Star ratings as a predictor for sentiment polarity in online reviews - a case study

Adriana-Mihaela Guran
Babeş-Bolyai Univeristy
1, M. Kogălniceanu, Cluj-Napoca
adriana.guran@ubbcluj.ro

Grigoreta-Sofia Cojocar
Babeş-Bolyai Univeristy
1, M. Kogălniceanu, Cluj-Napoca
Grigoreta.cojocar@ubbcluj.ro

Dan Cojocar
Babeş-Bolyai Univeristy
1, M. Kogălniceanu, Cluj-Napoca
dan.cojocar@ubbcluj.ro

ABSTRACT

Nowadays, the web is abundant in information related to product and services. People are using online reviews to support their decisions. Sometimes, consumers focus their attention on the star rating of a product, which is a quantitative assessment of the product, while the text of the review provides qualitative information, but more time consuming to read and process. Still, there are few studies verifying the consistency between the star rating of a product or service and the sentiment of the review, and no study for Romanian reviews. In this paper we discuss an approach on verifying the consistency between star ratings and sentiment polarity for a set of Romanian reviews gathered from an e-commerce site. We measure the consistency between the two series and the results show a good consistency for our data.

Author Keywords
sentiment analysis, star ratings, customer reviews

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., USI): Miscellaneous.

General Terms
Human Factors; Design; Measurement.

DOI: 10.37789/icusi.2025.22

INTRODUCTION

Online reviews are one of the most important sources of opinions on products and services for businesses and consumers. An online review is opinion information on a target product or service shared by a consumer on an online platform [19]. The target entity of reviews is a product or service. Electronic word of mouth (eWOM) is abundant in the web today and influences decision making of consumers in multiple domains. There are two sources of information for the consumers: the star ratings of the products and the review itself. Each of these sources provides different insight for the consumers: star ratings provide a quantitative indicator of a product or service, while the textual review may provide details on various aspects of it, supporting the decision-making. The star rating is a 1–5 scale numeric score assigned by consumers to a target product or service to express their opinion when writing a review. This feature is supported by most online review platforms. Online reviews can hold both positive and negative aspects, such that a deep analysis of the textual content of the review is needed to support the decision making. Sentiment analysis (SA) refers to text analytics techniques that try to automatically find the polarity of the text [15]. It is a field of study that analyses people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. The main motivation is to enable informed decisions and an understanding of the dynamics of discourse by businesses and researchers. Various approaches exist for understanding the tone and polarity of reviews and then correlating it with the star rating of a product regarding the English resources, described in [1], [2], [11], [12]. However, for the Romanian language, there is no study addressing the correlation between the star rating and the polarity of a product review.

RELATED WORK

Al-Natour & Turetken [1] have studied the alignment between the star ratings and textual reviews in the tourism domain using deep learning and natural language processing. They have used machine learning, lexicon based and hybrid techniques to perform SA. The conclusion of their work shows a correlation between the results of SA and the star ratings. The comparison was made for the English language. Approaches to correlating star ratings and SA for other languages are presented in [10] and [18] and [22] for Slovak language, in [5] for Russian language, in [8] for the Czech language, in [21] for Croatian, and in [9] for Lithuanian. For the Romanian language there are few attempts of applying SA techniques. In [4] authors propose a semi-supervised learning algorithm for SA in Romanian language, while in [20] the authors test the most popular classification methods applied to SA, including standard machine learning, deep learning and Bidirectional Encoder Representations from Transformers (BERT). Authors of [6] build a cross-domain dataset and after comparing multiple machine learning classification techniques, concludes the fitness of BERT models for the SA task. Authors of [23] have studied the correlation between star reviews and sentiment analysis from three nations (Japan, China and U.S.A.) related to restaurant reviews and found inconsistency of positive sentiment correlation with 5- and 4-star ratings generated by Japanese consumers especially. SA is influenced by the domain of the reviews, because terms that are positive in one domain can be negative in another one [14, 15]. Recently, Briciu et al. have compared multiple machine learning algorithms on SA at various level (document, aspect, sentence) on a Romanian product reviews dataset [3]. To our knowledge there is no research about the alignment between star ratings polarity and SA results for the Romanian language. Our work comes to fill this gap.

