

Cambridge University Press

978-1-107-03607-9 - Statistical Methods for Recommender Systems

Deepak K. Agarwal and Bee-chung Chen

Index

[More information](#)

Index

Note to index: An *f* following a page number indicates a figure; a *t* following a page number indicates a table.

- A/A test, 67. *See also* online bucket tests
- A/B test, 8, 55. *See also* online bucket tests
- accuracy metrics, 59–62, 63
- adaptive sequential design. *See* multiarmed bandit (MAB) problem
- Adomavicius, Gediminas, 261
- Agarwal, Deepak K., 97, 99, 107, 110, 147, 198, 210, 242, 250*t*, 251, 252, 253, 261
- algorithmic techniques, for recommender system, 5–7
- alternating least squares, 35
- AOL, 82*t*
- application settings, 85–87
 - context-related characteristics, 87
 - item-related characteristics, 85
 - response-related characteristics, 87
 - user-related characteristics, 85–86
- architecture. *See* system architecture
- area under ROC curve (AUC), 62
- ARS-based EM for logistic response, 157–161
 - E-step, 157–160
 - centering, 160
 - initial points for ARS, 160
 - M-step, 160–161
- article recommendation, 9, 24, 85, 95, 134, 177, 186. *See also* Yahoo! Today module; Yahoo! Buzz; news article recommendation, multifaceted
- AUC (area under ROC curve), 62
- Auer, P., 46–47, 108, 109
- bag-of-words, 5
 - item characterization, 18–21
 - dense format, 19*f*
 - dimension reduction, 19–20
 - phrases and entities, 19
 - sparse format, 18–19*f*
 - term frequency-inverse document frequency version, 18
 - term frequency version, 18
 - unweighted version, 18
- bandit scheme, 42. *See also* multiarmed bandit (MAB) problem
- Basu, Sugato, 261
- Bayardo, Roberto J., 261
- Bayes 2×2 scheme, 107
- BayesGeneral scheme, 106–107, 112*t*
- Bayesian approach, to multiarmed bandit (MAB) problem, 42–46
- Bayesian solution, most-popular recommendation, 98–107
 - BayesGeneral scheme, 106–107, 112
 - Gamma-Poisson (GP) model, 99
 - general solution, 104–107
 - dynamic set of candidate items, 105–107
 - nonstationary CTR, 107
 - two-stage approximation, 105
 - $K \times 2$ case, 102–104
 - convexity, 104
 - Lagrange relaxation, 103–104
 - near-optimal solution, 104
 - separability, 104
 - one-step look-ahead, 99

Cambridge University Press

978-1-107-03607-9 - Statistical Methods for Recommender Systems

Deepak K. Agarwal and Bee-chung Chen

Index

[More information](#)

274

Index

- Bayesian solution, most-popular (*cont.*)
 - 2×2 case, 99–101, 112
 - normal approximation, 101–102
 - optimal solution, 101
 - properties of gain function, 102
- Bayes optimal scheme, most-popular recommendation, 98
- bcookies (browser cookies), 85–86
- Bergstrom, Peter, 32
- Best, N. G., 160
- bias, 67, 73, 75–76, 177–178, 208–209, 212–213
- bias smoothed tensor (BST) model, 212, 217, 226, 229–231
- bi-grams, 19
- bilinear models. *See* Fast Online Bilinear Factor Model (FOBFM); matrix factorization
- binary ratings, 28, 64, 166
- Blei, David M., 21
- BM25, 26, 227, 228
- Bookcrossing, 201, 203
- Booth, James G., 151
- bootstrap sampling, 67
- Borchers, A., 37
- bounce clicks, 7
- bounce rate, 261
- B-POKER, 110–111*f*, 112, 114*f*
- browser cookies (bcookies), 85–86
- BST (bias smoothed tensor) model, 212, 217, 226, 229–231
- B-UCB1, 108–109, 110–111*f*, 112, 113, 114*f*, 115, 116*f*, 117*f*
- Burges, Christopher J. C., 261
- categorization, 17–18, 24
- Cesa-Bianchi, N., 46–47, 108, 109
- Chen, Bee-Chung, 97, 99, 107, 110, 147, 198, 210, 242, 250*t*, 251, 252, 253, 261
- Chernoff-Hoeffding bound, 47
- classical bandit, 42
- classical methods, 15–38
- click lift, 172, 177, 178–179*t*, 180*f*, 253
- click-through rate (CTR), 5, 7, 10–14, 13*f*, 68, 69, 241
- clustering, 49
 - hierarchical, 51*f*
 - k -means, 249, 250*t*, 251, 256, 260
 - TopItem, 249, 250*t*, 254
- CMF (collapsed matrix factorization), 216
- cold start, 36, 37, 122, 142, 179*t*, 199, 204
- collaborative filtering, 31–36
 - augmented by artificial feature-based ratings, 37
 - based on item-item similarity, 32–33
 - based on user-user similarity, 31–32
 - dimension reduction, 52
 - matrix factorization, 33–36
 - optimization methods, 35–36
 - regularization, 34–35
- collapsed matrix factorization (CMF), 216
- content-based profile, 24–25
- content filtering and understanding, 5, 7
- context-dependent recommendation, 88–89, 206–236
 - hierarchical shrinkage, 211–218
 - location-dependent, 206
 - multiapplication, 206
 - multicategory, 206
 - multifaceted news article recommendation, 218–233
 - related-item recommendation, 206, 233–235
 - combining predicted response/relatedness, 235
 - response prediction, 234–235
 - semantic relatedness, 233–234
- tensor factorization models, 207–211
 - discussion, 210–211
 - model fitting, 209–210
 - modeling, 208–209; bias and popularity, 208–209; feature-based aggression, 208; regression priors, 209; three-way interaction of factors, 209; two-way interactions of factors, 209
- context items, 233
- context-related characteristics, 87
- contextual information, 30–31
- control bucket, 55
- coordinate descent, 30
- COS (cosine similarity model), 227, 228
- cosine similarity, 26
- cross-validation, 12, 58, 164–165, 178, 197, 203, 226, 252
- crowdsourcing, to define/flag low-quality content, 5
- CTR lift, 253
- cumulative logit model, 28–29
- cut-point vector, 30
- data
 - exploratory analysis, 219–226
 - partitioning, 170–171

