

LIST OF TABLES

Chapter 1

A COMMON LOGIC APPROACH TO DATA MINING AND PATTERN RECOGNITION, by A. Zakrevskij

		1
<i>Table 1.</i>	The Dependency of E on r Under Fixed n and m	14
<i>Table 2.</i>	The Dependency of the Maximum Rank r_{max} on the Parameters n and m	15
<i>Table 3.</i>	Finding All the Occurring Pairs of the Attribute Values Generated by the Element 01001	17
<i>Table 4.</i>	Finding All the Occurring Pairs of the Attribute Values Generated by the Selection F	17
<i>Table 5.</i>	Forecasting the Value of the Attribute x_i	29

Chapter 2

THE ONE CLAUSE AT A TIME (OCAT) APPROACH TO DATA MINING AND

KNOWLEDGE DISCOVERY, by E. Triantaphyllou

		45
<i>Table 1.</i>	Continuous Observations for Illustrative Example	54
<i>Table 2(a).</i>	The Binary Representation of the Observations in the Illustrative Example (first set of attributes for each example)	56
<i>Table 2(a).</i>	The Binary Representation of the Observations in the Illustrative Example (second set of attributes for each example)	57

Chapter 3

AN INCREMENTAL LEARNING ALGORITHM FOR INFERRING LOGICAL RULES FROM EXAMPLES IN THE FRAMEWORK OF THE COMMON REASONING PROCESS, by X. Naidenova

		89
<i>Table 1.</i>	Example 1 of Data Classification	106
<i>Table 2.</i>	Structure of the Data	108
<i>Table 3.</i>	The Results of the Procedure DEBUT for the Examples of Class 2	120
<i>Table 4.</i>	The Result of Inferring GMRT's for the Examples of Class 2	121
<i>Table 5.</i>	The Number of Combinations C_n^2 , C_n^3 , C_n^4 , as a Function of N	121
<i>Table 6.</i>	The Intersections of Example t_2 with the Examples of Class 2	123

Table 7.	The Projection of the Example t_2 on the Examples of Class 2	123
Table 8.	The Result of Reducing the Projection after Deleting the Values 'Brown' and 'Embrown'	124
Table 9.	Example 2 of a Data Classification	125
Table 10.	The Projection of the Value 'Tall' on the Set $R(+)$	126
Table 11.	The Projection of the Value 'Tall' on the Set $R(+)$ without the Values 'Bleu' and 'Brown'	126
Table 12.	The Projection of the Value 'Tall' on the Set $R(+)$ without the Examples t_5 and t_7	127
Table 13.	The Result of Deleting the Value 'Tall' from the Set $R(+)$	127
Table 14.	The Result of Deleting t_5 , t_7 , and t_8 , from the Set $R(+)$	127
Table 15.	The Essential Values for the Examples t_5 , t_6 , t_7 , and t_8	128
Table 16.	The Data for Processing by the Incremental Procedure INGOMAR	136
Table 17.	The Records of the Step-by-Step Results of the Incremental Procedure INGOMAR	137
Table 18.	The Sets TGOOD (1) and TGOOD (2) Produced by the Procedure INGOMAR	137
Table 19.	The Set of the Positive Examples $R(+)$	139
Table 20.	The Set of the Negative Examples $R(-)$	139
Table 21.	The content of $S(\text{test})$ after the DEBUT of the Algorithm NIAGaRa	140
Table 22.	The Contents of the set $STGOOD$ after the DEBUT of the Algorithm NIAGaRa	140
Table 23.	The Set Q after the DEBUT of the Algorithm NIAGaRa	141
Table 24.	The Extensions of the Elements of $S(\text{test})$	141
Table 25.	The Sets $STGOOD$ and $TGOOD$ for the Examples in Tables 19 and 20	142
Table 26.	The Set $SPLUS$ of the Collections $splus(A)$ for all A's in Tables 19 and 20	142

Chapter 4

DISCOVERING RULES THAT GOVERN MONOTONE

PHENOMENA, by V.I. Torvik and E. Triantaphyllou 149

Table 1.	History of Monotone Boolean Function Enumeration	159
Table 2.	A Sample Data Set for Problem 3	170
Table 3.	Example Likelihood Values for All Functions in M_3	171
Table 4.	Updated Likelihood Ratios for $m_z(001) = m_z(001) + 1$	173
Table 5.	The Representative Functions Used in the Simulations of Problem 3	179
Table 6.	The Average Number of Stage 3 Queries Used by the	

Evaluative Criterion $\max \Delta \lambda (v)$ to Reach $\lambda > 0.99$ in Problem 3 Defined on $\{0,1\}$ with Fixed Misclassification Probability q	182
---	-----

