Lecture Notes in Artificial Intelligence

Subseries of Lecture Notes in Computer Science Edited by J. G. Carbonell and J. Siekmann

Lecture Notes in Computer Science Edited by G. Goos, J. Hartmanis, and J. van Leeuwen

Springer Berlin

Berlin Heidelberg New York Barcelona Hong Kong London Milan Paris Tokyo James J. Alpigini James F. Peters Andrzej Skowron Ning Zhong (Eds.)

Rough Sets and Current Trends in Computing

Third International Conference, RSCTC 2002 Malvern, PA, USA, October 14-16, 2002 Proceedings



Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

James J. Alpigini Penn State Great Valley School of Professional Graduate Studies 30 East Swedesford Road, Malvern, PA 19355, USA E-mail: jja7@psu.edu James F. Peters University of Manitoba, Department of Electrical and Computer Engineering Winnipeg, MB R3T 5V6, Canada E-mail: jfpeters@ee.umanitoba.ca

Andrzej Skowron Warsaw University, Institute of Mathematics Banacha 2, 02-097 Warsaw, Poland E-mail: skowron@mimuw.edu.pl

Ning Zhong Maebashi Institute of Technology Department of Systems and Information Engineering 460-1 Kamisadori-Cho, Maebashi-City 371-0816, Japan E-mail: zhong@maebashi-it.ac.jp

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Rough sets and current trends in computing : third international conference ; proceedings / RSCTC 2002, Malvern, PA, USA, October 14 - 16, 2002. James J. Alpigini ... (ed.). - Berlin ; Heidelberg ; New York ; Hong Kong ; London ; Milan ; Paris ; Tokyo : Springer, 2002

(Lecture notes in computer science ; Vol. 2475 : Lecture notes in artificial intelligence) ISBN 3-540-44274-X

CR Subject Classification (1998): I.2, F.4.1, F.1, I.5.1, I.4, H.2.8

ISSN 0302-9743 ISBN 3-540-44274-X Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag Berlin Heidelberg New York, a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

© Springer-Verlag Berlin Heidelberg 2002 Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Stefan Sossna e.K. Printed on acid-free paper SPIN: 10871364 06/3142 5 4 3 2 1 0

Preface

This volume contains the papers selected for presentation at the Third International Conference on Rough Sets and Current Trends in Computing (RSCTC 2002) held at Penn State Great Valley, Malvern, Pennsylvania, U.S.A., 14–16 October 2002. Rough set theory and its applications constitute a branch of soft computing that has exhibited a significant growth rate during recent years. RSCTC 2002 provided a forum for exchanging ideas among many researchers in the rough set community and in various areas of soft computing and served as a stimulus for mutual understanding and cooperation. In recent years, there have been a number of advances in rough set theory and applications. Hence, we have witnessed a growing number of international workshops on rough sets and their applications. In addition, it should be observed that one of the beauties of rough sets and the rough set philosophy is that it tends to complement and reinforce research in many traditional research areas and applications. This is the main reason that many international conferences are now including rough sets into the list of topics.

It is our great pleasure to dedicate this volume to Professor Zdzisław Pawlak, who created rough set theory over twenty years ago. The growth of rough set theory and applications owes a great deal to Professor Pawlak's vibrant enthusiasm and wit as well as his great generosity towards others, especially in encouraging and advising beginners in rough sets. The depth, breadth, and richness of current rough set research are directly traceable to Professor Pawlak's inventiveness and the richness of his many insights and ideas concerning data mining, machine learning, logic, and mathematics. The computational features of rough sets are also giving rise to new forms of neurocomputing based on rough sets and to a family known as rough processors for digital computers. We would also like to congratulate Professor Pawlak, who received an honorary doctorate (Doctor Honoris Causa) from Poznań Polytechnic University, Poznań, Poland on 10 April 2002.

We wish to express our gratitude to Professors Zdzisław Pawlak and Lotfi A. Zadeh, who accepted our invitation to serve as honorary chairs and to present keynote papers for this conference. We also wish to thank Professors J. Komorowski, T.Y. Lin, D.W. Russell, R. Slowiński, and I.B. Türksen for accepting our invitation to be plenary speakers at RSCTC 2002.