Cambridge University Press

978-1-107-03607-9 - Statistical Methods for Recommender Systems

Deepak K. Agarwal and Bee-chung Chen

Index

[More information](#)

- sparseness, 14, 19–20, 50–54, 112, 144
- splitting methods, 56–59
- user data store, 91
- database, lexical, 20
- decision trees, 49, 51, 147
- declared profile, 23–24
- deep learning, 21
- demographics, 5, 23–24
- de Moura, Edleno Silva, 261
- DGP (dynamic Gamma-Poisson) model, 107–108, 110
- dimension reduction, 19–20, 50–54, 123
- discounted-reward settings, 43, 44
- diversity, optimizing for, 7, 261
- domain-specific sites, 82
- dynamic Gamma-Poisson (DGP) model, 107–108, 110
- dynamic state-space model, 11
- EachMovie, 164, 165*t*–166
- e-commerce, 3, 5, 17, 206, 233
- editorial labeling, 5
- ϵ -greedy, 14, 47, 49, 52, 53, 109, 112, 113, 114*f*, 116*f*, 117*f*, 132–133
- Elango, Pradheep, 97, 99, 107, 110, 242, 250*t*, 251, 252, 253, 261
- EM. *See* expectation-maximization (EM)
- EMP. *See* estimated most popular (EMP) item
- ensemble, 37
- error, 9, 60–61, 113, 148, 155, 164, 203, 252
- estimated most popular (EMP) item, 112–114*f*
 - EMP regret, 113–114*f*
 - fraction of EMP display, 113–114*f*
 - non-EMP regret, 113–114*f*
- Euclidean distance, 21, 52, 185, 188
- EWMA (exponentially weighted moving average), 11–13, 49
- Exp3, 47, 109, 110, 111*f*, 112
- expectation-maximization (EM)
 - Monte Carlo EM, 151, 156, 186, 191–192, 197, 209–210, 213–214
 - personalization through factor models, 157–164
 - personalization through feature-based regression, 126–130
 - variational approximation, 161–164
- explicit ratings, 65
- explore-exploit, 39–54, 42, 88
 - Bayesian approach (*see* most-popular recommendation, Bayesian solution)
 - data sparsity, 50–54
 - introduction, 40–41
 - most-popular recommendation, 48–49
 - multiarmed bandit problem, 40–48
 - personalized recommendation, 49–50
 - exponentially weighted moving average (EWMA), 11–13, 49
- Facebook, 76, 82, 219, 220, 224. *See also* news article recommendation, multifaceted
- factorization
 - collapsed matrix, 216
 - FEAT-ONLY, 172–173
 - matrix, 33–36
 - separate matrix, 216–217, 226
 - tensor, 207–211, 213, 234
 - See also* factorization through latent Dirichlet allocation (fLDA); personalization through factor models; regression-based latent factor model (RLFM)
- factorization through latent Dirichlet allocation (fLDA), 185–205
- experiments, 198–204
 - Bookcrossing data set, 201, 203
 - MovieLens data, 198–199
 - Yahoo! Buzz application, 199–201
- model, 186–191
 - first-stage observation model, 188–189
 - LDA prior, 189–190
 - notation, 188
 - overview, 187–188
 - second-stage state model, 189
 - specification, 188–191
- related work, 203–204
- training and prediction, 191–198
 - fitting logistic regression, 197
 - model fitting, 192–197
 - Monte Carlo E-step, 193–195
 - M-step, 195–197
 - number of topics, 197
 - prediction, 197–198
 - regularizing regressions, 197
 - scalability, 197
- factors, defining, 33
- false positive, 11, 62*t*
- Fast Online Bilinear Factor Model (FOBFM)
 - illustration on Yahoo! data sets, 134–141
 - FOBFM, 135
 - FOBFM without offline bilinear terms, 140–141*f*
 - My Yahoo! data set, 135–137

