

Question Retrieval with Distributed Representations and Participant Reputation in Community Question Answering

Sam Weng^{1,2}, Kevin Chun-Kai Wu³, Yu-Chun Wang⁴, Richard Tzong-Han Tsai^{2*}

¹AsusTek Computer Inc., Taiwan

²Department of Computer Science and Information Engineering,
National Central University, Taiwan

³Department of Computer Science, National Tsing Hua University, Taiwan

⁴Department of Buddhist Studies, Dharma Drum Institute of Liberal Arts, Taiwan

*corresponding author

thtsai@csie.ncu.edu.tw

Abstract

In recent years, community-based question and answer (CQA) sites have grown rapidly in number and size. These sites represent a valuable source of online knowledge; however, they often suffer from the problem of duplicate questions. The task of question retrieval (QR) aims to find previously answered semantically similar questions in CQA archives. Nevertheless, synonymous lexical variations pose a big challenge for question retrieval. Some QR approaches address this issue by calculating the probability of correlation between new questions and archived questions. Much recent research has also focused on surface string similarity among questions. In this paper, we propose a method that first builds a continuous bag-of-words (CBoW) model with data from Asus's Republic of Gamers (ROG) forum and then determines the similarity between a given new question and the Q&As in our database. Unlike most other methods, we calculate the similarity between the given question and the archived questions and descriptions separately with two different features. In addition, we factor user reputation into our ranking model. Our experimental results on the ROG forum dataset show that our CBoW model with reputation features outperforms other top methods.

Keywords: question retrieval, QR, community-based question and answer, CQA

Acknowledgement

This research was supported in part by the Ministry of Science and Technology of Taiwan (MOST 106-2633-E-002-001), National Taiwan University (NTU-106R104045), Intel Corporation, and Delta Electronics, and Advantech.