Computing and Informatics, Vol. 37, 2018, 737–758, doi: 10.4149/cai_2018_3_737

SENTIMENT AND AUTHORITY ANALYSIS IN CONVERSATIONAL CONTENT

Kristína Machová, Martin Mikula
Martina Szabóová, Marián Mach

Department of Cybernetics and Artificial Intelligence
FEI, Technical University of Košice
Letná 9, 042 00 Košice, Slovakia
e-mail: {kristina.machova, martin.mikula, martina.tarhanicova, marian.mach}@tuke.sk

Abstract. This paper deals with mining conversational content from the social media. It focused on two issues: opinion and emotion classification and identification of authoritative reviewers. The paper also describes applications representing the results obtained in the given areas. Authority identification can be used by organizations to search for experts in their specific areas to employ them. The opinion and emotion analysis can be useful for providing decision-making support.

Keywords: Opinion analysis, emotion analysis, conversational content, conversation structure, authority identification

Mathematics Subject Classification 2010: 68U15

1 INTRODUCTION

The social web is a phenomenon of the present day where users create a huge amount of information called conversational content. It increases the number of interactions among web users affecting our decisions in real life situations. Therefore, application of both sentiment analysis and authority mining can be useful in the process of decision making. The input of this application can be a hypertext reference to the related web discussion or a key phrase from the area of interest being discussed. The output is a summarized opinion [2] (clearly positive, neutral or negative) represented in this discussion as well as information about authoritative discussants.

Authoritative users can influence us more than the others, therefore, having knowledge about them can improve the precision of sentiment analysis. In this paper, an authoritative user or authoritative reviewer will be referred to as "authority". The problem of *"authority identification"* can be solved through finding a function for estimation of the value of authority, which represents the measure of authoritativeness of a given discussant, and defining a threshold on the value of the variable authority. All discussants with the value of authority equal to or higher than the defined threshold will be *"identified as authorities"* of a given web discussion.

In this paper, we focus on two main issues. The first is sentiment analysis. The concept of the sentiment is represented by the opinion or some kind of emotion. Based on the expression of the sentiment we divided our work into opinion classification and emotion classification. We used two different dictionaries: one to classify opinions into positive and negative classes and the other to classify emotions into six basic emotions.

The second issue is based on authority identification. We define authority as a person, who understands the topic, writes comprehensive answers, has many good reactions to his/her comments and has a good reputation in the community. We analyse these features as well as the state threshold of the authority value and label all authors having a higher value than the authorities in the discussion.

Thus, the main contributions presented in this paper are new approaches to classifying opinions and emotions, and a completely original approach to estimation and identification of the authority in web discussions.

2 STATE OF THE ART

The tasks of both sentiment analysis and opinions and emotions analysis can be performed using two main approaches [22]:

- *Lexicon based approach*, which calculates the polarity of a comment from polarities of words or phrases in the text of the comment. According to [16] the lexicon-based approach uses the lexicons which can be created manually, semi-automatically or automatically by using another dictionary or corpus.
- *Classification based approach*, which builds classifiers from labelled examples of comments. This approach can use statistical or machine learning methods.

The *lexicon-based approach* takes into consideration only the words which can express the sentiment the best way and are stored in a classification dictionary. Based on the polarity of each word in the dictionary, the polarity of the whole comment and subsequently of the whole web discussion can be determined. The polarity of a word may also depend on the context. Consideration of the context requires using more complicated techniques. The simplest dictionaries enable to perform binary classification into positive or negative sentiment. More complex dictionaries can also determine the strength of this polarity. Such dictionaries can be created for a particular domain or application. Some well-known lexicons are, for example, WordNet[1], WordNet-Affect [26], SenticNet [6], SentiWordNet [1, 9], etc. An approach most similar to our solution of the sentiment analysis problem is presented in [27]. It uses a dictionary of words annotated with their polarity. This work splits this dictionary into four sub-dictionaries according to word classes (adjectives, nouns, verbs and adverbs), these dictionaries are checked for consistency and reliability. In this approach, more words in the dictionary can result in an increase of noise levels and subsequently decrease in the precision.

In the *classification based approach*, many of the well-known machine learning methods can be used. The machine learning methods estimate user's opinions or emotions on the basis of a set of training examples. These training examples are represented by annotations of comments to a web forum. The most common machine learning methods are, for example, Naïve Bayes classifier or Support Vector Machines (SVM), as well as statistical methods such as Maximal Entropy. All these algorithms were used in the work [10]. The author focused on the sentiment classification from the micro-blog service Twitter. When machine learning is used for sentiment analysis it is important to select the best features. Different kinds of features, such as pointwise mutual information, information gain, chi-quadrat and term frequency were tested in works [28, 32]. Another approach presented in [29] is focused on the SentiStrength detection algorithm, which solves some problems connected with sentiment analysis, for example: generation of the sentiment strength list, optimization of the sentiment word strengths, allocation of missing words, spelling correction, creation of the booster word list, negating word list and emoticon list, processing of the repeated letters and ignoring negative emotion in questions. Their approach is based on using machine learning techniques (Logistic Regression, Support Vector Machine, J48 Classification Tree, AdaBoost, Decision Table, Multilayer Perceptron and Naïve Bayes).

An important drawback of the machine learning approach is its dependency on a huge amount of annotated training examples. An annotation tool with some level of automation had to be utilized. The source [25] describes some annotation tools, for example, GATE (General Architecture for Text Engineering), SemTag (Semantic Annotation Tool), the annotation platform KIM (Knowledge and Information Management) and Luvak, or a more general semi-automatic annotation tool built on the Eclipse platform.

The second problem, which we tried to solve in relation to sentiment analysis, was the authority identification in some web discussions. The most known approaches determine the authority degree only from the conversation structure [4, 7, 31]. Our approach is not based on NLP (Natural Language Processing) or on information retrieval algorithm but on generating the estimation function for labeling the degree of authority prediction (see Section 4).

