ClasSi: Measuring Ranking Quality in the Presence of Object Classes with Similarity Information*

Anca Maria Ivanescu, Marc Wichterich, and Thomas Seidl

Data Management and Data Exploration Group
Informatik 9, RWTH Aachen University, 52056 Aachen, Germany
{ivanescu,wichterich,seidl}@cs.rwth-aachen.de
http://www.dme.rwth-aachen.de/

Abstract. The quality of rankings can be evaluated by computing their correlation to an optimal ranking. State of the art ranking correlation coefficients like Kendall's τ and Spearman's ρ do not allow for the user to specify similarities between differing object classes and thus treat the transposition of objects from similar classes the same way as that of objects from dissimilar classes. We propose *ClasSi*, a new ranking correlation coefficient which deals with class label rankings and employs a class distance function to model the similarities between the classes. We also introduce a graphical representation of *ClasSi* akin to the *ROC curve* which describes how the correlation evolves throughout the ranking.

Keywords: ranking, quality measure, class similarity, ClasSi.

1 Introduction and Related Work

Evaluating the performance of an algorithm by comparing it against others is an important task in many fields such as in data mining and information retrieval. There are several evaluation methods developed for this purpose which can be integrated in the algorithm design process to improve effectiveness. Data mining and information retrieval models often return a ranking of the database objects. This ranking can be evaluated by checking if relevant documents are found before non relevant documents. Available measures for this evaluation are *precision* and *recall* as well as their weighted harmonic mean, known as the *F-measure* [9]. Related evaluation measures include the *mean average precision* [8], the *ROC curve* and the *area under the ROC curve (AUC)* [2]. These measures are all limited to binary class problems, distinguishing only between relevant and non relevant objects. Extensions of ROC to multi-class problems such as *generalized AUC* [4], the *volume under the curve* [1], and the *scalable multi-class ROC* [5] are combinations of two-class problems and do not consider class similarities.

* The authors gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center (SFB) 686 "Model-Based Control of Homogenized Low-Temperature Combustion" and DFG grant SE 1039/1-3.

L. Cao et al. (Eds.): PAKDD 2011 Workshops, LNAI 7104, pp. 185–196, 2012.

When evaluating object rankings, statistical methods to measure the correlation between two rankings can also be employed (e.g. *Kendall's* τ [6] and *Spearman* rank correlation coefficient ρ [11]). A positive correlation coefficient indicates an agreement between two rankings while a negative value indicates a disagreement. Variants of τ such as τ_b, τ_c [7], gamma (Γ) [3], and Somers' asymmetric coefficients [10] address the case of tied objects through different normalizations. However, these rank correlation measures only take the order of objects into account and the degree of similarity between objects is ignored.

In this work we propose *ClasSi*, a rank correlation coefficient which is capable of handling rankings with an arbitrary number of class labels and an arbitrary number of occurrences for each label. The main advantage of *ClasSi* is that it incorporates a class similarity function by which the user is able to define the degree of similarity between different classes.

In Section 2.1 we describe existing rank correlation coefficients τ and ρ. Section 2.2 defines *ClasSi* and Section 2.3 examines its properties, showing that all requirements of a rank correlation coefficient are met. In Section 2.4 we discuss how to compute *ClasSi* for the first k ranking positions, obtaining a graphical representation similar to the *ROC curve*. Section 3 analyzes the behavior of *ClasSi* in an experimental setup. Finally, in Section 4 we conclude the paper.

2 Ranking Quality Measures for Objects in Classes

We propose a new ranking correlation coefficient which allows for a user defined class distance measure and also comes with a graphical representation.

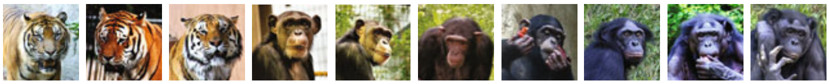
As mentioned before, the state of the art evaluation measures cannot handle class label rankings where objects in the database are grouped into classes according to some property that confers a notion of group similarity. For example, in an image similarity search system, if an object from a query test set is assigned the class label "bonobos", it only matters that other objects from the class "bonobos" are retrieved early on, but it does not matter which particular bonobo appears early on. In addition, objects from the similar class "chimpanzees" should appear before objects from the dissimilar class "tigers".

