Exceptional Preferences Mining

Cláudio Rebelo de Sá[1,2]($\boxtimes$), Wouter Duivesteijn[3], Carlos Soares[1],
and Arno Knobbe[2]

[1] INESCTEC, Porto, Portugal
claudio.r.sa@inesctec.pt, csoares@fe.up.pt
[2] LIACS, Universiteit Leiden, Leiden, The Netherlands
a.j.knobbe@liacs.leidenuniv.nl
[3] Data Science Lab and iMinds, Universiteit Gent, Ghent, Belgium
wouter.duivesteijn@ugent.be

Abstract. Exceptional Preferences Mining (EPM) is a crossover between two subfields of datamining: local pattern mining and preference learning. EPM can be seen as a local pattern mining task that finds subsets of observations where the preference relations between subsets of the labels significantly deviate from the norm; a variant of Subgroup Discovery, with rankings as the (complex) target concept. We employ three quality measures that highlight subgroups featuring exceptional preferences, where the focus of what constitutes 'exceptional' varies with the quality measure: the first gauges exceptional overall ranking behavior, the second indicates whether a particular label stands out from the rest, and the third highlights subgroups featuring unusual pairwise label ranking behavior. As proof of concept, we explore five datasets. The results confirm that the new task EPM can deliver interesting knowledge. The results also illustrate how the visualization of the preferences in a Preference Matrix can aid in interpreting exceptional preference subgroups.

1 Introduction

Consider a survey where detailed preferences of sushi types have been collected, along with demographic details of the respondents. For each example in the dataset, we have personal details (age, gender, income, etc.) as well as a set of sushi types, ordered by preference [16]. By mapping the demographic attributes and unusual preferences, marketeers would be able to target key demographics where specific sushi types have greater potential.

The study of preference data has been approached from a number of perspectives, grouped under the name *Preference Learning* (PL) (e.g., as Label Ranking [3,5,24]). Typically, the aim is to build a global predictive model, such that the preferences can be predicted for new cases. However, in several areas, such as marketing, there is also great value in identifying subpopulations whose preferences deviate from the norm. If some sushi type is markedly underpreferred by a certain age group or in a certain region, then the vendor can develop specific

strategies for those groups. Finding coherent groups of customers to focus on is an invaluable part of promotion strategies.

Arguably the most generic setting for discovering local, supervised deviations is that of Subgroup Discovery (SD) [18]. The aim of SD is to discover subgroups in the data for which the target shows an unusual distribution, as compared to the overall population [18]. SD is generic in the sense that the actual nature of the target variable can be quite diverse [1,15,23]. In this paper, we develop a Subgroup Discovery approach that focuses on a deviation target concept representing preferences over a fixed set of labels.

1.1   Main Contributions

This work provides focus specifically on the discovery of meaningful subgroups with exceptional preference patterns (see Sect. 4). We propose three quality measures for this purpose, reflecting different facets of interestingness one might have about the unusual preferences. All quality measures contrast the ranking of the labels in the subgroup with the ranking of the labels in the entire dataset; they differ in the granularity of the measured deviation. A subgroup is deemed interesting by the first quality measure if the overall ranking is exceptional, by the second quality measure if one particular label behaves exceptionally, and by the third quality measure if a single pair of labels displays exceptional behavior. Hence, Exceptional Preferences Mining provides subgroups displaying exceptional ranking behavior; different quality measures allow for this exceptional behavior to either encompass the entire label space, or focus on more local peculiarities.

2   Label Ranking

Label Ranking (LR) studies the problem of learning a mapping from instances to rankings over a finite number of predefined labels [13]. It can be considered a variant of the conventional classification problem [4]. However, in contrast to a classification setting, where the objective is to assign examples to a specific class, in LR we are interested in assigning a complete preference order of the labels to every example.

More formally, in classification, given an instance x from the instance space $\mathbb{X}$, the goal is to predict the label (or class) λ to which x belongs, from a predefined set $\mathcal{L} = \{\lambda_1, \ldots, \lambda_k\}$. In Label Ranking, the goal is to order the labels in $\mathcal{L}$ by their association with x. A ranking is a total order over $\mathcal{L}$ defined on the permutation space Ω. A total order can be represented as a permutation π of the set $\{1, \ldots, k\}$, such that $\pi(a)$ is the position of λ_a in π.

A total order

$$\lambda_{\pi(1)} \underset{x}{\succ} \lambda_{\pi(2)} \underset{x}{\succ} \ldots \underset{x}{\succ} \lambda_{\pi(k)}$$

is associated with every instance $x \in \mathbb{X}$, representing a ranking $\pi \in \Omega$. In cases where the orders are partial, they are represented as rankings with ties [11].

The goal in label ranking is to learn the mapping $\mathbb{X} \to \Omega$. The training data is defined as D, which is a bag of n records of the form $x = (a_1, \ldots, a_m, \pi)$, where $\{a_1, \ldots, a_m\}$ is set of values from m independent variables $\mathcal{A}_1, \ldots, \mathcal{A}_m$ describing instance x and π is the corresponding target ranking.

