

Errata to the paper “Search Results Clustering without External Resources”

Chris Staff, Joel Azzopardi, Colin Layfield, and Dan Mercieca

TABLE I. OUR RESULTS WITH NO-K-MEANS - NOLSA, WITHOUT AND WITH QUERY TERMS (QT). SCORES THAT BEAT THE BEST PERFORMERS TO DATE ARE IN BOLD.

QT	<i>SimThres</i>	F1	RI	ARI	J1	Ave. # clusters	Ave. clus. size
	GDI_Varied	71.78	68.30	26.19	35.13	8.00	9.49
	GDI_Fixed	70.19	67.77	26.40	36.26	7.14	9.84
	0.01	64.72	62.06	19.34	36.82	5.01	14.72
✓	GDI_Varied	56.69	43.43	3.55	38.49	2.47	27.74
✓	GDI_Fixed	59.77	51.08	8.54	36.33	3.47	22.19
✓	0.01	54.56	40.19	0.13	39.91	1.10	60.80

TABLE II. OUR RESULTS WITH NO-K-MEANS - WITHLSA, WITHOUT AND WITH QUERY TERMS (QT).

QT	<i>Mode</i>	F1	RI	ARI	J1	Ave. # clusters	Ave. clus. size
	GDI_Varied	64.33	60.63	19.71	38.57	3.90	21.07
	GDI_Fixed	67.92	62.74	18.02	27.44	5.69	11.41
✓	GDI_Varied	54.91	40.68	0.50	39.73	1.27	55.57
✓	GDI_Fixed	64.81	54.61	8.81	31.13	4.76	14.12

While extending the No-K-Means algorithm given in [1] the authors obtained results that were inconsistent with the results reported in [1]. We discovered that when we had included Generalised Dunn’s Index (GDI) to automatically select a cluster configuration, we had unwittingly introduced a bug that resulted in GDI miscalculating the cluster scores. This means that many of the No-K-Means results reported in [1] are erroneous. The correct results are provided in Tables I and II in this brief note. None of the claims made in [1] are effected, although one observation must be changed and another must be withdrawn. *simThres* has a value of 0.04, and not 0.05 as reported in [1], when GDI_m is greatest for NoLSA with query terms omitted. Given these results, the observation that “GDI_Fixed outperforms GDI_Varied on all measures” [1] is not valid.

REFERENCES

- [1] C. Staff, J. Azzopardi, C. Layfield, and D. Mercieca, “Search results clustering without external resources,” in *Proceedings of the 12th International Workshop on Text-based Information Retrieval*, Valencia, Spain, September 2015.