METHOD

The goal of the present research is to show if it is possible to derive the sentiment polarity of Romanian reviews only based on the star evaluation of the product, without analysing the text of the review itself or its title.

RQ 1 Are the star ratings, and the sentiment polarity of a Romanian product review correlated (aligned)?

Sample

To perform our study, we have built a dataset of reviews collected from an e-commerce website from Romania. 5829 reviews have been collected with their link, author name, date, title of review and number of stars. For each review category the dataset was balanced related to the number of stars.

Categories

The collected reviews are related to electronic products from the categories: vacuum cleaners, speakers, fitness bands, headphones, monitors, laptops, mouses, routers, smartwatches, tablets, keyboards and Apple mobile phones.

Procedure

Five reviewers, master students in Computer Science, have annotated the reviews. The reviews have been annotated with a sentiment from the following six categories: *Positive, Very Positive, Neutral, Negative, Very negative* and *I don't know* based on the content of the review. The annotators have performed the annotation individually. Each annotator has analysed all the items from the dataset. They had access only to the content of the review, without seeing the number of stars assigned to the product. To check the reviewers' agreement, we have computed the Fleiss' Kappa inter-agreement coefficient [7] on the entire dataset, and we have obtained a value of 0.5468, showing a moderate agreement [13].

At the end of the annotation process we merged the annotations performed for each product. For each item in the dataset a label has been assigned with the most frequent label from the set of the six mentioned above. If there was no such a majority chosen label, then the review has been furtherly processed manually. Of the 5829 items in the dataset, for 204 human analysis was needed, as there was no agreement between the annotators.

For the rest of the 5625 items, we have converted the labels into numbers, as follows: Very Good-5, Good-4, Neutral-3, Negative-2, Very Negative-1.

Results

Of the 204 reviews manually analysed, 34 reviews had an *equal number of positive and negative labels* (2 labels) and the gap between analysts' reviews can be explained by the following situations:

A. reviews that contain both positive and negative aspects (see Table 1) .

Table 1 – Example of review with both positive and negative aspects

Review in Romanian	Review in English
Nu se aud f bine, dar la prețul dat merita banii	*They don't sound very good, but at the price they are worth the money*

B. the review contains initially a positive review, followed by an *Update* or *Later Edit* (LE) which presents negative aspects (see Table 2)

Table 2 – Example of review with update

Review in Romanian	Review in English
Super, sunt multumită, sunet perfect, le recomand *UPDATE ❗ 15 Sept 2022 ❗* *Comandate in 6 IUNIE,* *de AZI* *15 SEPT,casca stanga SE Aude în sudină,(f. încet)în Timp ce dreapta,MERGE NORMAL(sunet tare).* *(Au 90%baterie).* *SE pare ca sunetul a revenit,si in STÂNGA,cand eram pe cale sa cumpär altele.Sper sa nu Mai pätesc asa .* 🤔	*Great, I'm satisfied, perfect sound, I recommend them* *UPDATE 15 ❗ Sept 2022 ❗* *Ordered on JUNE 6,* *TODAY* *15 SEPTEMBER, the left earbud is heard quietly, (v. slowly) while the right, FUNCTIONS NORMALLY.* *(They have 90% battery).* *It seems that the sound came back, and on the LEFT, when I was about to buy others. I hope it doesn't happen like this again.*

C. A negative review and a Later Edit (LE) which contains positive actions taken by the sales agents for that product (see Table 3).

Table 3 – Example of a negative review with LE

Review in Romanian	Review in English
Initial au mers super bine dar dupa vreo 2-3 luni o casca a incetat sa mai functioneze. Later edit: in urma acestui review, vanzatorul mi-a trimis gratuit inca o pereche de casti care, momentan, se comporta bine. Sa speram ca astea n-or sa	*Initially they worked super well, but after about 2-3 months stopped working.* *Later edit: following this review, the seller sent me another pair of headphones for free which, at the moment, behaves well. Let's hope these don't break anymore.*

se mai strice.	

ok	*otherwise it's ok)*

In the rest of 170 reviews, we have identified the following categories:

D. Reviews having an equal number of *positive* and *very positive* labels (see Table 4)

Table 4 – Example of a review with positive and very positive labels

Review in Romanian	Review in English
Este un produs superb. Este foarte util (cu multe funcții) și poți dormi liniștit cu el la încheietură deoarece nu ma deranjează deloc. Își merită toți banii.	*It's a superb product. It is very useful (with many functions) and I can sleep peacefully with it on my wrist because it does not bother me at all. It's worth all the money.*