The preferred order of object classes in a ranking then depends on the class of the object for which the other objects were ranked (e.g., a query object in a similarity search scenario). An optimal ranking in the presence of classes is one where objects from the same class as the query object come first, then objects from the neighboring classes and so on. For example, a query from the class "bonobos" may have following optimal ranking where the order of the objects within the classes is arbitrary (b = bonobo, c = chimpanzee, t = tiger):

r_{bc}:

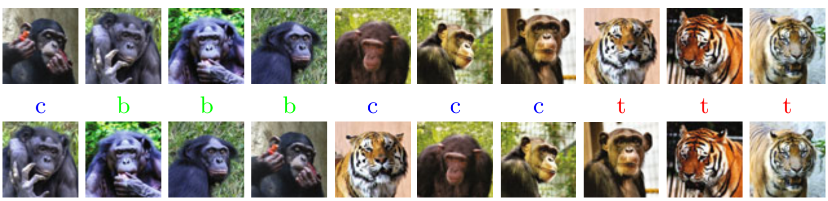
b b b c c c c t t t

A worst-case ranking is one in which objects from the most dissimilar class come first, then objects from one class closer and so on. The objects from the class coinciding with the query are the last ones in this ranking. For the same example query, a worst case ranking is:

r_{wc}:


To evaluate the quality of a computed ranking we look at its correlation to the optimal ranking. A positive correlation coefficient indicates a certain agreement between the two rankings. The higher the coefficient the more the ranking coincides with the optimal one, and if its value is 1 then it corresponds to it.

Additionally *ClasSi* accounts for the different degrees of dissimilarities between classes. For the example above, we can see that the classes "bonobos" and "chimpanzees" are more similar to each other, and both of them are dissimilar from the class "tigers". Consider the following two rankings:



Intuitively, the r_1 ranking coincides better with r_{bc} than r_2 does although both r_1 and r_2 have exactly one non bonobo object moved forward by three slots and r_1 has a non bonobo object at the very first position. The reason is that mistaking a tiger for a bonobo is a much bigger mistake than mistaking a chimpanzee for a bonobo.

2.1 Preliminaries: Measuring Ranking Quality

Throughout the paper we consider a database $DB = \{o_1, \ldots, o_m\}$ of cardinality m. A ranking of the objects in DB is defined as a bijective mapping r from DB to $\{1, \ldots m\}$ where $r(o)$ gives the position of object $o \in DB$ in the ranking and $r^{-1}(a)$ gives the a^{th} object from DB according to the ranking.

Measures such as Kendall's τ [6] and Spearman's rank correlation coefficient [11] assess the correlation between two ranking. As they will serve as the basis for the measure proposed in Section 2.2, they shall be reviewed here shortly.

Definition 1. ***Kendall's*** τ***.*** *Given a database* $DB = \{o_1, \ldots, o_m\}$ *and two ranking* r_1, r_2*, Kendall's correlation coefficient* τ *is defined as*

$$\tau = \frac{Con - Dis}{\frac{1}{2}m(m-1)} = 1 - \frac{2 \cdot Dis}{\frac{1}{2}m(m-1)}$$

with

$$Con = |\{\left(r_1^{-1}(a), r_1^{-1}(b)\right) \mid 1 \leq a < b \leq m \wedge r_1^{-1}(a) \prec_{r_2} r_1^{-1}(b)\}| \text{ and}$$

$$Dis = |\{\left(r_1^{-1}(a), r_1^{-1}(b)\right) \mid 1 \leq a < b \leq m \wedge r_1^{-1}(b) \prec_{r_2} r_1^{-1}(a)\}|$$

where $o_a \prec_r o_b \Leftrightarrow r(o_a) < r(o_b)$*.*

Kendall's τ can be used to evaluate the quality of a ranking r by comparing r to an optimal ranking r^*. Kendall's τ then measures the correlation of these two rankings by counting the number of concordant object pairs (those, that are sorted the same way by both rankings) minus the number of discordant object pairs (those that are sorted differently by the two rankings). The result is then divided by the total number of pairs ($\frac{1}{2}m(m-1)$) to normalize the measure between -1 and 1 where 1 is reached for identical rankings and -1 for reversed rankings.

