

Rasim M. Alguliyev¹, Irada Y. Alakbarova²

DOI: 10.25045/jpit.v09.i1.01

Institute of Information Technology ANAS, Baku, Azerbaijan

¹director@iit.science.az, ²airada.09@gmail.com

ABOUT METHODS OF ANALYZING THE TONALITY

The article overviews the interpretations of scientists on the theory of emotion and studies several systems analyzing the emotional tone of various types of information. A general conceptual model for assessing the emotional content of data stored in social media is given. The prospects for using Opinion Mining and Sentiment Analysis methods for analyzing emotions in the text are determined.

Keywords: *emotion, emotionology, social media, text tonality, Opinion Mining, Sentiment Analysis, Emoji language.*

Introduction

In the modern era of information society, new tools are developed for fast data transfer, retrieval from different sources, processing, efficient use and so forth. Large-scale information is stored in many different databases, on individual computers, in local and global networks as texts, photos, audio and video files. Disclosure of opinions and emotions contained in this information may be used in the process of data analysis for different purposes. In many cases, people concern about a question when making any important decision: What do others think of it? With the emergence of the Internet, it became easy to learn people's attitudes to an event or topic. Today, the Internet resources are widely used to study people's attitudes to the events going on around the world, the products offered by companies, and even to the prominent individuals and different political parties and groups [1].

As online information grows and new online resources become available, the number of active users in open encyclopedias, blogs, social networks and other multimedia resources continues to expand. The exchange of information is impossible to imagine without emotion. Any type of contains emotion. Emotion is a psychological process and a complex and complicated form of human perception manifested in biological, physiological changes and behaviors. It is derived from the Latin word "emoveo", which is according to psychologists, in most cases, is understood to be a human feeling experienced in relation to something or someone [1, 2].

Emotion is shown in the form of excitement, rage, fear, love, hate, sympathy, antipathy, consensus, disappointment and so forth. Emotion has been a subject area for research in psychology, philosophy, biology, neurophysiology, psycholinguistics and some other scientific fields for many years. With the development of information and communication technologies (ICT), the theory of emotionology and effective computing methods have gradually become one of the subject fields of informatics. It is possible to identify the society's mood and problems, and predict future changes and even social explosions by analyzing the information gathered in the virtual space, primarily in the social media, [2-5].

Basic elements of emotionology

Charles Darwin expressed interesting ideas on the role of emotions in society, the psychological and physiological state of man, and proposed new approaches to the determination of the emotional state of man in his book *The Expression of the Emotions of Man and Animals* published in 1872. With this approach, he founded the theory of emotions [2].

Emotionology refers to a research in various fields related to the study of human emotions and feelings. The theory of emotionology can also be considered as linguistics of emotion. Emotionology proves that the changes in the human emotions cause the behavioral changes and that this process is primarily dependent on the events and environment surrounding the human beings [2, 3].

Currently, various opinions about the analysis and classification of emotions are available. The main distinguished ideas in Emotionology are related to the basic elements of emotion. Experts state that emotionology has the basic elements and derivative emotions. For example, Professor Keith Oatley from the University of Toronto divides the basic emotions into five basic elements in accordance with the "emotional psychological approach" and calls these elements as the main focus in psychology [3]. Besides the emotions listed below, Keith Oatley distinguishes other derived emotions as follows:

1. Happiness - when the subject of emotion achieves its goal;
2. Sadness - when it is impossible to achieve the goal;
3. Fear - when the "self-preservation" is in danger;
4. Hate - arises against an object that causes the destruction of plans;
5. Aversion - a mixed form of hate and grief.

Professor of the University of California, physiologist Paul Ekman's research on the emotions is also very fascinating. In his article "*Expression and the Nature of Emotion*," the scientist focuses on two main topics: the physiological manifestation of emotions and the universality of emotion. The scientist believes that the physiological manifestations of emotion appear in two forms: mimic and autonomous nervous state. An autonomous nervous state refers to the physiological reaction of body: increased heart rate, blood pressure change, palpitations, and so forth. Whereas an indication of emotion includes the psychological state of man as anger, fear, anxiety, surprise, sadness and joy. Ekman notes that each emotion is unique and characterized only by its mimic and autonomous nervous state, and does not depend on the lifestyle, language, or culture of people. According to Ekman, fear, joy, hate and other emotions generate the same physiological state in all nations [4].

Professor of the University of South Florida, Robert Plutchick describes the emotions in the form of 4 pairs: anger - fear, admiration - aversion, joy - sorrow, hope – surprise [5]. Professor Carroll Ellis Izard from the University of Delaware describes emotions as 10 fundamental signs: anger, aversion, hate, sorrow, fear, sin, curiosity, joy, embarrassment, and surprise. He deems other emotions as the derivative emotions [6].