Cambridge University Press

978-1-107-03607-9 - Statistical Methods for Recommender Systems

Deepak K. Agarwal and Bee-chung Chen

Index

[More information](#)

276

Index

- Fast Online Bilinear Factor Model (*cont.*)
 - no-init, 135
 - offline, 135
 - PCR, 135
 - PCR-B, 135
 - test set, 134–135
 - training set, 134
 - Yahoo! front page data set, 137–140
- personalization, 122–126, 124*t*
 - dimensionality reduction, 123
 - feature-based initialization, 123
 - Gaussian model for numeric response, 125
 - logistic model for binary response, 125
 - online model selection, 124
 - online regression + feature-based offset, 125–126
 - rapid online regression, 124
 - reduced rank regression, 126
 - response model, 124–125
- FEAT-ONLY (feature-only factorization model), 172–173
- feature-based method overview, 25–26
 - contextual information, 30–31
 - notation, 30–31
 - supervised methods, 27–30
 - binary ratings, 28
 - numeric ratings, 28
 - ordinal ratings, 28–29
 - pairwise preference, 29
 - regularized maximum likelihood estimation, 29–30
 - unsupervised methods, 26–27
- feature-based regression, personalization through, 122–126, 124*t*
- featured modules (FMs), 83–84
- features
 - collaborative filtering as, 37
 - defining, 33
- feature vectors, 15–16, 19*f*, 212
- FilterBot, 37, 164, 165*t*, 199*t*
- Fischer, P., 46–47, 108
- five-fold cross-validation, 164–165, 226
- fixed-reward settings, 43, 44
- fLDA, *see* factorization through latent Dirichlet allocation
- FN (false negative), 62*t*
- Freund, Y., 47, 109
- Frobenius norms, 20, 173
- Gamma-Poisson model, 88, 99, 107, 117–118, 240, 252, 253
- Gaussian response model, 28, 34, 126, 128, 129–130, 143, 144–145, 149, 151–157, 160, 161, 162
- generalized linear model framework, 51–52, 121
- general portals, 82
- Gibbs sampling, 22, 152–153, 155–156, 160, 168, 190, 193, 197, 198, 210, 214–215
- Gilks, Walter R., 158, 160
- Gittins, John C., 44
- Gittins index, 44
- gradient-boosted tree ensemble, 147
- Graepel, T., 147
- ground-truth distribution, in evaluation methods, 71–72
- ground-truth model, in evaluation methods, 63–64, 70–73
- Hadoop, 91, 180
- Herbrich, R., 147
- Herlocker, J. L., 37
- heuristic bandit scheme, 47–48
 - ϵ -greedy, 47, 49, 52, 53
 - k-deviation UCB, 48, 53
 - SoftMax, 47–48, 52, 53
 - Thompson sampling, 48, 52–53
- hierarchical clustering, 51
- hierarchical shrinkage, 211–218, 234–235
 - locally augmented tensor model, 216–218
- model fitting, 213–216
 - E-step, 214–215
 - M-step, 216
- modeling, 212–213
 - bias smoothed tensor, 212, 217, 226, 229–231
 - feature vectors, 15–16, 19*f*, 212
 - hierarchical regression prior, 213
 - item bias, 213
 - tensor factorization, 213
 - user bias, 212–213
- Hobert, James P., 151
- hybrid approach, 36–37
- hypothetical run, 108–109
- Iacovou, Neophytos, 32
- ICM (iterated conditional modes) algorithm, 152
- implicit ratings, 16, 65
- information retrieval (IR) systems, 65
- input signals
 - extent of social sharing, 5

Cambridge University Press

978-1-107-03607-9 - Statistical Methods for Recommender Systems

Deepak K. Agarwal and Bee-chung Chen

Index

[More information](#)

