On Incomplete Label Ranking with IF-sets

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

[1] Interdisciplinary PhD Studies at the Polish Academy of Sciences,
Jana Kazimierza 5, 01-248 Warsaw, Poland
pawelladyz@wp.pl

[2] Systems Research Institute, Polish Academy of Sciences,
Newelska 6, 01-447 Warsaw, Poland
pgrzeg@ibspan.waw.pl

[3] Faculty of Mathematics and Information Science,
Warsaw University of Technology,
Koszykowa 75, 00-662 Warsaw, Poland

Abstract. Probabilistic models, like the Mallows model, are commonly used for label ranking. However, for incomplete preferences the existing methods are exhaustive in the learning step and therefore the applications of the Mallows model in practical label ranking problems or in recommender systems are limited. In this paper, we show how to improve the Mallows model using IF-sets so it may become more simple and more effective for analyzing vague preferences and creating recommendations.

Keywords: IF-sets, incomplete data, instance-based learning, label ranking, the Mallows model, recommender systems.

1 Introduction

Label ranking is an important task in many applications like information retrieval, rating products or recommender systems. It can be treated as a generalization of a classification problem, where, instead of a ranking of all labels, only a single label is requested as a prediction for given observation. Thus, in brief, the label ranking can be perceived as a problem of learning a mapping from instances to rankings over a finite set of predefined labels.

This problem can be solved in different ways. Existing methods often use binary classification algorithms so the ranking is obtained by pairwise comparisons (see [6]). Another approaches utilize probabilistic models defined on a class of all rankings. As prominent example one can mention the Mallows model [7].

In recommender systems, due to the large amount of rated items, we typically meet incomplete preferences for all users available in data bases. However, it is not obvious, how to cope with such incomplete or vague preferences. Therefore, we propose an algorithm that combines the Mallows model with IF-set theory to get an effective method of label ranking and create recommendations in the presence incomplete rankings.

The paper is organized as follows. In Sec. 2 we describe briefly the problem of instance based label ranking and the Mallows model. Next, in Sec. 3 we show how

P. Grzegorzewski et al. (eds.), *Strengthening Links between Data Analysis & Soft Computing*,
Advances in Intelligent Systems and Computing 315, DOI: 10.1007/978-3-319-10765-3_7

to apply IF-sets for modeling incomplete preferences and then how to enrich the classical Mallows model. Finally, in Sec. 4, we propose a new efficient algorithm for instance based label ranking and present results of the experimental study comparing our algorithm with other approaches.

2 Label Ranking and the Mallows Model

2.1 Basic Notions

Let $\mathbb{X}$, 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 $x \in \mathbb{X}$ with a total order of all class labels $\mathbb{Y} = \{y_1, \ldots, y_M\}$. Moreover, we say that $y_i \succ_x y_j$ indicates that y_i is preferred to y_j given the instance x.

A total order $\succ_x$ can be identified with a permutation π_x of the set $\{1, \ldots, M\}$, where π_x is defined such that $\pi_x(i)$ is the index j of the class label y_j put on the i-th position in the order. Hence, $\pi_x^{-1}(j) = i$ gives the position of the j-th label (see [2]). The class of permutations of $\{1, \ldots, M\}$ will be denoted by Ω.

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

The main goal in label ranking is to predict a ranking of labels $y_1, \ldots, y_M$ for a new instance x, 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 do not usually contain all labels, i.e our information is of the form $y_{\pi_x(1)} \succ_x \ldots \succ_x y_{\pi_x(k)}$, where $k < M$.

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 [2]).

2.2 The Mallows Model

Going back to the above mentioned probability distribution $\mathbb{P}(\cdot|x)$, we need a probabilistic model suitable for our considerations. In [2] the Mallows model was used in the context of an instance-based approach to label ranking.

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

$$\mathbb{P}(\pi|\theta, \pi_0) = \frac{exp(-\theta D(\pi, \pi_0))}{\phi(\theta)}, \tag{1}$$

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 label ranking problem is then solved by the maximum likelihood estimation connected with (1). In [2] parameters θ, π_0 are estimated using $\pi_1, \ldots, \pi_k$

rankings connected with k nearest neighbors of a new instance x in the training set. It works nicely when all rankings from the training set are complete. Unfortunately, such situation is unusual in the real world problems.