Most experts note that any emotion has a composition, that is, human beings depend on the environment, and certain emotions are aimed at the abolition of human relationships, whereas other emotions strengthen these relationships [3-6]. However, the composition of emotions in K. Izard's approach is ignored. Izard proposed the *Differential Emotions Theory* in the analysis of emotions. The theory is based on the following five bases [6, 7]:

- fundamental emotions constitute the main motivation system for human existence;
- each fundamental emotion has unique motivation and phenomenological features;
- fundamental emotions, such as joy, sorrow, anger and shame, cause the internal excitement and various forms of expression of these excitements;
- emotions are interrelated - one emotion can activate or weaken another;
- emotional processes lead to and stimulate perceptual, homeostatic and cognitive processes.

Corresponding member of the Russian Academy of Sciences, Professor Sergey Leonidovich Rubinstein states that emotions vary from time to time widely depending on the demands, interests and goals that arise in the process of human activity. Thus, any activity is expressed by certain emotion. Rubinstein expresses only two basic elements of emotion: joy and sorrow. Other emotions are described as derivatives of base elements [8].

Many experts provide emotional assessment as a major subject field of research. For example, Russian psycho-physiologist, biophysicist, academician Pavel Simonov first introduced the information theory of emotion. According to the scientist, although the theory does not demonstrate all the nature of emotion, it is the most precise approach to the exact interpretation of this phenomenon. In his works, Simonov mentions that emotion occurs due to emotional

deficiency or remains of restraint to meet demand. He emphasizes that if the demand and the information for this demand is superior, the value of emotion is even greater [9]. The explanation of this idea is provided as an "emotion formula":

$$E = T (I_t - I_m),$$

here, E - emotion; T - Demand; I_t - information needed to meet the demand; I_m - information provided by the subject when requested. The approach indicates that if the information needed for demand or to meet the demand equals to 0, the emotion must also equal to 0.

Stanley Schachter, a professor of the Columbia University and author of two-factor theory of emotion, reviews emotions as a phenomenon caused by two factors, which are a physiological excitement and a cognitive interpretation of this excitement. In his cognitive-physiological theory of emotions, he shows that if an event or situation causes anxiety in humans, then it is necessary to evaluate the essence of this event. According to Schachter's theory, as a result of emotions, the physiological changes take place in the human body, and he/she assesses the situation based on his/her experience, knowledge, needs and interests. S. Schachter notes that the memory and motivation of the person are fundamental for the emotional process, and the assessment of the situation allows the subject to identify the emotions he/she feels in a certain situation: joy, grief, anger, fear, and so forth [10].

When examining the scientists' approaches to the theory of emotions, it turns out that emotion is studied as a subject field in a number of scientific areas. Emotions can be classified for different characteristics as follows:

- Parameter: positive (joy), negative (anger, aversion), neutral;
- Intensity: feeling, mood, opinion etc.;
- Modality: happiness, surprise, anger, fear etc.;
- Emotive spaces: one-dimensional emotive space and multidimensional emotive space;
- Generation velocity: instantaneous, generating in a certain time;
- Power: high, medium and low to be felt etc.;
- Resistance: momentary, for a certain time interval etc.

Emotion is chiefly expressed through speech, text, mimics, pantomime, gesticulation and vocalization. It is possible to understand how people currently assess the situation and predict their following activity by determining their emotions.

Emotion in visual information

IT specialists have been reviewing the automated identification of emotions through the facial features for many years [11-13]. Many companies offer different software that identifies human emotions according to his/her mood. The automated identification of emotions is now known as "Pattern Recognition" and "Processing of visual information". Advanced technology used for automated identification of emotions is now available and attracts those interested in the methods of processing and identification of objects and their properties. Emotion recognition systems may include *Face Analysis System* software developed by MMER-Systems, *FaceSecurity* and software developed by Cognitec, and *Affective Computing Research Group* developed by US company Affectiva [11].

FaceReader system created by the Dutch company Noldus Information Technology is interesting software product [12]. The system defines different emotions by interpreting facial expressions. There is no need for training and additional adjustments before use. The software product uses the Active Template method, which is a part of the Computer Vision Technology [13]. This method identifies emotions through adding a deformation template to the image. The system also uses Active Appearance model to generate an artificial facial pattern and compare it with the samples in the database, taking into account the facial control points and details. The classification process of the system is implemented through a neural network model, which

provides a high accuracy. The average accuracy of the system of automated identification of emotions is 89%. The system can process statistical image files along with real-time webcam information [13].

Another popular system for emotion's identification based on the visual appearance of a human is the eMotion Software developed by the Dutch "*Visual Recognition*". The software designates six emotional states: anger, sadness, fear, surprise, aversion and happiness. It became popular after identifying emotions of popular "Mona Lisa" for the first time. The image analysis of the system found out that the portrait contains 83% happiness, 9% aversion, 6% fear and 2% anger. The system is open and can be accessed through the company's website. Emotions in the photos can be determined online only by uploading any image [14].