- item content and source, 4–5
- popularity indicators, 5
- user declared information, 5
- user interest profile, 5
- item-based splitting, 58
- item bias, 213
- item characterization, item feature
 - construction, 16–23
 - audio features, 23
 - bag-of-words, 18–21
 - image features, 23
 - location, 23
 - source, 23
 - topic modeling, 21–22
- item corpus, 18
- item feature vector, 16*f*
 - constructing, 18–21
- item index, 91
- item-item similarity, 32–33
- item pool, defining, 5
- item-related characteristics, 85
- item retrieval, 92
- item starvation problem, 11
- iterated conditional modes (ICM) algorithm, 152
- Jaakkola, Tommi S., 161
- Jaccard similarity, 26
- Jambor, Tamas, 261
- Johnson-Lindenstrauss lemma, 21
- Jones, David M., 44
- Jordan, Michael I., 21, 161
- $K \times 2$ case, Bayesian explore-exploit scheme, 102–104
 - convexity, 104
 - Lagrange relaxation, 103–104
 - near-optimal solution, 104
 - separability, 104
- k-deviation UCB, 48, 53
- Kendall's τ , 63–64
- keyword-based features, 24, 136
- k-means, 249, 250*t*, 251, 256, 260
- kMeans-QP-Dual, 250*t*, 254, 256
- kMeans-QP-Primal, 250*t*, 256
- Konstan, J. A., 37
- Kwon, Young, 261
- Lacerda, Anisio, 261
- Lagrangian duality, 245–248
- Laplace approximation, 128–129, 132
- large content pools, 117–118
- LASSO regression model, 147
- LAT, *see* locally augmented tensor (LAT) model
- latent Dirichlet allocation (LDA) topic model, 21–22
- latent factors, 33–34. *See also* factorization through latent Dirichlet allocation (fLDA); regression-based latent factor model (RLFM)
- L-BFGS, 30
- Liblinear, 226
- linear projection, 51–52
- LinkedIn, 11, 12, 76, 82*t*, 85*f*, 219
- LM (language model), 227, 228*f*
- localized multiobjective program (l-MOP), 242
- locally augmented tensor (LAT) model, 216–218
 - model fitting, 218
 - modeling, 217–218
- logistic models, 28, 128–129, 130, 145
- low-quality content, defining, 5
- MAB. *See* multiarmed bandit (MAB) problem
- machine learning, to define/flag low-quality content, 5
- MAE (mean absolute error), 60, 252
- Malkin, Robert G., 261
- Manouselis, Nikos, 261
- MAP (mean average precision), 64, 227
- MapReduce, 91, 169–170
- Markov decision process (MDP), 42
- matrix factorization, 33–36
 - optimization methods, 35–36
 - alternating least squares, 35
 - stochastic gradient descent, 35–36
 - regularization, 34–35
 - vs.* RLFM, 143–144
 - See also* factorization through latent Dirichlet allocation (fLDA)
- MCCEM, *see* Monte Carlo EM (MCEM) algorithm
- MCEM-ARS, 173
- MCEM-ARSID, 173
- MCEM-VAR, 173
- MDP (Markov decision process), 42
- mean absolute error (MAE), 60, 252
- mean average precision (MAP), 64, 227
- minimax approach, 46–47
- model store, 91
- Mohri, M., 109

Cambridge University Press

978-1-107-03607-9 - Statistical Methods for Recommender Systems

Deepak K. Agarwal and Bee-chung Chen

Index

[More information](#)

278

Index

- Monte Carlo EM (MCEM) algorithm, 151, 156, 186, 191–192, 197, 209–210, 213–214
- most-popular recommendation, 87–88, 94–119
 - Bayesian solution, 98–107
 - 2×2 case, 99–101, 112; gain function, 102; normal approximation, 101–102; optimal solution, 101
 - BayesGeneral scheme, 106–107, 112
 - Gamma-Poisson model, 99
 - general solution, 104–107; dynamic set of candidate items, 105–107; nonstationary CTR, 107; two-stage approximation, 105
 - $K \times 2$ case, 102–104; convexity, 104; Lagrange relaxation, 103–104; near-optimal solution, 104; separability, 104
 - one-step look-ahead, 99
- B-POKER, 109, 112
- B-UCB1, 108–109, 112
- Exp3, 47, 109, 110, 111f, 112
- large content pools, 117–118
- non-Bayesian solutions, 107–109, 112
- offline simulation, 71–72
- problem definition, 96–98
- Yahoo! Today empirical evaluation, 109–117
 - bucket test results, 116–117
 - comparative analysis, 110–112
 - scheme characterization, 112–114, 112–114f; EMP regret, 113–114f; fraction of EMP display, 113–114f; non-EMP regret, 113–114f
 - segmentation analysis, 114–115
- Yahoo! Today example, 94–119
 - batch serving, 96
 - dynamic set of items, 96
 - nonstationary CTR, 96
 - problem definition, 96–98; Bayes optimal scheme, 98; oracle optimal scheme, 98; regret, 98; serving scheme, 97–98
- MovieLens, 164–166, 165t, 172, 173–175, 174t, 198–199
- moving averages, 11
- MSN, 82
- multiarmed bandit (MAB) problem, 40–48
 - Bayesian approach, 42–46
 - optimization problem, 44–46
 - optimal policy, 43
 - reward, 43
 - state, 42
 - state transition, 43
- heuristic bandit schemes, 47–48
 - ε -greedy, 47, 49, 52, 53
 - k-deviation UCB, 48, 53
 - SoftMax, 47–48, 52, 53
 - Thompson sampling, 48, 52–53
- minimax approach, 46–47
- multiobjective program (MOP), 242
- multivariate regression, 147
- named entities, 19, 201
- nDCG (normalized discounted cumulative gain), 64
- nearest-neighbor model, 147
- neighborhood-based methods, 32, 147, 203
- Netflix, 60, 175, 198, 203, 204
- network-update modules (NMs), 84
- news article recommendation, multifaceted, 218–233
 - empirical evaluation, 226–233
 - bilinear model, 226
 - BM25, 227, 228
 - breakdown by facets, 230
 - breakdown by user activity level, 230–231
 - BST, 226
 - CMF, 226
 - COS, 227, 228
 - data, 227
 - evaluation metrics, 227
 - experimental setup, 227–228
 - IR model performance, 228
 - IR models used, 227
 - LAT, 226
 - LM, 227, 228
 - mean average precision, 227
 - overall performance, 228–230
 - paired t -test result, 229t
 - perceived differences among facets, 231–233
 - P@k, 227
 - ranking examples, 232t
 - SMF, 226
- exploratory data analysis, 219–226
 - article category/user segment variations, 225–226
 - article category variations, 224
 - correlations among post-read actions, 220–222f
 - data diversity, 220, 221f
 - data source, 219

