

Copyright © 2003, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**SOFT COMPUTING FOR INTERNET
AND BIOINFORMATICS**

by

Masoud Nikraves, Souad Souafi-Bensafi
and Lotfi A. Zadeh

Memorandum No. UCB/ERL M03/47

15 December 2003

**SOFT COMPUTING FOR INTERNET
AND BIOINFORMATICS**

by

Masoud Nikravesh, Souad Souafi-Bensafi
and Lotfi A. Zadeh

Memorandum No. UCB/ERL M03/47

15 December 2003

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

FLINT-CIBI 2003

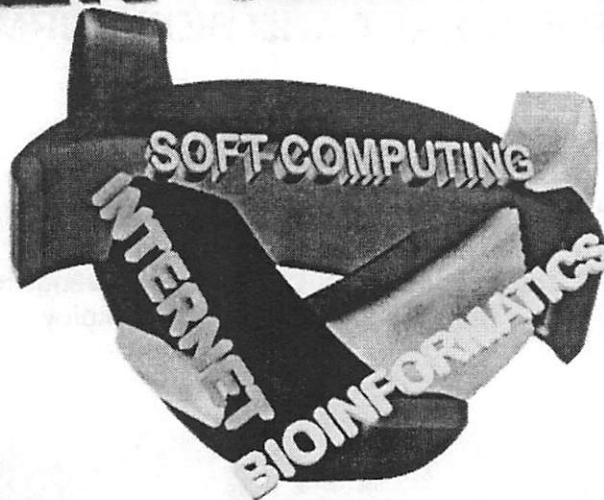
2003 BISC FLINT-CIBI INTERNATIONAL JOINT WORKSHOP ON SOFT COMPUTING FOR INTERNET AND BIOINFORMATICS

December 15-19, 2003
Hewlett-Packard Auditorium (306 Soda Hall)
Computer Science Division
Electrical and Computer Sciences Department
University of California-Berkeley



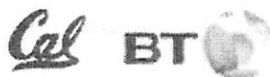
Sponsored by BISC Program, UC Berkeley ILP, BTEExact Technologies

FLINT-CIBI 2008



2003 BISC FLINT-CIBI INTERNATIONAL JOINT WORKSHOP ON SOFT COMPUTING FOR INTERNET AND BIOINFORMATICS

Berkeley Initiative in Soft Computing (BISC)
Computer Science Division
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, CA 94720-1776
Tel: (510) 643-4522, FAX: 510-642-5775
Email: zadeh@cs.berkeley.edu
Email: nikraves@cs.berkeley.edu
URL: <http://www-bisc.cs.berkeley.edu/>



Sponsored by BISC Program, UC Berkeley ILP, BTEExact Technologies

FLINT-CIBI 2008

2003 BISC FLINT-CIBI INTERNATIONAL JOINT WORKSHOP ON SOFT COMPUTING FOR INTERNET AND BIOINFORMATICS FRONTIERS

Honorary Chair:	Lotfi A. Zadeh (BISC-UC Berkeley, USA)
General Chairs:	Masoud Nikravesh (BISC-UC Berkeley, USA) Nik Kasabov (KEDRI, NZ) Tomohiro Takagi (MUSIC-Meiji, Japan)
Local Chairs:	Souad Souafi-Bensafi (BISC-UC Berkeley) Martine De Cock (BISC-UC Berkeley) Mahnaz Hassibi (BISC-UC Berkeley) Gamil Serag-Eldine (BISC-UC Berkeley) Sergio Guadarrama (BISC-UC Berkeley) Sema Alptekin (BISC-UC Berkeley)
Program Committee:	H. Adeli (USA) Ben Azvine (UK) Hamid Berenji (USA) Michael R. Berthold (USA) Antonio Di Nola (Italy) Ronald Fagin (USA) Marti Hearst (USA) Mohammad Jamshidi (USA) Janusz Kacprzyk (Poland) T. Y. Lin (USA) Vincenzo Loia (Italy) John A. Meech (Canada) Bouchon-Meunier (France) Detlef Nauck (UK) Charles Ortiz (USA) Gabriella Pasi (Italy) Fred Petry (USA) James Shanahan (France) Vincent S.M. Tseng (Taiwan) I. Burhan Turksen (Canada) John Yen (USA)

Greetings to Participants in FLINT-CIBI 2003 Workshop

FLINT-CIBI is the third in the sequence of workshops which started with FLINT-I (Fuzzy Logic and the Internet) Workshop in June,2001, in Berkeley.

The genesis of FLINT-I was the perception that fuzzy logic has the potential for making an important contribution to enhancement of web intelligence. The basis for this perception is that much of web knowledge is to some degree imprecise, uncertain and partially true. In web-related literature, such knowledge is dealt with through the use of adaptations of techniques drawn from bivalent logic and probability theory. Although considerable progress has been made in recent years through the use of such techniques, much more can be done through the use of fuzzy logic and, more generally, soft computing--a coalition of methodologies whose principal members are fuzzy logic, neurocomputing, evolutionary computing, probabilistic computing and machine learning.

Another important application-area for soft computing is bioinformatics -- a field that is rapidly growing in importance and visibility. Soft computing and its applications to the Internet and bioinformatics are the principal themes of the FLINT-CIBI Workshop.

Participants in the CIBI Workshop come from many countries and work in many fields. We should like to extend to them our warm welcome, and are looking forward to stimulating discussions and exchange of ideas and information in Berkeley's friendly and cosmopolitan environment.

**FLINT-CIBI 2003; Organizing Committee
BISC PROGRAM, Director
UC BERKELEY
CALIFORNIA
USA
DECEMBER 15, 2003**

**2003 BISC FLINT-CIBI INTERNATIONAL JOINT WORKSHOP
ON
SOFT COMPUTING FOR INTERNET AND BIOINFORMATICS**

December 15-18, 2003
306 Soda Hall; HP Auditorium, EECS
Berkeley Initiative in Soft Computing (BISC)
University of California, Berkeley, California, USA

The basic ideas underlying soft computing in its current incarnation have links to many earlier influences, among them Prof. Zadeh's 1965 paper on fuzzy sets; the 1973 paper on the analysis of complex systems and decision processes; and the 1979 report (1981 paper) on possibility theory and soft data analysis.

BISC Program is the world-leading center for basic and applied research in soft computing. The principal constituents of soft computing (SC) are fuzzy logic (FL), neural network theory (NN) and probabilistic reasoning (PR), with the latter subsuming belief networks, evolutionary computing including DNA computing, chaos theory and parts of learning theory. Some of the most striking achievements of BISC Program are: fuzzy reasoning (set and logic), new soft computing algorithms making intelligent, semi-unsupervised use of large quantities of complex data, uncertainty analysis, perception-based decision analysis and decision support systems for risk analysis and management, computing with words, computational theory of perception (CTP), and precisiated natural language (PNL).

SOFT COMPUTING FOR INTERNET AND BIOINFORMATICS

The 2003 BISC International Workshop on "SOFT COMPUTING FOR INTERNET BIOINFORMATICS" will be held in the Computer Science Division of the University of California, Berkeley, from December 15 -19. UC Berkeley is a preeminent research institution in the proximity of Silicon Valley. At the dawn of the new millennium, we can expect dramatic increase in the use of fuzzy logic in the Bioinformatics and internet applications, since we have to deal with an increasing amount of data, that is mainly unstructured and designed for human access. Therefore, it is usually hard to extract relevant information automatically. These aspects will be reflected in the subjects treated at this 2003 BISC International Workshop. The main purpose of the Workshop is to draw the attention of the fuzzy logic community as well as the Bioinformatics and internet community to the fundamental importance of specific Internet-related problems. The BISC Program invites researchers from around the world who are interested in advancing the frontiers of the Bioinformatics and Internet by use of fuzzy logic and soft computing methods. The Workshop will provide a unique opportunity for the academic and corporate communities to address new challenges, share solutions, and discuss research directions for the future.

FUZZY LOGIC AND THE INTERNET

World Wide Web search engines have become the most heavily-used online services, with millions of searches performed each day. Their popularity is due, in part, to their ease of use. The central tasks for the most of the search engines can be summarize as 1) query or user information request- do what I mean and not what I say!, 2) model for the Internet, Web representation-web page collection, documents, text, images, music, etc, and 3) ranking or matching function-degree of relevance, recall, precision, similarity, etc.

Design of any new intelligent search engine should be at least based on two main motivations:

- The web environment is, for the most part, unstructured and imprecise and much of world knowledge consists of perceptions. To deal with information in the web environment what is needed is a logic that supports modes of reasoning which are approximate rather than exact. While searches may retrieve thousands of hits, finding decision-relevant and query-relevant information in an imprecise environment is a challenging problem, which has to be addressed. In addition, dealing with perception-based information is more complex and more effort intensive than dealing with measurement-based information. In this instance, as in many others, complexity is the price that has to be paid to achieve superior performance.
- Another, and less obvious, is deduction in an unstructured and imprecise environment given the huge stream of complex information. For Example; Search engines, with Google at the top, have many remarkable capabilities. But what is not among them is the deduction capability—the capability to synthesize an answer to a query by drawing on bodies of information which are resident in various parts of the knowledge base. It is this capability that differentiates a question-answering system, Q/A system for short, from a search engine. The main thrust of the fuzzy-logic-based approach to question-answering is that to achieve significant question-answering capability it is necessary to develop methods of dealing with the reality that much of world knowledge—and especially knowledge about underlying probabilities is perception-based.

SOFT COMPUTING FOR INTERNET AND BIOINFORMATICS

Bioinformatics is the area of science concerned with the application of computational methods to problems in biology, that include analysis, modeling and discovery from biological and medical data (e.g. genetic-, proteomic-, clinical). Bioinformatics is a symbiosis of several areas of science: biological sciences and molecular biology in particular, bio-chemistry, medicine, pharmacy, mathematics, physics, computer and information science, engineering.

DNA encodes the basic blue print of life; however it is not known how to interpret most of this information. The volume of data produced in bioinformatics increases exponentially with time and that requires powerful and intelligent methods to deal with this vast amount of information.

To address this issue the BISC-Special Interest Group in Computational Intelligence and Bioinformatics is formed. This SIG is concerned with the application of the methods of computational intelligence to the analysis, modeling and knowledge discovery in bioinformatics

FLINT-CIBI 2003

and to the development of new methods and systems for solving various challenging problems in the specified area. The methods of computational intelligence include methods of artificial intelligence, neural networks, fuzzy logic, evolutionary algorithms, methods for data mining, statistical methods and chaos analysis, algebraic methods and finite automata, and combination of them.

The BISC-SIG-CIBI's focus is to take advantage of the wealth of new information and develop new techniques and methodology for intelligent information analysis, knowledge management and knowledge discovery and rule-based reasoning in an environment that is imprecise, uncertain, complex and not known to our knowledge as of today. These tools include computational methods for the identification of functional elements in complex genomes and the identification of patterns in large datasets (for example, microarray data). The program also supports development of intelligent search engine and query engine with deductive capability and reasoning that can be widely shared among different databases in order to facilitate intelligent information retrieval from genomes databases and DNA sequences. These bioinformatics resources will allow the scientific community efficient access to genomic data, which will enable new types of analyses for the computer modeling and experimental validation of the complex genome networks that ultimately determine the causes of many human diseases. We believe that the information and knowledge extracted from such system-wide analyses using the computational intelligence will fundamentally alter the way of our life

The SIG is aiming at integrating efforts of biologists and computer scientists, graduate students and practitioners, who are interested in making progress both in terms of solving practical problems in bioinformatics and improving the existing computational techniques.

