Weighted Rank Correlation: A Flexible Approach Based on Fuzzy Order Relations

Sascha Henzgen and Eyke Hüllermeier(✉)

Department of Computer Science, University of Paderborn, Paderborn, Germany
{shenzgen,eyke}@upb.de

Abstract. Measures of rank correlation are commonly used in statistics to capture the degree of concordance between two orderings of the same set of items. Standard measures like Kendall's tau and Spearman's rho coefficient put equal emphasis on each position of a ranking. Yet, motivated by applications in which some of the positions (typically those on the top) are more important than others, a few weighted variants of these measures have been proposed. Most of these generalizations fail to meet desirable formal properties, however. Besides, they are often quite inflexible in the sense of committing to a fixed weighing scheme. In this paper, we propose a weighted rank correlation measure on the basis of fuzzy order relations. Our measure, called *scaled gamma*, is related to Goodman and Kruskal's gamma rank correlation. It is parametrized by a fuzzy equivalence relation on the rank positions, which in turn is specified conveniently by a so-called scaling function. This approach combines soundness with flexibility: it has a sound formal foundation and allows for weighing rank positions in a flexible way. The usefulness of our class of weighted rank correlation measures is shown by means of experimental studies using both synthetic and real-world ranking data.

1 Introduction

Rank correlation measures such as Kendall's tau [11] and Spearman's rho [20], which have originally been developed in non-parametric statistics, are used extensively in various fields of application, ranging from bioinformatics [1] to information retrieval [21]. In contrast to numerical correlation measures such as Pearson correlation, rank correlation measures are only based on the ordering of the observed values of a variable. Thus, measures of this kind are not limited to numerical variables but can also be applied to non-numerical variables with an ordered domain (i.e., measured on an ordinal scale) and, of course, to rankings (permutations) directly.

In many applications, such as Internet search engines, one is not equally interested in all parts of a ranking. Instead, the top positions of a ranking (e.g., the first 10 or 50 web sites listed) are typically considered more important than the middle part and the bottom. Standard rank correlation measures, however, put equal emphasis on all positions. Therefore, they cannot distinguish disagreements in different parts of a ranking. This is why *weighted* variants have been

A. Appice et al. (Eds.): ECML PKDD 2015, Part II, LNAI 9285, pp. 422–437, 2015.
DOI: 10.1007/978-3-319-23525-7_26

proposed for some correlation measures, as well as alternative measures specifically focusing on the top of a ranking [6,7,10,16,21]. Most of these generalizations fail to meet desirable formal properties, however. Besides, they are often quite inflexible in the sense of committing to a fixed weighing scheme.

In this paper, we develop a general framework for designing weighted rank correlation measures based on the notion of *fuzzy order relation*, and use this framework to generalize Goodman and Kruskal's gamma coefficient [8].[1] Our approach has a sound formal foundation and allows for weighing rank positions in a flexible way. In particular, it is not limited to monotone weighing schemes that emphasize the top in comparison to the rest of a ranking. The key ingredients of our approach, to be detailed further below, are as follows:

- *Fuzzy order relations* [4] are generalizations of the conventional order relations on the reals or the integer numbers: `SMALLER`, `EQUAL` and `GREATER`. They enable a smooth transition between these predicates and allow for expressing, for instance, that a number x is smaller than y *to a certain degree*, while to some degree these numbers are also considered as being equal. Here, the `EQUAL` relation is understood as a kind of similarity relation that seeks to model the "perceived equality" (instead of the strict mathematical equality).
- *Scaling functions* for modeling fuzzy equivalence relations [12]. For each element x of a linearly ordered domain X, a scaling function $s(\cdot)$ essentially expresses the degree $s(x)$ to which x can be (or should be) distinguished from its neighboring values. A measure of distance (or, equivalently, of similarity) on X can then be derived via accumulation of local degrees of distinguishability.
- *Fuzzy rank correlation* [5,17] generalizes conventional rank correlation on the basis of fuzzy order relations, thereby combining properties of standard rank correlation (such as Kendall's tau) and numerical correlation measures (such as Pearson correlation). Roughly, the idea is to penalize the inversion of two items (later on called a *discordance*) depending on how dissimilar the corresponding rank positions are: the more similar (less distinguishable) the positions are according to the `EQUAL` relation, the smaller the influence of the inversion on the rank correlation.

The rest of the paper is organized as follows. In the next two sections, we briefly recall the basics of fuzzy order relations and fuzzy rank correlation, respectively. Our weighted rank correlation measure, called *scaled gamma*, is then introduced in Section 5, and related work is reviewed in Section 6. A small experimental study is presented in Section 7, prior to concluding the paper in Section 8.

2 Rank Correlation

Consider $N \geq 2$ paired observations $\{(x_i, y_i)\}_{i=1}^{N} \subset \mathbb{X} \times \mathbb{Y}$ of two variables X and Y, where $\mathbb{X}$ and $\mathbb{Y}$ are two linearly ordered domains; we denote

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_N), \quad \boldsymbol{y} = (y_1, y_2, \ldots, y_N) \ .$$

[1] A preliminary version of this paper has been presented in [9], on the occasion of the German Workshop on Computational Intelligence, Dortmund, Germany, 2013.

In particular, the values x_i (and y_i) can be real numbers ($\mathbb{X} = \mathbb{R}$) or rank positions ($\mathbb{X} = [N] = \{1, 2, \ldots, N\}$). For example, $\boldsymbol{x} = (3, 1, 4, 2)$ denotes a ranking of four items, in which the first item is on position 3, the second on position 1, the third on position 4 and the fourth on position 2.

The goal of a (rank) correlation measure is to capture the dependence between the two variables in terms of their tendency to increase and decrease (their position) in the same or the opposite direction. If an increase in X tends to come along with an increase in Y, then the (rank) correlation is positive. The other way around, the correlation is negative if an increase in X tends to come along with a decrease in Y. If there is no dependency of either kind, the correlation is (close to) 0.