Cambridge University Press

978-1-107-03607-9 - Statistical Methods for Recommender Systems

Deepak K. Agarwal and Bee-chung Chen

Index

[More information](#)

- post-read action rates at different resolutions, 224
 - pre-read vs. post-read, 220
 - read vs. post-read, 222–224, 223*f*
 - user segments variations, 224, 225*f*
- Ng, Andrew Y., 21
- nonbatch schemes, for explore-exploit. *See* B-UCB1; WTA-POKER; WTA-UCB1
- normalized discounted cumulative gain (nDCG), 64
- novelty effect, 67, 70, 168, 261
- numerical features, 24
- numeric ratings, 28
- observed ratings, 16, 27
 - offline evaluation, 56–58
 - types of, 28–30
- offline bilinear feature-based regression model, 123
- offline evaluation
 - accuracy metrics, 59–61
 - log-likelihood, 60
 - mean absolute error, 60
 - normalized L_p norm, 60
 - root mean squared error, 60
 - translating to real system improvements, 61
 - as unsuitable for measuring ranking, 60–61
- data-splitting methods, 56–59
 - cross-validation, 58
 - splitting methods, 57–59
 - tuning set, 58–59
- observed ratings, 56–58
- predeployment stage, 55
- ranking metrics, 61–65
 - area under the ROC curve (AUC), 62
 - global ranking metrics, 61–62
 - Kendall's τ , 63–64
 - local ranking metrics, 64, 65
 - precision-recall (P-R) curve, 62
 - receiver operating characteristic (ROC) curve, 62
 - Spearman's ρ , 63
- offline learner, 89
- offline models, 88
- offline replay, 73–77
 - basic replay estimator, 73–75
 - extensions of replay, 76–77
 - unbiased estimator of expected reward, 75–76
- offline simulation, 70–73
- Okapi BM25, 26, 227, 228
- OLR (online logistic regression) model, 251–252
- one-step look-ahead, 99
- online bucket tests, 8, 66–70
 - bias, 67
 - bucket setup, 66–67
 - request-based bucket, 66–67
 - user-based bucket, 66–67
 - online performance metrics, 67–69
 - actions beyond clicks, 68
 - average number of clicks per user, 68
 - click-through rate, 68
 - fraction of clickers, 68
 - time spent, 68–69
- test results analysis, 69–70
 - histograms of user attributes, 69
 - impression statistics, 69
 - metrics split by user segments, 69–70; by time, 70; by user activity levels, 70; by user attribute, 70
 - numbers of impressions/clicks/users over time, 69
 - user visits by frequency, 69
- online evaluation, postdeployment stage, 55. *See also* online bucket tests
- online learner, 89
- online logistic regression (OLR) model, 251–252
- online models, 88
- optimization
 - long-term objectives, 5
 - metrics to optimize, 7
 - near-term objectives, 5
 - standard methods for, 30, 35–36
 - See also* optimization, multiobjective
- optimization, multiobjective, 237–262
 - application setting, 238–239
 - segmented approach, 239–242
 - approximation methods, 248–250
 - clustering, 249–250; k -means, 249, 250*r*, 256, 260; Top1Item, 249, 250*r*, 254
 - sampling, 249–250; random, 249, 250*r*; stratified, 249
 - personalized approach, 238, 243–248
 - Lagrangian duality, 245–248; conversion algorithm, 247–248; dual serving plan, 246–247
 - primal formulation, 243–248; challenges, 244; key idea, 244–245; objectives, 243–244; personalized serving plan, 243; revised primal program, 245; serving plan, 243
 - related work, 261–262

Cambridge University Press

978-1-107-03607-9 - Statistical Methods for Recommender Systems

Deepak K. Agarwal and Bee-chung Chen

Index

[More information](#)

