

# Testing Fairness Using the Log-likelihood Ratio

Massimo Melucci  
University of Padua, Italy

## 1 Indexing

The test collection was indexed by Elasticsearch [1]. The default indexing configuration was utilized. The JSON files were opened and each record was processed for all files by using `json.loads`. If a valid document identifier was found in the record, the document was added to the index by using `elasticsearch.Elasticsearch.index` for the document body.

## 2 Training

The training dataset was processed to produce an array of features for each record. The following features were produced for abstract, entities, title, and venue: average idf, average tfidf, average saturation term, average BM25, where saturation term is the function of term frequency that multiplies the IDF of the BM25 weight. To reduce the high variability, the logarithm function was applied to average idf, average tfidf, and average BM25. Moreover, the number of authors per author group was produced. Finally, the relevance assessment, the number of in-citations, the number of out-citations, and the number of query terms was added to each output record.

The aforementioned feature file was processed to compute the proportion of authors for each relevance assessment and for each author group, that is:

$$s_j = \frac{f_{0,j}}{\sum_{i=1}^7 f_{0,i}} \quad r_j = \frac{f_{1,j}}{\sum_{i=1}^7 f_{1,i}} \quad t_j = \frac{f_{0,j} + f_{1,j}}{\sum_{i=1}^7 f_{0,i} + f_{1,i}} \quad j = 1, \dots, 7 \quad (1)$$

where  $j$  is the author group and  $f_{\ell,i}$  is the number of documents authored by authors of group  $i$  and assigned the relevance assessment  $\ell = 0, 1$ .

## 3 Ranking

The retrieved evaluation documents were reranked as follows. The probabilities (1) were smoothed to avoid computational issues and the following estimates

were utilized:

$$\hat{s}_j = s_j + \frac{1}{2} \quad \hat{r}_j = r_j + \frac{1}{2} \quad j = 1, \dots, 7 \quad (2)$$

The features for the sample evaluation file were computed as described in Section 2 and a similar evaluation feature file was produced. For each record of the evaluation feature file, the number of authors for each author group was extracted. For each author group, the following log-likelihood were computed:

$$L_{0,j} = f_{0,j} \log \hat{s}_j \quad L_{1,j} = f_{1,j} \log \hat{r}_j . \quad (3)$$

was computed for each document. Then, the following

$$L_0 = \sum_{i=1}^7 L_{0,i} \quad L_1 = \sum_{i=1}^7 L_{1,i} \quad (4)$$

The fairness score was computed as follows:

$$F = L_1 - L_0 \quad (5)$$

and then it was summed to the BM25 scores computed for abstract, title, and entities.

## References

- [1] C. Gormley and Z. Tong. *Elasticsearch: The Definitive Guide*. O'Reilly Media, Inc., 2015.