EVALUATION OF RANKING SIMILARITY IN ORDINAL RANKING PROBLEMS

Jiří Mazurek

Klíčová slova:
Korelační koeficient, podobnost pořadí, pořadí, problém s ordinálním pořadím, rozhodování.

Keywords:
correlation coefficient, decision making, ordinal ranking problem, ranking, ranking similarity

Abstrakt:
V problémech s ordinálním pořadím jsou objekty, alternativy, produkty, služby apod. seřazeny několika experty a cílem je získat z těchto (obecně odlišných) pořadí jedno výsledné (konsensuální) pořadí. Dosažení tohoto cíle nicméně závisí na stupni shody mezi jednotlivými pořadími. Pokud jsou pořadí náhodná, nemůžeme očekávat dosažení smysluplné shody, ale pokud jsou si pořadí „blízká“ a vyjadřují shodu mezi experty, pak má výsledné konsensuální pořadí smysl. Cílem tohoto článku je ukázat způsoby hodnocení podobnosti pořadí, které je založeno na použití Kendallova τ a W, Spearmanova ρ, Pearsonova r a skalárního součinu vektorů. Jsou zde diskutovány případy bez remíz (shodných hodnocení) i s remízami, stejně tak jako problém podobnosti mezi dvěma neúplnými pořadími (nejlepšími k pořadími). Vysvětlení jsou založena na příkladech.

Abstract:
In ordinal ranking problems objects, alternatives, products, services, etc. are ranked by several experts and the goal is to convert a set of (generally different) rankings into the final group consensus ranking. However, this goal depends on a degree of agreement among rankings. With random rankings one cannot expect to get meaningful consensus, but if rankings are „close“ and represent agreement between experts, then the final group consensus has much more sense. The aim of this article is to present the evaluation of similarity among rankings, which is based on Kendall's τ and W, Spearman's ρ, Pearson's r and dot product of vectors. Cases without and with ties are discussed as well as a problem of similarity between incomplete rankings (top k lists). Explanations are based on examples.

1 Introduction

In many situations people, products, services, companies, countries, etc. are ranked (rated) from the best to the worst with regard to some set of criteria such as achievement, price, profit, population, etc. In general, such settings form ordinal ranking problems dealing with ordinal variables, which assign numbers to objects or events which represent their rank order (1^{st}, 2^{nd}, 3^{rd}, etc.) according to some criterion. With more than one ranking or more than one criterion two problems arise:

- To get one final consensus ranking of objects (this constitutes well-known *ordinal consensus ranking problem*).
- To evaluate agreement (consistency) among experts by evaluating similarity among rankings.

As for the first problem, many methods converting set of rankings into one consensus ranking were proposed, such as Borda-Kendall's method of marks, CRM, DCM, MAH etc., and there exists comprehensive literature on the topic ([2], [9]). However, little attention have been paid to examination of agreement among these methods which do not always yield the same results [9]. It can be only speculated that the methods give the same (or very close) results when there is a high degree of agreement among experts (and rankings are similar) while the methods differ in cases with random (nonsimilar) expert's rankings. There are only few papers on consistency among experts' ratings too (e.g. [8]).

Therefore, the aim of this article is to focus on the second problem, which has been rather neglected so far. In Section 2 concepts of similarity, dissimilarity and distance are introduced. Evaluation of similarity between two rankings without ties is involved in Section 3 and with ties in Section 4. Section 5 examines similarity of more than two rankings and finally Section 6 is devoted to brief discussion of similarity between two incomplete rankings. For better understanding explanations are based on examples.

2 Similarity, dissimilarity and its measurement

Similarity is a measure of proximity between two or more objects or variables. On the other hand, dissimilarity is a measure of difference (distance) between objects or variables. Both measures are conventionally normalized to take the value from intervals $\langle 0,1\rangle$ or $\langle -1,1\rangle$. The dissimilarity can be measured by a function called *distance* or *metric*.

A *metric* on a set X is a function $d : X \times X \rightarrow \mathbf{R}$, which satisfies the following conditions:

1. $d(x, y) \geq 0$
2. $d(x, y) = 0$ if and only if $x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) \leq d(x, y) + d(y, z)$, for $x, z, y \in X$.

Function d satisfying only conditions 1. to 3. is called *distance*.

Ordinal data can be represented as (ordered) sets, vectors or matrices. For all formats there exist suitable distances. Distance between two sets A nad B can be measured e.g. via Jaccard metric [6]:

$$d(A,B) = \frac{|A \Delta B|}{|A \cup B|},$$

where $A\Delta B \equiv (A-B)\cup(B-A)$ is a symmetric difference between sets A and B.