280

Index

- optimization, multiobjective (*cont.*)
 - segmented approach, 237–238, 239–242
 - objective optimization, 240–242
 - problem setup, 239–240
 - segmented serving scheme, 240
 - serving scheme, 240
 - Yahoo! front page experiment (*See* Yahoo! front page optimization experiment)
- oracle optimal scheme, 98
- ordinal ratings, 28–29
- overfitting, avoiding
 - algorithms, 152, 197
 - learning processes, 21
 - locally augmented tensor model, 217, 230
 - matrix factorization, 143
 - optimization metrics, 7
 - regression-based latent factor model, 144, 146–147
 - regularized maximum likelihood estimation, 30
 - resistance of sampling to, 192
- paired *t*-test, 229*t*
- pairwise preference, 29
- Pang, Bo, 210
- parallel matrix factorization, 170
- Pearson correlation, 32
- personalization through factor models, 142–182
 - cold start example, 164–167
 - EachMovie data, 165*t*–166
 - methods, 164
 - MovieLens data, 164–165*t*
 - Yahoo! front page data, 166–167
 - fitting algorithms, 150–164
 - number of Gibbs samples, 155–156
 - problem definition, 150–151
 - regularization in M-step, 155
 - remarks, 155–157
 - scalability, 156
 - shrinkage estimator for factors, 156–157
- large-scale problems, example, 172–182
 - large Yahoo! front page data, 177–180;
 - baseline methods, 178;
 - category-profile, 178; data, 177;
 - experimental results, 178–179;
 - experimental setup, 177; item-profile, 178; results, 180–182; unbiased evaluation, 177–178; variation approximation issue, 179–180
 - methods, 172–173; FEAT-ONLY, 172–173; MCEM-ARS, 173; MCEM-ARSID, 173; MCEM-VAR, 173; SGD, 173
 - MovieLens-1 M data, 173–175; data, 173–174; MCEM-ARS/MCEM-VAR comparison, 174; SGD comparison, 175
 - small Yahoo! front page data, 175–177;
 - comparison with SGD, 176; different partition methods, 176–177; number of partitions, 176; single-machine results, 175
- large-scale/time-sensitive items, 167–172
 - identifiability issues, 171–172
 - online models for items, 168–169
 - online models for users, 169
 - parallel fitting algorithm, 169–170
 - partitioning data, 170–171
 - periodic offline training, 168
- regression-based latent factor model, 142–150
 - matrix factorization vs. RLFM, 143–144
 - model specification, 144–149
 - stochastic process, 149–150
- personalization through feature-based regression, 120–141
 - expectation-maximization, 126–130
 - E-step, 127–130; Gaussian model, 128, 129–130; logistic model, 128–129, 130
 - M-step, 128
- Fast Online Bilinear Factor Model (FOBFM), 122–126, 124*t*
 - dimensionality reduction, 123
 - feature-based initialization, 123
 - online model selection, 124
 - rapid online regression, 124
 - response model, 124–125; Gaussian model for numeric response, 125; logistic model for binary response, 125
 - online regression + feature-based offset, 125–126
 - reduced rank regression, 126
- FOBFM on Yahoo! data sets, example, 134–141
- FOBFM, 135
- FOBFM without offline bilinear terms, 140–141*f*

Cambridge University Press

978-1-107-03607-9 - Statistical Methods for Recommender Systems

Deepak K. Agarwal and Bee-chung Chen

Index

[More information](#)

- My Yahoo! data set, 135–137;
 - effectiveness of online model selection, 137
- no-init, 135
- offline, 135
- PCR, 135
- PCR-B, 135
- test set, 134–135; log-likelihood, 135;
 - rank correlation, 135
- training set, 134
- Yahoo! front page data set, 137–140
- offline regression model, 121
- offline training, 126–131
 - scalability, 122, 130–131
- online learning, 131–134
 - explore-exploit themes, 132–133;
 - d -deviation UCB, 133; ϵ -greedy, 132–133; SoftMax, 133; Thompson sampling, 133
 - Gaussian model, 131
 - logistic model, 132
 - model selection, 133–134
- online regression model, 121
- personal portals, 82
- Pilász, István, 36
- POKER, 109
- policy, defining, 42
- positive explicit ratings, 5
- Posse, Christian, 261
- precision, mean average, 64, 227
- precision at rank K ($P@K$), 64, 277
- precision-recall (P-R) curve, 62
- primal plans, 253
- principal component analysis, 51–52
- problem settings, 81–89
 - application settings, 85–87
 - common recommendation modules, 82*t*–85
 - common statistical methods, 87–89
 - See also* system architecture
- product recommendation, 24
- profile modeling, 5–6*f*, 7
- pull model, 9–10
- push model, 9–10
- random bucket, 11, 13–14, 110, 177, 178, 251
- random forest, 147
- randomized serving scheme, 11
- random linear projection, 21
- random sampling
 - approximation methods for multiobjective optimization, 249, 250*t*
 - experiments of multiobjective optimization, 252, 253*t*, 254, 255–256*f*, 258–260, 259*f*
 - splitting methods, 57, 58
- random splitting, 57, 58
- ranking, 6–7, 9, 60–61, 93
- rating matrix, 16
- rating(s)
 - defining, 15–16
 - observed ratings, 16, 27, 56
 - accuracy metrics, 59–60, 61–62, 63
 - offline evaluation, 56–58
 - types of, 28–30
 - ordinal, 28–29
 - unseen, 56, 59
- recall, precision-recall (P-R) curve, 62
- receiver operating characteristic (ROC) curve, 62
- recommendation modules, common, 82*t*–85
 - domain-specific sites, 82
 - featured modules, 83–84
 - general portals, 82
 - network-update modules, 84
 - personal portals, 82
 - related-content modules, 85
 - social network sites, 82–83
- recommendation service, example system, 89, 92–93
- recommender systems
 - defining, 3
 - evaluation of, 8–9
 - example of typical, 3–4*f*
 - overview of, 4–10
- regression
 - multivariate, 147
 - offline bilinear feature-based, 123
 - online logistic regression, 251–252
 - supervised methods, 27–30
 - univariate, 147, 155
 - See also* personalization through feature-based regression
- regression-based latent factor model (RLFM), 142–150
 - cold start example, 164–167
 - EachMovie data, 165*t*–166
 - methods, 164
 - MovieLens data, 164–165*t*