Chapter 5

LEARNING LOGIC FORMULAS AND RELATED

ERROR DISTRIBUTIONS, by G. Felici, F. Sun, and K. Truemper 193

Table 1. Estimated $\hat{F}_A(z)$ and $\hat{G}_B(z)$	215
--	-----

Chapter 6

FEATURE SELECTION FOR DATA MINING

by V. de Angelis, G. Felici, and G. Mancinelli 227

Table 1. Functions Used to Compute the Target Variable	242
Table 2. Results for Increasing Values of γ	243
Table 3. Results for Different Random Seeds for Classification Function A	244
Table 4. Results for Larger Instances for Classification Function A	244
Table 5. Results for Classification Functions B, C, and D	244
Table 6. Performances with Duplicated Features on Classification Function A	245
Table 7. Solution Times for Different Size Instances and Parameters for Classification Function A	245
Table 8. Solution Times for Different Size Instances and Parameters for Classification Functions D, E, F	246
Table 9. Logic Variables Selected by FSM3-B with $k_1=5$, $k_2=20$ and $\gamma=0$	247
Table 9. Logic Variables Selected by FSM3-B with $k_1=10$, $k_2=20$ and $\gamma=2.00$	247

Chapter 7

TRANSFORMATION OF RATIONAL AND SET DATA TO

LOGIC DATA, by S. Bartnikowski, M. Granberry, J. Mugan, and K. Truemper 253

Table 1. σ as a Function of N for $\sigma \leq 10$ and $p = q = 0.5$	269
Table 2. Performance of Cut Point vs. Entropy	275

Chapter 8

DATA FARMING: CONCEPTS AND METHODS, by A. Kusiak 279

Chapter 9

RULE INDUCTION THROUGH DISCRETE SUPPORT

VECTOR DECISION TREES, by C. Orsenigo and C. Vercellis	305
<i>Table 1.</i> Accuracy Results – Comparison among FDSDT _{SLP} and Alternative Classifiers	320
<i>Table 2.</i> Accuracy Results – Comparison among FDSDT _{SLP} and its Variants	322
<i>Table 3.</i> Rule Complexity – Comparison among Alternative Classifiers	323

Chapter 10

MULTI-ATTRIBUTE DECISION TREES AND DECISION RULES, by J.-Y. Lee and S. Olafsson	327
<i>Table 1.</i> A Simple Classification Problem	344

Chapter 11

KNOWLEDGE ACQUISITION AND UNCERTAINTY IN FAULT DIAGNOSIS: A ROUGH SETS PERSPECTIVE, by L.-Y. Zhai, L.-P. Khoo, and S.-C. Fok	359
<i>Table 1.</i> Information Table with Inconsistent Data	367
<i>Table 2.</i> Information Table with Missing Data	367
<i>Table 3.</i> A Comparison of the Four Approaches	373
<i>Table 4.</i> A Typical Information System	378
<i>Table 5.</i> Machine Condition and Its Parameters	385
<i>Table 6.</i> Machine Condition after Transformation	385
<i>Table 7.</i> Rules Induced by ID3 and the RClass System	386
<i>Table 8.</i> Process Quality and Its Parameters	386
<i>Table 9.</i> Process Quality (after Transformations)	387
<i>Table 10.</i> Rules Introduced by ID3 and the RClass System for the Second Illustrative Example	388

Chapter 12

DISCOVERING KNOWLEDGE NUGGETS WITH A GENETIC ALGORITHM, by E. Noda and A.A. Freitas	395
<i>Table 1.</i> Accuracy Rate (%) in the Zoo Data Set	420
<i>Table 2.</i> Accuracy Rate (%) in the Car Evaluation Data Set	420
<i>Table 3.</i> Accuracy Rate (%) in the Auto Imports Data Set	421
<i>Table 4.</i> Accuracy Rate (%) in the Nursery Data Set	422
<i>Table 5.</i> Rule Interestingness (%) in the Zoo Data Set	424
<i>Table 6.</i> Rule Interestingness (%) in the Car Evaluation Data Set	424
<i>Table 7.</i> Rule Interestingness (%) in the Auto Imports Data Set	425
<i>Table 8.</i> Rule Interestingness (%) in the Nursery Data Set	425
<i>Table 9.</i> Summary of the Results	427

Chapter 13**DIVERSITY MECHANISMS IN PITT-STYLE****EVOLUTIONARY CLASSIFIER SYSTEMS, by M. Kirley,****H.A. Abbass and R.I. McKay****433**

