

INDEX

- ABC model, 29, 32
Ab initio methods, 148
Activated B-like DLBCL, 91
Acute myeloid leukemia (AML), 102, 106
Adenine, 71
Affinity coefficient (AC), 332, 333
AGAVE, 176, 180
Agglomerative hierarchical clustering, 333
Alignment-based kernels, 158, 162
Amarzguioui's rule, 132, 135
Amino acids, electronic properties of, 273
Annotation-centric format, 194
Antisense, 63
 identification techniques, 58
 regulation identification, 63
 transcripts, 63, 64, 66
Arabidopsis genome, 58, 69
Arabidopsis thaliana, 58
AraC-type transcriptional regulators, 213
Association rule, 27, 29, 31, 33, 34
Astral database, 161
AT1G26290, 64
AT2G27740, 64
Average clustering coefficient, 254, 256, 326
Bader coefficient, 329
Balloting scheme, 6
 β -Barrel proteins, 267, 268
Batch update, 275
 rule, 281, 282
 strategy, 279
Baum–Welch algorithm, 4
Bayes error, 269, 272–274
Bayesian network
 methods, 136
 model, 131
biCluster, 305, 307, 308
 algorithm, 306
BiCluster graph, 307, 309, 315
Biclustering, 295, 300, 302, 303
Bidirectional recurrent neural network
 (BRNN), 6
Binary search tree (BST), 122
Biological sequence markup language
 (BSML), 176, 177, 179, 180
BioMolQuest, 180, 181, 188, 189
BLAST search tool, 188, 192
BLOSUM encoding, 11
Blosum62 substitution matrix, 3, 12

- Boolean constraints, 202
- Boolean constraint technique, 205
- Boolean expression, 202
- Boolean operator, 202
- Bridge concepts, 28
- Brill's tagger, 200
- Brownian particle, 330, 331
- B-term centered ranking (BCR), 27, 32, 37, 42

- Caenorhabditis elegans*, 114
- Candidate keyphrase, 199
- Catalase C-terminal domain, 211, 212, 226
- Child structure, 239, 241–243
- Chi-square, 1, 29, 31
 - ABC, 32, 33, 35
 - association rule ABC, 33, 34, 37
 - dependency graph matrix, 32, 34, 36
 - filtering, 32, 38, 43
 - link ABC, 33, 36, 40, 42
- Chitin, 58, 61
- Chou–Fasman algorithm, 2
- Classical back-propagation algorithm, 6
- CLICK, 302
- CLIQUE, 302
- Cluster, span of a, 308
- Clustering algorithm, 98, 106, 301
- Clustering coefficient, 254, 255, 326
- Cluster mean, 333
- Coil regions, 146
- Collection programming language (CPL), 191
- Combined classifier, 17
- Comparative modeling, 148
- Complete search, 77
- Conditional probability estimation, 334
- Consistency measures, 75
- Core-clustering coefficient, 339
- CoreNucleotide, 183
- Correlation coefficient, 75, 80, 329
 - Matthew's, 20
 - Pearson's, 75, 134, 300
- Correlation measures, 73, 75
- COSA, 302
- Cox model, 94
- Cox proportional-hazards regression, 93
- Cox score, 95
- Critical assessment for structure prediction (CASP), 145
- Crossover, 46, 50, 51, 78

- Cross-reference databases, 186
- Cross-validation test, 4, 6, 13, 18, 100
- Curcumin, 35, 38
- Cyclooxygenase-2 inhibitor(s), 53
- Cytosine, 71, 72

- DAG tertiary classifier, 17
- Database drivers, 191
- Data warehouse model, 182
- dbEST, 183
- dbGSS, 183
- Declarative query interface, 183
- Degree distribution(s), 254, 256, 326
- Degree of overlap, 269, 273
- Dense subgraphs, 335, 336
- Deoxyribonucleic acid (DNA), 71
- Depth-first search (DFS), 305
- Deterministic heuristic search, 77
- Dice coefficient, 329
- Dicer, 114–116, 120, 127, 138
- Direct kernel function, 158
- Disassortativity, 327
- Disjunctive normal form (DNF), 198, 202
- Distance-based clustering, 328, 334, 351
- Distance from first occurrence (DFO), 199, 200
- Distance measures, 74, 78, 332
- Distribution-free models, 150
- Drosophila melanogaster*, 130
- Dynalign, 235