While Kendall's τ takes the number of discordant pairs into account Spearman's rank correlation coefficient ρ explicitly considers the difference in ranking positions when comparing two rankings.

Definition 2. ***Spearman's*** ρ***.*** *Given a database* $DB = \{o_1, \ldots, o_m\}$ *and two ranking* r_1, r_2*, Spearman's rank correlation coefficient* ρ *is defined as*

$$\rho = 1 - \frac{6 \sum_{o \in DB}(r_1(o) - r_2(o))^2}{m(m^2-1)}.$$

Similar to Kendall's τ, Spearman's ρ is normalized between -1 and 1. Even though the difference in ranking position is considered by ρ, it is conceivable that two mismatches (each for example by ten ranking positions) may be of notably differing importance to the quality of the rankings. In the following we propose to incorporate knowledge on cross object (class) similarity into the evaluation of rankings. This allows for a more meaningful assessment of for instance similarity search results, which currently are mostly evaluated using either simple precision/recall measures or ranking quality measures that ignore class similarity information.

2.2 Class Similarity Ranking Correlation Coefficient ClasSi

We next introduce according class labeling and comparison functions that help us assess the quality of rankings in this scenario. The function $l : DB \rightarrow C$ assigns a class label from $C = \{c_1, \ldots c_n\}$ to each object $o \in DB$. The class distance function $d : C \times C \rightarrow \mathbb{R}$ conveys the notion of (dis)similarity between individual classes (e.g., $d(\text{bonobos}, \text{bonobos}) = 0$ and $d(\text{bonobos}, \text{chimpanzees}) < d(\text{bonobos}, \text{tigers})$). Based on the class distance function and a query object, the best case and worst case rankings are defined as follows.

Definition 3. ***Optimal class ranking.*** *For a query object q with class label c a ranking r_{bc} is said to be optimal iff*

$$d(c, l(r^{-1}(a))) \leq d(c, l(r^{-1}(a+1))) \qquad \forall a \in \{1, \ldots, m-1\}$$

where $l(r^{-1}(a))$ is the label of the a^{th} object according to ranking r.

Definition 4. ***Worst-case class ranking.*** *For a query object q with class label c a ranking r_{wc} is said to be a worst-case ranking iff*

$$d(c, l(r^{-1}(a))) \geq d(c, l(r^{-1}(a+1))) \qquad \forall a \in \{1, \ldots, m-1\}.$$

The *ClasSi* ranking correlation coefficient not only takes into consideration the number of discordant pairs but also as their dissimilarities. The dissimilarity of a discordant pair of class labels c_i and c_j is appraised by looking at their distances to the query class label c_q.

$$cost(i, j) = \begin{cases} 0, & \text{if } d(c_q, c_i) \leq d(c_q, c_j), \\ d(c_q, c_i) - d(c_q, c_j), & \text{else.} \end{cases}$$

For a given class ranking r as defined above, we compute *ClasSi* by iterating through all positions and sum up the dissimilarity cost for each discordant pair.

Definition 5. *Given a database $DB = \{o_1, \ldots, o_m\}$, a class distance function $d : C \times C :\longrightarrow \mathbb{R}$, a query object q which defines best and worst case rankings r_{bc} and r_{wc}, and the dissimilarity cost function $cost : C \times C :\longrightarrow \mathbb{R}$, the ClasSi correlation coefficient between an arbitrary ranking r and r_{bc} is defined as*

$$ClasSi = 1 - \frac{2 \cdot DisCost_r}{DisCost_{r_{wc}}}$$

where $DisCost_r$ is the cost generated by the discordant pairs of r compared to r_{bc} and $DisCost_{r_{wc}}$ is the according cost generated by the worst case ranking:

$$DisCost_r = \sum_{a=1}^{m} \sum_{b=a+1}^{m} cost(l(r^{-1}(a)), l(r^{-1}(b)))$$

$$DisCost_{r_{wc}} = \sum_{a=1}^{m} \sum_{b=a+1}^{m} cost(l({r_{wc}}^{-1}(a)), l({r_{wc}}^{-1}(b)))$$

