

Zbigniew W. Ras and William Ribarsky (Eds.)

Advances in Information and Intelligent Systems

Prof. Zbigniew W. Ras
University of North Carolina at Charlotte
College of Computing and Informatics
Charlotte, N.C. 28223
USA
E-mail: ras@uncc.edu

William Ribarsky
University of North Carolina at Charlotte
College of Computing and Informatics
Charlotte, N.C. 28223
USA

ISBN 978-3-642-04140-2

e-ISBN 978-3-642-04141-9

DOI 10.1007/978-3-642-04141-9

Studies in Computational Intelligence

ISSN 1860-949X

Library of Congress Control Number: 2009934304

© 2009 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 in acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

The College of Computing and Informatics (CCI) at UNC-Charlotte has three departments: Computer Science, Software and Information Systems, and Bioinformatics and Genomics. The Department of Computer Science offers study in a variety of specialized computing areas such as database design, knowledge systems, computer graphics, artificial intelligence, computer networks, game design, visualization, computer vision, and virtual reality. The Department of Software and Information Systems is primarily focused on the study of technologies and methodologies for information system architecture, design, implementation, integration, and management with particular emphasis on system security. The Department of Bioinformatics and Genomics focuses on the discovery, development and application of novel computational technologies to help solve important biological problems.

This volume gives an overview of research done by CCI faculty in the area of Information & Intelligent Systems. Presented papers focus on recent advances in four major directions: Complex Systems, Knowledge Management, Knowledge Discovery, and Visualization. A major reason for producing this book was to demonstrate a new, important thrust in academic research where college-wide interdisciplinary efforts are brought to bear on large, general, and important problems. As shown in the research described here, these efforts need not be formally organized joint undertakings (though parts could be) but are rather a convergence of interests around grand themes. Taken together, the sections and chapters in this book reveal a variety of methods and tools that can be applied to quite diverse types of data and analysis problems but where each produces knowledge and features that can be combined, organized, explored, understood, and used in common ways (for example, through interactive visualization interfaces).

The first part of the book contains three contributions in the area of Complex Systems.

- The first chapter is written by J. Studnicki, C. Eichelberger, and J. Fisher and it presents a novel application of complex, multi-agent simulation methods to a study of how informed patient choices can influence the distribution of surgical volume for complex procedures. Their exploratory model suggests that multi-agent simulation methods can be helpful in understanding the

complex interactions which are operative within the U.S. healthcare industry. The authors have focused upon the relationship involved in the performance of complex surgeries, especially those for which there is a significant likelihood of an adverse outcome in the form of a post surgical complication, or even death.

- The next chapter, written by W.J. Tolone explores a particular form of modeling and simulation, called integrated modeling and simulation, which is well-suited to empower users to make sense of complex systems. The author demonstrates its potential through the Integrated Modeling Environment. He also illustrates how the design of the IME recognizes the implications of sensemaking on integrated modeling and simulation.
- In the chapter titled “*Towards a General Tool for Studying Threshold Effects Across Diverse Domains*”, the authors present the hypothesis that by modeling differing complex systems we can use the known causes and mechanisms in one domain to gain insight into the controlling properties of similar effects in another domain. To that end, they create a general Complex Adaptive Systems model so that it can be individually tailored and mapped to phenomena in various domains. They describe how this model applies to two domains: cancer/immune response and political dissent.

The second part of the book contains four contributions in the area of Knowledge Discovery.

- Web-accessible databases (e.g., book databases from Amazon and Barnes & Noble) are growing at a phenomenal rate. As a result, it is becoming increasingly important to be able to effectively search over these databases. The first chapter in this section studies the problem of integrating query interfaces of databases, a key step toward providing uniform access to the databases. It proposes novel hierarchical modeling of query interfaces, to address limitations of existing modeling solutions. A novel spatial clustering-based algorithm, to effectively discover interface models, is presented. It also describes results of extensive experiments over real-world data sets, to demonstrate the utility of the proposed solution.
- The purpose of the next chapter is to investigate how well an edge based graph randomization approach can protect node identities and sensitive links in publishing social networks. The authors quantify both identity disclosure and link disclosure when adversaries have the vertex degree background knowledge. Their studies show that edge randomization is a necessity in addition to node anonymization in order to preserve privacy in the released graph.
- The third chapter, written by W. Jiang, A. Cohen, and Z.W. Ras, presents a multi-labeled classification system for polyphonic music retrieval that estimates multiple timbre information from polyphonic sounds according to a similarity measure based on both feature vectors and spectrum envelope. In order to achieve a higher estimation rate, they introduce the

