Modeling Vague Preferences in Recommender Systems

Paweł P. Ładyżyński[1], and Przemysław Grzegorzewski[2,3]

[1] Institute of Computer Science, Polish Academy of Sciences, ul. Jana Kazimierza 5, 01-248 Warsaw, Poland
[2] Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland
[3] Warsaw University of Technology, Faculty of Mathematics and Information Science, ul. Koszykowa 75, 00-662 Warsaw, Poland

Abstract.

Efficient analysis of consumer's preferences is a crucial problem in recommender systems. However, in practice we often have to deal with different types of vagueness of the data. Due to the large number of products in the databases, the knowledge of each user is usually incomplete and the ratings are often uncertain (i.e. the same ratings for different products). In this paper we discuss several IF-sets based methods, that are helpful in vague preferences modeling and use full available knowledge about user preferences, to support customers decisions in most appropriate way. We show how to improve algorithms applied in recommender systems using these methods.

We also show some IF-set based modifications of the probabilistic models applied in the instance-based label ranking algorithms, which improve their performance and make them applicable in content-based recommender systems.

Finally, we 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, e.g. corresponding to his attitude to risk.

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. Two basic architectures of recommender systems may be highlighted: 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). The problem of preference modeling is common in both types of recommender systems. These preference systems in real world applications are usually vague due to the large number of products rated by users with only partial knowledge about the whole set of analyzed items. In this paper we show, that IF-set based model can be a very useful tool while representing vague preferences and can be successfully applied in recommendation algorithms.

In this contribution we also discuss the choice of a recommendation strategy. The majority of methods proposed in the literature, in general, focus on providing the final recommendation to the user. The final recommendation strategy is assumed at the level of designing the recommender system, which means that it must be chosen to fit the individual decision making strategies of the users. This is rather a difficult problem. 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, whether 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 2 items? We can use i.e. logarithmic weighting to differentiate the impact of more and less competent users, but if a product was rated only by inexperienced users, the overall rating of it, without any additional information, might still be misleading . Providing user with some aggregated value of the level of experience of people who rated this item seems to be a natural solution to this problem. There are also some contributions proposing trust-aware recommendations (see [1],[2]), 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. We show how to use this new idea in collaborative filtering, and how to apply it to build computationally efficient predictive models in content-based recommender systems.

2 Modeling preference systems

Let $\mathbb{X} = \{x_1, \ldots, x_n\}$ denote a set of objects (e.g. films, books or other goods). A user A is associated to a vector $R_A = (A_1, \ldots, A_n)$, where A_i describes a position of element x_i among all other elements in $\mathbb{X}$ in his preference system according to objects from $\mathbb{X}$.

If all elements $\{x_1, \ldots, x_n\}$ create the total order (complete information and no ties), then R_A is just a ranking. For example, if we get the following vector $R_A = (1, 3, 4, 2, 5, 6)$ then it means that in A's opinion the most favorite element in the set $\{x_1, \ldots, x_6\}$ is x_1, the next one is x_4, then we have x_2, x_3, x_5, and the worst is x_6.

However, in real databases used in recommender systems, due to the large number of products, users knowledge about all of them is limited and their rankings may be incomplete or some elements may be indistinguishable. Thus, in general, a vector R_A may be not a representation of the total order of n objects. Suppose, e.g., that we get $R_A = (1, 2, 3, \mathrm{NA}, \mathrm{NA}, 2)$. In this case x_1 is the most favorite element. Then observer A indicates two elements x_2 and x_6 but he cannot decide which one of these two objects is better. The next one is x_3 and finally, there are two non-classified elements x_4 and x_5, described by NA, (i.e. "not available"). Further on, we will reserve the word "ranking" only to vectors describing linearly ordered elements. A general case that allows also partial orders we will call a preference system.

There are several methods of dealing with preference systems with missing information and indistinguishable elements. One possibility is omit such data. Appropriate imputation method to transform a vague preference system into a ranking may also be used. The first method leads to loss of information about the amount of knowledge possessed by users, whereas the second may be criticized for unavoidable subjectivity. Thus, we propose to use the model admitting vague preference systems that was proposed by Grzegorzewski (see [3],[4]). This model deals with both well-ordered items, possible ties, missing ranks and non-comparable elements. The key idea in construction proposed in [3],[4] is to represent a vague preference system by the appropriate IF-set. Due to such kind of representation, we can take advantage of the broad apparatus of mathematical methods defined for IF-sets.

Let $\mathbb{U}$ denote a usual set, called the universe of discourse. An IF-set (Atanassov's intuitionistic fuzzy set, see [5]) is given by a set of ordered triples $\tilde{C} = \{(u, \mu_{\tilde{C}}(u), \nu_{\tilde{C}}(u)) : u \in \mathbb{U}\}$, where $\mu_{\tilde{C}}, \nu_{\tilde{C}} : \mathbb{U} \to [0, 1]$ stand for the membership and nonmembership functions, respectively. It is assumed that $0 \le \mu_{\tilde{C}}(u) + \nu_{\tilde{C}}(u) \le 1$ for each $u \in \mathbb{U}$.

