DATA MANAGEMENT FOR MULTIMEDIA RETRIEVAL

K. Selçuk Candan
Arizona State University

Maria Luisa Sapino
University of Torino
Contents

Preface

1 Introduction: Multimedia Applications and Data Management Requirements
   1.1 Heterogeneity 1
   1.2 Imprecision and Subjectivity 8
   1.3 Components of a Multimedia Database Management System 12
   1.4 Summary 19

2 Models for Multimedia Data
   2.1 Overview of Traditional Data Models 21
   2.2 Multimedia Data Modeling 32
   2.3 Models of Media Features 34
   2.4 Multimedia Query Languages 92
   2.5 Summary 98

3 Common Representations of Multimedia Features
   3.1 Vector Space Models 99
   3.2 Strings and Sequences 109
   3.3 Graphs and Trees 111
   3.4 Fuzzy Models 115
   3.5 Probabilistic Models 123
   3.6 Summary 142

4 Feature Quality and Independence: Why and How?
   4.1 Dimensionality Curse 144
   4.2 Feature Selection 145
   4.3 Mapping from Distances to a Multidimensional Space 167
   4.4 Embedding Data from One Space into Another 172
   4.5 Summary 180
5 Indexing, Search, and Retrieval of Sequences 181
5.1 Inverted Files 181
5.2 Signature Files 184
5.3 Signature- and Inverted-File Hybrids 190
5.4 Sequence Matching 191
5.5 Approximate Sequence Matching 195
5.6 Wildcard Symbols and Regular Expressions 202
5.7 Multiple Sequence Matching and Filtering 204
5.8 Summary 206

6 Indexing, Search, and Retrieval of Graphs and Trees 208
6.1 Graph Matching 208
6.2 Tree Matching 212
6.3 Link/Structure Analysis 222
6.4 Summary 233

7 Indexing, Search, and Retrieval of Vectors 235
7.1 Space-Filling Curves 238
7.2 Multidimensional Index Structures 244
7.3 Summary 270

8 Clustering Techniques 271
8.1 Quality of a Clustering Scheme 272
8.2 Graph-Based Clustering 275
8.3 Iterative Methods 280
8.4 Multiconstraint Partitioning 286
8.5 Mixture Model Based Clustering 287
8.6 Online Clustering with Dynamic Evidence 288
8.7 Self-Organizing Maps 290
8.8 Co-clustering 292
8.9 Summary 296

9 Classification 297
9.1 Decision Tree Classification 297
9.2 k-Nearest Neighbor Classifiers 301
9.3 Support Vector Machines 301
9.4 Rule-Based Classification 308
9.5 Fuzzy Rule-Based Classification 311
9.6 Bayesian Classifiers 314
9.7 Hidden Markov Models 316
9.8 Model Selection: Overfitting Revisited 322
9.9 Boosting 324
9.10 Summary 326

10 Ranked Retrieval 327
10.1 k-Nearest Objects Search 328
10.2 Top-k Queries 337
11 Evaluation of Retrieval

11.1 Precision and Recall 381
11.2 Single-Valued Summaries of Precision and Recall 381
11.3 Systems with Ranked Results 383
11.4 Single-Valued Summaries of Precision-Recall Curve 384
11.5 Evaluating Systems Using Ranked and Graded Ground Truths 386
11.6 Novelty and Coverage 390
11.7 Statistical Significance of Assessments 390
11.8 Summary 397

12 User Relevance Feedback and Collaborative Filtering

12.1 Challenges in Interpreting the User Feedback 400
12.2 Alternative Ways of Using the Collected Feedback in Query Processing 401
12.3 Query Rewriting in Vector Space Models 404
12.4 Relevance Feedback in Probabilistic Models 404
12.5 Relevance Feedback in Probabilistic Language Modeling 408
12.6 Pseudorelevance Feedback 411
12.7 Feedback Decay 411
12.8 Collaborative Filtering 413
12.9 Summary 425

Bibliography 427

Index 473

Color plates follow page 38