Active learning enhanced semi-automatic annotation tool for aspect-based sentiment analysis

Miroslav Smatana¹, Peter Koncz¹, Peter Smatana¹, Ján Paralič¹
¹Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic
miroslav.smatana@student.tuke.sk, {peter.koncz, jan.paralic}@tuke.sk, psmatana@gmail.com

Abstract—Aspect-based sentiment analysis has become popular research field which allows the quantification of textual evaluations of different aspects of products and services. Methods of aspect-based sentiment analysis built on machine learning usually depend on manually annotated training corpora. In order to facilitate the processes of their creation, annotation tools dedicated to this purpose are needed. In this work we proposed a semi-automatic annotation tool which uses active learning to increase the effectiveness of the documents annotation. The use of active learning adapted to the needs of aspect-based sentiment analysis is the main difference between the proposed solution and existing annotation tools. We applied it in the domain of hotels evaluations. The results of realized experiments confirmed the faster increase of the annotation suggestions quality in terms of F1-measure in comparison to the scenario without active learning.

I. INTRODUCTION

During the last years, the sentiment analysis has become popular research field devoted to the methods of automated quantification of the information contained in unstructured or weakly structured textual evaluations [1]. The earlier works usually aimed to the identification of the overall sentiment of the contribution or its parts regardless to the topic to which it is related. However, the evaluations are usually related to many different objects and their different features. The aspect-based sentiment analysis is an emerging subfield of sentiment analysis which goes beyond the usual sentiment analysis and tries to anchor the sentiment to the concrete aspect of the evaluation. Part of its methods is built on the methods of machine learning which usually depends on the manually annotated training corpora. In order to make the annotation process more effective, dedicated annotation tools are needed. However, the current annotation tools are either general or dedicated to the other domains. The aim of this work is to present an annotation tool adapted to the needs of aspect-based sentiment analysis which utilizes the methods of active learning. The rest of the work is divided as follows. The second chapter is devoted to the related works in which we overview the existing approaches in the related fields. In the third chapter, the proposed solution is described. Consequently, the fourth chapter describes the experimental verification of our approach. Finally, the last part is devoted to the conclusions.

II. RELATED WORKS

The majority of the existing works devoted to the sentiment analysis focuses only on the classification of sentiment. However, we argue that the ability to relate the content polarity to the concrete objects and their features is crucial in order to leverage the potential of the sentiment analysis. The existing solutions usually divide the aspect-based sentiment analysis into the phases of aspect identification and sentiment analysis [2]. The solution for the aspect-based sentiment analysis described in the work of Hu and Liu [3] is divided into the identification of nouns representing the object attributes and the identification of sentiment by using the dictionaries of positive, negative and negating words. A similar solution was proposed in the work of Thet et al. [4] where the aspects were identified by the use of grammar rules and sentiment by the use of dictionaries. Yang et al. [5] also identified the product attributes on the basis of frequent nouns. The opinion of their context has been classified by the use of dictionaries of positive and negative words derived from the training data. The dictionary based methods for both mentioned phases were used in the work of Kieu and Pham [6]. However, the above mentioned methods used the dictionaries for the sentiment analysis phase that can be, as in the case of the classical sentiment analysis, realized by the use of machine learning algorithms. The support vector machines have been used for the sentiment classification in the aspect-based sentiment analysis solution proposed by Jiang et al. [7]. The methods of the aspect-based sentiment analysis based on the machine learning depend on the manually annotated corpora. To support their creation, semi-automatic annotation tools can be utilized.

Dill et al. [8] proposed an automatic semantic annotation tool called SemTag. It detects the occurrences of entities and then it uses the so called Taxonomy Based Disambiguation algorithm. Kiryakov et al. [9] described in their work the semantic annotation platform called KIM (Knowledge and Information Management). KIM, similarly as TagSem, uses the ontology based annotation in which the named entities are annotated and linked to the concepts of ontology. KIM uses the GATE (General Architecture for Text Engineering) for information extraction as well as for the content and annotations management. In the experiment realized in [9] the gazetteers, the shallow analysis of the text and the pattern matching grammars were used for information extraction, however, the proposed architecture of KIM allows integration of other GATE pipelines. Song et al. [10] discuss in their work a tool called Semantator. Semantator annotates the parts of biomedical texts according to the ontologies and this annotation can be confirmed or corrected by users. To support the biomedical texts
curation, Rak et al. [11] developed a tool called Argo, based on UIMA (Unstructured Information Management Architecture). There are also general platforms for annotation tools, for example the mentioned GATE and UIMA. An overview of annotation tools with the automated creation of annotations can be also found in the work of Kiyavitskaya et al. [12].

