Classification of opinions in conversational content

Martin Mikula and Kristína Machová
Department of cybernetics and UI, FEI, Technical University Košice,
Letná 9, 042 00, Košice,
Martin.Mikula@tuke.sk, Kristina.Machova@tuke.sk

Abstract—**Nowadays, with enhancing possibilities of the Internet usage, the number of its users grows as well. People use it more and more to communicate among themselves. This kind of communication plays a significant role in the decision-making process. Based on this finding, a need to analyze the content of the ample web discussions (so-called conversational content) using the computers arose. Therefore, the following article deals especially with the issue of opinion analysis, more specifically the classification of opinions. We have created an algorithm, which allows determining the polarity of the text. With the analysis of text we can also process the intensification, negation and their combinations. We have created 4 classification dictionaries divided according to the types of words they contain. We have subsequently tested the algorithm with average accuracy of 86%.**

***Keywords*—opinion classification, conversational content, web discussion, dictionary approach**

I. Introduction

Contemporary web allows communication of a great amount of people all around the world. People share information and their experiences. This information exchange also influences others who utilize the Internet. For example, if somebody wants to buy a new product, at first, they often search through the available information and read a few web discussions or comments that include reviews of the owners, of the product's functionality and the performance of this product. However, these discussions are often too broad, so it is useful to develop an application enabling automatic classification of reviews on the web.

The opinion classification, sentiment classification, is a process of analyzing user's opinions and sentiments on a given topic. The opinion is determined by the evaluation factor (can be positive or negative) and strength. The strength of the evaluating factor is dependent on the level of the intensity of the word's polarity and their quantity. Under the term topic, we understand e.g. rating products, persons, books, etc. The author and owner of the opinion is a person, who has a specific opinion on a given object. An object is defined as a topic to which a given opinion relates.

II. Procedure of Opinion Classification

In principle, there are two different basic approaches to solve the problem of opinion classification [7]. The first approach is based on lexicons. This approach involves calculating orientation for document from the semantic orientation of words or phrases. These words and phrases are saved in dictionaries. The dictionary is called the opinion lexicon. The approach of using opinion words to determine orientation of text is called the lexicon-based approach. Dictionaries for this method can be created in different ways: manually, using existing dictionaries or semi-automatically by using resources like WordNet or SentiWordNet. SentiWordNet is a lexicon for sentiment analyzing derived from WordNet by leveraging word relationships and word glosses. SentiWordNet sets opinion score of words with given meaning using semi-automated method where small bag of words are manually labeled. The remaining database is derived using an automated method. The dictionary can also be produced automatically via association. Score of a new word is calculated using the frequency of the proximity of that word with respect to one or more seed words. Seed words is a small group of words, which are strongly positive or negative. The association is usually calculated by Turney's method for computing mutual information [9]

The second approach is based on statistical or machine-learning approach. Machine-learning uses mainly SVM (Support Vector Machine) and KNN (K-Nearest Neighbors) methods. Classifiers, using SVN, are trained on a particular data set using features such as unigrams or bigrams, with or without POS (part-of-speech) labels. The most successful features seem to be basic unigrams [8]. Classifiers built on supervised methods reach a high accuracy in detecting the polarity of a text. Although such classifiers have very good results in the domain that they are trained on, their performance drops down when the same classifier is used on different domain.

Sometimes this division is described as division of endogenous and exogenous methods [4]. Endogenous methods don't require any external information instead of information that is a part of algorithm or sample data. Exogenous approach needs an external data like dictionaries or manually created rules.

We decided to use lexicon-based method, in which we use dictionaries of words annotated with the word's semantic orientation.

A. Word subjectivity

We do not need all the words for classification. The only interesting words are those, which express a certain attitude towards the subject. Most of them are adjectives, but we also use verbs, nouns and adverbs. The dictionary approach focuses on finding these kinds of words in the text of posts, while ignoring the words that do not have subjectivity. Determination of the word's subjectivity represents a certain form of preprocessing of the text when the words are sorted into ones that are useful for further classification and others.

B. Word orientation (polarity)

In the following step, the polarity of each subjective word is determined. In the simplest case, the polarity of the word can

978-1-4799-8221-9/15/$31.00 ©2015 IEEE

be positive, negative and neutral. Determination of the word's polarity is based on comparison with the dictionary prepared in advance. However, the determination of expressions with negation is not that trivial. We differentiate 2 types of determining the polarity of the negation:

- switch negation - the value of the word changes into a word with equal strength but of opposite polarity
- shift negation - the value of the word is determined by the shift towards the opposite polarity by an exact value (e.g. 4)

It is an interesting aspect of negation that negative statements tend to be perceived as more marked than their affirmative counterparts, both pragmatically and psychologically. This markedness is true in terms of linguistic form, with negative forms being marked across languages, and it is also manifested as frequency distribution, with negatives being less frequent. Negation tends to be expressed in euphemistic ways, which makes negative sentiment more difficult to identify in general [9].

