



Lexicon-Based Sentiment Analysis of Online Customer Ratings as a Quinary Classification Problem

Claudia Hösel^(✉), Christian Roschke, Rico Thomanek, and Marc Ritter

University of Applied Sciences Mittweida, 09648 Mittweida, Germany
hoesel@hs-mittweida.de

Abstract. Online customer reviews are not only an important decision-making tool for customers, they are also used by e-commerce providers as a source of information to analyze customer satisfaction. In order to reduce the complexity of evaluation comments, written reviews are additionally represented by evaluation stars in many evaluation systems. Numerous studies address the sentiment recognition of written reviews and view polarity recognition as a binary or ternary problem. This study presents the first results of a holistic approach, which takes up the combination of customer reviews with evaluation points realized in platform-dependent evaluation systems. Sentiment analysis is regarded as a quinary classification problem. In this study, 5,000 customer evaluations are analyzed with lexicon-based sentiment analysis at document level with the target to predict the evaluation points based on the determined polarity. For sentiment analysis the data mining tool RapidMiner is used and the categorization of the sentiment polarity is realized by using different NLP techniques in combination with the sentiment dictionary SentiWordNet. The supervised learning algorithms k-Nearest Neighbor, Naïve Bayes and Random Forest are used for classification and their classification quality is compared. Random Forest achieves the most accurate results in conjunction with NLP techniques, while the other two classifiers provide worse results. The results suggest that a stronger scaling of polarity requires a stronger differentiation between classes and thus a more intensive lexical preprocessing.

Keywords: Natural language processing · Sentiment analysis · Classification · Supervised learning methods

1 Introduction

In business-to-consumer e-commerce, customer evaluations represent an important source of information – both for customers and for e-commerce providers. At the customer level, the social proof of customer evaluations is an important decision-making aid for purchasing decisions [1]. Around two thirds of online shoppers read customer reviews before buying products in online shops [2]. For

e-commerce providers, customer reviews are an important component in the sale of products and services, as they allow conclusions to be drawn about customer satisfaction, among other things. E-commerce providers such as *Amazon.com, Inc.* [3], therefore operate – mostly platform-dependent – rating systems in order to extract moods and emotions from the reviews. An intellectual extraction of this information is hardly possible, especially for large e-commerce providers, due to the large number of reviews. Consequently, an automated evaluation is required, which is realized by means of opinion mining. Opinion mining, a field of text mining, deals with the automated extraction and evaluation of opinions from texts and uses various techniques for sentiment recognition [4]. One opinion mining technique is the sentiment analysis. Sentiment analysis uses various natural language processing (NLP) methods, such as tokenization or stemming, to analyze the sentiment of a text or the emotional attitude to an object contained in the text [5]. A fundamental problem of sentiment analysis is the categorization of sentiment polarity. Natural-language texts can contain valence shifts – for example through negation or intensification – which are usually easily understood by humans, but not by computational systems [6]. Recent research attempts to address this problem through lexicon-based preprocessing, mostly using existing sentiment dictionaries.

Prakoso et al. 2018 explicitly investigate in their study the effects of lexicon-based preprocessing on the accuracy of sentiment classification when using supervised machine learning algorithms and note that sentiment analysis with lexicon-based preprocessing achieves higher accuracy in all classification models [7]. Existing approaches from the domain of lexicon-based sentiment analysis, such as Alkalbani et al. 2017 [8], Fang and Zahn 2015 [9], Lin et al. 2018 [10], regard sentiment recognition as a binary or ternary classification problem. However, a stronger scaling of the polarity seems to make sense especially for the sentiment recognition of customer ratings, since existing – especially platform-dependent – rating systems mostly combine metrics such as written review and integer rating system. Research approaches that scale the polarity more strongly usually refer to selected social media channels and can only be transferred to other fields of application to a limited extent due to their specific functionalities. El Alaoui et al. 2018 present a lexicon-based approach for the sentiment analysis of tweets, which distinguishes seven polarity classes and uses the specific functionalities of the microblogging service, such as re-tweets or likes, in addition to a sentiment lexicon, when determining polarity [11].