E. Reviews having an equal number of *negative* and *very negative l*abels (see Table 5)

Table 5 – Example of a review with negative and very negative labels

Review in Romanian	Review in English
O copie slaba dupa Xiaomi Mi Band 5, cu o aplicatie foarte prost optimizata. Voi prefera oricand Xiaomi in fata Amazfitului, chiar daca sunt produse sub aceeasi licenta.	*A poor copy of the Xiaomi Mi Band 5, with a very poorly optimized application. I will always prefer Xiaomi over Amazfit, even if they are produced under the same license.*

In these cases, the polarity can be considered as positive or negative, only that the annotators have labelled the review differently.

Another category of reviews had an equal number of *positive* and *neutral* labels, or an equal number of *negative* and *neutral* labels. This kind of reviews are usually containing a list of positive aspects of the product, followed by a "but" ("dar" in Romanian) and some negative aspects (see Table 6 and Table 7).

Table 6 – Example of a positive review followed by "but"

Review in Romanian	Review in English
Imi place, dar ma asteptam la mai mult	*I like it, but I expected more*

Table 7– Example of a negative review followed by "but" and positive aspects

Review in Romanian	Review in English
Se mai deconectează cand și când de la aplicație dar în rest e	*It disconnects now and then from the application, but*

Next, we investigated the correlation between the number of stars and the sentiment labels for the reviews in the dataset containing a most frequent label. We have computed the Cohen's kappa coefficient [17] for the two series (the star ratings and the sentiment label) and we have obtained a value of **0.832**, indicating a good correlation. As such, we can answer our research question, by saying that there is a strong correlation between the star ratings and the polarity of the sentiment in our dataset for the reviews that had a clear sentiment label (the most frequent one). Still, there is a need to further analyse the reviews on which an agreement between annotators has not been possible.

THREATS TO VALIDITY

Our reported results are based on an empirical study, thus being subject to certain threats to validity.

Internal validity refers to factors that could have influenced the obtained results. The training of the annotators – annotators involved in the annotation process have no previous experience with labelling process and sentiment analysis, Moreover, they have no language-oriented studies, being computer science master students. As such, their technical knowledge may influence the way they interpreted the products' reviews.

External validity concerns the generalization of our findings. The size of the data – the number of reviews that we have gathered is not so big, thus it might influence the quality of our research and the generalizability of our results. The specific categories of products – the reviews that we have analysed address the category of electronic devices, and the words and expressions used in these reviews belong to a specific category of the language. Choosing another product category may change the obtained results. Cultural influences: Romanian culture has a far higher uncertainty avoidance index [16]. An important feature of a high uncertainty avoidance culture is avoiding confrontation, which could even lead to no action on unsatisfactory products for Romanian customers.

COMCLUSIONS AND FUTURE WORK

In this paper we have presented the results of initial research on the correlation between star ratings and the sentiment of the textual description of an online review. In the future we intend to experiment more on using automatic approaches to SA and to include aspects related to fake reviews, people with inconsistent ratings, misclicks, which may negatively influence consumers' choices.

REFERENCES

1.Al-Natour, S., Turetken, O., A comparative assessment of sentiment analysis and star ratings for consumer reviews, *International Journal of Information Management*, Volume 54, 2020, 102132, ISSN 0268-4012, https://doi.org/10.1016/j.ijinfomgt.2020.102132.

2.Bigne, E., Ruiz, C., Perez-Cabañero, C. et al. Are customer star ratings and sentiments aligned? A deep learning study of the customer service experience in tourism destinations. *Serv Bus* 17, 281–314 (2023). https://doi.org/10.1007/s11628-023-00524-0

3.Briciu, A., Călin, A.-D., Miholca, D.-L., Moroz-Dubenco, C., Petrașcu, V., Dascălu, G. Machine-Learning-Based Approaches for Multi-Level Sentiment Analysis of Romanian Reviews. *Mathematics* 2024, 12, 456. https://doi.org/10.3390/math12030456