- Edge connectivity, 340
- Effective frequencies, 151
- Electronic effects, 273
- Electronic polarizability, 273
- Elongation, 252, 255, 258
- EMBL database, 172, 173, 180
- EMBLxml format, 176, 180
- Entrez, 191
- Entropy, 75, 136, 156, 201
- Enumeration, 335
- Error surface, 280, 281
- EST (cDNA), 64
- Eukaryotic genomes, 58
- Eulerian graph representation, 137
- Evolutionary granular kernel tree (EGKT), 45
- Evolutionary voting kernel machines (EVKMs), 46, 51
- Extensible markup language (XML), 176

- False discovery rate (FDR), 105
- FASTA, 170–172, 184, 194
- Feature-based clustering, 300
- Feature subset selection, 76
- File transfer protocol (FTP) site, 173, 180
- Filter models, 73, 79, 84
- Finger-PRINTScan access point, 189
- Fisher criterion, 74, 78, 99
- FOLDALIGN, 234
- Fold change parameter, 105
- Fold prediction, 148, 149, 157, 161
- Frequency matrix encoding, 11
- F*-statistic, 78, 79, 83, 85
- Functional modules, 327, 338, 348

- G1/S transition, 313
- GAIN, 200, 201
- GDB, 191
- Genbank, 171, 172, 174, 183, 191
- Gene
 - expression
 - arrays, 323
 - data, 348
 - family knockdown, 129, 136, 138
 - identification, 98, 108
 - knockdown network, 138
 - selection, 72, 73, 78, 81–85
 - set, 303
 - subset selection, 73, 79–85
- Gene-expression signature, 96
- Gene ontology (GO), 313
 - index, 62
- General repository for interaction datasets (GRID), 254
- Generation procedure, 73, 76
- Generic API, 183
- Genetic algorithm(s) (GA), 4, 77, 83
- Genetic regulation, 117
- Genome information management system (GIMS), 182
- Genome-wide model, 253
- Genomic databases, 169
- GenPept, 171
- Germinal-center B-cell-like DLBCL, 91, 98
- Glutamine amidotransferases (GAT), 211
- Gram matrix, 150
- Grantham scale, 272, 275
- Granular kernel trees (GKTs), 45, 46
- Granular kernel tree structure evolving system (GKTSES), 46, 53
- Graph-based clustering, 300, 301, 320, 328, 335, 340
- Graph notation, 253
- Graph-theoretic approach, 253
- Graph-theoretic concepts, 340
- Greenwood's formula, 93
- Guanine, 71, 72

- Hadamard power, 345
- Hazard rate, 94
- Hazard ratio, 94
- HCS, 302, 340, 341
- α -Helical proteins, 267, 268
- Hidden Markov model(s) (HMM), 3, 157
- Hierarchy constraint, 238, 239
- Hodgkin's disease, 39
- Hubs, 326, 327
- Huesken's motifs, 133, 135
- Human genome project chromosome 22's database system (HGPC22), 191
- Huntington's disease, 113, 119
- Hybrid encoding, 11
- Hypothetical protein, 216

- Individual feature ranking, 76, 79
- Inflation, 345, 346, 348
- Information
 - gain, 76, 200–202
 - index, 154, 156
 - measure ranking, 27, 31, 32
- Information retrieval (IR), 136, 197, 206
- INSDseq, 176, 177, 180
- Internal ribosome entry sites (IRESs), 252
- International prognostic index (IPI), 89
- Intracellular protease, 214, 216, 228, 213
- Inverted file, 122

- JA, 135
- Jaccard coefficient, 328, 329
- Jagla's rules, 133, 134
- Joint probability distributions, 270–273