Distance between vectors $A = (a_1,\ldots,a_n)$ and $B = (b_1,\ldots,b_n)$ can be defined as l_p metric [6]:

$$d(A,B) = \left(\sum_{i=1}^{n} (a_i - b_i)^p \right)^{1/p}$$

Analogically, the distance between matrices $A = (a_{ij})$ and $B = (b_{ij})$, where $i = 1$ to n, $j = 1$ to m, $p \in N$, is given as:

$$d(A,B) = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} \left| a_{ij} - b_{ij} \right|^{p} \right)^{1/p}$$

Distance between ordered sets (rankings) can be measured by Kendall's, Spearman's or Pearson's correlation coefficients, and by dot product of two vectors in n-dimensional Euclidean space.

3 Similarity between two rankings without ties

In ordinal ranking problem, preferences of an expert can be *strong* or *weak*. In the former case all pairs of objects are compared and form *totally ordered set*. In the latter case some pairs of objects are not compared and this situation is regarded as a „tie", objects form *partially ordered set* (*poset*).

Let N be the set of n objects, for example N = {A, B, C, D}, with binary relation r „is better than" which is antisymmetric, transitive and total:

r = {[B, A], [C,A], [A,D], [B,C], [B,D], [C,D]},

where for example [B,A] means "B is better than A". Then the set O = [B, C, A, D] is the *totally ordered set*, objects A, B, C, D have *ranks* 3, 1, 2, 4 and the vector (3, 1, 2, 4) is the *ranking* (the *ranking vector*). In general, rankings take format $A = (a_1, a_2, ..., a_n)$, where a_i is the rank of the object i.

3.1 Kendall's rank correlation coefficient τ

Let A and B be two rankings on the same domain (on the same set of objects). Kendall's rank correlation coefficient τ is defined as [1]:

$$\tau = \frac{n_c - n_d}{\frac{1}{2} n(n-1)}, \tag{1}$$

where n_c is the number of concordant pairs and n_d is the number of discordant pairs. Concordant (discordant) pair is the ordered pair of objects, which has the same (opposite) order in both rankings. Kendall's τ is normalized to interval $\langle -1,1 \rangle$. In the case of maximum similarity between two rankings $\tau = 1$ (rankings are identical). In the case of maximum dissimilarity $\tau = -1$ (one ranking is reverse of the other). Interpretation of the correlation coefficient's size for positive values according to Chráska (see [5]) is given in Table 1.

Table 1. Interpretation of the correlation coefficient's size for positive values. Source: [5].

Correlation coefficient	**Dependance between variables**
1	absolute
0.9 - 1	very high
0.7 - 0.9	high
0.4 - 0.7	medium
0.2 - 0.4	low
0 - 0.2	very low
0	none

3.2 Spearman's rank correlation coefficient ρ

Another measure of a correlation between two rankings of n objects is Spearman's rank correlation coefficient ρ given as [5]:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \tag{2}$$

where d_i is the difference in rankings for each object i, $i \in \{1,2,...,n\}$. Speaman's ρ is normalised to interval $\langle -1,1 \rangle$ and its interpretation is analogous to Kendall's τ. In contrast to Kendall τ, Spearman's ρ can be used for problems with ties. Also, it can be used for the data which are not normally distributed [8].

3.3 Pearson's correlation coefficient r

Pearson correlation coefficient r between two variables X and Y is defined as [7]:

$$r_{xy} = \frac{E[(X-\mu_X)\cdot(Y-\mu_Y)]}{\sigma_X \cdot \sigma_Y}, \tag{3}$$

where E is the expected value operator and σ_x, σ_y are standard deviations. Again, r takes value from interval $\langle -1,1 \rangle$ and its interpretation is the same as above. Pearson's r can be used for problems with ties too.

Pearson's correlation coefficient is a measure of linear dependance of two variables while Spearman's correlation coefficient measures the extent to which both variables can be described by monotonic (not necessary linear!) function. If both variables increase or decrease (one increases and the other decreases), then correlation coefficient is positive (negative).

3.4 Dot product of vectors

Another approach to similarity stems from geometry. If we imagine rankings as n-dimensional vectors in n-dimensional Euclidean space with initial points in 0 (see Fig. 1), then the angle between vectors $\vec{a}$ and $\vec{b}$ can be computed by dot product of vectors:

$$\cos\varphi = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}|\cdot|\vec{b}|} \tag{4}$$

This measure is standardized, because $\cos\varphi \in \langle -1,1 \rangle$, and it can be shown it is equivalent to Pearson's correlation coefficient for centered data (data with a mean equal to zero) [7]. If vectors are identical, then $\varphi = 0$ and $\cos\varphi = 1$, and if vectors are opposite, then $\varphi = 180°$ and $\cos\varphi = -1$, as desired. The smaller is φ, the closer are both vectors.