Consider any finite set of objects $\mathbb{X} = \{x_1, \ldots, x_M\}$. Given any user A let us define two functions $w_x, b_x : \mathbb{X} \to \{0, 1, \ldots, M-1\}$ as follows: for each $x_i \in \mathbb{X}$ let $w_A(x_i)$ denote a number of elements in $\mathbb{X}$ surely worse than x_i, while $b_x(x_i)$ let denote a number of elements surely better than x_i, with respect to the preference related to user A. Next let

$$\mu_{\tilde{A}}(x_i) = \frac{w_A(x_i)}{M-1}, \quad \nu_{\tilde{A}}(x_i) = \frac{b_A(x_i)}{M-1}. \tag{1}$$

denote a membership and nonmembership function, respectively, of the IF-set $\tilde{A} = \{(x_i, \mu_{\tilde{A}}(x_i), \nu_{\tilde{A}}(x_i)) : x_i \in \mathbb{X}\}$ describing the preference system connected with user A.

Using above representation (1), the following vector corresponding to the preference system of one of users $R_1 = (1, 2, \mathrm{NA}, 2,\ 3, \mathrm{NA})$ can be represented in the form of an IF-set where the values for membership function are equal to $\mu_{\tilde{R}_1} = (0.6, 0.2, 0, 0.2, 0, 0)$ and the non-membership function $\nu_{\tilde{R}_1} = (0, 0.2, 0, 0.2, 0.6, 0)$ respectively for elements $x_1, \ldots, x_6$.

3 Preferences in Collaborative Filtering - Finding Similar Users

Measuring similarity between preferences is a crucial problem for collaborative filtering recommender systems. This task becomes significantly harder when preferences are incomplete or somehow vague.

In previous section we have shown the way of modeling preference systems by IF-sets. Since our main goal is to compare preference systems hence one may ask about methods for IF-sets comparison. This topic seems to be interesting not only in our context. Three general types of comparison measures were discussed in [6]: IF-distances, IF-dissimilarities and IF-divergences, and some relationships between them were also examined.

In [7], [8], we discussed the problem of choosing similarity measure between preference systems, with appropriate properties to be applied in collaborative filtering recommender systems. Below we mention the list of requirements that were proposed in [8]:

(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.

After analyzing several types of similarity measures, we propose two similarity measures with desired properties proved in [8]:

$$S_E(R_1, R_2) = 1 - \sqrt{\frac{3(n-1)}{n(n+1)}} D_E(\tilde{R}_1, \tilde{R}_2) \tag{2}$$

and

$$S_H(R_1, R_2) = 1 - \frac{2(n-1)}{n^2} D_H(\tilde{R}_1, \tilde{R}_2). \tag{3}$$

where

$$D_E(\tilde{R}_1, \tilde{R}_2)^2 = \frac{1}{2}\sum_{i=1}^{n}((\mu_{\tilde{R}_1}(x_i) - \mu_{\tilde{R}_2}(x_i))^2 + (\nu_{\tilde{R}_1}(x_i) - \nu_{\tilde{R}_2}(x_i))^2), \tag{4}$$

and

$$D_H(\tilde{R}_1, \tilde{R}_2) = \frac{1}{2}\sum_{i=1}^{n}(|\mu_{\tilde{R}_1}(x_i) - \mu_{\tilde{R}_2}(x_i)| + |\nu_{\tilde{R}_1}(x_i) - \nu_{\tilde{R}_1}(x_i)|). \tag{5}$$

Both (2) and (3) reach their maximal values if and only if R_1 and R_2 are perfectly discordant. However, although (2) reaches its minimal value if and only if R_1 and

R_2 are perfectly concordant, (3) reaches its minimal value not only for perfectly discordant preference systems.

According to property (C-3), we compared the behavior of measures (2) and (3) with the generalized Kendall's τ and Spearman's ρ defined in [4], [9]. The generalized Kendall correlation coefficient ([4]) is defined as follows

$$\tilde{\tau} = \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} [sgn(\mu_A(x_j) - \mu_A(x_i)) \cdot sgn(\mu_B(x_j) - \mu_B(x_i)) \tag{6}$$
$$+ sgn(\nu_A(x_j) - \nu_A(x_i)) \cdot sgn(\nu_B(x_j) - \nu_B(x_i))],$$

The generalized Spearman coefficient ([9]) is defined by

$$\tilde{r}_s(A,B) = 1 - \frac{3(n-1)}{n(n+1)} \sum_{i=1}^{n} [(\mu_A(x_i) - \mu_B(x_i))^2 + (\nu_A(x_i) - \nu_B(x_i))^2]. \tag{7}$$

In [8], we proved the following property of measures S_E and S_H:

Proposition 1. *Let $A, B, C, D \in \mathbb{IFS}(\mathbb{X})$ describe preference systems with respect to elements of $\mathbb{X} = \{x_1, \ldots, x_n\}$. If $\tilde{r}_s(A,B) \leq \tilde{r}_s(C,D)$, then $S_E(A,B) \leq S_E(C,D)$.*

and the following lemma for measure S_E:

Lemma 1. *Let $A, B, C, D \in \mathbb{IFS}(\mathbb{X})$ describe preference systems with respect to elements of $\mathbb{X} = \{x_1, \ldots, x_n\}$. If $\tilde{r}_s(A,B) < \tilde{r}_s(C,D)$, then $S_E(A,B) < S_E(C,D)$.*

From 1 and 1 we can observe that measures (2), (3) posses desired properties connected with (C-3).