---

[1] `http://wordnet.princeton.edu`

3 SENTIMENT ANALYSIS

3.1 Opinion Classification

Our opinion analysis approach has focused on the classification of web users' opinion. It summarizes positive, neutral or negative polarity across the discussion. This classification is provided in several steps. At first, the polarity of particular words is identified. Next, the polarity of particular lexical units is distinguished and consequently the polarity of the whole comment is stated. The final step is classification of the whole discussion to positive, neutral or negative polarity. Other web discussions concerning the discussed topic can be also processed. It means that the resulting opinion can be compiled from more web sources.

The basic problems can be simply solved using classification dictionaries. These dictionaries focus on words, which express sentiment very well – mainly adjectives (e.g. "extraordinary") [3] and adverbs (e.g. "awfully") [17]. On the other hand, some other words must be also taken into account in order to achieve the satisfactory precision, for example nouns (e.g. "crash") or verbs (e.g. "damage") [27]. All of these words are identified in the text; they usually enter the dictionaries with their degree of polarity.

For the purpose of deeper processing of a text, *shifters* can be used, which are words that can change the polarity of a word. The positive influence of the shifters processing was studied in work [14]. Shifters can be divided into two groups:

- negation,
- intensification.

There are two basic approaches to *negation processing*: a switch negation and a shift negation. The *switch negation* is simply reversion of the polarity of a lexical unit. In this case, the reversion is changing the sign of a number, which represents the polarity degree (from minus to plus and vice versa). There are many words related to negation such as *not*, *any*, *never*, *nothing*. They are usually located next to the related word. Other negations as *without, don't, lack* etc., which can be situated at a significant distance from the lexical item should also be considered. These negations can be hardly processed by the switch negation. In some cases, the switch negation may not be sufficiently precise because the negation of a strong positive word is rarely a strong negative word and vice versa. More often the negation of a strong positive word is a slightly negative word. The *shift negation*, instead of changing the sign, shifts the polarity degree towards the opposite polarity by a fixed value (e.g. value 4 in the implementation [27]). For example: "She's not terrific ($5 - 4 = 1$) but not terrible ($-5 + 4 = -1$) either."

The *intensification processing* assumes the existence of a dictionary of intensifiers. An intensifier is a word, which can increase (or decrease) the intensity of polarity. According to [27], intensifiers can be of two categories: *amplifiers* (e.g. *very*) increase the semantic intensity of a neighbouring lexical unit, whereas *downtowners* (e.g. *slightly*) decrease it. All intensifiers are stored in the dictionary together with a sign and a value. The value represents the percentage of the change in the polarity intensity and the sign represents the type of this change ("plus" represents the increase in the polarity value by an amplifier and "minus" represents the decrease in the polarity value by a downtowner).

3.1.1 N-Grams Approach to the Opinion Classification

An n-gram can be defined as a series of items from a sequence. From the semantic point of view, it can be a sequence of characters or words. In practice, n-gram as a sequence of words is the most common. Our approach uses n-grams for splitting web discussion comments into lexical units and that is a dictionary-based approach. This application works with Slovak texts. The dictionary consists of two parts. The first part contains adjectives, nouns and verbs. The second part contains adverbs and negations. The first part of the dictionary is used to solve the basic problems of the opinion classification. The second part of the dictionary is used in negations and intensification processing, because only adverbs can increase ("*surprisingly nice*") or decrease ("*extremely low-class*") the intensity of a related word polarity. The first (basic) dictionary also contains some emoticons, which naturally can express emotions and opinions very well. In the case, when the analysed text is less clear, the emoticons can increase the precision of the classification. All words and emoticons from the first dictionary are quantified to the polarity degree from the interval $\langle -3, 3 \rangle$ (Table 1). Intensifiers (adverbs) in the second part of the dictionary are assigned values from the interval $\langle -0.5, 1 \rangle$ and negations are represented by the value -2 (Table 2).

Polarity Degree	Words and Emoticons
3	:D, godlike, extraordinary
2	:), super, excellent
1	nice, functional, OK
−1	unpleasant, weak
−2	:(, shocking, miserable
−3	:((, fatal, catastrophic

Table 1. Polarity degrees of example words and emoticons (the first part of the dictionary)

Polarity Degree	Words
1	very, totally, extraordinarily
0.5	suitably, really, actually
−0.5	little, overly, unnecessarily
−2	Negations: no, not, don't

Table 2. Polarity degrees of negations and intensifications (second part of the dictionary)

All the words from the analysed comments are compared with all words stored in the first and in the second part of the dictionary. In case that words match with the first dictionary, the values of all matching words are summed up. The resulting sum represents the solution of the basic problems of opinion classification (the first sum in the formula (1)). The second sum takes into account negations and intensifications of the related words incorporated in the first sum. This is the reason why multiplication (not an aggregation) was proposed to be used between the first and the second sum in the formula (1). The second sum aggregates the values obtained from the second dictionary for intensifiers and negations. But if comments do not contain any negation or any intensification, the second sum will be zero, and consequently the resulting polarity of the comment will be zero. Thus, value "1" was added to the second sum as a neutral value. This idea is represented by the formula (1):

$$P = \sum v(w_i^1) \left[1 + \sum v(w_J^2)\right] \tag{1}$$

where:

- P is the polarity degree of analysed text,
- $v(w_i^1)$ is the value of word w_i from the text found in the first dictionary,
- $v(w_i^2)$ is the value of word w_J from the text found in the second dictionary,
- $\sum v(w_i^1)$ is the first sum from the first dictionary (solution of the basic problem),
- $\sum v(w_J^2)$ is the second sum from the second dictionary (solution of the negation and intensification).

