Weighted Similarity: A New Similarity Measure for Document Ranking Features

Mehrnoush Barani Shirzad[1(✉)] and Mohammad Reza Keyvanpour[2]

[1] Data Mining Research Laboratory, Department of Computer Engineering, Alzahra University, Tehran, Iran
mb.shirzad@yahoo.com
[2] Department of Computer Engineering, Alzahra University, Tehran, Iran
keyvanpour@alzahra.ac.ir

Abstract. Many ranking features are utilized by information systems. Several ranking methods act similarly to each other and thus provide similar information. Some information retrieval systems need to select privilege ranking methods and eliminate redundant rankers. To deal with redundant features, the present work introduces a new feature similarity measure, which is based on documents distance. Then the measure is weighted by relevance degree of documents. Experiments are conducted on two data sets MQ2008 and OHSUMED for all features pairs. We adopt two methods of similarity measures in order to compare them with our similarity measure. Results show that our method has correlation with other measures and with MAP.

Keywords: Information retrieval · Similarity measure · Ranking similarity

1 Introduction

Ranking the results is a challenge in every information retrieval system. As an example of ranking, ranker algorithms provide a list of documents according to their relevance. Many algorithms have been developed for document ranking, including vector space algorithms and probabilistic algorithms. The rankers are compared with each other according to their results with the aim of deducing their similarity. Evaluating the ranked list is conventionally performed by evaluation measures including NDCG, MAP, ERR and can be applied for comparison. Similarity measures show how much a pair of rankers are correlated. A similarity measure indicates the agreement of two rankers with a value which is often in range of [0..1] or [−1..1]. This similarity value can apply under feature selection task in systems that apply a variety of rankers to compare the results of two search engines. In this paper the first task is considered.

Various similarity methods have been introduced. Kendal's τ, Pearson correlation coefficient, Spearman rank correlation coefficient are well-known correlation coefficient measures. Kendal's τ considers priority between each two objects and Spearman shows the linear correlation coefficient for two ranked lists. Studies in [1–8] mentioned problems with Kendal's τ and attempted to solve the problems. Yilmaz et al. in [1] view the problem of Kendal's τ as ignoring the importance of top ranks. In [3] the problem with Spearman's footrule and is discussed and the relevance of elements and

R. Silhavy et al. (eds.), *Artificial Intelligence Trends in Intelligent Systems*,
Advances in Intelligent Systems and Computing 573, DOI 10.1007/978-3-319-57261-1_27

positional information are taken into consideration. In [1] a correlation measure was applied to determine the similarity between two search engines. We investigate the issue of feature similarity, which can be applied for feature selection.

In this paper, we propose a similarity method to measure the agreement between two ranked lists based on the difference in the position of documents on each sorted list. Moreover, we consider the difference between relevance degrees of documents in each position. The experiments were conducted on two standard datasets from Letor3 and Letor4 for all features pairs. Our method contributes to correlated performance with respect to other similarity measures.

This paper is organized as follows: in Sect. 2 related work focusing on similarity measures are reviewed. In Sect. 3 we present our method, introducing the weighted similarity measure, and in Sect. 4 empirical results are reported, and we conclude in Sect. 5.

2 Related Work

In this section, common rank correlation coefficient metric and the algorithms introduced to improve them are reviewed.

The aim of a similarity measure is to evaluate the correlation of two features that rank a list of documents. Several proposals have been made for rank correlation. Kendal's τ correlation coefficient and Pearson correlation coefficient are two widely used correlation coefficient measures that are also applied for feature selection for learning to rank.

Kendal's τ [9] indicates the number of paired documents of the same list sorted by two rankers that take equal preferences order in two ranked lists, over total number of documents pairs. We apply a version of Kendal's τ as follows. Kendal's -τ (TAU) for ranking a query q and two features f_i and f_j, is defined as follows:

$$Tau(f_i, f_j) = \frac{\#\{(x_s, x_t) \in X_q | x_s <_{x_i} x_t \text{ and } x_s <_{x_j} x_t\}}{\#\{(x_s, x_t) \in X_q\}} \quad (1)$$

Where the numerator is the number of paired documents related to query q that takes equal preference according to two features fi and fj, and the denominator is the number of whole pairs of documents associated with the query.

To measure the similarity, Pearson (PCC) is defined as follows:

$$PCC(f_i, f_j) = \frac{\text{cov}(f_i, f_j)}{\sqrt{\text{var}(f_i).\text{var}(f_j)}} \quad (2)$$

Where $cov(f_i, f_j) = \sum_{k=1}^{n} \left(f_i^{(k)} - \overline{f_i}\right)\left(f_j^{(k)} - \overline{f_j}\right)$ is the covariance of two features and $var(f_i) = \sum_{k=1}^{n} \left(f_i^{(k)} - \overline{f_i}\right)^2$ is the variance of a feature.