2.1 Concordance and Discordance

Many rank correlation measures are defined in terms of the number C of *concordant*, the number D of *discordant*, and the number T of *tied* data points. Let $\mathcal{P} = \{(i, j) \,|\, 1 \leq i < j \leq N\}$ denote the set of ordered index pairs. We call a pair $(i, j) \in \mathcal{P}$ concordant, discordant or tied depending on whether $(x_i - x_j)(y_i - y_j)$ is positive, negative or 0, respectively. Thus, let us define three $N \times N$ relations $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{T}$ as follows:

$$\mathcal{C}(i,j) = \begin{cases} 1 & (x_i - x_j)(y_i - y_j) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$\mathcal{D}(i,j) = \begin{cases} 1 & (x_i - x_j)(y_i - y_j) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$\mathcal{T}(i,j) = \begin{cases} 1 & (x_i - x_j)(y_i - y_j) = 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The number of concordant, discordant and tied pairs $(i, j) \in \mathcal{P}$ are then obtained by summing the entries in the corresponding relations:

$$\begin{aligned} C &= \sum_{(i,j)\in\mathcal{P}} \mathcal{C}(i,j) = \frac{1}{2} \sum_{i\in[N]} \sum_{j\in[N]} \mathcal{C}(i,j) \\ D &= \sum_{(i,j)\in\mathcal{P}} \mathcal{D}(i,j) = \frac{1}{2} \sum_{i\in[N]} \sum_{j\in[N]} \mathcal{D}(i,j) \\ T &= \sum_{(i,j)\in\mathcal{P}} \mathcal{T}(i,j) = \frac{1}{2} \sum_{i\in[N]} \sum_{j\in[N]} \mathcal{T}(i,j) - \frac{N}{2} \end{aligned}$$

Note that

$$\mathcal{C}(i,j) + \mathcal{D}(i,j) + \mathcal{T}(i,j) = 1 \tag{4}$$

for all $(i, j) \in \mathcal{P}$, and

$$C + D + T = |\mathcal{P}| = \frac{N(N-1)}{2} \ . \tag{5}$$

2.2 Rank Correlation Measures

Well-known examples of rank correlation measures that can be expressed in terms of the above quantities include Kendall's tau [11]

$$\tau = \frac{C - D}{N(N-1)/2} \tag{6}$$

and Goodman and Kruskal's gamma coefficient [8]

$$\gamma = \frac{C - D}{C + D} \ . \tag{7}$$

As will be detailed in the following sections, our basic strategy for generalizing rank correlation measures such as γ is to "fuzzify" the concepts of concordance and discordance. Thanks to the use of fuzzy order relations, we will be able to express that a pair (i, j) is concordant or discordant to a certain degree (between 0 and 1). Measures like (7) can then be generalized in a straightforward way, namely by accumulating the degrees of concordance and discordance, respectively, and putting them in relation to each other.

3 Fuzzy Relations

3.1 Fuzzy Equivalence

The notion of a fuzzy relation generalizes the standard notion of a mathematical relation by allowing to express "degrees of relatedness". Formally, a (binary) fuzzy relation on a set $\mathbb{X}$ is characterized by a membership function $\mathcal{E} : \mathbb{X} \times \mathbb{X} \longrightarrow [0, 1]$. For each pair of elements $x, y \in \mathbb{X}$, $\mathcal{E}(x, y)$ is the degree to which x is related to y.

Recall that a conventional equivalence relation on a set $\mathbb{X}$ is a binary relation that is reflexive, symmetric and transitive. For the case of a fuzzy relation $\mathcal{E}$, these properties are generalized as follows:

- reflexivity: $\mathcal{E}(x, x) = 1$ for all $x \in \mathbb{X}$
- symmetry: $\mathcal{E}(x, y) = \mathcal{E}(y, x)$ for all $x, y \in \mathbb{X}$
- $\top$-transitivity: $\top(\mathcal{E}(x, y), \mathcal{E}(y, z)) \leq \mathcal{E}(x, z)$ for all $x, y, z \in \mathbb{X}$

A fuzzy relation $\mathcal{E}$ having these properties is called a fuzzy equivalence relation [5]. While the generalizations of reflexivity and symmetry are rather straightforward, the generalization of transitivity involves a triangular norm (t-norm) $\top$, which plays the role of a generalized logical conjunction [13]. Formally, a function $\top : [0, 1]^2 \longrightarrow [0, 1]$ is a t-norm if it is associative, commutative, monotone increasing in both arguments, and satisfies the boundary conditions $\top(a, 0) = 0$ and $\top(a, 1) = a$ for all $a \in [0, 1]$. Examples of commonly used t-norms include the minimum $\top(a, b) = \min(a, b)$ and the product $\top(a, b) = ab$. To emphasize the role of the t-norm, a relation $\mathcal{E}$ satisfying the above properties is also called a $\top$-equivalence.

3.2 Fuzzy Ordering

The notion of an order relation $\leq$ is similar to that of an equivalence relation, with the important difference that the former is antisymmetric while the latter is symmetric. A common way to formalize antisymmetry is as follows: $a \leq b$ and $b \leq a$ implies $a = b$. Note that this definition already involves an equivalence relation, namely the equality $=$ of two elements. Thus, as suggested by Bodenhofer [2], a fuzzy order relation can be defined on the basis of a fuzzy equivalence relation. Formally, a fuzzy relation $\mathcal{L} : \mathbb{X} \times \mathbb{X} \longrightarrow [0, 1]$ is called a *fuzzy ordering* with respect to a t-norm $\top$ and a $\top$-equivalence $\mathcal{E}$, for brevity $\top$-$\mathcal{E}$-*ordering*, if it satisfies the following properties for all $x, y, z \in \mathbb{X}$:

- $\mathcal{E}$-reflexivity: $\mathcal{E}(x, y) \leq \mathcal{L}(x, y)$
- $\top$-$\mathcal{E}$-antisymmetry: $\top(\mathcal{L}(x, y), \mathcal{L}(y, x)) \leq \mathcal{E}(x, y)$
- $\top$-transitivity: $\top(\mathcal{L}(x, y), \mathcal{L}(y, z)) \leq \mathcal{L}(x, z)$

Furthermore a $\top$-$\mathcal{E}$-ordering $\mathcal{L}$ is called *strongly complete* if

$$\max\big(\mathcal{L}(x, y), \mathcal{L}(y, x)\big) = 1$$

for all $x, y \in \mathbb{X}$. This is expressing that, for each pair of elements x and y, either $x \leq y$ or $y \leq x$ should be fully true.