To illustrate the topic, we give a few examples:

- A sentence containing only positive and negative words without negations and intensifications can be processed using only the first dictionary
  - "*The mouse is nice but the processing is miserable and globally it is unsuccessful.*" is processed in the following way. Only three words from the first dictionary were found in the sentence with their degree of polarity in the parentheses:

    $$\textit{nice}\ (+1) + \textit{miserable}\ (-3) + \textit{unsuccessful}\ (-1).$$

    There are polarity degrees were the sum and the final polarity of the sentence equals $P = -3$.

  A sentence containing negation

  - "*It is not a good solution.*" is processed in the following way. Only the word *good* (+1) was found in the first dictionary, so the result of the first sum is "1". The sentence contains also negation *not* (−2) from the second dictionary, so the result of the second sum is "−2". According to formula (1) the final polarity is: $P = 1 * [1 + (-2)] = -1$.

A sentence containing intensification

- "*Globally, the processing is very polite.*" is processed in the following way. Only the word *polite* (+1) was found in the first dictionary, so the result of the first sum is "1". The sentence contains also intensification *very* (+1) from the second dictionary, so the result of the second sum is "1". According to formula (1) the final polarity is: $P = 1 * [1 + 1)] = +2$.

More information on this approach can be found in [20]. In our approach, the dictionary can be created directly from web discussions. It increases the precision of the opinion classification in the given domain. Our approach also uses a different method for the processing of negations and intensifications based on the formula (1). Our approach does not need an intensifier (negation) to be located in the neighbourhood of the related word. They can take any position around the lexical unit. This length is limited by the value "4" in 4-gram approach. There are two possibilities for the location of the intensifier and negation. They can be located before and/or after the related word.

The value $n = 4$ in the n-gram was determined experimentally. Experiments with the value n from 2 to 4 were performed to find the best value of n. The ideal value had to be sufficiently big to avoid isolation of the processed word but not too big to cover the whole sentence. The experiments showed that $n = 4$ was sufficient. Thus, the greater value, for example $n = 5$, would only make the processing more complex. When we compared 4-grams with other n-grams, 2-grams and 3-grams did not bring any benefit, because 4-grams could also cover phrases of two or three words. The work [15] also uses 4-grams in the sentiment analysis engine called Umigon for the sentiment analysis of tweets. The texts of tweets are decomposed into a list of 4-grams, and they are compared with terms in lexicons. Each term searched in the lexicon is processed using heuristics and decision rules for 4-gram polarity determination.

The n-grams approach applied to the opinion classification was tested on a set of discussion comments from the portal `http://www.mojandroid.sk` (discussion thread related to reviews of the mobile telephones HTC One X and HCT One S) and `http://www.pocitace.sme.sk` (discussion thread related to reviews of two products Asus Transformer Prime TF201 and Asus Transformer Pad TF300T). 200 comments including 100 positive and 100 negative were used in the experiments. Thus, objects were equally divided into the two classes. The evaluation of n-gram implementation was based on the pair of the well known metrics – precision and recall. These measures depended on the numbers of true positive, false positive and false negative classifications of our implementation in comparison with the opinion of a human expert on real positivity or negativity of the evaluated cases. The resulting precision and recall of our n-gram implementation is given in Table 4. The achieved precision and recall, mainly recall of negative comments, was low. These results were influenced by comments containing irony or the polarity hidden in the context. On the other hand, this implementation had quite a good precision in the processing of positive comments.

3.1.2 Opinion Classification Based on a Special Lexicon

Because of the unsatisfactory results of the implementation based on n-grams, we improved our dictionary-based approach. Our aim was to increase the efficiency of this approach and to extend it by a special lexicon[2] for the Slovak language. The lexicon with Slovak words was created by the translation of its available version in English used in work [13]. We added synonyms and negations of words from the Slovak thesaurus into this lexicon. The new lexicon contains 1 430 words (598 positive words, 772 negative words, 41 intensifiers and 19 negations). As in the previous case, our lexicon contains the following word classes: adjectives, adverbs, nouns and verbs. Each word has two attributes. The first attribute is a degree of polarity within the range from -3 to 3, where -3 is the most negative polarity and 3 is the most positive polarity. The second attribute denotes the type of a given word from four possibilities:

- p – positive word,
- n – negative word,
- i – intensifier,
- o – opposite/negation.

Negation words are assigned the value -1. Intensifiers have values of polarity degree in the range $\langle 1.0, 2.0 \rangle$. Examples of words and their values of the polarity degree and the type of word are illustrated in Table 3.

Polarity Degree	Type of Word	Word
3	p	extra, genius, super, brilliant
2	p	better, advanced, success
1	p	good, ok, strong, smart
-1	n	bad, weak, boring
-2	n	dangerous, hostile, stupid
-3	n	terrible, waste, worst
-1	o	no, not, never, haven't
1.25	i	really, rather
1.5	i	middle, pretty
1.75	i	too, complete
2	i	very, total, absolute

Table 3. Examples of words in the dictionary with the polarity degree and the type of word

[2] `http://klanaz.studenthosting.sk/sa.html`

The analysed comment text was split into individual sentences and diacritic marks were removed from the text. All words of the text were converted into nominative of the plural using the modified version of the Lancaster stemming algorithm[3]. The converted words were compared with the words in the dictionary. The sentence was processed word by word. Each positive or negative word was multiplied by intensifications and negations. Then, all polarities of positive and negative words were summed up according to the formula (2). In the case, when a comment did not contain any intensification or negation, the values of intensification and negation were set to 1.

$$P = \sum \left[w_w \prod w_i \prod w_n \right] \tag{2}$$

where:

- P – polarity degree of the analysed sentence,
- w_w – value of positive or negative word,
- w_i – value of intensifier,
- w_n – value of negation,
- $\prod w_i$ – multiplication of all intensifiers,
- $\prod w_n$ – multiplication of all negations.