In order to facilitate the process of annotation even more, the active learning methods can be utilized. The basic idea of active learning is the selection of unlabeled examples for manual annotation on the basis of their informativeness for the classification task. The particular active learning methods differ in the way how this informativeness is calculated. The research devoted to the active learning methods is usually realized on the single-labeled classification tasks [13]. However, in case of the aspect-based sentiment analysis, one sentence can be assigned to more than one aspect; hence this is a multi-label classification task. The comparison of different strategies for multi-label active learning in the context of the text classification can be found in the work of Esuli and Sebastiani [13]. The authors compared the different methods for ranking of unlabeled documents according to their informativeness. These methods differed in the way how the informativeness, according to particular classes, was evaluated, how the informativeness values for particular classes were combined and in the weighting of classes. Among the evaluated strategies, the minimum across classification confidences and the equal weights of classes have been shown as the best strategy. The extensive experimental evaluation of 12 methods for the multi-label learning over 11 benchmark multi-label datasets can be found in the work of Madjarov et al. [14]. Li et al. [15] in their work proposed an active learning approach for the imbalanced sentiment classification tasks. Yang et al. [16] proposed an active learning method for multi-label classification. The proposed method used the support vector machines to obtain the list of labels ordered on the basis of their confidence and prediction of the number of labels on the basis of logistic regression.

In this work the results from the mentioned research areas are combined in order to enhance the process of documents annotation for the needs of the aspect-based sentiment analysis.

III. PROPOSED SOLUTION

A. Active learning method

In this section, an active learning method applicable in the context of aspect-based sentiment analysis is proposed. The aim of this method is to select the best examples for manual annotation on the basis of the coefficients called informativeness. The overall informativeness of the example is computed as a combination of two informativeness coefficients, one for the aspect classification and the other for the sentiment classification.

Firstly, the local informativeness for aspects is computed. The local informativeness is the informativeness according to particular class. This is based on the fact that the aspect identification is a multi-label classification task for which the binary relevance method is used. The binary relevance method transforms the multi-label classification task with the multiple possible labels to a binary classification tasks, one for each label. For the creation of binary classifiers, several machine learning methods can be used. The Naïve Bayes Classifier (NBC), used in our experimental evaluation of the proposed method, is an established machine learning method for the text classification tasks including the sentiment analysis. Besides the labels assigned by classifiers, we also need the estimated values of the classification confidence, i.e. the certainty whether the prediction of class is correct. In case of the NBC, the probability of classes is computed.

In case of the multi-label classification, it is necessary to combine the values of probabilities for particular classes expressed by means of the local informativeness into the global informativeness of classified example. In our case, this was solved by the simple sum of deviations of probabilities for each class from the average probability of the example using equation 1.

\[
I = 1 - \sum_{i=1}^{n} \left[ \frac{\sum_{j=1}^{n} p_{ij} - p_{i}}{n} \right]
\]

In this equation \( p_{i} \) and \( p_{ij} \) are the probabilities of aspects and \( n \) is the number of aspects.

For each aspect, there is a classifier to classify the sentiment. This is based on the assumption that the classifiers trained for particular aspect provide more accurate classification of sentiment than general classifier. In the realized experiments, the sentiment classifiers classified the sentence on the five-point scale (from very bad to very good). We used the NBC also for the sentiment classification. The global informativeness for the sentiment was computed by the use of the above mentioned equation and it was computed for each aspect. Then, the final global informativeness of the example was computed as the average of global informativeness values for aspects.

When we obtained two informativeness values for each example, one for the aspects and the other for the sentiment, we could select examples for annotation that fits these two conditions best. We used the min-max normalization for the interval \((-\infty, 1]\) for both informativeness values and we sum them for each example.

B. Annotation tool

The active learning method described in the previous section was used in the proposed annotation tool which architecture is depicted in Figure 1. The main parts of this annotation tool are the preprocessing module, the proposed active learning method, the classification module and the manual annotation module. The annotation tool is created as an extension of the tool called Luwak\(^1\). The Luwak is a semi-automatic annotation system for textual documents. It is built on the Eclipse platform and it uses the GATE Embedded. The main function of the Luwak is the utilization of the Eclipse plugin architecture for the customization of user interface according to the needs of task. The knowledge engineer is able to customize the text preprocessing steps, the conceptual model for extraction, the graphical user interface for manual annotation and the automatic extraction algorithms.