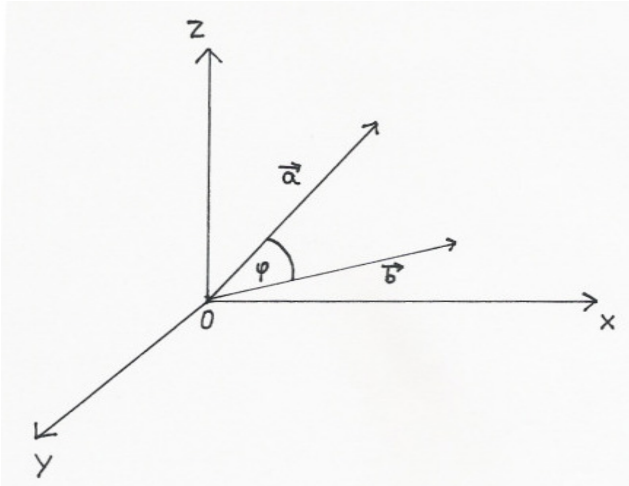

Fig. 1. Two vectors **a** and **b** containing angle φ. Source: own.

3.5 Example

Two (fictional) magazines, Auto Magazine (AM) and Car Revue (CR), have evaluated car reliability of eight car manufacturers and ranked them from the best (1. place) to the worst (8. place), see Table 2. Evaluate Kendall's, Spearman's and Pearson's correlation coefficients between rankings. Use dot product too.

Table 2. Car manufacturers reliability ranking. Source: own.

Rank	Auto Magazine	Car Revue
1	Mazda	Mazda
2	BMW	Honda
3	Honda	BMW
4	Audi	Audi
5	Toyota	Ford
6	VW	VW
7	Ford	Toyota
8	Nissan	Nissan

- *Kendall's τ*

For n objects there are $\binom{n}{2}$ pairs to compare, for $n = 8$ the number of pairs is 28. Discordant pairs are: {(BMW,Honda), (Toyota,Ford), (Toyota,VW), (VW,Ford)}, so $n_c = 24$ and $n_d = 4$. From (1) we get:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} = \frac{24-4}{\frac{1}{2}8 \cdot 7} = 0.714$$

The similarity between both rankings is rather high and this result can be interpreted so that experts from both magazines mostly agree on the topic.

The next issue is whether this agreement is accidental or not. Distribution of τ for larger values of n converges to normal distribution with a mean 0 and variance $\frac{2(2n+5)}{9n(n-1)}$ [1]. A null hypothesis H_0: „Agreement between both rankings is accidental.“ can be tested with standardized Z value [1]:

$$Z = \frac{\tau}{\sigma} = \frac{\tau}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} \tag{5}$$

For n = 8 and significance level α = 0.05 the critical value is 0.571 [5]. From (5) we get Z = 2.47, which exceeds critical value and we reject the null hypothesis. Agreement between both rankings is not accidental (is statistically significant).

- *Spearman's ρ*

We rearrange rankings according to manufacturer (see Table 3) and compute difference d and its square d^2 in rankings of each manufacturer.

Table 3. Rearrangement of Table 2. Source: own.

Manufacturer	AM rank	CR rank	d	d^2
Audi	4	4	0	0
BMW	2	3	-1	1
Ford	7	5	2	4
Honda	3	2	1	1
Mazda	1	1	0	0
Nissan	8	8	0	0
Toyota	5	7	-2	4
VW	6	6	0	0

From Table 3 and (2) we get:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d^2}{n(n^2-1)} = 1 - \frac{6 \cdot 10}{8(64-1)} = 0.881$$

Similarity between both rankings is high.

- *Pearson's r*

From (3) and Table 3 we get: $r_{xy} = \frac{E[(X-\mu_X)\cdot(Y-\mu_Y)]}{\sigma_X \cdot \sigma_Y} = \frac{4.625}{2.291 \cdot 2.291} = 0.881$

Because the data are ordinal, Pearson‘s r is equal to the Spearman‘s ρ.

- *Dot product*

From (4) we get:

$$\cos\varphi = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|} = \frac{199}{204} = 0.975$$

Similarity between both rankings is very high. The angle between both vectors φ = 12°43'. (The value 0.975 differs from Pearson's correlation coefficient because the data are not centered).