The resulting polarity of the sentence was adjusted using logarithmic function in the formula (3) to avoid processing huge numbers – values of P. When the value of the sentence polarity before adjusting was 0, the adaptation was not used. When the value of the sentence polarity was lower than 0, the absolute value was adjusted and multiplied by -1 to maintain the negative polarity of the comment.

A dataset did not contain comments with just facts. Such comments were not collected and added to the dataset during its creation. Sentences which contained no sentiment words were evaluated as mistakes. For example, the sentence which was labelled as positive but contained no sentiment word was evaluated in the same way as a negative one. The resulting sentiment value was 0 in these cases.

$$P_l = 1 + \log_{10} |P| \tag{3}$$

where

- P_l – the new value of sentence polarity,
- P – the polarity value of sentence obtained by the formula (2).

The following examples illustrate the above computation of the polarity degree value. A sentence containing only positive and negative words without intensifications and negations

[3] `https://goo.gl/STmHi0`

- *"The mouse is nice but the processing is miserable and globally it is unsuccessful."* is processed in following way. Three words from the dictionary were found in the sentence with their degree of polarity in parentheses:

$$\textit{nice}\ (+1) + \textit{miserable}\ (-2) + \textit{unsuccessful}\ (-1).$$

  These degrees of polarity were the sum and the final polarity of the sentence is $P = -2$. After adjustment according to the formula (3) $P_l = -1.3$.

A sentence with intensification (very) and negation

- *"It is not a very good solution."* is processed in following way. Three words from the dictionary were found in the sentence with their degree of polarity. The first word was *good* (+1), so $w_w = 1$. The sentence also contains intensification *very* (2), so $w_i = 2$ and negation *not* (−1), so $w_n = -1$. According to the formula (2) the final polarity is $P = -1 * 2 * 1 = -2$. After adjustment according to the formula (3) $P_l = -1.3$.

This approach is a little different in comparison with the previous n-gram based application. It uses only one dictionary for all types of words and this dictionary contains mainly adjectives and nouns in nominative plural (the dictionary also contains verbs in the form as they appear in the original text without any modification). In comparison with the n-grams implementation, this approach uses a different method to process intensification and negation. Moreover, this approach can process multiple intensifications (e.g. *very very good*).

Our approach was tested on two datasets. The first dataset was the same as the dataset for n-grams. It allowed to compare our method with the previous one based on n-grams. A new dataset was created to test our approach. This second dataset contained collected comments from different areas (e.g. films, electronics, politics, etc.) Each comment was labelled by a human annotator. The neutral ones and facts were removed from this dataset. The second experiment was performed on the set of 5 242 comments and 182 645 words, where 2 573 comments were positive and 2 669 negative so that the objects were equally divided into available classes. The dataset is available on the website[4].

The results of testing given in Table 4 show that quite a good recall was achieved for positive comments. We compared our approach with the n-grams approach in the first test. The precision obtained for positive comments was lower, but the results for other indicators were better. The reason could be the dictionary, which contains more words and more comprehensive computation of the polarity degree. This approach is more similar to human understanding of the text. In the second experiment, a good recall was achieved for positive comments, but it was low for negative comments. The results could be influenced by irony, sarcasm or description of opinions without polarity words. The number of missclassified comments was also increased by comments, which did not contain polarity words. These comments

[4] `http://klanaz.studenthosting.sk/dataset.txt`

were evaluated as mistakes and added to the same class as incorrectly evaluated comments.

Experiment	Precision (pos)	Precision (neg)	Overall Precision	Recall (pos)	Recall (neg)	Overall Recall
NGR 1	0.830	0.570	0.700	0.652	0.214	0.433
SD 1	0.654	0.727	0.691	0.850	0.471	0.661
SD 2	0.561	0.675	0.618	0.802	0.396	0.599

Table 4. The precision and recall of tests achieved by the n-grams approach (NGR) and the approach based on the special dictionary (SD)

The results of our experiments with opinion classification showed that using our first approach (Section 3.1.1) the best value of precision obtained was 0.830 and the best value of recall 0.652. Our experiments with modified approach (Section 3.1.2) achieved better results in recall (0.850) than in precision (0.727). In comparison with another method of opinion classification using an approach that is very close to the processing negation and intensifiers [27], our results were slightly worse. The results of F1 (combining precision and recall using equal weights) presented in the study by Taboada, were in the range from 0.58 to 0.89 according to types of reviews.

3.2 Emotion Classification

Emotion analysis is a similar problem as the opinion classification. Both of them (opinion and emotion classification) represent the sentiment analysis. Accordingly, in emotion classification words from dictionaries representing emotions (joy, anger, disappointment, ...) are searched for in the input text. The type of emotion expressed by these words denotes the kind of emotion presented in the given text.

According to [11], three major directions in emotion computing can be recognized: categorical/discrete, dimensional and appraisal based approaches. Despite the existence of other models, the categorical and dimensional approaches are the most commonly used models for automatic analysis and prediction of the emotion in the continuous input.

The Categorical Approach claims there is a small number of basic emotions that are hard-wired in our brain, and recognized across the world. Emotional states are classified by a single category. However, a couple of researchers proved that people show non-basic, subtle and rather complex emotional states that could be impossible to handle, such as embarrassment or depression [12].

The Dimensional Approach is based on Wundt's [30] proposal that feelings (which he distinguishes from emotions) can be described as pleasantness – unpleasantness, excitement – inhibition and tension – relaxation, as well as Osgood's work [21] on the dimensions of affective meaning (arousal, valence, and potency). Most recent models concentrate on only two dimensions, valence and arousal. Valence (pleasure/displeasure) depicts how positive or negative emotions can be. Arousal (activation/deactivation) depicts how exciting or apathetic emotions can be.

For our research purposes we decided to choose the categorical approach, i.e. Ekman's [8] six basic emotions: happiness (positive), sadness (negative), surprise (positive/negative), fear (negative), disgust (negative), and anger (negative).