The papers contributed to this volume reflect advances in rough sets as well as complementary research efforts in the following areas:

- Rough set foundations
- Rough sets and fuzzy sets
- Rough neurocomputing
- Rough sets and probabilistic reasoning
- Rough set methods

- Rough set applications in biology, classification, dynamical systems, image processing, medical diagnosis, musicology, neurology, pattern recognition, robotics and robotic control systems, signal analysis, software engineering, and web mining
- Computing with words and granular computing
- Machine learning and pattern recognition
- Data mining

We wish to express our thanks to the members of the Advisory Board: N. Cercone, J. Grzymała-Busse, T.Y. Lin, A. Nakamura, S.K. Pal, L. Polkowski, R. Slowiński, H. Tanaka, S. Tsumoto, Y.Y. Yao, and W. Ziarko for their contribution to the scientific program of this conference. We also wish to thank the local committee for their help in organizing this conference: P. McFadden, C. Neill, K. Patel, F. Ramsey, and S.S. Slish.

The accepted papers that appear in this volume were selected from over 100 submitted draft papers. These papers were divided into regular communications (each allotted 8 pages) and short communications (each allotted 4 pages) on the basis of reviewer evaluations. Most papers received three or more reviews. The reviewing process itself rested with the RSCTC 2002 Program Chairs, members of the RSCTC 2002 Advisory Board, and the following members of the Program Committee: P. Apostoli, M. Beynon, H.D. Burkhard, G. Cattaneo, J.S. Deogun, P. Doherty, D. Dubois, I. Duentsch, S. Greco, X. Hu, M. Inuiguchi, J. Järvinen, J. Komorowski, B. Kostek, J. Koronacki, M. Kryszkiewicz, C.-J. Liau, P. Lingras, B. Matarazzo, E. Menasalvas, Z. Michalewicz, R. Michalski, N. Michinori, S. Miyamoto, M. Moshkov, T. Murai, H.S. Nguyen, E. Orłowska, W. Pedrycz, M. Quafafou, S. Ramanna, Z. Raś, J. Stefanowski, J. Stepaniuk, Z. Suraj, A. Szałas, M. Szczuka, A. Wakulicz-Deja, G. Wang.

We also would like to acknowledge help in reviewing from P. Balbiani, J. Bazan, K. Dembczynski, D. Niwiński, G. Góra, Y. Kawahara, R. Latkowski, J. Małuszyński, H. Midelfart, J.M. Peña, A. Radzikowska, D. Ślezak, P. Synak, S. Wilk, A. Wojna, J. Wróblewski, M. Zawadowski.

Special thanks are also extended to Marcin Szczuka, Jakub Wróblewski, and Dominik Śleżak for their help in typesetting and preparing the RSCTC 2002 proceedings.

Our special thanks go to all individuals who submitted valuable papers for the RSCTC 2002 conference and to all conference participants.

We also wish to express our thanks to Alfred Hofmann at Springer-Verlag for his support and cooperation.

October 2002

James Alpigini James F. Peters Andrzej Skowron Ning Zhong

RSCTC 2002 Conference Committee

Honorary Chairs:	Zdzisław Pawlak, Lotfi A. Zadeh
Conference Chair:	James Alpigini
Program Chairs:	James F. Peters, Andrzej Skowron, Ning Zhong

Advisory Board

Nick Cercone Jerzy Grzymała-Busse Tsau Young Lin Akira Nakamura Sankar K. Pal Lech Polkowski Roman Slowiński Hideo Tanaka Shusaku Tsumoto Yiyu Yao Wojciech Ziarko

