# Contents

1 An Introduction to Recommender Systems .......................... 1
   1.1 Introduction ................................... 1
   1.2 Goals of Recommender Systems ........................ 3
      1.2.1 The Spectrum of Recommendation Applications ............ 7
   1.3 Basic Models of Recommender Systems .................... 8
      1.3.1 Collaborative Filtering Models ........................ 8
         1.3.1.1 Types of Ratings ............................ 10
         1.3.1.2 Relationship with Missing Value Analysis ........... 13
         1.3.1.3 Collaborative Filtering as a Generalization of Classification and Regression Modeling .................. 13
      1.3.2 Content-Based Recommender Systems .................. 14
      1.3.3 Knowledge-Based Recommender Systems ............... 15
         1.3.3.1 Utility-Based Recommender Systems ............... 18
      1.3.4 Demographic Recommender Systems .................. 19
      1.3.5 Hybrid and Ensemble-Based Recommender Systems ..... 19
      1.3.6 Evaluation of Recommender Systems ................ 20
   1.4 Domain-Specific Challenges in Recommender Systems .......... 20
      1.4.1 Context-Based Recommender Systems .................. 20
      1.4.2 Time-Sensitive Recommender Systems ................ 21
      1.4.3 Location-Based Recommender Systems ................ 21
      1.4.4 Social Recommender Systems ........................ 22
         1.4.4.1 Structural Recommendation of Nodes and Links .......... 22
         1.4.4.2 Product and Content Recommendations with Social Influence ........................................ 23
         1.4.4.3 Trustworthy Recommender Systems ................ 23
         1.4.4.4 Leveraging Social Tagging Feedback for Recommendations ........................................ 23
   1.5 Advanced Topics and Applications ........................... 23
      1.5.1 The Cold-Start Problem in Recommender Systems ........ 24
      1.5.2 Attack-Resistant Recommender Systems ............... 24
      1.5.3 Group Recommender Systems ........................ 24
1.5.4 Multi-Criteria Recommender Systems ........................................... 24
1.5.5 Active Learning in Recommender Systems ........................................ 25
1.5.6 Privacy in Recommender Systems .................................................. 25
1.5.7 Application Domains ........................................................................ 26
1.6 Summary ............................................................................................... 26
1.7 Bibliographic Notes ............................................................................... 26
1.8 Exercises ............................................................................................... 28

2 Neighborhood-Based Collaborative Filtering ............................................. 29
2.1 Introduction ............................................................................................. 29
2.2 Key Properties of Ratings Matrices ......................................................... 31
2.3 Predicting Ratings with Neighborhood-Based Methods ......................... 33
  2.3.1 User-Based Neighborhood Models .................................................... 34
    2.3.1.1 Similarity Function Variants .................................................... 37
    2.3.1.2 Variants of the Prediction Function ....................................... 38
    2.3.1.3 Variations in Filtering Peer Groups ....................................... 39
    2.3.1.4 Impact of the Long Tail ....................................................... 39
  2.3.2 Item-Based Neighborhood Models .................................................... 40
  2.3.3 Efficient Implementation and Computational Complexity ................... 41
  2.3.4 Comparing User-Based and Item-Based Methods ............................... 42
  2.3.5 Strengths and Weaknesses of Neighborhood-Based Methods ............... 44
  2.3.6 A Unified View of User-Based and Item-Based Methods .................... 44
2.4 Clustering and Neighborhood-Based Methods .......................................... 45
2.5 Dimensionality Reduction and Neighborhood Methods .......................... 47
  2.5.1 Handling Problems with Bias .......................................................... 49
    2.5.1.1 Maximum Likelihood Estimation ......................................... 49
    2.5.1.2 Direct Matrix Factorization of Incomplete Data .................... 50
2.6 A Regression Modeling View of Neighborhood Methods .......................... 51
  2.6.1 User-Based Nearest Neighbor Regression ....................................... 53
    2.6.1.1 Sparsity and Bias Issues ...................................................... 54
  2.6.2 Item-Based Nearest Neighbor Regression ....................................... 55
  2.6.3 Combining User-Based and Item-Based Methods ............................... 57
  2.6.4 Joint Interpolation with Similarity Weighting .................................... 57
  2.6.5 Sparse Linear Models (SLIM) .......................................................... 58
2.7 Graph Models for Neighborhood-Based Methods .................................... 60
  2.7.1 User-Item Graphs ............................................................................ 61
    2.7.1.1 Defining Neighborhoods with Random Walks ....................... 61
    2.7.1.2 Defining Neighborhoods with the Katz Measure ..................... 62
  2.7.2 User-User Graphs ............................................................................ 63
  2.7.3 Item-Item Graphs ............................................................................ 66
2.8 Summary ............................................................................................... 67
2.9 Bibliographic Notes ............................................................................... 67
2.10 Exercises ............................................................................................. 69

