Studies in Computational Intelligence, Volume 288

Editor-in-Chief
Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 267. Ivan Zelinka, Sergej Celikovský, Hendrik Richter, and Guanrong Chen (Eds.)
Evolutionary Algorithms and Chaotic Systems, 2009
ISBN 978-3-642-10706-1

Vol. 268. Johann M.Ph. Schumann and Yan Liu (Eds.)
Applications of Neural Networks in High Assurance Systems, 2009
ISBN 978-3-642-10689-7

Vol. 269. Francisco Fernández de de Vega and Erick Cantú-Paz (Eds.)
Parallel and Distributed Computational Intelligence, 2009
ISBN 978-3-642-10674-3

Vol. 270. Zong Woo Geem
Recent Advances In Harmony Search Algorithm, 2009
ISBN 978-3-642-04316-1

Vol. 271. Janusz Kacprzyk, Frederick E. Petry, and Adnan Yazici (Eds.)
Uncertainty Approaches for Spatial Data Modeling and Processing, 2009
ISBN 978-3-642-10662-0

Vol. 272. Carlos A. Coello Coello, Clarisse Dhaenens, and Laetitia Jourdan (Eds.)
Advances in Multi-Objective Nature Inspired Computing, 2009
ISBN 978-3-642-11217-1

Vol. 273. Fatos Xhafa, Santi Caballé, Ajith Abraham, Thanasis Daradoumis, and Angel Alejandro Juan Perez (Eds.)
Computational Intelligence for Technology Enhanced Learning, 2010
ISBN 978-3-642-11223-2

Vol. 274. Zbigniew W. Raś and Alicja Wieczorkowska (Eds.)
Advances in Music Information Retrieval, 2010
ISBN 978-3-642-11673-5

Vol. 275. Dilip Kumar Pratihar and Lakhmi C. Jain (Eds.)
Intelligent Autonomous Systems, 2010
ISBN 978-3-642-11675-9

Vol. 276. Jacek Mańdziuk
Knowledge-Free and Learning-Based Methods in Intelligent Game Playing, 2010
ISBN 978-3-642-11677-3

Vol. 277. Filippo Spagnolo and Benedetto Di Paola (Eds.)
European and Chinese Cognitive Styles and their Impact on Teaching Mathematics, 2010
ISBN 978-3-642-11679-7

Vol. 278. Radomir S. Stankovic and Jaakko Astola
From Boolean Logic to Switching Circuits and Automata, 2010
ISBN 978-3-642-11681-0

Vol. 279. Manolis Wallace, Ioannis E. Anagnostopoulos, Phivos Mylonas, and Maria Bielikova (Eds.)
Semantics in Adaptive and Personalized Services, 2010
ISBN 978-3-642-11683-4

Vol. 280. Chang Wen Chen, Zhu Li, and Shiguo Lian (Eds.)
Intelligent Multimedia Communication: Techniques and Applications, 2010
ISBN 978-3-642-11685-8

Vol. 281. Robert Babuska and Frans C.A. Groen (Eds.)
Interactive Collaborative Information Systems, 2010
ISBN 978-3-642-11687-2

Vol. 282. Husrev Taha Sencar, Sergio Velastin, Nikolaos Nikolaidis, and Shiguo Lian (Eds.)
Intelligent Multimedia Analysis for Security Applications, 2010
ISBN 978-3-642-11754-1

Vol. 283. Ngoc Thanh Nguyen, Radoslaw Katarzyniak, and Shyi-Ming Chen (Eds.)
Advances in Intelligent Information and Database Systems, 2010
ISBN 978-3-642-12089-3

Vol. 284. Juan R. González, David Alejandro Pelta, Carlos Cruz, Germán Terrazas, and Natalio Krasnogor (Eds.)
Nature Inspired Cooperative Strategies for Optimization (NICSO 2010), 2010
ISBN 978-3-642-12537-9

Vol. 285. Roberto Cipolla, Sebastiano Battiato, and Giovanni Maria Farinella (Eds.)
Computer Vision, 2010
ISBN 978-3-642-12847-9

Vol. 286. Alexander Bolshoy, Zeev (Vladimir) Volkovich, Valery Kirzhner, and Zeev Barzily
Genome Clustering, 2010
ISBN 978-3-642-12951-3

Vol. 287. Dan Schonfeld, Caifeng Shan, Dacheng Tao, and Liang Wang (Eds.)
Video Search and Mining, 2010
ISBN 978-3-642-12899-8

Vol. 288. I-Hsien Ting, Hui-Ju Wu, Tien-Hwa Ho (Eds.)
Mining and Analyzing Social Networks, 2010
ISBN 978-3-642-13421-0

I-Hsien Ting, Hui-Ju Wu, Tien-Hwa Ho (Eds.)

Mining and Analyzing Social Networks

Springer

Dr. I-Hsien Ting
Department of Information
Management, No. 700
National University of Kaohsiung
Kaohsiung University Rd.
Kaohsiung, 811
Taiwan 5.
E-mail: iting@nuk.edu.tw

Dr. Tien-Hwa Ho
Department of Information
Management, No. 700
National University of Kaohsiung
Kaohsiung University Rd.
Kaohsiung, 811
Taiwan 5

Dr. Hui-Ju Wu
Department of Information
Management, No. 700
National University of Kaohsiung
Kaohsiung University Rd.
Kaohsiung, 811
Taiwan 5

ISBN 978-3-642-13421-0 e-ISBN 978-3-642-13422-7

DOI 10.1007/978-3-642-13422-7

Studies in Computational Intelligence ISSN 1860-949X

Library of Congress Control Number: 2010928121

© 2010 Springer-Verlag Berlin Heidelberg

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, reuse of illustrations, recitation, broadcasting, reproduction on microfilm 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. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typeset & *Cover Design:* Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

Mining social networks has now becoming a very popular research area not only for data mining and web mining but also social network analysis. Data mining is a technique that has the ability to process and analyze large amount of data and by this to discover valuable information from the data. In recent year, due to the growth of social communications and social networking websites, data mining becomes a very important and powerful technique to process and analyze such large amount of data. Thus, this book will focus upon Mining and Analyzing social network.

