Improvement of Estimation of Objective Scores of Answer Statements Posted at Q&A Sites

Yuya Yokoyama, Graduate School of Information Science, Kyoto Institute of Technology, Kyoto, Japan
Teruhisa Hochin, Graduate School of Information Science, Kyoto Institute of Technology, Kyoto, Japan
Hiroki Nomiya, Graduate School of Information Science, Kyoto Institute of Technology, Kyoto, Japan

ABSTRACT

To eliminate mismatches between the intentions of questioners and respondents of Question and Answer (Q&A) sites, the authors have clarified the characteristics of the question and answer statements. It has been shown that the impression of the statements could be captured by nine factors, and the factor scores could be estimated from the feature values of the statements. Here, the objective scores of answer statements are provided. So far the authors have tried to estimate the objective scores of answer statements through multiple regression analysis. They are estimated from the factor scores estimated by using multiple regression formulas already obtained. As a result, with the consideration of natural logarithm, objective scores of the statements on love counseling & human relationships could be estimated good. The estimation accuracy of the statements on Yahoo! Auction and PC, however, is required to be improved. Therefore, this paper tries to improve the estimation accuracy of objective scores of Yahoo! Auction and PC. Here, the authors adopt Steven’s power law as well as Fechner’s law as new explanatory variables. In considering multicollinearity, the authors selected as many explanatory variables as possible. As a result, objective scores of the statements of PC are estimated fairly good.

Keywords: Fechner’s Law, Multicollinearity, Multiple Regression Analysis, Question and Answer Sites, Steven’s Power Law

INTRODUCTION

Recently, the number of people using Question and Answer (Q&A) sites on the Internet has been increasing (Yahoo! Answers, 2013; Yahoo! Chiebukuro, 2013). Q&A sites are online communities where users can manually post questions and answers. Hence, these sites can be considered as databases containing enormous amounts of knowledge that can be used to solve various problems. When a user posts a question, other users may respond.
The questioner selects the most appropriate response as the “Best Answer” and awards the respondent with some points as a form of fee. The Best Answer is the response statement that the questioner subjectively finds most satisfying. Several research efforts have attempted to estimate the Best Answer (Blooma, Chua, & Goh, 2008; Agichtein, Castillo, Donato, Gioni, & Mishne, 2008; Wang, Tu, Feng, & Zhang, 2009; Kim, Oh, & Oh, 2007; Nishihara, Matsumura, & Yachida, 2008).

As the number of users of Q&A sites increases and more questions are posted, it becomes harder for respondents to select questions that match their specialty and interests. Consequently, a question posed by a user may not be seen or answered by qualified respondents. Moreover, if an appropriate respondent is not encountered, mismatching may occur, which may cause the following problems:

- A questioner may acquire incorrect knowledge from inappropriate answers;
- Respondents may not have the necessary knowledge to properly answer the question, and thus the problem remains unsolved;
- Users may be offended by answers that contain abusive words, slanders, or statements against public order and standards of decency.

The authors’ objective is to present questions to qualified users who can appropriately answer them, thus avoiding the problems described above. Specifically, the authors used the impressions of 60 statements posted on Yahoo! Chiebukuro (Yokoyama, Hochin, Nomiya, & Satoh, 2011), a Q&A site in Japan, and conducted an impression evaluation experiment. By applying factor analysis to the scores obtained in the experiment, nine factors have been obtained (Yokoyama et al., 2011).

However, using this approach the authors can only obtain factor scores for the statements used in the experiment. To estimate the factor scores of other statements, multiple regression analysis is applied to the feature values of the statements. The authors adopt the syntactic information of the statements, such as word classes (such as nouns and verbs), and the number of appearances (or the percentage) of alphanumeric characters and kanji (Yokoyama et al., 2011), which is one of Chinese characters and is the Japanese writing system (“Text See Manual”, 2013). Moreover, word imageability, closing sentence expressions, word familiarity, and notation validity are also adopted as feature values (Sakuma, Ijuin, Fushimi, Tatsumi, Tanaka, Amano & Kondoh, 2008; Amano & Kondoh, 2003). It is shown that the overall estimation accuracy is good. The authors have confirmed the validity of estimating the scores of each factor by obtaining the major feature values (Amano & Kondoh, 2003).

Estimation of Best Answers is difficult because these are subjectively selected as described before. Here, National Institute of Informatics, Japan (National Institute of Informatics, 2013), has given a kind of quality to answer statements. This quality indicates how an answer is good for the question, and/or suitable to it. The degree of quality is determined by four assessors in the range of nine levels. As four assessors have determined the degrees, these are considered to be objective. These may be estimated more easily than the Best Answers. With the consideration of objective viewpoints as well as subjective ones, the issues caused by the mismatch between questioners and respondents could be avoided.

So far the authors have estimated the objective scores of answer statements through multiple regression analysis (Yokoyama, Hochin, Nomiya, & Satoh, 2013). They are estimated from the factor scores estimated by using multiple regression formulas already obtained. As a result of estimation, objective scores of the statements of Yahoo! Auction and PC are estimated fairly good. Those of love counseling & human relationships could be estimated well by taking the natural logarithms of the factor scores (Yokoyama et al., 2013). As to Yahoo! Auction and PC, however, the estimation accuracy of objective scores of those statements is required to be improved.
Related Content

Analysis of Existing Software Cognitive Complexity Measures
[www.igi-global.com/article/analysis-of-existing-software-cognitive-complexity-measures/204524?camid=4v1a](www.igi-global.com/article/analysis-of-existing-software-cognitive-complexity-measures/204524?camid=4v1a)

Unfolding Models for Asymmetric Dissimilarity Data With External Information Based on Path Structures
[www.igi-global.com/article/unfolding-models-for-asymmetric-dissimilarity-data-with-external-information-based-on-path-structures/207725?camid=4v1a](www.igi-global.com/article/unfolding-models-for-asymmetric-dissimilarity-data-with-external-information-based-on-path-structures/207725?camid=4v1a)
Non-Intrusive Adaptation of System Execution Traces for Performance Analysis of Software Systems
www.igi-global.com/chapter/non-intrusive-adaptation-of-system-execution-traces-for-performance-analysis-of-software-systems/108632?camid=4v1a

Experiences with Requirements Model Reuse: The NFR Catalogue for Ubiquitous Systems
www.igi-global.com/chapter/experiences-requirements-model-reuse/66468?camid=4v1a