

DATA MINING IN FINANCE

Advances in Relational and
Hybrid Methods

by

BORIS KOVALERCHUK

Central Washington University, USA

and

EVGENII VITYAEV

Institute of Mathematics
Russian Academy of Sciences, Russia

KLUWER ACADEMIC PUBLISHERS

Boston/ Dordrecht/London

TABLE OF CONTENTS

Foreword by Gregory Piatetsky-Shapiro	xi
Preface	xiii
Acknowledgements	xv
1. The Scope and Methods of the Study	
1.1 Introduction	1
1.2 Problem definition	3
1.3 Data mining methodologies	4
1.3.1 Parameters	4
1.3.2 Problem ID and profile	6
1.3.3 Comparison of intelligent decision support methods.	7
1.4 Modern methodologies in financial knowledge discovery	9
1.4.1 Deterministic dynamic system approach	9
1.4.2 Efficient market theory	10
1.4.3 Fundamental and technical analyses	11
1.5 Data mining and database management	12
1.6 Data mining: definitions and practice	14
1.7 Learning paradigms for data mining	17
1.8 Intellectual challenges in data mining	19
2. Numerical Data Mining Models with Financial Applications	
2.1. Statistical, autoregression models	21
2.1.1. ARIMA models	22
2.1.2. Steps in developing ARIMA model	25
2.1.3. Seasonal ARIMA	27
2.1.4. Exponential smoothing and trading day regression	28
2.1.5. Comparison with other methods.	28
2.2. Financial applications of autoregression models	30
2.3. Instance-based learning and financial applications	32
2.4. Neural networks	36
2.4.1. Introduction	36
2.4.2. Steps	38
2.4.3. Recurrent networks	39
2.4.4. Dynamically modifying network structure	40
2.5. Neural networks and hybrid systems in finance	40
2.6. Recurrent neural networks in finance	42
2.7. Modular networks and genetic algorithms	44
2.7.1. Mixture of neural networks	44
2.7.2. Genetic algorithms for modular neural networks	45
2.8. Testing results and the complete round robin method	47
2.8.1. Introduction	47
2.8.2. Approach and method	47
2.8.3. Multithreaded implementation	52
2.8.4. Experiments with SP500 and neural networks	54
2.9. Expert mining	58
2.10. Interactive learning of monotone Boolean functions	66
2.10.1. Basic definitions and results	66

2.10.2.	Algorithm for restoring a monotone Boolean function	67
2.10.3.	Construction of Hansel chains	69
3.	Rule-Based and Hybrid Financial Data Mining	
3.1.	Decision tree and DNF learning	71
3.1.1.	Advantages	71
3.1.2.	Limitation: size of the tree	72
3.1.3.	Constructing decision trees	81
3.1.4.	Ensembles and hybrid methods for decision trees	84
3.1.5.	Discussion	87
3.2.	Decision tree and DNF learning in finance	88
3.2.1.	Decision-tree methods in finance	88
3.2.2.	Extracting decision tree and sets of rules for SP500	89
3.2.3.	Sets of decision trees and DNF learning in finance	93
3.3.	Extracting decision trees from neural networks	95
3.3.1.	Approach	95
3.3.2.	Trepan algorithm	96
3.4.	Extracting decision trees from neural networks in finance	97
3.4.1.	Predicting the Dollar-Mark exchange rate	97
3.4.2.	Comparison of performance	99
3.5.	Probabilistic rules and knowledge-based stochastic modeling	102
3.5.1.	Probabilistic networks and probabilistic rules	103
3.5.2.	The naïve Bayes classifier	106
3.5.3.	The mixture of experts	107
3.5.4.	The hidden Markov model	108
3.5.5.	Uncertainty of the structure of stochastic models	111
3.6.	Knowledge-based stochastic modeling in finance	112
3.6.1.	Markov chains in finance	112
3.6.2.	Hidden Markov models in finance	114
4.	Relational Data Mining (RDM)	
4.1.	Introduction	115
4.2.	Examples	118
4.3.	Relational data mining paradigm	123
4.4.	Challenges and obstacles in relational data mining	127
4.5.	Theory of RDM	129
4.5.1.	Data types in relational data mining	129
4.5.2.	Relational representation of examples.	130
4.5.3.	First-order logic and rules	135
4.6.	Background knowledge	140
4.6.1.	Arguments constraints and skipping useless hypotheses	140
4.6.2.	Initial rules and improving search of hypotheses	141
4.6.3.	Relational data mining and relational databases	144
4.7.	Algorithms: FOIL and FOCL	146
4.7.1.	Introduction	146
4.7.2.	FOIL	147
4.7.3.	FOCL	150
4.8.	Algorithm MMDR	151
4.8.1.	Approach	151
4.8.2.	MMDR algorithm and existence theorem	154
4.8.3.	Fisher test	159

4.8.4	MMDR pseudocode	162
4.8.5	Comparison of FOIL and MMDR	165
4.9	Numerical relational data mining	166
4.10	Data types	169
4.10.1	Problem of data types	169
4.10.2	Numerical data type	174
4.10.3	Representative measurement theory	174
4.10.4	Critical analysis of data types in ABL	175
4.11	Empirical axiomatic theories: empirical contents of data	179
4.11.1	Definitions.	179
4.11.2	Representation of data types in empirical axiomatic theories	181
4.11.3	Discovering empirical regularities as universal formulas	186
5	Financial Applications of Relational Data Mining	
5.1.	Introduction	189
5.2.	Transforming numeric data into relations	191
5.3.	Hypotheses and probabilistic "laws"	193
5.4.	Markov chains as probabilistic "laws" in finance	196
5.5.	Learning	199
5.6.	Method of forecasting	202
5.7.	Experiment 1	204
5.7.1.	Forecasting Performance for hypotheses H1-H4	204
5.7.2.	Forecasting performance for a specific regularity	207
5.7.3.	Forecasting performance for Markovian expressions	209
5.8.	Experiment 2	212
5.9.	Interval stock forecast for portfolio selection	213
5.10.	Predicate invention for financial applications: calendar effects	215
5.11.	Conclusion	218
6	Comparison of Performance of RDM and other methods in financial applications	
6.1.	Forecasting methods	219
6.2.	Approach: measures of performance	220
6.3.	Experiment 1: simulated trading performance	222
6.4.	Experiment 1: comparison with ARIMA	225
6.5.	Experiment 2: forecast and simulated gain	227
6.6.	Experiment 2: analysis of performance	227
6.7.	Conclusion	229
7.	Fuzzy logic approach and its financial applications	
7.1.	Knowledge discovery and fuzzy logic	231
7.2.	"Human logic" and mathematical principles of uncertainty	235
7.3.	Difference between fuzzy logic and probability theory	239
7.4.	Basic concepts of fuzzy logic	240
7.5.	Inference problems and solutions	248
7.6.	Constructing coordinated contextual linguistic variables	252
7.6.1.	Examples	252
7.6.2.	Context space	259
7.6.3.	Acquisition of fuzzy sets and membership function	262
7.6.4.	Obtaining linguistic variables	265
7.7.	Constructing coordinated fuzzy inference	266

7.7.1. Approach	266
7.7.2. Example	268
7.7.3. Advantages of "exact complete" context for fuzzy inference	270
7.8. Fuzzy logic in finance	278
7.8.1. Review of applications of fuzzy logic in finance	278
7.8.2. Fuzzy logic and technical analysis	281
REFERENCES	285
Subject Index	299