<i>Table 1.</i>	Results for the Five Data Sets-Percentage and Standard Deviations for Accuracy, Coverage and Diversity from the Stratified Ten- Fold Cross-Validation Runs Using Island Model and Fitness Sharing	450
<i>Table 2.</i>	Results for the Five Data Sets -Percentage and Standard Deviations for Accuracy, Coverage and Diversity from the Stratified Ten-Fold Cross-Validation Runs Using Island Model Without Island Model	451
<i>Table 3.</i>	Results for the Five Data Sets-Percentage and Standard Deviations for Accuracy, Coverage and Diversity from the Stratified Ten-Fold Cross-Validation Runs Using Fitness Sharing without Island Model	451
<i>Table 4.</i>	Results for the Five Data Sets-Percentage and Standard Deviations for Accuracy, Coverage and Diversity from the Stratified Ten-Fold Cross-Validation Runs without Fitness Sharing or Island Model	452

Chapter 14**FUZZY LOGIC IN DISCOVERING ASSOCIATION****RULES: AN OVERVIEW, by G. Chen, Q. Wei and E.E. Kerre****459**

<i>Table 1.</i>	Example of a Transaction Dataset T and a Binary Database D	466
<i>Table 2.</i>	Database D with Continuous Domains	467
<i>Table 3.</i>	Database D' Transformed from D by Partitioning Domains	467
<i>Table 4.</i>	Database D'' (in part) with Fuzzy Items	468
<i>Table 5.</i>	Example of Extended Database D_G in Accordance with G in Figure 2 (a)	471
<i>Table 6.</i>	Example of Extended Database D_{Gf} in Accordance with G_f in Figure 2(b)	472

Chapter 15**MINING HUMAN INTERPRETABLE KNOWLEDGE WITH FUZZY MODELING METHODS: AN OVERVIEW,****by T.W. Liao****495**

Chapter 16

DATA MINING FROM MULTIMEDIA PATIENT RECORDS,

by A.S. Elmaghraby, M.M. Kantardzic, and M.P. Wachowiak 551

Table 1. Integrated Computerized Medical Record Characteristics 562

Chapter 17

LEARNING TO FIND CONTEXT BASED SPELLING

ERRORS, by H. Al-Mubaid and K. Truemper 597

Table 1. Text Statistics 609

Table 2. Learning Cases 610

Table 3. Large Testing Text Cases 611

Table 4. Error Detection for Large Testing Texts 611

Table 5. Small Testing Texts 612

Table 6. Small Testing Text Cases 612

Table 7. Error Detection for Small Testing Texts 613

Table 8. Performance of Ltest Compared with BaySpell
and WinSpell 614

Chapter 18

INDUCTION AND INFERENCE WITH FUZZY RULES

FOR TEXTUAL INFORMATION RETRIEVAL, by J. Chen,

D.H. Kraft, M.J. Martin-Bautista, and M. –A., Vila 629

Table 1. Portions of the EDC Database Used in
this Study 636

Table 2. The Prediction of Interestingness of Unseen Web Pages 645

Table 3. The Precisions of Queries vs. Number of Top Web Pages 645

Chapter 19

STATISTICAL RULE INDUCTION IN THE PRESENCE OF PRIOR INFORMATION: THE BAYESIAN RECORD

LINKAGE PROBLEM, by D.H. Judson 655

Table 1. An Illustration of the Comparison Between
Two Records 662

Table 2. Parameters for the Simulated Data 670

Table 3. Three Fictional Address Parsings, and the
Comparison Vector Between Record One and
Record Two 672

Table 4. Parameters, Computer Notation, and Their
Interpretation for the Simulated Data 673

Table 5. Results from the MCMC Estimation of Posterior
Distributions of Simulated Parameters 675

Table 6. Estimated Posterior Probability that the Records
are a Match, for All Possible Field Configurations and

	the Estimated Logistic Regression Parameters- Relatively Uninformative Priors Condition	676
<i>Table 7.</i>	Results from the MCMC Estimation of Posterior Distributions of Simulated Parameters	677
<i>Table 8.</i>	Estimated Posterior Probability that the Records are a Match, for All Possible Field Configurations and the Estimated Logistic Regression Parameters Informative Priors Condition	678
<i>Table 9.</i>	Associated Matching Fields, Parameters, Computer Notation and their Interpretation for CPS Address Data	682
<i>Table 10.</i>	Results From the MCMC Estimation of Posterior Distributions of CPS Address Field Parameters	684
<i>Table 11.</i>	Posterior Median Estimates Converted to Approximate Probabilities	685
<i>Table 12.</i>	Posterior Probability Calculations for all Obtained Comparison Vectors	687

Chapter 20

**FUTURE TRENDS IN SOME DATA MINING AREAS,
by X. Wang, P. Zhu, G. Felici, and E. Triantaphyllou**

695