For the example above we define following class distance measure: $d(b, c) = 1$, $d(b, t) = 6$. The dissimilarity costs between classes in this case are: $cost(c, b) = 1$, $cost(t, b) = 6$, $cost(t, c) = 5$, and 0 for all other cases. To compute *ClasSi* we iterate through the ranking positions and sum up the cost of the discordant pairs. For *ClasSi* between r_1 and r_{bc} we count 3 discordant pairs: there are 3 labels b which occur after a c label, thus $DisCost_{r_1} = 3 \cdot cost(c, b) = 3$. Between r_{bc} and r_{wc} all possible discordant pairs occur. The corresponding cost is $DisCost_{wc} = 4 \cdot 3 \cdot cost(t, c) + 3 \cdot 3 \cdot cost(t, b) + 3 \cdot 4 \cdot cost(c, b) = 126$. The $ClasSi$ correlation coefficient is then $ClasSi_{r_1} = 1 - \frac{2 \cdot 3}{126} = 0.95$. For r_2 there are 3 labels c which occur after a t label, thus $DisCost_{r_2} = 3 \cdot cost(t, c) = 15$. We obtain $ClasSi_{r_2} = 1 - \frac{2 \cdot 15}{126} = 0.76$ which is considerable smaller than $ClasSi_{r_1}$, since the dissimilarity between a tiger and a bonobo is much higher than the dissimilarity between a chimpanzee and a bonobo.

2.3 Properties of ClasSi

After introducing $ClasSi$ in the previous section, we now show that it has all the properties of a correlation coefficient. With $ClasSi$ we measure the correlation between an arbitrary ranking r and the optimal ranking r_{bc} of a set of class labels. If r is also an optimal ranking then $ClasSi$ equals 1 as the two are perfectly correlated. If r is a worst case ranking then $ClasSi$ equals -1 as they perfectly disagree. Finally, if r is a random ranking then the expected value of $ClasSi$ is 0 which means that the two rankings are independent.

Theorem 1. *The $ClasSi$ correlation coefficient between ranking r and the optimal ranking r_{bc} is 1 if r corresponds to r_{bc}:*

$$l(r^{-1}(a)) = l(r_{bc}^{-1}(a)) \quad \forall a \in 1, ..., m$$

Proof. If r corresponds to the optimal ranking, then there are no discordant pairs and thus no dissimilarity cost. In this case:

$$ClasSi = 1 - \frac{2 \cdot 0}{DisCost_{r_{wc}}} = 1$$

Theorem 2. *The $ClasSi$ correlation coefficient between ranking r and the optimal ranking r_{bc} is -1 if r corresponds to r_{wc}:*

$$l(r^{-1}(a)) = l(r_{wc}^{-1}(a)) \quad \forall a \in 1, ..., m$$

Proof. If r corresponds to the worst case ranking, then $DisCost_r = DisCost_{r_{wc}}$ and in this case:

$$ClasSi = 1 - \frac{2 \cdot DisCost_{r_{wc}}}{DisCost_{r_{wc}}} = -1$$

Theorem 3. *The expected correlation coefficient $E(ClasSi)$ between a random ranking r and the optimal ranking r_{bc} is 0.*

Proof. Assume w.l.o.g. that there are m_i objects with label c_i. Then for each object with label c_i there are Dis_i possible objects with a different label which are more similar to the query object and would be discordant if they were to be ranked after the c_i-labeled objects. More formally:

$$Dis_i = | \{o_a \mid d(c_q, c_i) > d(c_q, l(o_a)), \forall 1 \leq a \leq m\} |$$