Local Committee

Patty McFadden	Kalpesh Patel	Sally Sue Slish
Colin Neill	Fred Ramsey	

Program Committee

Peter Apostoli Ali Arsanjani Malcolm Beynon Hans Dieter Burkhard Mihir Kr. Chakraborty Liu Chunnian Andrzej Czyżewski Jitender S. Deogun Patrick Doherty Dieder Dubois Ivo Duentsch Maria C. Fernandez Fernando Gomide Salvatore Greco Xiaohua Hu Masahiro Inuiguchi Jouni Järvinen Janusz Kacprzyk Daijin Kim Jan Komorowski

Bożena Kostek Jacek Koronacki Marzena Kryszkiewicz Churn-Jung Liau Pawan Lingras Jiming Liu Qing Liu Solomon Marcus Benedetto Matarazzo Ernestina Menasalvas **Z**bigniew Michalewicz Ryszard Michalski Nakata Michinori Sadaaki Miyamoto Mikhail Moshkov Tetsuva Murai Nguyen Hung Son Ewa Orłowska Witold Pedrycz Mohamed Quafafou

Sheela Ramanna Zbigniew W. Raś Ron Shapira Jerzy Stefanowski Jarosław Stepaniuk Zbigniew Suraj Roman Świniarski Andrzei Szałas Marcin Szczuka Francis E.H. Tay Helmut Thiele Mihaela Ulieru Alicja Wakulicz-Deja Guoyin Wang Anita Wasilewska Michael Wong Xindong Wu Hussein Zedan

Table of Contents

Keynote Papers

In Pursuit of Patterns in Data Reasoning from Data – The Rough Set Way Zdzisław Pawlak	1
Toward a Theory of Hierarchical Definability (THD) (Causality Is Undefinable) Lotfi A. Zadeh	10
Plenary Papers	
Modelling Biological Phenomena with Rough Sets Jan Komorowski	13
Database Mining on Derived Attributes (Granular and Rough Computing Approach) <i>Tsau Young Lin</i>	14
A Proposed Evolutionary, Self-Organizing Automaton for the Control of Dynamic Systems David W. Russell	33
Rough Set Analysis of Preference-Ordered Data Roman Słowiński, Salvatore Greco, Benedetto Matarazzo	44
Fuzzy Sets, Multi-valued Mappings, and Rough Sets	60
Foundations and Methods I	
Investigating the Choice of <i>l</i> and <i>u</i> Values in the Extended Variable Precision Rough Sets Model	61
A Quantitative Analysis of Preclusivity vs. Similarity Based Rough Approximations <i>Gianpiero Cattaneo, Davide Ciucci</i>	69

Heyting Wajsberg Algebras as an Abstract Environment Linking	
Fuzzy and Rough Sets	77
Gianpiero Cattaneo, Davide Ciucci	

Dominance-Based Rough Set Approach Using Possibility and Necessity Measures 85 Salvatore Greco, Masahiro Inuiguchi, Roman Słowiński
Generalized Decision Algorithms, Rough Inference Rules, and Flow Graphs
Generalized Rough Sets and Rule Extraction
Towards a Mereological System for Direct Products and Relations
On the Structure of Rough Approximations
Modification of Weights of Conflict Profile's Elements and Dependencies of Attributes in Consensus Model
Reasoning about Information Granules Based on Rough Logic
Foundations and Methods II
A Rough Set Framework for Learning in a Directed Acyclic Graph 144 Herman Midelfart, Jan Komorowski
On Compressible Information Systems 156 Mikhail Moshkov
Functional Dependencies in Relational Expressions Based on Or-Sets 161 Michinori Nakata, Tetsuya Murai
On Asymptotic Properties of Rough–Set–Theoretic Approximations. Fractal Dimension, Exact Sets, and Rough Inclusion in Potentially Infinite Information Systems
About Tolerance and Similarity Relations in Information Systems 175
J.A. Pomykała
J.A. Pomykała Rough Sets, Guarded Command Language, and Decision Rules 183 Frederick V. Ramsey, James J. Alpigini

