Question Quality Analysis and Prediction in Community Question Answering Services with Coupled Mutual Reinforcement

ABSTRACT:

Community Question Answering Services (CQAS) (e.g., Yahoo! Answers) provides a platform where people post questions and answer questions posed by others. Previous works analyzed the Answer Quality (AQ) based on answer-related features, but neglect the question-related features on AQ. Previous work analyzed how asker- and question-related features affect the Question Quality (QQ) regarding the amount of attention from users, the number of answers and the question solving latency, but neglect the correlation between QQ and AQ (measured by the rating of the best answer), which is critical to Quality of Service (QoS). We handle this problem from two aspects. First, we additionally use QQ in measuring AQ, and analyze the correlation between a comprehensive list of features (including answer-related features) and QQ. Second, we propose the first method that estimates the probability for a given question to obtain high AQ. Our analysis on the Yahoo! Answers trace confirmed that the list of our identified features exert influence on AQ, which determines QQ. For the correlation analysis, the previous classification algorithms cannot consider the mutual interactions between multiple (>2) classes of features. We then propose a novel Coupled Semi-Supervised Mutual Reinforcement-based Label Propagation (CSMRLP) algorithm for this purpose.

INTRODUCTION:

SEarch engines are widely used in our daily life to find the answers to questions. However, they require users to know the effective search keywords to questions, without which users may spend an extremely long time in searching for answers. Since some user questions are typically personal, heterogeneous, extremely specific and open-ended, search engines are usually not intelligent enough to find a single web page that can directly answer such questions . Since real humans are believed to understand and answer better than a machine, Community Question Answering Services (CQAS) provide a platform to allow people to post questions and answer questions posed by others. In a CQAS, a question is open for receiving answers during a certain

time period. An asker can select the best answer for his/her question along with a rating in a given range . Also, each question has an attribute of "tag-of-interests", which represents the number of users interested in this question. Although CQAS provides more effective ways to find answers by using human resources, it is found that a large portion of questions remain unanswered in such systems, referred to as question starvation phenomenon .To solve this problem, it is important to improve the Question Quality (QQ) in order to attract more attention to a question from users. A previous work measured QQ with consideration of the amount of attention from users, the number of answers and the question solving latency. However, this work neglects the correlation between QQ and AQ, which is critical to QoS. Agichtein et al. showed that QQ correlates with AQ. In addition to the number of responses and response latency, AQ also is a critical factor that determines the Quality of Service (QoS) of CQAS. Thus, we argue that QQ measurement should also consider AQ. As a result, enhancing QQ will automatically improve AQ and hence QoS.