The examination of the above-mentioned systems reveals that the following methods are used for automated identification of emotion out of visual information [13, 14]:

1. Method based on the key points location;
2. Method based on texture information;
3. Combined method.

When it comes to the disadvantages of the system, it is important to note that in these systems, the accuracy of recognition of children under 5 is very low. Besides, the system fails to identify emotions of the facial images with glasses, head veil, and different hairstyle. As some systems do not take into account the change of the image at a certain angle, the resulting accuracy is not always satisfactory. For example, the most popular identification system of emotions based on visual appearance is FaceReader system, evaluates people's emotions differently depending on the color of their skin [13].

Emotion in audio information

Emotionality of speech plays an important role in the communication between people. From this point of view, the automated and objective diagnostics of the emotional state of a man is of great importance. Various commercial and scientific organizations are engaged in research in this area. Depending on the emotions, the quality, power, and other characteristics of a voice change during the speech. The pace of the speech, pauses and emphasis can also determine the emotional state of a person. Starting from the 80s of the last century, the issue of automated identification of emotions by voice has been studied. Vladimir Morozov, Professor of the Russian Academy of Sciences, has analyzed the emotional tones in music, and he coined a term "emotional music sense". In Morozov's theory, emotional music sense is referred to the identification of emotional state of a person in accordance with his/her speech and singing [15].

Most systems for the automated identification of emotion in speech accept two basic elements: neutral and aggressive. For example, *Emo-DB* system, developed by the researchers of the Berlin Technical University, runs based on the records of emotional speeches in German. The speeches stored in the database belong to different actors. An informative signs is assigned for each audio record and this sign is evaluated for the classification effectiveness [16].

The information used for the automated identification of emotion depends on the cognitive schemes of emotion, i.e., it is closely linked to a set of signs representing one or another emotion. The main difficulties of the problem solution are as follows [16]:

1. The lack of parameters precisely identifying the emotion of the voice;
2. The importance of precise study of the relationship between the acoustic characteristics of the voice and the emotional state of the performer;
3. The importance of predefined parameters for speech analysis: frequency of voice, pace of speech (number of words expressed within the time unit), the contour of the main voice tone, the voice power estimation.

The US and some European countries are currently working on the computer detector of emotions based on *Voice-Stress Analysis* (VSA). VSA-based systems are widely used in the

government and law enforcement agencies [17]. For example, *Computerized Voice Stress Analyzer* (CVSA) is one of the most popular and initial software packages designed to identify voice emotion. Although CVSA was launched in 1986, it is still relevant and used for a lie detection in more than 500 US law enforcement agencies [19]. Other VSA-based systems may include VSA-15, VSA Mark I000, TVSA3 and so forth. The systems chiefly use *Bayesian hypothesis method*, Neural networks, system analysis and various classification methods for the identification of the emotional stresses [17, 18].

Emotion on the keyboard

Although modern computers left far behind people in solving various complex mathematical problems, they still fail in identifying emotions. IT professionals have always concerned about the issues related to emotional behavior during the interaction of people with computer. They are conducting various studies in this field to develop an intelligent computer that identifies the emotional state of a person. For example, experts from the Islamic University of Technology in Bangladesh have decided to use the information acquired from the movements of the fingers on the keyboard to solve the problem of automated identification of emotional state through computers.

The study case uses 19 parameters that recognize emotional state: pace of typing, button pressing interval, button pressing power, etc. In the course of the survey, 7 different emotion elements were automatically identified. Emotional state of a person and quality of his/her work were also automatically identified by analyzing his/her behavior behind the keyboard. Studies show that emotions related to joy (87% accuracy) and sadness (81% accuracy) perform more precise results [19].

Emotional tonality in the text

Automated identification of textual emotional tone means that this text is divided into the fragments expressing positive and negative emotions in regard with the object of the emotional valuation, i.e., the object of the tonality. In other words, the emotional value expressed in the text is called the textual tone or sentiment. The emotional expression conveyed at the communicative fragment or lexeme level is called lexical tonality or lexical sentiment. The tone of the text is determined by the lexical tonality of its units and by their combination rules [20, 21]. The identification of the emotional tone helps to detect the emotionality of lexicon and learn the author's attitude towards the object. The tone of the text is assessed taking into account 3 factors [22]:

1. Subject of the tone – a person or author who created the text;
2. Evaluation of the tone: positive, neutral, negative - emotional attitude of the author to the object;
3. Object of the tone – an object or person which the author refers the text to.