Pairwise comparisons have been used to decompose LR or Multi-Label problems into binary problems [13]. In LR, the most relevant approach is Ranking by Pairwise Comparisons (RPC) [10], which decomposes the LR problem into a set of binary classification problems. Then, a learning method is trained with all examples for which either $\lambda_i \succ \lambda_j$ or $\lambda_j \succ \lambda_i$ is known [10]. The resulting predictions are then combined to predict a total or partial ranking [3].

Recently, some approaches have been suggested for mining preferences and ranks [12,19]. These approaches tackle different problems from the one we propose in this paper. In [12], the authors suggest an approach to mine the rankings with association rules that search for *subranking* patterns, while our approach relates the ranking patterns with descriptors (otherwise referred to as independent variables). From a different perspective, [19] suggests a *ranked tiling* approach to search for rank patterns, whereas we are interested in the preference relations derived from the ranks.

3   Subgroup Discovery and Exceptional Model Mining

Subgroup Discovery (SD) [18] is a data mining framework that seeks subsets (satisfying certain user-specified constraints) of the dataset where something exceptional is going on. In SD, we assume a flat-table dataset D, which is a bag of n records of the form $x = (a_1, \ldots, a_m, t_1, \ldots, t_\ell)$. We call $\{a_1, \ldots, a_m\}$ the *descriptors* and $\{t_1, \ldots, t_\ell\}$ the *targets*, and we denote the collective domain of the descriptors by $\mathcal{A}$. We are interested in finding interesting subsets, called *subgroups*, that can be formulated in a *description language* $\mathcal{D}$. In order to formally define subgroups, we first need to define the following auxiliary concepts.

Definition 1 (Pattern and coverage). *Given a description language $\mathcal{D}$, a pattern $p \in \mathcal{D}$ is a function $p : \mathcal{A} \to \{0,1\}$. A pattern p covers a record x iff $p(a_1, \ldots, a_m) = 1$.*

Patterns induce subgroups, and subgroups are associated with patterns, in the following manner.

Definition 2 (Subgroup). *A subgroup corresponding to a pattern p is the bag of records $S_p \in D$ that p covers:*

$$S_p = \{x \in D \mid p(a_1, \ldots, a_m) = 1\}$$

For simplicity, we will loosely identify pattern and subgroup with each other.

The exact choice of the description language is left to the domain expert or analyst. A typical choice is the use of conjunctions of conditions on attributes.

Restricting the findings of SD from all subsets to only subgroups that can be defined in such a way, ensures results of the following form:

$$\text{Age} \geq 30 \;\wedge\; \text{Likes} = \text{Salmon Roe} \quad \text{is unusual}$$

Restricting the search from subsets to subgroups, combined with a sensible choice of description language, ensures that SD delivers subgroups that are defined in terms of attributes of the dataset. This means that the results are delivered in a form with which dataset domain experts are familiar. In other words, the focus of SD lies on delivering *interpretable* results.

Formally, the interestingness of a subgroup can be measured using all information available in its associated pattern. In practice, it depends on the task we are trying to solve. Therefore, we should define one or more *quality measures* to assess the interestingness we want to explore.

Definition 3 (Quality Measure). *A* quality measure *is a function* $\varphi : \mathcal{D} \to \mathbb{R}$.

In the most common form of pattern mining, *frequent itemset mining* [2], interestingness is measured by the frequency of the pattern. *Subgroup Discovery* [18], on the other hand, measures interestingness in a *supervised* form. One designated target t_1 is identified in the dataset, and subgroup interestingness is gauged by an unusual distribution of that target. Hence, considering that a poll revealed that the majority of Japanese people like *Fatty tuna* sushi, an interesting subgroup could refer to a group of people for which the majority prefers *Tuna roll*:

$$\text{Age} \geq 30 \;\wedge\; \text{Lives in region} = \text{Hokkaido} \quad \Rightarrow \quad \text{Likes} = \text{Tuna roll}$$

If instead of a single target, multiple targets $t_1, \ldots, t_\ell$ are available, and if we are not interested in finding unusual target distribution, but unusual target interaction, we can employ *Exceptional Model Mining* (EMM) [6,7] instead of SD. EMM is instantiated by selecting two things: a model class and a quality measure. Typically, a model class is defined to represent the unusual interaction between multiple targets we are interested in. A specific quality measure that employs concepts from that model class must be defined to express exactly when an interaction is unusual and, therefore, interesting.

The target concept at hand in this paper has only one target object t, which resembles SD. However, that target object is a label ranking $\pi \in \Omega$, as defined in Sect. 2. Hence it represents unusual interactions between multiple individual labels, which is more consistent with EMM.

3.1   Traversing the Search Space

Typically, *subgroups* are found by a level-wise search through attribute space [21]. We define constraints on single attributes and define the corresponding subgroups as those records satisfying each one of those constraints.

The actual phenomenon of the data that a given quality measure favors, depends on the target concept (binary, numeric, preferences, . . .). For very small

subgroups, one easily finds an unusual distribution of the target. To favor larger subgroups, one defines the quality measure such that it balances the exceptionality of the target distribution with the size of the subgroup.