Based on this approach, the question arises to what extent the specific functionalities of evaluation platforms can be used for the sentiment determination of customer evaluations. The first step is to determine to what extent the points or stars awarded by customers via rating systems reflect the opinions expressed in the written evaluations. The present study takes up this problem and presents a first state of work. Starting from the assumption that platform-dependent rating systems from e-commerce providers combine rating comments with a five-level star-scaled rating system, sentiment recognition is regarded as a quinary classification problem.

2 Methods

This study uses a data set from “Kaggle” [12] with around 400,000 customer ratings of unlocked mobile phones. Each of these devices was reviewed on *Amazon.com, Inc.* [3] and also rated by customers there. The data set is adjusted for unneeded attributes as well as missing values and balanced with regard to the attribute “Rating”. This attribute is based on an integer star-scaled system where the highest rating can be five stars and the lowest rating one star. After cleansing and normalizing the data set, each rating level is represented by the same amount of elements. For resource-related reasons, 1,000 elements are used per rating level; the data set used in this study therefore contains 5,000 elements.

The methodology applied is divided into four process steps. As shown in Fig. 1, various NLP techniques are used during data preprocessing to clean up the text, structure it and convert it into a machine-readable form. The tokens of the document are then used to generate a vector that represents the document numerically and thus makes it usable for mathematical operations. The weighting of the terms is done by the combined method “Term Frequency – Inverted Document Frequency” (TF-IDF). This method takes into account the frequency distribution of terms in the corpus and weights terms on the basis of frequency and differentiations [13]. In the next step, the features extracted from the texts are used to predict the sentiment. For classification the machine supervised models k-Nearest Neighbor (k-NN), Naïve Bayes and Random Forest will be implemented, evaluated by a tenfold leave-one-out cross validation and the quality of classification will be compared between the models.

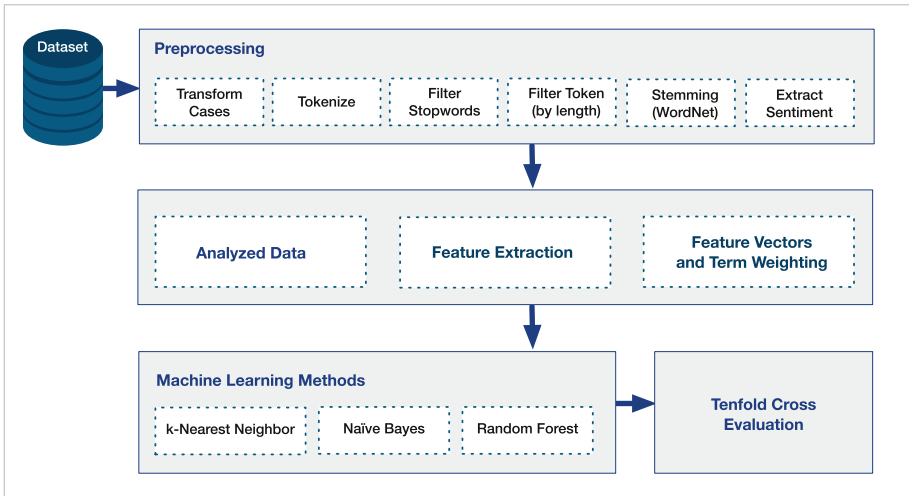


Fig. 1. Flow of process methods

The entire process was realized with the data mining tool RapidMiner 9.1 [14]. For polarity detection, the extension SentiWordNet 3.0 [15] is implemented in RapidMiner. SentiWordNet is an open source lexical resource developed for opinion mining applications [16]. The mood-bearing expressions in the text are identified and coded in relational scale.

3 Results and Discussion

In this study, a sentiment analysis with lexicon-based preprocessing of online customer ratings is carried out with the aim of predicting the integer star-scaled ratings provided by the customers based on the written reviews. For the quinary classification problem the classifiers k-Nearest Neighbor, Naïve Bayes and Random Forest are used and their accuracy is compared. Figure 2 shows the results of the classifiers.