Some chapters in this book are extended from the papers that presented in MSNDS2009 (the First International Workshop on Mining Social Networks for Decision Support) and SNMABA2009 ((The International Workshop on Social Networks Mining and Analysis for Business Applications)). In addition, we also sent invitations to researchers that are famous in this research area to contribute for this book. The chapters of this book are introduced as follows:

In chapter 1-Graph Model for Pattern Recognition in Text, Qin Wu et al. present a novel approach that uses a weighted directed multigraph for text pattern recognition. In the proposed methodology, a weighted directed multigraph model has been set up by using the distances between the keywords as the weights of arcs as well a keyword-frequency distance based algorithm has also been introduced. Case studies are also included in this chapter to show the performance is better than traditional means.

In chapter 2-Information Retrieval in Wikis using an Ontology, Carlos Miguel Tobar et al. presented an system which is designed based on the ideas from the semantic Web combined with adaptive mechanisms and a modification of the classic vector model for information retrieval. This system can be used to extract relevant information from huge amount of txt, such as wiki.

In chapter 3-Ego-centric Network Sampling in Viral Marketing Applications, Huaiyu (Harry) Ma et al. describe a study about ego-centric network sampling to show the network structure can be captured accurately. The Stanford-Berkeley network to show that the approach can capture the underlying structure with a minimal amount of data.

In chapter 4- Integrating SNA and DM Technology into HR Practice and Research: Layoff Prediction Model, Hui-Ju Wu et al. proposed a new application direction to combine the techniques of SNA and DM into the research area of Human Resource Management. In this chapter, a valuable dataset has been used to analyze the social structure in a organization and by this to discover the reasons behind layoff.

In chapter 5-Actor Identification in Implicit Relational Data Sources, Michael Farrugia and Aaron Quigley presents a study of a range of techniques that can be employed to identify unique actors when inferring networks from non explicit network data sets. They also present methods for unique node identification of social network actors in a business scenario. A real world case study has also been included in this chapter.

In chapter 6- Perception of Online Social Networks, Travis Green and Aaron Quigley examine data derived from an application on Facebook.com that investigates the relations among members of their online social network. It confirms that online social networks are more often used to maintain weak connections but that a subset of users focus on strong connections, determines that connection intensity to both connected people predicts perceptual accuracy, and shows that intra-group connections are perceived more accurately.

In chapter 7- Ranking Learning Entities on the Web by Integrating Network-based Features, Yingzi Jin et al. propose an algorithm to generate and integrate network-based features systematically from a given social network that is mined from the world-wide web. After learning a model for explaining target rankings researchers' productivity based on social networks confirms the effectiveness of our models. This chapter specifically examines the application of a social network that exemplifies the advanced use of social networks mined from the web.

In chapter 8-Discovering Proximal Social Intelligence for uality Decision Support, Yuan-Chu Hwang focus on discovering the proximal social intelligence or quality decision support. The author illustrates a case of leisure recommendatory e-service for bicycle exercise entertainment in Taiwan as well as introduces the proximity e-service as well as its theoretical support.

In chapter 9- Discovering User Interests by Document Classification, Loc Nguyen propose I propose a new approach for discovering user interest based on document classification. The basic idea is to consider user interests as classes of documents. The process of classifying documents is also the process of discovering user interests.

In chapter 10- Network Analysis of Opto-electronics Industry Cluster: A Case of TAIWAN, Ting-Lin LEE provides a study to describe supply chain relationships networks of opto-electronics industry in STSP as fully as possible, tease out the prominent patterns in such networks, and discover what effects these relationships and networks have on organizations performance. The results of this study contribute to a better understanding of how firms can utilize network benefits to enhance their innovation performance. Furthermore, "coreness centrality" is the most interpretable position variable for innovation performance.

In summary, this book's content sets out to highlight the trends in the research area in Mining and Analysis of Social Networks. Through integrating the two research areas of social networks analysis and data mining, more and more applications and research ideas can be rised.

I-Hsien Ting
Hui-Ju Wu
Tien-Hwa Ho

Contents

Graph Model for Pattern Recognition in Text 1
Qin Wu, Eddie Fuller, Cun-Quan Zhang

Retrieving Wiki Content Using an Ontology 21
Carlos Miguel Tobar, Alessandro Santos Germer, Juan Manuel Adán-Coello, Ricardo Luís de Freitas

Ego-Centric Network Sampling in Viral Marketing Applications 35
Huaiyu (Harry) Ma, Steven Gustafson, Abha Moitra, David Bracewell

Integrating SNA and DM Technology into HR Practice and Research: Layoff Prediction Model 53
Hui-Ju Wu, I-Hsien Ting, Huo-Tsan Chang

Actor Identification in Implicit Relational Data Sources 67
Michael Farrugia, Aaron Quigley

Perception of Online Social Networks 91
Travis Green, Aaron Quigley

Ranking Learning Entities on the Web by Integrating Network-Based Features 107
Yingzi Jin, Yutaka Matsuo, Mitsuru Ishizuka

Discovering Proximal Social Intelligence for Quality Decision Support 125
Yuan-Chu Hwang

Discovering User Interests by Document Classification 139
Loc Nguyen

Network Analysis of Opto-Electronics Industry Cluster: A Case of Taiwan 161
Ting-Lin LEE

Author Index 183