

Index

A

activities of daily living (ADLs) 244
 Akaike information criterion (AIC) 248
 a metric incremental clustering algorithm (AMICA)
 16

B

Bayesian information criterion (BIC) 248
 bioinformatics 86–105
 biosequences 85–105
 classifier function 90
 conservation function 90
 data structures 91
 suffix trees 91
 tries 91
 formalization and approaches 92
 e-neighbor pattern 100
 extraction constraints 93
 multi-period tandem repeats (MPTRs) 97
 variable length tandem repeats (VLTRs) 97
 future trends 101
 pattern discovery problems 95
 string, suffix, and don't care 90
 blocks 128
 algorithms 131
 complexity issues 134
 generation for single measure values 132
 processing interval-based blocks 132

refining the computation of 135
 support and confidence of 129

C

categorical data 14
 incremental clustering 14
 dendrogram 18
 features 18
 feature selection 18
 conflict of interests (COI) 170
 correlation-based feature (CSF) 21
 customer relationship management (CRM) 221

D

data mining 1–31
 decision trees 9
 J48 technique 12
 metric splitting criteria 9
 metric methods 1–31
 partitions, metrics, entropies 3
 geometry 6
 metric space 9
 data visualization 106–123
 classification 117
 visualization methods 108
 VizRank 108
 experimental analysis 112
 projection ranking 110
 deviance information criterion (DIC) 249

discretization 23
 a metric approach 23

F

frequent
 itemset mining (FIM) 32
 scalability tests 50
 path bases (FPB) 40
 pattern mining 32–56
 constraints 33–35
 anti-monotone 33–35
 bi-directional pushing 35
 monotone 33–35
 fuzzy
 partitions 126
 support definition 130

G

geographical information systems (GIS) 227

H

hierarchical Bayesian mixed-membership models (HBM-MMS) 241–275
 characterization 245
 dirichlet process prior 248
 relationship with other data mining methods 247
 strategies for model choice 248
 the issue of model choice 242
 two case studies 243
 disability survey data (1982-2004) 244
 PNAS biological sciences collection (1997-2001) 243
 hyperclique patterns 57–84
 all-confidence measure 59
 definition 61
 equivalence between all-confidence measure and H-confidence measure 62
 experimental results 72
 for identifying protein functional modules 80
 h-confidence 61
 as a measure of association 66
 cross-support property 63
 for measuring the relationship among several objects 68
 relationship with correlation 66
 relationship with Jaccard 66
 item clustering approach 72, 78
 miner algorithm 69
 detailed steps 69
 scalability of 77
 the pruning effect 74
 effect of cross-support pruning 76
 quality of 77

I

inductive logic programming (ILP) 182
 instrumental activities of daily living (IADLs) 244

L

leap algorithms 38
 closed and maximal patterns 38
 COFI-Leap 38
 with constraints 41
 COFI-trees 40
 HFP-Leap 38
 load distribution strategy 52
 parallel BifoldLeap 44
 sequential performance evaluation 47
 impact of P() and Q() selectivity 50

M

machine learning (ML) 199
 mining
 ontology 171
 XML documents 198–219
 classification and clustering 208
 document representation 209
 by a set of paths 211
 edge-centric approaches 204
 feature selection 209
 frequent tree discovery algorithms 204
 frequent tree structure 202
 modeling documents with Bayesian networks 213
 stochastic generative model 212
 tile-centric approach 205
 tree-based complex data structure 200
 Monte Carlo Markov chain (MCMC) 252
 multidimensional databases 127

N

National Long Term Care Survey (NLTCs) 263

O

on-line analytical processing (OLAP) 125

P

principal component analysis (PCA) 120

S

singular-value-decomposition (SVD) 229
 social
 network analysis (SNA) 150, 165
 applications 166
 authoritativeness 165

Index

- networking services (SNSs) 150
- network mining 149–175
 - advanced mining methods 156
 - co-occurrence
 - affiliation network 161
 - co-occurrence 160
 - from the Web 152
 - keyword extraction 160
- spatio-textual association rules 181–197
 - document descriptions 183
 - document management systems 177
 - image analysis 177
 - mining with SPADA 186
 - reference objects (RO) 182
 - task-relevant objects (TRO) 182

T

- toll-like receptors (TLR) 89
- topic and cluster evolution 220–239
 - evolving topics in clusters 222
 - tasks of topic detection and tracking 222
 - tracing changes in summaries 222
 - for a stream of noisy documents 228
 - application case (automotive industry) 228
 - monitoring changes in cluster labels 223
 - monitoring cluster evolution 225
 - spatiotemporal clustering 226
 - remembering and forgetting in a stream of documents 225
 - topic evolution monitoring 230
 - visualization of linked clusters 233
- translation initiation sites (TIS) 89