A New Method for Determining of Extensions and Restrictions of Information Systems
Wojciech Rząsa, Zbigniew Suraj
A Logic Programming Framework for Rough Sets 205 Aida Vitória, Jan Małuszyński
Attribute Core of Decision Table
Foundations and Methods III
Signal Analysis Using Rough Integrals
 How Much Privacy? — A System to Safe Guard Personal Privacy while Releasing Databases
Rough Clustering: An Alternative to Find Meaningful Clusters by Using the Reducts from a Dataset
Concept Learning with Approximation: Rough Version Spaces
Variable Consistency Monotonic Decision Trees
Importance and Interaction of Conditions in Decision Rules
Time Complexity of Rough Clustering: GAs versus K-Means
Induction of Decision Rules and Classification in the Valued Tolerance Approach
Time Series Model Mining with Similarity-Based Neuro-fuzzy Networks and Genetic Algorithms: A Parallel Implementation 279 Julio J. Valdés, Gabriel Mateescu

Granular and Neuro Computing

Closeness of Performance Map Information Granules: A Rough Set Approach
Granular Computing on Binary Relations (Analysis of Conflict and Chinese Wall Security Policy)
 Measures of Inclusion and Closeness of Information Granules: A Rough Set Approach
Rough Neurocomputing: A Survey of Basic Models
of Neurocomputation
Rough Neurocomputing Based on Hierarchical Classifiers
Using Granular Objects in Multi-source Data Fusion
Induction of Classification Rules by Granular Computing
Probabilistic Reasoning
Acquisition Methods for Contextual Weak Independence
A Method for Detecting Context-Specific Independence in Conditional Probability Tables
Properties of Weak Conditional Independence
A Proposal of Probability of Rough Event Based on Probability of Fuzzy Event
Approximate Bayesian Network Classifiers
Accuracy and Coverage in Rough Set Rule Induction

Statistical Test for Rough Set Approximation Based on Fisher's Exact Test
Triangulation of Bayesian Networks: A Relational Database Perspective
Data Mining, Machine Learning, and Pattern Recognition
A New Version of Rough Set Exploration System
Local Attribute Value Grouping for Lazy Rule Induction
Incomplete Data Decomposition for Classification
Extension of Relational Management Systems with Data Mining Capabilities
Reducing Number of Decision Rules by Joining
Scalable Classification Method Based on Rough Sets
Parallel Data Mining Experimentation Using Flexible Configurations 441 José M. Peña, F. Javier Crespo, Ernestina Menasalvas, Victor Robles
 An Optimization of Apriori Algorithm through the Usage of Parallel I/O and Hints
Patterns in Information Maps
Discernibility Matrix Approach to Exception Analysis
Gastric Cancer Data Mining with Ordered Information

Web Mining

A Granular Approach for Analyzing the Degree of Affability of a Web Site
Comparison of Classification Methods for Customer Attrition Analysis 487 Xiaohua Hu
User Profile Model: A View from Artificial Intelligence
Mining the Client's Life Cycle Behaviour in the Web
PagePrompter: An Intelligent Web Agent Created Using Data Mining Techniques
VPRSM Approach to WEB Searching
Applications I
Rough Set Approach to the Survival Analysis
The Identification of Low-Paying Workplaces: An Analysis Using the Variable Precision Rough Sets Model
A Search for the Best Data Mining Method to Predict Melanoma 538 Jerzy W. Grzymała-Busse, Zdzisław S. Hippe
Towards the Classification of Musical Works: A Rough Set Approach 546 Monika P. Hippe
Segmentation of Medical Images Based on Approximations in Rough Set Theory
Adaptive Robust Estimation for Filtering Motion Vectors
Rough Set Feature Selection and Diagnostic Rule Generation for Industrial Applications

Applications II

$\lambda\text{-}\mathrm{Connected}$ Approximations for Rough Sets
Adaptive Classifier Construction: An Approach to Handwritten Digit Recognition
The Application of Support Diagnose in Mitochondrial Encephalomyopathies
Obstacle Classification by a Line-Crawling Robot: A Rough Neurocomputing Approach
Rough Neural Network for Software Change Prediction
Handling Spatial Uncertainty in Binary Images: A Rough Set Based Approach
 Evolutionary Algorithms and Rough Sets-Based Hybrid Approach to Classificatory Decomposition of Cortical Evoked Potentials
Rough Mereological Localization and Navigation
Author Index