SD approaches have been developed for binary, nominal [1] and numeric target variables [14,15], as well as for targets encompassing multiple attributes [23]. However, none of the previous approaches is able to capture all the sets of preferences that can be derived from rankings within an SD framework.

4   Exceptional Preferences Mining

Exceptional Preferences Mining (EPM) is the search for subgroups with deviating preferences. Exactly what constitutes an interesting deviation in preferences is governed by the employed quality measure, and can be inspired by the application at hand.

When the number of labels is large, the search for preference patterns can be hard to analyze and visualize. A real world example is the Sushi dataset [16], which represents the preferences of 5 000 persons over 10 types of sushi. Even this relatively modest number of sushi types can be ranked in a large number of combinations: more than 98 % of the 5 000 rankings present in this dataset are unique. This illustrates why it can be more difficult to directly learn a ranker that associates a reliable complete ranking for any subset in $\mathbb{X}$ when the number of labels is large.

In EPM, we want to search for strong preference behavior. However, in cases like the Sushi dataset, it is difficult to get strong total orders, due to the low number of ranking repetitions. In other words, searching for subgroups where all types of sushi are consistently ranked in this exact same order can be unfruitful. For this reason, we also propose lower-granularity measures that focus on one label versus the others (Labelwise). That is, we look for subgroups where at least one type of sushi is often preferred to all the others. As an example, if a subgroups ranks *tekka-maki* consistently in the top 3 while the majority in the dataset ranks it in the last 3, this measure will find it to be very interesting. We also propose a measure of even lower granularity, focusing on label versus label (Pairwise) preferences. This means that, if most people display a preference $tamago \succ kappa\text{-}maki$, a subgroup where most people prefer $kappa\text{-}maki \succ tamago$ will be deemed interesting by this measure.

Our assumption is that, even though over 98 % of the total rankings in the Sushi dataset are unique, there is plenty of information present in these rankings: the partial orders and pairwise comparisons can reveal interesting subgroups.

4.1   Preference Matrix

Let us define a function, ω, assigning a numeric value to the pairwise comparison of the labels λ and $\hat{\lambda}$:

$$\omega\left(\lambda, \hat{\lambda}\right) = \begin{cases} 1 & \text{if } \lambda \succ \hat{\lambda} \ (\lambda \text{ preferred to } \hat{\lambda}) \\ -1 & \text{if } \lambda \prec \hat{\lambda} \ (\hat{\lambda} \text{ preferred to } \lambda) \\ 0 & \text{if } \lambda \sim \hat{\lambda} \ (\lambda \text{ indifferent to } \hat{\lambda}) \\ n/a & \text{if } \lambda \perp \hat{\lambda} \ (\lambda \text{ incomparable to } \hat{\lambda}) \end{cases}$$

Note that, by definition, $\omega\left(\lambda, \hat{\lambda}\right) = -\omega\left(\hat{\lambda}, \lambda\right)$. We can use ω to represent a ranking π as a *Preference Matrix* (PM), M_π:

$$M_\pi\left(i, j\right) = \omega_\pi\left(\lambda_i, \lambda_j\right)$$

M_π is, by definition, an antisymmetric matrix with $tr\left(M_\pi\right) = 0$. PMs can natively represent partial or incomplete orders but can also be aggregated to represent sets of rankings from an entire dataset D or subgroup S. To aggregate the entries, the *mean* or the *mode* can be used.

Table 1. Example dataset $\hat{D}$. The first column is the only descriptor. The subsequent four columns represent the preferences among four labels, by providing their ranks. An alternative representation is presented in the rightmost section of the table.

$\mathcal{A}_1$	π				Alternative π
	λ_1	λ_2	λ_3	λ_4	
0.1	4	3	1	2	$\lambda_3 \succ \lambda_4 \succ \lambda_2 \succ \lambda_1$
0.2	3	2	1	4	$\lambda_3 \succ \lambda_2 \succ \lambda_1 \succ \lambda_4$
0.3	1	4	2	3	$\lambda_1 \succ \lambda_3 \succ \lambda_4 \succ \lambda_2$
0.4	1	3	2	4	$\lambda_1 \succ \lambda_3 \succ \lambda_2 \succ \lambda_4$

Aggregation of a PM for Sets of Rankings. The PM of a set of rankings from a dataset D with n rankings, M_D, aggregated with the mean is:

$$M_D\left(i, j\right) = \frac{1}{n} \sum_{\pi \in D} M_\pi\left(i, j\right)$$

The PM of the example dataset $\hat{D}$ (cf. Table 1) is the following:

$$M_{\hat{D}} = \begin{bmatrix} 0 & 0 & 0 & 0.5 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ -0.5 & 0 & -1 & 0 \end{bmatrix}$$

This representation enables easy detection of strong partial order relations in a set. If row i has all the values very close to 1, then λ_i is highly preferred in this group. If entry $M_{\hat{D}}(i, j) = 1$ or $M_{\hat{D}}(i, j) = -1$, then all rankings in $\hat{D}$ agree that $\lambda_i \succ \lambda_j$ or $\lambda_i \prec \lambda_j$, respectively.

