Closed-Ended Questionnaire Data Analysis

Leuo-Hong Wang[1], Chao-Fu Hong[1], and Chia-Ling Hsu[2]

[1] Evolutionary Computation Laboratory,
Department of Information Management, Aletheia University, Taiwan
{wanglh, cfhong}@email.au.edu.tw
[2] Centre for Teacher Education, Tamkang University, Taiwan
clhsu@mail.tku.edu.tw

Abstract. A KeyGraph-like algorithm, which incorporates the concept of structural importance with association rules mining, for analyzing closed-ended questionnaire data is presented in this paper. The proposed algorithm transforms the questionnaire data into a directed graph, and then applies association rules mining and clustering procedures, whose parameters are determined by gradient sensitivity analysis, as well as correlation analysis in turn to the graph. As a result, both statistically significant and other cryptic events are successfully unveiled. A questionnaire survey data from an instructional design application has been analyzed by the proposed algorithm. Comparing to the results of statistical methods, which elicited almost no information, the proposed algorithm successfully identified three cryptic events and provided five different strategies for designing instructional activities. The preliminary experimental results indicated that the algorithm works out for analyzing closed-ended questionnaire survey data.

1 Introduction

The questionnaire survey with closed-ended questions is one of the most common used tools for user information elicitation. The data collected are usually analyzed by various statistical techniques [1]. However, if no statistically significant events exist in the data, little information can be extracted by these statistical techniques.

Well-designed questionnaires, especially close-ended, rating scaled questionnaires, always try to capture the intended information as much as possible by carefully wording each question. Sentences in such questions therefore contain a certain keywords to appropriately represent the subjects of questions. Respondents convey how intensive they feel about these keywords at the same time when they answer questions to express their opinions. Hence, dealing with the collected data via the keywords viewpoint gives us an alternative way for eliciting information from survey data. If each keyword in the questions is, in other words, treated as an item and the relationships between these keywords can be appropriately defined, then data mining algorithms such as the a-priori algorithm for association rule mining [2] can be applied. As a result, the dependency

B. Gabrys, R.J. Howlett, and L.C. Jain (Eds.): KES 2006, Part III, LNAI 4253, pp. 1–7, 2006.

of keywords can be calculated. That implies the relationship between variables, which are observed by the questionnaire, is capable of being found.

Although dealing with data via the a-priori algorithm seems to work out for survey data just as good as other statistical techniques, only the most significant events would be identified. Eliciting little information is still a problem for such an analysis scheme. Therefore, we have designed a new algorithm integrating the concept of finding rare but structurally important events, which have been discussed by a handful of researchers [3][4][5][6] in *chance discovery* [7] discipline recently, with the a-priori algorithm for analyzing closed-ended questionnaires in this paper. Actually, the algorithm is similar to KeyGraph algorithm [4], except the measures of *support* and *confidence* used by association rules mining are introduced. In addition, a sensitivity analysis procedure which is applied to find the threshold values of these measures is also integrated with the process of identifying chances. With these modifications, the new algorithm can automatically identify chances and is suitable for questionnaire data analysis.

The rest of this paper is organized as follows. Section 2 presents the new algorithm to identify significant events and discover chances from closed-ended questionnaire survey data. Section 3 starts with introducing the instructional design application and then defines its structurally important events. The experimental results of applying our algorithm to the application are also shown in this section. We would summarize the paper in the final section.

2 The A-Priori Based Chance Discovery Algorithm

To deal with questionnaire data, we use the *support* and *confidence* measures which usually used by data mining applications to identify clusters in our algorithm. To investigate the structural importance of the survey data, a weighted, directed graph $G = (V, A)$ where V is a set of nodes and A is a set of arcs between nodes must be constructed first. As mentioned in Section 1, a well-designed questionnaire is always carefully wording each question, so the set of nodes V for the graph G intuitively consists of every question of the questionnaire. However, question numbers but not keywords are used for representing the nodes in the following for simplicity. Meanwhile, each arc in the set A represents an association rule between two questions. Moreover, each node and each arc are assigned a weight. The weight of a node is the summation score given by all respondents. The weight of arc is defined as the co-occurrence frequency, which would be described later, of two adjacent questions. Once the graph is constructed, our algorithm can then be applied to identify the rare but structurally important nodes.

2.1 Preprocessing- Constructing a Weighted, Directed Graph

Assume a questionnaire consisting of N closed-ended, 1-to-5 scale rating questions is given to M respondents for evaluation. Then an M by N score matrix $S = s_{mn}$, where s_{mn} is the score of the n^{th} question given by the m^{th} respondent, can be filled. Each row of S is rewritten to a string record consisting of