As measures (2), (3) pretend to behave properly in recommender system environment, we can consider the further steps of creating recommendations. The simplest way of recommending a new item to a user A is to find another users (say $B_1, \ldots, B_k$) with preferences similar to A and to suggest A some resources highly preferred by $B_1, \ldots, B_k$ which are yet not known to A.

However, during our experiments, we noticed that the situation where several users have identical preference systems (also no additional products known by some of $B_1, \ldots, B_k$) is quite common. What is more, in [8], we proved the following property of measure S_H:

Proposition 2. *Let R_1 and R_2 denote two preference systems with respect to n objects from the set $\mathbb{Y} = \{x_1, \ldots, x_n\}$. Suppose, that at least one element of $\mathbb{Y}$ got no opinion according to both preference systems R_1 and R_2. Moreover, let R_2^* denote a preference system which is identical to R_2 up to one element $x_i^* \in \mathbb{Y}$ which is ranked according to R_2^* but not considered by R_1 and R_2. Then*

$$0 \leq S_H(R_1, R_2) - S_H(R_1, R_2^*) < \frac{2}{n}. \tag{8}$$

The simple deduction from proposition 2 is, that this measure does not promote users who know a lot about many different products not yet known by A. In fact, we can not prove similar property for measure S_E (it does not hold in general) however, in [8] we show the experimental evaluation based on 10 million randomly generated experiments (different parameters i.e. number of products, fraction of missing values were considered) to analyze the distribution of $(S_E(R_1, R_2) - S_E(R_1, R_2^*))$. The results of our experiment show that for measure S_E, the inequality from Proposition 2 does not hold for less than 2% cases (see fig. 1, 2, detailed description can be found in [8]).

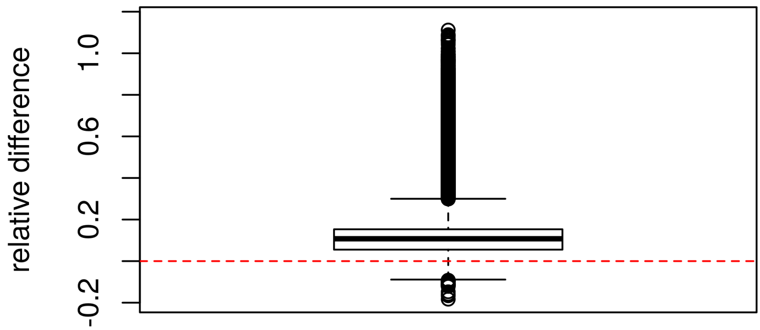

Fig. 1. A simulation for measure S_E. A boxplot for $\mathrm{Diff}(R_1, R_2, R_2^*)$ obtained for 10 millions of random pairs of preference systems, fraction of missing ranks: q=0.5.

Therefore, as we are interested in finding users not only similar to A but who also differ from A in a sense that they could provide an information on items not known yet by A, we had to modify these measures to make them possess the property of promoting those customers similar to a new user, who have a broad knowledge on the items not seen yet by this new user. The general idea is to modify the similarity measures by including some penalty connected with those users whose knowledge is not sufficient. We decided to take an advantage from the two main types of the entropies that appear in the IF-set environment. The first one is connected with the fuzziness of given IF-set, while the second with the hesitancy and the lack of knowledge connected with this IF-set (see i.e. [10], [11]). We use the following definition of two-tuple entropy (see [11]):

Definition 1. *Let $E_F, E_{HLK} : \mathbb{IFS}(\mathbb{X}) \to [0,1]$ denote two mappings. A pair (E_F, E_{HLK}) is said to be a two-tuple entropy if E_F and E_{HLK} satisfy the following conditions:*

(i) $E_F(A) = 0$ *if and only if A is crisp or $\mu_A(x) = \nu_A(x) = 0$ for every $x \in A$,*
(ii) $E_F(A) = 1$ *if and only if $\mu_A(x) = \nu_A(x) = 0.5$ for every $x \in A$,*

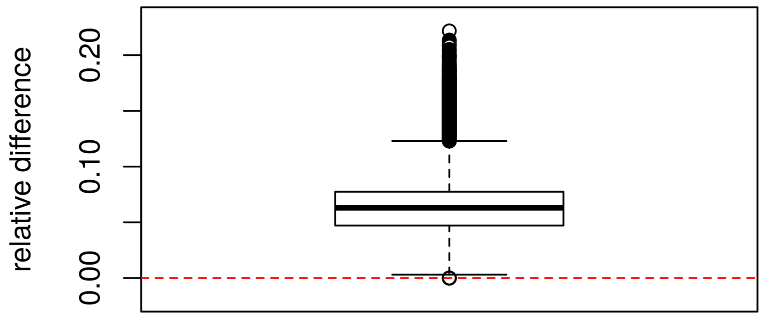

Fig. 2. A simulation for measure S_E. A boxplot for Diff(R_1, R_2, R_2^*) obtained for 10 millions of random pairs of preference systems, fraction of missing ranks: q=0.3.