A fuzzy relation $\mathcal{L}$ as defined above can be seen as a generalization of the conventional "smaller or equal" on the real or the integer numbers. What is often needed, too, is a "stricly smaller" relation $<$. In agreement with the previous formalizations, a relation of that kind can be defined as follows: A binary fuzzy relation $\mathcal{R}$ is called a *strict fuzzy ordering* with respect to a $\top$-norm and a $\top$-equivalence $\mathcal{E}$, or strict $\top$-$\mathcal{E}$-*ordering* for short, if it has the following properties for all $x, x', y, y', z \in \mathbb{X}$ [5]:

- irreflexivity: $\mathcal{R}(x, x) = 0$
- $\top$-transitivity: $\top(\mathcal{R}(x, y), \mathcal{R}(y, z)) \leq \mathcal{R}(x, z)$
- $\mathcal{E}$-extensionality: $\top(\mathcal{E}(x, x'), \mathcal{E}(y, y'), \mathcal{R}(x, y)) \leq \mathcal{R}(x', y')$

3.3 Practical Construction

The above definitions provide generalizations $\mathcal{E}$, $\mathcal{L}$ and $\mathcal{R}$ of the standard relations $=$, $\leq$ and $<$, respectively, that exhibit reasonable properties and, moreover, are coherent with each other. Practically, one may start by choosing an equivalence relation $\mathcal{E}$ and a compatible t-norm $\top$, and then derive $\mathcal{L}$ and $\mathcal{R}$ from the corresponding $\top$-equivalence.

More specifically, suppose the set $\mathbb{X}$ to be a linearly ordered domain, that is, to be equipped with a standard (non-fuzzy) order relation $\leq$. Then, given a $\top$-equivalence $\mathcal{E}$ on $\mathbb{X}$, the following relation is a coherent fuzzy order relation, namely a strongly complete $\top$-$\mathcal{E}$-ordering:

$$\mathcal{L}(x, y) = \begin{cases} 1 & \text{if } x \leq y \\ \mathcal{E}(x, y) & \text{otherwise} \end{cases}$$

Moreover, a strict fuzzy ordering $\mathcal{R}$ can be obtained from $\mathcal{L}$ by

$$\mathcal{R}(x, y) = 1 - \mathcal{L}(y, x) \tag{8}$$

The relations thus defined have a number of convenient properties. In particular, $\min(\mathcal{R}(x, y), \mathcal{R}(y, x)) = 0$ and

$$\mathcal{R}(x, y) + \mathcal{E}(x, y) + \mathcal{R}(y, x) = 1 \tag{9}$$

for all $x, y \in \mathbb{X}$. These properties can be interpreted as follows. For each pair of elements x and y, the unit mass splits into two parts: a degree $a = \mathcal{E}(x, y)$ to which x and y are equal, and a degree $1 - a$ to which either x is smaller than y or y is smaller than x.

4 Fuzzy Relations on Rank Data

Since we are interested in generalizing rank correlation measures, the underlying domain $\mathbb{X}$ is given by a set of rank positions $[N] = \{1, 2, \ldots, N\}$ (equipped with the standard $<$ relation) in our case. As mentioned before, this domain could be equipped with fuzzy relations $\mathcal{E}$, $\mathcal{L}$ and $\mathcal{R}$ by defining $\mathcal{E}$ first and deriving $\mathcal{L}$ and $\mathcal{R}$ afterward. Note, however, that the number of degrees of freedom in the specification of $\mathcal{E}$ is of the order $O(N^2)$, despite the constraints this relation has to meet.

4.1 Scaling Functions on Rank Positions

In order to define fuzzy relations even more conveniently, while emphasizing the idea of weighing the importance of rank positions at the same time, we leverage the concept of a *scaling function* as proposed by Klawonn [12]. Roughly speaking, a scaling function $w : \mathbb{X} \longrightarrow \mathbb{R}_+$ specifies the dissimilarity of an element x from its direct neighbor elements, and the dissimilarity between any two elements x and y is then obtained via integration of the local dissimilarities along the chain from x to y. In our case, a scaling function can be defined as a mapping $w : [N-1] \longrightarrow [0, 1]$ or, equivalently, as a vector

$$\boldsymbol{w} = \Big(w(1), w(2), \ldots, w(N-1)\Big) \in [0, 1]^{N-1} \quad . \tag{10}$$

Here, $w(n)$ can be interpreted as the degree to which the rank positions n and $n-1$ are distinguished from each other; correspondingly, $1 - w(n)$ can be seen as the degree to which these two positions are considered to be equal. From the local degrees of distinguishability, a global distance function is derived on $\mathbb{X}$ by defining

$$d(x, y) = \min\left(1, \sum_{i=\min(x,y)}^{\max(x,y)-1} w(i)\right) \quad . \tag{11}$$

Put in words, the distance between x and y is the sum of the degrees of distinguishability between them, thresholded at the maximal distance of 1. In principle, accumulations of the degrees of distinguishability other than the sum are of course conceivable. For example, the maximum could be used as well:

$$d(x,y) = \max\Big\{w(i) \,|\, i \in \{\min(x,y),\ldots,\max(x,y)-1\}\Big\} \ . \tag{12}$$

In general, $d(x,y)$ is supposed to define a pseudo-metric on $\mathbb{X}$. Under this condition, it can be shown that the fuzzy relation $\mathcal{E}$ defined as

$$\mathcal{E}(x,y) = 1 - d(x,y)$$

for all $x, y \in \mathbb{X}$ is a $\top_L$-equivalence, where $\top_L$ is the Łukasiewicz t-norm $\top_L(a,b) = \max(0, a+b-1)$ [3]. Relations $\mathcal{L}$ and $\mathcal{R}$ can then be derived from $\mathcal{E}$ as described in Section 3.3. In particular, we obtain

$$\mathcal{R}(x,y) = \begin{cases} d(x,y) & \text{if } x < y \\ 0 & \text{otherwise} \end{cases}$$

According to our discussion so far, the only remaining degree of freedom is the scaling function s. Obviously, this function can also be interpreted as a *weighing function*: the more distinguishable a position n from its neighbor positions, i.e., the larger $w(n-1)$ and $w(n)$, the higher the importance of that position.