A GLIMPSE INTO THE FUTURE

The successful applications of soft computing and the rapid growth of BISC suggest that the impact of soft computing will be felt increasingly in coming years. Soft computing is likely to play an especially important role in science and engineering, but eventually its influence may extend much farther. In many ways, soft computing represents a significant paradigm shift in the aims of computing - a shift which reflects the fact that the human mind, unlike present day computers, possesses a remarkable ability to store and process information which is pervasively imprecise, uncertain and lacking in categoricity. In this perspective, what is important about BISC is that it provides a platform for the advancement of soft computing - a platform which lowers barriers between the constituents of soft computing and facilitates international cooperation on a global scale.

**FLINT-CIBI 2003
BISC PROGRAM
UC BERKELEY
CALIFORNIA
USA
DECEMBER 15, 2003**

MONDAY, DECEMBER 15, 2003**Soft Computing Techniques/Methodologies/Tools**

8:00 - 8:10	Welcome Prof. Shankar Sastry (EECS-Chair) Prof. Jitendra Malik (CS-Chair)	306 Soda Hall
8:10 - 8:15	Opening Remarks Prof. Lotfi A. Zadeh (BISC Director) Dr. Masoud Nikravesh (FLINT-CIBI Chair) Prof. Nikola Kasabov (FLINT-CIBI Chair)	
8:15 - 09:00	Prof. Mo Jamshidi <i>Applications of Soft Computing: Control, Robotics, Simulation and Image Processing</i>	306 Soda Hall
9:00 - 10:15	Dr. Dilip Kotak, Prof. William A. Gruver <i>Collaborative Agent Systems for Distributed Automation</i> Dr. John A. Meech <i>Computational Intelligence and Modelling in Materials Science and Processing</i>	
10:15 - 10:30	Break	
10:30 - 11:15	Prof. Sankar K. Pal <i>Machine Intelligence, Granular Computing and Data Mining</i>	
11:15 - 12:15	Prof. Saeid Nahavandi, Prof. Mohammad Khoshnevisan <i>An Evolutionary Computing Approach to Minimize Dynamic Hedging Error</i> Prof. Mohammad Khoshnevisan and Sukanto Bhattacharya <i>A Proposed Fuzzy Optimal Control Model to Minimize Target Tracking Error in a Dynamic Hedging Problem with a Multi-Asset, Best-of Option</i> Prof. Florentin Smarandache, Prof. Mohammad Khoshnevisan <i>Fuzzy Logic, Neutrosophic Logic, and Applications</i>	306 Soda Hall
12:15 - 1:30	Lunch Break	
1:30 - 2:15	Prof. Burhan Turksen <i>Belief and Fuzziness: Re-structuring Epistemology</i>	306 Soda Hall
2:15 - 3:15	Prof. Syozo Yasui	

	<u>Pruning, Selective Binding and Emergence of Internal Models: Applications to ICA and Analogical Reasoning</u> Dr. Souad Bensafi, Gamil Serag-Eldin, Dr. Masoud Nikravesh <u>BISC-DSS: Evolutionary Computing-based Multi-Aggregator Fuzzy Decision Trees and optimization</u>	
3:15 - 3:30	Break	
3:30 - 4:15	Dr. Juan C. Meza <u>Molecular Conformation, Uncertainty Quantification and Optimization in Statistics</u>	306 Soda Hall
4:15 - 5:30	Prof. Mihaela Ulieru <u>Emergent Information Infrastructures for Intelligent Distributed Systems : Mirroring the Universe and Life in Cyberspace</u> Prof. Thomas Sudkamp <u>Fuzzy temporal predicates in data mining</u> Victor Korotkikh (15 minutes) <u>Fuzzy Logic in the Modelling of Structures and Functions of Protein Sequences</u>	
5:30 - 5:45	Break	
5:45 - 6:30	Panel Discussion: Soft Computing W. Gruver D. Kotak J. Meza S. Pal S. Yasui B. Turksen	306 Soda Hall
7:00 - 9:00p	Reception	Wozniak Loun

TUESDAY, DECEMBER 16, 2003

Computation Intelligence Techniques/Methodologies/Tools

08:00 - 08:15	Opening Remark	
08:15 - 09:00	Dr. Enrique H. Ruspini <i>Qualitative Description of Complex Objects</i>	
09:00 - 10:15	Prof. Ruzena Baicsy <i>CIRTIS Program</i> Prof. Lotfi A. Zadeh <i>World Knowledge and Fuzzy Logic</i>	306 Soda Hall
10:15-10:30	Break	
10:30 -11:15	Dr. Frank Olken <i>Graph Data Management for Biology</i>	
11:15 -12:15	Prof. Lawrence J. Mazlack <i>Causality Recognition For Data Mining In An Inherently Ill Defined World</i> Prof. S-M. Vincent Tseng <i>A Framework for Effective Gene Expression Analysis and Biological Knowledge Discovery</i>	306 Soda Hall
12:15 - 1:30	Lunch Break	
1:30 - 2:15	Prof. Amit Sheth <i>Semantics Powered Bioinformatics: Semantic Search, Integration, Processes and Analytics</i>	
2:15 - 3:15	Prof. Miguel Delgado <i>Mining Fuzzy Association Rules: an Overview</i> Galina Korotkikh (15 minutes) <i>Fuzzy Spectral Patterns for Information Integration in a Search Engine</i> F. Olcay Cirit, Prof. Sema Alptekin, Dr. Masoud Nikravesh (15 minutes) <i>Automated Consumer Profiling using Fuzzy Query and Social Network Techniques</i>	306 Soda Hall
3:15 - 3:30	Break	
3:30 - 4:15	Dr. Horst D. Simon <i>Soft Computing and Current Trends in Supercomputing</i>	306 Soda Hall

PLINT-CIBI 2008

4:15 - 5:30	Prof. Mo Jamshidi <i>Some Opportunities for Soft Computing Applications</i> Prof. Paul P. Wang <i>Fuzzy Logic for Genetic Networking</i> Yonathan Asfaw, Dr. Martine De Cock, Dr. Masoud Nikravesh (15 minutes) <i>Performing region detection using simple fuzzy agents</i>	
5:30 – 5:45	Break	
5:45 - 6:30	Panel Discussion: Computational Intelligence R. Bajcsy M. Jamshidi E. Ruspini A. Sheth H. Simon L. A. Zadeh	306 Soda Hall

WEDNESDAY, DECEMBER 17. 2003

Bioinformatics

08:15 - 09:00	Prof. Nikola Kasabov <i>Knowledge Based Neurocomputing for Data Analysis and Knowledge Discovery in Bioinformatics</i>	
09:00 - 10:15	Dr. Chris Ding <i>Principal Component and Self-aggregation Clustering of Gene Expressions and Protein Interactions</i> Dr. Michael R. Berthold <i>Mining Imprecise Discriminative Molecular Fragments</i>	306 Soda Hall
10:15 - 10:30	Break	
10:30 - 11:15	Prof. C. Anthony Hunt <i>A Decentralization Method For Modeling The Multiple Levels Of Organization And Function Within The Liver</i>	
11:15 - 12:15	Prof. Voula C. Georgopoulos <i>Augmented Fuzzy Cognitive Maps based on Case Based Reasoning for Decisions in Medical Informatics</i> Prof. Adam Gaweda <i>Soft Computing Methods for Drug Dosing in Chronic Anemia</i>	306 Soda Hall
12:15 - 1:30	Lunch Break	
1:30 - 2:15	Prof. Nikola Kasabov <i>NeuCom – A Neuro Computing Environment for Evolving Intelligence</i>	
2:15 - 3:15	Dr. Yangqing Zhang <i>Hybrid Fuzzy Neural Networks for Protein Secondary Structure Prediction</i> Prof. Rainer Unland <i>Multi-Agent system and soft computing technologies as a fruitful symbiosis for reliable medical diagnoses</i>	306 Soda Hall
3:15 - 3:30	Break	
3:30 - 4:15	Prof. James Keller <i>Soft Computing Tools for Gene Similarity Measures in Bioinformatics</i>	306 Soda Hall
4:15 - 5:30	Dr. Isabelle Guyon	

PLINT-CIEA 2008

	<u>Applications of Variable and Feature Selection in Bioinformatics</u> Prof. Sunanda Mitra, Shuyu Yang <u>Adaptive Clustering in Medical Image Segmentation as a Diagnostic Tool</u>	
5:30 - 5:45	Break	
5:45 - 6:30	Panel Discussion: <i>Bioinformatics</i> M. Berthold V. C. Georgopoulos I. Guyon A. Hunt N. Kasabov Y. Zhang	306 Soda Hall
7:00 - 9:00	Dinner	Pomegranate Mediterranean*

* 1585 University Ave., Berkeley, CA 94703, (510) 665-5567

THURSDAY, DECEMBER 18, 2003

Search Engine & Internet

08:15 - 09:00	Dr. Peter Norvig <u>GoogleTM</u>	
09:00 - 10:15	Dr. Marcus Thint <u>Research and Applications of Soft Computing at BT Exact</u> Dr. Robert I. John <u>The Role of User Modeling in Information Retrieval from the WWW</u>	306 Soda Hall
10:15 - 10:30	Break	
10:30 - 11:15	Babak Hodjat <u>Evolution of the Laws that Deal with the Utilization of Information Networks</u>	
11:15 - 12:15	Prof. Alexander Rylov <u>Basic Principles of Information Monitoring Systems</u> Prof. Tomohiro Takagi, Dr. Masoud Nikraves <u>BeMySearch: Concept-Based Search Engine for Multimedia</u>	306 Soda Hall
12:00 - 01:30	Lunch Break	
01:30 - 02:15	Danis Yadegar <u>ARSIN</u>	
02:15 - 03:15	Dr. Charles Ortiz <u>Context-directed information access</u> Prof. Olfa Nasraoui <u>Web user profiling and web personalization based on data mining and soft computing</u>	306 Soda Hall
03:15 - 03:30	Break	
03:30 - 04:15	Dr. James G. Shanahan <u>Mining opinion</u>	
04:15 - 05:30	Dr. Martine De Cock, Sergio Guadarrama, Dr. Masoud Nikraves <u>Constructing Fuzzy Thesauri for the WWW : application to BeMySearch</u>	306 Soda Hall

PLANT-CIEI 2008

	<p><u>Andrey Kolobov, Daniel Kuo, Dr. Martine De Cock, Dr. Masoud Nikraves</u> <u>On Similarity Measures and Pruning Strategies for Fuzzy Ontologies</u></p> <p><u>Yuchun Tang, Yanqing Zhang</u> <u>Genetic Fuzzy Neural Agents Using Type-2 Fuzzy Reasoning for Intelligent Web Information Search Task</u></p>	
05:30 - 05:45	Break	
05:45 - 06:30	<p>Panel Discussion: <i>Internet and Industry</i> B. Hodjat P. Norvig T. Takagi M. Thint J. Shanahan D. Yadegar</p>	306 Soda Hall

FRIDAY, DECEMBER 19, 2003**TUTORIALS**

9:00 - 10:30	<u>Dr. M. Berthold</u> Intelligent Data Analysis	306 Soda Hall
10:30 - 10:45	Break	
10:45 - 12:00	<u>Dr. C. Ding</u> Bioinformatics	306 Soda Hall
12:00 - 1:30	Lunch Break	
1:30 - 3:00	<u>Dr. Isabelle Guyon</u> Feature Extraction and Feature Selection Challenge	306 Soda Hall
3:00 - 3:15	Break	
3:15 - 4:30	<u>Prof. L.A. Zadeh and Dr. M. Nikravesh,</u> Soft Computing Techniques/Methodologies/Tools	306 Soda Hall

MONDAY, DECEMBER 15

**SOFT COMPUTING
TECHNIQUES/METHODOLOGIES/TOOLS**

Abstracts of MONDAY, DECEMBER 15:

Soft Computing Techniques/Methodologies/Tools

Applications of Soft Computing: Control, Robotics, Simulation and Image Processing

Mo Jamshidi

One of the biggest challenges of any control paradigm is being able to handle complex systems under unforeseen uncertainties. A system may be called complex here if its dimension (order) is too high and its model (if available) is nonlinear, interconnected, and information on the system is uncertain such that classical techniques cannot easily handle the problem. Knowledge about such systems is a key attribute which is not often exploited for design and synthesis. Soft computing, a consortium of fuzzy logic, neuro-computing, genetic algorithms and genetic programming, has proven to be powerful tools for analysis and design of many complex systems. For such systems the size soft computing control architecture will be nearly infinite. Examples of complex systems are power networks, space robotic colonies, national air traffic control system, an integrated manufacturing plant, Hubble Telescope, a swarm of robotic agents, etc. In this talk a survey of many soft computing based applications in control, robotics, simulation and image processing will be given.

