Expert Systems With Applications 42 (2015) 9402–9411

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Vague preferences in recommender systems

Paweł Ładyżyński [a], Przemyslaw Grzegorzewski [b,c,*]

[a] Interdisciplinary PhD Studies, Polish Academy of Sciences, Jana Kazimierza 5, 01-248 Warsaw, Poland
[b] Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland
[c] Faculty of Mathematics and Information Science, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland

ARTICLE INFO

Keywords:
Collaborative filtering
Correlation
Entropy
IF-set
Ordering
Preference
Rank
Recommender systems
Similarity

ABSTRACT

Measuring similarity between preferences is a crucial problem for recommender systems. This task becomes significantly harder when preferences are incomplete or somehow vague. In this paper we show how to model vague preferences using IF-sets and then how to quantify similarity between preferences in a way that might be useful in collaborative filtering. We consider some comparison measures between IF-sets to find those possessing properties desirable in recommender systems. Then we construct some measures that might be useful in finding other customers somehow similar to our new user of a recommender system and in promoting those customers who have an extensive knowledge on many products not yet familiar to this new user. We also suggest how to combine the aforementioned methodology with some new entropy-based analytical and graphical tools to create recommendations and support customer's decisions. The proposed graphical method for comparing possible recommendations due to several aspects enables to choose a recommendation that fits best to individual decision-making strategy of each user.

1. Introduction

The main goal of a recommender system is to generate meaningful recommendations for items or products that might be interesting for a user (see Melville & Sindhwani, 2010). Two basic architectures are usually applied in recommender systems: content-based filtering (focused on the similarity of items determined by measuring the similarity in their properties) and collaborative filtering systems (focused on the similarity of items determined by the similarity of the ratings of those items rated by the users).

Measuring similarity between preferences of the users is a crucial point in the collaborative filtering. Actually, a recommender system suggests to a user available resources using similarity of his/her preferences to those of the other customers. These preferences may be represented by an ordering set by a user of the items he/she have seen earlier. If an ordering specific to user A is similar to an ordering of another user B then the recommender system tends to suggest to user A another resources highly preferred by B and yet not known to A.

Preferences for certain items in a recommender system are stored in the so-called utility matrix with rows corresponding to the users, columns representing items, and the entries showing ratings attributed by each user to particular items.

There are many methods to compare rankings, like Spearman's rho or Kendall's tau, etc., that enable to state whether preferences under study are concordant or discordant and to evaluate the strength of the possible association (for some recent results we refer the reader to Al-Shamri, 2014). However, these coefficients work perfectly for ratings formed by linear orders. Unfortunately, real utility matrices are sparse since many entries are unknown. Such so-called vague preferences require special tools for measuring similarities. The generalized version of Kendall's tau for vague preferences was suggested by Grzegorzewski (2009) who has also proposed the generalized version of Kendall's coefficient of concordance (Grzegorzewski, 2006). The generalized version of Spearman's rho was suggested (Grzegorzewski & Ziembińska, 2011).

In this paper we propose a novel approach for measuring similarity between preferences. We discuss it in the context of IF-based model of vague preferences (Grzegorzewski, 2006; 2009; Grzegorzewski & Ziembińska, 2011), IF-distances, IF-dissimilarities and IF-divergences (Montes, Janiš, Montes, & Greco, 2012; Montes, Iglesias, Janiš, & Montes, 2012). The suggested methodology for modeling preferences by specific IF-sets provides a single frame comprising diverse possible situations including perfect knowledge, incomplete data, ties, vague answers, etc. Moreover, expressing preferences in a form of IF-sets we can take advantage of a variety of tools including different measures for comparing IF-sets. In this paper we conduct a study to indicate such comparison measures that possess

* Corresponding author at.: Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland. Tel.: +48 223810207; fax: +48 223810105.
E-mail addresses: pawelladyz@wp.pl (P. Ładyżyński), pgrzeg@ibspan.waw.pl (P. Grzegorzewski).

http://dx.doi.org/10.1016/j.eswa.2015.08.006
0957-4174/

properties desirable in recommender systems. Next, we use some selected measures with appropriate properties for quantifying similarity between preferences and we suggest how to create recommendations to support customer's decisions. We also propose a novel methodology for graphical summarizing of possible recommendations that enables a user to choose such recommendation that fits best to his individual decision-making strategy (corresponding to his degree of aversion or risk, etc.).

The paper is organized as follows. Section 2 presents the related work and motivates this study. In Section 3 we establish general requirements desired for similarity measures in recommender systems. In Section 4 we recall briefly some information on IF-sets and then, in Section 5 we describe how to apply them for modeling preferences. The problem of vague preference systems comparison is discussed in Section 6. Next, in Section 7, we suggest some tools that might be helpful in creating recommendations, including entropy-based similarity measures and graphical methods for supporting customer's decision. Then, in Section 8, we describe a simulation study and present the results of the experiment performed to evaluate the suggested methodology. Finally, Section 9 contains conclusions and discussion of some further research directions.