An example of a scaling function for $N = 12$ is shown in Figure 1. This function puts more emphasis on the top and the bottom ranks and less on the middle part. According to (11), the distinguishability between the positions 4 and 7 is $d(4,7) = 0.4 + 0.2 + 0.2 = 0.8$ (sum of the weights $w(i)$ in the shaded region). Thus, 4 is strictly smaller than 7 to the degree of $\mathcal{R}(4,7) = 0.8$, while both positions are considered equal to the degree $\mathcal{E}(4,7) = 0.2$.

Note that, with $w(i) = [\![i < k]\!]$, we also cover the top-k scenario as a special case. Here, the standard $<$ relation is recovered for all elements on the first k positions, whereas the remaining positions are considered as fully equivalent, i.e., these elements form an equivalence class in the standard sense.

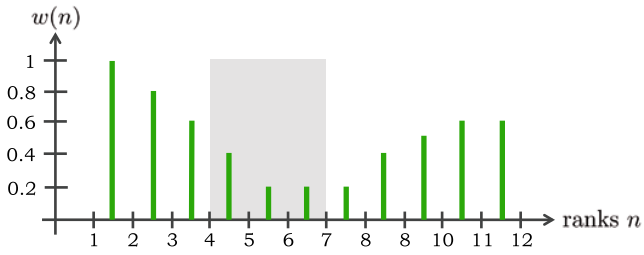


Fig. 1. Example of a scaling function.

5 Weighted Rank Correlation

Our approach to generalizing rank correlation measures is based on the "fuzzification" of the relations (1–3) and, correspondingly, the number of concordant, discordant and tied item pairs. The tools that are needed to do so have already been introduced in the previous sections. In particular, suppose a fuzzy equivalence relation $\mathcal{E}$ and a "strictly smaller" relation $\mathcal{R}$ to be derived from a scaling function w on $\mathbb{X}$, based on the procedure outlined above. For notational convenience, we assume the same scaling function (and hence the same relations) to be used on both domains $\mathbb{X}$ and $\mathbb{Y}$. In principle, however, different functions w_X and w_Y (and hence relations $\mathcal{E}_X$, $\mathcal{R}_X$ and $\mathcal{E}_Y$, $\mathcal{R}_Y$) could be used.

Now, according to (1), a pair $(i,j) \in \mathcal{P}$ is concordant if both x_i is (strictly) smaller than x_j and y_i is smaller than y_j, or if x_j is smaller than x_i and y_j is smaller than y_i. Using our fuzzy relation $\mathcal{R}$ and a t-norm $\top$ as a generalized conjunction, this can be expressed as follows:

$$\tilde{\mathcal{C}}(i,j) = \top\big(\mathcal{R}(x_i,x_j),\mathcal{R}(y_i,y_j)\big) + \top\big(\mathcal{R}(x_j,x_i),\mathcal{R}(y_j,y_i)\big) \tag{13}$$

The discordance relation can be expressed analogously:

$$\tilde{\mathcal{D}}(i,j) = \top\big(\mathcal{R}(x_i,x_j),\mathcal{R}(y_j,y_i)\big) + \top\big(\mathcal{R}(x_j,x_i),\mathcal{R}(y_i,y_j)\big) \tag{14}$$

Finally, the degree to which (i,j) is tied is given by

$$\tilde{\mathcal{T}}(i,j) = \bot\big(\mathcal{E}(x_i,x_j),\mathcal{E}(y_i,y_j)\big) \ ,$$

where $\bot$ is the t-conorm associated with $\top$ (i.e., $\bot(u,v) = 1 - \top(1-u,1-v)$), serving as a generalized logical disjunction. Generalizing (4), the three degrees sum up to 1, i.e.,

$$\tilde{\mathcal{C}}(i,j) + \tilde{\mathcal{D}}(i,j) + \tilde{\mathcal{T}}(i,j) \equiv 1 \ , \tag{15}$$

and either $\tilde{\mathcal{C}}(i,j) = 0$ or $\tilde{\mathcal{D}}(i,j) = 0$. In other words, a pair (i,j) that has originally been concordant (discordant) will remain concordant (discordant), at least to some extent. However, since $\mathcal{E}$ may introduce a certain indistinguishability between the positions x_i and x_j or the positions y_i and y_j, the pair could also be considered as a partial tie.

Given the above fuzzy relations, the number of concordant, discordant and tied data points can be obtained as before, namely by summing over all ordered pairs $(i,j) \in \mathcal{P}$:

$$\tilde{C} = \sum_{(i,j)\in\mathcal{P}} \tilde{\mathcal{C}}(i,j) \ , \quad \tilde{D} = \sum_{(i,j)\in\mathcal{P}} \tilde{\mathcal{D}}(i,j) \ , \quad \tilde{T} = \sum_{(i,j)\in\mathcal{P}} \tilde{\mathcal{T}}(i,j) \ .$$

According to (15),

$$\tilde{C} + \tilde{D} + \tilde{T} = |\mathcal{P}| = \frac{N(N-1)}{2} \ ,$$

which generalizes (5). Using these quantities, rank correlation measures expressed in terms of the number of concordant and discordant pairs can be generalized in a straightforward way. In particular, a generalization of the gamma coefficient (7) is obtained as

$$\tilde{\gamma} = \frac{\tilde{C} - \tilde{D}}{\tilde{C} + \tilde{D}} \ . \tag{16}$$

It is worth mentioning that the weighted rank correlation measure thus defined exhibits a number of desirable formal properties, which it essentially inherits from the general fuzzy extension of the gamma coefficient; we refer to [17], in which these properties are analyzed in detail.

