# Quality assessment of answers with user-identified criteria and data-driven features in social Q&A

Hengyi Fu[a], Sanghee Oh[b],*

[a] *School of Information, Florida State University, FL, USA*
[b] *Department of Library and Information Science, Chungnam National University, Daejeon, South Korea*

ARTICLE INFO

ABSTRACT

The purpose of the current study is to identify the user criteria and data-driven features, both textual and non-textual, for assessing the quality of answers posted on social questioning and answering sites (social Q&A) across four different knowledge domains—Science, Technology, Art and Recreation. A comprehensive review of literature on quality assessment of information produced in social contexts was carried out to develop the theoretical framework for the current study. A total of 23 user criteria and 24 data features were proposed and tested with high-quality answers obtained from four social Q&A sites in Stack Exchange. Findings indicate that content-related criteria and user and review features were the most frequently used in quality assessments, while the importance of user criteria and data features was variable across the knowledge domains. In the Technology Q&A site containing mostly self-help questions, the utility class was the most frequently used group of criteria. The popularity of the socio-emotional class was more apparent in discussion-oriented topic categories such as Art and Recreation, where people seek others' opinions or advice. Users of Art and Recreation Q&A sites in Stack Exchange appear to place more value on answerers' efforts and time, good attitudes or manners, personal experience, and the same taste. The importance of user features and the emphasis on answerer's expertise on the Science Q&A site was observed. Examining the connection or gap between user quality criteria and data features across the knowledge domains could help to better understand users' evaluation behaviors for their preferred answers, and identify the potential of social Q&A for user education/intervention in answer quality evaluation. This examination also offers practical guidance for designing more effective social Q&A platforms, considering how to customize community support systems, motivate contributions, and control content quality.

## 1. Introduction

Social questioning and answering sites (social Q&A) are one of the major information and knowledge sources that have emerged from Internet-mediated social practice. As an example of collective intelligence, social Q&A allows users to post questions, provide answers and comments, and evaluate questions and answers among peers. The increasing importance of social Q&A sites has resulted in an extensive body of literature aimed at evaluating and predicting content quality. Previous research has investigated quality assessment on social Q&A sites from either user or data perspectives. The first type of research has focused on user-based quality assessment—how users interpret quality and what are the criteria they use to evaluate quality (Kim, 2010; Kim & Oh, 2009; Kim, Oh,

& Oh, 2007). The second type has examined textual and non-textual features to predict quality of answers in certain settings, while the selection of features varies across studies (Dalip, Gonçalves, Cristo, & Calado, 2013; Harper, Raban, Rafaeli, & Konstan, 2008; Shah & Pomerantz, 2010; Surdeanu, Ciaramita, & Zaragoza, 2008). Both bodies of research have provided useful insights for answer assessment on social Q&A sites, but previous studies have explored the user criteria and data-driven features separately. There is still a lack of research regarding whether and how user-identified quality criteria are related to or reflected by the data-driven features of answers.

Another gap in previous research is knowledge domain (topic) as a factor in assessing and predicting quality from user and data perspectives. The criteria that users apply to evaluate answers to science-related questions may be different from those of art-related questions. The usefulness of a particular quality criterion in predicting answer quality may also vary across knowledge domains. Harper et al. (2008) concluded that topics had a small and marginally significant effect on answer quality, but did not go further to study how it could influence quality assessment. The effectiveness of textual and non-textual features may vary among different knowledge domains as well. Fu, Wu, and Oh (2015) have discovered that high-quality answer evaluation in different knowledge domains is associated with a variety of quality features.

We believe examining the connection or gap between user quality criteria and data features could help to better understand users' evaluation behaviors in relation to their preferred answers, and identify the potential of social Q&A for user education/intervention in answer quality evaluation. This type of investigation could also highlight practical ways to design more effective social Q&A platforms, including interface and community tools. Findings from our study about answer quality assessment across knowledge domains could provide insights for building classifiers to select high-quality answers based on knowledge domain, and for guiding social Q&A designers on how to customize community support systems, motivate contributions, and control content quality.

## 2. Purpose of the study

The purpose of this study is to fill the research gaps mentioned above. We examine both user-identified quality criteria and data-driven features (hereinafter user criteria and data features) of high-quality answers from four social Q&A sites in different knowledge domains—Science, Technology, Art and Recreation—hosted on Stack Exchange. The research questions we propose are as follows.

- RQ1: What are the user criteria and data features that are applied when assessing quality of answers in social Q&A?
- RQ2: How are the user criteria and data features of social Q&A applied differently across knowledge domains, especially Science, Technology, Art, and Recreation when assessing quality of answers in social Q&A?

We had an assumption that good quality answers could be highly recognizable by the social Q&A users. We used Stack Exchange Q&A network as our testbed for a social Q&A site and Stack Exchange allows users to up-vote or down-vote on answers that they like or dislike. Each answer can have an answer score which is calculated by the differences between the number of up-votes and down-votes of an answer receives. We used this answer score as an indication of showing the quality of answers: The higher the scores, the higher the answer quality was considered by users.

User criteria indicate the quality dimensions that social Q&A users value. We identified 23 user criteria and applied them when analyzing users' comments on the high-quality answers. In addition, we examined 24 data features—14 textual features and 9 non-textual features. Textual features are those that can be captured directly from the answer content such as answer length, answer structure, and writing style. Non-textual features are those that cannot be extracted directly from the textual content but can be recognized from the information about users (especially those who provide answers) or answer reviews (e.g., users' comments on answers or number of answer revisions, etc.). We then compared both user criteria and data features of high-quality answers across four domains of Science, Technology, Art, and Recreation.

## 3. Related work

We carried out a comprehensive review of literature, not limiting our review to the studies about social Q&A but including those about Wikipedia. Both answers and Wikipedia articles are user-generated information and knowledge. Users of these sites may apply similar user criteria and data features when evaluating the answers or the articles, although the content depth and structure of the answers and the articles may differ from one another. This extended scope of literature review helped to discover various user criteria and data features tested in previous studies.

### 3.1. User criteria for assessing quality of answers in social Q&A

We carried out a comparative analysis across the criteria tested in previous user studies of social Q&A answer quality assessment (Chai, Potdar, & Dillon, 2009; Choi & Shah, 2016; Emamjome, Rabaa'i, Gable, & Bandara, 2013; Ge, Helfert, & Jannach, 2011; Kim, 2010; Shah & Pomerantz, 2010; Sun, Zhao, & Zhu, 2015; Zhu, Bernhard, & Gurevych, 2009). Participants recruited for these quality assessment studies ranged from subject experts to ordinary users. User studies require more time and effort but quality criteria proposed by Q&A users can better reflect their information needs and understandings of the information problems. More recently, a new variation of user-based quality assessment has emerged in which the main goal is to examine if quality criteria identified from general Q&A sites can be transferred to Q&A features on other social networking platforms (Jenders, Krestel, & Naumann, 2016; Jeng, DesAutels, He, & Li, 2017). We grouped the criteria by the meanings/definitions although their names are different. We found