C. Strength of polarity

After determining the polarity of the subjective word, it is often necessary to determine the strength of its polarity. The strength of the word's polarity may be graded directly in the classification dictionary, where words are classified not only towards a positive or negative polarity, but directly to a certain degree of polarity. Sometimes the power of the word's polarity is changed (increased or decreased) by another word, the so-called intensifier, which precedes the processed word.

D. Overview of existing applications

There is a large number of works dealing with the classification of views. Taboada et al. focuses on the sentiment analysis using dictionaries [9]. In his work he uses intensification and shift negation. They created dictionary manually and compared the results with other manual or automatic dictionaries. When they tested the accuracy of the solution, its success ranged from 65% to 81%. The work of Ohama et al. deals with the possibility of using SentiWordNet on opinion mining [6]. They compared manually created dictionary with a method utilizing the SentiWordNet. SentiWordNet obtained good results in comparison with manually created lexicons. This lexicon could be used for creating prediction models. The application, which deals with the classification of opinions in the Slovak language, is KLAN [5] created at TUKE. KLAN works on the dictionary principle and uses a dynamic coefficient to process the intensification and negation. The application reached an 86.2% precision for positive contributions and 69.2% for negative contributions. An interesting possibility is to link the dictionary methods with the methods of machine learning. Lexicon-based approaches give good precision but low recall. For recall improvement, we can apply methods of machine-learning. This process described Zhang et al. in his work [10]. First, the dictionary method is applied. From results of lexicon-based method they exploited information which was used as an input for the machine learning methods. They achieved average accuracy of 85.4% in their application. Choi [2] in his work presents machine-learning approach based on semantic orientation. He incorporates structural inference motivated by compositional semantics into the learning procedure. They discovered in their experiments, that simple heuristics based on compositional semantics can perform better then learning-based methods that do not incorporate compositional semantics. They achieved accuracy of 89.1%. A method that integrates compositional semantics into learning performs better than all other alternatives (90.7%). Other works dealing with this topic in the Czech language include Sentiment analysis in posts on the social network Twitter [1]. This work describes an application that uses n-grams for the sentiment analysis. The machine learning methods are also applied in the work of Koktan [3], where he uses SVM methods, Naive Bayes and Maximum entropy for sentiment analysis.

III. Proposal of approach to opinion classification

We have created an algorithm for the classification of opinions, which works in three steps. The first step is to obtain the text we want to analyze. The second step splits the text into sentences and words. The words are modified and compared with the words in the dictionary. In case of compliance, the words in the text are assigned an appropriate level of polarity from the dictionary. The third step determines the resulting strength of polarity of the entire contribution as the sum of polarities of individual words.

To determine the strength of the polarity we have chosen a scale with values ranging from -3 to 3 (see Table 1.). We solved the use of negation in the text through the so-called switch negation (e.g. a strong positivity becomes a strong negativity). For intensifiers, we have chosen the values from 1:00 to 2:00 depending on the strength with which they enhance the polarity (e.g. an extraordinary - 1.5). This way it is possible to increase the intensity depending on the strength of the word we want to intensify (words with stronger polarity will be more enhanced than words with weaker polarity).

A. Obtaining and processing of posts

The texts of posts, which are to be analyzed, can be input in the application in three different ways. Either the text is entered manually using the keyboard, or a group of posts from a text file is loaded, or a web address of a discussion is entered and the comment texts are downloaded directly using HTML tags.

During the processing, the text is split into individual sentences and subsequently the diacritic is removed from the text. Then individual sentences are split into elementary units - words. Adjectives are usually the main carrier of the polarity.

TABLE I. DEGREES STRENGTH OF THE POLARITY

Strength of the polarity	Polarity
3	strong positive
2	moderate positive
1	weak positive
0	neutral
-1	weak negative
-2	moderate negative
-3	strong negative

Therefore a modified version of Lancaster stemming algorithm, which is available on the web page[1], was used. It converted them into nominative of the plural. Prefixes such as kilo-, mega-, mini, mili-, etc., are removed from the words. Then suffixes are found and replaced by the predefined. Since in Slovak the same case endings in primary form can have "y" and "i" at the end, a nominative plural was selected, where only "i" is always assigned at the end.

