Estimation of Factor Scores of Impressions of Question and Answer Statements

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, we have conducted impression evaluation experiment and nine factors are obtained. Factor scores of the statements used in the experiments are obtained. Those of other statements, however, cannot be obtained. This paper proposes a method of estimating the factor scores of other statements through multiple regression analysis applied to those feature values of statements. Syntactic information of statements, word imageability, closing sentence expressions, word familiarity and notation validity are adopted as feature values of statements. It is shown that overall estimation accuracy is very good for all the factors. This research uses the data of “Yahoo! Chiebukuro” given to National Institute of Informatics by Yahoo Japan Corporation.

Keywords: Factor Estimations, Mismatching, Notation Validity, Question and Answer Statements, Word Familiarity, Word Imageability

INTRODUCTION

Recently, the number of people using Q&A sites on the Internet has been increasing. 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.

DOI: 10.4018/ijsi.2013040105
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.

In this paper, our objective is to present questions to qualified users who can appropriately answer them, thus avoiding the problems described above. Specifically, we use the impressions of 60 statements posted on Yahoo! Chiebukuro (Yokoyama et al., 2011) and conduct an impression evaluation experiment. By applying factor analysis to the scores obtained in the experiment, we have obtained nine factors (Yokoyama et al., 2011). It has been shown that the Best Answers can be identified by using the factor scores. However, we can only obtain factor scores for the statements used in this experiment.

This paper proposes a method of estimating the factor scores of other statements through multiple regression analysis applied to the feature values of the statements. Feature values adopted are the syntactic information of the statements, such as word classes (such as nouns and verbs), the number of appearances (or the percentage) of Chinese and alphanumeric characters, word imageability, closing sentence expressions, word familiarity and notation validity. It is shown that estimation accuracy is very good when these feature values are used.

The remainder of this paper is organized as follows. First, we describe related works. Factors of Question and Answer statements are summarized in the following section. Feature values of statements used are introduced afterwards. Next, we present the obtained estimation results. Finally, we provide considerations and we conclude the paper.

RELATED WORKS

Several attempts at estimating the Best Answers have been reported [3–7]. Nishihara et al. used textual and non-textual features to predict the Best Answers (Nishihara, Matsumura, & Yachida, 2008). They used five textual and five non-textual features. It was found that textual features influence the quality of answers more than non-textual features. Accuracy, completeness, language, reasonableness, and length are considered as textual features. Agichtein et al. used the content and usage features of questions and answers to assess their quality (Agichtein et al., 2008). Of the twenty major features determining the quality of a question, eleven were related to web information and nine were obtained through questions. The analogical reasoning approach (Wang et al., 2009) finds the Best Answer by using links between questions and answers contained in previous knowledge. In their approach, three textual features, seven statistical features, and five user interactions were used. Kim et al. have proposed the Best Answer selection criteria (Kim, Oh, & Oh, 2007), which consist of seven categories: content value, cognitive value, socio-emotional value, extrinsic value, information source value, utility, and general statement. For information-type questions, content values are important. Utility is important for suggestion-type questions, while socio-emotional values are vital for opinion-type questions.

Nishihara et al. have also proposed a method of detecting the Best Answer to a question (Nishihara, Matsumura, & Yachida, 2008). They obtained the Best Answer to a question by noticing the affinity between closing sentence expressions of questioners and respondents, and the clustering combinations...
Related Content

Security Requirements Engineering for Evolving Software Systems: A Survey
[www.irma-international.org/article/security-requirements-engineering-evolving-software/39009/](www.irma-international.org/article/security-requirements-engineering-evolving-software/39009/)

Using Attack Graphs to Analyze Social Engineering Threats
[www.irma-international.org/article/using-attack-graphs-to-analyze-social-engineering-threats/136466/](www.irma-international.org/article/using-attack-graphs-to-analyze-social-engineering-threats/136466/)

On Spam Susceptibility and Browser Updating
Eric Luong, Toan Huynh and James Miller (2012). *International Journal of Systems and Service-Oriented Engineering* (pp. 44-57).
[www.irma-international.org/article/spam-susceptibility-browser-updating/64198/](www.irma-international.org/article/spam-susceptibility-browser-updating/64198/)

Applying a Model-Driven Framework for Gap Analysis: Towards Business Service Engineering
[www.irma-international.org/chapter/applying-a-model-driven-framework-for-gap-analysis/96617/](www.irma-international.org/chapter/applying-a-model-driven-framework-for-gap-analysis/96617/)

Basics to Multirate Systems
[www.irma-international.org/chapter/basics-multirate-systems/27211/](www.irma-international.org/chapter/basics-multirate-systems/27211/)