To handle incomplete rankings in the training data it was proposed in [2] to maximize the probability

$$\mathbb{P}(\pi|\theta, \pi_0) = \sum_{\pi^* \in E(\pi)} \mathbb{P}(\pi^*|\theta, \pi_0), \quad (2)$$

where $E(\pi)$ - set of linear extensions of π. However, calculations with (2) are rather exhaustive. Therefore, we suggest below another method based on IF-modeling of incomplete rankings proposed by Grzegorzewski (see [3,4]).

3 IF-sets and Incomplete Preferences

Let $\mathbb{U}$ denote a usual set, called the universe of discourse. An IF-set (Atanassov's intuitionistic fuzzy set, see [1]) 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 $u \in \mathbb{U}$.

In [3,4,5] Grzegorzewski proposed how to model preference systems admitting ties and missing ranks. The key idea is to represent a preference system by an appropriate IF-set. Consider any finite set of labels $\mathbb{Y} = \{y_1, \ldots, y_M\}$. Given any instance $x \in \mathbb{X}$ let us define two functions $w_x, b_x : \mathbb{Y} \to \{0, 1, \ldots, M-1\}$ as follows: for each $y_i \in \mathbb{Y}$ let $w_x(y_i)$ denote a number of elements in $\mathbb{Y}$ surely worse than y_i, while $b_x(y_i)$ let denote a number of elements surely better than y_i, with respect to the preference related to instance x. Next let

$$\mu_{\tilde{x}}(y_i) = \frac{w_x(y_i)}{M-1}, \quad \nu_{\tilde{x}}(y_i) = \frac{b_x(y_i)}{M-1}. \quad (3)$$

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

Having any two instances $x_1, x_2 \in \mathbb{X}$ we may compute a correlation between preference systems $\tilde{x}_1, \tilde{x}_1$ generated by these instances, using the generalized Kendall's tau, admitting incomplete preferences (see [4]):

$$\tilde{\tau} = \frac{1}{2M(M-1)} \sum_{i=1}^{M} \sum_{j=1}^{M} [sgn(\mu_{\tilde{x}_1}(y_j) - \mu_{\tilde{x}_1}(y_i)) \cdot sgn(\mu_{\tilde{x}_2}(y_j) - \mu_{\tilde{x}_2}(y_i)) \quad (4)$$
$$+ sgn(\nu_{\tilde{x}_1}(y_j) - \nu_{\tilde{x}_1}(y_i)) \cdot sgn(\nu_{\tilde{x}_2}(y_j) - \nu_{\tilde{x}_2}(y_i))].$$

In Sec. 2.1 we have identified preferences with an adequate permutation π_x of labels $\mathbb{Y}$. For possibly incomplete preferences we get incomplete permutation $\tilde{\pi} = \tilde{\pi}_x$ which might be identified with the corresponding IF-set $\tilde{x}$. Thus for any two instances $x_1, x_2 \in \mathbb{X}$ we have $\tilde{\tau} = \tilde{\tau}(\tilde{x}_1, \tilde{x}_2) = \tilde{\tau}(\tilde{\pi}_1, \tilde{\pi}_2)$. Hence, using (4), we may consider the following measure

$$D_{\tilde{\tau}}(\tilde{\pi}_1, \tilde{\pi}_2) = \frac{1 - \tilde{\tau}(\tilde{\pi}_1, \tilde{\pi}_2)}{2}, \tag{5}$$

which seems to be useful in the generalized Mallows model (1) admitting incomplete rankings and defined as follows

$$\tilde{\mathbb{P}}(\tilde{\pi}|\theta, \tilde{\pi}_0) = \frac{exp(-\theta D_{\tilde{\tau}}(\tilde{\pi}, \tilde{\pi}_0))}{\phi(\theta)}. \tag{6}$$

Of course, when modeling preferences by IF-sets one can also consider other substitutes for the measure D in (1), including different distances, dissimilarity measures or divergences (see, e.g., [8]). 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., [2]).