In Pursuit of Patterns in Data Reasoning from Data – The Rough Set Way

Zdzisław Pawlak

Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, ul. Bałtycka 5, 44 100 Gliwice, Poland zpw@ii.pw.edu.pl

Abstract. This paper concerns some aspects of rough set based data analysis. In particular rough set look on Bayes' formula leads to new methodology of reasoning from data and shows interesting relationship between Bayes' theorem, rough sets and flow graphs. Three methods of flow graphs application in drawing conclusions from data are presented and examined.

> MOTTO: "It is a capital mistake to theorise before one has data" Sherlock Holmes In: A Scandal in Bohemia

1 Introduction

No doubt that the most famous contribution to reasoning from data should be attributed to the renowned Mr. Sherlock Holmes, whose mastery of using data in reasoning has been well known world wide for over hundred years.

More seriously, reasoning from data is the domain of inductive reasoning, which uses data about sample of larger reality as a starting point of inference – in contrast to deductive reasoning, where axioms expressing some universal truths are used as a departure point of reasoning.

In the rough set approach granular structure of data imposed by the indiscernibility relation is used do discover patterns in data. In rough set theory patterns in data can be characterized by means of approximations, or equivalently by decision rules induced by the data. With every decision rule in a decision table three coefficients are associated: the *strength*, the *certainty* and the *coverage factors* of the rule. It is shown that these coefficients satisfy Bayes' theorem and the total probability theorem. This enables us to use Bayes' theorem to discover patterns in data in a different way from that offered by standard Bayesian inference technique employed in statistical reasoning, without referring to prior and posterior probabilities, inherently associated with Bayesian inference methodology. Besides, a new form of Bayes' theorem is introduced, based on the strength of decision rules, which simplifies essentially computations.

[©] Springer-Verlag Berlin Heidelberg 2002

Furthermore, it is shown that the decision rules define a relation between condition and decision granules, which can be represented by a flow graph. The certainty and coverage factors determine a "flow of information" in the graph, ruled by the total probability theorem and Bayes' theorem, which shows clearly the relationship between condition and decision granules determined by the decision table. This leads to a new class of flow networks, unlike to that introduced by Ford and Fulkerson [1]. The introduced flow graphs may have many applications not necessarily associated with decision tables, but this requires further study.

The decision structure of a decision table can be represented in a "decision space", which is Euclidean space, in which dimensions of the space are determined by decision granules, points in the space are condition granules and coordinates of the points are strengths of the corresponding rules. Distance in the decision space between condition granules allows to determine how "distant" are decision makers in view of their decisions. This idea can be viewed as a generalization of the indiscernibility matrix [7], basic tool to find reducts in information systems. Besides, the decision space gives a clear insight in the decision structure imposed by the decision table.

A simple tutorial example is used to illustrate the basis ideas discussed in the paper.

2 Basic Concepts

In this section we recall basic concepts of rough set theory [4,5,6,7].

An information system is a pair S = (U, A), where U and A, are non-empty finite sets called the *universe*, and the set of *attributes*, respectively such that $a: U \to V_a$, where V_a , is the set of all values of a called the *domain* of a. Any subset B of A determines a binary relation I(B) on U, which will be called an *indiscernibility relation*, and defined as follows: $(x, y) \in I(B)$ if and only if a(x) =a(y) for every $a \in A$, where a(x) denotes the value of attribute a for element x. Obviously I(B) is an equivalence relation. The family of all equivalence classes of I(B), i.e., a partition determined by B, will be denoted by U/I(B), or simply by U/B; an equivalence class of I(B), i.e., block of the partition U/B, containing x will be denoted by B(x) and called B-granule induced by x.

If (x, y) belongs to I(B) we will say that x and y are *B*-indiscernible (indiscernible with respect to B). Equivalence classes of the relation I(B) (or blocks of the partition U/B) are referred to as *B*-elementary sets or *B*-granules.