Algorithms in [1–3] investigate the problem with Kendal's tau in ranking. In [1] Yilmaz et al. the weakness of Kendal's tau in mirroring the error and precision in top rank is discussed. They proposed τ_{AP} a variation of Kendal's tau based on the average precision that gives more weight to the errors at high rankings. τ_{AP} has gained a considerable interest [2, 3], Stefani et al. [2] apply a τ_{AP} in Mallows model as the distribution function for the problem of learning probabilistic models for permutations. Urabno et al. in [3] also apply statistical estimators Kendal's and τ_{AP} to estimate the expected correlation of test collection against true ranking.

Carterette et al. in [4] solves the problem using Kendall's tau that works regardless of actual correlation between the measurements, proposing a rank correlation based distance between rankings. In [5] Kumar et al. view the problem with Spearman's footrule and Kendall's tau as overlooking objects relevance and positional information. They extend Spearman's footrule and Kendall's tau to element weights, position weights, and element similarities. They list five principles for similarity metric; richness, simplicity, generalization, basic properties and correlation with other metrics. Richness evolves three concepts: element weights, position of elements and diversity. In [6] Luchen et al. propose a family of similarity measures based on maximization effectiveness difference. Effectiveness measures include MAP, NDCG and ERR applied. Webber et al. in [7] address the problem of indefinite rankings in which two lists lack the same items and only have some common items. Gao et al. in [8] by considering the top of the ranked list, propose a head-weighted measure. Their metric evaluates the gap between system scores, and is effected by the gap at the top of the ranked lists. In the next section, our method is presented based on distance and relevance degree.

3 Our Method

We consider feature (ranker) similarity under feature selection problem.

3.1 Similarity Based on Distances

A similarity measure is applied to evaluate the different between ranker's results. To evaluate the similarity between features (ranking method) according to their ranked list, one way is to consider the distance between two ranked lists. We also consider the relevance (importance) degree of each document on the ranked list. Figure 1 shows the weighted similarity measure pattern.

We apply the definition of distance from Spearman's ρ, and present a new similarity measure. The distance of instances in two ranked lists of the same document, provided by two features (ranker) is defined as the difference in the two positions on the two ranked lists. The following demonstrates an example for the distance of documents in two lists ranked on the same set with two different features.

Feature f_i: $d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9$
Feature f_j: $d_9, d_8, d_5, d_7, d_6, d_1, d_3, d_4, d_2$
Distance of d_3 = position d_3 in list 1 - list 2 ($|3–7|$)

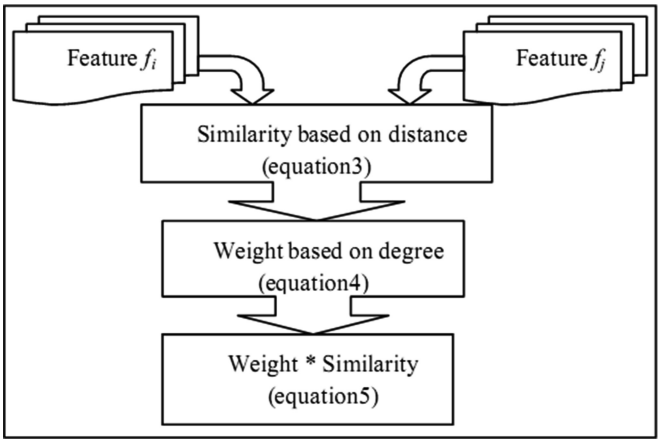


Fig. 1. Weighted similarity measure pattern

To define similarity the following normalization is used:

$$sim(f_i,f_j) = \frac{\max(d) - \sum d}{\max(d)} \tag{3}$$

Where max (d) is the maximum distance which, for n instances, equals $\left\lfloor \frac{n^2}{2} \right\rfloor$. The above definition measures the similarity according to difference in distance, but the relevance of each document in each position is important in ranking.

3.2 Similarity Based on Degree

We consider another distance, which shows the difference of weight in two ranked lists. For two ranked lists the relevance of each document in each position is considered.