The probability for the occurrence of a discordant pair can be modeled by means of the *hypergeometric distribution*. For a sequence of s drawings without replacement from a statistical population with S entities, out of which M have a certain property, the hypergeometric distribution describes the probability that k is the number of successful draws, i.e. the number of draws having that property:

$$P(X = k) = \frac{\binom{M}{k}\binom{S-M}{s-k}}{\binom{S}{s}} \qquad E(X) = s\frac{M}{S}$$

Let us consider position $m - e$ in the class ranking r which is followed by e entries and assume label $l(r^{-1}(m - e)) = c_i$ at this position. The probability that there are k discordant entries among the e following entries, is according to the hypergeometric distribution

$$P(Dis = k|c_i) = \frac{\binom{Dis_i}{k}\binom{m-1-Dis_i}{e-k}}{\binom{m-1}{e}} \tag{1}$$

The expected number of discordant entries occurring in the remaining e entries is then:

$$E_e(Dis|c_i) = \sum_{k=0}^{e} k\frac{\binom{Dis_i}{k}\binom{m-1-Dis_i}{e-k}}{\binom{m-1}{e}} = e\frac{Dis_i}{m-1} \tag{2}$$

For each object with label c_i we compute the average cost of a discordant pair:

$$\overline{cost}_i = \frac{1}{Dis_i} \cdot \sum_{\substack{1 \le j \le n \\ d(c_q,c_i) > d(c_q,c_j)}} m_j \cdot cost(i,j).$$

The expected associated cost at position $m - e$ is obtained by multiplying the expected number of discordant pairs with the average cost $\overline{cost}_i$ of a discordant pair for the label c_i:

$$E_e(DisCost_r|c_i) = e\frac{Dis_i \cdot \overline{cost}_i}{m-1} \tag{3}$$

For an arbitrary label and e entries left, the expected associated cost is:

$$E_e(DisCost_r) = \sum_{i=1}^{n} p_i e\frac{Dis_i \cdot \overline{cost}_i}{m-1} \tag{4}$$

where $p_i = \frac{m_i}{m}$ denotes the a priori class probability. If the expected costs at each position are summed up, then the expected costs generated by the expected number of discordant entries is obtained as

$$\begin{aligned} E(DisCost_r) &= \sum_{e=1}^{m-1}\sum_{i=1}^{n} p_i \cdot e\frac{Dis_i \cdot \overline{cost}_i}{m-1} \\ &= \frac{1}{m-1}\sum_{e=1}^{m-1} e \sum_{i=1}^{n} p_i \cdot Dis_i \cdot \overline{cost}_i \\ &= \frac{1}{m-1}\frac{(m-1)m}{2}\sum_{i=1}^{n} p_i \cdot Dis_i \cdot \overline{cost}_i \end{aligned}$$

$$= \frac{m \sum_{i=1}^{n} p_i \cdot Dis_i \cdot \overline{cost}_i}{2}$$

Knowing that m_i is the number of entries with label c_i

$$E(DisCost_r) = \frac{m \sum_{i=1}^{n} p_i \cdot Dis_i \cdot \overline{cost}_i}{2} = \frac{\sum_{i=1}^{n} m_i \cdot Dis_i \cdot \overline{cost}_i}{2}$$

At this point we obtained the following expected correlation coefficient:

$$E(ClasSi) = 1 - \frac{2 \cdot \frac{\sum_{i=1}^{n} m_i \cdot Dis_i \cdot \overline{cost}_i}{2}}{DisCost_{r_{wc}}}$$

Considering that for r_{wc} we have all possible discordant pairs, the associated dissimilarity cost can be computed by iterating over all n classes and taking all their objects, their possible discordant pairs, and their average cost of discordant pairs into account:

$$DisCost_{r_{wc}} = \sum_{i=1}^{n} m_i \cdot Dis_i \cdot \overline{cost}_i$$

Thus, the expected correlation coefficient between r and r_{bc} is

$$E(ClasSi) = 1 - \frac{2 \cdot \frac{\sum_{i=1}^{n} m_i \cdot Dis_i \cdot \overline{cost}_i}{2}}{\sum_{i=1}^{n} m_i \cdot Dis_i \cdot \overline{cost}_i} = 0.$$

Thus a ranking returned by a similarity measure can be also assessed by considering the $ClasSi$ correlation coefficient to the optimal ranking. Since for a random ranking the expected $ClasSi$ value is 0, a computed ranking should have a higher $ClasSi$ correlation coefficient to the optimal ranking.