4 The similarity of two rankings with ties

Kendall's τ was found to be flawed in cases with weak preferences (ties) by Emond and Mason [3]. They proposed modified correlation coefficient τ_x which is equivalent to Kendall's τ for strong preferences and for weak preferences ties are given score 1 rather than 0 [3]:

$$\tau_x(A,B)=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\cdot b_{ij}}{n(n-1)}$$

In the notation above, preferences A and B are given as matrices of order n with $a_{ij} = 1$ if object i is preferred to object j and $a_{ij} = -1$ vice versa. Diagonal elements are zero.

Pearson's and Spearman's correlation coefficients given by (3) and (4) treat ties well. In the case of ties, equal objects have the same rank which is the mean of their ranks if they would follow one after another.

4.1 Example

Evaluate similarity (correlation) between two rankings in Table 4 by Pearson's r.

Table 4. Rankings with ties. Source: own.

Rank	Auto Magazine	Rank	Car Revue
1.	Mazda	1.-2.	Mazda, Honda
2.-3.	BMW, Honda	3.	BMW
4.	Audi	4.	Audi
5.	Toyota	5.	Ford
6.-7.	VW, Ford	6.	VW
8.	Nissan	7.-8.	Toyota, Nissan

We rearrange rankings according to manufacturer (Table 5). For tied ranks we use halves: 1.5, 2.5, etc.

Table 5. Rearrangement of Table 4. Source: own.

Manufacturer	AM rank	CR rank
Audi	4	4
BMW	2.5	3
Ford	6.5	5
Honda	2.5	1.5
Mazda	1	1.5
Nissan	8	7.5
Toyota	5	7.5
VW	6.5	6

Using Table 5 and (3) we get:

$$r_{xy} = \frac{E[(X-\mu_X)\cdot(Y-\mu_Y)]}{\sigma_X \cdot \sigma_Y} = \frac{4.469}{2.264 \cdot 2.264} = 0.872$$

The similarity between the rankings is high.

5 The similarity of more than two rankings

5.1 Kendall's coefficient of concordance (W)

For assessing agreement among more than two rankings on the same domain without or with ties Kendall's coefficient of concordance (Kendall's *W*) is used. It ranges from 0 – no agreement, to 1 – complete agreement. It is given by [5]:

$$W = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{\frac{1}{12} \cdot k^2 \cdot \left(n^3 - n\right)}, \tag{6}$$

where X_i is the sum of ranks for object *i*, *k* is the number of rankings and *n* is the number of objects.

Statistical significance of Kendall's *W* can be evaluated by χ^2 test with $n - 1$ degrees of freedom [5]:

$$\chi^2 = W \cdot k \cdot (n-1) \tag{7}$$

5.2. Example

Eight car manufacturers were ranked by four (fictional) magazines: Auto Magazine (AM), Car Revue (CR), World of Motors (WoM) and Moto Sport (MS), see Table 6. Evaluate similarity among rankings using Kendall's *W*.

Table 6. Four rankings of eight car manufacturers. Source: own.

Manufacturer	**AM rank**	**CR rank**	**WoM rank**	**MS rank**	**Sum of X_i**	**Sum of X_i^2**
Audi	4	4	3	4	15	225
BMW	2	3	2	1	8	64
Ford	7	5	5	6	23	529
Honda	3	2	4	3	12	144
Mazda	1	1	1	2	5	25
Nissan	8	8	8	7	31	961
Toyota	5	7	6	5	23	529
VW	6	6	7	8	27	729

Sum of X_i = 144, sum of X_i^2 = 3206, $k = 4$, $n = 8$. From (6) we get:

$$W=\frac{\sum_{i=1}^{n}X_i^2-\frac{\left(\sum_{i=1}^{n}X_i\right)^2}{n}}{\frac{1}{12}\cdot k^2\cdot\left(n^3-n\right)}=\frac{3206-\frac{20736}{8}}{\frac{1}{12}\cdot 16\cdot\left(8^3-8\right)}=\frac{614}{672}=0.914$$

Thus, there is very high agreement among rankings.

For statistical significance of Kendall's W relation (7) is used. We can test the null hypothesis H_0: „Computed Kendall's W does not evidence agreement among rankings ($W = 0$).", against alternative hypothesis H_A: „Computed Kendall's W evidence agreement among rankings ($W \neq 0$)." In our case:

$$\chi^2 = W\cdot k\cdot\left(n-1\right)=0{,}914\cdot 4\cdot 7=25.59$$

Critical value for $n - 1 = 7$ degrees of freedom and significance level $\alpha = 0.05$ is $\chi^2_{0,05}(7)$ = 14.07 [5]. Thus we reject the null hypothesis H_0 and accept the alternative hypothesis H_A. There is statistically significant agreement among rankings.