(iii) $E_F(A) = E_F(A^C)$, *where* $A^C = \{(x, \nu_A(x), \mu_A(x)) : x \in \mathbb{X}\}$ *is the complement of* A,
(iv) $E_F(A) \leq E_F(B)$ *if* $\mu_A(x) \leq \mu_B(x) \leq 0.5$ *and* $\nu_A(x) \geq \nu_B(x) \geq 0.5$ *for* $\mu_B(x) \leq \nu_B(x)$ *or if* $\mu_A(x) \geq \mu_B(x) \geq 0.5$ *and* $\nu_A(x) \leq \nu_B(x) \leq 0.5$ *for* $\mu_B(x) \geq \nu_B(x)$,
(v) $E_{HLK}(A) = 0$ *if and only if* $A \in FS(\mathbb{X})$,
(vi) $E_{HLK}(A) = 1$ *if and only if* $\mu_A(x) = \nu_A(x) = 0$,
(vii) $E_{HLK}(A) = E_{HLK}(A^C)$,
(viii) $E_{HLK}(A) \geq E_{HLK}(B)$ *if* $\mu_A(x) + \nu_A(x) \leq \mu_B(x) + \nu_B(x)$ *for every* $x \in \mathbb{X}$.

It is seen that E_F is strictly related to fuzziness while E_{HLK} is connected with the hesitancy and the lack of knowledge. In [8], we propose to use the following expression:

$$E_{HLK}(A) = \frac{1}{n} \sum_{i=1}^{n} [1 - (\mu_A(x_i) + \nu_A(x_i))], \tag{9}$$

and we show how to decompose it into two parts where one is connected with lack of knowledge and would later be used as a penalty for users whose knowledge is not rich enough:

Proposition 3. *Let* $\tilde{R} \in \mathbb{IFS}(\mathbb{X})$ *describe a preference system* R *with respect to* n *objects from the set* $\mathbb{X} = \{x_1, \ldots, x_n\}$. *Suppose that* t *different ranks* $(1 < t \leq n)$ *were attributed to elements of* $\mathbb{X}$ *in such way that* k_i *denote the number of elements which obtained* i*-th rank according to* R. *Moreover, let* m *denote the number of objects in* $\mathbb{X}$ *not ranked according to* R, *where* $m + \sum_{i=1}^{t} k_i = n$. *Then*

$$E_{HLK}(\tilde{R}) = E_H(\tilde{R}) + E_{LK}(\tilde{R}), \tag{10}$$

where

$$E_H(\tilde{R}) = \frac{1}{n}\sum_{i=1}^{t}\frac{k_i(k_i-1)}{n-1}, \tag{11}$$

$$E_{LK}(\tilde{R}) = \frac{1}{n}\left[\frac{(n-m)m}{n-1} + m\right]. \tag{12}$$

In proposition 12, E_{LK} quantifies the lack of knowledge impact connected with unavailable opinions on evaluated objects. Thus we proposed in [8] the following form of modified S_E measure, including some penalty for not sufficient knowledge about many different products:

$$S_E^{pen}(R_1, R_2) = S_E(R_1, R_2)\cdot(1 - E_{LK}(\tilde{R}_2)). \tag{13}$$

Now, the proposed measure (13) promotes, in the process of finding users similar to A due to their preference systems, users with broad knowledge about many products. Such tool is quite satisfactory to be used in collaborative filtering. The next important step of creating recommendation is reasoning from multiple preference systems of several users, which are the most similar to the new user A due to measure S_E^{pen}. The question how to aggregate different preferences about product not yet known by A is not trivial. Formally, we can express the preferences of chosen users, say $B_1, \ldots, B_k$, represented by appropriate IF-sets $\tilde{B}_1, \ldots, \tilde{B}_k$, in the form of their membership and non-membership functions $(\mu_{\tilde{B}_j}(x_1), \mu_{\tilde{B}_j}(x_2), \ldots, \mu_{\tilde{B}_j}(x_n))$ and $(\nu_{\tilde{B}_j}(x_1), \nu_{\tilde{B}_j}(x_2), \ldots, \nu_{\tilde{B}_j}(x_n))$ respectively for $j = 1, \ldots, k$.

Further performance of the recommender system depends strongly on the choice of the aggregation method for these preferences and the final recommendation strategy. What is more, the strategy of decision making may be very specific for different group of users, which may affect the overall accuracy of the system. Exemplary "ready to use" algorithms of creating final recommendation were mentioned in [8]. However, we also proposed in [8] the new idea of a graphical tool that summarizes properties of possible recommendations and may be used in the form of interaction with users to let them choose the recommendation which best fits their characteristic of individual decision making strategy. It may also be helpful for the designers of fully automatic recommender systems to analyze different possible algorithms and choose the one, that is fitted to the specificity of the group of users they consider. The idea of that graph is to provide the user with a value of a special score function calculated for different items together with some information on the strength (or credibility) of the score. Exemplary results of summarizing two possible recommendations using proposed method can be seen in fig. 3.

