 4.

The recognition of the emotional polarity of the sentences is conducted with very good performance by the three classifiers and this is a reason for the superior performance of the ensemble classifier. The results of the classifiers on the different data are presented in Table 5.

The results show that the ensemble schema is performing very well in recognizing the emotional polarity of the sentences. The three classifiers report quite good performance in recognizing the emotional polarity of headlines and article posts and a lower in Twitter posts. The ensemble classifier performance is better than the sole classifiers in headlines and posts while in Twitter its performance is almost the same with the Naïve Bayes classifier.

The knowledge based tool also indicates a better performance in headlines and articles than in Tweets. Its performance to specify the polarity of emotional sentences is associated with the tool's way to handle negations. The tool handles negations in a more systematic approach than the statistical learners and in syntactically well-structured sentences is able to correctly trace negations that inverse the status of the emotional words and also the sentence's polarity. However, its performance decreases in Tweets, since the tweets are very short, consisting of less than 10 words, in most cases, and have arbitrary and flighty structure. In addition, emotions in tweets are expressed in many cases in an ironic mode, based on comments or events of previous posts, and the emotional identification of such cases necessitates analysis and deeper understanding of the topic at hand and the conversation about it. Indeed, in contrast to headlines and articles that state and present facts and events, the sentimental analysis of tweets in many cases also necessitates the sentimental analysis of the topic of the conversation. In both experiments, the MaxEnt and Naïve Bayes learners achieve a better performance than the knowledge based tool in the analysis of the user post in Twitter. So, sentimental analysis of tweets is better to be conducted by statistical approaches, since the grammatical structure of the posts are flighty. The performance of the statistical classifiers in tweets may due to their training phase, since ISEAR and Affective text datasets consist of different types of sentences, which have different characteristics from the users' tweets. Also, the addition of more emotional annotated tweets in the training phase could increase the statistical classifiers performance.

6. Conclusions and future work

In this paper, an ensemble classifier system for the sentiment analysis of textual data is presented. It is based on three main classifiers, a naïve Bayes learner, a maximum entropy one and a knowledge based tool, which are combined via a majority voting approach. The naïve Bayes and the Maximum Entropy classifiers were trained on ISEAR and Affective text datasets. The knowledge-based tool performs a deep analysis of the natural language structure, specifies word dependencies and determines the way that words are connected. It utilizes Wordnet Affect to specify words that convey emotional content and specifies the sentence's emotional status based on the sentence's dependency graph in an approach where the overall sentence emotional state is derived by the emotional affinity of the sentence's emotional parts. The

Table 2
Classifiers performance.

Metric	KBtool	N.B.	MaxEnt	Ensemble classifier
Accuracy	0.77	0.85	0.80	0.87
Precision	0.72	0.89	0.85	0.91
Sensitivity	0.91	0.88	0.86	0.89
Specificity	0.59	0.78	0.68	0.82

Table 4
Evaluation results of emotional status.

Metric	KBtool	N.B.	MaxEnt	Ensemble classifier
Accuracy	0.77	0.84	0.81	0.86
Precision	0.82	0.89	0.86	0.87
Sensitivity	0.79	0.79	0.83	0.85
Specificity	0.74	0.87	0.82	0.87

Table 3
Recognizing emotion presence.

Metric	Headlines				Articles				Tweets			
	KBtool	N.B.	MaxEnt	Ensemble classifier	KBtool	N.B.	MaxEnt	Ensemble classifier	KBtool	N.B.	MaxEnt	Ensemble classifier
Accuracy	0.82	0.87	0.82	0.89	0.79	0.86	0.82	0.89	0.7	0.81	0.77	0.82
Precision	0.77	0.93	0.89	0.94	0.72	0.91	0.87	0.93	0.68	0.84	0.78	0.85
Sensitivity	0.92	0.86	0.86	0.9	0.92	0.9	0.85	0.9	0.9	0.87	0.86	0.88
Specificity	0.66	0.85	0.75	0.9	0.58	0.81	0.7	0.87	0.52	0.67	0.6	0.68

Table 5
Evaluation results of emotional status.

