Erratum: On Theoretically Valid Score Distributions in Information Retrieval

Ronan Cummins and Colm O'Riordan

Dept. of Information Technology,
National University of Ireland, Galway, Ireland
`ronan.cummins@nuigalway.ie`

Abstract. We correct the results to the experiments presented in our paper. We discovered an error in our code that only significantly affects one of the valid score distribution (SD) models (i.e. that of the two-lognormal SD model). The corrected results are included here. As a consequence we slightly modify the conclusions of the paper. For the three models studied in this paper, we conclude that adhering to the RFCH is practically useful for all three models. In general, there is no loss in performance as measured by the task of inferring average precision for the three models when adhering to the RFCH (a slight increase for the two-normal model is noted). Adhering to the RFCH reduces the parameters in these models and is therefore, a useful constraint. The best SD model for the task and data outlined in this paper is the two-lognormal model.
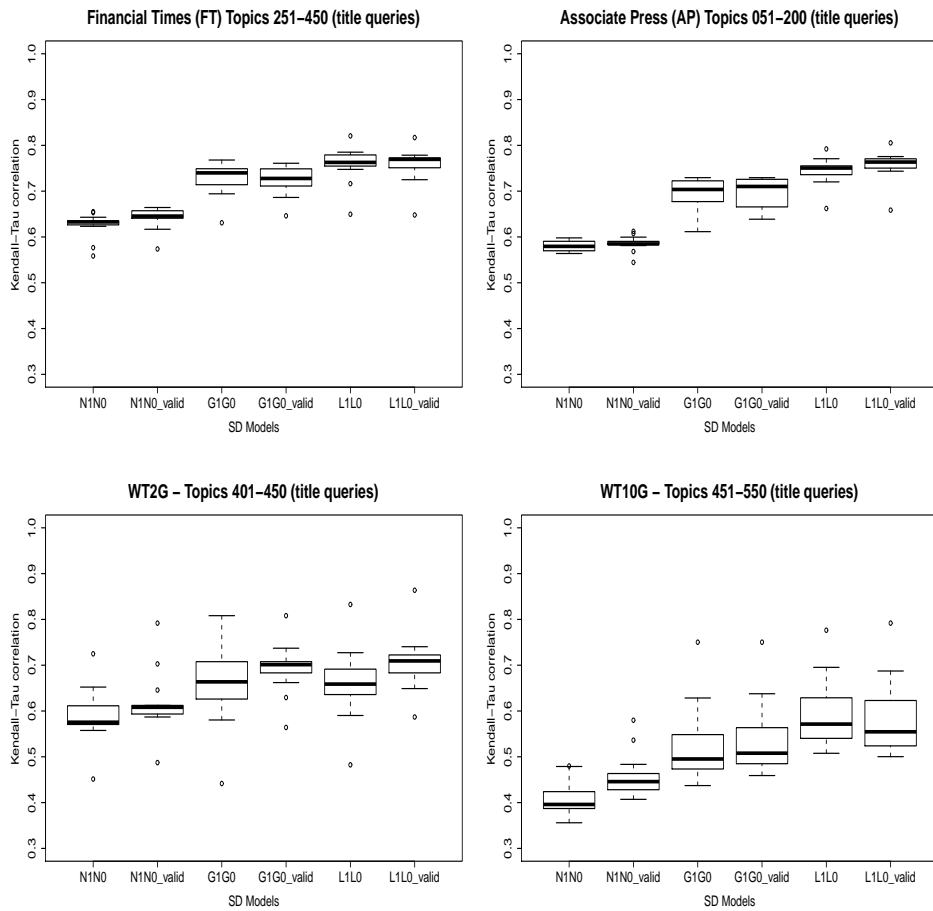
1 Corrected Results For Fig. 1

Fig. 1. Kendall's τ Correlations for mixtures that violate the RFCH and those that adhere to the RFCH for title queries on two Newswire and two Web collections