Accepted Manuscript

Ranking online consumer reviews

Sunil Saumya, Jyoti Prakash Singh, Abdullah Mohammed Baabdullah, Nripendra P. Rana, Yogesh K. Dwivedi

PII: S1567-4223(18)30035-8
DOI: https://doi.org/10.1016/j.elerap.2018.03.008
Reference: ELERAP 783

To appear in: *Electronic Commerce Research and Applications*

Received Date: 11 January 2018
Revised Date: 28 March 2018
Accepted Date: 28 March 2018

Please cite this article as: S. Saumya, J.P. Singh, A.M. Baabdullah, N.P. Rana, Y.K. Dwivedi, Ranking online consumer reviews, *Electronic Commerce Research and Applications* (2018), doi: https://doi.org/10.1016/j.elerap.2018.03.008

RANKING ONLINE CONSUMER REVIEWS

Sunil Saumya
National Institute of Technology Patna, Bihar, India
Sunils.cse15@nitp.ac.in

Jyoti Prakash Singh
National Institute of Technology Patna, Bihar, India
jps@nitp.ac.in

Abdullah Mohammed Baabdullah
King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia
baabdullah@kau.edu.sa

Nripendra P. Rana
Swansea University Bay Campus, Swansea, UK
nrananp@gmail.com

Yogesh k. Dwivedi
Swansea University Bay Campus, Swansea, UK
ykdwivedi@gmail.com

Last revised: March 29, 2018

ABSTRACT

Product reviews are posted online by the hundreds and thousands for popular products. Handling such a large volume of continuously generated online content is a challenging task for buyers, sellers and researchers. The purpose of this study is to rank the overwhelming number of reviews using their predicted helpfulness scores. The helpfulness score is predicted using features extracted from review text, product description, and customer question-answer data of a product using the random-forest classifier and gradient boosting regressor. The system classifies reviews into low or high quality with the random-forest classifier. The helpfulness scores of the high-quality reviews are only predicted using the gradient boosting regressor. The helpfulness scores of the low-quality reviews are not calculated because they are never going to be in the top *k* reviews. They are just added at the end of the review list to the review-listing website. The proposed system provides fair review placement on review listing pages and makes all high-quality reviews visible to customers on the top. The experimental results on data from two popular Indian e-commerce websites validate our claim, as 3-4 newer high-quality reviews are placed in the top ten reviews along with 5-6 older reviews based on review helpfulness. Our findings indicate that inclusion of features from product description data and customer question-answer data improves the prediction accuracy of the helpfulness score.

Keywords: Big data challenge; e-commerce; helpfulness; machine learning; online reviews

---

Acknowledgments. The author would like to acknowledge the Ministry of Electronics and Information Technology (MeitY), Government of India for the financial support during research work through "Visvesvaraya PhD Scheme for Electronics and IT".