Metric	Headlines				Articles				Tweets			
	KBtool	N.B.	MaxEnt	Ensemble classifier	KBtool	N.B.	MaxEnt	Ensemble classifier	KBtool	N.B.	MaxEnt	Ensemble classifier
Accuracy	0.81	0.87	0.84	0.89	0.8	0.85	0.82	0.87	0.71	0.81	0.78	0.83
Precision	0.85	0.91	0.87	0.90	0.77	0.89	0.85	0.85	0.84	0.88	0.86	0.87
Sensitivity	0.85	0.84	0.82	0.91	0.82	0.76	0.80	0.86	0.71	0.77	0.87	0.79
Specificity	0.77	0.90	0.85	0.89	0.74	0.86	0.85	0.86	0.70	0.85	0.77	0.86

ensemble classifier performance emotion recognition on sentence level and so, a new text is initially split in sentences and each sentence is forwarded to the ensemble classifier schema, where features are extracted, represented as bag-of-words, are lemmatized and then handled by the statistical classifiers. The ensemble classifier determines whether a sentence is emotional or neutral, and in case it is emotional, specifies the underling emotional polarity. The experimental studies conducted and the results gathered indicate quite satisfactory performance regarding the ability to recognize emotion presence in text and also to identify the text's emotions polarity. The work indicates that ensemble technique is an effective way to combine different classification algorithms for better textual emotional classification. The ensemble schema performs better in both tasks than the sole classifiers.

As a future work, a larger scale evaluation will give us a deeper insight of the system’s performance. Also, the knowledge base of the tool currently utilizes Affective WordNet and manually added words to identify sentence's emotional words. An extension of system's knowledge base could be made by adding more lexical resources, such as General Inquirer and SentiWordNet resources. Also, another direction for future work will be to properly extend the training phase of the classifiers and the dependencies rules of the knowledge based tool in order to face cases that currently fail to be classified correctly, such as sentences “I laughed at him” and “He laughed at me”, which evoke different emotions depending on the first person's perspective. Furthermore, another extension may concern the associations of classifier weights in the voting approach, which could represent to some degree their classification confidence and the strength of the emotion specified. Also, a further research study may focus on the specification of the emotional polarity of the sentences and we plan to gradually extend the system to contribute in recognizing emotions in a more fine-grained scale and also assist in deeper understanding of emotion-expressing phenomena. Exploring this direction is a key aspect of our future work.

Acknowledgments

Part of this work was supported by the Research Committee of the University of Patras, Greece, Program “Karatheodoris”, Project no. C901.

References

Alm, C.O., Roth, D., Sproat, R., 2005. Emotions from text: machine learning for text-based emotion prediction. In: Proceedings of the conference on human language technology and empirical methods in natural language processing, pp. 579–586.

Anusha, V., Sandhya, B., 2015. A learning based emotion classifier with semantic text processing. In: El-Alfy, E.M., Thampi, S.M., Takagi, H., Piramuthu, S., Hanne, T. (Eds.), Advances in Intelligent Informatics. Springer International Publishing, pp. 371–382.

Baccianella, S., Esuli, A., Sebastiani, F., 2010. SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. LREC 10, 2200–2204.

Bradley, M.M., Lang, P.J., 1999. Affective norms for english words (ANEW): instruction manual and affective ratings (Technical Report C-1). The Center for Research in Psychophysiology, University of Florida, pp. 1–45.

Brilis, S., Gkatzou, E., Koursoumis, A., Talvis, K., Kermanidis, K.L., Karydis, I., 2012. Mood classification using lyrics and audio: a case-study in Greek music. In: Iliadis, L., Maglogiannis, L., Papadopoulos, H., Karatzas, K., Sioutas, S. (Eds.), Artificial Intelligence Applications and Innovations. Springer, Heidelberg, pp. 421–430.

Calvo, R.A., D'Mello, S., 2010. Affect detection: an interdisciplinary review of models, methods, and their applications. In: Proceedings of IEEE Transactions on Affective Computing, vol. 1, 1, pp. 18–37.

Cambria, E., Schuller, B., Xia, Y., Havasi, C., 2013. New avenues in opinion mining and sentiment analysis. IEEE Intell. Syst. 28 (2), 15–21.