Any text or a part of a text may be given as an object of emotional assessment. At the same time, any certain name or common subject in the sentence is also designated as an object of emotional assessment [23]. The name of any product or organization, any service, and type of profession can be an object of the tonality.

In Computer Linguistics, *Content Analysis* methods are widely used for the analysis of emotionality of the text, which in many cases is called textual tone analysis or *Sentiment Analysis* (SA). SA is a content-analysis method designed for automated identification of emotionally-sensitive content of the text and the author's attitude towards the object. The SA method is based on Opinion Mining Technologies [21-23].

Practical application of SA methods on contents is carried out in the following areas:

1. Sociology - gathering all information from social networks to identify opinions and emotions;
2. Political science - gathering information about the political view of the population from the blogs;
3. Marketing - gathering information about advertising and sales in tweets and microblogs;

4. Medical field and psychology - gathering information showing the psychological state of social network users and disturbing the citizens.

Social media, blogs, open encyclopedias, microblogs, and other social media projects developed owing to the widespread use of Web 2.0 technologies show emotions, attitudes and approaches of the people to any event altering their lifestyle, outlook, communication skills and habits. The identification of emotional tone of the content of the social media may be used for the determination of the mood of the society and the attitude of people to the state policy. Thus, determining the attitude of the population to the subject field of the research through SA methods, the social studies may also predict how these relationships will change. For example, the project "*Pulse of the Nation*" has been developed at *Harvard University* (USA). The goal of the project is to identify the mood of the citizens of the country studying and analyzing daily articles and comments on Twitter.

The identification of emotionality using SA methods based on collected texts from the social media is primarily intended for the following purposes:

1. Large and small companies want to be aware of the opinions of people about their products from the social media, particularly from the social networks. For example, 88% of the respondents participated in the survey conducted by the US *Dimensional Research* believe that any positive or negative feedback on any product on the Internet affects their decision to purchase this product [25].
2. Content-analysis methods are widely used to detect information about illegal activities in the field of public security (terrorist threats, information impact on the virtual space, economic dangers, those who share prohibited resources on sites).

Google, Twitter, FaceBook, and other companies with huge databases and analytical systems identify emotions based on web technologies. Google is trying to achieve a smarter search engine using emotions' identification technology. Twitter assesses the mood of users in online blogs, whereas the Tagger service offered to FaceBook users is aimed at automated recognition and registration of the users in the social network, and the identification of their emotions. The use of various icons and graphic images in the products of these companies, generally on all social networks, has become a widespread and crucial phenomenon in determining the author's emotional state. In many cases, an emotional state of the exchange of information can be defined even without reading the text. Some of the most common icons of emotion and social networking are as follows:

- positive icons: “:)””, “=)”, “:))”, “=))”, “-)””, “-))”, “:-)””, “:-))”, “:D” and etc.
- negative icons: “:(””, “=(””, “:((””, “=((””, “-((””, “-((””, “-((””, “:-((””, “:-((” and etc.

Emoji graphics is used to express opinions, emotions, and ideas in the modern web and mobile apps. There is one or more standard codes (*American Standard Code for Information Interchange, ASCII*) in this language for the graphic icon representing any emotion. Emoji does not only convey an emotion. People discuss the news on the social networks, make appointments, express their attitude, greetings and desires through these icons. Since 2009, Emoji characters have been included into Unicode 5.2. This graphic language widely uses emoticons (smileys, animals, plants, etc.). The idea of using emoticons to express emotions in the text was first proposed by the Japanese mobile phone operators [26, 27].

Today, emotions are becoming an integral part of modern global culture. At present, over one thousand emoticons are available and their number is increasing. However, it should be taken into consideration that not all texts contain icons and emoticons, and in some cases misleading icons are also added. Experts have revealed that only 1/4 of the tweets use emoticons. In social networks, only 37.6% of users show emotions through emoticons [28].

Some approaches to the sentiment analysis

The manner of using words in the text determines the author's attitude to the topic. Emotions in the text can be given through lexicon, punctuation, structure of sentences, exclamation mark,

introductory word, application form, and graphic emphasis on a part of the text. Studies show that the tone of the text is determined either by a binary scale or by a multidimensional scale [29, 30]. In the simplest case, text is classified by its tone in two ways - positive and negative emotional assessments. The binary scale uses only two values to solve the problem. This approach is often called a use of single-dimensional emotive space [29]. However, this does not always definitely identify the emotion of the document content. In this case, the use of multi-dimensional scale, i.e. multi-dimensional emotive space is more preferable. This approach involves a broader sentiment classification and the main purpose is to identify negative or positive comments expressed and to classify the contradictions, namely emotional states - calm, nervous, sad, happy, and so forth [30].

The sentiment classification is called emotional text evaluation. Emotional text evaluation is characterized in relation to other objects, while the tone of the lexical units within the text is determined by their relationships. The key approaches to the sentiment classification include the followings:

- Rule-based approach;
- Dictionary-based approaches;
- Machine learning methods.