3 Model-Based Collaborative Filtering ......................................................... 71
3.1 Introduction ............................................................................................. 71
3.2 Decision and Regression Trees .............................................................. 74
  3.2.1 Extending Decision Trees to Collaborative Filtering ......................... 76
### 3.3 Rule-Based Collaborative Filtering
- 3.3.1 Leveraging Association Rules for Collaborative Filtering
- 3.3.2 Item-Wise Models versus User-Wise Models

### 3.4 Naive Bayes Collaborative Filtering
- 3.4.1 Handling Overfitting
- 3.4.2 Example of the Bayes Method with Binary Ratings

### 3.5 Using an Arbitrary Classification Model as a Black-Box
- 3.5.1 Example: Using a Neural Network as a Black-Box

### 3.6 Latent Factor Models
- 3.6.1 Geometric Intuition for Latent Factor Models
- 3.6.2 Low-Rank Intuition for Latent Factor Models
- 3.6.3 Basic Matrix Factorization Principles
- 3.6.4 Unconstrained Matrix Factorization
  - 3.6.4.1 Stochastic Gradient Descent
  - 3.6.4.2 Regularization
  - 3.6.4.3 Incremental Latent Component Training
  - 3.6.4.4 Alternating Least Squares and Coordinate Descent
  - 3.6.4.5 Incorporating User and Item Biases
  - 3.6.4.6 Incorporating Implicit Feedback
- 3.6.5 Singular Value Decomposition
  - 3.6.5.1 A Simple Iterative Approach to SVD
  - 3.6.5.2 An Optimization-Based Approach
  - 3.6.5.3 Out-of-Sample Recommendations
  - 3.6.5.4 Example of Singular Value Decomposition
- 3.6.6 Non-negative Matrix Factorization
  - 3.6.6.1 Interpretability Advantages
  - Observations about Factorization with Implicit Feedback
- 3.6.7 Understanding the Matrix Factorization Family

### 3.7 Integrating Factorization and Neighborhood Models
- 3.7.1 Baseline Estimator: A Non-Personalized Bias-Centric Model
- 3.7.2 Neighborhood Portion of Model
- 3.7.3 Latent Factor Portion of Model
- 3.7.4 Integrating the Neighborhood and Latent Factor Portions
- 3.7.5 Solving the Optimization Model
- 3.7.6 Observations about Accuracy
- 3.7.7 Integrating Latent Factor Models with Arbitrary Models

### 3.8 Summary

### 3.9 Bibliographic Notes

### 3.10 Exercises

### 4 Content-Based Recommender Systems
- 4.1 Introduction
- 4.2 Basic Components of Content-Based Systems
- 4.3 Preprocessing and Feature Extraction
  - 4.3.1 Feature Extraction
    - 4.3.1.1 Example of Product Recommendation
# 6 Ensemble-Based and Hybrid Recommender Systems

