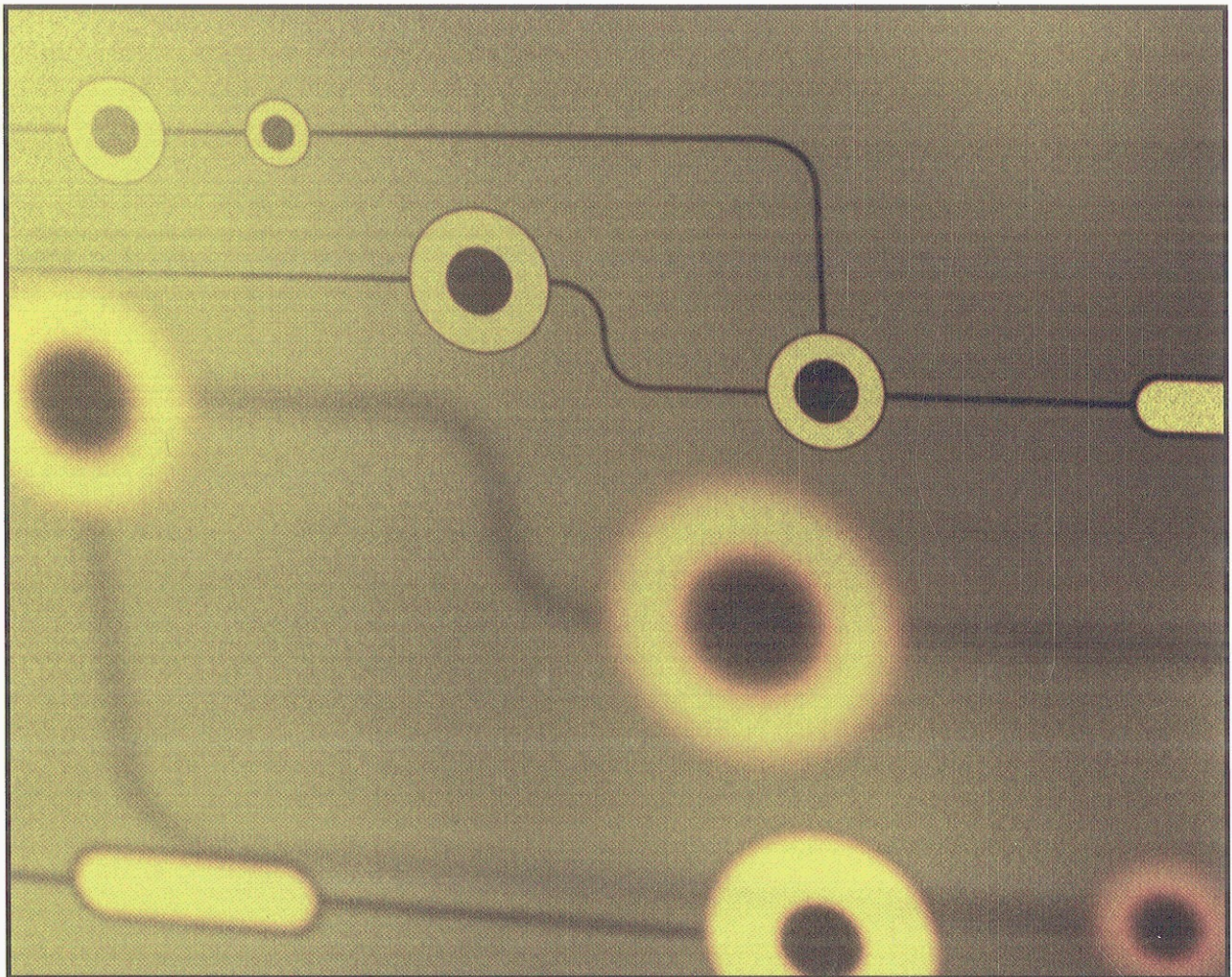


PREMIER REFERENCE SOURCE

DATA MINING AND MEDICAL KNOWLEDGE MANAGEMENT

CASES AND APPLICATIONS



Petr Berka, Jan Rauch, & Djamel Abdelkader Zighed

Detailed Table of Contents

Foreword	xiv
Preface	xix
Acknowledgment	xxiii

Section I **Theoretical Aspects**

This section provides a theoretical and methodological background for the remaining parts of the book. It defines and explains basic notions of data mining and knowledge management, and discusses some general methods.