- zlej $\rightarrow$ zlý $\rightarrow$ zli
- lepšej $\rightarrow$ lepší $\rightarrow$ lepsi

As already stated, once the conformity of a word with a dictionary is found, the word is assigned its intensity of polarity based on the data from the dictionary. After each new word, the polarity value of the processed sentence is modified. If the processed word is an intensifier, then the algorithm searches for the next word with positive or negative polarity to which the intensifier relates and the polarity value of the found word is multiplied according to the strength of the intensifier. If the program finds a negation, it switches the value of the negated word (switch). In case all options appear in the text, the calculation is carried out in the way that the value of the current word is multiplied by the strength of the intensification and negation (-1). The resulting polarity of the sentence is adjusted:

$$log\_value = 1 + log_{10}(sentence\_value) \tag{1}$$

If the resulting value before adjustment is equal to 0, then no adaptation occurs. If the value is less than 0, we count with absolute value, which is again multiplied by the -1 to maintain a negative attitude in your post.

B. Dictionary

The proposed method represents a dictionary approach. The dictionary contains keywords of a discussion domain. Our dictionary was created by a translation from the English language. Consequently synonyms of all the words were found. We have also added intensifiers and negators. If the word ends in a vowel other than "o", it is written in the dictionary in nominative plural. If the word ends in a consonant, or the vowel "o", it is saved without modifications. Each word in the dictionary is assigned polarity based on the scales described above. Words, which invert polarity are assigned zero because when the algorithm identifies a negation, it automatically rotates the polarity.

IV. Experiments

The proposed algorithm was implemented and tested. We performed the testing using 4 dictionaries. At the moment, there is no standard dataset in the Slovak language for testing sentiment analysis. We wanted to create big dataset with more than 250 comments. Therefore we have decided to select the testing data from discussions of the high rated movies The Green Mile and Forrest Gump (www.csfd.cz). We downloaded comments from web page. Than the comments were translated into Slovak, since originally they had been written mainly in Czech, and in order to maintain their objectivity, grammar and stylistic errors were left. After this language translation, expert read all of these comments and assigned positive or negative polarity. We removed neutral comments from dataset. We collected them into text file with spetial formatting for automated testing. The sample contained 2500 entries. We are working on setting up a standard dataset that will be used for further testing in the future.

We considered the evaluation of the comment to be correct (good, valid) if it coincided with the evaluation done by the expert. Based on the conformity or disagreement, the precision and recall of the method was calculated. A positive precision is a ratio of correctly evaluated positive comments (tp) to all posts marked by the algorithm as positive (tp + fp). A positive recall is a ratio of correctly evaluated positive comments (tp) to all posts marked by the expert as positive (tp + fn). The same method of calculation was used to calculate the precision and recall for neutral and negative comments. A negative precision is a ratio of correctly evaluated negative comments (tn) to all posts marked by the algorithm as negative (tn + fn). A negative recall is a ratio of correctly evaluated negative comments (tn) to all posts marked by the expert as negative (tn + fp). Accuracy is a ratio of correctly evaluated positive and negative (tp + tn) posts to all posts (tp + tn + fp + fn).

$$Precision = \frac{tp}{(tp + fp)} \tag{2}$$

$$Recall = \frac{tp}{(tp + fn)} \tag{3}$$

$$Accuracy = \frac{tp + tn}{(tp + fp + tn + fn)} \tag{4}$$

A. Discussion testing

The discussion contains 2324 positive and 176 negative comments. We distinguished only positive and negative comments. When algorithm set contribution as neutral we marked it as bad result. The application achieved high precision and recall of the positive comments (see Table 3). In average the precision for positive comments was around 95.3% and 89.3% for the recall. Precision of negative contributions was around 23.5% and recall was in average 45.5%. We achieved accuracy about 86% which means that a small number of negative contributions did not affect the accuracy.

Presented results were mainly affected by numbers of positive a negative comments. When algorithm marked 10 positive comments incorrectly, it was only little percent (around 0.5%). But when we marked 10 negative comments incorrectly, it was more than 5%. Positive opinion was expressed directly by words, that were in dictionary. In opposite of this, negative opinion was expressed by description of parts of the films. Next problem was also irony, when positive description had

TABLE II. Contigency table for precision and recall

	Positive set	Negative set
Positive by algorithm	true positive (tp)	false positive (fp)
Negative by algorithm	false negative (fn)	true negative (tn)

[1]http://www.comp.lancs.ac.uk/computing/research/stemming/index.htm

TABLE III. EVALUATION OF PRECISION AND RECALL OF THE DISCUSSION

Measures	Precision(%)		Recall(%)		Accuracy(%)
Dictionary	Pos.	Neg.	Pos.	Neg.	
Dictionary (pos. & neg.)	95.2	25	89.7	41	86.2
Dictionary (intensifier)	95.3	22.8	89.2	42	85.9
Dictionary (nagetive)	95.4	23.5	89.4	43.2	86.1
Dictionary (all)	95.4	23	88.9	43.7	85.7

negative meaning. Problematic were comments where author described film as positive but he changed his opinion after he had seen the movie a few times.

V. CONCLUSION

The web services, which deal with the analysis of web discussions, are becoming more and more popular. It is mainly because today's busy person no longer has time to read the whole discussion on a topic.

