

Contents

- 1 Introduction** 1
 - 1.1 The Rationale Behind Data Fusion 1
 - 1.2 Several Typical Data Fusion Methods 2
 - 1.3 Several Issues in Data Fusion 4
- 2 Evaluation of Retrieval Results** 7
 - 2.1 Binary Relevance Judgment 7
 - 2.2 Incomplete Relevance Judgment 10
 - 2.3 Graded Relevance Judgment 11
 - 2.4 Score-Based Metrics 14
 - 2.5 Uncertainty of Data Fusion Methods on Effectiveness Improvement 18
- 3 Score Normalization** 19
 - 3.1 Linear Score Normalization Methods 20
 - 3.1.1 The Zero-One Linear Method 20
 - 3.1.2 The Fitting Method 21
 - 3.1.3 Normalizing Scores over a Group of Queries 22
 - 3.1.4 Z-Scores 22
 - 3.1.5 Sum-to-One 23
 - 3.1.6 Comparison of Four Methods 24
 - 3.2 Nonlinear Score Normalization Methods 27
 - 3.2.1 The Normal-Exponential Mixture Model 27
 - 3.2.2 The CDF-Based Method 27
 - 3.2.3 Bayesian Inference-Based Method 28
 - 3.3 Transforming Ranking Information into Scores 29
 - 3.3.1 Observing from the Data 29
 - 3.3.2 The Borda Count Model 30
 - 3.3.3 The Reciprocal Function of Rank 31
 - 3.3.4 The Logistic Model 31

3.3.5	The Cubic Model	32
3.3.6	The Informetric Distribution	33
3.3.7	Empirical Investigation	35
3.4	Mixed Normalization Methods	40
3.5	Related Issues	41
3.5.1	Weights Assignment and Normalization	41
3.5.2	Other Relevance Judgments	41
4	Observations and Analyses	43
4.1	Prior Observations and Points of View	43
4.2	Performance Prediction of CombSum and CombMNZ	46
4.2.1	Data Fusion Performance	48
4.2.2	Improvement over Average Performance	51
4.2.3	Performance Improvement over Best Performance	52
4.2.4	Performance Prediction	54
4.2.5	The Predictive Ability of Variables	55
4.2.6	Some Further Observations	57
4.2.7	Summary	61
4.3	Comparison of CombSum and CombMNZ	62
4.3.1	Applying Statistical Principles to Data Fusion	63
4.3.2	Further Discussion about CombSum and CombMNZ	64
4.3.3	Lee's Experiment Revisited	66
4.4	Performance Prediction of the Linear Combination Method	70
5	The Linear Combination Method	73
5.1	Performance Related Weighting	73
5.1.1	Empirical Investigation of Appropriate Weights	75
5.1.2	Other Features	81
5.2	Consideration of Uneven Similarity among Results	87
5.2.1	Correlation Methods 1 and 2	87
5.2.2	Correlation Methods 3 and 4	88
5.2.3	Correlation Methods 5 and 6	89
5.2.4	Empirical Study	91
5.2.5	Some More Observations	94
5.3	Combined Weights	97
5.3.1	Theory of Stratified Sampling and Its Application	97
5.3.2	Experiments	99
5.4	Deciding Weights by Multiple Linear Regression	102
5.4.1	Empirical Investigation Setting	103
5.4.2	Experiment 1	104
5.4.3	Experiment 2	109
5.4.4	Experiment 3	111
5.5	Optimization Methods	114
5.6	Summary	116

- 6 A Geometric Framework for Data Fusion** 117
 - 6.1 Relevance Score-Based Data Fusion Methods and Ranking-Based Metrics 117
 - 6.2 The Geometric Framework 119
 - 6.3 The Centroid-Based Data Fusion Method 120
 - 6.4 The Linear Combination Method 126
 - 6.5 Relation between the Euclidean Distance and Ranking-Based Measures 129
 - 6.6 Conclusive Remarks 132
- 7 Ranking-Based Fusion** 135
 - 7.1 Borda Count and Condorcet Voting 135
 - 7.2 Weights Training for Weighted Condorcet Voting 137
 - 7.3 Experimental Settings and Methodologies 139
 - 7.4 Experimental Results of Fusion Performance 140
 - 7.5 Positive Evidence in Support of the Hypothesis 144
- 8 Fusing Results from Overlapping Databases** 149
 - 8.1 Introduction to Federated Search 149
 - 8.2 Resource Selection 152
 - 8.2.1 A Basic Resource Selection Model 152
 - 8.2.2 Solution to the Basic Resource Selection Model 156
 - 8.2.3 Some Variants of the Basic Model and Related Solutions . . . 159
 - 8.2.4 Experiments and Experimental Results 163
 - 8.2.5 Further Discussion 165
 - 8.3 Results Merging 166
 - 8.3.1 Several Result Merging Methods 167
 - 8.3.2 Evaluation of Result Merging Methods 169
 - 8.3.3 Overlap between Different Results 177
 - 8.3.4 Conclusive Remarks 178
- 9 Application of the Data Fusion Technique** 181
 - 9.1 Ranking Information Retrieval Systems with Partial Relevance Judgment 181
 - 9.1.1 How Is the Problem Raised? 182
 - 9.1.2 Other Options Than Averaging All the Values for Ranking Retrieval Systems 185
 - 9.1.3 Evaluation of the Four Ranking Methods 186
 - 9.1.4 Remarks 189
 - 9.2 Ranking Information Retrieval Systems without Relevance Judgment 189
 - 9.2.1 Methodology 191
 - 9.2.2 Experimental Settings and Results 192

- 9.3 Applying the Data Fusion Technique to Blog Opinion Retrieval 193
 - 9.3.1 Data Set and Experimental Settings 194
 - 9.3.2 Experimental Results 195
 - 9.3.3 Discussion and Further Analysis 202
- 9.4 TREC Systems via Combining Features and/or Components 203
 - 9.4.1 A Hierarchical Relevance Retrieval Model for Entity Ranking 203
 - 9.4.2 Combination of Evidence for Effective Web Search 205
 - 9.4.3 Combining Candidate and Document Models for Expert Search 206
 - 9.4.4 Combining Lexicon-Based Methods to Detect Opinionated Blogs 207
 - 9.4.5 Combining Resources to Find Answers to Biomedical Questions 210
 - 9.4.6 Access to Legal Documents: Exact Match, Best Match, and Combinations 210
 - 9.4.7 Summary 212
- References 213
- A Systems Used in Section 5.1 221
- B Systems Used in Section 5.2.4 225
- C Systems Used in Sections 7.3-7.5 227