- Kaplan–Meier plot, 91, 93
- Kaplan–Meier survival curve, 91, 93, 108
- Kernel-based methods, 8
- Kernel function, 9, 11, 80, 135, 150, 154, 158, 159

- Kernel principal component analysis (KPCA), 8
- Keypphrase-based technique, 204
- k*-length
 - pathratio, 330
 - pathstrength, 330
- k*-nearest neighbors, 82, 100
- Korbel coefficient, 329
- KR data set, 134

- Laboratory information management system (LIMS), 188
- Lagrange multipliers, 80
- Large-B-cell lymphoma (DLBCL), 89, 90, 95, 98
- Latent semantic indexing (LSI), 30
- Leave-one-out method, 350
- Link-grammar, 200
- Line graph generation, 346
- Lipid transport, 313
- Liu–Deber scale, 270, 275
- Log-likelihood score, 130
- Log-rank test, 93, 100, 106, 108
- Loner proteins, 255, 258, 261
- Long-term survivors, 103, 104
- Lung adenocarcinoma, 97, 98
- Lymph-node signature, 96
- Lymphochip DNA microarrays, 95

- Mad cow disease, 113
- Major-histocompatibility-complex (MHC) class II signature, 96
- Mantel–Haenszel test, 93
- Many-to-many (M-to-M) relationship, 181
- MAP kinase activity, 313
- Marker genes, 72, 73, 82, 84, 85
- Markov clustering (MCL), 344
- Markov matrix, 345
- Maryland Bridge coefficient, 329
- Mass genome sequencing projects, 169
- Mass spectrometry (MS), 322, 323
- Maximal clique enumeration, 309
- Maximal tricluster, 298, 299
- Maximum likelihood (ML), 4
- Maximum-margin hyperplane, 8
- MCODE algorithm, 339
- Mean path, 326
- Mean squared errors (MSE), 135
- Median cut, 102
- Median rate of false positives (mRFPs), 162
- Mercer’s conditions, 150
- Messenger RNA(s) (mRNA), 63, 114, 138, 234, 324
- MFOLD, 234
- Microarray, 57
 - expression data, 295
- Microarray-based cancer classification, 72
- MicroRNA (miRNA), 114, 116, 138, 234
- Minimumcut (Mincut), 340
- MIPS functional categories, 254
- Mismatch kernel, 158
- Model-based gene expression clustering, 301
- Modeler programs, 148
- Model training step, 334
- Molecular complex detection (MCODE), 339
- Monte Carlo approach, 336
- Monte Carlo optimization, 336
- Multiple-database access points, 186
- Mutation, 50
- Mutual information, 30, 31, 33, 38, 143
- MySQL DBMS database server, 182

- Naive cost function, 342
- Name equals value pairs, 172
- Natural antisense transcript (NAT), 69
- Nearest-neighbor method, 2
- Nearest shrunken centroid model, 101, 102
- Network topology, 254
- Neural network (NN), 6, 133, 134, 152
 - methods, 5
- Neuron(s), 5, 277–279, 282
- NLParser, 200
- Nodes, 5, 325–327
- Noncoding RNAs (ncRNAs), 234
- Nondeterministic heuristic search, 77
- Nonontology
 - noun phrase category, 202
 - proper noun phrase category, 202
- Non-scale-free topology, 256
- Nontransmembrane (NOT TM), 275, 283, 287
- Nucleophilic elbow, 214
- Nucleotide, 183
- Null hypothesis, 31, 32, 34, 93

- Object-oriented database (OODB), 182, 183
- ODMB Java binding, 183
- Off-target effect(s), 120, 125, 138

- Off-target error, 122, 123, 128
- Oligomeric proteins, 210
- Oligomerization states, 210, 214
- Ontological analyses, 58
- Ontology phrase category, 202, 203
- OP-Cluster, 303
- Open reading frame (ORF), 63
- OPSM, 303
- ORCLUS, 301
- Output expansion process, 6
- Overlapping clusters, 296
- Overlap ratio(s), 269, 271
- OWL database, 181, 189
- Oxidoreductase activity, 313