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Introduction</td>
<td>199</td>
</tr>
<tr>
<td>6.2 Ensemble Methods from the Classification Perspective</td>
<td>204</td>
</tr>
<tr>
<td>6.3 Weighted Hybrids</td>
<td>206</td>
</tr>
<tr>
<td>6.3.1 Various Types of Model Combinations</td>
<td>208</td>
</tr>
<tr>
<td>6.3.2 Adapting Bagging from Classification</td>
<td>209</td>
</tr>
<tr>
<td>6.3.3 Randomness Injection</td>
<td>211</td>
</tr>
<tr>
<td>6.4 Switching Hybrids</td>
<td>211</td>
</tr>
<tr>
<td>6.4.1 Switching Mechanisms for Cold-Start Issues</td>
<td>212</td>
</tr>
<tr>
<td>6.4.2 Bucket-of-Models</td>
<td>212</td>
</tr>
<tr>
<td>6.5 Cascade Hybrids</td>
<td>213</td>
</tr>
<tr>
<td>6.5.1 Successive Refinement of Recommendations</td>
<td>213</td>
</tr>
<tr>
<td>6.5.2 Boosting</td>
<td>213</td>
</tr>
<tr>
<td>6.5.2.1 Weighted Base Models</td>
<td>214</td>
</tr>
<tr>
<td>6.6 Feature Augmentation Hybrids</td>
<td>215</td>
</tr>
<tr>
<td>6.7 Meta-Level Hybrids</td>
<td>216</td>
</tr>
<tr>
<td>6.8 Feature Combination Hybrids</td>
<td>217</td>
</tr>
<tr>
<td>6.8.1 Regression and Matrix Factorization</td>
<td>218</td>
</tr>
<tr>
<td>6.8.2 Meta-level Features</td>
<td>218</td>
</tr>
<tr>
<td>6.9 Mixed Hybrids</td>
<td>220</td>
</tr>
<tr>
<td>6.10 Summary</td>
<td>221</td>
</tr>
<tr>
<td>6.11 Bibliographic Notes</td>
<td>222</td>
</tr>
<tr>
<td>6.12 Exercises</td>
<td>224</td>
</tr>
</tbody>
</table>

# 7 Evaluating Recommender Systems

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Introduction</td>
<td>225</td>
</tr>
<tr>
<td>7.2 Evaluation Paradigms</td>
<td>227</td>
</tr>
<tr>
<td>7.2.1 User Studies</td>
<td>227</td>
</tr>
<tr>
<td>7.2.2 Online Evaluation</td>
<td>227</td>
</tr>
<tr>
<td>7.2.3 Offline Evaluation with Historical Data Sets</td>
<td>229</td>
</tr>
<tr>
<td>7.3 General Goals of Evaluation Design</td>
<td>229</td>
</tr>
<tr>
<td>7.3.1 Accuracy</td>
<td>229</td>
</tr>
<tr>
<td>7.3.2 Coverage</td>
<td>231</td>
</tr>
<tr>
<td>7.3.3 Confidence and Trust</td>
<td>232</td>
</tr>
<tr>
<td>7.3.4 Novelty</td>
<td>233</td>
</tr>
<tr>
<td>7.3.5 Serendipity</td>
<td>233</td>
</tr>
<tr>
<td>7.3.6 Diversity</td>
<td>234</td>
</tr>
<tr>
<td>7.3.7 Robustness and Stability</td>
<td>235</td>
</tr>
<tr>
<td>7.3.8 Scalability</td>
<td>235</td>
</tr>
<tr>
<td>7.4 Design Issues in Offline Recommender Evaluation</td>
<td>235</td>
</tr>
<tr>
<td>7.4.1 Case Study of the Netflix Prize Data Set</td>
<td>236</td>
</tr>
<tr>
<td>7.4.2 Segmenting the Ratings for Training and Testing</td>
<td>238</td>
</tr>
<tr>
<td>7.4.2.1 Hold-Out</td>
<td>238</td>
</tr>
<tr>
<td>7.4.2.2 Cross-Validation</td>
<td>239</td>
</tr>
<tr>
<td>7.4.3 Comparison with Classification Design</td>
<td>239</td>
</tr>
<tr>
<td>7.5 Accuracy Metrics in Offline Evaluation</td>
<td>240</td>
</tr>
<tr>
<td>7.5.1 Measuring the Accuracy of Ratings Prediction</td>
<td>240</td>
</tr>
<tr>
<td>7.5.1.1 RMSE versus MAE</td>
<td>241</td>
</tr>
<tr>
<td>7.5.1.2 Impact of the Long Tail</td>
<td>241</td>
</tr>
</tbody>
</table>
10 Structural Recommendations in Networks

