https://doi.org/10.1109/CSCI46756.2018.00059.

2018 International Conference on Computational Science and Computational Intelligence (CSCI)

Sentilyzer: Aspect-Oriented Sentiment Analysis of Product Reviews

* Short Paper submitted to CSCI18: Symposium on Computational Science (CSCI-ISCS)

Wladislav Schulz, Johannes Zenkert, Christian Weber, André Klahold, Madjid Fathi
Department of Electrical Engineering and Computer Science
Institute of Knowledge Based Systems and Knowledge Management
University of Siegen
Siegen, Germany

***Abstract*—Aspect-oriented sentiment analysis is used in particular on textual product reviews to identify positive or negative product characteristics. This information is beneficial not only for customers related to purchasing decisions, but also used by manufacturers to improve their products. In this short paper, we introduce Sentilyzer, a system which performs state-of-the-art aspect-oriented sentiment analysis based on automatically generated category-specific sentiment and aspect dictionaries from Amazon product reviews.**

***Index Terms*—Sentiment Analysis, Product Reviews, Aspect Orientation, Knowledge-based System**

I. Introduction

The digital age is characterized by constant information growth. In particular, the emergence of user-generated content has increased due to the strong growth of social media such as blogs, forums, rating portals and social networks. People increasingly tend to share their views, opinions and experiences on such platforms, for example by writing product reviews, writing (micro) blog entries or participating in discussions in forums or other social networks. This creates huge amounts of subjective information in digital form that are of great benefit to both customers and manufacturers.

In order to evaluate subjective information effectively, systems are needed that can perform analysis steps automatically and help to generate sentiment and aspect dictionaries. This paper introduces the Sentilyzer system, designed to automate aspect-oriented sentiment analysis using Amazon product reviews.

II. Background

Various terminologies can be found in the literature for the concept of sentiment analysis. Frequently used terms are "opinion mining", "sentiment analysis" and "subjectivity analysis" [1]. The inconsistent terminology is mainly due to the different backgrounds of the researchers. While the term "opinion mining" has its origin in the fields of web search and information retrieval, the terms "sentiment analysis" and "subjectivity analysis" come from Natural Language Processing (NLP) research [1]. Although these terms may be used for slightly different tasks or perspectives, they represent the same field of research [2] [3]. Broß defines sentiment analysis as a field of research that uses NLP techniques to automatically identify and analyze subjective information in natural language texts [4]. The goal is to determine which subjective information is expressed about which objects in the text [4]. Subjective information can manifest itself both explicitly through the expression of opinions or facts and explicitly in the attitude of the author [4]. In the literature, sentiment analysis is essentially examined on the following three levels:

a) Document: At this level, the task of sentiment analysis is to determine whether an entire document expresses a positive or negative sentiment [3]. This form is known as "document-level sentiment classification" and can be used, for example, to determine whether a product review as a whole expresses a positive or negative sentiment about the product [3]. Since several topics and thus different sentiments are possible for a document, the assignment of a single sentiment for the entire document is usually too imprecise.

b) Sentence: The task of sentiment analysis at sentence level is to recognize which sentences in a document express a positive, negative, or neutral sentiment [3]. In the product review example, sentiment analysis at sentence level allows a system to examine which sentences in the review contain product information and express a positive, negative, neutral or no sentiment information.

c) Entity and aspect: Sentiment analysis at the entity and aspect level (aspect-oriented sentiment analysis) is a fine-granular form of sentiment analysis. Unlike the other levels, it analyzes what aspects the author likes or dislikes [3]. The focus is not on documents, paragraphs or sentences as a whole, but on the opinion itself [3]. It is based on the assumption that an opinion consists of a sentiment and a goal to which the sentiment is expressed [3]. Targets of sentiment can be entities or their different aspects [3]. In the example of a product review, the reviewed product corresponds to an entity, while product properties represent aspects of the entity.

