Advanced Information and Knowledge Processing

Lipo Wang · Xiuju Fu

Data Mining with Computational Intelligence

With 72 Figures and 65 Tables

Springer

Lipo Wang
Nanyang Technological University
School of Electrical and Electronical Engineering
Block S1, Nanyang Avenue,
639798 Singapore, Singapore
elpwang@ntu.edu.sg

Xiuju Fu
Institute of High Performance Computing,
Software and Computing, Science Park 2,
The Capricorn
Science Park Road 01-01
117528 Singapore, Singapore
fuxj@pmail.ntu.edu.sg

Series Editors
Xindong Wu
Lakhmi Jain

Library of Congress Control Number: 200528948

ACM Computing Classification (1998): H.2.8., I.2

ISBN-10 3-540-24522-7 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-24522-3 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springeronline.com

Cover design: KünkelLopka, Heidelberg
Typesetting: Camera ready by the authors
Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig
Printed on acid-free paper 45/3142/YL - 5 4 3 2 1 0

Preface

Nowadays data accumulate at an alarming speed in various storage devices, and so does valuable information. However, it is difficult to understand information hidden in data without the aid of data analysis techniques, which has provoked extensive interest in developing a field separate from machine learning. This new field is data mining.

Data mining has successfully provided solutions for finding information from data in bioinformatics, pharmaceuticals, banking, retail, sports and entertainment, etc. It has been one of the fastest growing fields in the computer industry. Many important problems in science and industry have been addressed by data mining methods, such as neural networks, fuzzy logic, decision trees, genetic algorithms, and statistical methods.

This book systematically presents how to utilize fuzzy neural networks, multi-layer perceptron (MLP) neural networks, radial basis function (RBF) neural networks, genetic algorithms (GAs), and support vector machines (SVMs) in data mining tasks. Fuzzy logic mimics the imprecise way of reasoning in natural languages and is capable of tolerating uncertainty and vagueness. The MLP is perhaps the most popular type of neural network used today. The RBF neural network has been attracting great interest because of its locally tuned response in RBF neurons like biological neurons and its global approximation capability. This book demonstrates the power of GAs in feature selection and rule extraction. SVMs are well known for their excellent accuracy and generalization abilities.

We will describe data mining systems which are composed of data preprocessing, knowledge-discovery models, and a data-concept description. This monograph will enable both new and experienced data miners to improve their practices at every step of data mining model design and implementation.

Specifically, the book will describe the state of the art of the following topics, including both work carried out by the authors themselves and by other researchers:

- Data mining tools, i.e., neural networks, support vector machines, and genetic algorithms with application to data mining tasks.
- Data mining tasks including data dimensionality reduction, classification, and rule extraction.

Lipo Wang wishes to sincerely thank his students, especially Feng Chu, Yakov Frayman, Guosheng Jin, Kok Keong Teo, and Wei Xie, for the great pleasure of collaboration, and for carrying out research and contributing to this book. Thanks are due to Professors Zhiping Lin, Kai-Ming Ting, Chunru Wan, Ron (Zhengrong) Yang, Xin Yao, and Jacek M. Zurada for many helpful discussions and for the opportunities to work together. Xiuju Fu wishes to express gratitude to Dr. Gih Guang Hung, Liping Goh, Professors Chongjin Ong and S. Sathiya Keerthi for their discussions and supports in the research work. We also express our appreciation for the support and encouragement from Professor L.C. Jain and Springer Editor Ralf Gerstner.

Singapore, *Lipo Wang*
May 2005 *Xiuju Fu*

Contents