6 Related Work

Weighted versions of rank correlation measures have not only been studied in statistics but also in other fields, notably in information retrival [6,10,16,21]. Most of them are motivated by the idea of giving a higher weight to the top-ranks: in information retrieval, important documents are supposed to appear in the top, and a swap of important documents should incur a higher penalty than a swap of unimportant ones.

Kaye [10] introduced a weighted, non-symmetric version of Spearman's rho coefficient. Costa and Soares [6] proposed a symmetric weighted version of Spearman's coefficient resembling the one of Kaye. Another approach, based on average precision and called *AP correlation*, was introduced by Yilmaz *et al.* [21]. Maturi and Abdelfattah [16] define weighted scores $W_i = w^i$ with $w \in (0, 1)$ and compute the Pearson correlation coefficient on these scores. All four measures give higher weight to the top ranks.

Two more flexible measures, not restricted to monotone decreasing weights, have been proposed by Shieh [19] and Kumar and Vassilivitskii [14]. In the approach of Shieh [19], a weight is manually given to every occurring concordance or discordance through a symmetric weight function $w : [N] \times [N] \longrightarrow \mathbb{R}_+$:

$$\tau_w = \frac{\sum_{i<j} w_{ij} C_{ij} - \sum_{i<j} w_{ij} D_{ij}}{\sum_{i<j} w_{ij}} = \frac{\sum_{i<j} w_{ij} (C_{ij} - D_{ij})}{\sum_{i<j} w_{ij}}. \tag{17}$$

The input parameter for w are the ranks of a reference ranking π_{ref}, which is assumed to be the natural order $(1, 2, 3, \ldots, N)$. Therefore, this approach is not symmetric. To handle the quadratic number of weights, Shieh proposed to define them as $w_{ij} = v_i v_j$ with v_i the weight of rank i.

Kumar and Vassilivitskii [14] introduce a generalized version of Kendall's distance. Originally, they proposed three different weights: element weights, position weights, and element similarities. The three weights are defined independently of each other, and each of them can be used by its own for weighting discordant pairs. Here, we focus on the use of position weights. Like in our approach, Kumar and Vassilivitskii define $N - 1$ weights $\delta_i \geq 0$, which are considered as costs for swapping two elements on adjacent positions $i + 1$ and i. The accumulated cost

of changing from position 1 to $i \in \{2, \ldots, N\}$ is $p_i = \sum_{j=1}^{i-1} \delta_j$, with $p_1 = 0$. Moreover,

$$\bar{p}_i(\pi_1, \pi_2) = \frac{p_{\pi_1(i)} - p_{\pi_2(i)}}{\pi_1(i) - \pi_2(i)} \tag{18}$$

is the average cost of moving element i from position $\pi_1(i)$ to position $\pi_2(i)$; if $\pi_1(i) = \pi_2(i)$ then $\bar{p}_i = 1$.The weighted discordance of a pair (i, j) is then defined in terms of the product of the average costs for index i and j:

$$\hat{D}_\delta(i,j) = \begin{cases} \bar{p}_i(\pi_1, \pi_2)\bar{p}_j(\pi_1, \pi_2) & \text{if } (i,j) \text{ is discordant} \\ 0 & \text{otherwise} \end{cases} . \tag{19}$$

Finally, the weighted Kendall distance K_δ is given by

$$K_\delta = \tilde{D}_\delta = \sum_{i=1}^{N-1} \sum_{i+1}^{N} \hat{D}_\delta(i,j) . \tag{20}$$

Note that (20) is indeed a distance and not a correlation measure. To enable a comparison with τ_ω and $\tilde{\gamma}$ in the next section, we define

$$\hat{C}_\delta(i,j) = \begin{cases} \bar{p}_i(\pi_1, \pi_2)\bar{p}_j(\pi_1, \pi_2) & \text{if } (i,j) \text{ is concordant} \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

as the weighted concordance of a pair (i, j), and finally another weighted version of gamma:

$$\tilde{\gamma}_\delta = \frac{\tilde{C}_\delta - \tilde{D}_\delta}{\tilde{C}_\delta + \tilde{D}_\delta} .$$

7 Experiments

Needless to say, an objective comparison of weighted rank correlation measures is very difficult, if not impossible. Even in the case of standard measures, one cannot say, for example, that Kendall's tau is "better" than Spearman's rho. Instead, these are simply different measures trying to capture different types of correlation in the data.

Nevertheless, we conducted some controlled experiments with synthetic data, for which there is a natural expectation of how the measures are supposed to behave and what results they should ideally produce. We compare our approach with those of Shieh as well as Kumar and Vassilivitskii, since these are able to handle non-monotone weight functions, too. For the purpose of these experiments, our measure $\tilde{\gamma}$ was instantiated with the maximum in (12) and the product t-norm in (13) and (14).[2]

[2] Of course, other instantiations are conceivable; however, tuning our measure by optimizing the choice of operators was beyond the scope of the experiments.