All the elements of $\hat{D}$ reveal distinct total preferences, but λ_3 is always preferred to λ_2, which is easily verified by checking that $M_{\hat{D}}(3, 2) = 1$. In the ranking representation of $\hat{D}$, this fact follows from four distinct combinations of ranks: rank $3 > 1$, rank $2 > 1$, rank $4 > 2$ and rank $3 > 2$ (this information is found in the two columns below λ_2 and λ_3). Conversely, λ_4 is never preferred to λ_3, which is represented by $M_{\hat{D}}(4, 3) = -1$. In some cases, the overall trend is not as clear (e.g. λ_1 is preferred to λ_4 but not always) and in other cases, there is no trend at all (e.g. λ_1 and λ_2).

Representing a set of rankings as a PM has another advantage over the traditional permutation representation: it enables simple measurement of *labelwise* (by rows/columns of the PM) and *pairwise* (by single entries of the PM) distances (see Sect. 4.2).

From the PM of a subgroup S, one can derive a new ranking π_S. How to do so is a non-trivial question, which has received a lot of attention in several research fields with similar types of matrix [13]. The straightforward way is to sum the rows of the PM and then assign a score to each corresponding label. Higher values correspond to a relatively more preferred label.

The generation of a PM is basically a pairwise decomposition problem. The complexity is $\mathcal{O}(sk^2)$ per subgroup, where s is the size of the subgroup and k the number of labels in the ranking. Even though any number of labels is theoretically permitted in label ranking, in practice the number of labels is usually smaller than 20. Hence, the generation of PMs should not be an issue in terms of computational time.

We use a visual representation of PM that is a set of colored tiles (cf. Fig. 1). Each tile represents an entry of the PM. The entries of a PM can vary from -1 to 1. The negative entries of the matrix are represented with red tiles, the positive with green tiles, and 0 is represented in white. The colored tiles fade out as they get closer to 0.
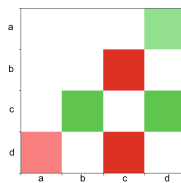

Fig. 1. PM representation of the set of rankings in $\hat{D}$ (cf. Table 1). Dark green tiles represent 1 and dark red tiles respresent -1. (Color figure online)

4.2   Characterizing Exceptional Subgroups

The table has now been set to formally define the quality measures for EPM, which will evaluate how exceptional the preferences are in the subgroups. A subgroup can be considered interesting both by the amount of deviation (distance) and by its size (number of records *covered* by the subgroup, cf. Sect. 3) [9]. Since, reasonable quality measures should take both these factors into account, we divide the quality measures into two parts: the *distance* component and the *size* component.

$$QM_S = size_S \cdot distance_S$$

In order to allow direct comparisons between different quality measures, both components are normalized to the interval $[0, 1]$. A common measure for the size in Subgroup Discovery is $\sqrt{s}$ [17]. To normalize, we use the square root of the fraction of the dataset covered by S: $size_S = \sqrt{s/n}$.

Before introducing the distance components, let us first define a distance matrix L_S, as the distance matrix between the PMs M_S and M_D:

$$L_S = \frac{1}{2} (M_D - M_S)$$

where $S \subseteq D$ (division by 2 limits the distance to the interval $[-1, 1]$). We can measure different properties of L_S and represent them with a numeric value. This way we get an indicator of the quality of the distance of preferences for a subgroup. Consider the subgroup $\hat{S}_1 : A_1 \geq 0.3$, which covers the last two cases from our example dataset $\hat{D}$. Its PM is:

$$M_{\hat{S}_1} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ -1 & 0 & -1 & 0 \\ -1 & 1 & 0 & 1 \\ -1 & 0 & -1 & 0 \end{bmatrix}$$

The first row clearly reveals that λ_1 is always preferred to all other labels in this subgroup. If we compute the difference matrix $L_{\hat{S}_1}$ we get:

$$L_{\hat{S}_1} = \begin{bmatrix} 0 & -0.5 & -0.5 & -0.25 \\ 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 \\ 0.25 & 0 & 0 & 0 \end{bmatrix}$$

The difference matrix $L_{\hat{S}_1}$ shows that the behavior of λ_1 is exceptional in $\hat{S}_1$.

Only subgroups for which we can infer at least one pairwise preference are considered interesting in Exceptional Preferences Mining. That is, subgroups with a PM containing only zeros are not considered interesting.

As we are interested in subgroups with exceptional preferences, we use the distance matrix L_S to measure exceptionality. The distance measures we employ here typically consider a particular subset of the cells of the distance matrix L_S.

Norm. Maximizing the distance of preferences is also maximizing the magnitude of L_S. The most fundamental mathematical way to measure the magnitude of a vector or matrix is the *norm*. Hence we can use the *Frobenius norm* of L_S as a distance measure.

$$\text{Norm}(S) = \sqrt{s/n} \cdot ||L_S||_F = \sqrt{s/n} \cdot \sqrt{\sum_{i=1}^{k} \sum_{j=1}^{k} L(i,j)^2}$$

If one is searching for preference deviations in general, one should use the *Norm* quality measure, as it considers all the PM entries at the same time. After the subgroups are found, ideally, we can derive a complete ranking from their PMs. The overall deviation can be due to one label deviating strongly or from multiple labels deviating less strongly.