978-1-7281-1360-9/18/$31.00 ©2018 IEEE
DOI 10.1109/CSCI46756.2018.00059

III. Aspect-Oriented Sentiment Analysis: Automatic Generation of Sentiment and Aspect Dictionaries

Aspect-oriented sentiment analysis can be subdivided into the subgoals identification of aspect sentiment tuples, classification of sentiment and aggregation [5]. The first subgoal deals with the identification of aspects mentioned in a text related to an entity and the localization of the associated sentiment to form aspect sentiment tuples. The second subgoal is the classification of the selected tuples. Different binary or multi-value classification schemes can be used for this purpose. An example of a binary schema is the classification into one of the two categories ”positive” or ”negative”. In a multi-value scheme, the classes ”rather negative”, ”neutral” or ”rather positive” would also be conceivable. It is often necessary or desirable to aggregate the classified tuples by aspect. This means that a summarizing sentiment should be calculated for all the same aspects that have been identified. The methodology used for this depends on the application context in which the aspect-oriented sentiment analysis is carried out. The methods for identifying aspects distinguish between frequency-based methods, syntax-based methods, methods based on supervised and unsupervised machine learning [6], [7] and hybrid methods [8].

IV. Sentilyzer: System for aspect-oriented sentiment analysis of Amazon product reviews

Product reviews from the online retailer Amazon[1] are used as the data basis for generating the resources (sentiment and aspect lexicons) of the Sentilyzer sentiment analysis system. Amazon product reviews are suitable for the generation of resources due to the following characteristics: Amazon has a large variety of products and numerous reviews of the products offered. These reviews are rich in subjective information that can be analyzed. All products are organized in product categories which is well suited to capture domain-specific characteristics. The structure of the reviews is particularly interesting. In addition to the headline and review text, each review contains an overall numerical rating, information about the usefulness of the review, information about the rank of the author of the review, measured by the number of useful reviews written, and information about whether the purchase has been verified. Since the reviews are user-generated content, it is to be expected that informal language will be used. Both the syntax and the spelling of the texts can be wrong. In addition, emoticons, excessive use of punctuation, and certain spellings of words can be used to express an opinion.

The developed concept represents a system which essentially has two tools, the sentiment analysis tool and the lexicon generator. The sentiment analysis tool uses domain-specific sentiment and aspect lexicons, which are generated beforehand using the lexicon generator. Domain-specific here means that the dictionaries are intended for a specific product category.