hierarchical structured classification model under the inspiration of the human perceptual process. This cascade classification system would first estimate the higher level decision attribute, which stands for the musical instrument family. Then, further estimation is done within that specific family range. Experiments show better performance of cascade system than the flattened classifiers.

- The fourth chapter, written by J. Liu and K. Subramanian, proposes a new centerline extraction method that employs a Gaussian type probability model to estimate the boundaries of medical objects. The model is computed using an integration of the image gradient field. Probabilities assigned to boundary voxels are then used to compute a more robust distance field that is less sensitive to noise. Distance field algorithms are then applied to extract the centerline.

The third part of the book contains five contributions in the area of Knowledge Management.

- In the first chapter, authors incorporate a novel visual analytics framework to design a human-centered multimedia computing environment. Their visual analytics framework can allow users to obtain better understanding of the hypotheses, which means they can further incorporate their personal preferences to make more suitable hypotheses for achieving personalized classifier training.
- The second chapter, written by R. Gandhi and D.C. Wilson, examines a multi-strategy approach for improving the quality of geo-entity extraction. The implemented experimental framework is targeted for web data, and it provides a comparative evaluation of individual approaches and parameterizations of their multi-strategy method. Presented results show that the multi-strategy approach provides a significant benefit in terms of accuracy, domain independence, and adaptability.
- The emerging Monte-Carlo Tree Search (MCTS) paradigm is bringing an exciting breakthrough in computer Go toward challenging human experts, especially on smaller Go boards. The third chapter, written by K.-H. Chen, D. Du, and P. Zhang, gives an overview of both classical and MCTS approaches to computer Go. MCTS techniques as implemented in Go Intellect are discussed. Solving capturing problems in Go using MCTS is presented in detail. A novel approach for tuning program parameters using genetic algorithms guided by confidence bounds is introduced.
- The fourth chapter, written by R. Gandhi and S.-W. Lee, presents ongoing research on the transformation of informal sources (in the problem space) into a representation that supports well-defined metrics (in the solution space) through a combination of knowledge engineering and requirements engineering techniques. Their research outlines a methodological approach for metrics development and understanding using the structured representation of regulatory security requirements in a problem domain ontology. The metrics derived from the domain ontology create a

traceable chain of analytical thoughts with software artifacts (e.g., requirements, design, and code).

- Botnets have historically used centralized architectures for their command and control systems. While deployment and logical construction of these systems is simplistic and efficient, a critical weak-point exists in the central server used to coordinate messages and route traffic. Recently, the introduction of decentralized architectures with peer-to-peer (P2P) routing has provided malware authors with increased resilience and location obfuscation for command distribution points. To date, botnets with these topologies have been difficult for the defenders to accurately enumerate and effectively remediate. In the last chapter, written by B. Kang and C. Nunnery, authors describe the architectures, capabilities, functional behaviours, and current mitigation efforts for the Nugache, Storm, and Mayday botnets.

The last part of the book contains four contributions in the area of Visualization.