- Pantothenate biosynthesis, 313
- Parent circle, 238
- Parent structure, 239, 340
- Partial likelihood, 94
- Partition mean, 333
- Part-of-speech (POS)
 - phrase categorization, 198
 - tagging, 200
- Path, 326
- PathRatio, 330
- PathStrength, 330
- Pattern-based clustering, 300
- Pattern proximity, 327
- PCKimmo, 200
- pCluster, 302, 303
- Per-instance error surfaces, 281, 282
- Per-state Matthews correlation coefficients, 155
- Phase transition, 138
- PHD, 152
- Pleurisy, 39
- Poisson distribution, 256, 326
- Polarity, 3, 271
- Polyacrylamide gel electrophoresis (PAGE), 323
- Portable graymap (PGM), 333
- Posttranscriptional gene silencing (PTGS), 113, 114, 116, 139
- Posttranslational modifications (PTMs), 253, 323
- Potato virus Y (PVY), 118
- Power-law degree distribution, 326
- Power-law regression, 257
- Precedence constraint, 238

- Precision, 123, 136
- Prediction analysis for microarrays (PAM), 101
- Prediction model construction, 98
- Primary distance, 333
- PRINTS-S database, 181, 189
- Probabilistic greedy algorithm, 129
- PROCLUS, 301
- Profile kernel, 157, 158
- Proliferation signature, 96
- Protein(s)
 - biosynthesis, 252
 - complexes, 327
 - family profiles, 148
 - functional categories of, 262
 - microarray, 323
 - phosphatase regulator activity, 313
 - phosphorylation, 261
 - physicochemical analysis of, 268
 - translation, 251
 - networks, 251, 253, 255, 258
- Protein-protein interaction(s) (PPI), 209, 319
- Proteomics, 320, 323
- Pseudoknots, 234, 236
- Pseudo-relevance feedback, 197
- PSI-BLAST, 11, 148, 211, 224, 228
- PSIPRED, 152, 157
 - method, 6
- PSSM encoding, 11, 13
- PTGS, 116, 139

- Quadratic programming (QP), 9
- Quaternary structures, 210, 214
- Query structure, 233, 238

- Radial basis function (RBF), 9, 11
- Range multigraph, 303, 305, 315
- Ratio range, 303, 304
- Raynaud's disease, 27, 35
- RdRP, 116, 139
- Receiver operating characteristic (ROC) scores, 162
- Relational databases, 180, 181
- Relevance feedback, 197
- Respiratory syncytial virus (RSV), 118
- Restricted neighborhood search clustering algorithm (RNSC), 342, 343
- Retinal neoplasm, 39
- Reynolds' rule(s), 132, 134, 135

- Ribonucleic acid (RNA), 63, 71, 234
 - mining operations, 233
 - secondary structures, 233
 - string kernels, 124, 139
- Ribonucleoprotein complex (miRNP), 116, 138
- Ribosomal RNA (rRNA), 234
- RIBOSUM substitution matrices, 236
- Risk
 - index, 95
 - score, 96, 104
- RNA-binding proteins, 213
- RNAdistance, 236
- RNAfold, 130, 234
- RNAforester, 236
- RNAi kinetics, 119, 139
- RNA-induced silencing complex (RISC), 114–116, 121, 139
- RNA interference (RNAi), 113, 114, 139
- RNA–RNA duplexes, 131
- Root, 59
- Roulette wheel method, 49
- RSmatch, 233, 336, 237, 247
- RSview, 233, 244, 245, 247