7.1 First Study

In a first experiment, we generated rank data by sampling from the Plackett-Luce (PL) model, which is a parameterized probability distribution on the set of all rankings over N items. It is specified by a parameter vector $\boldsymbol{v} = (v_1, v_2, \dots v_N) \in \mathbb{R}_+^N$, in which v_i accounts for the "skill" of the i^{th} item. The probability assigned by the PL model to a ranking represented by a permutation π is given by

$$\mathbf{P}(\pi \mid \boldsymbol{v}) = \prod_{i=1}^{N} \frac{v_{\pi^{-1}(i)}}{v_{\pi^{-1}(i)} + v_{\pi^{-1}(i+1)} + \ldots + v_{\pi^{-1}(N)}} \ , \tag{22}$$

where $\pi(i)$ is the position of item i in the ranking, and $\pi^{-1}(j)$ the index of the item on position j. This model is a generalization of the well-known Bradley-Terry model [15], a model for the pairwise comparison of alternatives, which specifies the probability that "a wins against b" in terms of $v_a/(v_a + v_b)$. Obviously, the larger v_a in comparison to v_b, the higher the probability that a is chosen. Likewise, the larger the parameter v_i in (22) in comparison to the parameters v_j, $j \neq i$, the higher the probability that the i^{th} item appears on a top rank. Moreover, the more similar the skill parameters, the more likely two items are reversed. Thus, a ranking drawn from a PL model is more stable, and hence more "reliable", in regions in which the difference between the skill values (sorted in decreasing order from highest to lowest) is large, and less stable in regions in which this difference is small.

Instead of defining the skills $\boldsymbol{v}$ directly, it is more convenient to define them via the representation of PL as a Thurstone model with scores following a Gumble distribution. The means μ_i of this distribution translate into PL-parameters via $v_i = \exp(\frac{\mu_i}{\beta})$, with β the scaling parameter of the Gumble distribution.

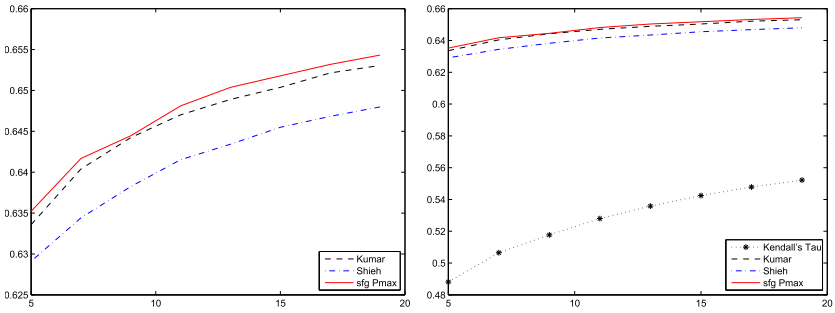
For our experimental study, we generated mixtures of $c = 4$ PL distributions, i.e., data sets consisting of four clusters. To this end, c reference rankings were first generated by sampling from the PL distribution with $\boldsymbol{\mu} = (30, 29, \dots, 1)$ and $\beta = 0.3$, i.e., these references are perturbations of the identity $\pi_{id} = (1, 2, \dots, n)$. Then, a score vector

$$\boldsymbol{\mu}^{(0)} = (18, \dots 14, 13.1, 12.3, \dots, 9.6, \dots, 9.6, 9.5, 9.3, \dots, 5.9, 5, \dots, 1)$$

is defined, which reflects high stability in the top and bottom ranks, and low stability in the middle ranks, and new score vectors $\boldsymbol{\mu}^{(i)}$, $i = 1, \dots, c$, are generated by permuting $\boldsymbol{\mu}^{(0)}$ according to the reference rankings; each of these rankings $\boldsymbol{\mu}^{(i)}$ defines the center of a cluster. Finally, 200 rankings are sampled from each of the PL models with parameter $\boldsymbol{\mu}^{(i)}$ and $\beta = 0.3$, and these rankings are assigned label i.

We produced 100 such data sets with rankings of length 30. As a weight vector, which corresponds to the scaling function (10) for $\tilde{\gamma}$ and defines the transitions costs δ_i for $\tilde{\gamma}_\delta$, we used

$$\boldsymbol{w} = (1, 1, 1, 1, 0.9, \dots, 0.1, 0, 0, 0, 0.1, \dots, 0.9, 1, 1, 1, 1) \ .$$



(a) (b)

Fig. 2. Both plots show the average accuracy against the neighborhood size k. The right plot additionally shows the results for classical Kendall's tau.

This weight vector seeks to account for the fact that, according to our construction, the middle positions of the observed rankings are less reliable and, therefore, should have a lower weight in the computation of similarities between rankings. The weight vector for τ_ω is derived from $\boldsymbol{w}$, so as to make it maximally comparable, and is given by $v_i = (w_{i-1} + w_i)/2$ with $w_{-1} = 1$ and $w_n = 1$.

For each correlation measure and each data set, we applied a k-nearest neighbor classifier with the correlation as a similarity measure. Here, the idea is the following: the better the similarity between rankings is reflected by a correlation measure, the stronger the performance of the classifier is supposed to be. The classifiers were validated by averaging one hundred repetitions of a 10-fold cross validation. In the end, we also averaged over all data sets. The results are shown in Figure 2. As can be seen, $\tilde{\gamma}_\delta$ and $\tilde{\gamma}$ are performing more or less on par, with a slight advantage for $\tilde{\gamma}$. Moreover, they both outperform τ_ω, which is nevertheless much better than the classical Kendall's tau (Figure 2(b)).

In Figure 3, two exemplary data sets are visualized using a kernel-PCA [18] for dimensionality reduction, using the different correlation measures to produce the similarity matrices. Every data point is colored according to its original class membership. As can be seen, the classical Kendall's tau is hardly able to separate the classes, whereas $\tilde{\gamma}_\delta$, $\tilde{\gamma}$, and τ_ω are at least able to separate three of the four classes. Despite following quite different approaches, the results of these three measures appear to be surprisingly similar.