Cambridge University Press

978-1-107-03607-9 - Statistical Methods for Recommender Systems

Deepak K. Agarwal and Bee-chung Chen

Index

[More information](#)

282

Index

- regression-based latent factor model (*cont.*)
 - Yahoo! front page data, 166
 - Yahoo! front page results, 166–167
- large-scale/time-sensitive items, 167–172
 - identifiability issues, 171–172
 - online models for items, 168–169
 - online models for users, 169
 - parallel fitting algorithm, 169–170
 - partitioning data, 170–171
 - periodic offline training, 168
- matrix factorization vs. RLFM, 143–144
- model specification, 144–149
 - factorization, 146
 - log likelihood, 148
 - model fitting and prediction, 148–149
 - overfitting problem, 146–147
 - regression functions, 147
 - regression priors, 147
 - response model, 144–145
- stochastic process of RLFM, 149–150
- regret, 44, 98, 113–114f
- regularized maximum likelihood estimation, 29–30
- rejection sampling (RS), 158
- related-content modules (RMs), 85
- Resnick, Paul, 32
- response rate, 39
- response-related characteristics, 87
- return/retention rates, optimizing for, 5
- revenue, optimizing for, 5
- Ribeiro, Marco T., 261
- Riedl, John, 32, 37
- RLFM, *see* regression-based latent factor model
- RMSE (root mean squared error), 60
- ROC (receiver operating characteristic) curve, 62
- Rodriguez, Mario, 261
- root mean squared error (RMSE), 60
- sampling
 - bootstrap, 67
 - Gibbs sampling, 22, 152–153, 155–156, 160, 168, 190, 193, 197, 198, 210, 214–215
 - stratified, 249
- scalability
 - factor models, 156, 197
 - MapReduce, 91, 169–170
 - regression, 122, 130–131
 - segmentation, 241–242
- Schapiro, R. E., 47, 109
- scoring, 6, 7–9, 10f–14, 12f
- Sculley, D., 261
- segmented approach, 49–50, 72–73, 114–115, 237–238, 240, 241–242
- segment-LP, 250t, 255f, 256f
- separate matrix factorization (SMF), 216–217
- serendipity, 7
- serving bucket, 11, 13–14
- serving scheme, 97–98
- SGD (stochastic gradient descent), 30, 35–36, 173, 176
- Shanmugasundaram, Jayavel, 261
- similarity
 - item-item, 32–33
 - user-user, 31–32
- similarity-based collaborative filtering, 37
- singular value decomposition (SVD), 20–21, 51–52
- SMF (separate matrix factorization), 216–217, 226
- social actions, optimizing for, 5
- social network sites, 82–83
 - See also* Facebook; LinkedIn; Twitter
- SoftMax, 47–48, 52, 53, 109, 112, 113–114f, 115
- sparseness, data, 14, 18–20, 50–54, 112, 144
- Spearman's ρ , 63
- starvation problem, 11
- state parameter, 42
- state transition rule, 43
- statistical methods, common, 87–89
 - context-dependent recommendation, 88–89
 - most-popular recommendation, 87–88
 - personalized recommendation, 88
- status-quo algorithm, 241
- Stern, D. H., 147
- stochastic gradient descent (SGD), 30, 35–36, 173, 176
- storage systems, 91
- stratified-kMeans, 250t, 255f, 259f, 260
- stratified sampling, 249
- stratified-Top1Item, 250t, 255f, 259f, 260
- subscriptions, optimizing for, 5
- Suchak, Mitesh, 32
- SVD (singular value decomposition), 20–21, 51–52
- Svore, Krysta M., 261
- synonym expansion, 21
- system architecture, 89–93
 - example, 89–93, 90f

Cambridge University Press

978-1-107-03607-9 - Statistical Methods for Recommender Systems

Deepak K. Agarwal and Bee-chung Chen

Index

[More information](#)