There are four major approaches to emotion classification in the text: dictionary-based methods, machine learning methods, knowledge-based methods and hybrid methods. Our main interest is the *dictionary-based methods*. As our target language is Slovak, and we are not aware of any Slovak lexicon for emotional words, we created one. Every word (see Table 5) in the lexicon is labelled by an appropriate emotion (happiness, sadness, surprise, fear, disgust, and anger), part of speech (noun, adjective, adverb, verb) and intensity (in the range $\langle -3; 3\rangle$, -3 being the most negative and 3 the most positive). Emotions can be typically positive or negative. For example, sadness is a negative emotion, but a surprise can be both positive and negative. In such a case, the resulting polarity depends on the context in the form of the surrounding words. The dictionary used in our study was created using a web based application and contains about 19 000 words with information about polarity and emotions (see Table 5). We consider the wisdom of the crowd being the most straightforward way of obtaining data from users.

Word	Part of Speech	Emotion	Polarity
horlivo (eagerly)	adverb	joy	3
nenávidím (hate)	verb	angry	−3
horšia (worse)	adjective	sadness	−3
bezstarostnosť (carelessness)	noun	joy	0

Table 5. Examples of Slovak words in the dictionary with the part of speech tag, type of emotion and polarity degree

The experiment was performed on the same dataset, which was used in the second experiment SD2 within "Opinion Classification Based on a Special Lexicon". This dataset contained 5 242 comments and 182 645 words, where 2 573 comments were positive and 2 669 negative. Thus, in this experiment, objects were also equally divided into classes. The results of tests are given in Table 6. The highest recall was achieved for the emotion "happiness" and the lowest for the emotion "surprise". It could be caused by the fact that the emotion "surprise" is hard to label (for every other emotion it is easy to determine either positivity or negativity but surprise can represent both classes). The labelling of emotions was based on computing the probability of each emotion for the given text. For the rest of the emotions, we needed to improve our dictionary by adding new words to it because the values of precision and recall were low.

The results of our experiments also showed that we might need to reconsider using Plutchik's wheel of emotions illustrated in Figure 1 which adds two basic emotions (anticipation and trust) to Ekman's six emotions. It could cover a wider range of words and also increase a recall for each emotion.

In the tests presented in Table 6, the evaluation through precision and recall was based on comparison of our implementation of "opinion" on the resulting emotion

Emotion	Precision	Recall
Happiness	0.651	0.701
Sadness	0.589	0.590
Surprise	0.423	0.382
Fear	0.566	0.506
Disgust	0.473	0.424
Anger	0.622	0.651

Table 6. Achieved precision and recall of tests of the designed approach

with an "opinion" of human experts on the emotion presented in the text. These comparisons were made by a contingency table, from which all values of precision and recall were calculated. In this contingency table, all emotions were represented by six classes. Of course, there can be more than one emotion in one review. If that is the case, the resulting emotion is the emotion, which is the most probable, because

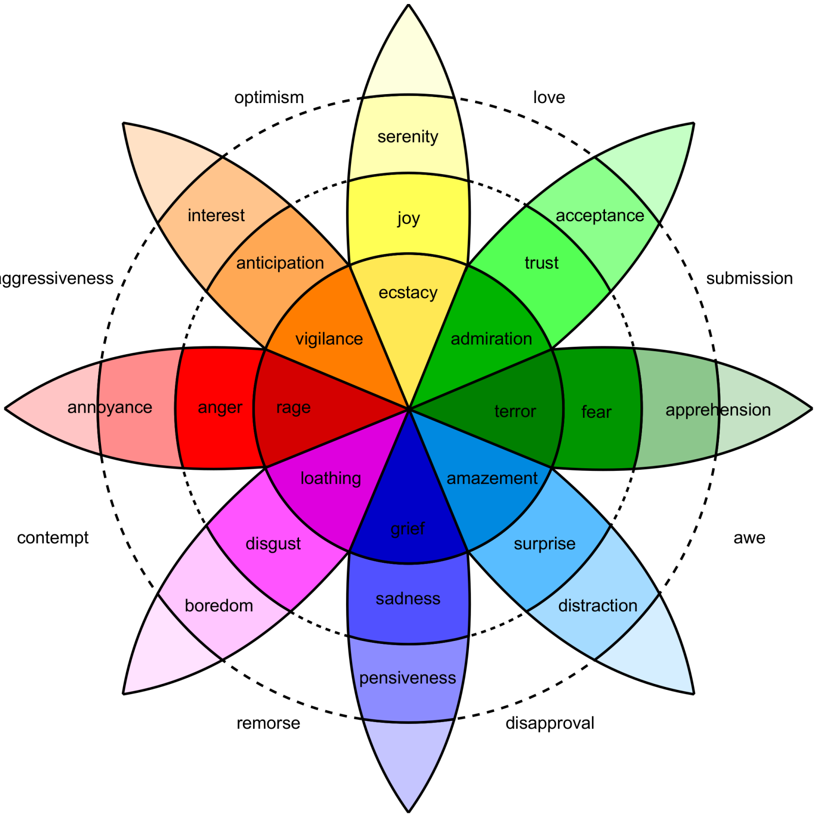

Figure 1. Plutchik's wheel of emotion [23]

the given review contains most words labelled with the considered emotion. When more than one emotion achieves the same number of votes, then all these emotions are taken into consideration. The result is that the given input text contains more emotions and this fact has to be a part of using the emotion classification for practical purposes. For example, a robot should adapt its behaviour to all these emotions. A problem can arise only when these emotions are in a contradiction. In this case, the result is not usable (e.g. in human–robot interactions).

4 AUTHORITY IDENTIFICATION

This section is focused on the identification of the authority of persons, who comment a topic on a web discussion forum. Authority identification represents a different problem than authorship identification. It does not answer the question "Who is the author?". Authority identification answers the question "Does the author have a deep knowledge of the topic?".