7.2 Second Study

The second experiment is meant to explore the behavior of the rank correlation coefficients when comparing two rankings of a specific type. We compared a ranking $\pi_{id} = (1, 2, \ldots, 11, 12)$ with rankings $\pi_{i\to 1} = (2, 3, \ldots i, 1, i+1, \ldots, 12)$ in which the i^{th} item is moved from rank i to rank 1 and all items with index smaller i are shifted one position to the right. Each time the index i is incremented, another discordant pair is created, hence the similarity between π_{id} and

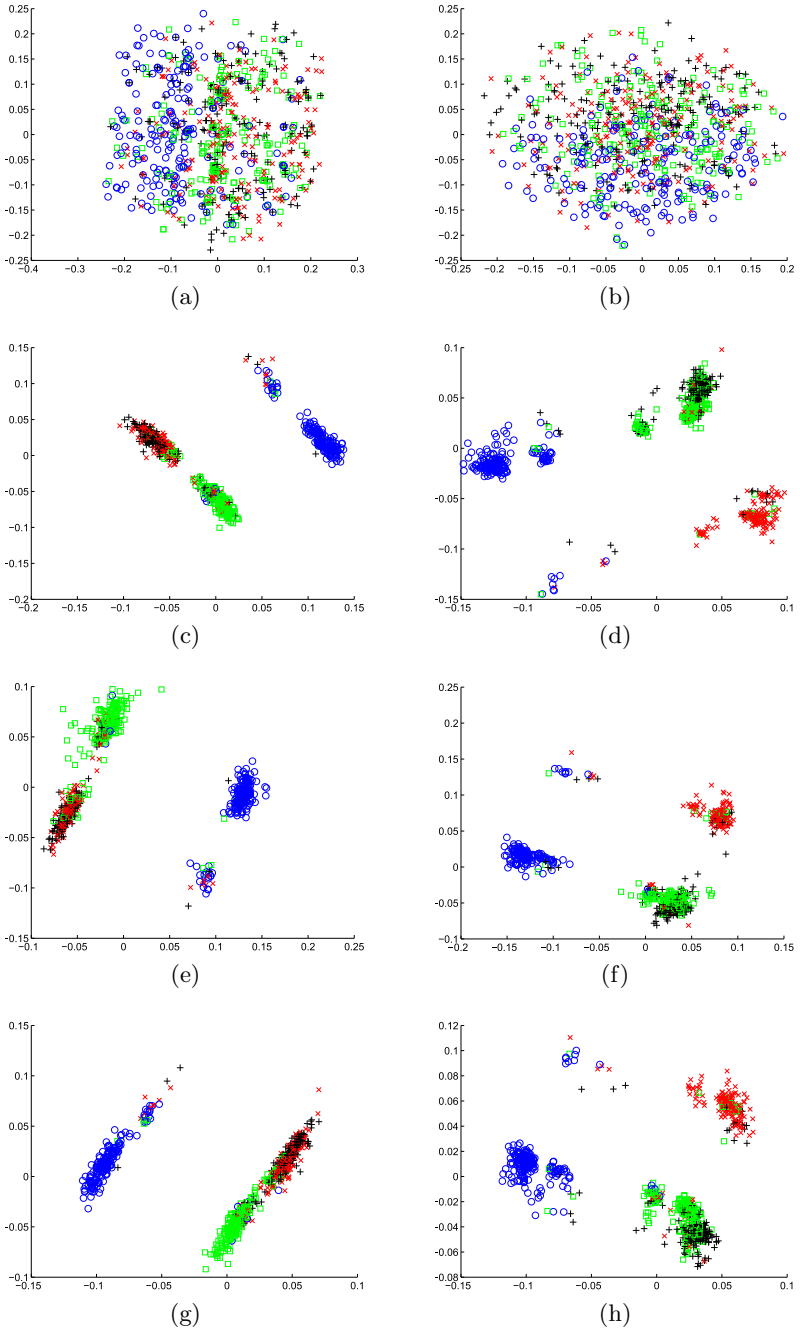


Fig. 3. Visualization of 30-dimensional ranking data sets. Every column shows one data set, every row one rank correlation coefficient. (a) – (b) Kendall's tau, (c) – (d) $\tilde{\gamma}_\delta$, (e) – (f) τ_ω, (g) – (h) $\tilde{\gamma}$.eps