Chapter I

Data, Information and Knowledge.....	1
--------------------------------------	---

Jana Zvárová, Institute of Computer Science of the Academy of Sciences of the Czech Republic v.v.i., Czech Republic; Center of Biomedical Informatics, Czech Republic
Arnošt Veselý, Institute of Computer Science of the Academy of Sciences of the Czech Republic v.v.i., Czech Republic; Czech University of Life Sciences, Czech Republic
Igor Vajda, Institutes of Computer Science and Information Theory and Automation of the Academy of Sciences of the Czech Republic v.v.i., Czech Republic

This chapter introduces the basic concepts of medical informatics: data, information, and knowledge. It shows how these concepts are interrelated and can be used for decision support in medicine. All discussed approaches are illustrated on one simple medical example.

Chapter II

Ontologies in the Health Field.....	37
-------------------------------------	----

Michel Simonet, Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France
Radja Messai, Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France
Gayo Diallo, Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France
Ana Simonet, Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France

This chapter introduces the basic notions of ontologies, presents a survey of their use in medicine, and explores some related issues: knowledge bases, terminology, information retrieval. It also addresses the issues of ontology design, ontology representation, and the possible interaction between data mining and ontologies.

Chapter III

Cost-Sensitive Learning in Medicine..... 57

Alberto Freitas, University of Porto, Portugal; CINTESIS, Portugal

Pavel Brazdil, LIAAD - INESC Porto L.A., Portugal; University of Porto, Portugal

Altamiro Costa-Pereira, University of Porto, Portugal; CINTESIS, Portugal

Health managers and clinicians often need models that try to minimize several types of costs associated with healthcare, including attribute costs (e.g. the cost of a specific diagnostic test) and misclassification costs (e.g. the cost of a false negative test). This chapter presents some concepts related to cost-sensitive learning and cost-sensitive classification in medicine and reviews research in this area.

Chapter IV

Classification and Prediction with Neural Networks..... 76

Arnošt Veselý, Czech University of Life Sciences, Czech Republic

This chapter describes the theoretical background of artificial neural networks (architectures, methods of learning) and shows how these networks can be used in medical domain to solve various classification and regression problems.

Chapter V

Preprocessing Perceptrons and Multivariate Decision Limits..... 108

Patrik Eklund, Umeå University, Sweden

Lena Kallin Westin, Umeå University, Sweden

This chapter introduces classification networks composed of preprocessing layers and classification networks, and compares them with “classical” multilayer perceptrons on three medical case studies.

Section II General Applications

This section presents work that is general in the sense of a variety of methods or variety of problems described in each of the chapters.

Chapter VI

Image Registration for Biomedical Information Integration..... 122

Xiu Ying Wang, BMIT Research Group, The University of Sydney, Australia

Dagan Feng, BMIT Research Group, The University of Sydney, Australia; Hong Kong Polytechnic University, Hong Kong

In this chapter, biomedical image registration and fusion, which is an effective mechanism to assist medical knowledge discovery by integrating and simultaneously representing relevant information from diverse imaging resources, is introduced. This chapter covers fundamental knowledge and major methodologies of biomedical image registration, and major applications of image registration in biomedicine.

Chapter VII

ECG Processing 137

Lenka Lhotská, Czech Technical University in Prague, Czech Republic

Václav Chudáček, Czech Technical University in Prague, Czech Republic

Michal Huptych, Czech Technical University in Prague, Czech Republic

This chapter describes methods for preprocessing, analysis, feature extraction, visualization, and classification of electrocardiogram (ECG) signals. First, preprocessing methods mainly based on the discrete wavelet transform are introduced. Then classification methods such as fuzzy rule-based decision trees and neural networks are presented. Two examples - visualization and feature extraction from Body Surface Potential Mapping (BSPM) signals and classification of Holter ECGs – illustrate how these methods are used.