On the vertical axis of Figure 3 we place the aggregated values of μ and ν functions in the form of interval valued fuzzy set, i.e. $[\mu_{agg}^L(x_i), \mu_{agg}^R(x_i)]$, where $\mu_{agg}^L(x_i) = \mu_{agg}^A(x_i)$ and $\mu_{agg}^R(x_i) = 1 - \nu_{agg}^A(x_i)$ (since IF-sets are isomorphic with interval-valued fuzzy sets, see e.g., [12]), where

$$\mu_{agg}^{A,k}(x_i) = \frac{1}{\sum_{j=1}^{k} \mathbb{I}(\mu_{\tilde{B}_j}(x_i) \vee \nu_{\tilde{B}_j}(x_i) > 0)} \sum_{j=1}^{k} \mu_{\tilde{B}_j}(x_i), \tag{14}$$

$$\nu_{agg}^{A,k}(x_i) = \frac{1}{\sum_{j=1}^{k} \mathbb{I}(\mu_{\tilde{B}_j}(x_i) \vee \nu_{\tilde{B}_j}(x_i) > 0)} \sum_{j=1}^{k} \nu_{\tilde{B}_j}(x_i), \tag{15}$$

and $\mathbb{I}(\cdot)$ denotes the indicator.

This interval can be interpreted as the aggregated degree of membership to the set of highly preferred items. Due to the different level of knowledge of several users, the information they provide about products they already know may be more or less precise. The general interpretation is that the thinner is the interval, the more experienced users (with knowledge about many different products) rated the product and thus, the recommendation is more trustful. Quite a different aspect is the confidence of the recommendation. Even if the product was rated by experienced users, but only a small number of them, we can suspect that the recommendation does not have appropriate level of confidence. Thus, on the horizontal axis, we present the fraction of nearest neighbors that have any knowledge about product presented in the graph (in our example 0.4 NN know product x_1 and 0.8 of NN know product x_2). In fig. 3, we can observe that x_2 is more confident recommendation, which is also known by users with higher amount of knowledge than x_1 (thinner interval). It is worth noticing that the interval reduces to the single point if and only if this item is rated by all users "similar" to our customer, it is considered on the same position by all of them and they have full knowledge about all products in the database. On the other hand, product x_1, can still be potentially the best product in our data set ($\mu_{agg}^{R}(x_1)$=1 means that due to the users that know x_1, none of other products is better than it). Considering this example, the optimistic person with low aversion to risk would probably choose product x_1 (due to available knowledge), whereas users, who need more confident recommendations, would choose x_2.

In [8], we show the experimental results of evaluation of collaborative filtering recommender system based on similarity measure (13). We analyzed 3 possible different strategies of decision making and the simulated user interaction strategy with usage of proposed graphical tool. The results were promising, so in this contribution, we decided to apply this method in content based recommender system, and compare the results with one of well performing algorithms we proposed in [13].

4 Modeling Preferences in Content Based Recommender Systems

We will now focus on creating recommendations in a different situation, where some meta-data about users of a recommender systems are available. Let $\mathbb{U}$,

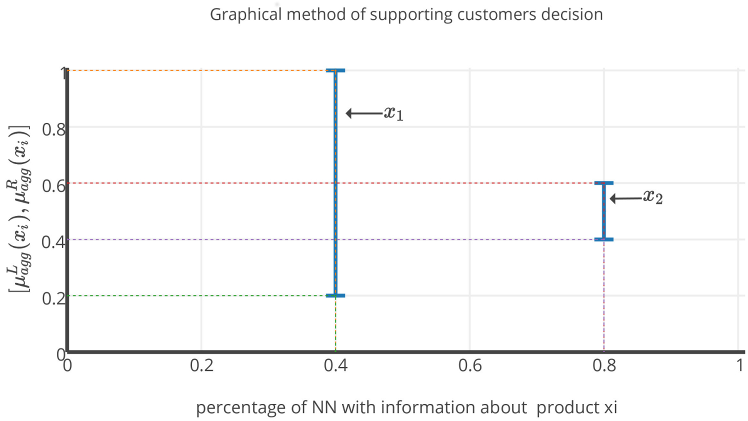

Fig. 3. Graphical method for supporting customer's decision

called an instance space, denote a set of elements (users, patients etc.) characterized by several attributes. Suppose that instead of classifying instances into separate classes, we associate each instance $u \in \mathbb{U}$ with a total order of all class labels $\mathbb{Y} = \{y_1, \ldots, y_M\}$. Moreover, we say that $y_i \succ_u y_j$ indicates that y_i is preferred to y_j given the instance u. A total order $\succ_u$ can be identified with a permutation π_u of the set $\{1, \ldots, M\}$, where $\pi_u(i)$ is the index j of the class label y_j put on the i-th position in the order. The class of permutations of $\{1, \ldots, M\}$ will be denoted by Ω.