For each position the absolute difference of relevance degrees of two ranked lists is computed. This distance is based on the degree of relevance. The following example illustrates the distance of degrees.

docs degree: $d_1 = 0, d_2 = 1, d_3 = 2, d_4 = 0, d_5 = 2, d_6 = 1, d_7 = 3, d_8 = 0$
Feature f_i: $d_3, d_2, d_3, d_4, d_7, d_5, d_6, d_8$
Feature f_j: $d_1, d_2, d_3, d_8, d_5, d_6, d_7, d_4$
Distance of degrees: 2, 0, 0, 0, 1, 2, 0

Sum of this distance shows the weighting difference of two ranked lists. For two identical rankings, this value is zero and the maximum value is (r. $\overline{d}$) in which r is the number of degrees and d is the total object or position. Then d is normalized to the following equation to define similarity:

$$w(f_i,f_j) = \frac{\max(\overline{d}) - \sum \overline{d}}{\max(\overline{d})} + \alpha \tag{4}$$

Where α is a threshold to prevent zero, which equals 1 for identical rank. The similarity measure becomes w.sim.

$$w.sim(f_i,f_j) = \left(\frac{\max(\overline{d}) - \sum \overline{d}}{\max(\overline{d})} + \alpha\right).\frac{\max(d) - \sum d}{\max(d)} \tag{5}$$

4 Experiments

In this session, first the datasets and evaluation measures are introduced and then the algorithm settings are presented.

4.1 Dataset

In order to investigate the proposed method we conduct our experiment on two datasets from Letor benchmark. MQ2008 from Letor 4.0 has in total 784 queries, containing 15211 document-query pairs, for which 3 relevance degrees have been provided and 46 features have been extracted. OHSUMED dataset from Letor 3.0 consists of 106 queries, 45 features extracted for each document query pair. It consists of 16 140 document-query pairs, and 3 relevance degrees are supplied. For the purpose of cross-validation, each dataset is folded into five folds and each fold contains a training set, a validation set and a test set.

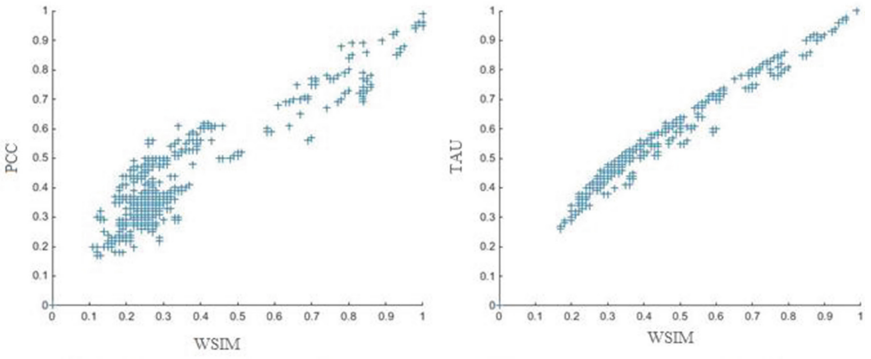
To compare the proposed method with other correlation strategies we applied two similarity methods, Kendal's -τ Eq. (1) and Pearson correlation coefficient Eq. (2).

To assess the proposed measure and show that similar features provide similar results according to the accuracy measure, a comparison between the results of weighted similarity against MAP is necessary. The problem is the correlation between two ranked lists, none of which is the ground truth. Therefore, we compare weighted similarity against their similarity in MAP. For this evaluation we define MAP similarity of two features as MAPSIM $(f_i,f_j) = 1 - |\text{MAP}(f_i) - \text{MAP}(f_j)|$.

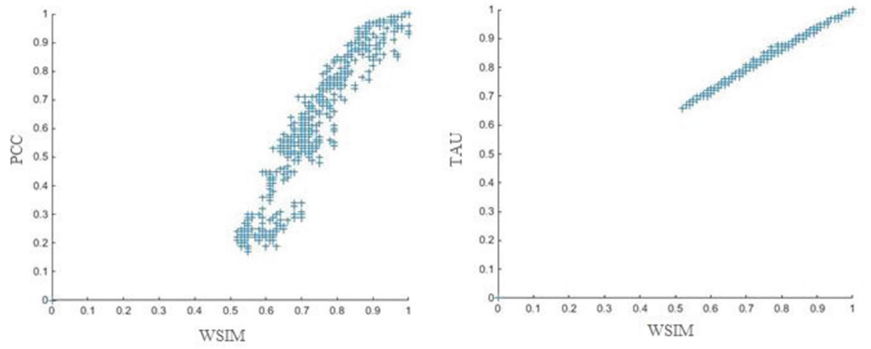
In the following WSIM shows the weighted similarity, PCC is used to refer to Pearson correlation coefficient and TAU shows Kendal's tau. MAP represents MAPSIM.