2. Related work

The scope of this paper can be divided into two major streams. The first one is strongly connected with modeling vague preferences. Particularly, we focus on methods of comparing vague preference systems and their applicability in creating recommendations. The second stream concerns a novel methodology of recommending a product that fits best to user expectations.

Incomplete data are a serious problem in diverse applications including, but not limited, to recommender systems. There are different reasons for missings in data. For example, respondents may refuse or may be unable to answer some questions, etc. Sometimes analysts exclude units that have missing values from the study but such strategy, called a *complete-case analysis*, is generally inappropriate, since we lose too much information available in records containing missings. Therefore, methods to handle missing data are of interest. For the detailed description and characterization of these methods we refer the reader to the famous monograph by Little and Rubin (2002).

In recommender systems the lack of actual ratings is addressed, e.g., in Kim and Choi (2014), Lee, Park, and Park (2007) or Melville, Mooney, and Nagarajan (2002). In Lee et al. (2007) it is proposed how to fill the rating matrix using implicit feedback from the user. Another approach is suggested in Melville et al. (2002), where a content model is used to estimate ratings of all objects for all users and then the collaborative filtering utilizes the matrix with all ratings. In Kim and Choi (2014) a Bayesian binomial mixture model is used for the collaborative prediction, where the generative process for data and missing data mechanism are jointly modeled to handle non-random missing data.

The aim of our contribution is to use the available data in their existing form without using any imputation or missing ranks refilling. This step is crucial in the next part of creating recommendation where the information about missing labels is required to determine the confidence of the recommendation as well as on the aggregated general knowledge of the users that exerts an influence on the recommendation. Thus we take advantage of the IF-set theory (Atanassov, 1986) which delivers flexible tools for constructing mathematical models that make possible to represent vague information. IF-sets (also called intuitionistic fuzzy sets) which are a generalization of fuzzy sets introduced by Zadeh (Zadeh, 1965) appear very useful in describing and handling imprecise information. Contrary to the usual fuzzy sets, IF-sets allow to model bipolar information which may appear when expressing positive and negative attitude or opinions, expressed by respondents. This property of IF-sets was utilized by Grzegorzewski to propose a model of vague preferences comprising not only precise and imprecise answers but ties, missings, unknown or non-comparable outputs, etc. (Grzegorzewski, 2006; 2009; Grzegorzewski & Ziembińska, 2011). In this paper we develop this approach to propose similarity measures which possess appropriate properties to be used in collaborative filtering recommender systems.

Numerous contributions devoted to that topic, offering diverse tools and approaches, have appeared recently in the literature (just to mention some recent results, see e.g., Lee & Lee, 2015; Mazurowski, 2013; Porcel, Lopez-Herrera, & Herrera-Viedma, 2009; Son, 2014). Some work concerning the confidence of proposed recommendations have already been done (see Mazurowski, 2013). Methods proposed in the literature, in general, focus on providing the final recommendation to the user. Such approaches need some assumptions about the whole recommendation strategy. However, we can imagine a very common situation when two items are rated by significantly different number of users. The decision how to deal with aggregation of ratings for such item has to be made before providing a user with a final recommendation. Another point concerns the case of different degree of knowledge or experience connected with every user. One may ask, weather we should treat ratings given by the user who has experience with 100 products equally to the rating given by the user who knows only two items? There are some contributions proposing trust-aware recommendations (see Martinez-Cruz, Porcel, Bernabé-Moreno, & Herrera-Viedma, 2015; Moradi & Ahmadian, 2015), but these approaches do not take into consideration the experience of the user with respect to his previous products history.

In this paper, we propose some new tools like entropy-based similarity or a graphical method for comparing recommendations, that may be considered interesting and turn out useful in solving mentioned problems.

3. The evaluation criteria for vague preferences comparison measures

Let us briefly discuss some conditions that each reasonable similarity measure for comparing two (vague) preference systems should possess. Here the word "similar" has an informal meaning but widespread in a common usage. Although each person can specify his own requirements, it seems that the following two conditions are obvious:

- **(C-1)** A similarity measure between preference systems A and B takes its maximal value if and only if A and B are perfectly concordant rankings.
- **(C-2)** A similarity measure between preference systems A and B takes its minimal value if and only if A and B are perfectly discordant rankings.
- **(C-3)** A similarity measure between two preference systems A and B is larger than between C and D if and only if a correlation between A and B is stronger than between C and D.

The third requirement corresponds to the intuitive property that similarity between two preference systems decreases as they differ more and more. Hence the next condition relates to monotonicity with respect to assumed correlation measure. Since the traditional correlation coefficients cannot be applied for vague preferences, later on we will use the generalized versions of the classical Spearman's ρ and Kendall's τ proposed in Grzegorzewski (2009) and Grzegorzewski and Ziembińska (2011).

Before we examine possible candidates for a similarity measure between two vague preference systems we have to adopt a mathematical model for vague preferences. However, since our model utilizes IF-sets, first let us briefly recall basic information on IF-set theory.