Contents

Data Trading: Theory and Technology

English translation of the original table of contents

Part I. Introduction to Data Trading 1

Chapter 1. Introduction to Data Trading 3
1.1 Factors of Production 3
1.1.1 Traditional Factors of Production 4
1.1.2 Data as a Factor of Production and Its Characteristics 8
1.2 Significance of Data Trading 11
1.2.1 The Necessity of Data Trading 12
1.2.2 Economic Significance of Data Trading: Comparative Advantage and Specialization 13
1.2.3 Strategic Significance of Data Trading 19
1.3 Fundamental Concepts of Data Trading 22
1.3.1 Definition of Data Trading 22
1.3.2 Participants in Data Trading 23
1.3.3 Objects of Data Trading 24
1.3.4 Forms of Data Trading 26
1.3.5 Data Trading Framework 27
1.4 Core Requirements for Data Trading 29
1.4.1 Property Rights Issues 30
1.4.2 Pricing Issues 30
1.4.3 Fairness Issues 32
1.4.4 Security Issues 33
1.4.5 Book Structure and Technical Roadmap 33

Part II. Theoretical Foundations 39

Chapter 2. Fundamentals of Game Theory 41
2.1 Basic Concepts of Game Theory 41
2.1.1 Introduction: Some Examples 41
2.1.2 Basic Definition of a Game 44
2.2 Static Games of Complete Information and Basic Solution Concepts 46
2.2.1 Dominant Strategies 46
2.2.2 Nash Equilibrium 49
2.2.3 Mixed Strategies 54
2.3 Dynamic Games 59
2.3.1 Extensive-Form Games 59
2.3.2 Subgame-Perfect Equilibrium 60
2.4 Games of Incomplete Information 65
2.4.1 Basic Model of Games of Incomplete Information 65
2.4.2 Equilibria in Games of Incomplete Information 67
2.5 Fundamentals of Cooperative Game Theory 69
2.5.1 From Noncooperative to Cooperative Games 69
2.5.2 The Core 72
2.5.3 The Shapley Value 74

Chapter 3. Fundamentals of Auctions and Mechanism Design 85
3.1 Fundamentals of Auction Theory 85
3.1.1 Ways of Selling an Item 85
3.1.2 Common Auction Formats 85
3.1.3 Basic Theory of Single-Item Sealed-Bid Auctions 87
3.1.4 Overview of Other Topics in Auction Theory 94
3.2 Basic Concepts of Mechanism Design 95
3.2.1 General Framework of Mechanism Design 95
3.2.2 The Revelation Principle 96

3.3 Design of Welfare-Maximizing Auction Mechanisms . . . 101
3.3.1 The VCG Mechanism . . . 101
3.3.2 Properties of the VCG Mechanism . . . 104
3.4 Optimal Auction Mechanism Design . . . 105
3.4.1 Virtual Welfare Maximization . . . 105
3.4.2 Derivation of the Optimal Mechanism . . . 109

Chapter 4. Selected Topics in Games and Information . . . 119
4.1 Bayesian Persuasion . . . 119
4.1.1 Bayesian Persuasion: An Example . . . 119
4.1.2 General Model and Problem Reformulation . . . 121
4.1.3 Optimal Signaling Mechanism . . . 125
4.1.4 Receiver's Utility in Bayesian Persuasion . . . 129
4.2 Information Elicitation Mechanisms . . . 130
4.2.1 Fundamentals of Information Theory . . . 130
4.2.2 Scoring Rules . . . 134
4.2.3 Peer Prediction . . . 138

Chapter 5. Fundamentals of Multi-Armed Bandit Algorithms . . . 153
5.1 The Multi-Armed Bandit Problem . . . 153
5.2 Stochastic Multi-Armed Bandits . . . 155
5.2.1 Problem Description . . . 155
5.2.2 An Initial Attempt . . . 156
5.2.3 Upper Confidence Bound Algorithm . . . 159
5.2.4 Thompson Sampling Algorithm . . . 161
5.3 Adversarial Multi-Armed Bandits . . . 162
5.3.1 Problem Description . . . 162
5.3.2 Full-Information Feedback and the Multiplicative Weights Algorithm . . . 163
5.3.3 Adversarial Multi-Armed Bandits and the Exp3 Algorithm . . . 166