Another important property of $ClasSi$ is that it not only considers the number of discordant pairs, but also the degree of their dissimilarity. By specifying the class distance function, the user specifies different degrees of dissimilarity for the discordant pairs. Nevertheless, only the relative differences matter.

Theorem 4. *Let d and d' be two class distance functions such that $d'(c_i, c_j) = \alpha \cdot d(c_i, c_j)$, and $ClasSi^{(d)}$ and $ClasSi^{(d')}$ be the corresponding rank correlation coefficients, then:*

$$ClasSi^{(d)} = ClasSi^{(d')}$$

Proof. From the relationship between the class distance functions we also obtain following relationship between the dissimilarity cost functions

$$cost'(c_i, c_j) = \alpha \cdot cost(c_i, c_j)$$

Thus the scaling of $DisCost'_r = \alpha \cdot DisCost_r$ and of $DisCost'_{r_{wc}} = \alpha \cdot DisCost_{r_{wc}}$ cancel each other.

2.4 ClasSi on Prefixes of Rankings

Up to now, the *ClasSi* measure has been computed for a complete ranking of objects in a database yielding a single value that reflects the overall quality of the ranking. In some situations, it might be more interesting to have a quality score for a subset of the ranked objects. The first k positions of a ranking are of particular interest, since only these results might either be presented to a user (e.g., in a retrieval system) or be considered in a data mining process. The proposed *ClasSi* measure can easily be adapted to suite this need in a meaningful manner. Instead of measuring the cost of misplaced objects for the whole ranking, the *ClasSi* measure restricted to the top k positions measures the guaranteed cost of objects placed in the first k positions. That is, for each object o within the top k objects, it is checked how much cost will be generated due to objects o' appearing after o when they were supposed to appear before o in the ranking. Likewise, the cost of the worst case scenario is restricted to the cost guaranteed to be generated by the top k objects of the worst case ranking.

Definition 6. *Given a database $DB = \{o_1, \ldots, o_m\}$, a class distance function $d : C \times C :\rightarrow \mathbb{R}$, a query object q which defines best and worst case rankings r_{bc} and r_{wc}, and the dissimilarity cost function $cost : C \times C :\rightarrow \mathbb{R}$, the ClasSi correlation coefficient for the top k positions between a ranking r and r_{bc} is defined as*

$$ClasSi_k = 1 - \frac{2 \cdot DisCost_{r,k}}{DisCost_{r_{wc},k}}$$

where $DisCost_{r,k}$ is the cost generated by the discordant pairs of r rooted within the top k positions of r and $DisCost_{r_{wc},k}$ is the according cost generated by the worst case ranking:

$$DisCost_{r,k} = \sum_{a=1}^{k} \sum_{b=a+1}^{m} cost(l(r^{-1}(a)), l(r^{-1}(b)))$$

$$DisCost_{r_{wc},k} = \sum_{a=1}^{k} \sum_{b=a+1}^{m} cost(l({r_{wc}}^{-1}(a)), l({r_{wc}}^{-1}(b)))$$

Algorithms 1 and 2 return arrays filled with cumulative discordant pair costs and ClasSi values in $O(k * m)$ time. By keeping track of the number of objects seen for each class up to position k in $O(k * n)$ space, it is possible to speed up the computation to $O(k * n)$ if the number of objects per class is known a priori.

By plotting $ClasSi_k$ for all $k \in \{1, \ldots, m\}$ we obtain a curve, which describes how the correlation evolves. If the first $ClasSi_k$ values are small and the curve is growing, this means that most of the discordant pairs are at the beginning and towards the end the ranking agrees with the optimal one. If the curve is decreasing, this means that the quality of the ranking is better at the beginning and decreases towards the end of the ranking.