If we distinguish in the information system two disjoint classes of attributes, called *condition* and *decision attributes*, respectively, then the system will be called a *decision table* and will be denoted by S = (U, C, D), where C and D are disjoint sets of condition and decision attributes, respectively and $C \cup D = A$.

C(x) and D(x) will be referred to as the condition granule and the decision granule induced by x, respectively.

An example of a decision table is shown in Table 1.

Fact no.		Driving conditions		Consequence	N
	weather	road	time	accident	
1	misty	icy	day	yes	80
2	foggy	icy	night	yes	140
3	misty	not icy	night	yes	40
4	sunny	icy	day	no	500
5	foggy	icy	night	no	20
6	misty	not icy	night	no	200

 Table 1. An example of decision table

In the table, 6 facts concerning 980 cases of driving a car in various driving conditions are presented. In the table columns labeled *weather*, *road* and *time*, called *condition attributes*, represent driving conditions. The column labeled by *accident*, called *decision attribute*, contains information whether an accident has occurred or not. N denotes the number of analogous cases.

3 Decision Rules

Each row of the decision table determines a decision rule, e.g., row 1 determines the following decision rule *"if weather is misty and road is icy and time is day then accident occurred"* in 80 cases.

Let S = (U, C, D) be a decision table. Every $x \in U$ determines a sequence $c_1(x), \ldots, c_n(x), d_1(x), \ldots, d_m(x)$ where $\{c_1, \ldots, c_n\} = C$ and $\{d_1, \ldots, d_m\} = D$.

The sequence will be called a *decision rule induced by* x (in S) and denoted by $c_1(x), \ldots, c_n(x) \to d_1(x), \ldots, d_m(x)$ or in short $C \to_x D$.

The number $supp_x(C, D) = |C(x) \cap D(x)|$ will be called a support of the decision rule $C \to_x D$ and the number

$$\sigma_x\left(C,D\right) = \frac{supp_x\left(C,D\right)}{|U|},$$

will be referred to as the *strength* of the decision rule $C \rightarrow_x D$, where |X| denotes the cardinality of X.

With every decision rule $C \to_x D$ we associate a *certainty factor* of the decision rule, denoted $cer_x(C, D)$ and defined as follows:

$$cer_{x}\left(C,D\right) = \frac{\left|C\left(x\right)\cap D\left(x\right)\right|}{\left|C\left(x\right)\right|} = \frac{\sigma_{x}\left(C,D\right)}{\pi\left(C\left(x\right)\right)},$$

where $C(x) \neq \emptyset$ and $\pi(C(x))$.

The certainty factor may be interpreted as conditional probability that y belongs to D(x) given y belongs to C(x), symbolically $\pi_x(D|C)$, i.e., $cer_x(C,D) = \pi_x(D|C)$. If $cer_x(C,D) = 1$, then $C \to_x D$ will be called a *certain decision rule*; if $0 < cer_x(C,D) < 1$ the decision rule will be referred to as an *uncertain decision rule*.

Besides, we will also use a *coverage factor* (see [8]) of the decision rule, denoted $cov_x(C, D)$ defined as

$$cov_{x}\left(C,D\right) = \frac{\left|C\left(x\right)\cap D\left(x\right)\right|}{\left|D\left(x\right)\right|} = \frac{\sigma_{x}\left(C,D\right)}{\pi\left(D\left(x\right)\right)},$$

where $D(x) \neq \emptyset$ and $\pi(D(x)) = \frac{|D(x)|}{|U|}$. Similarly $cov_x (C, D) = \pi_x (C|D)$.

If $C \to_x D$ is a decision rule then $D \to_x C$ will be called an *inverse decision* rule. The inverse decision rules can be used to give *explanations* (reasons) for a decision.