Part III. Building Data Trading Platforms 173

Chapter 6. Data Acquisition and Governance . . . 175
6.1 Data Acquisition for Data Management Tasks . . . 175
6.1.1 Data Acquisition . . . 175
6.1.2 Data Labeling . . . 177
6.1.3 Existing Data . . . 178
6.2 Data Selection for Machine Learning Tasks . . . 180
6.2.1 Heuristic Data Selection Methods . . . 181
6.2.2 Principled Data Selection Methods . . . 183
6.2.3 A Unified Framework and Limitations of Data Selection Methods . . . 185
6.2.4 Section Summary . . . 187
6.3 Data Acquisition with Strategic Data Providers . . . 187
6.3.1 Mechanism Design for Crowdsourced Data Labeling . . . 187
6.3.2 Mechanism Design for Truthful Data Labeling . . . 191
6.3.3 Mechanism Design for Truthful Dataset Acquisition . . . 200

Chapter 7. Data Security and Compliance . . . 209
7.1 Zero-Knowledge Proofs . . . 209
7.2 Data Watermarking . . . 210
7.3 Federated Learning . . . 212
7.4 Secure Multiparty Computation . . . 213
7.5 Data Masking and Anonymization . . . 214
7.6 Differential Privacy . . . 215
7.6.1 Definition of Differential Privacy . . . 215
7.6.2 The Laplace Mechanism . . . 216
7.6.3 The Exponential Mechanism . . . 219
7.6.4 Properties of Differential Privacy . . . 221
7.6.5 Local Differential Privacy . . . 223
7.7 Secure Signaling Mechanisms . . . 224
7.8 Protection Mechanisms Against Strategic Buyers . . . 226

Part IV. Data Pricing and Valuation **233**

Chapter 8. Fundamental Data Pricing Strategies, Part I . . . 235
8.1 Real-World Data Pricing Strategies as a Starting Point . . . 235
8.1.1 Basic Methods of Data Pricing . . . 235
8.1.2 Fundamental Data Pricing Strategies . . . 236
8.1.3 Insights from Real-World Pricing Strategies . . . 238
8.2 Version-Based Pricing of Database Queries . . . 239
8.2.1 Basic Ideas and Concepts . . . 239
8.2.2 DPS Arbitrage-Free Pricing Function . . . 242
8.2.3 APS Arbitrage-Free Pricing Function . . . 248
8.2.4 QPS Arbitrage-Free Pricing Function . . . 255
8.2.5 Efficient Computation of Pricing Functions . . . 260
8.3 Version-Based Pricing of Machine Learning Models . . . 268
8.3.1 Versioning Methods . . . 268
8.3.2 Arbitrage-Free Conditions . . . 271
8.3.3 Profit Maximization . . . 274

Chapter 9. Fundamental Data Pricing Strategies, Part II . . . 285
9.1 Auction-Based Data Pricing . . . 285
9.1.1 Auctions for Machine Learning Models . . . 285
9.1.2 Prior-Free Mechanism Design for Unlimited-Supply Auctions . . . 287
9.1.3 Differential Privacy and Mechanism Design . . . 294
9.2 Dynamic Data Pricing Strategies . . . 297
9.2.1 Multi-Armed Bandit-Based Data Pricing Strategies . . . 297
9.2.2 Ellipsoid Method-Based Data Pricing Strategies . . . 305
9.3 Free-Trial Strategies for Data . . . 314
9.3.1 Two Basic Models . . . 314
9.3.2 Uncertainty in the Utilities of Data Buyers and Sellers . . . 315

