Fusion: Rank-Based

COMP3009J: Information Retrieval

Dr. David Lillis (david.lillis@ucd.ie)

UCD School of Computer Science
Beijing Dublin International College

Rank-Based Fusion: Interleaving

It dose not use chorus effect

- **Interleaving** is perhaps the simplest fusion algorithm of all*
- Here, we take one document from the top of each input set in a "**round robin**" fashion and add it to the fused result set.
  - The document chosen is the highest-ranked document that is not yet included in the fused result set.
- The effectiveness of this technique is, however, poor.
- There is an assumption that every result set is of equal quality, which can have the result that the better result sets are diluted by being merged with non-relevant documents from poorer systems.

* Voorhees, E. M., Gupta, N. K., & Johnson-Laird, B. (1994). The Collection Fusion Problem. In *Proceedings of the Third Text REtrieval Conference (TREC-3)* (pp. 95–104)

Rank-Based: Interleaving (Example)

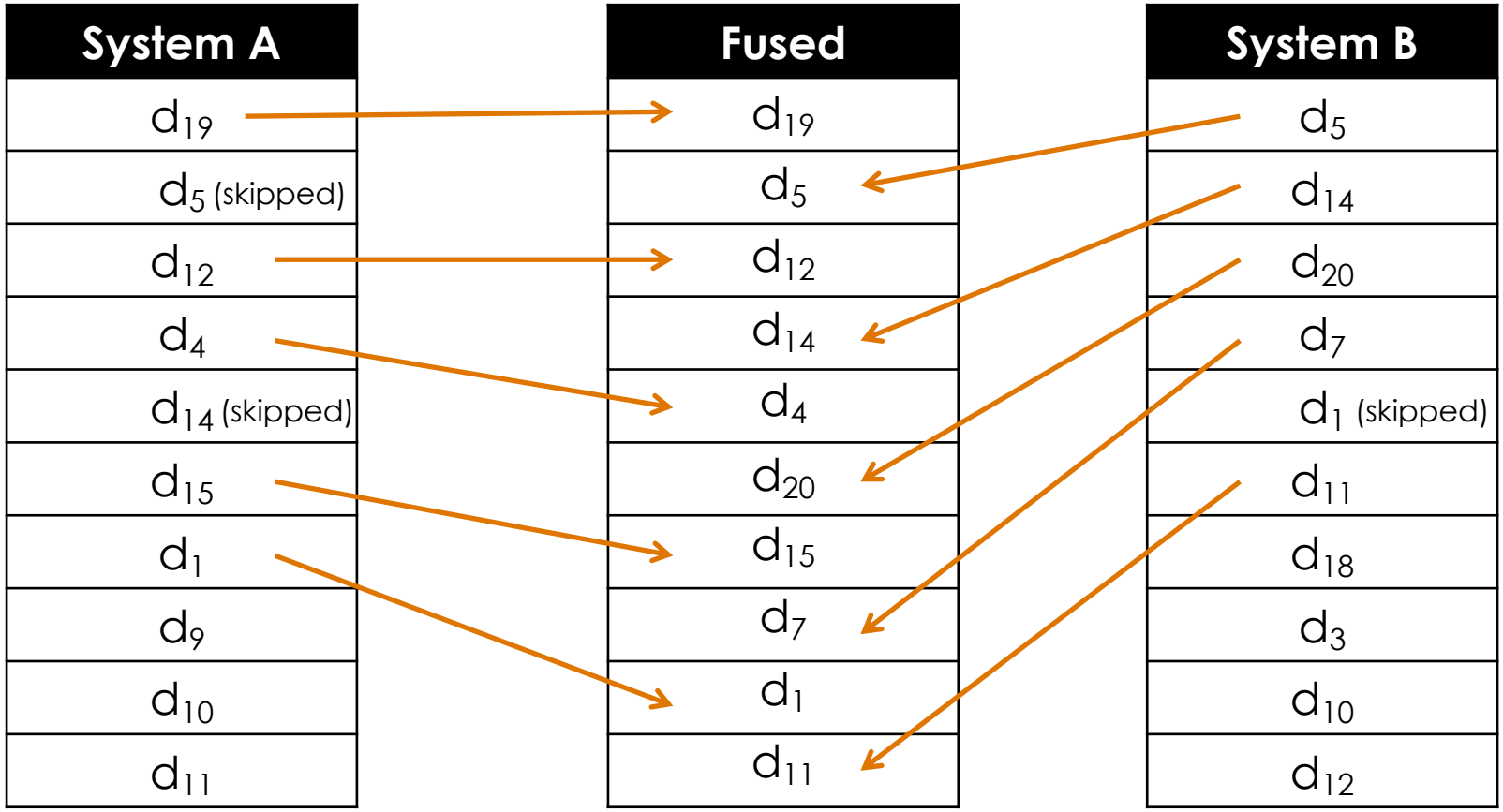
System A
d_{19}
d_5
d_{12}
d_4
d_{14}
d_{15}
d_1
d_9
d_{10}
d_{11}

