Ranking learning-to-rank methods

Ranking Learning-to-Rank Methods*

Djoerd Hiemstra
University of Twente
Enschede, The Netherlands
hiemstra@cs.utwente.nl

Niek Tax
Eindhoven University of Technology
Eindhoven, The Netherlands
n.tax@tue.nl

Sander Bockting
KPMG Netherlands
Amstelveen, The Netherlands
bockting.sander@kpmg.nl

ABSTRACT

We present a cross-benchmark comparison of learning-to-rank methods using two evaluation measures: the Normalized Winning Number and the Ideal Winning Number. Evaluation results of 87 learning-to-rank methods on 20 datasets show that ListNet, SmoothRank, FenchelRank, FSMRank, LRUF and LARF are Pareto optimal learning-to-rank methods, listed in increasing order of Normalized Winning Number and decreasing order of Ideal Winning Number.

CCS CONCEPTS

• **Information systems → Learning to rank**;

1 INTRODUCTION

Like most information retrieval methods, learning-to-rank methods are evaluated on benchmark datasets, such as the many datasets provided by Microsoft and the datasets provided by Yahoo and Yandex. These learning-to-rank datasets offer feature set representations of the to-be-ranked documents instead of the documents themselves. Therefore, any difference in ranking performance is due to the ranking algorithm and not the features used. This opens up a unique opportunity for cross-benchmark comparison of learning-to-rank methods. In this paper, we compare learning to rank methods based on a sparse set of evaluation results on many benchmark datasets.

2 DATASETS AND METHODS

Evaluation results of 87 learning-to-rank methods on 20 well-known benchmark datasets are collected using a systematic literature review [1]. We included papers that report the mean average precision or nDCG at 3, 5 or 10 documents retrieved. Papers that used different or additional features, or that reported no baseline performance that allowed us to check validity of the results, were excluded from the analysis.

The Winning Number of a learning-to-rank method is defined as the number of other methods that a method beats over the set of datasets. So, a method with a high Winning Number beats many other methods on many datasets. For every method, we find a different set of datasets on which the method was evaluated. The Ideal Winning Number is the maximum Winning Number that the method can achieve on all datasets on which it was evaluated. The Normalized Winning Number is the Winning Number divided by the Ideal Winning Number. The Normalized Winning Number gives insight in the ranking accuracy of the learning to rank method. The Ideal Winning Number gives insight in the degree of certainty concerning the ranking accuracy. We report the best performing methods by Normalized Winning Number and Ideal Winner Number.

3 RESULTS

Figure 1 shows the Normalized Winning Number as function of the Ideal Winning Number for 87 learning-to-rank methods over 20 datasets and all investigated evaluation measures: Mean Average Precision and nDCG at 3, 5, 10. The figure labels the Pareto optimal algorithms and also the Rank-2 Pareto optima in a smaller font, which are the labels of the algorithms with exactly one algorithm having a higher value on both axes. In addition, Linear Regression and the ranking method of simply sorting on the best single feature are labeled as baselines.

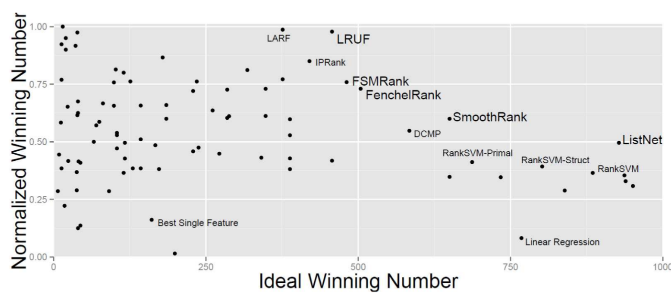

Figure 1: Winning numbers of 87 learning to rank methods.

The figure shows that LRUF beats almost all other methods with an Ideal Winning Number of almost 500 measures and datasets. If we move to the right of the figure, we increase our confidence in the results. That is, we are more confident about the results of ListNet as its Ideal Winning Number is close to 1000 measures and datasets. However, ListNet is outperformed on half, so about 500, of the datasets and measures.

4 CONCLUSION

Based on a cross-benchmark comparison of 87 learning-to-rank methods on 20 datasets, we conclude that ListNet, SmoothRank, FenchelRank, FSMRank, LRUF and LARF are Pareto optimal learning-to-rank methods, listed in increasing order of Normalized Winning Number and decreasing order of Ideal Winning Number [1].

*The full version of this work was published by Tax, Bockting and Hiemstra [1].

LEARNER'17, October 1, 2017, Amsterdam, The Netherlands

REFERENCES

[1] Niek Tax, Sander Bockting, and Djoerd Hiemstra. 2015. A cross-benchmark comparison of 87 learning to rank methods. *Information processing & management* 51, 6 (2015), 757–772. *(Awarded IPM Best Paper of 2015)*