Labelwise. An interesting task in the PL field is the *labelwise* analysis [3]. Instead of focusing on a whole ranking, it focuses on the preference behavior from the perspective of individual labels. A data analyst might be interested in finding if a particular label λ behaves substantially different according to most members in a subgroup S, compared to its behavior on the overall dataset. Hence, the fact that only one label behaves differently, disregarding the interaction between the other labels, can also be interesting. We can measure the distance of each label, in subgroup S, by computing the *norm* of the rows from L_S. Since in this case we are interested in exceptionality of only one label, we consider the maximum value found:

$$\text{Labelwise}(S) = \sqrt{s/n} \cdot \max_{i=1,\ldots,k} \frac{1}{(k-1)} \sum_{j=1}^{k} L(i,j)$$

Pairwise. Another well-studied task in PL is Pairwise Preferences [13] which decomposes the preferences into pairs label-vs-label. In situations where there are not even exceptional labelwise preferences, one can still search for localized preference strongholds. If we are interested in subgroups with, at least one pair with distinctive preference behavior, we can employ the following quality measure:

$$\text{Pairwise}(S) = \sqrt{s/n} \cdot \max_{i,j=1,\ldots,k} L(i,j)$$

This quality measure is the least restrictive of this set: a subgroup is interesting if one pair of labels interacts unusually, disregarding all other label interactions.

5   Experiments

We incorporate Exceptional Preferences Mining in the Cortana[1] software package [22]. This package delivers a generic framework for SD, implements several SD instances, and offers many generic features allowing for different SD

---

approaches. The description language consists of logical conjunctions of conditions on single attributes.

Our experiments use a standard *beam search* approach. Since the Subgroup Discovery algorithm itself is not the topic of this paper, we will skip over the algorithmic details, but they can be found elsewhere: the relevant pseudo-code is given in [6, Algorithm 1]. The most influential parameters are set as follows: we use a relatively generous search width w (also known as beam width or beam size) of 100, allowing for a relatively broad (albeit heuristic) search, and a maximum search depth d of 2, which keeps the resulting subgroups interpretable. We explore some striking subgroups found with the quality measures on a variety of datasets, providing evidence of the versatility of our work.

All the findings we present in this paper have gone through the DFD validation procedure [8] with 100 copies, and all have been found significant at a significance level of $\alpha = 1\%$.

5.1   Datasets

Statistics regarding the datasets used in this work are shown in Table 2. The majority are Label Ranking datasets from the KEBI Data Repository at Philipps University of Marburg [4]. These datasets were adapted from multi-class and regression problems both from the UCI repository [20] and the Statlog collection [4]. In the process, the features were normalized, and their names were replaced by $A1, A2, \ldots, Am$. Therefore, on these datasets, the reported subgroups cannot be interpreted on the original dataset domain, whereas for general datasets, this interpretability is a key feature of Exceptional Preference Mining. We choose to experiment with these datasets anyway, since they are well-known in the preference learning community.

For illustrating domain-specific interpretation of the results, we experiment with two further datasets. We adapt the COIL 1999 Competition Data from UCI [20]. This dataset concerns the frequencies of algae populations in different environments. We refer to this dataset as *Algae*. The original COIL dataset consists of 340 examples, each representing measurements of a sample of water from different European rivers in different periods. The measurements include concentrations of chemical substances such as nitrogen (in the form of nitrates, nitrites and ammonia), oxygen and chlorine. Also the pH, season, river size and flow velocity are registered. For each sample, the frequencies of 7 types of algae are also measured. In this work, we consider the algae concentrations as preference relations by ordering them from larger to smaller concentrations. Those with 0 frequency are placed in last position and equal frequencies are represented with ties. Missing values are set to 0.

Our final dataset is the Sushi preference dataset [16], which is composed of demographic data about 5 000 people and sushi preferences. Each person sorts a set of 10 different sushi types by preference. The 10 types of sushi, are (a) shrimp, (b) sea eel, (c) tuna, (d) squid, (e) sea urchin, (f) salmon roe, (g) egg (h) fatty tuna, (i) tuna roll and (j) cucumber roll. Since the attribute names were not transformed in this dataset, we can make a richer analysis of it.

**Table 2.** Dataset details. The column U_π represents the percentage of unique rankings.

Datasets	#Examples	#Labels	#Attributes	U_π
Cpu-small	8 192	5	6	1 %
Elevators	16 599	9	9	1 %
Wisconsin	194	16	16	100 %
Algae (COIL)	316	7	10	72 %
Sushi	5 000	10	10	98 %

For all the experiments, all results and statistical tests are completed in less than 5 min on an Intel Core 2 Duo CPU @ 2.93 GHz with 4 GB RAM.

5.2   Results

We start this section by presenting a discovery which provides an exemplary demonstration of one advantage of the PM representation.