Algorithm 1. DisCost(Prefix k, Ranking r, Labeling l, Costs $cost$)

```
dc = Float[k];
for a = 1 to k do     // iterate over position pairs (a,b) in ranking r
    dc[a] = 0;
    for b = a+1 to m do
        dc[a] += cost[l(r^-1(a))][l(r^-1(b))];     // sum discordant pair costs
    end for
    if (a > 1) then dc[a] += dc[a-1];
end for
return dc;
```

Algorithm 2. ClasSi(Prefix k, Rankings r, r_{wc}, Labeling l, Costs $cost$)

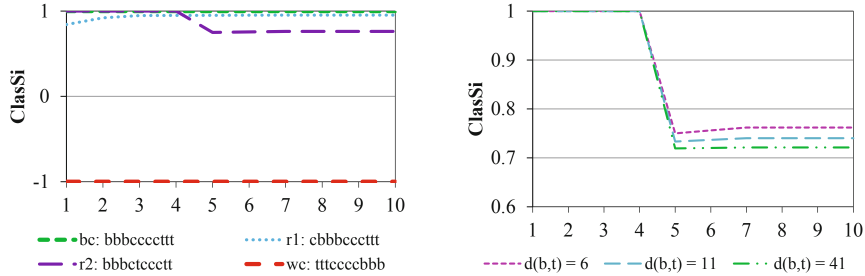
```
classi = Float[k];
rc = DistCost(k, r, l, cost);
wc = DistCost(k, r_wc, l, cost);
for a= 1 to k do
    classi[a] = 1 - (2 * rc[a] / wc[a]);
end for
return classi;
```

3 Examples

Using the values given by Algorithm 2, it is possible to track the progression of the $ClasSi$ value for ascending values of k. Figure 1(a) shows graphs for four rankings as described in Section 2.2. The best case ranking results in a constant graph at value +1.0. Analogously, the $ClasSi_k$ values for the worst case ranking are constant at -1.0. Ranking r_1 with one c (i.e., chimpanzee) moved forward by 3 positions to the first position results in a graph that starts at 0.84 and then increases up to a value of 0.95 as all further positions do not include any more discordant pairs while the number (and cost) of potential discordant pairs grows. Ranking r_2 on the other hand starts with four objects ranked identically to the best case ranking, thus the resulting curve starts at 1. On the fifth position an object with high cost for discordant pairs appears and significantly reduces the quality of the ranking to 0.75.

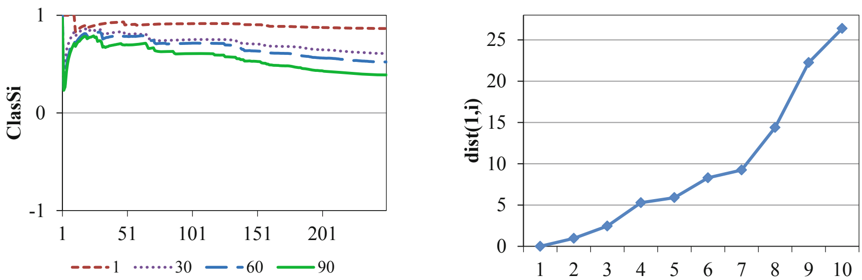
The dissimilarity between "bonobos" and "tigers" is specified by the user through the class distance function. In Figure 1(b) we see how the $ClasSi$-curve for r_2 drops when the distance $d(b, t)$ is increased while $d(b, c)$ remains constant. The higher $d(b, t)$, the smaller $ClasSi$ gets.

We further investigate the behavior of $ClasSi$ for an increasing number of discordant pairs. We consider a synthetically generated optimal ranking and another one, which emerged from the optimal one by randomly choosing a pair of objects and switching their position. The rankings have 250 entries and are divided into 10 classes. The randomly generated class distance function from the target class to the other classes is plotted in Figure 2(b). In Figure 2(a) $ClasSi$ curves are plotted for an increasing number of discordant pairs and we can see that the $ClasSi$ values decrease with an increasing number of discordant pairs.



