

[DOI: 10.20472/IAC.2019.048.020](https://doi.org/10.20472/IAC.2019.048.020)

CHIH LI HUNG

Chung Yuan Christian University, Taiwan

JHENG-HUA HUANG

Chung Yuan Christian University, Taiwan

DEVELOPING AN ADAPTIVE CHINESE NEAR-SYNONYM CORPUS FOR WORD OF MOUTH CLASSIFICATION

Abstract:

Word of mouth (WOM) is the subjective opinion of consumers for a brand, a product or a service. Its impact on consumer's purchasing decision is greater than the marketing activities of a product. Word of mouth classification is an effective means for document organization in an era of big data. However, the existing tasks of WOM classification are mainly dependent on the bag of word (BOW) in the vector space model (VSM), which usually suffers from the curse of dimensionality while dealing with large amounts of documents. We compared characters, context, and homophones, and integrated thesauruses to establish an adaptable Chinese near-synonym corpus. Subsequently, lexical replacement was applied, and the adaptable Chinese near-synonym corpus was created for classifying WOM documents. Two static corpora, the Ministry of Education's Revised Mandarin Chinese Dictionary and the Extended Chinese Synonym Forest, were used as the benchmarks of comparison for the proposed adaptable near-synonym corpus in the classification and evaluation stage. Evaluations were conducted by calculating recall, precision, F-measure, accuracy, and area under receiver operating characteristic curves (AUC). The results indicate that the classification accuracy of the adaptable near-synonym corpus proposed in the research exceeds that of static corpora when used in the fields of movie, leisure and travel, food, and cosmetics.

Keywords:

Near-Synonym, Adaptive Corpus, Word of Mouth Classification

JEL Classification: C63