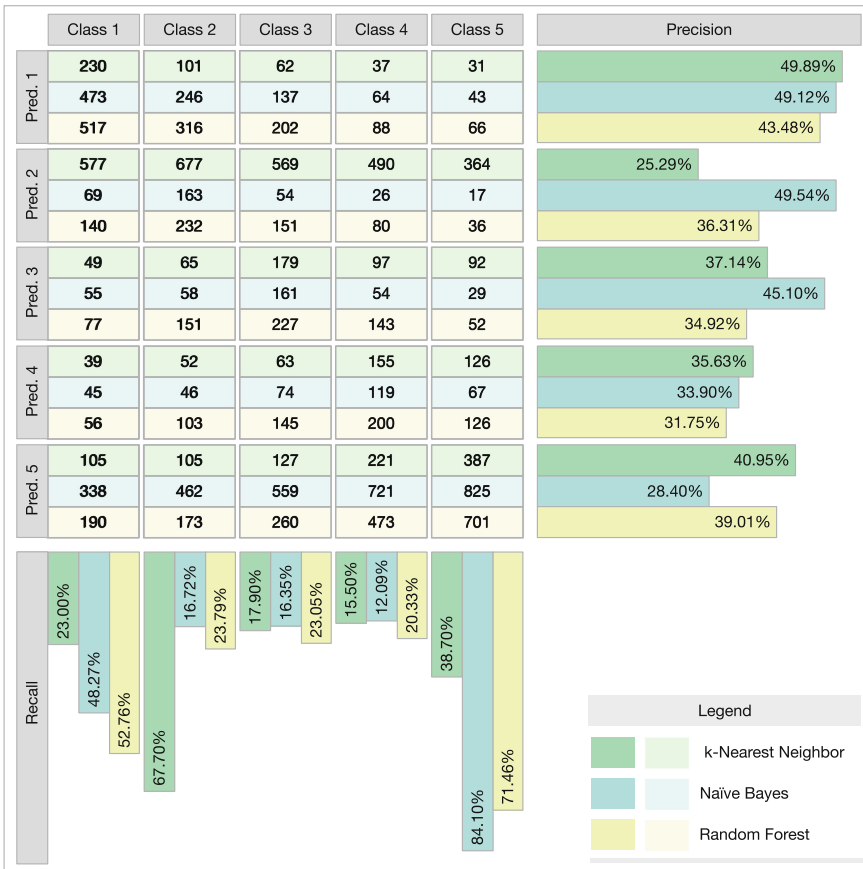


Fig. 2. Results of the classifiers k-Nearest Neighbor, Naïve Bayes and Random Forest

The classifier *k-Nearest Neighbor* classifies customer ratings by sentiment with an accuracy of 32.56%. The highest precision of 49.89% is achieved by k-NN in class 1, but the probability that a customer rating actually belonging to class 1 will be recognized and correctly classified is only 23.00%. The lowest precision of 25.29% is achieved in class 2. However, the corresponding recall value shows that a customer rating belonging to this class is recognized and correctly classified by k-NN with a probability of 67.70%.

Naïve Bayes achieves an accuracy of 35.49% on the quinary classification problem. The highest precision with 49.54% is achieved by the classifier in class 2. Nevertheless, the probability that a customer rating actually belonging to class 2 is recognized and correctly classified is only 16.72%. The lowest precision with 28.40% is class 5. However, the corresponding recall value of 84.10% shows that a customer rating belonging to this class is highly likely to be recognized and correctly classified.

The *Random Forest* classifier achieves a total accuracy of 38.27%. The highest precision of 43.48% is achieved in class 1. In addition, a customer rating belonging to this class is recognized with a probability of 52.76% and classified correctly. The lowest precision of 31.75% is achieved by the classifier in class 4. The associated recall value of 20.33% shows that a customer rating belonging to this class is recognized and correctly classified with a relatively low probability.

The low precision and recall values achieved by the classifications indicate that the individual classes could not be clearly distinguished from each other. The reason for this could be the small amount of training data used. Due to technical limitations, the data set was limited to 5,000 elements. Each class contained 1,000 elements and thus a relatively small number for training. In addition, the lexicon-based preprocessing was carried out with a cross-domain sentiment dictionary, which could have led to the fact that the sentiment of domain-dependent terms was not correctly recorded and classified.