10.1 Introduction ................................... 309
10.2 Ranking Algorithms .............................................. 311
10.2.1 PageRank ................................ 311
10.2.2 Personalized PageRank ......................... 314
10.2.3 Applications to Neighborhood-Based Methods ............ 316
10.2.3.1 Social Network Recommendations ...................... 317
10.2.3.2 Personalization in Heterogeneous Social Media ....... 317
10.2.3.3 Traditional Collaborative Filtering ..................... 319
10.2.4 SimRank .............................................. 321
10.2.5 The Relationship Between Search and Recommendation .... 322
10.3 Recommendations by Collective Classification .............. 323
10.3.1 Iterative Classification Algorithm ................... 324
10.3.2 Label Propagation with Random Walks ................ 325
10.3.3 Applicability to Collaborative Filtering in Social Networks ..... 326
10.4 Recommending Friends: Link Prediction ................. 326
10.4.1 Neighborhood-Based Measures ..................... 327
10.4.2 Katz Measure ........................................ 328
10.4.3 Random Walk-Based Measures . ................... 329
10.4.4 Link Prediction as a Classification Problem ............. 329
10.4.5 Matrix Factorization for Link Prediction ................. 330
10.4.5.1 Symmetric Matrix Factorization .................. 333
10.4.6 Connections Between Link Prediction and Collaborative Filtering 335
10.4.6.1 Using Link Prediction Algorithms for Collaborative Filtering 336
10.4.6.2 Using Collaborative Filtering Algorithms for Link Prediction 337
10.5 Social Influence Analysis and Viral Marketing ............... 337
10.5.1 Linear Threshold Model ........................ 339
10.5.2 Independent Cascade Model ...................... 340
10.5.3 Influence Function Evaluation ..................... 340
10.5.4 Targeted Influence Analysis Models in Social Streams .... 341
10.6 Summary ..................................... 342
10.7 Bibliographic Notes .................................... 343
10.8 Exercises ..................................... 344

11 Social and Trust-Centric Recommender Systems 345

11.1 Introduction .............................................. 345
11.2 Multidimensional Models for Social Context .................. 347
11.3 Network-Centric and Trust-Centric Methods ................ 349
11.3.1 Collecting Data for Building Trust Networks ............. 349
11.3.2 Trust Propagation and Aggregation ................... 351
11.3.3 Simple Recommender with No Trust Propagation ......... 353
11.3.4 TidalTrust Algorithm .......................... 353
11.3.5 MoleTrust Algorithm ........................................ 356
11.3.6 TrustWalker Algorithm ................................... 357
11.3.7 Link Prediction Methods .................................. 358
11.3.8 Matrix Factorization Methods .............................. 361
  11.3.8.1 Enhancements with Logistic Function ............... 364
  11.3.8.2 Variations in the Social Trust Component ........... 364
11.3.9 Merits of Social Recommender Systems .................... 365
  11.3.9.1 Recommendations for Controversial Users and Items . 365
  11.3.9.2 Usefulness for Cold-Start .......................... 366
  11.3.9.3 Attack Resistance .................................. 366
11.4 User Interaction in Social Recommenders ..................... 366
  11.4.1 Representing Folksonomies ............................ 367
  11.4.2 Collaborative Filtering in Social Tagging Systems ...... 368
  11.4.3 Selecting Valuable Tags ................................ 371
11.4.4 Social-Tagging Recommenders with No Ratings Matrix .... 372
  11.4.4.1 Multidimensional Methods for Context-Sensitive Systems 372
  11.4.4.2 Ranking-Based Methods ........................... 373
  11.4.4.3 Content-Based Methods ............................ 374
11.4.5 Social-Tagging Recommenders with Ratings Matrix .............................. 377
  11.4.5.1 Neighborhood-Based Approach ......................... 378
  11.4.5.2 Linear Regression ................................ 379
  11.4.5.3 Matrix Factorization ................................ 380
  11.4.5.4 Content-Based Methods ............................ 382
11.5 Summary .......................................................... 382
11.6 Bibliographic Notes ........................................... 382
11.7 Exercises .......................................................... 384