Collaborative Agent Systems for Distributed Automation

Dilip Kotak and William A. Gruver

Computer and communication technologies are very rapidly shrinking the world. Manufacturing and supply chains, the service industry and the infrastructure are becoming highly distributed and also highly integrated. This creates new challenges from the perspectives of systems, humans, and cybernetics. Fortunately, there is a significant emergence of technologies including monitoring and active identification, wired and wireless communications, and intelligent agent systems. Brought together effectively these technologies could complement each other and be integrated to provide unique solutions to meet the needs of decision-makers.

In the past decade a number of international groups have been formed to develop collaborative agent, holonic, and multi-agent systems. Within the Intelligent Manufacturing Systems Program, for example, there have been several such projects. However, many of these initiatives have already concluded and others are coming to conclusion.

This presentation will describe a new initiative of the IEEE SMC Society that consists of four core competency areas and four horizontal market application areas as follows:

Competency Areas

- Agent Monitoring and Self Identification
- Agent Communications
- Agent Collaboration and Coordination

- Agent Architecture

Application Areas

- Manufacturing and Supply Chains
- Service Automation
- Transportation Infrastructure
- Energy Infrastructure

We will provide examples of the research challenges in the core competency and application areas from manufacturing and distributed energy systems. Because agent communications systems requires a distributed and scalable hardware infrastructure, we will also describe our research in distributed wireless 802.11 communication that provides multi-hop, peer to peer connectivity with adaptive routing and priority scheduling. Finally, the presentation will identify opportunities for using soft computing techniques and related technologies in our research projects.

We shall encourage dialogue and discussions among the participants to explore opportunities for collaboration.

Computational Intelligence and Modelling in Materials Science and Processing

John A. Meech

To a great extent the future welfare of the world depends on developing new materials and processes. In all fields of endeavour there is need for materials with new properties that support the necessary advances. These fields encompass among others: robotics, computer hardware/software, automobiles and transportation, communication systems, medicine and biological systems, environmental protection, energy production and distribution, aircraft and aerospace, rapid transit systems and railroads, undersea exploration, building construction and other structures, missiles and weapon systems, commercial "white" products, etc. As each of these fields evolve into new areas of science and technology, demands are made for increased speed, increased productivity or bandwidth, increased storage capacities, increased strength or mechanical resistance, increased efficiency or effectiveness, adaptability to increased complexity, decreased size, decreased pollution levels and emissions, decreased costs, decreased cycle times, integration of existing processes, and doing this while incorporating "intelligence" within specific materials, processes or products.

This paper reviews some recent advances in the application of "intelligent" methods in the production of new materials and manufacture of new products. In some cases, the material embodies its "SMARTness" into an ability to self-adapt or respond to environmental changes. In other cases, the intellect is embodied into the processes used to derive the materials - primary processing, extraction, finishing, assembling, and delivery. Finally the use of "intelligence" to perform or support simulation models of new materials and/or their processes is presented. Both empirical data-derived models as well as First Principles modeling are discussed.

Machine Intelligence, Granular Computing and Data Mining

Sankar K. Pal

Emergence of data mining and knowledge discovery from pattern recognition point of view is illustrated. Significance of integrating various soft computing tools for efficient learning is described. Role of granular computing in data mining is given more emphasis. Two examples, demonstrating integrations of fuzzy sets, artificial neural networks, genetic algorithms and rough sets for efficient classification, rule generation and rule evaluation, and for granular case generation in case based reasoning problems, are provided along with their application specific merits on real life data. The significance of rough-fuzzy granulation for both reducing the computation time and improving the performance is explained. The talk concludes explaining the relation of rough-fuzzy case generation with the recently emerged computational theory of perception (CTP) and their applications in Web mining problems.

An Evolutionary Computing Approach to Minimize Dynamic Hedging Error

Saeid Nahavandi and Mohammad Khoshnevisan

The objective of our present paper is to derive a computationally efficient genetic pattern learning algorithm to evolutionarily derive the optimal rebalancing weights (i.e. dynamic hedge ratios) to engineer a structured financial product out of a multi-asset, best-of option. The stochastic target function is formulated as an expected squared cost of hedging (tracking) error which is assumed to be partly dependent on the governing Markovian process underlying the individual asset returns and partly on randomness i.e. pure white noise. A simple haploid genetic algorithm is advanced as an alternative numerical scheme, which is deemed to be computationally more efficient than numerically deriving an explicit solution to the formulated optimization model. An extension to our proposed scheme is suggested by means of adapting the Genetic Algorithm parameters based on fuzzy logic controllers.

A Proposed Fuzzy Optimal Control Model to Minimize Target Tracking Error in a Dynamic Hedging Problem with a Multi-Asset, Best-of Option

Mohammad Khoshnevisan and Sukanto Bhattacharya

The objective of our present paper is to derive a fuzzy optimal control model to explicitly derive the optimal rebalancing weights (i.e. dynamic hedge ratios) to engineer a structured financial product out of a multi-asset best-of option. The target function is actually the total cost of hedging (tracking) error over the investment horizon $t = 0$ to T , which is assumed to be partly dependent on the governing Markovian process underlying the individual asset returns and partly on randomness i.e. pure white noise. To derive

the necessary conditions for the fuzzy optimal control, we consider the problem: Minimize $\varepsilon = \int_0^T f_0(r, v, t) dt$, subject to the condition $\frac{d\mathbf{r}}{dt} = \mathbf{f}_1(r, v, t)$, where \mathbf{r} is the vector of n state variables i.e. the vector of returns on the n assets underlying the best-of option and \mathbf{v} is the vector of m fuzzy control variables i.e. the values of the replicating portfolio for m different choices of the dynamic hedge ratios. As the choice set is fuzzy, the fuzzy control vector is a subnormal fuzzy subset $[p_1/v_1, p_2/v_2 \dots p_n/v_n]$; p_j is the

membership grade of the j^{th} portfolio at time t .

Fuzzy Logic, Neutrosophic Logic, and Applications

Florentin Smarandache and M. Khoshnevisan

A Small Survey of the Evolution of Logics from Boolean to fuzzy and neutrosophic is presented. Afterwards the neutrosophic logic components are introduced followed by the definition of neutrosophic logic (NL), based on non-standard analysis, and neutrosophic logic connectors which are based on set operations.

Neutrosophic Logic Components:

Let T, I, F be standard or non-standard real subsets of $]0, 1^+]$,

with $\sup T = t_{\sup}, \inf T = t_{\inf},$

$\sup I = i_{\sup}, \inf I = i_{\inf},$

$\sup F = f_{\sup}, \inf F = f_{\inf},$

and $n_{\sup} = t_{\sup} + i_{\sup} + f_{\sup},$

$n_{\inf} = t_{\inf} + i_{\inf} + f_{\inf}.$

The sets T, I, F are not necessarily intervals, but may be any real sub-unitary subsets: discrete or continuous; single-element, finite, or (countably or uncountably) infinite; union or intersection of various subsets; etc. They may also overlap. The real subsets could represent the relative errors in determining t, i, f (in the case when the subsets T, I, F are reduced to points). In the paper, T, I, F , called *neutrosophic components*, will represent the truth value, indeterminacy value, and falsehood value respectively referring to neutrosophy, neutrosophic logic, neutrosophic set, neutrosophic probability, neutrosophic statistics. This representation is closer to the human mind reasoning. It characterizes/catches the *imprecision* of knowledge or linguistic inexactitude received by various observers (that's why T, I, F are subsets - not necessarily single-elements), *uncertainty* due to incomplete knowledge or acquisition errors or stochasticity (that's why the subset I exists), and *vagueness* due to lack of clear contours or limits (that's why T, I, F are subsets and I exists; in particular for the appurtenance to the neutrosophic sets). One has to specify the superior (x_{\sup}) and inferior (x_{\inf}) limits of the subsets because in many problems arises the necessity to compute them.

Definition of Neutrosophic Logic: A logic in which each proposition is estimated to have the percentage of truth in a subset T , the percentage of indeterminacy in a subset I , and the percentage of falsity in a subset F , where T, I, F are defined above, is called *Neutrosophic Logic*. We use a subset of truth (or indeterminacy, or falsity), instead of a number only, because in many cases we are not able to exactly determine the percentages of truth and of falsity but to approximate them: for example a proposition is between 0.3-0.4 true and between 0.6-0.7 false, even more imprecise: between 0.3-0.4 or 0.45-0.50 true (according to various analyzers), and 0.6 or between 0.66-0.70 false. The subsets are not necessary intervals, but any sets (discrete, continuous, open or closed or half-open/half-closed interval, intersections or unions of the previous sets, etc.) in accordance with the given proposition.

A subset may have one element only in special cases of this logic.

The differences between IFL and NL [and the corresponding Intuitionistic Fuzzy Set (IFS) and Neutrosophic Set (NS)] are:

a) Neutrosophic Logic can distinguish between *absolute truth* (truth in all possible worlds, according to Leibniz) and *relative truth* (truth in at least one world), because $NL(\text{absolute truth})=1^+$ while $NL(\text{relative truth})=1$. This has application in philosophy (see the neutrosophy). That's why the unitary standard interval $[0, 1]$ used in IFL has been extended to the unitary non-standard interval $]0, 1^+]$ in NL. Similar distinctions for absolute or relative falsehood, and absolute or relative indeterminacy are allowed in NL.

b) In NL there is no restriction on T, I, F other than they are subsets of $]0, 1^+]$, thus:
 $0 \leq \inf T + \inf I + \inf F \leq \sup T + \sup I + \sup F \leq 3^+$. This non-restriction allows paraconsistent, dialetheist, and incomplete information to be characterized in NL {i.e. the sum of all three components if they are defined as points, or sum of superior limits of all three components if they are defined as subsets can be >1 (for paraconsistent information coming from different sources) or < 1 for incomplete information}, while that information can not be described in IFL because in IFL the components T (truth), I (indeterminacy), F (falsehood) are restricted either to $t+i+f=1$ or to $t^2 + f^2 \leq 1$, if T, I, F are all reduced to the points t, i, f respectively, or to $\sup T + \sup I + \sup F = 1$ if T, I, F are subsets of $[0, 1]$.

c) In NL the components T, I, F can also be *non-standard* subsets included in the unitary non-standard interval $]0, 1^+]$, not only *standard* subsets included in the unitary standard interval $[0, 1]$ as in IFL.

d) NL, like dialetheism, can describe paradoxes, $NL(\text{paradox}) = (1, I, 1)$, while IFL can not describe a paradox because the sum of components should be 1 in IFL.

