A Characterized Rating Recommend System

Yao-Tsung Lin and Shian-Shyong Tseng

Department of Computer and Information Science, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C. {gis88801, sstseng}@cis.nctu.edu.tw

Abstract. In recent years, due to the rapid growth of Internet usage, the problem of how to avoid inappropriate Internet contents accessing becomes more and more important. To solve the problem, a Collaborative Rating System [3, 4] based upon PICS protocol has been proposed. However, since the users usually would like to consult the opinions of the user group with similar rating tendency rather than the common opinions from the majority, it means the opinion of second majority with sufficient number of voters should also be considered. So does third majority, and so on. In order to provide a characterized rating service, a Characterized Rating Recommend System is designed to provide more precise and proper rating service for each user. Also, in this work, a questionnaire is designed to get users’ opinions, and some experimental results show that the system can provide acceptable rating service.

1 Introduction

In recent years, due to the rapid growth of Internet usage, the problem of how to avoid inappropriate Internet contents accessing becomes more and more important. The concept of content selection is proposed to solve this problem, and there are many previous researches about content selection; e.g., PICS [5] protocol, which is proposed by W3C [8]. But there are still some problems; e.g., problem of rating information collecting. To solve this problem, a Collaborative Rating System [3, 4] has been proposed. However, users’ rating tendencies can not be considered in the system. In order to provide a Characterized Rating Service which take care of this problem, in this work, the opinions of participants will be first represented in well-structured data, Rating Vectors, and these Rating Vectors will be clustered into Rating Groups, corresponding to different rating opinions. Then the properties of each Rating Groups will be mined by using Rating Decision Tree Constructing Algorithm. To prevent the problem of over-fitting in a decision tree, a Precision and Support Based Decision Pruning Algorithm will be applied. Finally, the rules about Rating Groups generated will be used to provide users characterized rating services. Based on these concepts, a Characterized Rating Recommend System is designed. In the experiment, a questionnaire is designed to efficiently get opinions about content rating. 700 participants are

1 This work was supported by Ministry of Education and National Science Council of the Republic of China under Grand No. 89-E-FA04-1-4, High Confidence Information Systems.

asked to answer the questionnaire, 616 of the filled questionnaires are useable in the clustering without missing data. Some experimental results with cross validation will also be shown in this paper.

2 Related Works

In order to solve the problem of how to avoid inappropriate Internet content, many researches are proposed [1, 3, 4, 7], and among these researches, Collaborative Rating System [3, 4] based upon PICS [5] protocol, provides practical solution. And Characterized Rating Recommend System is proposed to make Collaborative Rating System more adaptive to different user requirements.

2.1 PICS, Platform for Internet Content Selection [5]

To solve the problems of selecting appropriate or desired content via the Internet, many researches have been proposed. In these researches, the PICS[5] protocol was proposed by W3C[8] organization and provided a systematic architecture for document rating system, and it also provides the methods of rating information collecting. In PICS protocol, the rating information is provided by two methods, self-labeling and third-party labeling. In self-labeling method, the rating information is provided by the content providers of each web page. In third-party labeling, the rating information is provided by specific groups or organizations instead of the content provider. However, these rating information collecting methods seem to be too weak, since there is no obligation for content providers to provide the rating, it is impossible to rate all documents by few voluntary or non-profit organizations, and it is hard to design an acceptable automatic rating system.

2.2 Collaborative Rating System [3, 4]

To solve the issues on rating information collecting, a Collaborative Rating System [3, 4] was proposed, which collects rating information by the help of huge amount of volunteers. In Collaborative Rating System, participants are asked to rate web contents as they browse web pages according to a selected rating category, and their ratings will be collected and used to conclude a more objective result. The attributes of collected rating data consist of:

<table>
<thead>
<tr>
<th>Category</th>
<th>Web Page</th>
<th>User</th>
<th>Level</th>
</tr>
</thead>
</table>

In the collected rating data, Category represents the selected rating category of this rating, and Web Page indicates the address of target web page. The information in User attribute records who had made the rating, and Level attribute is the rated level in the selecting rating category the user thought. The rating data of the same web page