[1]http://www.amazon.com

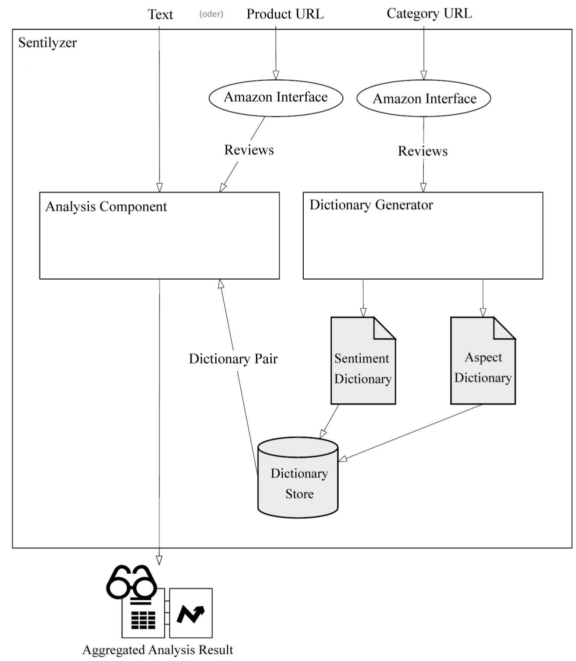

Fig. 1. Sentilyzer: Architectural Overview

The sentiment dictionary lists n-grams with their determined sentiment values. While the generation of dicionaries in these procedures is often manual or partially automated, the solution designed in the sentilyzer approach can generate all required resources automatically. The generation of the sentiment dictionary comprises two steps. In the first step, a tuple is generated for each adjective within the reviews. The adjective is assigned a numeric sentiment value corresponding to the number of stars in the review. If there is an adverb in front of the adjective at a distance of three words, it is added to the adjective as a modifier. The distance of three words defined here can also be changed in the system settings. By saving the modifier, the system is able to classify the expressions ”beautiful”, ”not beautiful” and ”very beautiful” separately from each other to deal with negations and other modifiers. The advantage of this approach is that the effect of a modifier on the sentiment value does not have to be defined manually. The result of the first step is a list of tuples with a sentiment expression and sentiment value. In the second step, the tuples are aggregated using the sentiment expression so that they are only contained once per sentiment dictionary. First, all uppercase letters are replaced by lowercase letters in order to be able to summarize sentiment expressions despite different uppercase and lowercase spelling. Then, for each sentiment expression, the number of times it occurs in positive reviews and the number of times it occurs in negative reviews are counted, i.e. how often the sentiment value 5 is assigned to the sentiment expression and how often the sentiment value 1 is assigned. These numbers can be used to calculate the aggregated sentiment value $S(w)$ of word w using the following equation: $S(w) = 4 * \frac{rP(w)}{rN(w)+rP(w)} + 1$ where $rP(w)$ is the relative frequency of how often w occurs in positive reviews and $rN(w)$ is the relative frequency of how often w occurs in negative reviews. The result of the aggregation is a list of unique sentiment expressions to which a numeric sentiment value between 1 and 5 is assigned. This list is referred to as the sentiment dictionary. Together with the transferred product category, the sentiment dictionary is permanently stored as a category-specific sentiment dictionary. In the analysis method, the aspect lexicon is then used to identify product aspects. Instead of using a lexicon, it would be possible to implement a frequency-based aspect identification method as a step in the analysis method. However, the following advantages are seen in the use of an aspect lexicon: (1) Repeated frequency-based calculation for aspect recognition is avoided. This reduces the runtime of the method. (2) An aspect dictionary is generated on the basis of many products of the same product category. It thus contains potential product aspects that may not be identified at the analysis runtime due to a lack of reviews from the individual product. In this way, the so-called cold start problem can be counteracted. To analyze user-defined texts, the user can select a sentiment dictionary whose domain corresponds to the topic of the text or, alternatively, not select a domain. In the latter case, a universal, domain-open sentiment dictionary is used, which is created by the system by combining all domain-specific sentiment dictionaries of a language. The aspects are then identified at runtime. Figure 3 gives an overview of the Sentilyzer interface. A user can read in reviews and receives a visualization of derived sentiments.

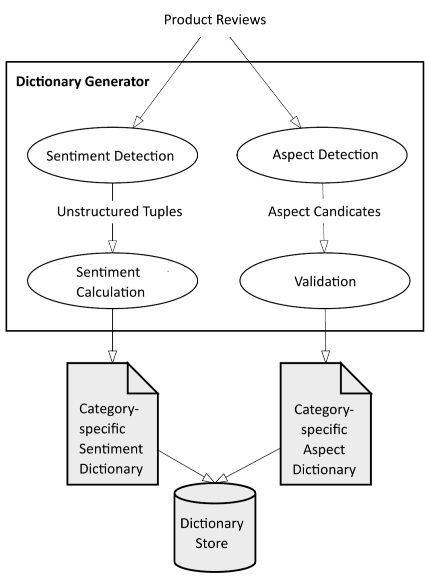

Fig. 2. Dictionary Generator: Generating Sentiment Dictionary and Aspect Dictionary from Product Reviews

V. Evaluation

The evaluation of the sentiment classification should show how precise sentiment values are classified by the Sentilyzer