Applications of Fuzzy and Neutrosophic Logics. Further one gives examples of fuzzy and neutrosophic logics used in the reconciliation between theoretical and market prices of long-term options contracts, in extension of the MASS model as a cost-optimal relative allocation of facilities technique by the incorporation of neutrosophic statistics and the DSm (Dezert-Smarandache) combination rule, and in conditional probability of actually detecting a financial fraud – a neutrosophic extension to the application of Benford's first-digit law.

Original work consists in the definition of neutrosophic logic (NL) and neutrosophic logical connectors as an extension of intuitionistic fuzzy logic (IFL) and the comparison among NL and other logics, especially the IFL, and in their applications.

Belief and Fuzziness: Re-structuring Epistemology

I. Burhan Turksen

We review briefly the basic components of the philosophical grounding of fuzzy theory from ontology to epistemology to applications. Within the context of epistemology, the notion of "belief" plays an essential role in foundationalism. In order to generalize such foundational concerns of epistemology, we propose that it be re-structured to include the language of fuzzy theory. Thus "Precisiated Natural Language, PNL" proposed by Lotfi A. Zadeh ought to be considered the sixth link in the evolutionary chain of scientific languages where meta-languages, speech, writing, mathematics, and computing languages represent the usual evolutionary links. It is well understood that all forms of language have both a communications and informatics dimension that facilitates human thought and decision making and therefore have an impact on computing with words and perceptions. In this regards, we review the connection between "belief" and "fuzzy sets" as well as classical probability and fuzzy sets.

Pruning, Selective Binding and Emergence of Internal Models: Applications to ICA and Analogical Reasoning

Syozo Yasui

Pruning of multi input/output neural networks is discussed and a pruning algorithm called CSDF is described. CSDF acts to induce internal models as a result of redundancy elimination and selective

bindings. CSDF is used in a new ICA method based on an auto-encoder performing sensor-signal identity mapping. An internal model of the external signalmixing situation emerges due to the CSDF pruning, and the hidden units that survive the CSDF pruning reconstruct the blind source signals. This ICA method which requires no pre-processing such as whitening is characterized by its high adaptability and robustness, as is demonstrated by trouble cases such as sudden increase of the source signals, sudden failure of sensors and so on. As another example, CSDF is applied in a neural network for analogical learning/inference. Internal abstraction models together with abstraction/de-abstraction bindings are generated as a result of the CSDF structural learning coupled with the backpropagation training. The internal abstraction model acts as an attractor for new relevant dataset, a process corresponding to analogical memory retrieval.

BISC Decision Support System (BISC-DSS)

Masoud Nikravesh

A key component of any autonomous system is a decision module, which is capable of handling large volume of data at high speed and high reliability. This paper will be focused, in the main, on the development of a decision module, which is capable of functioning in an environment of imprecision, uncertainty and imperfect reliability. There will be two principal tasks: Task A, which will be aimed at the development of novel methods of analysis and design; and Task B, which will be focused on the development of a decision-support system for ranking of decision alternatives. Following is an outline of the principal issues, which will be addressed in Task A.

In this study, we introduced the BISC decision support system as an alternative for ranking and predicting the risk which utilizes an imprecise and subjective process. The BISC decision support system key features are 1) intelligent tools to assist decision-makers in assessing the consequences of decision made in an environment of imprecision, uncertainty, and partial truth and providing a systematic risk analysis, 2) intelligent tools to be used to assist decision-makers answer "What if Questions", examine numerous alternatives very quickly and find the value of the inputs to achieve a desired level of output, and 3) intelligent tools to be used with human interaction and feedback to achieve a capability to learn and adapt through time.

Acknowledgements

Funding for this research was provided by the British Telecommunication (BT) and the BISC Program of UC Berkeley.

References

M. Nikravesh (2001) Credit Scoring for Billions of Financing Decisions, Joint 9th IFSA World Congress and 20th NAFIPS International Conference. IFSA/NAFIPS 2001 "Fuzziness and Soft Computing in the New Millenium", July 25-28, 2001

M. Nikravesh and B. Azvine, "Fuzzy queries, search, and decision support system", *Soft Computing*, vol. 6, pp. 373-399, 2002.

M. Nikravesh, B. Azvine, R. Yager and Lotfi A. Zadeh (2003b) "New Directions in Enhancing the power of the Internet", Editors:, to be published in the Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer (August 2003)

M. Nikraves and Lotfi A. Zadeh, "Fuzzy Logic an the Internet", to be published in the Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer (August 2003)

M. Nikraves, B. Azvine, R. Yagar and Lotfi A. Zadeh, "New Directions in Enhancing the power of the Internet", to be published in the Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer (Dec 2003)

Zadeh L. A. (1999), From Computing with Numbers to Computing with Words -- From Manipulation of Measurements to Manipulation of Perceptions, IEEE Transactions on Circuits and Systems, 45, 105-119, 1999.

Zadeh, L. and Kacprzyk, J. (eds.) (1999a), *Computing With Words in Information/Intelligent Systems 1: Foundations*, Physica-Verlag (1999a).

Zadeh, L. and Kacprzyk, J. (eds.) (1999b), *Computing With Words in Information/Intelligent Systems 2: Applications*, Physica-Verlag,(1999b).

Zadeh, L.A. (1996) Fuzzy Logic = Computing with Words," *IEEE Trans. on Fuzzy Systems* (1996) 4, 103-111.

Zadeh, L. A. and M. Nikraves (2002), Perception-Based Intelligent Decision Systems, AINS; ONR Summer 2002 Program Review, 30 July-1 August, UCLA.

Web-Based BISC Decision Support System **Gamil Serag-Eldin and Masoud Nikraves**

Most of the existing search systems 'software' are modeled using crisp logic and queries. We introduce fuzzy querying and ranking as a flexible tool allowing approximation where the selected objects do not need to match exactly the decision criteria resembling natural human behavior. The model consists of five major modules: the Fuzzy Search Engine, the Application Templates, the User Interface, the Database and the Evolutionary Computing. The system is designed in a generic form to accommodate more diverse applications and to be delivered as stand-alone software to academia and businesses.

In this study, we introduced fuzzy query, fuzzy aggregation, evolutionary computing and the BISC decision support system as an alternative for ranking and predicting the risk for credit scoring, university admissions, and several other applications, which currently utilize an imprecise and subjective process. The BISC decision support system key features are 1) intelligent tools to assist decision-makers in assessing the consequences of decision made in an environment of imprecision, uncertainty, and partial truth and providing a systematic risk analysis, 2) intelligent tools to be used to assist decision-makers answer "what if", questions examine numerous alternatives very quickly and find the value of the inputs to achieve a desired level of output, and 3) intelligent tools to be used with human interaction and feedback to achieve a capability to learn and adapt through time. In addition, the following important points have been found in this study: 1) no single ranking function works well for all contexts, 2) most similarity measures work about the same regardless of the model, 3) there is little overlap between successful ranking functions, and 4) the same model can be used for other applications such as the design of a more intelligent search engine which includes the user's preferences and profile (Nikraves, 2001a and 2001b).

Acknowledgment

Funding for this research was provided by the British Telecommunication (BT) and the BISC Program of UC Berkeley.

References

Fagin R. (1998) Fuzzy Queries in Multimedia Database Systems, Proc. ACM Symposium on Principles of Database Systems, pp. 1-10.

Fagin R. (1999) Combining fuzzy information from multiple systems. J. Computer and System Sciences 58, pp 83-99.

J. R. Koza, Genetic Programming : On the Programming of Computers by Means of Natural Selection, Cambridge : MIT Press 1992, 819 pages.

John H. Holland. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence. MIT Press, 1992. First Published by University of Michigan Press 1975.

Mizumoto M. (1989) Pictorial Representations of Fuzzy Connectives, Part I: Cases of T-norms, T-conorms and Averaging Operators, Fuzzy Sets and Systems 31, pp. 217-242.

Nikraves M. (2001a) Perception-based information processing and re-trieval: application to user profiling, 2001 research summary, EECS, ERL, University of California, Berkeley, BT-BISC Project. (<http://zadeh.cs.berkeley.edu/> & <http://www.cs.berkeley.edu/~nikrav>

<http://www.bisc.cs.berkeley.edu/>).

Nikraves M. (2001b) Credit Scoring for Billions of Financing Deci-sions, Joint 9th IFSA World Congress and 20th NAFIPS International Conference. IFSA/NAFIPS 2001 "Fuzziness and Soft Computing in the New Millenium", Vancouver, Canada, July 25-28, 2001.

Masoud Nikraves and Ben Azvine (2002), Fuzzy Queries, Search, and Decision Support System, Journal of Soft Computing, Volum 6, # 5, August 2002.

Masoud Nikraves, B. Azvine, R. Yagar, and Lotfi A. Zadeh (2003) "New Directions in Enhancing the power of the Internet", Editors:, to be published in the Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer (August 2003).

BISC-DSS: Evolutionary Computing-based Multi-Aggregator Fuzzy Decision Trees and optimization

Souad Souafi-Bensafi and Masoud Nikraves

In the world of information processing, we are faced with increasingly complex multi-domain problems containing either real-world or computer-generated data. To consider these problems the classical data processing tools may not be sufficient and more advanced approaches need to be developed. A unified approach based on Soft Computing will help in solving such problems by combining methodologies (Fuzzy Logic, Neuro-Computing, Evolutionary Computing and Probabilistic Reasoning), which collectively provide a foundation for the conception of Intelligent Systems.

Our aim is to develop intelligent computing techniques that address the problem of multi-criteria decision making dealing with subjective and imprecise data. This kind of problem requires conception of intelligent systems able to replace a human with expertise in a specific domain in a decision making process. So, the intelligent system should take into account the subjective and imprecise character of data on the one hand, and represent the user or expert's preferences and knowledge on the other hand. For this purpose, we developed a generic multi-criteria model based on fuzzy logic concepts for decision support systems. Our goal is to build such a model by 1) fitting real-world data and 2) representing the preferences of specific-domain users or experts. Toward this end, we used Evolutionary Computation techniques. Initially, we worked on a first order aggregation model and performed its learning using Genetic Algorithms, in which these preferences have been represented by a weighting vector associated with the variables involved in the aggregation process. This has been used in a specific application related to university admissions. Then, we developed a more advanced multi-aggregation model based on a hierarchical decision trees and for the learning process of this model, we developed a technique inspired from Genetic Programming. In this model tree nodes represent aggregators, terminals or leaves correspond to variables, and weight values are added to the children branches for each aggregator. The aggregation result overall the variables is then obtained by running recursively the root aggregator of the tree.

The parameters characterizing this multi-aggregation model are aggregators, weights and their combination in form of a tree structure. In this case, the learning process has to find the optimal combination of these parameters based on training data. In this learning process, the evolution principle remains the same as in a conventional GP but the DNA encoding needs to be defined according to the considered problem. We need to define a more complex tree structure representing the multi-aggregation model. In addition to the weights that have to be added to the classical tree, the nodes that represent the aggregators require a variable number of arguments. This is because the number of arguments cannot be known before the tree creation. Therefore, during the evolution process, trees are generated randomly by selecting aggregators for the nodes, and at the same time, the corresponding numbers of arguments are randomly chosen. Moreover, weight values are fixed randomly in each branch of the tree during its creation.