The main goal required while creating recommendation is to predict a ranking of labels $y_1, \ldots, y_M$ for a new instance u, given some instances with known rankings of labels as a learning set. In practical issues, especially in recommender systems where the amount of available products is large, preference on instances known from the learning set does not usually contain all labels, i.e our information is of the form $y_{\pi_u(1)} \succ_u \ldots \succ_u y_{\pi_u(k)}$, where $k < M$.

Several methods are available in the literature but many of them are computationally exhaustive in a presence of vague data. In [13], we proposed the IF-set modification of an algorithm based on the Mallows model. The proposed modification effected in significant improvement in performance.

Below, we highlight some important details of a mentioned algorithm.

We may assume that every instance is associated with a probability distribution over Ω, i.e. for each instance $u \in \mathbb{U}$ there exists a probability distribution $\mathbb{P}(\cdot|u)$ such that, for every $\pi \in \Omega$, $\mathbb{P}(\pi|u)$ is the probability that $\pi_u = \pi$.

To evaluate the predictive performance of a label ranker a suitable loss function on Ω is needed, e.g. based on Kendall's tau (see [14]).

The Mallows model is a distance-based probability model defined by

$$\mathbb{P}(\pi|\theta, \pi_0) = \frac{exp(-\theta D(\pi, \pi_0))}{\phi(\theta)}, \tag{16}$$

where the ranking $\pi_0 \in \Omega$ is the location parameter (center ranking), D is a distance measure on rankings, $\phi = \phi(\theta)$ is a constant normalization factor and θ stands for a spread parameter which determines how quickly the probability decreases with the increasing distance between π and π_0.

The main idea of our modification proposed in [13] is to replace measure D in (16) with a substitute that admits vague data.

Having any two instances $u_1, u_2 \in \mathbb{U}$ we may compute a correlation between preference systems $\tilde{u}_1, \tilde{u}_1$ generated by these instances, using the generalized Kendall's tau, admitting incomplete preferences (see [4]):

$$\tilde{\tau} = \frac{1}{2M(M-1)} \sum_{i=1}^{M} \sum_{j=1}^{M} [sgn(\mu_{\tilde{u}_1}(y_j) - \mu_{\tilde{u}_1}(y_i)) \cdot sgn(\mu_{\tilde{u}_2}(y_j) - \mu_{\tilde{u}_2}(y_i)) \tag{17}$$
$$+ sgn(\nu_{\tilde{u}_1}(y_j) - \nu_{\tilde{u}_1}(y_i)) \cdot sgn(\nu_{\tilde{u}_2}(y_j) - \nu_{\tilde{u}_2}(y_i))].$$

For possibly incomplete preferences we get incomplete permutation $\tilde{\pi} = \tilde{\pi}_u$ which might be identified with the corresponding IF-set $\tilde{u}$. Thus for any two instances $u_1, u_2 \in \mathbb{U}$ we have $\tilde{\tau} = \tilde{\tau}(\tilde{u}_1, \tilde{u}_2) = \tilde{\tau}(\tilde{\pi}_1, \tilde{\pi}_2)$. Hence, using (17), we may consider the following measure

$$D_{\tilde{\tau}}(\tilde{\pi}_1, \tilde{\pi}_2) = \frac{1 - \tilde{\tau}(\tilde{\pi}_1, \tilde{\pi}_2)}{2}, \tag{18}$$

which seems to be useful in the generalized Mallows model (16) admitting incomplete rankings and defined as follows

$$\tilde{\mathbb{P}}(\tilde{\pi}|\theta, \tilde{\pi}_0) = \frac{exp(-\theta D_{\tilde{\tau}}(\tilde{\pi}, \tilde{\pi}_0))}{\phi(\theta)}. \tag{19}$$

Of course, when modeling preferences by IF-sets one can also consider other substitutes for the measure D in (16), including different distances, dissimilarity measures or divergences (see, e.g., [6]). However, we have chosen a measure based on the generalized Kendall's tau because it is common to use distances utilizing the classical Kendall's coefficient in the Mallows model (see, e.g., [14]).

One of proposed in [13] algorithms can be described as follows:

Algorithm 1 ***Mallows Best Probability Algorithm (MBP)***
{***Input****: u - new instance, U - learning set of instances,* $\bar{\pi}$ *- permutations of labels connected with instances, k - number of nearest neighbors*}
1. Find k nearest neighbors of u in U.
2. For (j in $1 : M$*) calculate* $\sum_{\pi^* \in \bar{\pi}_{kNN(u)}} \tilde{\mathbb{P}}(y_j^{best}|\theta, \pi^*)$
3. MBP-rank $<-$ *Sort labels according to the values obtained in step 2 (in case of ties a label with lower index is better in the ranking).*
{***Output****: MBP-rank*}

where

$$\tilde{\mathbb{P}}(y_j^{best}|\theta, \pi^*) = \frac{exp(-\theta D^*(y_j^{best}, y_j^{\pi^*}))}{\phi(\theta)}, \quad (20)$$

where D^* is the Euclidean distance between IF-sets given by

$$D^*(y_j^{best}, y_j^{\pi^*}) = \sqrt{\frac{1}{2}\sum_{i=1}^{n}((\mu_{y_j^{best}} - \mu_{\pi^*}(y_j))^2 + (\nu_{y_j^{best}} - \nu_{\pi^*}(y_j))^2)}. \quad (21)$$

as we apply the Mallows model to express the probability corresponding to the best label.