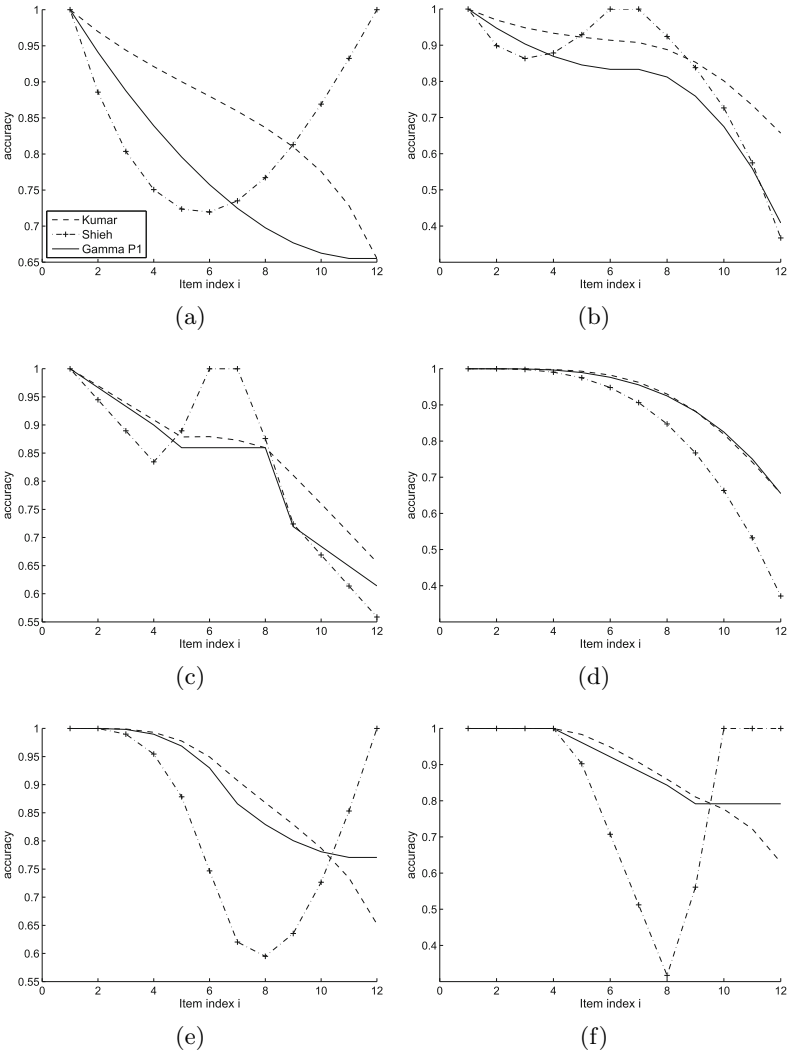


Fig. 4. Behavior of $\tilde{\gamma}_\delta$, τ_ω, and $\tilde{\gamma}$ in the "item i on rank 1" setting. The following weight vectors are used: (a) $(1\,0.9, \ldots, 0.1, 0)$, (b) $(1\,0.8, \ldots, 0, 0.2, \ldots, 1)$, (c) $(1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1)$, (d) $(0, 0.1, \ldots, 0.9, 1)$, (e) $(0, 0.2, \ldots, 1, 0.8, \ldots, 0)$, (f) $(0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0)$

$\pi_{i \to 1}$ should be monotone decreasing in i. Moreover, the higher the weight of the position i, the more pronounced the decrease should be.

Only $\tilde{\gamma}$ meets this expectation for all 6 weight vectors that have been considered (Figure 4). For instance, in Figure 4(a), $\tilde{\gamma}_\delta$ shows an increasing weighting of discordance with an increasing item index, although the weights are decreasing.

The strange behavior of τ_ω can be explained by the way in which weights $w_{ij} = v_i v_j$ are generated. In particular, as soon as one of the items has a small weight, all discordances in which this item is involved will have a small influence, too.

8 Conclusion and Future Work

We introduced a new approach to weighted rank correlation based on fuzzy order relations, as well as a concrete measure called *scaled gamma*. The latter allows for specifying the importance of rank positions in a quite flexible and convenient way by means of a scaling function. Thanks to the underlying formal foundation, such a scaling function immediately translates into a concrete version of our measure, in which the rank positions are processed within an appropriate weighting scheme.

First experimental studies with synthetic data are promising and suggest the usefulness of our approach. Experiments of this type will be continued in future work, not only with synthetic but also with real data. Moreover, let us again highlight that our extension of gamma is actually not a single measure but a family of measures, which is parameterized by the weight function w as well as the generalized logical conjunction (t-norm) used to define concordance and discordance. While the former will typically be specified as an external parameter by the user, the (fuzzy) logical operators offer an interesting degree of freedom that could be used to optimally adapt the measure to the application at hand. Again, this is an interesting direction for future work. Finally, going beyond the gamma coefficient, we also intend to apply our generalization to other rank correlation measures.

References

1. Balasubramaniyan, R., Hüllermeier, E., Weskamp, N., Kämper, J.: Clustering of gene expression data using a local shape-based similarity measure. Bioinformatics **21**(7), 1069–1077 (2005)
2. Bodenhofer, U.: A similarity-based generalization of fuzzy orderings preserving the classical axioms. Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems **8**(5), 593–610 (2000)
3. Bodenhofer, U.: Representations and constructions of similarity-based fuzzy orderings. Fuzzy Sets and Systems **137**, 113–136 (2003)
4. Bodenhofer, U., Demirci, M.: Strict fuzzy orderings with a given context of similarity. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems **16**(2), 147–178 (2008)
5. Bodenhofer, U., Klawonn, F.: Robust rank correlation coefficients on the basis of fuzzy orderings: Initial steps. Mathware & Soft Computing **15**, 5–20 (2008)
6. Pinto da Costa, J., Soares, C.: A weighted rank measure of correlation. Australian & New Zealand Journal of Statistics **47**(4), 515–529 (2005)
7. Fagin, R., Kumar, R., Sivakumar, D.: Comparing top k lists. SIAM Journal on Discrete Mathematics **17**(1), 134–160 (2003)
8. Goodman, L.A., Kruskal, W.H.: Measures of Association for Cross Classifications. Springer-Verlag, New York (1979)

9. Henzgen, S., Hüllermeier, E.: Weighted rank correlation measures based on fuzzy order relations. In: Hoffmann, F., Hüllermeier, E. (eds.) Proceedings 23. Workshop Computational Intelligence, pp. 227–236. KIT Scientific Publishing, Dortmund, Germany (2013)
10. Kaye, D.: A weighted rank correlation coefficient for the comparison of relevance judgements. Journal of Documentation **29**(4), 380–389 (1973)
11. Kendall, M.G.: Rank correlation methods. Charles Griffin, London (1955)
12. Klawonn, F.: Fuzzy sets and vague environment. Fuzzy Sets and Systems **66**, 207–221 (1994)
13. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Kluwer Academic Publishers (2002)
14. Kumar, R., Vassilvitskii, S.: Generalized distances between rankings. In: Proc. WWW, 19. International Conference on World Wide Web, pp. 571–580 (2010)
15. Marden, J.I.: Analyzing and Modeling Rank Data. CRC Press (1996)
16. Maturi, T.A., Abdelfattah, E.H.: A new weighted rank correlation. Journal of Mathematics and Statistics **4**(4), 226 (2008)
17. Dolorez Ruiz, M., Hüllermeier, E.: A formal and empirical analysis of the fuzzy gamma rank correlation coefficient. Information Sciences **206**, 1–17 (2012)
18. Schölkopf, B., Smola, A., Müller, K.R.: Kernel principal component analysis. In: Advances in Kernel Methods: Support Vector Learning, pp. 327–352. MIT Press (1999)
19. Shieh, G.S.: A weighted Kendall's tau statistic. Statistics & Probability Letters **39**(1), 17–24 (1998)
20. Spearman, C.: The proof and measurement for association between two things. Amer. Journal of Psychology **15**, 72–101 (1904)
21. Yilmaz, E., Aslam, J.A., Robertson, S.: A new rank correlation coefficient for information retrieval. In: Proc. 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 587–594. ACM (2008)