12 Attack-Resistant Recommender Systems .......................... 385
12.1 Introduction ...................................................... 385
12.2 Understanding the Trade-Offs in Attack Models ............... 386
  12.2.1 Quantifying Attack Impact ............................... 390
12.3 Types of Attacks .................................................. 392
  12.3.1 Random Attack ........................................... 393
  12.3.2 Average Attack .......................................... 393
  12.3.3 Bandwagon Attack ....................................... 394
  12.3.4 Popular Attack ........................................... 395
  12.3.5 Love/Hate Attack ......................................... 395
  12.3.6 Reverse Bandwagon Attack ................................ 396
  12.3.7 Probe Attack ............................................... 396
  12.3.8 Segment Attack ........................................... 396
  12.3.9 Effect of Base Recommendation Algorithm .................. 397
12.4 Detecting Attacks on Recommender Systems ...................... 398
  12.4.1 Individual Attack Profile Detection ...................... 399
  12.4.2 Group Attack Profile Detection ......................... 402
    12.4.2.1 Preprocessing Methods ............................ 402
    12.4.2.2 Online Methods .................................... 403
12.5 Strategies for Robust Recommender Design ..................... 403
  12.5.1 Preventing Automated Attacks with CAPTCHAs .............. 403
  12.5.2 Using Social Trust ....................................... 404
### Contents

12.5.3 Designing Robust Recommendation Algorithms .................................................. 404
12.5.3.1 Incorporating Clustering in Neighborhood Methods ........................................... 405
12.5.3.2 Fake Profile Detection during Recommendation Time ......................................... 405
12.5.3.3 Association-Based Algorithms ............................................................................ 405
12.5.3.4 Robust Matrix Factorization ............................................................................... 405
12.6 Summary .................................................................................................................... 408
12.7 Bibliographic Notes .................................................................................................. 408
12.8 Exercises ................................................................................................................... 410

13 Advanced Topics in Recommender Systems .................................................................. 411
13.1 Introduction ................................................................................................................ 411
13.2 Learning to Rank ........................................................................................................ 413
13.2.1 Pairwise Rank Learning ......................................................................................... 415
13.2.2 Listwise Rank Learning ......................................................................................... 416
13.2.3 Comparison with Rank-Learning Methods in Other Domains .............................. 417
13.3 Multi-Armed Bandit Algorithms ................................................................................ 418
13.3.1 Naive Algorithm .................................................................................................... 419
13.3.2 $\epsilon$-Greedy Algorithm ...................................................................................... 420
13.3.3 Upper Bounding Methods ..................................................................................... 421
13.4 Group Recommender Systems .................................................................................. 423
13.4.1 Collaborative and Content-Based Systems ............................................................ 424
13.4.2 Knowledge-Based Systems .................................................................................... 425
13.5 Multi-Criteria Recommender Systems ...................................................................... 426
13.5.1 Neighborhood-Based Methods .............................................................................. 427
13.5.2 Ensemble-Based Methods ...................................................................................... 428
13.5.3 Multi-Criteria Systems without Overall Ratings .................................................... 429
13.6 Active Learning in Recommender Systems .................................................................. 430
13.6.1 Heterogeneity-Based Models .................................................................................. 431
13.6.2 Performance-Based Models ................................................................................... 432
13.7 Privacy in Recommender Systems ............................................................................ 432
13.7.1 Condensation-Based Privacy .................................................................................. 434
13.7.2 Challenges for High-Dimensional Data .................................................................. 434
13.8 Some Interesting Application Domains ....................................................................... 435
13.8.1 Portal Content Personalization .............................................................................. 435
13.8.1.1 Dynamic Profiler ............................................................................................... 436
13.8.1.2 Google News Personalization ............................................................................. 436
13.8.2 Computational Advertising versus Recommender Systems .................................. 438
13.8.2.1 Importance of Multi-Armed Bandit Methods ...................................................... 442
13.8.3 Reciprocal Recommender Systems ....................................................................... 443
13.8.3.1 Leveraging Hybrid Methods .............................................................................. 444
13.8.3.2 Leveraging Link Prediction Methods .................................................................. 445
13.9 Summary .................................................................................................................... 446
13.10 Bibliographic Notes .................................................................................................. 446

Bibliography ....................................................................................................................... 449

Index .................................................................................................................................. 493
Recommender Systems
The Textbook
Aggarwal, C.C.
2016, XXI, 498 p. 79 illus., 18 illus. in color., Hardcover
ISBN: 978-3-319-29657-9