Index

ϵ-*decreasing* strategy, 163
ϵ-*first* strategy, 163
ϵ-*greedy* strategy, 163
k-shot learning, 225
n-way k-shot learning, 225, 239

A3R, 31
absorption, 287
abstract operators, 132
acquisition function, 110
active testing (AT), 93
adaptive batch classifiers, 203
adaptive capping, 108
adaptive dataset similarity, 117
adaptive ensembles, 203
adequacy of configuration spaces, 146
aggregate functions, 278
aggregate operator
  in Rapid Miner, 279
aggregating data, 278
AI planning, 131
algorithm
  configuration (AC), 104
  recommendation, 19
  selection, 19
algorithm recommendation
  for data streams, 201
algorithm selection, 77
ALMA, 196
AlphaD3M system, 136
approximate ranking trees (ARTs), 93
AR*, 32
arbitrating, 182
ART forests, 93
ART trees, 93
attentive recurrent comparators, 241, 247
Auto-WEKA
  parameter space, 127
Automated Statistician, 279
automatic feature construction
  in DNNs, 288
automatic machine learning (AutoML), 104
automating
  data preprocessing, 274
  data science, 269
  report generation, 279
  the design of complex systems, 283
  workflow design, 123
AutoML, 104
average ranking, 27

bagging, 170
base-level operators, 132
batch-incremental classifiers, 203
best classifier on last interval, 209, 210
best-first search, 107
bias
  control, 155
  domain-specific, 155
  weak/strong, 155
bias vs. variance in metalearning, 231
Blast, 209, 210
Boltzmann exploration, 163
boosting, 172
bounds
  for domain adaptation, 231
  on generalization error, 229

© The Editor(s) (if applicable) and The Author(s) 2022

P. Brazdil et al., *Metalearning*, Cognitive Technologies,
https://doi.org/10.1007/978-3-030-67024-5

using algorithmic stability, 230
budget, 24
business problem, 270

cascade generalization, 175
cascading, 177
CASH, 104
catastrophic forgetting, 226
CHADE, 196
changing the granularity of representation, 278
characterizing diversity, 159
by correlation of rankings, 159
choice of base-classifiers, 213
CITRUS, 136
class overlap, 59
class separability, 59
classification models at meta-level, 83
classifier output difference (COD), 153, 193, 213
classifiers
adaptive, 203
batch-incremental, 203
clause structure grammar, 290
cleaning, 277
clustering trees, 81
CMA-ES, 107
COD, 153, 213
cold-start problem, 113
combined algorithm selection and hyperparameter optimization (CASH), 104
combined measure, 31
combining
accuracy and time, 31
combining base-learners, 169
competence region, 195
complete metadata, 159
concept drift, 202
detector, 215
conditional hyperparameter, 103, 145
conditional neural processes, 241, 255
conditional operators, 285
configuration space, 144
continuous, 145
discrete, 145
configuration spaces, 98, 143
adequacy, 146
principles of constitution, 147
reduction, 151
constructive induction, 286
context-free grammars (CFGs), 128
induction, 130
limitations, 130
contextual-bandit problem, 164
continuous configuration space, 145
control layer
in two-layered architecture, 215
controlling
bias, 155
cuboids, 279

data
characteristics, 53
preprocessing, 274
science, 269
transformation, 275
data characterization
algorithm-specific, 68
dynamic, 68
in clustering, 63
in regression tasks, 60
in time series prediction, 62
in unsupervised learning, 63
iterative, 68
task-dependent, 54
useful for ranking pairs of algorithms, 69
data envelopment analysis, 33
data preprocessing
cleaning, 277
instance selection, 277
outlier elimination, 277
data stream, 201
ensembles, 210
data wrangling, 275
dataset, 25
repositories, 157, 298
segmentation, 158
similarity, 25, 34
variants, 157
datasets
2D footprints, 158
synthetic, 157
which are needed?, 156
with discriminatory power, 158
DEA, 33
declarative bias, 125
default settings, 104
defining