Rule-based approach. In rule-based approach, classification is specific to each organization and company and is based on a pre-defined rule-based approach. Analysts are not interested in this approach, as the experts implement the tonality of texts basing on rules. The advantage of the approach is its good performance based on a set of rules. The disadvantage is generating a set of rules, which is time-consuming and requires effort. In addition, when the size of the data increases, the difficulties may arise in the course of problem solution [31].

Dictionary-based approach. In this approach, the identification of emotions in the text is based on the words in the dictionaries. The analyzer uses one or two linguistic devices: affective lexicons and templates lexicons. Depending on the emotional values, words are collected in various affective lexicons. Previously developed definite templates are also used to get more precise results. Dictionaries-based approaches can classify the key stages as follows:

- Removing special terms from the text;
- Identifying the tonality;
- Associative analysis of relationships.

The most common and simplest approach in this area is Breen's Approach [32]. To determine the value of emotion in the approach, the number of negative words in the text is subtracted from the number of positive words:

$$E_q = P_i - N_i,$$

where E_q is the value of the emotion, P_i - the number of positive words, and N_i - the number of negative words.

If $E_q > 0$, the tonality of the text is positive. Otherwise, tonality is considered to be negative. If $E_q = 0$, the tonality of the text is neutral. The dictionary of positive and negative English words is available on a special web page [34]. The dictionary of negative and positive words offered on the web site is separately compiled and can be downloaded as a text file on a computer.

Another popular dictionary-based approach is based on ANEW (*Affective Norms for English Words*) dictionary. This approach searches the emotional lexical tonality in the text, according to previously tonal valuation dictionaries. ANEW dictionary algorithm is implemented in the following steps [33-35]:

- Each word in the text is assigned a tonality value from ANEW dictionary;
- The total tonality of the text is calculated;
- If the calculated value is higher than the pre-determined value, it is positive, and if smaller, it is negative.

The key shortcomings in dictionary-based approaches are related to the difficulties involved in the preparation of the dictionary. Thus, in order to classify the document with high accuracy, the terms in the dictionary must have acceptable weight to the subject area. For example, if the

term "light work" is "light" positively, then the same word in the term "light minded" refers to a negative emotion. Obviously, the words may relate to different emotions, depending on the subject area, consequently this approach is ignored.

Machine training. The main goal of machine learning method is to achieve essential and sufficient rules for classifying new information according to the text that constitutes the learning set. Professors from Cornell University *Bo Pang* and *Lillian Lee*, who have contributed to the field of the emotional tonality analysis in textual data, propose a new method and algorithms in the book *Opinion Mining and Sentiment Analysis* published in 2008. They prove that the use of machine learning in the sentiment is more effective and productive than a dictionary-based approach [21]. Opinion Mining is an automated intelligent identification of emotions and moods from textual data. Opinion Mining Algorithms are a part of Text Mining, and are widely used in real-time social media-based assessment (positive, negative, and neutral) of events, products, organizations, individuals and countries [21].

Opinion Mining is based on various scientific theories. These theories may include *Theory of Emotion and Social Life* [36], *Psychological Theory of Emotions* [37], *Lexical Semantic Language Theory* [38], *Intelligent Text Processing* [39], *Automatic Text Analysis* [40], *Artificial Intelligence* [41], *Natural Language Analysis (NLP)* [42] and so forth.

SA methods may include *Dynamics SA*, *Visual SA*, *Large Scale SA*, *Deeper SA*, *Online SA*, *Multilingual SA*, *Semantic Orientation Method*, *Semantic Structure Networks*, and so forth [43-45].

Studies have shown that *Support Vector Machine* [46, 47], *Natural Language Processing (NLP)* [48], *Naive Bayes Classifier* [49], *Fuzzy C-Means Algorithm* [40], *linguistic regression* [50] are widely used for the classification of texts by tonality. The SVM method used to classify text by tonality compares the text with the predefined reference framework by its selected proximity measure and affiliation. This comparative analysis refers the text to negative or positive emotions. The NLP method is universal and widely used in both dictionary-based approach and machine learning methods. NLP is a computerized analysis and synthesis of natural language being implemented with the joint use of artificial intelligence and mathematical linguistics. In other words, NLP teaches the computer to read, listen, and understand text as a human being. The linguistic structure of the natural language is complex and its parameters are constantly changing. In the native language, the context of the text defines the meaning of the expression, which does not always allow it to be interpreted in the same sense. NLP is based on lexicon of the text and used to define syntax structures in the semantic relationships. Regional dialects, non-standard sentence structuring, and morphological expressions complicate the identification of emotions in the text. Regardless of these difficulties, NLP-based software is widely used in the recognition and interpretation of the language. NLP performs high accuracy in the analysis of large-scale data. According to its abovementioned opportunities, NLP is often used for structuring data prior to SA [46-48]. Naive Bayes model calculates a posteriori probability based on the position of the words in the text. Therefore, it predicts the probability that the given set of properties belongs to a particular class. The classifier works on Bayesian theorem [49]:

$$P(i, x) = \frac{P(x, i) \cdot P(i)}{P(x)},$$

where $P(i, x)$ - is the probability of the properties x to belong to the class i . $i = 0..1$. $P(x, i)$ - is the probability of the properties x to present in all texts belonging to the class i . $P(i)$ - is the probability of the new object to belong to the class i . $P(x)$ - is the probability of the properties x to belong to all classes.

The classification using Naive Bayes methods (e.g., *Multinomial model*, *Bernoulli model*, *Begging algorithm*) performs accuracy within 85.5% and 87.7%, which is related to all machine learning algorithms. The disadvantage of this approach depends on the quantity and quality of learning algorithms used. In addition, the sequence of words in the text is not taken into account in this approach, and the objects and subjects of the tonality are impossible to be identified [51].

Common Sentiment analysis

According to the study, machine learning methods are based on three approaches for the sentiment analysis: statistical approach, linguistic approach and blended method. The key features of these approaches are as follows:

1. One of the strategic issues in the statistical approach is to classify the texts by tonality class. In the course of the problem solution, the text collections are pre-evaluated by the tonality and machine training is realized.
2. The objective of the content analysis in the linguistic approach is to automatically find the dictionary containing texts' emotions and to achieve the author's opinion regarding the object in the text. This approach searches an emotive lexical tonality in the text based on previously structured tonality dictionaries (sample lists). Based on the sum of the emotive lexicon, the text is evaluated for the scale showing the number of positive and negative lexicon.
3. Blended method involves the combination of the first and second approaches.

The main stages of the sentiment analysis include language identification, design of signs dictionary, sentiment classification and linguistic analysis (Fig. 1). In the process of sentiment analysis, the information containing emotion should be identified and structured.

The text written in natural language is non-structured information. Therefore, the text is structured prior to the analysis. The common algorithm used for the sentiment analysis can be explained as follows:

- authentication of the language of the text;
- each document is given as a vector of signs;
- type of tonality is identified for each document;
- method is specified for the classifier training;
- achieved model is used to identify the tonality of the new documents.

If the algorithm uses machine learning methods for the sentiment analysis, the problem should be solved in the following stages:

1. Selecting a method for the evaluation of the effectiveness of the algorithm;
2. Selecting signs used for the classification;
3. Identifying the classification algorithm;
4. Selecting samples appropriate to training;
5. Trial.

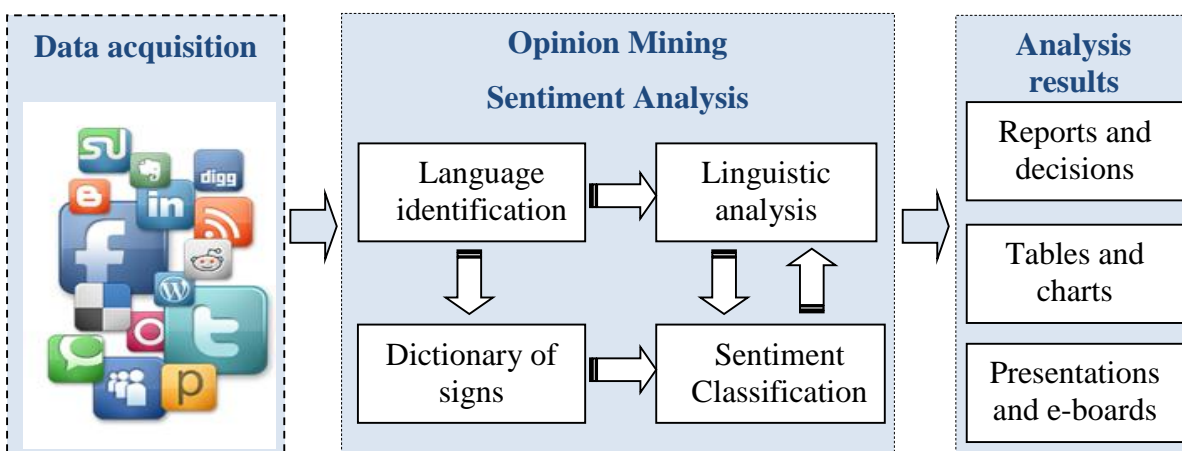


Figure 1. Common model for sentiment analysis

Although modern computers successfully perform speech recognition, semantic analysis of the text, the problems related to the automated identification of emotions in the text and "subtext phrases" reading still remain. A person intuitively senses the emotion in the text, while computer needs new methods and algorithms to identify emotion. SA methods can contribute to the development of rules for natural languages and solutions for computers to understand humans.