These encoding properties allow a large search space to solve our problem. If we need to simplify this tree structure according to some application constraints, we can add these constraints to the problem specification and they will be checked during the tree generation.

We pursue this work by considering many other applications and we aim at carrying out the multi-aggregation model base on decision trees. We defined this model as a basis for a more general and a more complex form of the considered problems which operates with linguistic variables. Our approach is a first attempt toward the use of the Computing with Words and Perceptions.

Molecular Conformation, Uncertainty Quantification and Optimization in Statictics **Juan C. Meza**

As head of the High Performance Computing Research Department, Dr. Meza oversees work in scientific data management, visualization, computational grids, numerical algorithms, and application development. He is responsible for developing short- and long-term research and development plans and proposing new technology directions.

Prior to joining Lawrence Berkeley National Laboratory, Dr. Meza held the position of Distinguished Member of the Technical Staff at Sandia National Laboratories and served as the manager of the

Computational Sciences and Mathematics Research department. In this capacity, he acted as the Research Foundation Network Research program manager, the ASCI Problem Solving Environment Advanced Software Development Environment program manager and served as a member of the Sandia California site Research Council. Dr. Meza worked at Sandia since 1987 in various other capacities. As the manager of Distributed Computing, he helped in developing the Distance/Distributed Computing program plan and the ASCI Verification and Validation program plan. He was also the technical liaison to the Department of Energy's Strategic Alliances program for the Stanford Center for Integrated Turbulence Simulations. Dr. Meza was the project lead for the ASCI Problem-solving Environment/Application Development Support and the DOE 2000 Advanced Computational Toolkit. He was also a principal investigator for the Mathematical, Information and Computational Sciences project on Large-scale Optimization for Scientific and Engineering Design.

Dr. Meza serves on numerous external committees including, the Department of Energy's Advanced Scientific Computing Advisory Committee, NPACI's External Advisory Committee, MSRI's Human Resources Advisory Committee and was formerly on the Institute for Mathematics and its Applications' Board of Governors.

Dr. Meza holds Ph.D. and M.S. degrees in Mathematical Sciences from Rice University. He also holds M.S. and B.S. degrees in Electrical Engineering (cum laude) from Rice University. His current research interests include parallel nonlinear optimization and methods for uncertainty quantification.

Emergent Information Infrastructures for Intelligent Distributed Systems : Mirroring the Universe and Life in Cyberspace

Mihaela Ulieru

We propose a fuzzy-evolutionary approach to self-organization that emulates social behavior and immunity in Cyberspace and has applications to emergency response management (Fig. 1) distributed manufacturing, medical informatics and Cybersecurities. Organized in a nested hierarchy (Fig. 2) distributed throughout the network the system consists of a hybrid mixture of static and mobile agents behaving like a *Cyberorganism* capable to react to unexpected changes / attacks in an optimal manner. Computational intelligence techniques endow the MAS with learning and discovery capabilities. By 'cloning' real-life entities into software agents, the proposed paradigm can be easily extended to the creation of emergent dynamic information infrastructures that are autonomous and proactive, capable of ensuring ubiquitous (optimal) resource discovery and allocation while at the same time self-organizing their resources to optimally accomplish the desired objectives.

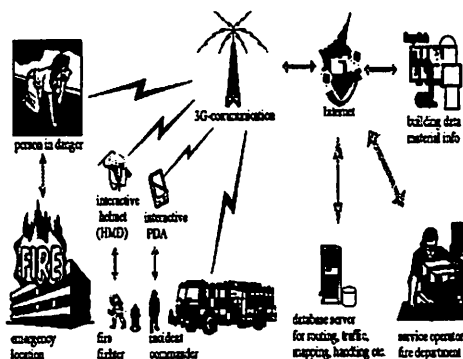


Fig. 1: Emergency Scenario

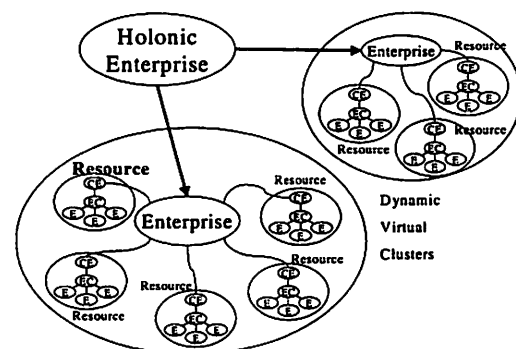


Fig. 2: Holonic System

Fuzzy temporal predicates in data mining

Thomas Sudkamp

The structure and precision of the data affects the type of relationships that can be analyzed by knowledge discovery and data mining algorithms. Data items are frequently sets, tuples in relational databases, or elements from a temporal stream. The former two representations consider individual data items as independent while temporality induces an ordering on the data. When data items are independent, data mining algorithms seek to discover frequently recurring associations within individual items. When distinct data items are related, the task of knowledge discovery is to identify relationships between the items.

In this presentation we review the types of relationships produced by standard data mining techniques. In particular we are interested in the representation of imprecision in both the data and the resulting associations. For independent data, fuzzy sets have been used to partition attribute domains to extend quantitative association rules to imprecise categories. Temporal relationships consider data from one or multiple sources in which the sole link between the data may be the time value. Association rules for temporal data may include imprecise durations and temporal constraints. Temporal durations specify the time an event is required to continue to satisfy a proposition, for example event *A* occurs for "a long period" of time. A constraint specifies the temporal relationship between the occurrence of distinct events; *B* occurs shortly after *A*. We will discuss the constraints, data representations, and search strategy needed for analyzing fuzzy temporal associations in data from multiple sources.

Fuzzy Logic in the Modelling of Structures and Functions of Protein Sequences

Victor Korotkikh

A major aim of bioinformatics is to contribute to the understanding of the interrelationships between protein sequence, structure and function. In this context approaches that can suggest structures and processes associated with sequences may be of special interest. Moreover, if such a structure and a process turn to work consistently with observations then it is become even possible to speculate about them as laws of the phenomena. It appears then naturally to ask why the structure and the process are special to the extent that they are able to describe the nature of living things. In particular, whether they are fundamental or it is possible to explain them in terms of deeper structures and processes. This may explain why irreducible structures and processes would come to play a special role in the modeling. Indeed, there can be no further reductions possible in their instances only.

In the paper we present an approach within which sequences become associated with a hierarchical structure that has integer relations as the elements. The ultimate building blocks of the structure are the integers. The other elements are made of them by an organizing principle. The structure is irreducible in the sense that its existence is based on the integers only and the other elements are made in the full control of arithmetic. This reduces the explanations about the elements, and why it is they are made the way they are, to explanations that reveal why it is integers exist and arithmetic holds the way it does.

A new type of process can be revealed as the result of an interpretation of the relationships existing between the elements of the structure. In particular, the relationships between the integer relations are specified the organizing principle, which describes how an element of one level of the structure can be

produced from elements of the previous one. Notably, this may be interpreted as a formation process. The structure has a natural ability to integrate processes. This may be helpful to model how the function of protein sequences as a new whole emerges from the functions of these sequences as separate entities.

In general, the approach is contrasting, because it relies on structures, with integer relations as the elements, in order to model structures and functions of protein sequences. Although the structures and the processes are rigorously defined mathematically, in computations however they resist to be easily described and processed. In the talk we discuss that fuzzy logic can propose an efficient solution to this problem.

Panel Discussion: MONDAY, DEC15

Soft Computing

The successful applications of soft computing and the rapid growth of BISC suggest that the impact of soft computing will be felt increasingly in coming years. Soft computing is likely to play an especially important role in science and engineering, but eventually its influence may extend much farther. In many ways, soft computing represents a significant paradigm shift in the aims of computing - a shift which reflects the fact that the human mind, unlike present day computers, possesses a remarkable ability to store and process information which is pervasively imprecise, uncertain and lacking in categoricity. In this perspective, what is important about BISC is that it provides a platform for the advancement of soft computing - a platform which lowers barriers between the constituents of soft computing and facilitates international cooperation on a global scale.

Panelists:

W. Gruver, D. Kotak, J. Meza, S. Pal, S. Yasui and B. Turksen

Moderator:

Masoud Nikravesh; University of California, Berkeley

TUESDAY, DECEMBER 16

**COMPUTATION INTELLIGENCE
TECHNIQUES/METHODOLOGIES/TOOLS**

Abstracts of TUESDAY, DECEMBER 16:

Computation Intelligence Techniques/Methodologies/Tools

Qualitative Description of Complex Objects

Enrique H. Ruspini

Enrique H. Ruspini received his degree of Licenciado en Ciencias Matemáticas from the University of Buenos Aires, Argentina, and his doctoral degree in System Science from the University of California at Los Angeles. Prior to joining SRI (formerly *Stanford Research Institute*), Dr. Ruspini held positions at the University of Buenos Aires, the University of Southern California, UCLA's Brain Research Institute, and Hewlett-Packard Laboratories.

Dr. Ruspini is one the earliest contributors to the development of fuzzy-set theory and its applications, having introduced its use to the treatment of numerical classification and clustering problems. He has also made significant contributions to the understanding of the foundations of fuzzy logic and approximate-reasoning methods. His recent research has focused on the application of fuzzy-logic techniques to the development of systems for intelligent control of teams of autonomous robots, information retrieval, qualitative description of complex objects, and knowledge discovery in large databases.

Dr. Ruspini, who has lectured extensively in the United States and abroad and is the author of over 100 original research papers, is a Fellow of the Institute of Electrical and Electronics Engineers, a First Fellow of the International Fuzzy Systems Association, a Fulbright Scholar, and a SRI Institute Fellow. Dr. Ruspini was the General Chairman of the Second IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'93) and of the 1993 IEEE International Conference on Neural Networks (ICNN'93). He is one of the founding members of the North American Fuzzy Information Processing Society and the recipient of that society's King-Sun Fu Award.

Dr. Ruspini is a member of the IEEE Board of Directors (Division X, Systems and Control), the Past-President (President-2001) of the IEEE Neural Networks Council and its past Vice-president of Conferences. Dr. Ruspini, who has led numerous IEEE technical, educational, and organizational activities, is also a member of the Transnational Committee of the IEEE Technical Activities Board. During his tenure as President of the IEEE Neural Networks Council, he led the successful effort for its transition to Society status.

Dr. Ruspini is also the Editor in Chief (together with P.P.Bonissone and W. Pedrycz) of the Handbook of Fuzzy Computation, published by Institute of Physics Publishing, a member of the Advisory Boards of the IEEE Transactions on Fuzzy Systems, and of the International Journal of Fuzzy Systems, and a member of the Editorial Board of the International Journal of Approximate Reasoning, the International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems, Fuzzy Sets and Systems, the Journal of Applied Nonstandard Logics, Mathware and Soft Computing, Soft/Ware, and the Journal of Advanced Computational Intelligence and Intelligent Informatics.

World Knowledge and Fuzzy Logic

Lotfi A. Zadeh

Informally, world knowledge is the knowledge acquired through experience, education and communication. World knowledge has a position of centrality in human cognition, and especially in summarization, assessment of relevance, deduction and search.

Centrality of world knowledge in human cognition entails its centrality in issues related to web intelligence. In the existing literature, world knowledge is dealt with through the use of knowledge-representation methods based on bivalent logic. The problem is that much of world knowledge, e.g., "California has a temperate climate" and "Most Finns are honest" is perception-based and/or dispositional. Such knowledge does not fit the conceptual structure of bivalent logical systems -- logical systems which are intolerant of imprecision, uncertainty and partial truth.