In the examples below we compare the suggested methodology with the results obtained using the distance based on the classical Kendall's tau for all linear extensions of incomplete rankings.

Example 1. Consider $M = 6$ labels and the following two ranking: $\pi_0 : y_1 \succ y_3 \succ y_4 \succ y_2 \succ y_5 \succ y_6$ and $\pi : y_3 \succ y_1 \succ y_5 \succ y_4 \succ y_6$. It is seen at once that the first ranking is complete, while the second one is incomplete because of unknown location of label y_2.

To perform the classical Mallows model one may consider possible six different locations of y_2 with respect to other labels. Using notation introduced in Sec. 2.1 if we put, e.g. $\pi^{-1}(2) = k$, which means that label y_2 is located on the k-th position in the complete ranking ($k = 1, \ldots, 6$), then for all labels y_j such that $\pi^{-1}(j) \geq k$, their position in the new ranking shifts to the right, so we get $\pi^{-1}(j) := \pi^{-1}(j) + 1$. The probabilities calculated for the classical Mallows model according to formula (1) for all possible location of the unknown label y_2 are given in Table 1. In these calculations the classical Kendall's τ was applied in (5) and the spread parameter $\theta = 1$ was assumed.

Table 1. Values of $\mathbb{P}(\pi|\theta, \pi_0)$ for different locations of y_2

$\pi^{-1}(2)$	1	2	3	4	5	6
$\mathbb{P}(\pi\|\theta, \pi_0)$	0.09049159	0.09673	0.1033985	0.09673	0.1033985	0.09673

On the other hand, we may construct IF-sets $\tilde{x}_0$ and $\tilde{x}$ describing preferences generated by π_0 and π, respectively. By (3) we get

$$\tilde{x}_0 = \{(y_1, 1, 0), (y_2, 0.4, 0.6), (y_3, 0.8, 0.2), (y_4, 0.6, 0.4), (y_5, 0.2, 0.8), (y_6, 0, 1)\}$$
$$\tilde{x} = \{(y_1, 0.6, 0.2), (y_2, 0, 0), (y_3, 0.8, 0), (y_4, 0.2, 0.6), (y_5, 0.4, 0.4), (y_6, 0, 0.8)\}$$

If we calculate probability (6) for the complete ranking π_0 and incomplete π using formula (5) based on the generalized Kendall's tau (4) then we get $\tilde{\mathbb{P}}(\pi|\theta, \pi_0) = 0.09673$. As we can see, (6) approximates possible probabilities quite well. We tried many other examples and the results were similarly good. □

Example 2. Now we will check what happen if there are more missing values in label ranking. Let us consider $M = 7$ labels and the following two ranking: $\pi_0 : y_1 \succ y_2 \succ y_3 \succ y_4 \succ y_5 \succ y_6 \succ y_7$ and $\pi : y_4 \succ y_1 \succ y_3 \succ y_7 \succ y_5$. So now the first ranking is complete, while the second one is incomplete because of two unknown location of labels y_2 and y_6.

Using the suggested methodology based on IF-sets and the generalized Kendall's tau (4) the probability value of (6) for the complete ranking π_0 and incomplete π equals $\tilde{\mathbb{P}}(\pi|\theta, \pi_0) = 0.05330688$.

However, if we apply the traditional approach based on possible linear extensions $\pi^* \in E(\pi)$ (see Sec. 2.2) we get $\min_{\pi^* \in E(\pi)} \{\mathbb{P}((\pi^*|\theta, \pi_0)\} = 0.0451233$ and $\max_{\pi^* \in E(\pi)} \{\mathbb{P}((\pi^*|\theta, \pi_0)\} = 0.0629746$, while the arithmetic mean and the median for all possible linear extensions $\{\mathbb{P}((\pi^*|\theta, \pi_0) : \pi^* \in E(\pi)\}$ equals 0.0551824 and 0.05459133, respectively. Hence again, IF-set based approach appears to be helpful in approximating the probability (1) for incomplete rankings. □

4 Incomplete Knowledge and the Mallows Model in Designing Recommendations

4.1 Main Idea

As we have mentioned above, our aim is to predict a ranking of labels for a given new instance x. Unfortunately, estimation of π from (6) is not very simple. However, in many applications it is not necessary to identify a whole ranking but it suffices to indicate only those labels which are located on the highest positions in the ranking. It is a typical case found in recommender systems.

