

Using Score Distributions to Compare Statistical Significance Tests for Information Retrieval Evaluation

Título Using Score Distributions to Compare Statistical Significance Tests for Information Retrieval Evaluation

Autores Javier Parapar, [David E. Losada](#), Manuel A. Presedo Quindimil, Álvaro Barreiro

Tipo Artículo de revista

Fonte  [Journal of the Association for Information Science and Technology](#) Wiley, 2019.

ISSN 2330-1635

DOI [10.1002/asi.24203](https://doi.org/10.1002/asi.24203)

Abstract Statistical significance tests can provide evidence that the observed difference in performance between two methods is not due to chance. In Information Retrieval, some studies have examined the validity and suitability of such tests for comparing search systems. We argue here that current methods for assessing the reliability of statistical tests suffer from some methodological weaknesses, and we propose a novel way to study significance tests for retrieval evaluation. Using Score Distributions, we model the output of multiple search systems, produce simulated search results from such models, and compare them using various significance tests. A key strength of this approach is that we assess statistical tests under perfect knowledge about the truth or falseness of the null hypothesis. This new method for studying the power of significance tests in Information Retrieval evaluation is formal and innovative. Following this type of analysis, we found that both the sign test and Wilcoxon signed test have more power than the permutation test and the t-test. The sign test and Wilcoxon signed test also have a good behavior in terms of type I errors. The bootstrap test shows few type I errors, but it has less power than the other methods tested.

Palabras clave Information Retrieval, Evaluation, Statistical Significance Tests

LIGAZÓNS

 [Versión da editorial](#)

DESCARGAS

 [Referencia BibTex](#)

PROXECTOS DE INVESTIGACIÓN

PPIAS: MODELOS DE LINGUAXE PROBABILISTICOS PARA RANKINGS PERSONALIZADOS EN SISTEMAS DE ACCESO Á INFORMACIÓN

PROGRAMAS CIENTÍFICOS

Aprendizaxe automática
Tecnoloxías da Linguaxe Natural