6 The similarity between incomplete rankings (top k lists)

In previous sections all rankings were defined on the same set of objects (domain), but there are situations were ranking domains are not the same. For example Internet search engines give different top 10 lists of items on the first page. To measure distance between two rankings on different domains generalized Kendall's tau was proposed by Fagin et al. [4]. They define generalized Kendall's distance K between top k lists τ_1 and τ_2 as [4]:

$$K^{(p)}(\tau_1,\tau_2)=\sum\overline{K}^{(p)}_{i,j}(\tau_1,\tau_2)\ ,$$

where $\overline{K}_{i,j}$ is a *penalty function* for each pair (i, j) of objects. In general, there are four cases that must be resolved separately [4]:

a) Objects i and j are in both rankings. Then penalty function $\overline{K} = 1$, if they are ranked in the opposite order and $\overline{K} = 0$, if they are ranked in the same order.
b) Objects i and j appear in one ranking and exactly one object (e.g. i) appears in the second ranking. Then penalty function $\overline{K} = 0$ if i is ranked above j in the first ranking and $\overline{K} = 1$ otherwise.
c) Object i (but not j) appears in one ranking and object j (but not i) appears in the second ranking. Then $\overline{K} = 1$.
d) Objects i and j appear in one ranking but nor i or j appears in the second ranking. Then we don't know the order of i and j in the second ranking and penalty function $\overline{K}$ = p, where p is an optional parameter. One can assign $p = 0$ (optimistic approach), p = 0.5 (neutral approach) or $p = 1$ (pessimistic approach) for such a pair (i, j).

Analogically, other measures such as Spearman's footrule metric or Hausdorff metric can be generalized and converted into generalized correlation coefficients (for details see [4]).

7 Conclusion

Similarity among rankings can be measured by appropriate distance or metric functions, and their normalized quantities, such as Pearson's, Kendall's and Spearman's correlation coefficients, or dot product of vectors. These measures evaluate similarity among rankings, thus providing useful information on agreement or disagreement between experts who compiled them. In ordinal consensus ranking problem, where a final consensus ranking from a given set of rankings have to be achieved, it can be assumed that rankings with higher correlation (reflecting better agreement among experts) may naturally lead to more meaningful final consensus ranking and better agreement among consensus ranking methods such as Borda-Kendall's method of marks, CRM, DCM or MAH. However, the relationship between ranking's similarity and methods' agreement have to be established by future research.

References:

[1] ABDI, H. The Kendall Rank Correlation Coefficient. *In: Neil Salkind (Ed.). Encyclopedia of Measurment and Statistics*, 2007.

[2] COOK, W. D. Distance-Based and Ad Hoc Consensus Models in Ordinal Preference Ranking. *European Journal of Operational Research*, Nr. 172 (2006), p. 369-385.

[3] EMOND, E. J., MASON, D. A New Rank Correlation Coefficient with Application to Consensus Ranking Problem. Journal of Multicriteria Decision Analysis, Vol. 11, No 1., 2002, p. 17-28.

[4] FAGIN, R., KUMAR, R., SIVAKUMAR, D. Comparing k-top lists. *SIAM, Journal of Discrete Mathematics*, Vol. 17, No. 1, p. 134-160.

[5] CHRÁSKA, M. *Metody pedagogického výzkumu*. Havlíčkův Brod: Grada, 2007. ISBN 978-80-247-1369-4.

[6] KOSMÁK, L., POTŮČEK, R. *Metrické prostory*. Academia. Praha: 2004. ISBN: 80-200-1202-8.

[7] MELOUN, M., MILITKÝ, J. Kompendium statistického zpracování dat, Metody a řešené úlohy včetně CD, Praha: Academia, 2002. ISBN:

[8] STEMLER, S. E. A Comparison of Consensus, Consistency and Measurement Approaches to Estimating Interrater Reliability. *Practical Assessment, Research and Evaluation*, Vol. 9, Nr. 4, 2004. ISSN: 1531-7714.

[9] TAVANA, M., LoPINTO, F., SMITHER, J. W. A Hybrid Distance-Based Ideal-Seeking Consensus Ranking Model. *Journal of Applied Mathematics and Decision Sciences*, vol. 2007, 18 pages.

Classification JEL: C10, D70

Mgr. Jiří Mazurek, Ph.D.
Odborný asistent
Katedra matematických metod v ekonomiii
Slezská univerzita v Opavě
Obchodně podnikatelská fakulta v Karviné
Univerzitní náměstí 1934/3
733 40 Karviná
e-mail: mazurek@opf.slu.cz