Elevators Dataset. Figure 2 shows the subgroup with highest score found with the *Norm* quality measure in the Elevators dataset. Considering the base matrix, which has information from all the rankings in the dataset, we conclude that e, f, g, h have fixed relative positions: $e \succ g \succ f \succ h$. This information is not easy to obtain with the usual representations of rankings, but is clearly revealed in the PM representation. In fact, 13 403 from a total of 16 599 rankings have $e \succ g \succ f \succ h$. This illustrates how the visual ranking representation in a PM can be very useful for supporting predictive methods and for data exploration. The subgroup, $A6 \geq 0.436$, covering 7 048 instances, had a norm of 0.0028. It shows a distinct behavior between the sets of labels a, b, c, d and the set e, f, g, h. In the whole data, labels a, b, c, d are a bit more desirable than e, f, g, h. However, in the subgroup, the latter are clearly preferred to a, b, c, d.
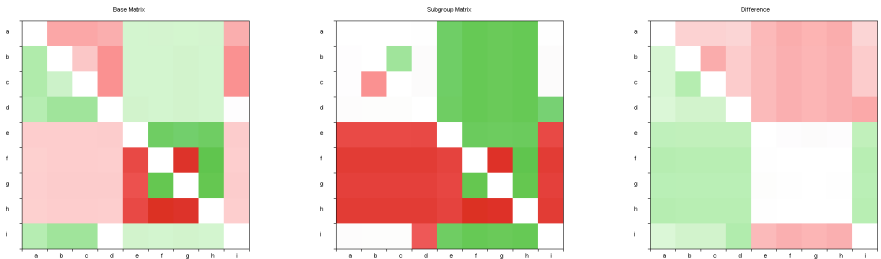
Fig. 2. PM representation of the dataset Elevators (base matrix), the subgroup $A6 \geq 0.436$ (subgroup matrix) and the difference (difference matrix).
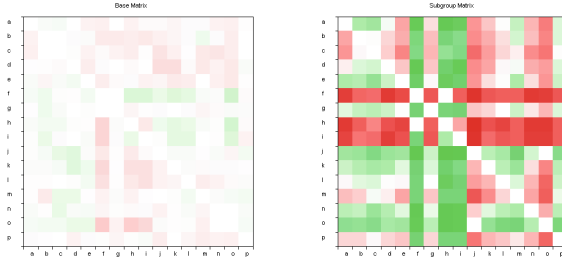
Fig. 3. PM representation of the dataset Wisconsin (Base Matrix) and the subgroup $A_5 \leq -0.527$ (Subgroup Matrix). (Color figure online)

Wisconsin. Using the *Norm* quality measure on the Wisconsin dataset, we obtain 30 subgroups, the 1st-ranked of which (it happens to occur at depth 1 in the search) is represented in Fig. 3. The base matrix reveals that the dataset has balanced preferences, by the low intensity of the colored tiles. The red rows of the PM of subgroup $A_5 \leq -0.527$ (Subgroup Matrix in Fig. 3) indicate a strong behavior of the labels f, h and i. The PM reveals that labels f, h, i are consistently ranked lower than the other labels in this specific subgroup. Since PMs are antisymmetric, the 3 green columns represent the same phenomena but from the perspective of the other labels. If we focus on these 3 labels, we can see that tile (f, h) is white, which means f and h are equivalent. On the other hand, tiles (i, f) and (i, h) are green, which means that $i \succ f$ and $i \succ h$. If one had to guess a reliable partial order from this subgroup using only the PM, a logical choice would be to say that $a, b, c, d, e, g, j, k, l, m, n, o, p \succ i \succ f, h$.

Algae. With the Algae dataset, we obtain results about the concentrations of algae with the *Norm* measure. One such example is that during *Spring*, the types of algae a, b and c are much more common in rivers than the others. This can be easily concluded by studying the PM representation of the subgroup
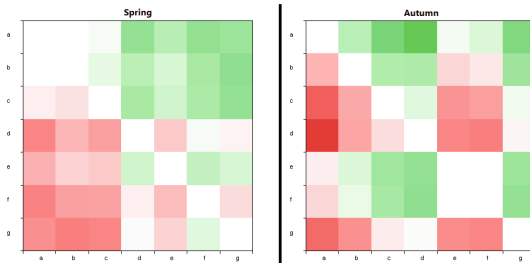

Fig. 4. PM representation of the subgroups $Season = Spring$ (left subgroup matrix) and $Season = Autumn$ (right subgroup matrix) from the Algae dataset.
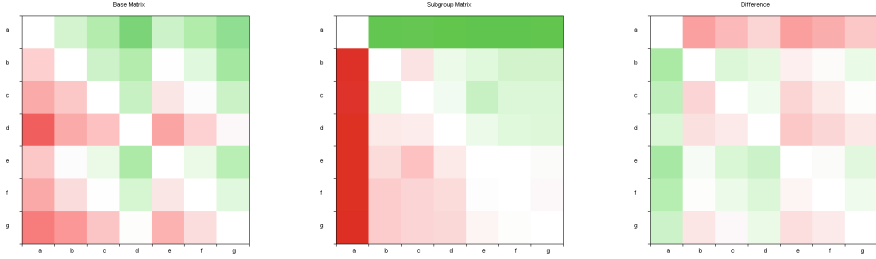
Fig. 5. PM representation of the dataset Algae (base matrix) and the subgroup $V10 \leq 59 \wedge V6 \leq 11.867$ (subgroup matrix), with difference matrix on the right.