4 Conclusion

The aim of this study was to make initial statements on the extent to which the points or stars awarded by customers via rating systems reflect the opinions expressed in the written reviews. The sentiment recognition of online customer ratings was considered a quinary classification problem. In order to gain initial insights, a lexicon-based sentiment analysis was combined with the machine learning algorithms k-Nearest Neighbor, Naïve Bayes and Random Forest. The results of the classifiers were evaluated with tenfold cross-validation and then compared. Random Forest achieved the highest accuracy with 38.27%, followed by Naïve Bayes with 35.49%. Although k-Nearest Neighbor delivered the lowest overall accuracy of 32.56%, it achieved the best predictive accuracy in three out of five classes. Naïve Bayes achieved the highest accuracy in two out of five classes. Due to the limitations described in the previous chapter, the focus of the continuation of this study will be on the delimitation of the individual classes. A first step could be to adapt the sentiment dictionary to the specific domain. The

sentiment of words or subsets can vary depending on the context. Words that are positive in one domain (e.g. the horror movie was scary) may be negative in another domain. The use of a domain-specific dictionary therefore seems to be useful for the differentiation of the individual classes. In the present study, a sentiment analysis was carried out at document level. In future research, opinion mining techniques will also be applied at sentence and aspect level in order to obtain more precise results.

References

1. Plottek, K., Herold, C.: Micro moments als entscheidender moment im rahmen einer zunehmend fragmentierteren customer journey. In: Rusnjak, A., Schallmo, D.R.A. (eds.) *Customer Experience im Zeitalter des Kunden*, pp. 143–176. Springer, Wiesbaden (2018). https://doi.org/10.1007/978-3-658-18961-7_5
2. Bitkom: Shopping digital – Wie die Digitalisierung den Handel tiefgreifend verändert. Technical report (2017). <https://www.bitkom.org/sites/default/files/file/import/171124-Studienbericht-Handel-Web.pdf>
3. Amazon. <https://www.amazon.com>
4. Scholz, T.: Opinion mining für verschiedene Webinhalte. In: Scherfer, K., Volpers, H. (eds.) *Methoden der Webwissenschaft*, pp. 63–81. Schriftenreihe Webwissenschaft, Lit, Berlin (2013)
5. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng. J.* **5**(4), 1093–1113 (2014)
6. Ziegler, C.N.: Automated capture of strategic knowledge on the web, p. 160. Albert-Ludwigs-Universität Freiburg i. Br., Freiburg i. Br. (2010). <http://www2.informatik.uni-freiburg.de/~ziegler/papers/Habil-Thesis.pdf>
7. Aryo Prakoso, A., Winantesa Yananta, B., Fitra Setyawan, A., Muljono: A lexicon-based sentiment analysis for Amazon web review. In: 2018 International Seminar on Application for Technology of Information and Communication, pp. 503–508. IEEE (2018)
8. Alkalbani, A.M., Gadhvi, L., Patel, B., Hussain, F.K., Ghamry, A.M., Hussain, O.K.: Analysing cloud services reviews using opining mining. In: IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), pp. 1124–1129. IEEE (2017)
9. Fang, X., Zhan, J.: Sentiment analysis using product review data. *J. Big Data* **2**(1), 1–14 (2015)
10. Lin, B., Zampetti, F., Bavota, G., Di Penta, M., Lanza, M., Oliveto, R.: Sentiment analysis for software engineering: how far can we go? In: The 40th International Conference, pp. 94–104. ACM Press (2018)
11. El Alaoui, I., Gahi, Y., Messoussi, R., Chaabi, Y., Todoskoff, A., Kobi, A.: A novel adaptable approach for sentiment analysis on big social data. *J. Big Data* **5**(1), 1–18 (2018)
12. Kaggle. <https://t1p.de/rk88>
13. Baeza-Yates, R., Ribeiro-Neto, B.: *Modern Information Retrieval*, pp. 29–30. Addison Wesley (1999)
14. Rapidminer. <https://rapidminer.com>
15. SentiWordNet. <http://sentiwordnet.isti.cnr.it/>
16. Pang, B., Lee, L.: Briefly noted. *Comput. Linguist.* **35**(2), 311–312 (2009)