When Truncated Rankings Are Better and How to Measure That – Abstract*

Enrique Amigó[1], Stefano Mizzaro[2] and Damiano Spina[3]

[1]UNED NLP & IR Group, Madrid, Spain
[2]University of Udine, Italy
[3]RMIT University, Melbourne, Australia

Abstract

In this work we provide both theoretical and experimental contributions for the truncated ranking evaluation, where systems have a stopping criteria to *truncate* the ranking at the right position to avoid retrieving those irrelevant documents at the end. We first define formal properties to analyze how effectiveness metrics behave when evaluating truncated rankings. Our theoretical analysis shows that de-facto standard metrics do not satisfy desirable properties to evaluate truncated rankings: only Observational Information Effectiveness (OIE) – a metric based on Shannon's information theory – satisfies them all. We then perform experiments to compare several metrics on nine TREC data sets. According to our experimental results, the most appropriate metrics for truncated rankings are OIE and a novel extension of Rank-Biased Precision that adds a user effort factor penalizing the retrieval of irrelevant documents.

Keywords

Information Retrieval, Evaluation, Evaluation measures, Ranking Cutoff

CIRCLE (Joint Conference of the Information Retrieval Communities in Europe) 2022 is the second joint conference of the information retrieval communities, July 4-7, 2022, Toulouse, France

*This work has been published at the SIGIR 2022 main conference.

enrique@lsi.uned.es (E. Amigó); mizzaro@uniud.it (S. Mizzaro); mizzaro@uniud.it (D. Spina)
https://sites.google.com/view/enriqueamigo/home (E. Amigó); https://www.dimi.uniud.it/mizzaro (S. Mizzaro); https://www.dimi.uniud.it/mizzaro (D. Spina)
0000-0003-1482-824X (E. Amigó); 0000-0002-2852-168X (S. Mizzaro); 0000-0001-9913-433X (D. Spina)