Authority can be formal (stated by a measure of power) and informal (stated by a measure of favour). The formal authority is given by a position or function in the organization. The informal authority is given by his/her credibility, wisdom, orientation, ability of good decision making, etc. This authority is enforced by other people's respect. Our approach has focused on the informal authority identification in web discussion forums. During the process of an online discussion, the structure of a conversation is created. This structure can be represented by an acyclic graph – the conversation tree.

People have various reasons for contributing to a discussion forum. Many of contributors are people, who want to find answers to their questions. These contributors create a core of the discussion, but they are not very authoritative. A smaller group of contributors is the group of troublemaking actors. They are provocateurs seeking an opportunity to present their opinions and invoking conflicts. They are not authoritative as well. They should be eliminated from the discussions. The last group of contributors are actors, who express their knowledge and share their ideas or opinions. These actors enter into the discussion seriously, add only truthful information, and join only if they are familiar with the topic. They are really authoritative contributors. We are interested in distinguishing them from the other actors.

The search for authoritative actors involves mining from web discussions. The input data contain the following aspects of comments:

- contributor name,
- reacting comment,
- length of the comment,
- position of the comment in the discussion represented graphically by the conversation tree.

The problem of authority identification is based on the estimation of authoritativeness *(A)*. The value of A is related to the contributors. Thus, data about each of 117 contributors in our dataset were collected including the following information:

- *NC* is the number of comments of a given contributor. We proposed that someone who understands the topic (authority) would contribute to the discussion more often than other actors.
- *ANR* is the average number of reactions to the comment(s) of a given author. This argument represents the number of reactions that support or negate a statement of the author, whose authority is examined. We started from the assumption that a more authoritative contributor could evoke a higher number of reactions.
- *AL* is the average number of all layers, at which the comments of a discussant are situated in the conversation tree. The conversation tree is a graphical representation of a web discussion. The *AL* represents the information, when the discussant joins the discussion, at the beginning or at the end. For example, a contributor, whose comments are located at the bottom level of the conversation tree, usually adds comprehensive comments answering all the questions. This can be the authoritative type of contributors.
- *NCH* is the number of characters, which represents the average length of comments of an author. This number is a common ratio of the number of all characters of the given contributor to the number of all his/her comments in the discussion. It penalizes authors with too short and thus less informative comments. We assume that an authoritative contributor does not post extremely short comments.
- *K* is karma of a contributor in the form of a number from 0 to 200, which represents the discussant's activity in the last 3 months from "`www.sme.sk`".
- *AE* is the average evaluation of a comment in the form of the ratio of the sum of all reactions (agree (+) and disagree (−)) to this comment of a given discussant to the number of all his/her comments. This average evaluation is available on the web discussion page. The AE range is a number from 0 to 80.

For the informal authority identification (detection), the main task is to estimate the function of authoritativeness A. In general, it is the function (4):

$$A = f(NC, ANR, AL, NCH, K, AE). \tag{4}$$

Firstly, we used a linear function with weights determined experimentally followed by regression analysis to compute weights of linear, polynomial and nonlinear functions. To compute these weights, it was necessary to know the values of the independent variables *NC*, *ANR*, *AL*, *NCH*, *K*, *AE*, as well as the values of the dependent variable *A*. The values of the variable *A* were derived from:

- evaluation of each discussant by "human expert",

- evaluation of each discussant by other discussants – it represents "wisdom of the crowd".

The following regression functions for authority estimation were generated in the process of learning:

- linear function learned from the "human expert",
- linear function learned from the "wisdom of the crowd",
- polynomial function learned from the "human expert",
- polynomial function learned from the "wisdom of the crowd",
- non-linear function learned from the "human expert",
- non-linear function learned from the "wisdom of the crowd".

All these 6 functions were tested and the validation was performed using the following measures:

- an average deviation – used for the validation of estimation functions,
- a precision – used for the validation of classification,
- a recall – used for the validation of classification.

We classified authors into two classes: Authority and Non-authority. A contributor with the estimated value of authoritativeness A greater than 70 was labelled as Authority otherwise he/she was labelled as Non-authority. The value of authoritativeness A could be in the interval (0, 100). The test dataset contained the data on 117 contributors. According to the results presented in Table 7, the best results were obtained by learning of linear function from the "wisdom of the crowd" in formula (5) and by learning of non-linear function from the "wisdom of the crowd" in formula (6). Surprisingly, learning of polynomial function from the "wisdom of the crowd" also provided good results but only in "recall".

$$A = 0.4385AE + 0.325K + 0.002NCH - 0.2928AL - 0.0853ANR + 1.0728NC, \tag{5}$$

$$A = 0.0185AE^{1.8135} + 141.5704K^{-78.39} + 0.0018NCH^{1.0457} - 0.0011AL^{3.7717} - 0.5562ANR^{0.0001} + 37.6642NC^{0.0038}. \tag{6}$$

Authority identification can be used in a variety of real situations. For example, an inexperienced web user searches for an authority that is able to provide him/her with advice and decision-making support. Another example – a technically oriented organization requires skilled employees, specialists who are authorities in the given field, and to whom such organization can offer interesting job positions. Thus, the person responsible for recruiting can search for authoritative users on web forums focused on the technologies used in this organization to fill in specific job positions.

	Deviation		Precision		Recall	
Version	Expert	Crowd	Expert	Crowd	Expert	Crowd
Linear	17.34	**3.29**	0.70	**0.98**	0.67	**0.80**
Polynomial	24.01	8.79	0.67	0.78	0.61	0.94
Non-linear	18.11	**6.56**	0.67	**0.97**	0.67	**0.80**

Table 7. Achieved average deviation, precision and recall of tests of the designed approach to the authoritativeness identification

5 CONCLUSIONS

The paper introduces a variety of approaches to social conversation data mining. The main attention is focused on two problems: sentiment analysis (opinion and emotion classification) and authority identification. It describes two approaches to opinion classification and one approach to authority identification.