(Fig. 4). This subgroup has a norm of 0.010647. On the other hand, we also see an interesting behavior during the *Autumn* season, with a norm of 0.01058.

With the *Labelwise* measure, we find more than 400 subgroups, the best of which is presented in Fig. 5. The PM clearly reveals the effect of the Labelwise quality measure: in the subgroup, the label *a* is strongly preferred over all others, while the image is much more nuanced over the whole dataset. If we ignore the label *a*, the PMs for both the overall dataset and the subgroup are rather bland, and their difference is not very pronounced. But for this one particular label *a*, the behavior on the subgroup is extremely clear-cut, and the Labelwise quality measure picks up on that effect.

Sushi. With the *Labelwise* measure, we find 149 subgroups on the Sushi dataset. We present the best subgroup using this measure in Fig. 6. The subgroup (Males over 30 years) shows a preference for Sea Urchin, since the majority of men rank this sushi type in the top 4. By contrast, in the whole population, more than half rate it between 5th to 10th, and every fifth person rate it in last place.
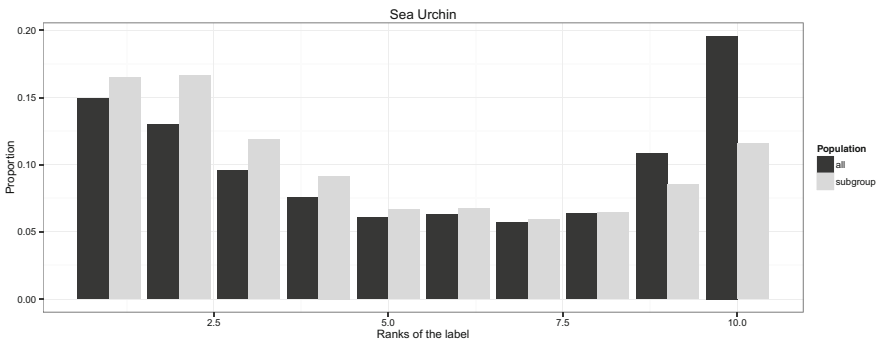

Fig. 6. Percentage of ranks for Sea Urchin (Sushi dataset) for all individuals in comparison to the subgroup (males older than 30 years).

Cpu-Small. On the Cpu-small dataset, the subgroup $A6 \geq 0.127$ ranks the best for the *Pairwise* quality measure. Around 80% of the $2\,221$ instances of this subgroup agree that $a \succ d$, in contrast to the 30% in the whole dataset of $8\,192$ instances.

6   Conclusions

We introduce Exceptional Preferences Mining (EPM), a supervised local pattern mining task where the target concept is a ranking of a fixed set of labels. The result of this task is a set of subgroups, which are coherent subsets of the dataset that can be described in terms of a conjunction of few conditions on an attribute, where the label preferences are exceptional in some sense.

The relevant statistics on a set of preference relations is collected in the cells of a Preference Matrix (PM). A PM is compiled for the entire dataset, and for each subgroup under consideration. A subgroup whose PM deviates significantly from the PM for the whole dataset is then considered to be interesting. We define three quality measures for EPM that instantiate this concept of 'interesting' to different levels of granularity. The *Norm* quality measure deems a subgroup interesting if the full set of preference relations is substantially displaced. The *Labelwise* quality measure highlights subgroups where any one label interacts exceptionally with the other labels, agnostic of how those other labels interact with each other. The *Pairwise* quality measure finds a subgroup interesting if any one pair of labels display exceptional preference relations. Hence, by choosing the appropriate quality measure, EPM delivers subgroups featuring preference relations that are exceptional at your preferred scope.

The experiments with the *Norm* quality measure on the Elevators dataset illustrate the value of the PM visualization. The PM, as displayed in Fig. 2, clearly indicates that there are strong relations between a subset of the available labels. We learn that quite frequently, labels e, f, g, h have fixed relative positions: $e \succ g \succ f \succ h$. This information is not easy to obtain with the usual representations of rankings, but is clearly revealed through the PM visualization. The experiments with the *Labelwise* quality measure on the Sushi dataset illustrate the relative merit of this quality measure: it focuses on subgroups where one particular label is exceptionally under- or overappreciated. The subgroup presented has a penchant for Sea Urchin (cf. Fig. 6). The *Pairwise* measure shows its potential on the Cpu-small dataset by identifying a subgroup with strong exceptional preferences with respect to the pair of labels a and d.

As we argued in Sect. 3, one of the main benefits of a local pattern mining method such as EPM is that it delivers interpretable results. That means that the resulting subgroups are ideally suited to instigate real-world policies and actions. However, due to the employed preprocessing in the KEBI datasets (cf. Sect. 5.1), interpretation of results on those datasets falters. Only the experiments on the Algae and Sushi datasets allow a more extensive exploration of interpretable results. In future work, we would be interested in evaluating EPM on more label ranking datasets that come with interpretable attributes.

