Knowledge-Based Systems 72 (2014) 96–107

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Game-theoretic rough sets for recommender systems

Nouman Azam, JingTao Yao *

Department of Computer Science, University of Regina, Regina, SK S4S 0A2, Canada

ARTICLE INFO

Article history:
Received 27 May 2014
Received in revised form 29 August 2014
Accepted 30 August 2014
Available online 20 September 2014

Keywords:
Game-theoretic rough sets
Probabilistic rough sets
Game theory
Recommender systems
Rough sets

ABSTRACT

Recommender systems guide their users in decisions related to personal tastes and choices. The rough set theory can be considered as a useful tool for predicting recommendations in recommender systems. We examine two properties of recommendations with rough sets. The first property refers to accuracy or appropriateness of recommendations and the second property highlights the generality or coverage of recommendations. Making highly accurate recommendations for majority of the users is a major hindrance in achieving high quality performance for recommender systems. In the probabilistic rough set models, these two properties are controlled by thresholds (α, β). One of the research issues is to determine effective values of these thresholds based on the two considered properties. We apply the game-theoretic rough set (GTRS) model to obtain suitable values of these thresholds by implementing a game for determining a trade-off and balanced solution between accuracy and generality. Experimental results on *movielen* dataset suggest that the GTRS improves the two properties of recommendations leading to better overall performance compared to the Pawlak rough set model.

1. Introduction

The internet or Web users are commonly confronted with situations where there are many potentially useful choices for selecting an item of interest. Making a suitable decision, such as a purchase decision, generally involves an urge to narrow down the search to a few but important choices. Web-based recommender systems (WRS) as a special kind of Web-based support systems, were introduced with the intent of handling relevant information in order to provide useful and customized recommendations to their users [2,30].

There are many techniques and approaches proposed in the literature that are used to develop different recommender systems. The most common approaches are content based, collaborative based, knowledge based and demographic based recommender systems [6]. In content based approach, the recommendations are based on contents or properties of items that are of interest to a user [32]. The collaborative based approach provides recommendations based on users having similar interests to the user in question [34]. The knowledge based approach makes use of the data pertaining to the user's needs and preferences to recommend a suitable option [7]. The demographic based approach aims to group or cluster the users based on personal attributes and make recommendations based on the demographic group a user belongs to [28]. Despite of some differences, all of these approaches require some sort of intelligent mechanisms to make effective recommendations. The collaborative based approach is comparatively more popular and successful way for building recommender systems [34]. In addition, the demographic based approach is sometimes treated and considered as an extension of the collaborative based approach [2]. For these reasons, we focus on collaborative and demographic based recommendations in this research.

Rough set theory, emerged in the early 1980s, is an important and useful mathematical approach to handle vague and imperfect knowledge [25,27]. It can effectively process uncertain, incomplete and insufficient information to make useful inferences and reasonings [26]. The conventional Pawlak model in rough set theory is of qualitative nature in the sense that it does not allow any errors in the positive and negative regions [26,41,44]. Researchers argued that the qualitative absoluteness or intolerance to errors can lead to problems and limitations in practical applications [44,46,47]. Quantitative generalizations and models of rough sets were introduced that generally resort to some measures and thresholds to express error tolerance [46]. The probabilistic rough set models represent one class of these quantitative models and include the decision-theoretic rough set model [40,45], the variable precision rough set model [47,48], the Bayesian rough set model [10,31], the information-theoretic rough set model [9] and the game-theoretic rough set model [12,39]. An important realization in these models is that a pair of thresholds (α, β) is used to define the rough set approximations and the resulting three regions [11,41,43]. The

* Corresponding author. Tel.: +1 306 585 4071.
E-mail address: jtyao@cs.uregina.ca (J.T. Yao).

http://dx.doi.org/10.1016/j.knosys.2014.08.030
0950-7051/

determination and interpretation of thresholds are two important issues in the probabilistic rough sets [43]. Some notable attempts in this regard can be found in references [4,9,12,16,19–21].

Three-way or ternary decisions are obtained with probabilistic rough set models when the positive, negative and boundary regions are interpreted as regions of acceptance, rejection and deferment decisions [42]. The configuration of thresholds play a crucial role in obtaining suitable decision regions [4]. Setting the thresholds in order to decrease the number of deferment decisions improves the generality but results in many incorrect acceptance and incorrect rejection decisions. On the other hand, adjusting the thresholds to obtain more accurate decisions leads to many deferment decisions. How to obtain thresholds in order to balance the properties of accuracy and generality is a major obstacle in obtaining an effective probabilistic model. We employ the game-theoretic rough set (GTRS) model for such a purpose.