The performance of proposed algorithm we show in Tables 1, 2 (see. [13]) for details).

Table 1. Comparison of label ranking algorithms for $p = 30\%$ missing labels in the learning set.

	accuracy			time [s]		
data set	IBLR	MBP	MMBP	IBLR	MBP	MMBP
glass (A)	0.781	0.784	0.788	3.504	0.26	3.7
vowel (A)	0.817	0.795	0.819	102.03	1.05	102.26
housing (B)	0.670	0.665	0.670	8.44	0.70	8.95
elevators (B)	0.622	0.617	0.624	1371.86	225.83	1583.55
wisconsin (B)	0.432	0.420	0.427	316.12	0.40	319.54
average	0.664	0.656	0.665	360.39	45.65	403.60

Table 2. Comparison of label ranking algorithms for $p = 50\%$ missing labels in the learning set.

	accuracy			time [s]		
data set	IBLR	MBP	MMBP	IBLR	MBP	MMBP
glass (A)	0.688	0.685	0.687	5.12	0.29	5.42
vowel (A)	0.725	0.700	0.715	119.84	0.95	126.04
housing (B)	0.579	0.570	0.573	12.53	0.7	13.12
elevators (B)	0.540	0.530	0.535	2326.23	272.67	2598.56
wisconsin (B)	0.381	0.351	0.363	502.22	0.37	508.74
average	0.583	0.567	0.575	593.19	55.00	650.38

Results given in Table 1 and Table 2 show that algorithms MBP, MMBP and IBLR have similar accuracy on our experimental sets. More precisely, MBP is usually slightly worse than the two other algorithms, but it is significantly faster which is crucial due to applications in recommender systems.

4.1 Fitting decision strategy to the user

One may notice that the decision making strategy in mentioned MBP algorithm is a strategy that we choose in the process of designing the algorithm. It gives satisfactory overall results, but the rule how to choose final recommendation is assumed without any survey about the users of a recommender system.

In [8], we considered three common strategies of decision making based on proposed graphical tool:

- Strategy 1: choose the product with μ_R^{agg} not lower than 0.5 with maximal p_{nn} and minimal difference between μ_R^{agg} and μ_L^{agg} - this strategy would fit users with the strongest aversion to risk. Taking into consideration highest p_{nn} and lowest size of the interval $[\mu_L^{agg}, \mu_R^{agg}]$ is equivalent to base the recommendation on decision of the largest possible number of the most experienced users (with the highest level of knowledge about many different products).
- Strategy 2: choose the product with maximal μ_R^{agg} - is a kind of the risky, optimistic strategy, that takes into consideration only the highest possible value of the μ function for the unknown product.
- Strategy 3: choose the product with maximal μ_L^{agg} - is a kind of the pessimistic strategy, that maximizes only the lowest possible value of the μ function for the unknown product.

We will now adapt the proposed graphical tool for valuating the possible recommendations. We will show what improvement can be obtained by using different strategies of creating recommendation and we will compare the results with MBP algorithm.

To perform our experiments, we use semi-synthetic label ranking datasets downloaded from `www.cs.uni-paderborn.de/fachgebiete/intelligente-systeme/software/label-ranking-datasets.htm`. Below, we extend our experimental evaluation, presented in [8], to analyze the behavior of discussed methodology (see [8] for more detailed description). This time, as we consider content-based recommendations, we combine preference systems and vector of attributes from *wisconsin* and *vowel* datasets randomly (assuming that labels and attributes from wisconsin dataset are always after labels and attributes from vowel dataset) to obtain dataset containing 528 instances with 37 attributes, where each instance is connected with ranking of 27 labels. We then generate the missing ranks (every element in each ranking is removed with probability 0.5). To analyze the performance and behavior of proposed method, we use leave-one-out cross-validation. For every observation A, we first find it's 5 nearest neighbors using all 37 attributes. After finding the set of nearest neighbors, we specify the IF-set representations of their preference systems and the aggregated values of μ and ν functions for every label using the formulas (14,15) are calculated. Having all statistics to use proposed graphical method and perform strategies $(1-3)$, we begin our experiment, taking into consideration also the best recommendation from MBP algorithm. The accuracy is calculated using the following formula:

$$acc_A(y_i) = 1 - \frac{1}{2}\sqrt{(\mu_A^*(y_i) - \mu_L^{agg}(y_i))^2 + (\nu_A^*(y_i) - (1 - \mu_R^{agg}(y_i)))^2}, \qquad (22)$$

where $\mu_A^*(y_i)$ and $\nu_A^*(y_i)$ are the values obtained by representing the preference system for a given user A in a form of IF-set after inputing the true rank for label y_i (from the learning set before the process of random removing the ranks) into his preference system.

Results are presented in Table 3. The simulation of the user interaction is obtained by choosing the strategy with the highest accuracy for every user and compare the results with the case when the recommendation strategy is fixed for every user. As MBP is rather too complicated strategy to be adapted individually by the user, it is not included in "simulated user strategy".

