

Accepted Manuscript

Sentiment labeling for extending initial labeled data to improve semi-supervised sentiment classification

Sangheon Lee, Wooju Kim

PII: S1567-4223(17)30077-7

DOI: <http://dx.doi.org/10.1016/j.elerap.2017.09.006>

Reference: ELERAP 729

To appear in: *Electronic Commerce Research and Applications*

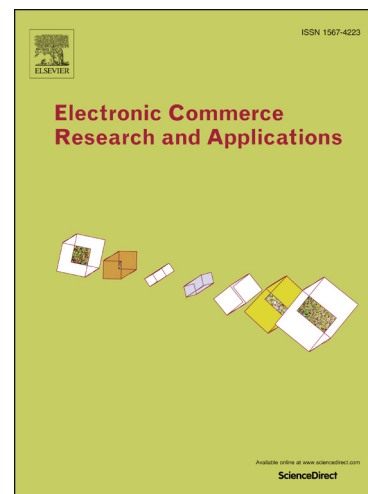
Received Date: 15 December 2016

Revised Date: 19 September 2017

Accepted Date: 19 September 2017

Please cite this article as: S. Lee, W. Kim, Sentiment labeling for extending initial labeled data to improve semi-supervised sentiment classification, *Electronic Commerce Research and Applications* (2017), doi: <http://dx.doi.org/10.1016/j.elerap.2017.09.006>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



**SENTIMENT LABELING FOR EXTENDING INITIAL LABELED DATA
TO IMPROVE SEMI-SUPERVISED SENTIMENT CLASSIFICATION****Sangheon Lee****Wooju Kim (corresponding author)**

Dept. of Information and Industrial Engineering
Yonsei University
50 Yonsei-ro, Seodaemun-gu, Seoul, Korea

Last revised: August 29, 2017

ABSTRACT

In recent decades, analyzing the sentiments in online customer reviews has become important to many businesses and researchers. However, insufficient amount of labeled training corpus is a bottleneck for machine learning approaches. Self-training is one of the promising semi-supervised techniques which does not require large amounts of labeled data. However, self-training also suffers from an incorrect labeling problem along with insufficient amount of labeled data. This study proposed a semi-supervised learning framework that adds only confidently predicted data to the training corpus in order to enrich the initial classifier in self-training. The experimental results indicate that the proposed method performed better than self-training.

Keywords: Concatenated vector, paragraph vector, self-training, sentiment classification, sentiment labeling, topic model

Acknowledgments. This work (Grant No. C0258734) was supported by the Business for Academic-industrial Cooperative Establishments funded by the Korea Small and Medium Business Administration in 2016. This work also was financially supported by the Korea Ministry of Land, Infrastructure and Transport (MOLIT) via the U-City Master and Doctor Course Grant Program.

Download English Version:

<https://daneshyari.com/en/article/4942482>

Download Persian Version:

<https://daneshyari.com/article/4942482>

[Daneshyari.com](https://daneshyari.com)