Fused

System B
d_5
d_{14}
d_{20}
d_7
d_1
d_{11}
d_{18}
d_3
d_{10}
d_{12}

Rank-Based: Interleaving (Example)

if the result of system A poor, the overall result will be worse



Rank-Based: Borda-Fuse

- **Borda-Fuse** is based on an election system for when a few voters(input systems) vote for many candidates (documents).
  - Each voter ranks a set of c candidates in order of preference.
  - For each voter, the top-ranked candidate is given c points, the second ranked is given *c-1* points, etc.
  - If a candidate is not ranked by a voter, the voter's remaining points are divided evenly among unranked candidates.

* Aslam, J. A., & Montague, M. (2001). Models for metasearch. In *SIGIR '01: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 276–284). New York, NY, USA.

Rank-Based: Borda Fuse

System A	Points
d_{19}	14
d_5	13
d_{12}	12
d_4	11
d_{14}	10
d_{15}	9
d_1	8
d_9	7
d_{10}	6
d_{11}	5

System B	Points
d_5	14
d_{14}	13
d_{20}	12
d_7	11
d_1	10
d_{11}	9
d_{18}	8
d_3	7

- 14 unique documents: c = 14
- System A gives 2.5 points to any documents it didn't choose: (4+3+2+1)/4.
- System B gives 3.5 points to documents it didn't choose: (6+5+4+3+2+1)/6

Rank-Based: Borda Fuse

System A	Points
d_{19}	14
d_5	13
d_{12}	12
d_4	11
d_{14}	10
d_{15}	9
d_1	8
d_9	7
d_{10}	6
d_{11}	5

System B	Points
d_5	14
d_{14}	13
d_{20}	12
d_7	11
d_1	10
d_{11}	9
d_{18}	8
d_3	7

Fused	Points (A)	Points (B)	Points (Borda)
d_5	**13**	**14**	27
d_{14}	**10**	**13**	23
d_1	**8**	**10**	18
d_{19}	**14**	3.5	17.5
d_{12}	**12**	3.5	15.5
d_4	**11**	3.5	14.5
d_{20}	2.5	**12**	14.5
d_{11}	**5**	**9**	14
d_7	2.5	**11**	13.5
d_{15}	**9**	3.5	12.5
d_{18}	2.5	**8**	10.5
d_9	**7**	3.5	10.5
d_{10}	**6**	3.5	9.5
d_3	2.5	**7**	9.5

Reciprocal Rank Fusion

- Reciprocal Rank Fusion is a simple rank-based method that has been shown to be effective in practice*.
- Given a set of documents *D* to be ranked, and a set of results *R*, the score for each document is calculated as follows:
  - $RRFscore(d \in D) = \sum_{r \in R} \frac{1}{k + r(d)}$
  - where *r(d)* is the rank of document *d* in result set *r*, and *k*=60 (set by experiment)

* Cormack, G. V., Clarke, C. L. A., Büttcher, S. (2009). Reciprocal Rank Fusion outperforms Condorcet and Individual Rank Learning Models. In *Proceedings of the 32nd international ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 758-759)

Other Rank-Based Techniques

- A variation on interleaving is to use **historical data** to estimate which input system(s) tends to perform better.
- A weighted version of interleaving is then used so that **more documents are taken from the better systems**. *
- Another election-based approach from the same authors as Borda-Fuse is the *Condorcet-Fuse*** algorithm. A weighted version of Borda-Fuse is also proposed, where the points from each input system are multipled by some weight.

* Voorhees et al. (1994)

** Montague, M., & Aslam, J. A. (2002). Condorcet fusion for improved retrieval. In *CIKM ' 02: Proceedings of the eleventh international conference on Information and knowledge management* (pp. 538–548). New York, NY, USA.