Conclusion

Solving the problems of emotional analysis requires the combination of computer science with psychology, sociology, psychophysiology and other sciences. The studies have shown that the interest in the automated sentiment identification is increasing along with the increasing capacities of the social media. Today, emotionology is attracting the experts from various areas as a separate field of science, and the sentiment analysis in pictures, sound and textual data is used for a variety of purposes. Available systems used to identify the emotional state of a person by different signs vary according to the list of identified emotions, the type of data used for the analysis, the method and classification algorithms. This difference complicates the comparison and evaluation of the obtained results.

Depending on the nature of the sentiment analysis, although different approaches are used, most studies mostly benefit from NLP, SVM and Naive Bayes models. The use of these models provides higher accuracy in the analysis process. The capabilities of sentiment analysis methods can play an important role in ensuring the security of the state information space, high performance in advertising and marketing of companies, and, finally, in improving the quality of machine translation.

References

1. Martinez A., Estrada H., Molina A., Mejia M., Perez J. Emotion-Bracelet: A Web Service for Expressing Emotions through an Electronic Interface // *Sensors* (online), 2016, 16(12). <http://mdpi.com/1424-8220/16/12/1980/>
2. Darwin C.R. The expression of the emotions in man and animals. London: John Murray, 1872, 472 p.
3. Oatley K. *Best Laid Schemes: the Psychology of Emotions*. Cambridge University Press, 1992, 445 p.
4. Ekman P. What Scientists Who Study Emotion Agree About // *Perspectives on Psychological Science*, 2016, vol.11. no.1, pp.31–34.
5. Plutchik R. *The Emotions*. University Press of America, 1991, 216 p.
6. Izard C.E. *The face of emotion*. Appleton-Century-Crofts, 1971, 468 p.
7. Izard C.E. Differential emotions theory and the facial feedback hypothesis of emotion activation: Comments on Tourangeau and Ellsworth's "The role of facial response in the experience of emotion // *Personality and Social Psychology*, 1981, vol.40, no.2, pp.350–354.
8. Rubinstein S.L. *The basics of general psychology*. Publisher: Piter, 2002, 720 pp.
9. Simonov P.V. *Emotional brain*. M.:Nauka, 1981, 166 c.
10. Schachter S., Singer J. Cognitive. Social and Physiological Determinants of Emotional State. *Psychological Review*, 1962, vol.69, no.5, pp.379–399.
11. <http://affect.media.mit.edu/>
12. www.noldus.com/facereader/facereader-online/
13. Den Uyl M., Van Kuilenberg H. The facereader: online facial expression recognition / *Proceedings of the 5th International Conference on Methods and Techniques in Behavioral Research*, 30 August–September 2005, Netherlands, Noldus Information Technology, pp.589–590.
14. Software sees Mona Lisa's emotions. <https://geek.com/news/software-sees-mona-lisas-emotions-558144/>