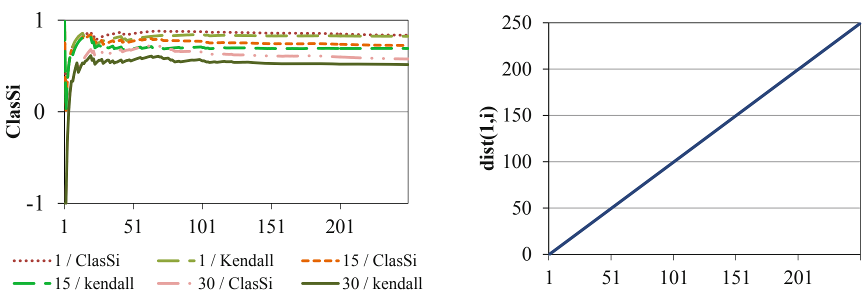
(a) (b)

Fig. 1. *ClasSi* in our example from Section 2 for varying k



(a) *ClasSi* for 1 up to 90 permutations (b) random class distance function

Fig. 2. *ClasSi* for an increasing number of object permutations



(a) *ClasSi* and *Kendall's* τ for 1 up to 30 permutations (b) Simple class distance function

Fig. 3. *ClasSi* for the simple case

Note that although *ClasSi* can deal with multiple entries with the same label and allows the user to define the class similarities, this coefficient can be also used for rankings in which each entry occurs only once and/or the user only specifies the desired order of classes. In this particular case, *ClasSi* behaves similarly to

Kendall's τ as it can be seen in Figure 3(a). The simple class distance function resulted from the specified class ordering is plotted in Figure 3(b).

4 Conclusion

In this paper we introduced a new measure to evaluate rankings of class labels by computing the correlation to an optimal ranking. It also allows for the user to specify different similarities between different classes. We have also proven that *ClasSi* has all the properties required for a correlation coefficient. *ClasSi* can be computed by iterating through the ranking and can be stopped at every position k, delivering an intermediate result $ClasSi_k$. By plotting these values we obtain a representation akin to the *ROC curve* from which we can recognize where the agreements and disagreements w.r.t. the optimal ranking occur.

References

1. Ferri, C., Hernández-Orallo, J., Salido, M.A.: Volume under the ROC surface for multi-class problems. In: Proceedings of 14th European Conference on Machine Learning, pp. 108–120 (2003)
2. Flach, P., Blockeel, H., Ferri, C., Hernández-Orallo, J., Struyf, J.: Decision support for data mining; introduction to ROC analysis and its applications. In: Data Mining and Decision Support: Integration and Collaboration, pp. 81–90. Kluwer Academic Publishers (2003)
3. Goodman, L.A., Kruskal, W.H.: Measures of association for cross classifications. Journal of the American Statistical Association 49(268), 732–764 (1954)
4. Hand, D.J., Till, R.J.: A simple generalisation of the area under the ROC curve for multiple class classification problems. Machine Learning 45, 171–186 (2001)
5. Hassan, M.R., Ramamohanarao, K., Karmakar, C.K., Hossain, M.M., Bailey, J.: A Novel Scalable Multi-Class ROC for Effective Visualization and Computation. In: Zaki, M.J., Yu, J.X., Ravindran, B., Pudi, V. (eds.) PAKDD 2010. LNCS, vol. 6118, pp. 107–120. Springer, Heidelberg (2010)
6. Kendall, M.: A new measure of rank correlation. Biometrika 30(1-2), 81–89 (1938)
7. Kendall, M., Gibbons, J.D.: Rank Correlation Methods. Edward Arnold (1990)
8. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press (2008)
9. van Rijsbergen, C.J.: Information Retrieval, 2nd edn. Butterworths, London (1979)
10. Somers, R.H.: A new asymmetric measure of association for ordinal variables. American Sociological Review 27(6), 799–811 (1962)
11. Spearman, C.: The proof and measurement of association between two things. The American Journal of Psychology 100, 441–471 (1987)