Chapter 10. Data Pricing from a Decision-Making Perspective . . . 329
10.1 The Value of Data for Decision-Making . . . 329
10.1.1 Introduction: A Simple Model . . . 329
10.1.2 A More Complex Model . . . 332
10.1.3 Implications and Discussion . . . 332
10.2 Optimal One-Round Selling Mechanism Design . . . 334
10.2.1 Basic Model . . . 334
10.2.2 Properties of the Optimal Menu . . . 336
10.2.3 Solution for Binary Actions and States . . . 341
10.3 Optimal Multi-Round Selling Mechanism Design . . . 352
10.3.1 Basic Model . . . 352
10.3.2 Mechanism Design with Independent Signals . . . 356
10.3.3 Mechanism Design with Dependent Signals . . . 362

Chapter 11. Data Pricing from a Privacy Perspective . . . 375
11.1 Measuring Privacy Costs . . . 375
11.1.1 A Utility Interpretation of Differential Privacy . . . 375
11.1.2 Other Measures of Privacy Costs . . . 376
11.2 Mechanisms for Acquiring Privacy-Sensitive Data . . . 378
11.2.1 Auction Mechanisms . . . 378
11.2.2 Problems with Auction Mechanisms . . . 383
11.2.3 Mechanisms with a Known Prior Distribution of Privacy Costs . . . 384
11.3 Welfare Effects of Privacy-Sensitive Data Trading . . . 390
11.3.1 Basic Model . . . 390
11.3.2 Extreme Markets . . . 395
11.3.3 Welfare Bounds . . . 397
11.4 Policies on Data Privacy . . . 399
11.4.1 Theoretical Research Perspective . . . 399
11.4.2 Empirical Research Perspective . . . 404

Chapter 12. Externalities in Data Pricing . . . 415
12.1 Data Externalities: The Buyer's Perspective . . . 415
12.1.1 Externality Model . . . 415

12.1.2 Posted-Price Data Markets . . . 418
12.1.3 Auction-Based Data Markets . . . 420
12.1.4 Data Externalities in Stock Markets . . . 427
12.2 Data Externalities: The Seller's Perspective . . . 437
12.2.1 Basic Model . . . 437
12.2.2 Existence and Properties of Equilibrium . . . 439
12.2.3 Welfare Analysis of Equilibrium . . . 442
12.2.4 Solutions . . . 445

Chapter 13. Data Valuation . . . 451
13.1 Fundamentals of Data Valuation . . . 451
13.1.1 Basic Framework for Data Valuation . . . 451
13.1.2 Leave-One-Out Method . . . 452
13.1.3 Semivalues . . . 453
13.2 Semivalue-Based Data Valuation Methods . . . 457
13.2.1 Data-Shapley . . . 457
13.2.2 Beta-Shapley . . . 458
13.2.3 Data-Banzhaf . . . 460
13.3 Computational Challenges in Data Valuation . . . 464
13.3.1 Computational Complexity . . . 464
13.3.2 Polynomial-Time Exact Solutions for Special Cases: KNN-Shapley . . . 468
13.3.3 Monte Carlo Algorithms . . . 470
13.3.4 Complementary Contribution Sampling . . . 473
13.3.5 KernelSHAP: A Regression-Based Approximation Method . . . 475
13.4 Data Manipulation Issues . . . 477

Appendix . . . 481
A.1 Fundamentals of Microeconomics . . . 481
A.1.1 Consumer Choice Theory . . . 481
A.1.2 Competitive Equilibrium and Welfare Analysis . . . 486
A.1.3 Welfare Analysis of Competitive Equilibrium . . . 487
A.1.4 Market Failures . . . 490
A.2 Integrated Implementation of Data Trading . . . 496
A.2.1 Data Acquisition Module . . . 496
A.2.2 Data Security Module . . . 496
A.2.3 Data Pricing Module . . . 496

Afterword . . . 499

Page references correspond to the original Chinese-language book.