In this contribution we apply the Mallows model to express the probability corresponding to the best label, i.e.

$$\tilde{\mathbb{P}}(y_j^{best}|\theta, \pi^*) = \frac{exp(-\theta D^*(y_j^{best}, y_j^{\pi^*}))}{\phi(\theta)}, \tag{7}$$

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

$$D^*(y_j^{best}, y_j^{\pi^*}) = \sqrt{\frac{1}{2}\sum_{i=1}^{n}((\mu_{y_j^{best}} - \mu_{\pi^*}(y_j))^2 + (\nu_{y_j^{best}} - \nu_{\pi^*}(y_j))^2)}. \tag{8}$$

In our case $\mu_{y_j^{best}} = 1$ and $\nu_{y_j^{best}} = 0$, as we want to calculate the probability that y_j is the best label for instance x. Then, as a final recommendation we assume

$$Y = \underset{y_j}{\text{argmax}}\{\sum_{\pi^* \in \bar{\pi}_{kNN(x)}} \tilde{\mathbb{P}}(y_j^{best}|\theta, \pi^*)\}, \tag{9}$$

where $\bar{\pi}_{kNN(x)}$ is the set of preference systems connected with k instances nearest to x. To predict the complete ranking for instance x we order labels $y_1, \ldots, y_M$ according to the values of (7).

4.2 Algorithms

We propose two algorithms based on the ideas discussed above. The first one is a direct implementation of the method proposed in Sec. 4.1.

Mallows Best Probability Algorithm (MBP)
{**Input**: x - new instance, X - learning set of instances, $\bar{\pi}$ - labels connected with instances, k - number of nearest neighbors}
1. *Find k nearest neighbors of x in X.*
2. *For (j in* $1 : M$*) calculate* $\sum_{\pi^* \in \bar{\pi}_{kNN(x)}} \tilde{\mathbb{P}}(y_j^{best}|\theta, \pi^*)$
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*}

The second algorithm is a modification of MBP that replaces missing labels in $\bar{\pi}_{kNN(x)}$ by the most probable extension of $\pi^* \in \bar{\pi}_{kNN(x)}$ with respect to (1). This replacement idea was suggested in IBLR algorithm given in [2].

Multistep Mallows Best Probability Algorithm (MMBP)
{**Input**: x - new instance, X - learning set of instances, $\bar{\pi}$ - labels connected with instances, k - number of nearest neighbors}
1. *Find k nearest neighbors of x in X.*
2. *For (j in* $1 : M$*) calculate* $\sum_{\pi^* \in \bar{\pi}_{kNN(x)}} \tilde{\mathbb{P}}(y_j^{best}|\theta, \pi^*)$
3. *MMBP-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).*
4. $\bar{\pi}^{mod}_{kNN(x)} <-$ *Find the most probable extensions of* $\pi^* \in \bar{\pi}_{kNN(x)}$ *with respect to (6).*
5. *For (j in* $1 : M$*) calculate* $\sum_{\pi^*_{mod} \in \bar{\pi}^{mod}_{kNN(x)}} \tilde{\mathbb{P}}(y_j^{best}|\theta, \pi^*)$
6. *MMBP-rankmod* $<-$ *Sort labels according to the values obtained in step 5 (in case of ties a label with lower index is better in the ranking)*
7. *If (MMBP-rankmod* $\neq$ *MMBP-rank) then (MMBP-rank* $<-$ *MMBP-rankmod, go to step 4) else (output(MMBP-rank)).*
{**Output**: *MMBP-rank*}

4.3 Experimental Results

To evaluate the proposed method we compared it with the IBLR algorithm given in [2]. Two types of data sets were used in our experiment: (A) For classification data, we followed the procedure proposed in [2], i.e. the naive Bayes classifier was first trained on the complete data set and then, for each example, all the labels present in the data set were ordered with respect to the predicted class probabilities. (B) For regression data a certain number of (numerical) attributes was removed from the set of predictors and each one was considered as a label. To obtain a ranking, the attributes were standardized and then ordered (see [2]). To obtain incomplete ranks we changed some ranks in every ranking into NA (non available). We considered different proportions p of missing values.