The rough set theory has recently gained some attention in the recommender systems research [5,8,15,17,24,33]. It has been used with different intends in designing, developing and implementing recommender systems, such as, to obtain rules for determining the competence of players in a game [8], to visualize user preferences in menu selection [17], to reduce attributes for determining useful keywords in web page recommendations [5] and to deal with the problem of missing ratings (also termed as the problem of sparsity) [15]. The work presented in this research is different from existing work in three ways. Firstly, we look at multiple aspects of recommendation decisions simultaneously. Secondly, we go beyond the basic capabilities of rough sets by merging and combining it with the field of game theory in a GTRS model. Thirdly, we focus on ternary decision making aspect of rough sets for recommendations.

The GTRS based thresholds can be used to obtain the three rough set regions which are helpful in applications for obtaining useful rules for decision support and reducing data processing time. In a recent article, the GTRS based three-way decisions were applied and analyzed in the medical field [37]. This article extends the GTRS to WRS for obtaining decision recommendations. Specifically, we focus on determining a tradeoff solution between the properties of accuracy and generality of rough set based recommendations. The GTRS provides benefits in at least two aspects. Firstly, the determination of thresholds based on a tradeoff solution between multiple criteria can comparatively lead to cost effective and moderate threshold levels [4,37]. Secondly, unlike other models where the users or experts are required to provide parameters or the notions of costs, risks or uncertainty are used to determine the thresholds, the GTRS obtains the thresholds based on the data itself [37].

The remaining of this article is organized as follows. Section 2 explains an architecture of WRS that incorporate a GTRS component for obtaining intelligent decision recommendations. Section 3 elaborates rough sets based recommendations and explains how the properties of accuracy and generality affect these recommendations. Section 4 describes the general GTRS model. In Section 5, a GTRS based approach for threshold determination is presented. Finally, Section 6 contains experimental results with the proposed approach on *movielens* dataset.

2. Architecture of Web-based recommender systems

The Web has been increasingly used as a platform for supporting, developing and deploying recommender systems [6]. In some sense, the WRS may be viewed as a branch or sub-class of Web-based support systems that provide assistance to their users in the form of recommendations. A WRS may consist of different components with many functionalities ranging from supporting end user activities and interaction through interface to maintaining and manipulating the knowledge within the system. Following our previous knowledge and understanding of the architecture of Web-based support systems [36,38], we consider the architecture of Web-based recommender systems as comprising of three fundamental layers, i.e. the interface, management and data layers.

The Web and internet make up the interface layer. A client interaction with the system that is deployed on the server side either completely or partially is made possible through the support provided by the Web and internet. The Web browsers play a major role in presenting the interface to the clients. The interface enables the users to enter any relevant information that the system may need. In addition, it also allows the users to receive useful information about items or services that may be of potential interest to them in the form of effective recommendations. An effective and carefully designed Web interface plays a critical role for the success of any WRS. Special attention has to be given for its clarity, completeness and consistency.

The management layer serves as a middleware in the three layer architecture. The information from the top and bottom layers are processed at this layer before being presented to an intended upper or lower layer. Some of the components that may be required to make up this layer include, a Database Management System, Knowledge Discovery/Data Mining and Control Facilities. The Database Management System is responsible for retrieving relevant information about the users and items from the data layer and make them available to other system components. The Knowledge Discovery component is responsible for discovering and mining important information based on users and items features. It is this component that ultimately determines the overall success of the system in the long term. Intelligent techniques such as logic, inference and reasoning about the data may be incorporated in this component to analyze data and make recommendations. We investigate the use of GTRS as a tool for making intelligent recommendations in this component. Finally, the Control Facility component may be needed to ensure that the system is being used in a right way. Access rights, permissions and confidential information should be properly handled by this component.

The third layer in the three layer architecture is the data layer which contains data necessary for the operation of the system. This layer contains data about the users of the system, the items and their respective features and the user choices, preferences or ratings for different items. The data related to users and ratings may be captured explicitly in the form of queries or may be implicitly gathered based on the users interaction with the system. The Knowledge Discovery component (at the management layer) can retrieve and access this information for analysis purpose. This layer may also have a knowledge base component that may contain the results of data analysis, such as rules or patterns or certain parameter values that are used in making recommendations.

3. Rough sets based recommendations

We consider recommendations with the probabilistic rough set model [41]. For the sake of completeness, we briefly review the main results of the Pawlak rough set model and the probabilistic rough set model.

The Pawlak rough set model approximates a concept or a set by a pair of lower and upper approximations. For a set C, the lower and upper approximations are defined as [25,26],

$$\underline{apr}(C) = \{x \in U \mid [x] \subseteq C\}, \tag{1}$$
$$\overline{apr}(C) = \{x \in U \mid [x] \cap C \neq \emptyset\}, \tag{2}$$

where U is the set of objects called universe and $[x]$ is an equivalence class (containing object x) and is based on an equivalence relation