- In the first chapter of this section authors propose a general framework for dimension management in high dimensional visualization. They provide guidelines for the design and development of dimension management functions in high dimensional visualization systems. Their recent work on dimension management in high dimensional visualization, namely the Hierarchical Dimension Management approach, the Value and Relation display, and the Multivariate Visual Explanation approach, are presented as examples to illustrate the proposed framework.
- The second chapter presents a new concept of knowledge templates for visualizing multi-field, time-varying 3D air quality data. There are three main contributions. First, the authors design a set of multi-level knowledge templates to capture important statistical data properties based on the distribution features of air quality data. Second, they have developed a fast template synthesis method to generate suitable templates according to user intentions. Third, they developed an integrated visualization system for visually comparing multiple templates and volume datasets.
- Information visualization has seen many applications recently in areas such as news analysis, network optimization, and financial fraud detection. There is also an undercurrent of artistic work in this field, that has inspired some new techniques, but does not quite fit into the data-driven computer science world nor to the pure art world. Visualization for communication clearly benefits from design and art influences, though. The third chapter, written by C. Ziemkiewicz and R. Kosara, attempts to place Information Visualization in a broader context by applying a number of formal criteria to information graphics, artistic visualization, scientific and information visualization, etc. Similar criteria also serve to differentiate within visualization, to better understand the differences between types of visualization, such as flow, volume and information visualization. Ideas from art theory turn out to mesh well with an analysis

of information loss and other more data-centric criteria to provide structure to a collection of fields that are clearly related to each other, but whose relations are poorly understood.

- The last chapter of this section develops and applies a geometric model of the ideal interaction volume for stereoscopic displays used for 3d user interfaces in multi-scale virtual environments. The authors discuss the model and the tradeoffs it must address and they use it to quantitatively analyse several physical display configurations. They present interaction volume management techniques which they integrate into an example multi-scale virtual environment that contains global terrain and volumetric weather data.

We wish to express our thanks to all the authors who contributed the above sixteen chapters to this book.

July 2009

Z.W. Raś
W. Ribarsky

Contents

Part I: Complex Systems

Complex Adaptive Systems: How Informed Patient Choice Influences the Distribution of Complex Surgical Procedures	3
<i>James Studnicki, Chris Eichelberger, John Fisher</i>	

Making Sense of Complex Systems through Integrated Modeling and Simulation	21
<i>William J. Tolone</i>	

Towards a General Tool for Studying Threshold Effects Across Diverse Domains	41
<i>Ted Carmichael, Mirsad Hadzikadic, Didier Dréau, Joseph Whitmeyer</i>	

Part II: Knowledge Discovery

Modeling and Extracting Deep-Web Query Interfaces	65
<i>Wensheng Wu, AnHai Doan, Clement Yu, Weiyi Meng</i>	

On the Quantification of Identity and Link Disclosures in Randomizing Social Networks	91
<i>Xiaowei Ying, Xintao Wu, Kai Pan, Ling Guo</i>	

Polyphonic Music Information Retrieval Based on Multi-label Cascade Classification System	117
<i>Wenxin Jiang, Amanda Cohen, Zbigniew W. Ras</i>	

Accurate and Robust Centerline Extraction from Tubular Structures in Medical Images	139
<i>Jianfei Liu, Kalpathi Subramanian</i>	

Part III: Knowledge Management

Incorporate Visual Analytics to Design a Human-Centered Computing Framework for Personalized Classifier Training and Image Retrieval . . .	165
<i>Yuli Gao, Chunlei Yang, Yi Shen, Jianping Fan</i>	
A Multi-strategy Approach to Geo-Entity Recognition	189
<i>Ruituraj Gandhi, David C. Wilson</i>	
Monte-Carlo Tree Search and Computer Go	201
<i>Keh-Hsun Chen, Dawei Du, Peigang Zhang</i>	
Ontology Guided Risk Analysis: From Informal Specifications to Formal Metrics	227
<i>Robin Gandhi, Seok-Won Lee</i>	
Decentralized Peer-to-Peer Botnet Architectures	251
<i>Brent ByungHoon Kang, Chris Nunnery</i>	

Part IV: Visualization

A Dimension Management Framework for High Dimensional Visualization	267
<i>Jing Yang, Scott Barlowe</i>	
Year-Long Time-Varying 3D Air Quality Data Visualization	289
<i>Aidong Lu, Wei Chen, William Ribarsky, David Ebert</i>	
Embedding Information Visualization within Visual Representation	307
<i>Caroline Ziemkiewicz, Robert Kosara</i>	
Interaction Volume Management in a Multi-scale Virtual Environment	327
<i>Zachary Wartell, Ernst Houtgast, Onno Pfeiffer, Chris D. Shaw, William Ribarsky, Frits Post</i>	
Author Index	351