- SAMBA, 302
- Scaled cost function, 342
- Scale-free network, 326
- Scale-free topology, 256
- Saccharomyces cerevisiae*, 120
- SCOP database, 161
- Scoring matrix, 151
- Search–deletion engine, 119
- Secondary distance, 333
- Secondary-structure prediction, 1
- Segment overlap measure (SOV), 21, 154
- Self-organized global ranking (SOGR)
 - algorithm, 275
- Self-organizing map (SOM), 12
 - algorithm, 275
- Semantic filtering, 30
- Semantic knowledge-based ranking, 32
- Sense, 63
- Sensitivity, 351
- Sequence centric format, 194
- Sequential backward selection, 77
- Sequential forward selection, 77
- Shared connections ranking, 32
- β -Sheets, 146
- Shifting cluster, 298
- Shortest path, 326
- Short hairpin RNA (shRNA), 116, 139
- Short temporal RNAs (stRNAs), 116, 139
- Short-term survivors, 103
- Sigma cross-reacting proteins, 213
- Signal-to-noise measure, 100
- Signal-to-noise statistic, 99
- Significance analysis of microarrays (SAM), 104
- Silencing efficacy, 139
- Similarity learning method, 334
- Simple voting tertiary classifier, 16
- Simpson coefficient, 329
- Simulated annealing (SA), 77
- Small interfering RNA (siRNA), 139, 234
 - design rules, 139
 - efficacy prediction, 139
- Small nuclear RNA (snRNA), 234
- Small nucleolar RNA (snoRNA), 234
- Small-world effect, 326
- SOGR-I, 277, 287, 289, 289
 - classifier, 287
- SOGR-IB, 278, 287, 289, 289
- SOGR algorithm, 276
 - variants of, 277
- Specificity, 351
- Spectrum kernel, 157
- Split ranges, 304
- S. pombe*, 126, 127
- SQ line, 176
- Standard derivation, 37
- Stochastic context-free grammars (SCFGs), 236
- Stochastic gradient update strategy, 279
- Structural classification of protein (SCOP), 349
- Structural risk minimization (SRM), 8
- Structure components, 238
- Subject structure, 239
- Substitution cost matrix (SCM), 14
- Substitution matrix-based kernels, 12
- Superparamagnetic clustering (SPC), 343
- Support vector(s) (SV), 93, 134
- Support vector machine(s) (SVM), 1, 100, 134
- Survival function, 93
- Swiss-Prot file transfer protocol (FTP) server, 180

- Swiss-Prot format, 172
Synset, 203
Synthetic protein translation networks, properties of, 261
- Tabu search (TS) algorithm, 78
TAMBIS system, 190
Tandem affinity purification (TAP), 325
Target frequencies, 151
TargetScan, 131
Termination, 252, 259
Tertiary classifier, 17
Threading, 149
Three-dimensional microarray data set, 296
Three-state per-residue accuracy, 154
Threshold, 101, 105
Thymine, 71
Tiling array, 57, 58
Transcriptional gene silencing (TGS), 116
Transcriptional regulation, 252
Transfer matrix, 331
Transfer RNA (tRNA), 63, 234
Transitive RNAi (tRNAi), 116, 139
Transmembrane (TM), 283
 proteins, 267
Tree-based tertiary classifier, 16
TrEMBL, 172
triCluster(s), 295, 296
Tuning parameter, 105
- Ubiquitin conjugating enzyme activity, 313
Ubiquitin cycle, 313
- Ui-Tei's rule, 132
Undiscovered public knowledge (UPK), 28
Unified medical language system (UMLS), 29, 206
Univariate model, 95, 108
Untranslated regions (UTRs), 69, 234
Unweighted pair group method with arithmetic mean (UPGMA), 333
UVCLUSTER, 332–334
- Valid cluster, 298
Vasoconstriction, 28
Verb phrase category, 202
Vertex-weighted graph, 339
Vienna RNA package, 234, 346
Viral gene knockdown, 118
Voting scheme-based evolutionary kernel machines (EVKMs), 46, 54
- Wald test, 96
Weighted-voting algorithm, 99
Whole-genome arrays, 58
Wilcoxon rank sum test, 99
Wrapper models, 73, 76, 80, 84
- xMotif, 303
- YASSPP, 152, 154
Yeast two-hybrid (Y2H) system, 322, 324
- Zero shifting, 159
Zimmerman–Eliezer–Simha scales, 273