Table 3. Values of averaged accuracy of recommendations for different recommendation strategies. "Acc simulated user" means the averaged accuracy of recommendation based on the best possible strategy for each user.

acc str1	acc str2	acc str3	acc MBP	acc simulated user
0.9133	0.8461	0.9056	0.9144	0.9343

We may notice that the simulation of user-interactive strategy gives the best results.

The results of analogical experiment in collaborative filtering context can be seen in Table 4 (see [8] for details).

Table 4. Values of averaged accuracy of recommendations for different recommendation strategies. "Acc simulated user" means the averaged accuracy of recommendation based on the best possible strategy for each user.

acc str1	acc str2	acc str3	acc simulated user
0.9047	0.8884	0.8967	0.9140

An interesting observation is the difference between accuracy for the same strategies with and without the information brought from the meta-data connected with our instances. The difference in the accuracy between the worst and the best strategy increases when we use the meta-data.

5 Conclusions and Future Work

In this paper we discuss several IFset based methods of dealing with vague preferences in different types of recommender systems. We show how the graphical tool of summarizing recommendations proposed in [8] can be applied to improve accuracy of content based recommender systems.

Some questions remain open and deserve future research. In particular, a natural desired extension of the proposed graphical method for comparing recommendations would be its special implementation in the form of the automatic user-adaptive algorithm. Such algorithm would learn user's decision-making strategy in order to propose him the appropriate recommendations automatically, even without his influence. We can imagine at least two types of data that can be used to train such algorithm. In the first case, the algorithm would learn the behavior of the user from his historical choices, e.g. from the proposals of selected recommendation presented in a graphical form. The second approach requires some meta-data connected with the user, like results of a psychological survey concerning his behavior in decision-making. Basing on such data we could deduce whether the user prefers something risky but with possible highest rates or a medium rated but well checked product.

Concerning the content based recommender systems, although the proposed MBP algorithm seems to be a promising candidate for creating recommendations especially in the presence of large number of labeled items, one may consider application of the proposed graphical tool to design global predictive models that would be more computationally efficient in the prediction step. The proposed method based on finding nearest neighbors is rather exhaustive when we have to deal with large databases of users. In this case consideration of i.e. GLM based models for estimating borders of intervals presented in proposed graphical method, seem to be desired extension of proposed algorithms.

Acknowledgments

Study was supported by research fellowship within "Information technologies: research and their interdisciplinary applications" project co-financed by European Social Fund (agreement no. POKL.04.01.01-00-051/10-00). [15]

References

1. Martinez-Cruz C., Porcel C., B.M.J.H.V.E.: A model to represent users trust in recommender systems using ontologies and fuzzy linguistic modeling. Information Sciences **311** (2015) 102–118
2. Moradi, P., A.S.: A reliability-based recommendation method to improve trust-aware recommender systems. Expert Systems with Applications **42** (2015) 7386–7398
3. Grzegorzewski, P.: The coefficient of concordance for vague data. Computational Statistics & Data Analysis **51** (2006) 314–322

4. Grzegorzewski, P.: Kendall's correlation coefficient for vague preferences. Soft Computing **13** (2009) 1055–1061
5. Atanassov, K.: Intuitionistic fuzzy sets: Theory and applications. Springer-Verlag (1999) 55–62
6. Montes S., Iglesias T., J.V..M.I.: A common framework for some comparison measures of if-sets. IWIFSGN 2012 (2012)
7. P.P. Ł adyżyński, P. Grzegorzewski, .: Comparing vague preferences in recommender systems. Proceedings of the 8th conference of the EUROFUSE 2013 Workshop on Uncertainty and Imprecision Modelling in Decision Making (2013) 149–156
8. P.P. Ł adyżyński, P. Grzegorzewski, .: Vague preferences in recommender systems. Expert Systems with Applications (2015) In press
9. Ziembińska, P.G..P.: Spearman's rank correlation coefficient. LNAI **7022** (2011) 342–353
10. E., G.P..M.: Some notes on atanassov's intuitionistic fuzzy sets. Fuzzy Sets and Systems **156** (2005) 492–495
11. Pal N.R., Bustince H., P.M.M.U.G.D..B.G.: Uncertainties with atanassov's intuitionistic fuzzy sets: Fuzziness and lack of knowledge. Information Sciences **228** (2013) 61–746
12. P. Grzegorzewski, M.E.: On the entropy of intuitionistic fuzzy sets and interval valued fuzzy sets. Proceedings of the Tenth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (2004) 1419–1426
13. P.P. Ł adyżyński, P. Grzegorzewski, .: On incomplete label ranking with if-sets. Strengthening Links Between Data Analysis and Soft Computing, Springer (2014) 55–62
14. W. Cheng, K. Dembczynski, E.H.: Decision tree and instance-based learning for label ranking. ICML (2009)
15. W. Cheng, E.H.: A nearest neighbor approach to label ranking based on generalized labelwise loss minimization. International Joint Conference on Artificial Intelligence (IJCAI-13) (2013)