abstract operators, 289
configuration space, 125
delegating, 180
descriptors of learning goals, 273
discounted cumulative gain (DCG), 51
discrete configuration space, 145
DM processes, 123
domain-specific
metafeatures, 271
double-loop architecture, 225
drift detection, 207
metafeatures, 207
drift detector
ADWIN, 207
DDM, 207
drill down/up, 278
dynamic
algorithm selection, 205
classifier selection, 190, 195
data characterization, 68
selection of models, 195

empirical risk, 228
ensemble learning, 189
generation, 192
integration, 195
iterative process, 193
phases, 192
pruning, 192
ensemble of surrogate models (SGPT), 116
ensemble recommendation
by ensemble learning, 192
by selection, 191
ensembles
accuracy-weighted, 214
adaptive, 203
for data streams, 210
heterogeneous, 191
hierarchical, 197
homogeneous, 191
in Auto-sklearn, 195
ensembles of algorithms, 169
entropy search, 111
error correlation, 153
establishing hyperparameter importance, 149
evaluation, 30
by correlation, 45
of rankings, 30, 39
of recommendations, 39
evolutionary computation (EC), 197
evolutionary search, 107
expected improvement (EI), 111
expected loss, 228
experiment design, 118
explainability, 284

fading factors, 211
faithful symbolic model, 284
feature drift, 213
feature-based transfer learning, 222
few-shot learning, 240
Flashfill wrangler, 277
folding and unfolding, 287
FOOFAH wrangler, 276
Friedman's test, 50
functional transfer, 221
in neural networks, 223

Gaussian mixture model (GMM), 159
Gaussian processes (GPs), 80, 111
with multi-kernel learning (MKL-GP), 115
goal dependency networks (GDN), 273
golden standard, 46
gradient descent, 108
granularity of representation, 285
graph neural networks, 241, 246
grid search, 105

Hasse diagrams, 20
heterogeneous ensembles, 191
heuristic ranker, 128
heuristic search methods, 107
hierarchical ensembles, 197
generation, 197
hierarchical planning, 132
abstract operators, 132
base-level operators, 132
HTN planning, 133
ML-Plan, 133
hill climbing, 107
Hoeffding bound, 203
Hoeffding tree, 203
homogeneous ensembles, 191
homogeneous transfer, 221
HTN planning, 133, 289
Hyperband, 109
hypergradients, 108

hyperparameter
  categorical, 103
  conditional, 103
  numeric, 103
  optimization (HPO), 104
hyperparameter importance, 148
  ablation analysis, 148
  forward selection, 148
  functional ANOVA, 148
hyperparameter log-scale, 145
hyperparameters
  establishing importance, 149

identifying the domain-specific
  background knowledge, 273
  data, 273
identifying the task domain, 272
  by classification, 273
  by matching descriptors, 272
incomplete metadata, 159
  effects, 162
  using estimates, 161
inductive logic programming (ILP), 274, 288
initialization
  of configuration, 113
inner loop, 225
input warping, 145
instance selection, 277
instance-based transfer learning, 222
Intelligent Discovery Assistant (IDA), 126
inter-construction, 288
interplay between subsymbolic and symbolic models, 284
interval estimation strategy, 164
interval of confidence, 22
introducing new concepts, 285
  autonomously, 286
  by clustering, 286
  from external sources, 285
introduction of new concepts
  absorption, 287
  folding and unfolding, 287
  inter-construction, 288
  propositionalization, 288
invariant transformations, 224
iterated local search, 108
iterative data characterization, 68
iterative learning, 290
iterative pairwise tests, 90