A thesis which is advanced in this talk is that to deal with world knowledge what is needed is fuzzy logic and, more particularly, computational theory of perceptions, dispositional logic and precisiated natural language. An outline of fuzzy-logic-based approach to world knowledge is presented and illustrated by examples.

From Search Engines to Question-Answering Systems

Lotfi A. Zadeh

Search engines, with Google at the top, have many remarkable capabilities. But what is not among them is the deduction capability - the capability to synthesize an answer to a query by drawing on bodies of information which are resident in various parts of the knowledge base. It is this capability that differentiates a question-answering system, Q/A system for short, from a search engine. Upgrading a search engine to a Q/A system is a complex, effort-intensive, open-ended problem. Semantic Web and related systems for upgrading quality of search may be viewed as steps in this direction. But what may be argued, as is done in the following, is that existing tools, based as they are on bivalent logic and probability theory, have intrinsic limitations. The principal obstacle is the nature of world knowledge.

The centrality of world knowledge in human cognition, and especially in reasoning and decision-making, has long been recognized in AI. The Cyc system of Douglas Lenat is a repository of world knowledge. The problem is that much of world knowledge consists of perceptions. More specifically, perceptions are f-granular in the sense that (a) the boundaries of perceived classes are fuzzy; and (b) the perceived values of attributes are granular, with a granule being a clump of values drawn together by indistinguishability, similarity, proximity or functionality. What is not widely recognized is that f-granularity of perceptions put them well beyond the reach of computational bivalent-logic-based theories.

Dealing with world knowledge needs new tools. A new tool which is suggested for this purpose is the fuzzy-logic-based method of computing with words and perceptions (CWP), with the understanding that perceptions are described in a natural language. A concept which plays a key role in CWP is that of Precisiated Natural Language (PNL). It is this language that is the centerpiece of our approach to reasoning and decision-making with world knowledge.

A concept which plays a key role in organization of world knowledge is that of an epistemic (knowledge-directed) lexicon (EL). Basically, an epistemic lexicon is a network of nodes and weighted links, with node i representing an object in the world knowledge database, and a weighted link from node i to node j representing the strength of association between i and j . The name of an object is a word or a composite word, e.g., car, passenger car or Ph.D. degree. An object is described by a relation or relations whose fields are attributes of the object. The values of an attribute may be granulated and associated with granulated probability and possibility distributions. For example, the values of a granular attribute may be labeled small, medium and large, and their probabilities may be described as low, high and low, respectively. Relations which are associated with an object serve as PNL-based descriptions of the world knowledge about the object. For example, a relation associated with an object labeled Ph.D. degree may contain attributes labeled Eligibility, Length.of.study, Granting.institution, etc. The knowledge associated with an object may be context-dependent. What should be stressed is that the concept of an epistemic lexicon is intended to be employed in representation of world knowledge - which is largely perception-based - rather than Web knowledge, which is not.

In conclusion, the main thrust of the fuzzy-logic-based approach to question-answering which is outlined in this abstract, is that to achieve significant question-answering capability it is necessary to develop methods of dealing with the reality that much of world knowledge is perception-based. Dealing with perception-based information is more complex and more effort-intensive than dealing with measurement-based information. In this instance, as in many others, complexity is the price that has to be paid to achieve superior performance.

Graph Data Management for Biology

Frank Olken

In this talk we shall discuss the use of graph data management for a variety of biological applications. Simple graphs are composed of a set of nodes and a binary relation comprising the edges which connect pairs of nodes.

We shall emphasize applications related to the representation and querying of biopathways databases, e.g., metabolic pathways, signal transduction pathways, and genetic regulatory networks. Other potential applications of graph data management to biology include: chemical structure graphs, protein interaction networks, phylogenetic trees, taxonomies of chemicals, proteins, enzymes and diseases, partonomies (e.g., in anatomy), topological adjacency relations (in image analysis), contact graphs (for 3D protein structure), bibliographic citation graphs, food webs, biogeochemical cycles, gene clusterings, partial order graphs for DNA multiple sequence alignments, genetic maps, operon and regulon structures, sequence overlap graphs for shotgun DNA sequencing, database schemas and mappings among schemas, data provenance (lineage), hypertext, semantic web applications, laboratory protocols, etc.

Graph data models for biology come in a number of variants: undirected and directed graphs, simple graphs, nested graphs, multigraphs, and hypergraphs. We will mention these variants and illustrate their applications.

Graph data management offers two major advantages for biopathways applications: naturalness of representation of pathway data, and ease of querying pathway data. It is the latter issue which is more important.

Graph data management systems permit users to frame queries in terms of graph operations, e.g., subgraph isomorphism, shortest paths, etc. which would be difficult to express or compute in conventional (relational) DBMS systems. We discuss a number of graph queries in the talk, e.g., subgraph homomorphism queries.

Graph data management systems typically treat individual fragments of the database more homogeneously (as either nodes or edges) than relational databases which partition the database into many specialized relations. In a GDMS the analog of relation structure is encoded as edges which indicate types of nodes. While the relational storage structures offer advantages in performance on fixed structure queries, the homogeneous graph data model is much easier to use in posing queries which allow paths to span many different possible relations (node types). It is this storage homogeneity which facilitates pattern matching and path queries in graph databases. In contrast, similar queries in relational DBMS involve large numbers of union queries over the various possible relations which might participate in a path or subgraph pattern match. We will discuss this issue and (briefly) some related comparisons to logic and object oriented database management systems.

We will illustrate the talk with references to some major biopathways databases. Time permitting we will also mention the role of RDF (Resource Description Framework) as a graph data model. We will conclude with a brief survey of some alternative approaches to implementation of graph database management systems.

This is joint work with Kevin D. Keck and Vijaya Natarajan (both at LBNL). Further information on the Biopathways Graph Data Manager Project may be found at <http://www.lbl.gov/~olken/graphdm/graphdm.htm> The work is funded by DARPA Biocomp Program (via the Biospice Project at LBNL (PI: A. Arkin)), and DOE Genomes to Life Program (via VIMSS GTL project at LBNL (PIs: A. Arkin and T. Hazen) and Synechococcus GTL project at Sandia National Lab (PI: G. Heffelfinger, LBNL PI: A. Shoshani). Submitted by Frank Olken on 2003-12-08.

Causality Recognition For Data Mining In An Inherently Ill Defined World

Lawrence J. Mazlack

1 Introduction

Commonsense causal reasoning occupies a central position in human reasoning. It plays an essential role in human decision-making. Considerable effort has been spent examining causation. Philosophers, mathematicians, computer scientists, cognitive scientists, psychologists, and others have formally explored questions of causation beginning at least three thousand years ago with the Greeks.

Whether causality can be recognized at all has long been a theoretical speculation of scientists and philosophers. At the same time, in our daily lives, we operate on the commonsense belief that causality exists.

Causal relationships exist in the commonsense world. If an automobile fails to stop at a red light and there is an accident, it can be said that the failure to stop was the accident's cause. However, conversely, failing to stop at a red light is not a certain cause of a fatal accident; sometimes no accident of any kind occurs. So, it can be said that knowledge of some causal effects is imprecise. Perhaps, complete knowledge of all possible factors might lead to a crisp description of whether a causal effect will occur. However, in our commonsense world, it is unlikely that all possible factors can be known. What is needed is a method to model imprecise causal models.

Another way to think of causal relationships is counterfactually. For example, if a driver dies in an accident, it might be said that had the accident *not* occurred; they would still be alive.

Our common sense understanding of the world tells us that we have to deal with imprecision, uncertainty and imperfect knowledge. This is also the case of our scientific knowledge of the world. Clearly, we need an algorithmic way of handling imprecision if we are to computationally handle causality. Models are needed to algorithmically consider causes. These models may be symbolic or graphic. A difficulty is striking a good balance between precise formalism and commonsense imprecise reality

1.1 Data mining, brief introduction

Data mining is an advanced tool for managing large masses of data. It analyzes data previously collected. It is *secondary* analysis. Secondary analysis precludes the possibility of experimentally varying the data to identify causal relationships.

There are several different data mining products. The most common are *conditional rules* or *association rules*. Conditional rules are most often drawn from induced trees while association rules are most often learned from tabular data. Of these, the most common data mining product is association rules; for example:

- *Conditional rule:*

IF Age < 20

THEN Income < \$10,000

with {belief = 0.8}

- *Association rule:*

Customers who buy beer and sausage

also tend to buy mustard

with {confidence = 0.8}

in {support = 0.15}

At first glance, these structures seem to imply a causal or cause-effect relationship. That is: A customer's purchase of both sausage and beer causes the customer to also buy mustard. In fact, when typically developed, association rules do not *necessarily* describe causality. Also, the strength of causal dependency may be very different from a respective association value. All that can be said is that associations describe the strength of joint co-occurrences. Sometimes, the relationship might be causal; for example, if someone eats salty peanuts and then drinks beer, there is probably a causal relationship. On the other hand, if a crowing rooster probably does not cause the sun to rise.

1.2 Naive association rules can lead to bad decisions

One of the reasons why association rules are used is to aid in making retail decisions. However, simple association rules may lead to errors. It is common for a food store to put one item on sale and then to raise the price of another item whose purchase is assumed to be associated. This may work if the items are truly associated; but it is problematic if association rules are blindly followed.

Example: At a particular store, a customer buys:

- *hamburger* 33% of the time
- *hot dogs* 33% of the time
- both *hamburger* and *hot dogs* 33% of the time
- *sauer-kraut* only if *hot dogs* are also purchased

This would produce the transaction matrix:

	hamburger	hotdog	sauerkraut
t ₁	1	1	1

t_2	1	0	1
t_3	0	1	0

This would lead to the associations:

- (hamburger, hot dog) = 0.5
- (hamburger, sauer-kraut) = 0.5
- (hot dog, sauer-kraut) = 1.0

If the merchant:

- reduced price of hamburger
- raised price of sauer-kraut to compensate
- the offset pricing compensation would not work.

1.3 False causality

Complicating causal recognition are the many cases of false causal recognition. For example, a coach may win a game when wearing a particular pair of socks, then always wear the same socks to games. More interesting, is the occasional false causality between music and motion. For example, Lillian Schwartz developed a series of computer generated images, sequenced them, and attached a sound track (usually Mozart). While there were some connections between one image and the next, the music was not scored to the images; however, on viewing them, the music appeared to be connected. All of the connections were observer supplied.

An example of non-computer illusionary causality is the choreography of Merce Cunningham. To him, his work is non-representational and without intellectual meaning. He often worked with John Cage, a randomist composer. Cunningham would rehearse his dancers, Cage would create the music; only at the time of the performance would music and motion come together. However, the audience usually conceived of a causal connection between music and motion and saw structure in both.

1.4 Recognizing causality basics

A common approach to recognizing causal relationships is by manipulating variables by experimentation. How to accomplish causal discovery in purely observational data is not solved. (Observational data is the most likely to be available for data mining analysis.) Algorithms for discovery in observational data often use correlation and probabilistic independence. If two variables are statistically independent, it can be asserted that they are not causally related. The reverse is not necessarily true.

Real world events are often affected by a large number of potential factors. For example, with plant growth, many factors such as temperature, chemicals in the soil, types of creatures present, etc., can all affect plant growth. What is unknown is what causal factors will or will not be present in the data; and, how many of the underlying causal relationships can be discovered among observational data.

Some define cause-effect relationships as: When α occurs, β always occurs. This is inconsistent with our commonsense understanding of causality. A simple environment example: When a hammer hits a bottle, the bottle *usually* breaks. A more complex environment example: When a plant receives water, it *usually* grows.