Chapter VIII

EEG Data Mining Using PCA 161

Lenka Lhotská, Czech Technical University in Prague, Czech Republic

Vladimír Krajča, Faculty Hospital Na Bulovce, Czech Republic

Jitka Mohylová, Technical University Ostrava, Czech Republic

Svojmil Petránek, Faculty Hospital Na Bulovce, Czech Republic

Václav Gerla, Czech Technical University in Prague, Czech Republic

This chapter deals with the application of principal components analysis (PCA) to the field of data mining in electroencephalogram (EEG) processing. Possible applications of this approach include separation of different signal components for feature extraction in the field of EEG signal processing, adaptive segmentation, epileptic spike detection, and long-term EEG monitoring evaluation of patients in a coma.

Chapter IX

Generating and Verifying Risk Prediction Models Using Data Mining 181

Darryl N. Davis, University of Hull, UK

Thuy T.T. Nguyen, University of Hull, UK

In this chapter, existing clinical risk prediction models are examined and matched to the patient data to which they may be applied using classification and data mining techniques, such as neural Nets. Novel risk prediction models are derived using unsupervised cluster analysis algorithms. All existing and derived models are verified as to their usefulness in medical decision support on the basis of their effectiveness on patient data from two UK sites.

Chapter X

Management of Medical Website Quality Labels via Web Mining.....	206
<i>Vangelis Karkaletsis, National Center of Scientific Research "Demokritos", Greece</i>	
<i>Konstantinos Stamatakis, National Center of Scientific Research "Demokritos", Greece</i>	
<i>Pythagoras Karampiperis, National Center of Scientific Research "Demokritos", Greece</i>	
<i>Martin Labský, University of Economics, Prague, Czech Republic</i>	
<i>Marek Růžička, University of Economics, Prague, Czech Republic</i>	
<i>Vojtěch Svátek, University of Economics, Prague, Czech Republic</i>	
<i>Enrique Amigó Cabrera, ETSI Informática, UNED, Spain</i>	
<i>Matti Pöllä, Helsinki University of Technology, Finland</i>	
<i>Miquel Angel Mayer, Medical Association of Barcelona (COMB), Spain</i>	
<i>Dagmar Villarroel Gonzales, Agency for Quality in Medicine (Aqumed), Germany</i>	

This chapter deals with the problem of quality assessment of medical Web sites. The so called "quality labeling" process can benefit from employment of Web mining and information extraction techniques, in combination with flexible methods of Web-based information management developed within the Semantic Web initiative.

Chapter XI

Two Case-Based Systems for Explaining Exceptions in Medicine.....	227
<i>Rainer Schmidt, University of Rostock, Germany</i>	

In medicine, doctors are often confronted with exceptions, both in medical practice or in medical research. One proper method of how to deal with exceptions is case-based systems. This chapter presents two such systems. The first one is a knowledge-based system for therapy support. The second one is designed for medical studies or research. It helps to explain cases that contradict a theoretical hypothesis.

Section III Specific Cases

This part shows results of several case studies of (mostly) data mining applied to various specific medical problems. The problems covered by this part, range from discovery of biologically interpretable knowledge from gene expression data, over human embryo selection for the purpose of human in-vitro fertilization treatments, to diagnosis of various diseases based on machine learning techniques.

Chapter XII

Discovering Knowledge from Local Patterns in SAGE Data.....	251
<i>Bruno Crémilleux, Université de Caen, France</i>	
<i>Arnaud Soulet, Université François Rabelais de Tours, France</i>	
<i>Jiří Kléma, Czech Technical University, in Prague, Czech Republic</i>	
<i>Céline Hébert, Université de Caen, France</i>	
<i>Olivier Gandrillon, Université de Lyon, France</i>	

Current gene data analysis is often based on global approaches such as clustering. An alternative way is to utilize local pattern mining techniques for global modeling and knowledge discovery. This chapter proposes three data mining methods to deal with the use of local patterns by highlighting the most promis-

ing ones or summarizing them. From the case study of the SAGE gene expression data, it is shown that this approach allows generating new biological hypotheses with clinical applications.