To compare algorithms we used two quality measures: their prediction accuracy and the evaluation times. As a measure of accuracy we used Kendall's tau. We evaluated the experiments using leave-one-out crossvalidation and according to the random effect of removing labels from complete rankings we repeated the evaluation 20 times for every chosen value of p. The results shown in Table 2 and Table 3 are the mean results for a given p.

All evaluations were performed using R package. We set the number of nearest neighbors to 5 (function *knn* from FNN library). The evaluation times, i.e. times of one full leave-one-out crossvalidation procedure for every algorithm, are measured using *proc.time()*. In Table 2 and Table 3 we show the mean times for all evaluations. To improve performance and parallelize our calculations, we used library *snowfall* with parameters *sfInit(cpus=4, parallel=TRUE)* on Intel core i5 2450M CPU. All data sets used for experiments were downloaded from `http://www.uni-marburg.de/fb12/kebi/research/repository/`

Table 2. 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

Results given in Table 2 and Table 3 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. MMBP algorithm, which can be perceived as the improved (in some sense) MBP, behaves more or less like IBLR both with respect to the accuracy and evaluation time. Therefore, one may conclude that our IF-set based method for handling incomplete label ranking seems to be very promising: it might be as accurate as IBLR (in MMBP version), but if we allow a slight lower accuracy then, using MBP version, we get desired results much faster than using IBLR.

Table 3. 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

5 Conclusions

In practice, the choice of the best method should be determined by the data structure. In recommender systems the 2% better accuracy is not as crucial as the time performance. Moreover, obviously the time consumed by all this methods increases with the number of labels and the number of missing values. The typical situation in recommender systems is that the number of labeled products is very large and therefore most of labels are missing for each user. Thus, the proposed MBP algorithm seems to be a promising candidate for creating recommendations especially in the presence of large number of labeled items.

Acknowledgments. Study was supported by research fellowship within "Information technologies: research and their interdisciplinary applications" project co-financed by European Social Fund (agreement number POKL.04.01.01-00-051/10-00).

References

1. Atanassov, K.: Intuitionistic Fuzzy Sets: Theory and Applications. Springer-Verlag (1999)
2. Cheng, W., Dembczynski, K., Hüllermeier, E.: Decision Tree and Instance-Based Learning for Label Ranking. In: ICML 2009, Montreal (2009)
3. Grzegorzewski, P.: The coefficient of concordance for vague data. Computational Statistics & Data Analysis 51, 314–322 (2006)
4. Grzegorzewski, P.: Kendall's correlation coefficient for vague preferences. Soft Computing 13, 1055–1061 (2009)
5. Grzegorzewski, P., Ziembińska, P.: Spearman's rank correlation coefficient for vague preferences. In: Christiansen, H., De Tré, G., Yazici, A., Zadrozny, S., Andreasen, T., Larsen, H.L. (eds.) FQAS 2011. LNCS, vol. 7022, pp. 342–353. Springer, Heidelberg (2011)
6. Hüllermeier, E., Fürnkranz, J., Cheng, W., Brinker, K.: Label ranking by learning pairwise preferences. Artificial Intelligence 172, 1897–1916 (2008)
7. Mallows, C.: Non-null ranking models. Biometrika 44, 114–130 (1957)
8. Montes, S., Iglesias, T.: Montes, S., Iglesias, T., Janiš, V., Montes, I.: A common framework for some comparison measures of IF-sets. In: IWIFSGN 2012, Warsaw (2012)