An important part of data mining is understanding whether there is a relationship between data items. Sometimes, data items may occur in pairs but may not have a deterministic relationship; for example, a grocery store shopper may buy both bread and milk at the same time. Most of the time, the milk purchase is not caused by the bread purchase; nor is the bread purchase caused by the milk purchase.

Alternatively, if someone buys strawberries, this may causally affect the purchase of whipped cream. *Some* people who buy strawberries want whipped cream with them; of these, the desire for the whipped cream varies. So, we have a conditional primary effect (whipped cream purchase) modified by a secondary effect (desire). How to represent all of this is open.

A largely unexplored aspect of mined rules is how to determine when one event causes another. Given that α and β are variables and there appears to be a statistical covariability between them, is this covariability a causal relation? More generally, when is any pair relationship causal? Differentiation between covariability and causality is difficult.

Some problems with discovering causality include:

- Adequately defining a causal relation
- Representing possible causal relations
- Computing causal strengths
- Missing attributes that have a causal effect
- Distinguishing between association and causal values
- Inferring causes and effects from the representation.

Beyond data mining, causality is a fundamentally interesting area for workers in intelligent machine based systems. It is an area where interest waxes and wanes; in part because of definitional and complexity difficulties. The decline in computational interest in cognitive science also plays a part. Activities in both philosophy and psychology [Glymour, 2001] overlap and illuminate computationally focused work. Often, the work in psychology is more interested in how people *perceive* causality as opposed to whether causality actually exists. Work in psychology and linguistics [Lakoff, 1990] [Mazlack, 1987] show that categories are often linked to causal descriptions. For the most part, work in intelligent computer systems has been relatively uninterested in grounding based on human perceptions of categories and causality. This paper is concerned with developing commonsense representations that are compatible in several domains.

2. Causality

Centuries ago, in their quest to unravel the future, mystics aspired to decipher the cries of birds, the patterns of the stars and the garbled utterances of oracles. Kings and generals would offer precious rewards for the information soothsayers furnished. Today, though predictive methods are different from those of the ancient world, the knowledge that dependency recognition attempts to provide is highly valued. From weather reports to stock market prediction, and from medical prognoses to social forecasting, superior insights about the shape of things to come are prized [Halpern, 2000].

Democritus, the Greek philosopher, once said: "Everything existing in the universe is the fruit of chance and necessity." This seems self-evident. Both randomness and causation are in the world. Democritus used a poppy example. Whether the poppy seed lands on fertile soil or on a barren rock is chance. If it takes root, however, it will grow into a poppy, not a geranium or a Siberian Husky [Lederman, 1993].

Beyond computational complexity and holistic knowledge issues, there appear to be inherent limits on whether causality can be determined. Among them are:

- *Quantum Physics*: In particular, Heisenberg's uncertainty principle
- Knowledge of the world might never be complete because we, as observers, are integral parts of what we observe
- *Gödel's Theorem*: Which showed in any logical formulation of arithmetic that there would always be statements whose validity was indeterminate. This strongly suggests that there will always be inherently unpredictable aspects of the future.
- *Turing Halting Problem*: Turning (as well as Church) showed that any problem solvable by a step-by-step procedure could be solved using a Turing machine. However, there are many routines where you cannot ascertain if the program will take a finite, or an infinite number of steps. Thus, there is a curtain between what can and cannot be known mathematically.

- *Chaos Theory*: Chaotic systems appear to be deterministic; but are computationally irreducible. If nature is chaotic at its core, it might be fully deterministic, yet wholly unpredictable [Halpern, 2000, 139].
- *Space-Time*: The malleability of Einstein's space time that has the effect that what is "now" and "later" is local to a particular observer; another observer may have contradictory views.
- *Arithmetic Indeterminism*: Arithmetic itself has random aspects that introduce uncertainty as to whether equations may be solvable. Chaitin [1987, 1990] discovered that Diophantine equations may or may not have solutions, depending on the parameters chosen to form them. Whether a parameter leads to a solvable equation appears to be random. (Diophantine equations represent well-defined problems, emblematic of simple arithmetic procedures.)

Given determinism's potential uncertainty and imprecision, we might throw up our hands in despair. It may well be that a precise and complete knowledge of causal events is uncertain. On the other hand, we have a commonsense belief that causal effects exist in the real world. If we can develop models tolerant of imprecision, it would be useful. Perhaps, the tools found in soft computing may be useful.

4. Epilogue

Causality occupies a central position in human commonsense reasoning. In particular, it plays an essential role in common sense human decision-making by providing a basis for choosing an action that is likely to lead to a desired result. In our daily lives, we make the commonsense observation that causality exists. Carrying this commonsense observation further, the concern is how to computationally recognize a causal relationship.

Data mining holds the promise of extracting unsuspected information from very large databases. Methods have been developed to build rules. In many ways, the interest in rules is that they offer the promise (or illusion) of causal, or at least, predictive relationships. However, the most common form of rules (association) only calculate a joint occurrence frequency; not causality. A fundamental question is determining whether or not recognizing an association can lead to recognizing a causal relationship.

An interesting question how to determine when causality can be said to be stronger or weaker. Either in the case where the causal strength may be different in two independent relationships; or, where in the case where two items each have a causal relationship on the other.

Causality is a central concept in many branches of science and philosophy. In a way, the term "causality" is like "truth" -- a word with many meanings and facets. Some of the definitions are extremely precise. Some of them involve a style of reasoning best supported by fuzzy logic.

Defining and representing causal and potentially causal relationships is necessary to applying algorithmic methods. A graph consisting of a collection of simple directed edges will most likely not offer a sufficiently rich representation. Representations that embrace some aspects of imprecision are necessary.

A deep question is when anything can be said to cause anything else. And if it does, what is the nature of the causality? There is a strong motivation to attempt causality discovery in association rules. The research concern is how to best approach the recognition of causality or non-causality in association rules. Or, if there is to recognize causality as long as association rules are the result of secondary analysis?

References

- G. Chaitin [1987] Algorithmic Information Theory, Cambridge University Press, Cambridge, United Kingdom
 G. Chaitin [1990] "A Random Walk In Arithmetic," New Scientist 125, n 1709 (March, 1990), 44-66
 C. Glymour [2001] The Mind's Arrows, Bayes Nets And Graphical Causal Models In Psychology, MIT Press, Cambridge, Massachusetts

- Hume, David [1748] *An Enquiry Concerning Human Understanding*, reprinted: Harvard Classics, v 37, copyright 1910, Collier and Son, New York
- G. Lakoff [1990] *Women, Fire, And Dangerous Things: What Categories Reveal About The Mind*, University of Chicago Press
- L. Lederman, D. Teresi [1993] *The God Particle: If the Universe Is the Answer, What Is the Question?* Delta, New York
- L. Mazlack [1987] "Machine Conceptualization Categories," *Proceedings 1987 IEEE Conference on Systems, Man, and Cybernetics*
- L. Zadeh [2000] "Abstract Of A Lecture Presented At The Rolf Nevanlinna Colloquium, University of Helsinki," reported to: *Fuzzy Distribution List*, fuzzy-mail@dbai.tuwien.ac.at, August 24, 2000
-

A Framework for Effective Gene Expression Analysis and Biological Knowledge Discovery

S. M. Vincent Tseng, Lien-Ching Chen, and Shih-Chiang Yang

The cDNA and Oligonucleotide microarray chips have become the most popular techniques for investigating gene expressions profiles. The high density DNA microarray technology can simultaneously monitor the expression levels of thousands of genes. How to analyze the huge experimental result and discover useful biological knowledge is an important topic. In this research we present a framework, namely GeneFilter, for effectively analyzing gene expression data and performing knowledge discovery. The main goals of the framework are as follows: 1) Design an effective model and flow for high-throughput data analysis, 2) Develop an integrated and efficient platform for gene expression analysis including data preprocessing, gene expression patterns mining and visualization modules, 3) Develop methodologies for biological knowledge discovery based on the gene expression analysis results.

The system architecture of GeneFilter is as shown in Figure 1. In the wet-Lab, biologists design the experiments for targeted diseases and conduct experiments with cDNA microarray, oligonucleotide microarray, Q-PCR, or 2D gel chip. The experimental results are then submitted to GeneFilter for the Dry-Lab analysis, including data preprocessing, statistical analysis, gene expression patterns mining, gene ranking and gene ontology analysis. Finally, the analysis results are validated by using more precise biological experiments like Q-PCR, and the validation results are feed back for refining the further analysis. In the following, we describe the main functions in each of the analysis process:

1. Data Preprocessing and Statistical Analysis

In this process, the quality of the microarray data is first examined via various statistical methods and visualization techniques like in [1]. The data of bad quality will be filtered out eventually. Then, various types of normalization methods, like Within-slide normalization, Paired-slides normalization, Multiple-slides normalization [2, 3], are provided for applications on the data. Various statistical graphs like MA Plots are also provided for examining the normalized results.

2. Gene Expression Patterns Mining

In this process, the interesting patterns hidden behind the microarray data are discovered automatically via the following data mining methods we developed based on our past experiences [7, 8, 9]: 1) Automatic and customized discovery of gene expression patterns (e.g., the up-regulated or down-regulated expression patterns), 2) Validation-based clustering for finding the nearly-optimal clustering of the gene expression in very short time, 3) Time-series analysis for discovering the activating relations between genes with tolerance of noisy data, 4) Classification analysis for modeling or contrasting gene expression patterns under different disease types (e.g., the dominating genes related to bladder cancer).

3. Gene Ranking

For the genes identified to be with specific patterns we are interested via the data mining methods, a

ranking mechanism is further applied for distinguishing the possible degree of importance of these genes. The ranking mechanism takes as inputs various kinds of information and then calculates a score for each gene based on the given weights for each type of information. The input information includes the significant amount of gene expression ratio value in the microarray, the degree of relevant to the targets based on the published literatures (e.g., Gene Expression Omnibus (GEO) web site (<http://www.ncbi.nlm.nih.gov/geo/> and [5]) releases the genes related to various kinds of cancers), Q-PCR results, etc.

4. Gene Ontology Analysis

For the interesting genes discovered through the above processes, three organizing networks of Gene Ontology (<http://www.geneontology.org/>) information, namely molecular function, biological process and cellular component, are provided for exploring the deeper knowledge on the genes. In GeneFilter platform, an agent program was designed for accessing the latest Gene Ontology information, and a querying system was provided for users to discover the Gene Ontology information related to specific genes.

5. Analysis Validations and Cyclic Refinement

Through the above analysis process, the most interesting genes related to the analysis targets will be discovered in an automatic and efficient way. To validate the analysis results, further biological experiments should be conducted via more precise biological experiments like Q-PCR [6]. The validation results will be feed back into the analysis platform again for refining the analysis by adjusting the policies and parameters in data preprocessing, expression pattern mining and gene ranking. Through the cyclic refinement process, more and more accurate analysis results will be obtained.

In real applications, we have conducted an extensive analysis of bladder cancer microarray on GeneFilter platform for discovering the genes most influential on the progress on different stages of bladder cancer. Multiple slides of microarray with more than 10,000 genes are analyzed and genes with interesting expression patterns (e.g., up-regulated) are discovered successfully. In particular, the whole analysis process can be done within very short period due to the nice features of the platform in terms of high degree of integration and automation.