\(^1\)http://sourceforge.net/projects/luwak/
1) **Preprocessing**

After the initial data gathering and creation of the appropriate domain model, the documents are preprocessed. We used three preprocessing steps: sentence splitting, tokenization, and stemming. The sentence splitter divides the entire text into the individual sentences that are processed. The tokenizer divides the text into the words or phrases that we call tokens. The porter stemmer cuts the suffixes from tokens. The preprocessed data are divided into the n-grams what are n words going in the sequence, the number of n was 1-3. We also manually created the seed set of documents containing the characteristic words for each aspect and each sentiment.

2) **Classification**

The classification module is designed for the classification of aspects and sentiments of examples. We used the Naïve Bayes Classifier both for the aspect classification as well as for the sentiment classification. To train these classifiers, we used the training set that was updated after each annotation. In the real application this can be done after the specified number of annotations. The n-grams from the preprocessing module create a large feature space. Hence feature selection was used before the classification of the examples. Several methods can be used for the feature selection. We used the information gain which has been shown as one of the best [17]. In case of the NBC, we computed the probability of the classes expressed in logarithms because the numerator and denoninator are very small numbers and we wanted to avoid the underflow.

3) **Manual annotation**

After the calculation of overall informativeness for each example according to the proposed active learning method, there are 10 examples selected with the highest informativeness and they are manually annotated. Moreover, the annotation tool provides the suggestions of annotations which can be accepted or corrected by the user. The newly annotated examples are added to the training set. The proposed tool also used the set of rules for the assignment of the aspect labels on the basis of the probabilities from the NBC.

IV. **Experiments**

In the series of experiments, we compared the proposed active learning solution to the baseline solution which was represented by the same architecture except for the active learning and the seed documents.

A. **Sample**

The sample described in this section was created from the dataset of Ganesan [18]. This dataset contains evaluations of hotels created by users. Each evaluation was divided into the sentences and this set of sentences represented the pool of 270 unlabeled sentences used as the sample. For the annotations, we selected the following attributes of hotels: cleanness, comfort, location, services, personnel, price-quality ratio. These attributes correspond with the aspects of evaluations. The test sample was balanced: it means that the number of sentences for each aspect and sentiment was the same.

For the first iteration of classifiers training, the seed set of documents for each aspect and each sentiment that was added to the training set was used. After the annotation of each sentence, it was added to the training set, so the number of features increased after each annotation.

B. **Results**

For the evaluation of proposed solution, we measured the increase of the F1-measure for the identification of the aspects according to the number of annotated documents as well as compared the F1-measure of the aspects classification with and without the active learning. The results achieved in the experiments are presented in the graphs in the Figures 2 and 3.

In the first graph, the increase of F1-measure for aspects according to the number of annotated sentences is presented. We can see that, at the beginning of the annotation, the F1-measure was relatively high. This was caused by the use of seeds for the first annotations. After the following decrease caused by the small number of annotated sentences and the lower importance of the seeds, we can see the continual increase of the F1-measure. The average F1-measure for all aspects achieved after the annotation of all 270 sentences was more than 62%.

In the second figure, the relation between the F1-measure and the number of annotated examples for the active learning enhanced annotation tool and annotation tool without the active learning is presented. As we can
see, the difference between the annotation with and without active learning, after annotation of 270 examples, was more than 6%. The active learning enhanced annotation tool needed only about 190 sentences for the same quality of annotation suggestions as the annotation tool without the active learning, after the annotation of 270 sentences.

The precision of sentiment for both annotation tools was also measured, where the precision of active learning enhanced annotation tool was more than 60% and the precision of annotation tool without active learning was about 57%.

![Image](Figure 2. Relation between the number of annotated sentences and the value of F1-measure of aspects classification)

![Image](Figure 3. Relation between the number of annotated sentences and the value of F1-measure of aspects classification with and without active learning)

V. CONCLUSION

In this paper we described an active learning enhanced semi-automatic annotation tool which could be used to create the annotations for the needs of aspect-based sentiment analysis. The achieved results show that the proposed active learning method increased the accuracy of the annotation suggestions. However, more detailed evaluation on bigger dataset is needed.

ACKNOWLEDGMENT

This work was partially supported by the Slovak Grant Agency of the Ministry of Education and Academy of Science of the Slovak Republic under grant No. 1/1147/12 and partially by the Slovak Research and Development Agency under the contract No. APVV-0208-10.

REFERENCES