Using the opinion classification approach, the comments were classified into two classes: positive and negative. The first opinion classification method used 4-grams to assign polarity to comments. It was able to process intensification and negation within the range of 4 words. This approach achieved good results for positive comments. The classification of negative comments was worse. That was the reason to develop the second opinion classification approach and create a new lexicon for this new approach. We also used a different method to process intensification and negation. This second method achieved better results than the one used previously.

The paper also describes the approach to emotion classification. Such approach mostly used to identify emotions in a text is similar to the approach used for identifying the polarity of the text. We focused on the lexicon-based approach in both cases – therefore, we created a lexicon that gave us information about emotions. By applying this lexicon we obtained interesting results. However, the results also showed that we might need to reconsider using Plutchik's wheel of emotions which adds two basic emotions (anticipation and trust) to Ekman's six emotions for the approach to be more precise in emotion classification. Changing the models also required reworking of the lexicon. In addition we had to take into consideration that emotions have no strict boundaries which means they often overlap each other, so it was a challenge to differentiate them properly.

Implementation of a new approach to authority identification in web discussions is presented and the resulting rating of authoritative contributors is provided. It should be noted that the linear model is better than other models. It is because of the character of input data (parameters of the web discussion) and also due to the character of the issue which is discussed on the web. Nevertheless, the linear model is sufficient for authority estimation. It is clear that learning from the "wisdom of the crowd" is better than learning from a "human expert". The reason might be that an expert's opinion can be biased whereas a combined opinion of many discussants is probably more objective.

In our future work, we want to combine knowledge gained from opinion classification and authority identification. We suppose that a more authoritative author has a greater influence on the resulting summarized sentiment. In the known approaches to sentiment analysis, each web forum comment has the same weight. We will use evolutionary algorithms [18, 5] to find an appropriate form of estimation function to calculate the authority value and then apply it to opinion classification. The comments written by authoritative users have higher weight and they will be classified with higher priority. We would like to apply the weighted opinion analysis in the domain where we could be able to recognize a person's aberration based on his/her written text [24] and also how to decrease the cognitive load for the web users [19].

Acknowledgements

The work presented in this paper was supported by the Slovak Grant Agency of the Ministry of Education and Academy of Science of the Slovak Republic under VEGA grant No. 1/0493/16 and by the Slovak Research and Development Agency under the contract No. APVV-017-0267.

REFERENCES

[1] Baccianella, S.—Esuli, A.—Sebastiani, F.: SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC '10), Valletta, 2010. European Language Resources Association (ELRA), 2010, pp. 2200–2204. ISBN 2-9517408-6-7.

[2] Balahur, A.—Kabadjov, M.—Steinberger, J.—Steinberger, R.—Montoyo, A.: Challenges and Solutions in the Opinion Summarization of User-Generated Content. Journal of Intelligent Information Systems, Vol. 39, 2012, No. 2, pp. 375–398, doi: 10.1007/s10844-011-0194-z.

[3] Benamara, F.—Cesarano, C.—Picariello, A.—Recupero, D. R.—Subramanian, V. S.: Sentiment Analysis: Adjectives and Adverbs are Better Than Adjectives Alone. Proceedings of the First International Conference on Weblogs and Social Media (ICWSM 2007), Boulder, 2007.

[4] Bouguessa, M.—Dumoulin, B.—Wang, S.: Identifying Authoritative Actors in Question-Answering Forums – The Case of Yahoo! Answers. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08), 2008, pp. 866–874, doi: 10.1145/1401890.1401994.

[5] Cadrik, T.—Mach, M.: Evolution Classifier Systems. Electrical Engineering and Informatics IV. Proceedings of the FEI Technical University of Košice, Košice, 2013, pp. 168–172 (in Slovak). ISBN 978-80-553-1440-2.

[6] Cambria, E.—Hussain, A.: Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis. Cham, Springer, 2015, 176 pp. ISBN 978-3-319-23654-4.

[7] Dom, B.—Eiron, I.—Cozzi, A.—Zhang, Y.: Graph-Based Ranking Algorithms for E-Mail Expertise Analysis. Proceedings of 8th ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD '03), 2003, pp. 42–48, doi: 10.1145/882082.882093.

[8] Ekman, P.: An Argument for Basic Emotions. Cognition and Emotion, Vol. 6, 1992, No. 3-4, pp. 169–200, doi: 10.1080/02699939208411068.

[9] Esuli, A.—Sebastiani, F.: SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. Proceedings of the 5th Conference on Language Resources and Evaluation (LREC '06), 2006, pp. 417–422.

[10] Go, A.—Bhayani, R.—Huang, L.: Twitter Sentiment Classification Using Distant Supervision. Stanford University, available on: `http://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf`, 2013.

[11] Grandjean, D.—Sander, D.—Scherer, K. R.: Conscious Emotional Experience Emerges as a Function of Multilevel, Appraisal-Driven Response Synchronization. Consciousness and Cognition, Vol. 17, 2008, No. 2, pp. 484–495, doi: 10.1016/j.concog.2008.03.019.

[12] Gunes, H.—Schuller, B.: Categorical and Dimensional Affect Analysis in Continuous Input: Current Trends and Future Directions. Image and Vision Computing, Vol. 31, 2013, No. 2, pp. 120–136.

[13] Hu, M.—Liu, B.: Mining and Summarizing Customer Reviews. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04), ACM, New York, 2004, pp. 167–177. ISBN 1-58113-888-1, doi: 10.1145/1014052.1014073.

[14] Kennedy, A.—Inkpen, D.: Sentiment Classification of Movie and Product Reviews Using Contextual Valence Shifters. Workshop on the Analysis of Informal and Formal Information Exchange During Negotiations (FINEXIN '05), Ottawa, 2005, pp. 11–22.