Chaffar, S., Inkpen, D., 2011. Using a heterogeneous dataset for emotion analysis in text. In: Butz, C., Lingras, P. (Eds.), Advances in Artificial Intelligence. Springer, Berlin Heidelberg, pp. 62–67.

Chaumartin, F.R., 2007. UPAR7: a knowledge-based system for headline sentiment tagging. In: Proceedings of the 4th International Workshop on Semantic Evaluations, Association for Computational Linguistics, pp. 422–425.

da Silva, N.F., Hruschka, E.R., Hruschka, E.R., 2014. Tweet sentiment analysis with classifier ensembles. Decis. Support Syst. 66, 170–179.

Danisman, T., Alpkocak, A., 2008. Feeler: emotion classification of text using vector space model. In: Proceedings of AISB 2008 Convention Communication, Interaction and Social Intelligence, vol. 1, 2008, p. 53.

De Choudhury, M., Gamon, M. and Counts, S., 2012. Happy, nervous or surprised? Classification of human affective states in social media. In: Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM).

de Marneffe, M.C., MacCartney, B., Manning, C.D., 2006. Generating typed dependency parses from phrase structure parses. In: Proceedings of LREC, vol. 6, pp. 449–454.

Oxford Dictionary, English, 2008. Oxford English dictionary.

Dietterich, T.G., 2000. Ensemble methods in machine learning. In: Proceedings of the First International Workshop on Multiple Classifier Systems, MCS'00, Springer-Verlag, London, UK, 2000, pp. 1–15.

Ekman, P., 1999. Basic emotions. Handb. Cogn. Emot., 45–60.

Feldman, R., 2013. Techniques and applications for sentiment analysis. Commun. ACM 56 (4), 82–89.

Finkel, J.R., Grenager, T., Manning, C., 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, pp. 363–370.

Ho, D.T., Cao, T.H., 2012. A high-order hidden Markov model for emotion detection from textual data. In: Richards, D., Kang, B.H. (Eds.), Knowledge Management and Acquisition for Intelligent Systems. Springer, Berlin Heidelberg, pp. 94–105.

Kanavos, A., Perikos, I., Vikatos, P., Hatzilygeroudis, I., Makris, C., Tsakalidis, A., 2014. Conversation Emotional Modeling in Social Networks. In: Proceedings of 26th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 478–484.

Kuncheva, L.I., 2004. Combining Pattern Classifiers: Methods And Algorithms. John Wiley & Sons.

Li, S., Zong, C., Wang, X., 2007. Sentiment classification through combining classifiers with multiple feature sets. In: Proceedings of International Conference on Natural Language Processing and Knowledge Engineering, NLP-KE, 2007, pp. 135–140.

Liu, B., 2015. Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press.

Liu, B., Zhang, L., 2012. A survey of opinion mining and sentiment analysis. In: Aggarwal, C.C., Zhai, C. (Eds.), Mining Text Data. Springer, US, pp. 415–463.

Loria, S.,2014. TextBlob: Simplified Text Processing. TextBlob. Np.

Medhat, W., Hassan, A., Korashy, H., 2014. Sentiment analysis algorithms and applications: a survey. Ain Shams Eng. J. 5 (4), 1093–1113.

Neviarouskaya, A., Prendinger, H., Ishizuka, M., 2007. Textual affect sensing for sociable and expressive online communication. In: Paiva, A.C.R., Prada, R., Picard, R.W. (Eds.), Affective Computing and Intelligent Interaction. Springer, Heidelberg, pp. 218–229.

Nigam, K., Lafferty, J., McCallum, A., 1999. Using maximum entropy for text classification. In: IJCAI-99 Workshop On Machine Learning For Information Filtering, vol. 1, pp. 61–67.

Orrite, C., Rodriguez, M., Martinez, F., Fairhurst, M., 2008. Classifier ensemble generation for the majority vote rule. In: Ruiz-Shulcloper, J., Kropatsch, W.G. (Eds.), Progress in Pattern Recognition, Image Analysis and Applications. Springer, Berlin Heidelberg, pp. 340–347.