In Table 2 the strength, certainty and coverage factors for Table 1 are given.

fact no.	Strength	Certainty	Coverage
1	0.082	1.000	0.308
2	0.143	0.877	0.538
3	0.041	1.167	0.154
4	0.510	1.000	0.695
5	0.020	0.123	0.027
6	0.204	0.833	0.278

 Table 2. Characterization of decision rules

4 Properties of Decision Rules

Decision rules have important probabilistic properties which are discussed next [2,3].

Let $C \to_x D$ be a decision rule. Then the following properties are valid:

$$\sum_{y \in C(x)} \operatorname{cer}_y(C, D) = 1 \tag{1}$$

$$\sum_{y \in D(x)} cov_y \left(C, D \right) = 1 \tag{2}$$

$$\pi \left(D\left(x\right) \right) = \sum_{y \in C(x)} cer_y\left(C, D \right) \cdot \pi \left(C\left(x \right) \right) =$$

$$= \sum_{y \in C(x)} \sigma_y\left(C, D \right)$$
(3)

In Pursuit of Patterns in Data Reasoning from Data – The Rough Set Way

$$\pi \left(C\left(x\right) \right) = \sum_{y \in D(x)} cov_y \left(C, D \right) \cdot \pi \left(D\left(y \right) \right) =$$

$$= \sum_{y \in D(x)} \sigma_y \left(C, D \right)$$
(4)

$$cer_{x}(C,D) = \frac{cov_{x}(C,D) \cdot \pi(D(x))}{\pi(C(x))} =$$

$$= \frac{\sigma_{x}(C,D)}{\pi(C(x))}$$
(5)

$$cov_{x}(C,D) = \frac{cer_{x}(C,D) \cdot \pi(D(x))}{\pi(D(x))} =$$

$$= \frac{\sigma_{x}(C,D)}{\pi(D(x))}$$
(6)

That is, any decision table, satisfies (1)-(6). Observe that (3) and (4) refer to the well known *total probability theorem*, whereas (5) and (6) refer to *Bayes'* theorem.

Thus in order to compute the certainty and coverage factors of decision rules according to formula (5) and (6) it is enough to know the strength (support) of all decision rules only.

Formulas (5) and (6) can be rewritten as

$$cer_x(C,D) = cov_x(C,D) \cdot \gamma_x(C,D)$$
(7)

$$cov_x \left(C, D \right) = cer_x \left(C, D \right) \cdot \gamma_x^{-1} \left(C, D \right) \tag{8}$$

where $\gamma_x(C,D) = \frac{|D(x)|}{|C(x)|} = \frac{cer_x(C,D)}{cov_x(C,D)}$ Let us observe that

$$cov_x (C, D) \cdot \pi (D(x)) = \sigma_x (C, D)$$
(9)

$$cer_{x}(C,D) \cdot \pi(C(x)) = \sigma_{x}(C,D)$$
(10)

5 Granularity of Data and Flow Graphs

With every decision table we associate a flow graph, i.e., a directed acyclic graph defined as follows: to every decision rule $C \rightarrow_x D$ we assign a directed branch x connecting the input node C(x) and the output node D(x). Strength of the decision rule represents a throughflow of the corresponding branch. The throughflow of the graph is governed by formulas (1),...,(6).

Classification of objects in this representation boils down to finding the maximal output flow in the flow graph, whereas explanation of decisions is connected with the maximal input flow associated with the given decision.

A flow graph for decision table shown in Table 1 is given in Figure 1.

5



Fig. 1. Flow graph

6 Decision Space

With every decision table having one n-valued decision attribute we can associate n-dimensional Euclidean space, where decision granules determine n axis of the space and condition granules determine points of the space. Strengths of decision rules are to be understood as coordinates of corresponding granules.

Distance $\delta(x, y)$ between granules x and y in the n-dimensional decision space is defined as

$$\delta(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_j)^2}$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are vectors of strengths of corresponding decision rules.

A decision space for Table 1 is given in Figure 2.

Distances between granules A, B, C and D are shown in Table 3.

Table 3. Distance matrix

	А	В	С	D
А				
в	0.064			
\mathbf{C}	0.208	0.210		
D	0.517	0.510	0.309	