Chapter XIII

Gene Expression Mining Guided by Background Knowledge..... 268

Jiří Kléma, Czech Technical University in Prague, Czech Republic
Filip Železný, Czech Technical University in Prague, Czech Republic
Igor Trajkovski, Jožef Stefan Institute, Slovenia
Filip Karel, Czech Technical University in Prague, Czech Republic
Bruno Crémilleux, Université de Caen, France
Jakub Tolar, University of Minnesota, USA

This chapter points out the role of genomic background knowledge in gene expression data mining. Its application is demonstrated in several tasks such as relational descriptive analysis, constraint-based knowledge discovery, feature selection and construction, or quantitative association rule mining.

Chapter XIV

Mining Tinnitus Database for Knowledge..... 293

Pamela L. Thompson, University of North Carolina at Charlotte, USA
Xin Zhang, University of North Carolina at Pembroke, USA
Wenxin Jiang, University of North Carolina at Charlotte, USA
Zbigniew W. Ras, University of North Carolina at Charlotte, USA
Pawel Jastreboff, Emory University School of Medicine, USA

This chapter describes the process used to mine a database containing data, related to patient visits during Tinnitus Retraining Therapy. The presented research focused on analysis of existing data, along with automating the discovery of new and useful features in order to improve classification and understanding of tinnitus diagnosis.

Chapter XV

Gaussian-Stacking Multiclassifiers for Human Embryo Selection..... 307

Dinora A. Morales, University of the Basque Country, Spain
Endika Bengoetxea, University of the Basque Country, Spain
Pedro Larrañaga, Universidad Politécnica de Madrid, Spain

This chapter describes a new multi-classification system using Gaussian networks to combine the outputs (probability distributions) of standard machine learning classification algorithms. This multi-classification technique has been applied to a complex real medical problem: The selection of the most promising embryo-batch for human in-vitro fertilization treatments.

Chapter XVI

Mining Tuberculosis Data..... 332

Marisa A. Sánchez, Universidad Nacional del Sur, Argentina
Sonia Uremovich, Universidad Nacional del Sur, Argentina
Pablo Acroglitano, Hospital Interzonal Dr. José Penna, Argentina

This chapter reviews current policies of tuberculosis control programs for the diagnosis of tuberculosis. A data mining project that uses WHO’s Direct Observation of Therapy data to analyze the relationship among different variables and the tuberculosis diagnostic category registered for each patient is then presented.

Chapter XVII

Knowledge-Based Induction of Clinical Prediction Rules..... 350

Mila Kwiatkowska, Thompson Rivers University, Canada

M. Stella Atkins, Simon Fraser University, Canada

Les Matthews, Thompson Rivers University, Canada

Najib T. Ayas, University of British Columbia, Canada

C. Frank Ryan, University of British Columbia, Canada

This chapter describes how to integrate medical knowledge with purely inductive (data-driven) methods for the creation of clinical prediction rules. To address the complexity of the domain knowledge, the authors have introduced a semio-fuzzy framework, which has its theoretical foundations in semiotics and fuzzy logic. This integrative framework has been applied to the creation of clinical prediction rules for the diagnosis of obstructive sleep apnea, a serious and under-diagnosed respiratory disorder.

Chapter XVIII

Data Mining in Atherosclerosis Risk Factor Data..... 376

Petr Berka, University of Economics, Prague, Czech Republic; Academy of Sciences of the Czech Republic, Prague, Czech Republic

Jan Rauch, University of Economics, Praague, Czech Republic; Academy of Sciences of the Czech Republic, Prague, Czech Republic

Marie Tomečková, Academy of Sciences of the Czech Republic, Prague, Czech Republic

This chapter describes goals, current results, and further plans of long-time activity concerning the application of data mining and machine learning methods to the complex medical data set. The analyzed data set concerns longitudinal study of atherosclerosis risk factors.

Compilation of References 398

About the Contributors 426

Index..... 437