In this work a dictionary approach for classification of opinions is presented. This method is able to process multiple intensification and negation and also intensify negation and negate intensification. The presented application achieved average precision of 95.3% for positive comments and about 23.5% for negative comments. We wanted to compare our approach with others. We chose one approach in Slovak language (Klan) and 4 approaches in English. Comparing with Slovak approach, we achieved better results for precision of positive comments than the Klan (86.2%) but worse precision of negative than the Klan application (86.2%), which is a significantly lower value. We compared our approach also with approaches in English. The main problem in this comparison is, that every approach was tested on different datasets, but we wanted to compare with current level of knowledge in the world. We can see accuracies in Table 4.

- Taboada - dictionary based approach, with intensification, shift negation and different dictionaries
- Zhang - mix of dictionary approach and machine learning approach
- Choi (sipmle) - heuristics based method on compositional semantics
- Choi (integrated) - integrates compositional semantics into machine learning
- Our approach - approach described in this work

The first method (Taboada) based on dictionary approach obtained lower accuracy than our approach (max 81% vs 86%). Method which combines lexicon methods and machine-learning also had lower accuracy (85% vs 86%). Next method

TABLE IV. COMPARISON OF ACCURACY FOR DIFFERENT APPROACHES

Approach	Accuracy (%)
Taboada	65-81
Zhang	85.4
Choi (simple)	89.1
Choi (integrated)	90.7
Our approach	86

was based on heuristics with compositional semantics (Choi simple). This approach had better accuracy than our method (89.1% vs 86%). The last method integrates compositional semantics into machine-learning (Choi integrated). It achieved the best accuracy (90.7%). We achieved better accuracy than approaches using lexicons and worse accuracy than methods based only on machine learning. Evaluation of positive comments achieved good results. They slightly improved using the intensifiers. Adding the words that turn the polarity, the results deteriorated. It was probably caused by the use of switch negation, which is not very accurate. The test results were also influenced by a low number of neutral and negative posts in proportion to the positively evaluated comments. Neutral comments were mostly the ones that rated the objective side of the movie, such as lines from the movie, commentaries from filming of the movie or Oscar nominations.

The most problematic were comments in which the author first described the movies as positive, but eventually changed his mind. The comment was therefore evaluated by the expert as negative, but the implementation evaluated it positively according to its initial positive description. Also, when the film was evaluated positively, it was usually a direct evaluation; while the negative evaluation was based mostly on descriptions of the parts that were not liked by the author of the comment. Another type of problematic comments was the one that rated the movie using points or percentages without further description. A problem is also the processing of irony and ambiguity.

ACKNOWLEDGMENT

This contribution was supported by the VEGA Agency under project no. 1/1147/12 "Methods of analysis of collaborative processes implemented through information systems".

REFERENCES

[1] J. Buryan, "Analýza sentimentu v příspěvcích na sociální síti Twitter," Bakalářská práce. Brno: Masarykova univerzita, Fakulta informatiky, pp 43, 2013 Available: <http://is.muni.cz/th/374039/fi_b/>.

[2] Y. Choi, C. Cardie, "Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis" in Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 793-801, 2008.

[3] M. Koktan, "Automatické rozpoznávání (analýza) sentimentu," Diplomová práce. Plzeň: Západočeská univerzita v Plzni, pp. 60, 2012.

[4] P. Koncz, "Aspektovo orientovaná analýza sentimentu," Písomná práca k dizertačnej skúške. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, pp. 59, 2012.

[5] K. Machová, M. Krajč, "Klasifikácia názorov vo vláknových diskusiách na webe," in: Znalosti 2011 : 10. ročník konference : Stará Lesná, Vysoké Tatry : 1. ledna - 2. února 2011, sborník příspěvků, Ostrava, VŠB-TU, pp. 136—147, 2011.

[6] B. Ohama, B. Tierney, "Opinion Mining with SentiWordNet," in Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains, pp. 266—286, 2010.

[7] B. Pang, L. Lee, "Opinion Mining and Sentiment Analysis," Foundation and Trends in Information Retrieval, Vol.2, No.1-2, pp. 1—135, 2008.

[8] B. Pang, L. Lee, S. Vaithyanathan "Thumbs up? Sentiment Classification using Machine Learning," in Proceeding of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 79—86, 2002.

[9] M. Taboada, et al., "Lexicon-based methods for sentiment analysis," Comput. Linguist., vol. 37, no. 2, pp. 267—307, 2011.

[10] L. Zhang, et al., "Combining Lexicon - based and Learning - based Methods for Twitter Sentiment Analysis," HP Laboratories, [cit 20.10.2014] 2011 Available: <http://www.hpl.hp.com/techreports/2011/HPL-2011-89.html>.