4.Buzea, M. C., Trăușan-Matu, Ș. and Rebedea, T., *A Three Word-Level Approach Used in Machine Learning for Romanian Sentiment Analysis*, 2019 18th RoEduNet Conference: Networking in Education and Research (RoEduNet), Galati, Romania, 2019, pp. 1-6, doi: 10.1109/ROEDUNET.2019.8909458

5.Chetviorkin, I., Loukachevitch, N., *Evaluating sentiment analysis systems in Russian.* In Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing, Sofia, Bulgaria, 8–9 August 2013;

6.Cioban, S., *Cross-Domain Cross-Domain Sentiment Analysis of the Natural Romanian Language*, In: Jallouli, R., Bach Tobji, M.A., Mcheick, H., Piho, G. (eds.) Digital Economy. Emerging Technologies and Business Innovation. ICDEc 2021. Lecture Notes in Business Information Processing, vol 431. Springer, Cham. https://doi.org/10.1007/978-3-030-92909-1_11

7.Fleiss, J. L., Measuring nominal scale agreement among many raters, *Psychological Bulletin*, 76 (5): 378–382, doi:10.1037/h0031619, 1971.

8.Habernal, I., Ptácek, T., Steinberger, J., Supervised sentiment analysis in Czech social media. *Inf. Process. Manag.* 2014, 50, 693–707.

9.Kapočiūtė-Dzikienė, J., Krupavičius, A., Krilavičius, T. *A comparison of approaches for sentiment classification on lithuanian ˙ internet comments*. In Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing, Sofia, Bulgaria, 8–9 August 2013; pp. 2–11.

10.Krchnavy, R., Simko, M., *Sentiment analysis of social network posts in Slovak language*. In Proceedings of the 2017 12th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), Bratislava, Slovakia, 9–10 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 20–25.

11.Kumar, N. & Hanji, B. Combined sentiment score and star rating analysis of travel destination prediction based on user preference using morphological linear neural network model with correlated topic modelling approach. *Multimedia Tools and Applications*. 1-32. 10.1007/s11042-023-17995-y, 2024.

12.Lai, X., Wang, F. and Wang, X. , Asymmetric relationship between customer sentiment and online hotel ratings: the moderating effects of review characteristics, *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 6, pp. 2137-2156. https://doi.org/10.1108/IJCHM-07-2020-0708

13.Landis, J. R., Koch, G. G, The measurement of observer agreement for categorical data, *Biometrics*, 33 (1): 159–174, 1977, doi:10.2307/2529310, JSTOR 2529310, PMID 843571.

14.Liu, B., *Sentiment Analysis and Subjectivity*. In: Indurkhya, N. and Damerauthe, F.J., Eds., Handbook of Natural Language Processing, 2nd Edition, Chapman and Hall/CRC, London, 1-38, 2010.

15.Liu, B., Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.

16.Marinescu, G., Uncertainty avoidance in Romanian organizational culture. *Journal of Business and Retail Management Research.* 8. 30-41, 2014

17.McHugh M. L, Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276–282, 2012.

18.Mojžiš, J., Krammer, P., Kvassay, M., Skovajsová, L., Hluchy, L., Towards Reliable Baselines for Document-Level Sentiment Analysis in the Czech and Slovak Languages. *Future Internet* 2022, 14, 300.

19.Mudambi, S. M. and Schuff, D., Research note: What makes a helpful online review? A study of customer reviews on Amazon. com, MIS Q., vol. 34, no. 1, pp. 185–200, 2010.

20.Neagu, D. C., Rus, A. B., Grec, M., Boroian, M. A., Bogdan, N., Gal, A., Towards Sentiment Analysis for Romanian Twitter Content. *Algorithms*. 2022; 15(10):357.

21.Rotim, L., Šnajder, J., *Comparison of short-text sentiment analysis methods for croatian*. In Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing, Valencia, Spain, 4 April 2017; pp. 69–75.

22.Sokolová, Z., Harahus, M., Juhár, J., Pleva, M., Staš, J., Hládek, D. Comparison of Machine Learning Approaches for Sentiment Analysis in Slovak. *Electronics* 2024, 13, 703. https://doi.org/10.3390/ electronics13040703 Academic Editors: Ark

23.Wan, Y. & Nakayama, M., *A Sentiment Analysis of Star-rating: a Cross-Cultural Perspective*. Proceedings of the 55th Hawaii International Conference on System Sciences, 2022,10.24251/HICSS.2022