Advanced Engineering Informatics 29 (2015) 813–829

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Automatic questionnaire survey by using the collective message over the Internet

Jiang-Liang Hou *, Yuh-Zong Chu [1]

Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu 300, Taiwan

ARTICLE INFO

Article history:
Received 2 January 2015
Received in revised form 8 August 2015
Accepted 2 September 2015
Available online 31 October 2015

Keywords:
Questionnaire survey
Artificial Neural Network
Question answering system

ABSTRACT

As to implement questionnaire survey, a researcher usually has to spend much time and efforts on questionnaire delivery, questionnaire collection and questionnaire analysis. In order to enhance efficiency and effectiveness in questionnaire survey, this research develops an automatic questionnaire survey model. In the proposed model, the questions in the questionnaires can be converted into structured components and the component characteristics can be obtained from the related online message. By using the Artificial Neural Network method, the relationship model between the online message and questionnaire questions can be derived and the derived model can be used to acquire the answers to the questions from the message highly related to the questions. Moreover, distributions of the answers can be visually depicted. By utilizing the proposed methodology, researchers can efficiently realize the results about their questionnaire derived from the online population.

On the basis of the proposed model, this research develops a platform for automatic questionnaire survey and analysis. Moreover, real-world questions and message, Yahoo! Answers, were used to evaluate feasibility and performance of the proposed model and platform. According to the experimental result question answers can be effectively acquired from the message highly related to the questions. The proposed method and platform can assist people quickly and effectively accomplish questionnaire survey.

1. Introduction

In the previous questionnaire survey process, researchers often need to design questionnaires regarding the research issues and deliver the questionnaires through the Internet to the target participants. After that, researchers can collect the survey results over the Internet as the participants finish the questionnaire. As the number of feasible results reached a pre-define target, the researchers can realize the results about their questionnaire survey by analyzing the questionnaire answers via statistic methods. As a matter of fact, efficiency and effectiveness of the traditional questionnaire survey process are not satisfactory to researchers owing to several issues. Firstly, before questionnaire delivery and collection, the researcher has to clarify propose of questionnaire in order to identify the target participants. After that, the researcher can deliver the questionnaire via different channels to the target participants and passively wait for replies from the participants. Secondly, after collecting the replies to the questionnaire, the researchers usually has to spend much time in examining the reliability of questionnaire replies and identify the effective replies. Finally, researcher must spend a lot of time and efforts to analyze the replies in order to derive the distribution of replies about the questionnaire survey. As a whole, the as-is model for the previous questionnaire survey can be shown in Fig. 1.

Owing to the recent development and popularization of the social network over the Internet in recent years, a large number of public messages are spread over the Internet nowadays. The public messages over the Internet can be easily acquired and can be utilized to extract the answers to the questionnaire. Therefore, in order to improve traditional questionnaire survey, this research develops an automatic questionnaire survey model and system to enhance efficiency and effectiveness of questionnaire survey. As a researcher would like to know the options and responses of the public on the target issues, he can simply design the questionnaire and import it onto our proposed automatic questionnaire survey system. After that, the proposed system will proactively search the related online message from different population over the Internet and extract the answers to the questions from the message highly related to the questions. The proposed system will also analyze the answers and replies via statistic approach, and

* Corresponding author. Tel.: +886 3 5742658; fax: +886 3 5722685.
E-mail addresses: adamhou@ie.nthu.edu.tw (J.-L. Hou), superflychu@gmail.com (Y.-Z. Chu).
[1] Tel.: +886 3 5715131x33981; fax: +886 3 5722685.