15. Morozov V.P. Emotional Hearing and Music Endowments / Proceedings of the International Scientific and Practical Conference "Development of the Scientific Heritage of Teplov B.M. in Domestic and World Science", Moscow: P. RAO, 2006. pp.198–203.
16. Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W., Weiss B. A. Database of German Emotional Speech / Proceedings of the Interspeech, 2005, pp.1517–1520.
17. Kelly R.D. Voice Stress Analysis: Only 15 Percent of Lies About Drug Use Detected in Field Test // National Institute of Justice Journal, 2008, no.259, pp.8–12.
18. Haddad D., Walter S., Ratley R., Smith M. Investigation and Evaluation of Voice Stress Analysis Technology, The Final Report of Department of Justice, 2002, 120 p.
19. Nahin N.H., Mohammad A.J., Mahmud H., Hasan K. Identifying emotion by keystroke dynamics and text pattern analysis // Behaviour and Information Technology, 2014, vol.33, no.9, pp.987–996.
20. Bing L. Sentiment Analysis and Opinion Mining. 2012, Morgan & Claypool Publishers, 167 p.
21. Pang B., Lee L. Opinion Mining and Sentiment Analysis // Foundations and Trends in Information Retrieval, 2008, vol.2, no.1–2, 135 p.
22. Feldman R., Sanger J. The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. 2006, Cambridge University Press, 410 p.
23. Wong M. Emotion Assessment in Evaluation of Affective Interfaces, 2006, www.cgl.uwaterloo.ca/wmcowan/research/essays/maria.pdf
24. Pulse of the Nation: U.S. Mood throughout the Day Inferred from Twitter, 2010. <http://infosthetics.com/archives/2010/07/>
25. Customer service and business results: a survey of customer service from mid-size companies, 2013. www.d16cvnquvjw7pr.cloudfront.net/
26. Piercy J. Symbols: A Universal Language, 2013, London:Michael O'Mara, 224 p.
27. Encyclopedia of Emoji: the meaning of all 1427 emoticons iOS 10.3. www.iphones.ru/iNotes/all-about-emoji-ios-10
28. Ljubešić N., Fiser D. A Global Analysis of Emoji Usage / Proceedings of the 10th Web as Corpus Workshop and the EmpiriST Shared Task, August 7-12, 2016, Berlin, pp.82–89.
29. Turney P. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews / Proceedings of the Association for Computational Linguistics, 2002, pp.417–424.
30. Chetviorkin I.I. Testing the sentiment classification approach in various domains / Proceedings of the International Conference "Dialog 2012", vol.2, no.11, pp.15–27.
31. Kan D. Rule-based approach to sentiment analysis / Sentiment Analysis Track at ROMIP, 2011. <http://dialog-21.ru/digests/dialog2012/materials/pdf/Kan.pdf>
32. www.cs.uic.edu/~liub/FBS/sentiment-analysis.html
33. Paroshina I. Natural language processing, difficulties of understanding and social networks // Computer-Online, 2014.
34. Montefinese M., Ambrosini E., Fairfield B., Mammarella N. The adaptation of the Affective Norms for English Words (ANEW) for Italian // Behavior Research Methods, 2014, vol.46, no.3, pp.887–903.
35. Bradley M.M., Lang P.J. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999, pp.1–45.
36. Warren D. TenHouten, A General Theory of Emotions and Social Life, 2006, Routledge, 336 p.
37. Sobkowicz P., Kaschesky M., Bouchard G. Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web // Government Information Quarterly, 2012, vol.29, no.4, pp.470–479.
38. Geeraerts D. Theories of Lexical Semantics, Oxford University Press, 2009, 384 p.

39. Alguliyev R.M., Aliguliyev R.M., Alakbarova I.Y. Extraction of hidden social networks from wiki-environment involved in information conflict // *International Journal of Intelligent Systems and Applications (IJISA)*, 2016, vol.8, no.2, pp.20–27.
40. Alguliev R.M., Aliguliyev R.M. Automatic text documents summarization through sentences clustering // *Automation and Information Sciences*, 2008, vol.40, no.9, pp.53–63.
41. Premchandani S., Pise M., Wankhede A. Artificial Intelligence and Opinion Mining // *Journal of Artificial Intelligence*, 2012, vol.3, no.2, pp.102–105.
42. Godbole N., Srinivasaiah M., Skiena S. Large-Scale Sentiment Analysis for News and Blogs / *Proceedings of the International Conference on Weblogs and Social Media*, March 26–28, 2007, pp.39–41.
43. Kanayama H, Hideo W., Nasukawa T. Deeper sentiment analysis using machine translation technology / *Proceedings of the 20th International conference on Computational Linguistics*, Tokyo, Japan, 2004. <http://aclweb.org/anthology/C04-1071>
44. Moraes R., Valiati J.F., Wilson P. Gavião N. Document-level sentiment classification: an empirical comparison between SVM and ANN // *Expert Systems with Application*, 2013, vol.40, issue 2, pp.621–633.
45. Hye-Jin Min, Jong C. Park. Identifying helpful reviews based on customer’s mentions about experiences // *Expert Systems with Application*, 2012, vol.39, issue15, pp.11830–11838.
46. Thelwall M., Buckley K., Paltoglou G., Cai D., Kappas A. Sentiment strength detection in short informal text // *Journal of the American Society for Information Science and Technology*, 2010, vol.61, issue 12, pp.2544–2558.
47. Yusupova N.I., Bogdanova D.R., Boyko M.V. Algorithmic support and software for the Sentiment analysis of text messages using machine learning // *Mathematical modeling, numerical methods and program complexes*, 2012, No6 (51), pp.91–99.
48. Poshevkin R.V., Bessmertny I.A. Application of sentiment analysis of texts for assessing public opinion // *Scientific and Technical News of Information Technologies, Mechanics and Optics*, 2015, Volume 15, No. 1, pp.169–171.
49. Kang H., Joon Yoo S., Han D. Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews // *Expert Systems with Application*, 2012, vol.39, issue 5, pp.6000–6010.
50. Wu Y., Kita K., Matsumoto K., Kang X. A Joint Prediction Model for Multiple Emotions Analysis in Sentences // *Computational Linguistics and Intelligent Text Processing*, 2013, vol.7817, pp.149–160.
51. Losada D.E., Azzopardi L. Assessing multivariate Bernoulli models for information retrieval // *ACM Transactions on Information Systems*, 2016, vol.26, issue 3, pp.53–65.