KDD process, 124, 270
Kendall's Tau correlation, 46
kernel density estimation, 112

landmarkers, 84
  in ensemble learning, 197
latent embedding optimization, 261
layered learning, 289
learning
  k-shot, 225
  inter-dependent tasks, 292
  multi-task, 223
  nested, 225
  rankings, 115
  similarity, 225
  with recurrent neural networks, 226
learning goals, 272
  schedule, 272
learning similarity, 225
learning-to-learn, 219, 229
loose accuracy, 50
loss curve, 31, 47
loss–time curve, 49
LSTM optimizer, 241, 257

macro-operators, 136
marginal, 149
marginal contribution, 195
  of algorithms, 148
  Shapley value, 148
  to performance, 148
matching descriptors (metafeatures), 272
matching networks, 241, 245
median rank, 26
memorization problem, 263
memory-augmented neural networks, 226, 241, 249
meta networks, 241, 251
meta-decision trees, 183
meta-level performance, 208
meta-model, 24
  instance-based (IBL), 80
  regression-based, 78
Meta-SGD, 261
MetaBags, 196
metadata, 24
  complete, 159
  incomplete, 159
  repositories, 297
metadistribution, 229

metafeature-based approaches, 205
metafeatures, 34, 53
  aggregated, 65
  complexity-based, 58
  concept variation, 58
  domain-specific, 271
  in ensemble learning, 194
  information-theoretic, 56
  landmarker sets, 58
  landmarkers, 57
  model-based, 56
  relational representation, 69
  relative landmarkers, 57
  selection, 67
  simple, 55
  statistical, 55
  subsampling landmarkers, 58
metalearning, 219, 225
  approaches, 19
  in ensemble construction, 189
  in ensemble learning, 189, 195
  metric-based, 237, 240, 241, 248, 263
  model-based, 237, 241, 248, 255, 263
  optimization-based, 237, 241, 256, 263
metalearning approaches, 77
metaparameter, 225
MiningMart, 136
ML-Plan, 133
MOA framework, 213
model, 24
  meta-level, 24
Model inference system (MIS), 290
model-agnostic meta-learning, 241, 259
model-based
  metafeatures, 56
Monte Carlo tree search (MCTS), 136
multi-armed bandits (MAB), 162
multi-criteria analysis, 33
multi-fidelity techniques, 108
multi-task learning, 223, 227
multitarget prediction, 81

Naive AutoML, 136
negative transfer, 220
Nemenyi test, 50
nested learning, 225
non-homogeneous transfer, 221
non-propositional approaches, 98
non-stationarity, 202
normalizing performance, 42, 80

obtaining data, 273
  by planning, 273
  from different sources, 274
  from OLAP data cube, 274
OLAP data cube, 274, 278
online
  bagging, 204
  boosting, 204
  learning, 202
  MAML, 203
ontologies, 126, 289
  DM ontology of Intelligent Discovery Assistant (IDA), 126
  shortcomings, 127
OpenAI Gym, 299
OpenML, 297, 299
  dataset characteristics, 299
  datasets, 299
  flows, 301
  integration in Python, 305
  integration in R, 303
  integration in Weka, 303
  runs, 302
  setups, 302
  tasks, 300
operators, 131
  abstract, 132
  base-level, 132
  conditional, 285
  manual selection, 131
  post-conditions/effects, 132
  preconditions, 132
  repeat, 285
optimistic reward estimate, 164
order of experiments
  does it matter?, 162
  multi-armed bandits, 162
  sequential model-based optimization, 162
outer loop, 225
outlier elimination, 277

pairwise comparisons of algorithms, 83
parameter-based transfer learning, 222
partial
  learning curves, 85
  ranking, 90
PCC-DES, 196
performance, 22
  aggregation, 26

comparable, 22
high-performance regions, 117
loss, 47
metadata, 24
models, 80
normalization, 80
planning
exploiting metaknowledge, 135
plans, 284
portfolio, 144
characterizing diversity, 159
of algorithms, 191
of ensemble elements, 191
of successful workflows, 137
reduction, 151
pricing strategies (POKER), 164
probability matching strategy, 163
probability of improvement (PI), 111
problem
description, 270
understanding, 270
procedural bias, 128
programmatic access to datasets, 298
propositionalization, 288
prototypical networks, 246