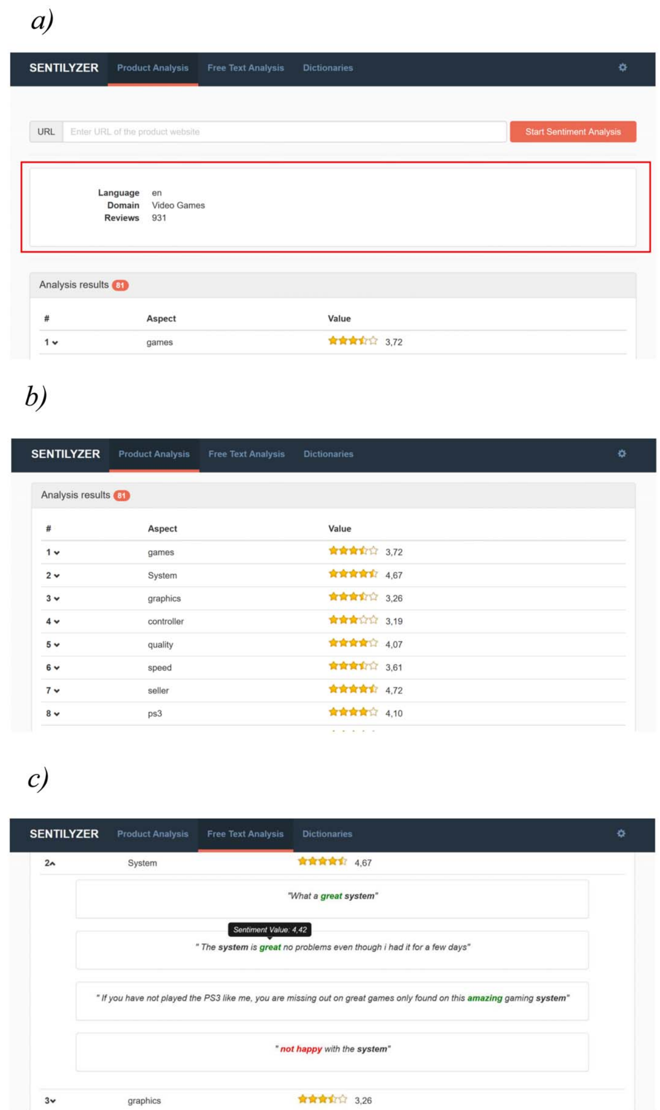

Fig. 3. Example for aspect oriented sentiment analysis with Sentilyzer: a) Analyzed domain ”video games” with 931 analysed reviews, b) Excerpt of recognized aspects, c) Preview of relevant, aspect related, identified text passages

system. Therefore, the identified sentiment classifications are evaluated using annotated data sets. A randomly selected test data set of 20 German and English annotated reviews has been analyzed. The English reviews are a subset of the annotated data set used in [9], [10]. The German dataset was created by random selection of reviews of a product and then annotating them and evaluate them using SentiWS [11], a German language resource for sentiment evaluation, as measurement for sentiment expressions.

The English reviews contain 47 positive sentiment terms and eight negative sentiment terms. From the positive sentiment terms, 37 were classified as positive (true positive, TP). The remaining ten positive sentiment terms were wrongly classified as negative (false negative, FN). Of the eight negative sentiment terms, seven were classified as such (true negative, TN), while one was classified as positive (false positive, FP). Within the German reviews there are 42 positive sentiment terms of which 34 were classified correctly (TP) and eight were classified incorrectly (FN). Furthermore, there are four negative sentiment terms in the reviews. Of these, three sentiment terms were correctly classified as negative (TN). One of the negative sentiment terms was wrongly classified as positive (FP). The following tables shows the results together with the calculated values of precision, recall, accuracy and F1 score. The achieved performance has been compared with other methods from the literature. Table II shows a comparison of the sentiment classification with other methods, while Table I shows a summary of the results.

TABLE I
EVALUATION RESULTS FROM SENTIMENT CLASSIFICATION ON THE BASIS OF ANNOTATED TEST DATA SETS

Language	**TP**	**FN**	**FP**	**TN**	**Precision**	**Recall**	**Accuracy**	**F1 score**
English	37	10	1	7	97,37%	78,72%	80,00%	87,06%
German	34	8	1	3	97,14%	80,95%	80,43%	88,31%