http://dx.doi.org/10.1016/j.aei.2015.09.001
1474-0346/

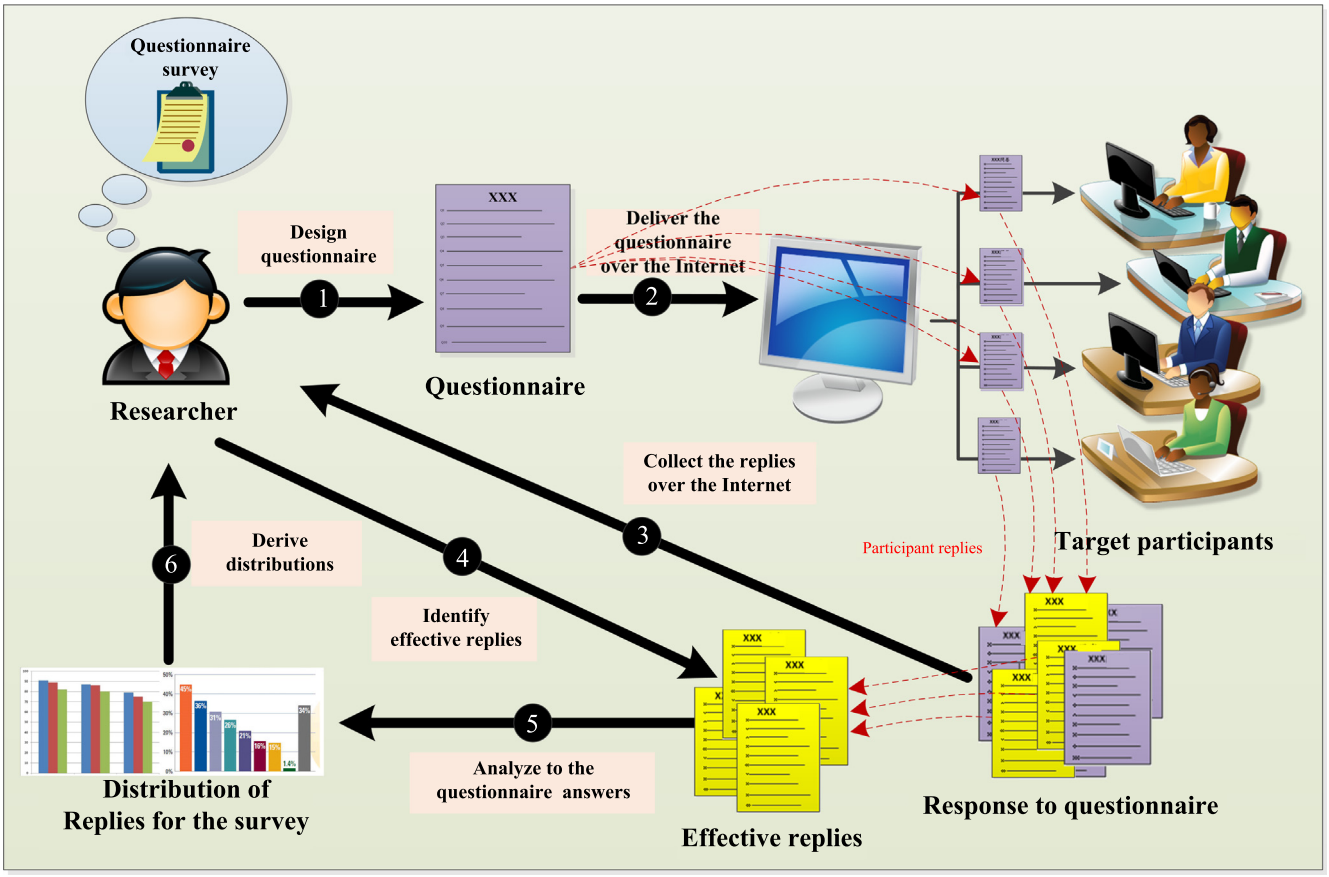

Fig. 1. As-is model for questionnaire survey.

visualize the distributions of answers with respect to different population. As a result, the researcher can efficiently realize the results of their questionnaire survey. The to-be model for questionnaire survey can be shown in Fig. 2.

2. Review of related works

In order to assist researchers to efficiently acquire the results for their questionnaire survey, this research develops an automatic questionnaire survey model and system by utilizing existing Internet message. Based on questionnaire questions, this system can proactively acquire related message and personal profiles that publish by the public over the Internet and automatically analyze feature of the related message and personal profiles. The results can also visualize for higher reliability. In brief, the related topics of this study are question and answer collection, extraction of answers to questions, and implementation Q&A systems. The related works about these subjects are reviewed in the following.

- ***Question and answer collection***

Question and answer collection is to analyze the representation characteristics and important information from different type of data (including the semantic expressions, images expressions and videos expressions) and construct the corresponding databases for those types of data on the basis of the analysis results [6,5,1]. In this topic, our study is highly related to analysis of semantic expressions, i.e., analysis of the semantics and structure of sentences and characteristics of domain terms. Based on the analysis results, a thesaurus is established for analysis questionnaire questions and replies. The traditional questionnaire survey focuses mainly on the question design and answer collection for specific application domains, such as health survey [13] and population estimation [3].

- ***Extraction of answers to questions***

The studies related to the issue of extraction of answers to questions can be further clarified into question classification, related document/information retrieval and answer extraction. In most studies, the questions provided by users are considered as target questions. The target questions are analyzed and followed by question classification, related document/information retrieval and answer extraction via the public resources [10], statistic approaches [14] and mapping mechanisms [8]. Concerning the existing challenges, dilemmas, and opportunities for survey researchers and social scientists, both technological and cultural changes influence choices for survey methodologies [11,2]. Different from the previous studies, this study will establish the relationship between online message and questionnaire questions by our proposed document/information retrieval model (based on the Artificial Neural Network method). Secondly, our study will present a methodology to extract answers to the questionnaire questions from the message highly related to the questions via answer extraction model (based on the established thesaurus and semantics). Finally, our model can classify the questionnaire questions via a question classification mechanism. Consequently, distributions of the answers corresponding to different population can be derived based on the classification result.