- application setting, 90
 - model, 90–91
 - offline learner, 91
 - offline/online learner synchronization, 92
 - online learner, 92
 - recommendation service, 92–93; item retrieval, 92; ranking, 93
 - storage systems, 91
 - main components, 89
 - See also* problem settings
- Tan, K. K. C., 160
- tensor models
 - bias smoothed, 212, 217, 226, 229–231
 - factorization models, 213, 234
 - locally augmented, 216–218
- TF-IDF (term frequency-inverse document frequency), 18
- TF (term frequency) version, 18
- Thompson, William R., 41
- Thompson sampling, 48, 52–53, 133
- three-fold cross-validation, 203
- Tikk, Domonkos, 36
- time-based splitting, 57–58
- time series methods, 11, 13, 69, 71, 96, 97, 107
- time spent on site, optimizing for, 5, 7
- $TN(\theta)$, 62*t*
- TopItem, 249, 250*t*, 254
- TopItem-QP-Dual, 250*t*, 254
- $TP(\theta)$, 62*t*
- transition probability, 43
- treatment bucket, 55
- trend effects, 70
- tri-grams, 19
- t*-test, 229*t*
- tuning parameters
 - Bayes 2×2 scheme, 107
 - describing, 58–59
 - empirical evaluation, 110–111
 - large content pools, 118
 - large scale problems, 173, 175
 - non-Bayesian solutions, 109
 - regularization, 30, 35
- Twitter, 12, 82*t*–83, 219
- 2×2 case, Bayesian explore-exploit scheme, 99–101, 107, 112
- normal approximation, 101–102
- optimal solution, 101
- properties of gain function, 102
- two-fold cross-validation, 252
- UCB1, 46–47, 109. *See also* B-UCB1; WTA-UCB1
- univariate regression, 147, 155
- unobserved ratings, 33
- unseen ratings, 56, 59
- unsupervised methods, 26–27
- user-based splitting, 58
- user bias, 212–213
- user characterization, 23–31
 - content-based profile, 24–25
 - current location, 25
 - declared profile, 23–24
 - demographics, 5, 23–24
 - interests, 24
 - item set, 25
 - recommendation application, 85–86*t*
 - search history, 25
 - usage-based features, 25
- user data store, 91
- user feature vector, 16*f*
- user item, computing score for, 15–16*f*
- user profile modeling, 5–6*f*, 7
- user-user similarity, 31–32
- U.S. Yahoo! News. *See* news article recommendation, multifaceted
- variational approximation, 22, 152, 192, 197
 - expectation-maximization, 161–164
 - E-step, 162–163
 - M-step, 163–164
 - large-scale problem, 172–175
 - potential issue in, 179–180
- Vassilvitskii, Sergei, 261
- vector space, 17–21, 26, 33, 146, 227
- Vee, Erik, 261
- Veloso, A., 261
- Vermorel, J., 109
- Volkovs, Maksims N., 261
- Wang, Jun, 261
- Wang, Xuanhui, 242, 250*t*, 251, 252, 253, 261
- warm start, 36, 37, 142, 175, 178–179*t*, 199, 204
- weighting, in user-user similarity methods, 32
- Whittle, P., 44

- WordNet, 20
- WTA-POKER, 109, 110–111*f*, 112
 - EMP regret vs. fraction of EMP, 113, 114*f*
- WTA-UCB1, 109, 110–111*f*
 - EMP regret vs. fraction of EMP, 113, 114*f*
- Yahoo! Buzz, 199–201
- Yahoo! front page data
 - personalization through factor models
 - FOBFM, 134–141
 - large-scale problems, 177–180
 - small-scale problems, 175–177
 - simple scoring model, 10*f*–14, 12*f*
- Yahoo! front page optimization experiment, 250–261
 - results, 252–261
 - advantage of personalized approach, 252–254
 - constraint satisfaction, 254–255
 - impact of y , 256–257
 - relaxation of per-property constraints, 257–258
 - sampling method comparison, 258–260
 - top N properties, 260–261
 - tradeoff comparison, 255–256
 - setup, 250–252
 - data, 250–251
 - estimation of p_{uj} and d_{uj} , 251–252
 - metrics, 251
- Yahoo! News. *See* news article
- recommendation, multifaceted
- Yahoo! Today module, 83
- empirical evaluation, 109–117
 - bucket test results, 116–117
 - comparative analysis, 110–112;
 - hypothetical scenario, 112; results
 - summary, 112; Yahoo! scenario, 110–111
- experimental setup, 116–117; Bayes
 - 2×2 , 116*f*–117*f*; B-UCBI, 116*f*, 117*f*; ϵ -greedy, 116*f*, 117*f*
- scheme characterization, 112–114*f*; EMP
 - regret, 113–114*f*; fraction of EMP
 - display, 113–114*f*; non-EMP regret, 113–114*f*
- segmentation analysis, 114–115
- example application, 95–96
 - batch serving, 96
 - dynamic set of items, 96
 - nonstationary CTR, 96
 - problem definition, 96–98
- Zhang, Ethan, 261
- Zhang, Liang, 147
- Ziviani, N., 261