TABLE II
OVERVIEW OF SENTIMENT CLASSIFICATION APPROACHES

Approach	**Precision**	**Recall**	**Accuracy**	**F1 score**
Hu & Liu (2004) [10]	-	-	-	84,20 %
Popescu & Etzioni (2005) [12]	84,80 %	89,28 %	-	-
Kobayashi et al. (2006) [13]	82,20 %	66,20 %	-	-
Choi & Cardie (2008) [14]	-	-	90,70 %	-
Jin et al. (2009) [15]	-	-	-	70,59%-77,15%
Moghaddam & Ester (2011) [16]	-	-	84%- 86%	-
Yu et al. (2011) [17]	-	-	-	71,7%-85,1%
Mukherjee & Liu (2012) [18]	78,00 %	73,00 %	-	-
Sauper & Barzilay (2013) [19]	74,30 %	86,30 %	-	-
Sentilyzer (English) test sample	*97,37 %*	*78,72 %*	*80,00 %*	*87,06 %*
Sentilyzer (German) test sample	*97,14 %*	*80,95 %*	*80,43 %*	*88,31 %*

VI. CONCLUSION

One goal of this work was the design and implementation of a complete system addressing problems of aspect-oriented sentiment analysis using domain-specific lexicon pairs. We see a high potential in the use of exchangeable domain-specific lexicon pairs, as they represent domain-specific information on the one hand and bring the same flexibility as domain-open resources on the other hand. This flexibility is achieved due to the interchangeability and the possibility of automatic generation of new lexicons. We also see a high potential in the novel procedure for the classification of sentiment words within the framework of sentiment dictionary generation. The evaluation results show that this method for sentiment classification has a high precision in terms of lexical resources and performs moderately well considering recall and F-score. However, a larger scale evaluation on bigger example sets are needed to accomplish full comparability.

REFERENCES

[1] B. Pang, L. Lee *et al.*, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
[2] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE Transactions on Knowledge & Data Engineering*, no. 1, pp. 1–1, 2016.
[3] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.
[4] J. Broß, "Aspect-oriented sentiment analysis of customer reviews using distant supervision techniques," 2013.
[5] M. Tsytsarau and T. Palpanas, "Survey on mining subjective data on the web," *Data Mining and Knowledge Discovery*, vol. 24, no. 3, pp. 478–514, 2012.
[6] M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh, and O. Qawasmeh, "Enhancing aspect-based sentiment analysis of arabic hotels reviews using morphological, syntactic and semantic features," *Information Processing & Management*, 2018.
[7] K. Schouten, O. van der Weijde, F. Frasincar, and R. Dekker, "Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data," *IEEE transactions on cybernetics*, vol. 48, no. 4, pp. 1263–1275, 2018.
[8] M. Z. Asghar, F. M. Kundi, S. Ahmad, A. Khan, and F. Khan, "T-saf: Twitter sentiment analysis framework using a hybrid classification scheme," *Expert Systems*, vol. 35, no. 1, p. e12233, 2018.
[9] M. Hu and B. Liu, "Mining opinion features in customer reviews," in *AAAI*, vol. 4, no. 4, 2004, pp. 755–760.
[10] ——, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004, pp. 168–177.
[11] R. Remus, U. Quasthoff, and G. Heyer, "Sentiws-a publicly available german-language resource for sentiment analysis." in *LREC*, 2010.
[12] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Natural language processing and text mining*. Springer, 2007, pp. 9–28.
[13] N. Kobayashi, R. Iida, K. Inui, and Y. Matsumoto, "Opinion mining on the web by extracting subject-aspect-evaluation relations." in *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, 2006, pp. 86–91.
[14] Y. Choi and 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, 2008, pp. 793–801.
[15] W. Jin, H. H. Ho, and R. K. Srihari, "Opinionminer: a novel machine learning system for web opinion mining and extraction," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 1195–1204.
[16] S. Moghaddam and M. Ester, "Ilda: interdependent lda model for learning latent aspects and their ratings from online product reviews," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011, pp. 665–674.
[17] J. Yu, Z.-J. Zha, M. Wang, and T.-S. Chua, "Aspect ranking: identifying important product aspects from online consumer reviews," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011, pp. 1496–1505.
[18] A. Mukherjee and B. Liu, "Modeling review comments," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics, 2012, pp. 320–329.
[19] C. Sauper and R. Barzilay, "Automatic aggregation by joint modeling of aspects and values," *Journal of Artificial Intelligence Research*, vol. 46, pp. 89–127, 2013.