[15] Levallois, C.: Umigon: Sentiment Analysis for Tweets Based on Lexicons and Heuristics. Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, Georgia, 2013, pp. 414–417.

[16] Liu, B.: Sentiment Analysis and Opinion Mining (Introduction and Survey). Morgan and Claypool Publisher, 2012, pp. 1–168.

[17] Lu, Y.—Kong, X.—Quan, X.—Liu, W.—Xu, Y.: Exploring the Sentiment Strength of User Reviews. Proceedings of the 11th International Conference on Web-Age Information Management (WAIM '10). Springer-Verlag, Berlin, Lecture Notes in Computer Science, Vol. 6184, 2010, pp. 471–482. ISBN 978-3-642-14245-1, doi: 10.1007/978-3-642-14246-8_46.

[18] Mach, M.: Evolution Algorithms – Problems Solving. FEI Technical University, Košice, 2013, 135 pp. (in Slovak). ISBN 978-80-553-1445-7.

[19] Machová, K.—Klimko, I.: Classification and Clustering Methods in the Decreasing of the Internet Cognitive Load. Acta Elektrotechnica et Informatica, Vol. 6, 2006, No. 2, pp. 52–56. ISSN 1335-8243.

[20] Machova, K.—Marhefka, L.: Opinion Classification in Conversational Content Using N-Grams. In: Badica, A., Trawinski, B., Nguyen, N. (Eds.): Recent Developments in Computational Collective Intelligence. Springer, Cham, Studies in Computational Intelligence, Vol. 513, 2013, pp. 177–186.

[21] Osgood, C. E.—May, W. H.—Miron, M. S.: Cross-Cultural Universals of Affective Meaning. University of Illinois Press, Urbana, 1975.

[22] Pang, B.—Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL '04). Association for Computational Linguistics, Stroudsburg, 2004, doi: 10.3115/1218955.1218990.

[23] Plutchik, R.: The Nature of Human Emotions. Scienceweek, 3 August 2001, pp. 1–2.

[24] Saloun, P.—Ondrejka, A.—Malčík, M.—Zelinka, I.: Personality Disorders Identification in Written Texts. In: Duy, V., Dao, T., Zelinka, I., Choi, H. S., Chadli, M. (Eds.): Recent Advances in Electrical Engineering and Related Sciences (AETA 2015). Springer Verlag, Lecture Notes in Electrical Engineering, Vol. 371, 2016, No. 1, pp. 143–154. ISBN 978-331927245-0. ISSN 1876-1100, doi: 10.1007/978-3-319-27247-4_13.

[25] Smatana, M.—Koncz, P.—Paralič, J.: Semi-Automatic Annotation Tool for Aspect-Based Sentiment Analysis. FEI Technical University of Košice, 2013, pp. 1–3.

[26] Strapparava, C.—Valitutti, A.: WordNet-Affect: An Affective Extension of WordNet. Proceedings of the 4th International Conference on Language ReSources and Evaluation. European Language Resources Association (ELRA), 2004, pp. 1083–1086.

[27] Taboada, M.—Brooke, J.—Tofiloski, M.—Voll, K.—Stede, M.: Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics, Vol. 37, 2011, No. 2, pp. 267–307, doi: 10.1162/COLI_a_00049.

[28] Tan, S.—Zhang, J.: An Empirical Study of Sentiment Analysis for Chinese Documents. Expert Systems with Applications, Vol. 34, 2008, No. 4, pp. 2622–2629. ISSN 0957-4174.

[29] 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, Vol. 61, 2010, No. 12, pp. 2544–2558, doi: 10.1002/asi.21416.

[30] Wundt, W.: Grundriss der Psychologie [Fundamentals of Psychology]. 7th revised edition. Engelmann, Leipzig, 1905 (in German).

[31] Zhang, J.—Ackerman, M. S.—Adamic, L.: Expertise Networks in On-Line Communities: Structure and Algorithms. Proceedings of the 16th ACM International World Wide Web Conference (WWW '07), 2007, pp. 221–230, doi: 10.1145/1242572.1242603.

[32] Zhang, Z. et al.: Sentiment Classification of Internet Restaurant Reviews Written in Cantonese. Expert Systems with Applications, Vol. 38, 2011, No. 6, pp. 7674–7682. ISSN 0957-4174.

Kristína Machová graduated (M.Sc.) in 1985 at the Department of Technical Cybernetics at the Technical University in Košice. She defended her Ph.D. thesis in the field of machine learning in 1996. She is Associate Professor at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at the Technical University of Košice. Her scientific research has focused on data mining within conversational web content, dictionary approach and machine learning approach to sentiment analysis, opinion and emotion classification to be applied in robotics, authority identification, text document processing using classification and clustering machine learning methods.

Martin Mikula received his M.Sc. degree in 2014 at the Department of Cybernetics and Artificial Intelligence at the Technical University in Košice. Currently, he is a postgraduate student. He studies business informatics at the Faculty of Electrical Engineering and Informatics at the Technical University in Košice. His research interests are big data, including knowledge discovery, data mining, text mining and sentiment analysis in conversational content.

Martina Szabóová graduated (M.Sc.) in 2012 at the Department of Cybernetics and Artificial Intelligence at the Technical University in Košice. She is in her last year of Ph.D. study. Her main research activities intersect two fields. The first one lays in natural language processing within the scope of emotion analysis from texts. The second field covers interactive learning systems in the social human-robot interaction.

Marián Mach graduated (M.Sc.) in 1985 at the Department of Technical Cybernetics at the Technical University in Košice. His Ph.D. thesis on uncertainty processing in expert systems was defended in 1992. He is Associate Professor at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at the Technical University of Košice. His scientific interests are knowledge management, data and web mining, opinion classification, information retrieval, semantic technologies, and knowledge modeling.