Q-statistic, 193
quantile regression, 117
quantile regression forest (QRF), 117
query set, 240

random forests (RFs), 80
as surrogate models, 112
random search, 106
rank correlation
Spearman's, 46
weighted, 47
ranking, 20
accuracy, 46
average, 27
complete, 20
evaluation, 30
incomplete, 20, 34, 35
linear, 23
meta-model, 28
of successful plans (workflows), 136
partial, 20
quasi-linear (weak), 23
total, 20
Rapid Miner
aggregate operator, 279
recommendation, 20
forms, 20
of algorithm, 20
of workflows (pipelines), 32
recurring meta-level models, 213
reducing
configuration spaces, 151
portfolios, 151
reduction
by envelopment curve, 154
reformulating
ontologies, 289
recursive definitions, 290
theories, 286
reformulation of theories
by specialization, 286
reinforcement learning optimizer, 241, 258
relation networks, 246
repeat operators, 285
report generation, 279
representation of goals, 273
representational transfer, 221
reproducibility, 298
reptile, 241, 261
rescaling, 42
reusing new concepts
in further learning, 289
to redefine ontologies, 289
reusing prior experiments, 298
reusing solutions, 289
Robot Scientist, 273
roll-up, 278

schedule for learning goals, 272
scheduling experiments, 162
ϵ-*decreasing*, 163
ϵ-*first*, 163
ϵ-*greedy*, 163
interval estimation strategy, 164
multi-armed bandits, 162
optimistic reward estimate, 164
pricing strategies (POKER), 164
probability matching, 163
SoftMax, 163
upper confidence bound (UCB), 164
search
for the best workflow, 279
search methods

that exploit metaknowledge, 114
selection
of preprocessing method, 278
of workflows, 278
sequential model-based
optimization (SMBO), 110
search, 110
SGPT, 116
Shapley value, 148
shifting window, 206
Siamese neural networks, 241
siamese neural networks, 244
significant wins, 27
similar learning curve, 86
similarity
correlation-based, 70
cosine-based, 70
performance-based, 97
similarity measures, 69
simple neural attentive meta-learner, 241, 253
SKILit, 290
sliding window, 206
SoftMax strategy, 163
source domain, 220
Spearman's rank correlation, 46
SPegasos, 203
stacking, 173
statistical test, 22
stochastic gradient descent (SGD), 108, 203, 226, 262
stream classifiers, 213
strong sufficiency, 274
subsampling landmarkers, 84
successive halving, 108
sum of wins and losses, 88
supervised
learning, 238
metalearning, 238
supervised transfer, 221
support set, 240
surrogate
collaborative tuning (SCoT), 115
model, 110
SVM
effects of hyperparameters settings, 306
SVMRank, 115
SYNTH, 277
Synth-a-Sizer wrangler, 277
synthetic datasets, 157
tabling, 136
target domain, 220
task
descriptor generation, 271
domain, 272
goals, 271
type, 271
task embedding, 263
task-dependent characterization, 54
templates, 198
top-*n* strategy, 28
TPOT, 133
transfer
across tasks, 219
by clustering, 227
functional, 221
homogeneous, 221
in kernel methods, 227
in neural networks, 223
in parametric Bayesian models, 227
learning, 219, 220
negative, 220
non-homogeneous, 221
of knowledge, 219
representational, 221
supervised, 221
unsupervised, 221
transfer acquisition function (TAF), 116
transfer learning
feature-based, 222
instance-based, 222
parameter-based, 222
transformation
of data, 274
tree-structured Parzen estimator (TPE), 112
Trifacta wrangler, 276
true ranking, 46
true risk, 228
trusted AI, 284
two-layered architecture
control layer, 215

unsupervised transfer, 221
upper confidence bound (UCB), 164
upper confidence bounds (UCB), 111

Vapnik–Chervonenkis (VC) dimension, 228
version spaces, 154

warm-start strategy, 113
window size, 208
workflow design
  manual, 131
  using planning, 132
workflow/pipeline design, 123
workflows, 284
wrangling, 275
  Flashfill, 277
  FOOFAH, 276
  Synth-a-Sizer, 277
  Trifacta, 276