References

1. Abudawood, T., Flach, P.: Evaluation measures for multi-class subgroup discovery. In: Buntine, W., Grobelnik, M., Mladenić, D., Shawe-Taylor, J. (eds.) ECML PKDD 2009, Part I. LNCS (LNAI), vol. 5781, pp. 35–50. Springer, Heidelberg (2009). doi:10.1007/978-3-642-04180-8_20

2. Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.: Fast discovery of association rules. Adv. Knowl. Disc. Data Min. **12**, 307–328 (1996)

3. Cheng, W., Henzgen, S., Hüllermeier, E.: Labelwise versus pairwise decomposition in label ranking. In: LWA 2013, Lernen, Wissen & Adaptivität, Workshop Proceedings Bamberg, 7–9 October 2013, pp. 129–136 (2013)

4. Cheng, W., Huhn, J.C., Hüllermeier, E.: Decision tree and instance-based learning for label ranking. In: Proceedings of the 26th Annual International Conference on Machine Learning, ICML 2009, Montreal, Quebec, Canada, 14–18 June 2009, pp. 161–168 (2009)

5. de Sá, C.R., Soares, C., Knobbe, A.J.: Entropy-based discretization methods for ranking data. Inf. Sci. **329**, 921–936 (2016)

6. Duivesteijn, W.: Exceptional model mining. Ph.D. Thesis, Leiden University (2013)

7. Duivesteijn, W., Feelders, A., Knobbe, A.: Exceptional model mining. Data Min. Knowl. Disc. **30**, 47–98 (2016)

8. Duivesteijn, W., Knobbe, A.J.: Exploiting false discoveries - statistical validation of patterns and quality measures in subgroup discovery. In: Proceedings of ICDM, pp. 151–160 (2011)

9. Dzyuba, V., Leeuwen, M.: Interactive discovery of interesting subgroup sets. In: Tucker, A., Höppner, F., Siebes, A., Swift, S. (eds.) IDA 2013. LNCS, vol. 8207, pp. 150–161. Springer, Heidelberg (2013). doi:10.1007/978-3-642-41398-8_14

10. Fürnkranz, J., Hüllermeier, E.: Pairwise preference learning and ranking. In: Lavrač, N., Gamberger, D., Blockeel, H., Todorovski, L. (eds.) ECML 2003. Lecture Notes in Artificial Intelligence (LNAI), vol. 2837, pp. 145–156. Springer, Heidelberg (2003). doi:10.1007/978-3-540-39857-8_15

11. Fürnkranz, J., Hüllermeier, E.: Preference Learning, 1st edn. Springer, New York (2010)

12. Henzgen, S., Hüllermeier, E.: Mining rank data. In: Japkowicz, N., Matwin, S. (eds.) DS 2014. LNCS (LNAI), vol. 8777, pp. 123–134. Springer, Heidelberg (2014). doi:10.1007/978-3-319-11812-3_11

13. Hüllermeier, E., Fürnkranz, J., Cheng, W., Brinker, K.: Label ranking by learning pairwise preferences. Artif. Intell. **172**(16–17), 1897–1916 (2008)

14. Jin, N., Flach, P., Wilcox, T., Sellman, R., Thumim, J., Knobbe, A.J.: Subgroup discovery in smart electricity meter data. IEEE Trans. Industr. Inf. **10**(2), 1327–1336 (2014)

15. Jorge, A.M., Pereira, F., Azevedo, P.J.: Visual interactive subgroup discovery with numerical properties of interest. In: Japkowicz, N., Matwin, S. (eds.) DS 2006. LNCS (LNAI), vol. 4265, pp. 301–305. Springer, Heidelberg (2006). doi:10.1007/11893318_31
16. Kamishima, T.: Nantonac collaborative filtering: recommendation based on order responses. In: Proceedings of KDD, pp. 583–588 (2003)
17. Klösgen, W.: Explora: a multipattern and multistrategy discovery assistant. In: Advances in Knowledge Discovery and Data Mining, pp. 249–271 (1996)
18. Klösgen, W., Zytkow, J.M. (eds.): Handbook of Data Mining and Knowledge Discovery. Oxford University Press, New York (2002)
19. Van, T., Leeuwen, M., Nijssen, S., Fierro, A.C., Marchal, K., Raedt, L.: Ranked tiling. In: Calders, T., Esposito, F., Hüllermeier, E., Meo, R. (eds.) ECML PKDD 2014. LNCS (LNAI), vol. 8725, pp. 98–113. Springer, Heidelberg (2014). doi:10.1007/978-3-662-44851-9_7
20. Lichman, M.: UCI Machine Learning Repository (2013)
21. Mannila, H., Toivonen, H.: Levelwise search and borders of theories in knowledge discovery. Data Min. Knowl. Discov. **1**(3), 241–258 (1997)
22. Meeng, M., Knobbe, A.: Flexible enrichment with cortana-software demo. In: Proceedings of BeneLearn, pp. 117–119 (2011)
23. Umek, L., Zupan, B.: Subgroup discovery in data sets with multi-dimensional responses. Intell. Data Anal. **15**(4), 533–549 (2011)
24. Vembu, S., Gärtner, T.: Label ranking algorithms: a survey. In: Fürnkranz, J., Hüllermeier, E. (eds.) Preference Learning, pp. 45–64. Springer, Heidelberg (2010)