4.2 Experimental Results

In Fig. 2(A) the plot illustrates the weighted similarity against Pearson correlation coefficient and Kendal's tau across all pairs of features for MQ2008. The plot shows a correlation between weighted similarity against Kendal's tau. Almost all data follow a similar pattern. The weighted similarity is partly correlated with Pearson correlation coefficient.



A) Weighted Similarity vs. Pearson correlation coefficient and Weighted Similarity vs. Kendal's tau on MQ20008 Data set



B) Weighted Similarity vs. Pearson correlation coefficient and Weighted Similarity vs. Kendal's tau on OHSUMED Data set

Fig. 2. Weighted similarity vs. Pearson correlation coefficient and weighted similarity vs. Kendal's tau

In Fig. 2(B) the plot shows the results for OHSUMED containing weighted similarity against Kendal's tau and Pearson correlation coefficient across all pairs of features. The plot shows again that for this data set the correlation between weighted similarity and Kendal's tau is significant. The weighted similarity and Pearson correlation coefficient are partly correlated.

Figure 3 plots the similarity between MAP of pairs of features against weighted similarity of corresponding feature pairs. According to plot A, there is a correlation between MAP and weighted similarity for OHSUMED data set. However, MAP similarity produced a higher value compared with weighted similarity. The linear relation between two measures for all pairs of features is visible, which shows similar features based on the proposed measure providing a similar accuracy according to MAP. A correlation between MAP and weighted similarity for MQ2008 data set is demonstrated in plot B. Though this correlation is linear and less gradual than the plot for OHSUMED, the data set shows that there is a clear relation between the two measures.

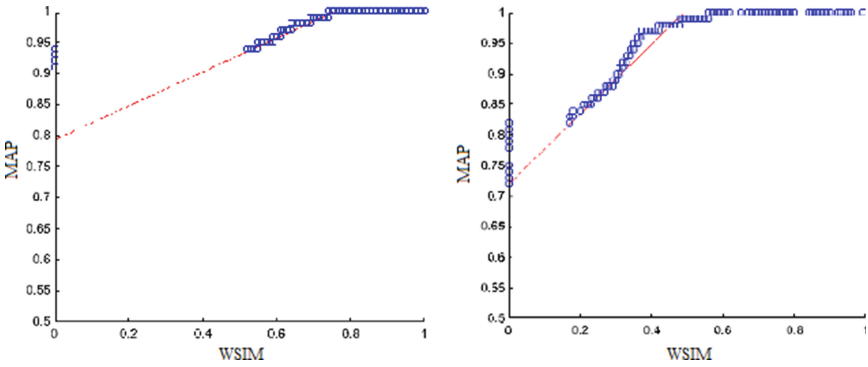


Fig. 3. (A) Weighted similarity vs. MAP similarity on OHSUMED Data set (B) Weighted Similarity vs. MAP Similarity on MQ2008 Data set

5 Conclusion

We introduced a new similarity measure that evaluates the rank correlation between two ranking features. We applied two methods in order to compare the proposed method as similarity measures. The empirical results showed that our method is correlated with other methods and with MAP. Also the proposed measure has other properties including simplicity and is weighted based on document relevance. We conducted our experiments on document retrieval; future work will include applying weighted similarity for other information retrieval applications.

References

1. Yilmaz, E., Aslam, J.A., Robertson, S.: A new rank correlation coefficient for information retrieval. In: Proceedings of the 31st International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 587–594. ACM (2008)
2. Stefani, L.D., Epasto, A., Upfal, E., Vandin, F.: Reconstructing hidden permutations using the average-precision (AP) correlation statistic. In: Proceedings of the 13th AAAI Conference on Artificial Intelligence, pp. 1526–1532. AAAI Press (2016)
3. Urbano, J., Marrero, M.: toward estimating the rank correlation between the test collection results and the true system performance. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1033–1036. ACM (2016)
4. Carterette, B.: On rank correlation and the distance between rankings. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 436–443. ACM (2009)
5. Kumar, R., Vassilvitskii, S.: Generalized distances between rankings, In: Proceedings of the 19th International Conference on World Wide Web, pp. 571–580 (2010)

6. Tan, L., Clarke, C.L.A.: A family of Rank similarity measures based on maximized effectiveness difference. IEEE Trans. Know. Data Eng. **27**, 2865–2877 (2014)
7. Webber, W., Moffat, A., Zobel, J.: A similarity measure for indefinite rankings. ACM Trans. Inf. Syst. **28**(4), 20 (2010)
8. Gao, N., Bagdouri, M., Oard, D.W.: Pearson rank: a head-weighted gap-sensitive score-based correlation coefficient. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 941–944. ACM (2016)
9. Kendall, M.: Rank Correlation Methods. Oxford University Press, Oxford (1962)