Ortony, A., Clore, G., Collins, A., 1988. The cognitive, Structure of Emotions. Cambridge University Press, Cambridge.
Osherenko, A., Andre, E., 2007. Lexical affect sensing: are affect dictionaries necessary to analyze affect?. In: Paiva, A.C.R., Prada, R., R.W., Picard (Eds.), Affective Computing and Intelligent Interaction, pp. 230–241.
Parrott, W.G., 2001. Emotions in Social Psychology: Essential Readings. Psychology Press.
Perikos, I., Hatzilygeroudis, I., 2013. Recognizing emotion presence in natural language sentences. In: Iliadis, L., Papadopoulos, H., Jayne, C. (Eds.), Engineering Applications of Neural Networks. Springer, Berlin Heidelberg, pp. 30–39.
Picard, R., 1997. Affective Computing. The MIT press, Cambridge, MA.
Plutchik, R., 2001. The nature of emotions. Am. Sci. 89 (4), 344–350.
Ptaszynski, M., Dokoshi, H., Oyama, S., Rzepka, R., Kurihara, M., Araki, K., Momouchi, Y., 2013. Affect analysis in context of characters in narratives. Expert Syst. Appl. 40 (1), 168–176.
Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R., Araki, K., 2009. A system for affect analysis of utterances in Japanese supported with web mining. J. Jpn. Soc. Fuzzy Theory Intell. Inform. 21 (2), 194–213.
Qiu, L., Lin, H., Ramsay, J., Yang, F., 2012. You are what you tweet: personality expression and perception on Twitter. J. Res. Pers. 46 (6), 710–718.
Reisenzein, R., Hudlicka, E., Dastani, M., Gratch, J., Hindriks, K., Lorini, E., Meyer, J.J., 2013. Computational modeling of emotion: toward improving the inter-and intradisciplinary exchange. Affect. Comput. IEEE Trans. 4 (3), 246–266.
Russell, J.A., 1980. A circumplex model of affect. J. Pers. Soc. Psychol. 39 (6), 1161.
Scherer, K.R., Wallbott, H.G., 1994. Evidence for universality and cultural variation of differential emotion response patterning. J. Pers. Soc. Psychol. 66 (2), 310.
Schmid, H., 1994. Probabilistic part-of-speech tagging using decision trees. In: Proceedings of the International Conference on New Methods in Language Processing, pp. 44–49.
Shaheen, S., El-Hajj, W., Hajj, H., Elbassuoni, S., 2014. Emotion recognition from text based on automatically generated rules. In: Proceedings of IEEE International Conference onData Mining Workshop (ICDMW), pp. 383–392.
Shanahan, J.G., Qu, Y., Wiebe, J., 2006. Computing Attitude And Affect In Text: Theory And Applications, vol. 20. Springer.
Stone, P.J., Dunphy, D.C., Smith, M.S., 1966. The General Inquirer: A Computer Approach to Content Analysis. The MIT Press.
Strapparava, C., Mihalcea, R., 2007. Semeval-2007 task 14: affective text. In: Proceedings of the 4th International Workshop on Semantic Evaluations, Association for Computational Linguistics, pp. 70–74.
Strapparava, C., Valitutti, A., 2004. WordNet-Affect: an affective extension of WordNet. In: Proceedings of LREC, vol. 4, pp. 1083–1086.
Thelwall, M., Buckley, K., Paltoglou, G., 2012. Sentiment strength detection for the social web. J. Am. Soc. Inf. Sci. Technol. 63 (1), 163–173.
Vinodhini, G., Chandrasekaran, R.M., 2012. Sentiment analysis and opinion mining: a survey. Int. J. 2, 6.
Wang, G., Sun, J., Ma, J., Xu, K., Gu, J., 2014. Sentiment classification: the contribution of ensemble learning. Decis. Support Syst. 57, 77–93.
Wang, Y., Pal, A., 2015. Detecting Emotions in Social Media: a constrained optimization approach. In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015), pp. 996–1002.
Xia, R., Zong, C., Li, S., 2011. Ensemble of feature sets and classification algorithms for sentiment classification. Inf. Sci. 181 (6), 1138–1152.