References

- [1] S. Dudoit, Y. H. Yang, T. P. Speed, and M. J. Callow, "Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments," in *Statistica Sinica*, Vol. 12, No. 1, p. 111-139, 2002.
- [2] Yee Hwa Yang, Sandrine Dudoit, Percy Luu and Terry Speed, "Normalization for cDNA Microarray Data," in *SPIE BiOS*, San Jose, California, January 2001.
- [3] X. Wang, M. J. Hessner, Y. Wu, N. Pati, S. Ghosh, "Quantitative quality control in microarray experiments and the application in data filtering, normalization and false positive rate prediction," in *Bioinformatics*, 19(0): 1-7, 2003.
- [4] Anil K. Jain & Richard C. Dubes (1988) Algorithms for Clustering Data. Prentice Hall.
- [5] Jung-Hsien Chiang and Hsu-Chun Yu, "MeKE: discovering the functions of gene products from biomedical literature via sentence alignment," *Bioinformatics* 2003 19: 1417-1422.
- [6] M.S. Rajeevan, Vernon S.D., Taysavang N., and Unger E.R. (2001) Validation of Array-Based Gene Expression Profiles by Real-time (Kinetic) RT-PCR. *J. Mol. Diagnostics* 3(1):26-30.
- [7] S. M. Tseng and L. J. Chen, "An Empirical Study of the Validity of Gene Expression Clustering," in *Proc. 2002 Int'l Conf. on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS)*, Las Vegas, USA, 2002.
- [8] S. M. Tseng and Ching-Ping Kao, "Mining and Validation of Gene Expression Patterns: An integrated approach and applications," in *Informatica*, vol. 27, no. 1, pp. 21-27, 2003.
- [9] S. M. Tseng and Y. L. Chen, "An Effective Approach for Resolving Fundamental Problems in Time-Series Gene Expression Mining", in *Proc. Int'l Workshop on Foundation and Direction in Data Mining (held with IEEE ICDM)*, USA, Nov., 2003.

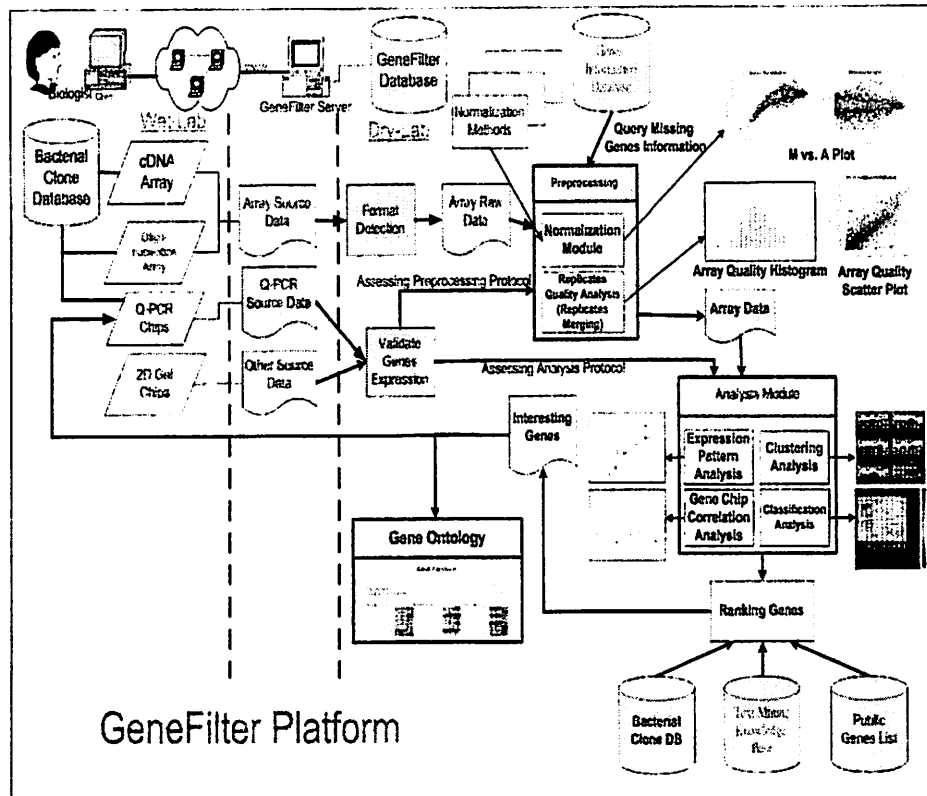


Figure 1. The GeneFilter Platform

Semantics Powered Bioinformatics: Semantic Search, Integration, Processes and Analytics

Amit Sheth

Bioinformatics is about providing biologists and those in allied disciplines with ability to exploit information information that is increasingly distributed, heterogeneous and massive. Bioinformatics has already succeeded in utilizing database management, workflow and information retrieval technologies, which have provided syntactic search, heterogeneous data access and sharing, and limited forms of integration. A good amount of effort has also involved use of statistical and syntactic techniques to support the essential tasks of finding patterns, similarities, and matches to identify building block structures. Looking to the future, we can realize more exciting potential of bioinformatics if we have more automated ways for analysis leading to insight and discovery to understand cellular components, molecular functions and biological processes, and more importantly complex interactions and interdependencies between them. And while lot of effort in the last decade focused on genes, next set of challenges involve more complex structures of protein and carbohydrates.

This talk focuses on semantics enabled bioinformatics. We outline increasing use of semantic techniques in bioinformatics for search, browsing, integration, analytics and discovery. Ontologies provide underpinning for most of today's semantic techniques and the Semantic Web research, and bioinformatics is one of the most aggressive adopters of ontologies among science and industry domains. We provide examples of use of multiple ontologies and Semantic Web Processes to investigate more automation in discovery and analysis. We will also weave a brief overview of the research and exciting

commercial state of the art in the semantic technology (specifically ontology driven information systems).

Sample background/related material:

- An example research project: Bioinformatics for Glycan Expression
<http://lsdis.cs.uga.edu/proj/glycomics/>
- Examples of Semantic Technology: SCORE (<http://lsdis.cs.uga.edu/lib/download/S+2002-SCORE-IC.pdf>) / Semantic Enhancement Engine (<http://lsdis.cs.uga.edu/lib/download/HSK02-SEE.pdf>), Taalee Semantic Search (<http://www.streamingmediaworld.com/gen/reviews/searchassociation/index.html>)
- An Example Commercial Product: Semagix Freedom (<http://www.semagix.com>)

Mining Fuzzy Association Rules: an Overview

Miguel Delgado

The theory of Fuzzy Sets has been recognized as a suitable tool to model several kinds of patterns that can hold in data. Particularly Fuzzy Sets Theory has been shown to be a very useful tool in Data Mining to representing a natural and human way the so called Association Rules. The objective of this paper is to present a revision of the most relevant results about the use of Fuzzy Sets in Data Mining, specially in relation with the discovery of Association Rules. First of all we will introduce the basic concepts of Data Mining to justify the need of Fuzzy Sets Theory. A historical revision on developments in this field is made too. Next we will present our researches about Fuzzy Association Rules, starting with the formulation of a general model to discover association rules among items in a (crisp) set of fuzzy transactions. This general model can be particularized in several ways; each particular instance corresponds to a different kind of pattern and/or repository of data. We describe some applications of this scheme, paying special attention to the discovery of fuzzy association rules in relational databases. Paper finishes with some suggestions about future researches and problems to be solved.

Fuzzy Spectral Patterns for Information Integration in a Search Engine

Galina Korotkikh

Recently, Lotfi Zadeh suggested that the development of a search engine with the ability to synthesize an answer to a query from different information elements would be required. In the search the information elements are specified by the query and their integration into an aggregated information element serves to provide a basis for the answer.

In general, the problem of information integration is very challenging and remains unresolved. In the Internet search context it may be characterized by a huge number of information elements involved. Therefore, an identification and understanding of possible patterns related to the information integration may turn to be useful in the development of a search engine.

In the paper we study this problem in terms of the eigenvalue spectra of the variance-covariance matrices. They are applied to binary sequences as information elements to model the information integration. The eigenvalue spectrum of the variance-covariance matrix is used to investigate the result of the information integration. It is suggested that some patterns of the eigenvalue spectra may be helpful in the understanding of the information integration.

We demonstrate that there exists a certain pattern dynamics of the eigenvalues within which the spectra of the variance-covariance matrices may be described. We reveal a mechanism that underlies the eigenvalue dynamics and show that its functioning can find efficient computational expression by a quadratic trace of the variance-covariance matrix. It is proposed how the pattern of the eigenvalue spectra, specified by the dynamics and the mechanism, may be interpreted to help us in the understanding of the information integration.

It is also discussed that the fuzzyfication of the eigenvalues has played a crucial role in the identification of the dynamics and the mechanism. This may support that fuzziness is an integral part of the information integration.

Automated Consumer Profiling Using Fuzzy Query and Social Network Techniques

F. Olcay Cirit, Sema Alptekin and Masoud Nikravesh

Decades of research in the field of Social Network Analysis have established the existence of a phenomenon known as *homophily*: the tendency of people of similar behavior and social condition to co-associate. One can observe this phenomenon on a small scale when girls in the same sorority tend to favor the same brands of products, while on a national scale one finds that people in high-income brackets overwhelmingly tend to have friends who are also in high income-brackets.

Many internet companies today, among them America Online™, Friendster™, and Yahoo!™, maintain large databases of information about their users, and, if they do not already, are at least in a position to capture information about how each user connects socially to the other users of the system. A web site could gather this information directly by correlating lists of friends that each user provides, or indirectly by correlating usage of online discussion forums and chat rooms. This project examines the possibility of applying the social connectivity information thus gleaned to profile potential consumers of a product or service. Also discussed are techniques for using the BISC/DSS, a fuzzy query-based decision support system, to implement an automated profiling.

For ease of conceptualization, we compare being a consumer of a product or service to being a member of a social club. The task of consumer profiling, then, can be thought of as the search for potential new club members. Alternatively, one can ask the following question: given a set of connected persons, which of them are most likely to belong to club X? One possible approach to solving these problems involves constructing a profile of the members already in the club, and then calculating the similarity of each user to the profile. Static information about the individuals in the club, such as age, income, interests, etc., could be supplemented with social network metrics such as connectedness, betweenness, closeness, etc. to provide a social grounding for the profile. Such an approach is distinctly suited for use with the DSS because the definition of the profile corresponds exactly with a fuzzy query. Targeted internet advertising and customer relationship management could make profitable use of the consumer profiling techniques outlined here. A proposed implementation of an automated consumer profiling system that integrates with the BISC/DSS is examined in detail, including a discussion of methods for collecting social connectivity information, and an assessment of the pitfalls of profiling.

Soft Computing and Current Trends in Supercomputing

Horst D. Simon

Supercomputing today in 2003/2004 is no longer based on custom built expensive vector supercomputers which were associated with the field more than a decade ago. Many of the high-end computers today are

built from inexpensive commodity parts, using open source software. Consequently the application of this high-end computing technology to floating-point intensive applications in science and engineering has become a dynamic and vibrant field. Clusters and related technology are used everywhere from science and engineering departments in universities, to commercial applications in banking, telecoms, and biotech. Surprisingly the field of soft computing as represented by the topics of this workshop, has not been a strong participant in this revolution of high end computing. With some notable exceptions such as Google's search engine that runs on commodity clusters, parallel supercomputing and soft computing seem to be living in parallel universes.

In this talk, I will first survey the state of the field in supercomputing, and analyze current developments based on data from the TOP500 list. I will demonstrate with a few examples from NERSC's users community, how powerful compute resources have transformed science, and how computational science has become accepted as the third leg of science in many disciplines. Then I will offer an analysis, why these developments have had little impact on computational problems in soft computing. Based on this analysis, I will conclude with some suggestions on how both fields could interact and learn from each other in the future.

Some Opportunities for Soft Computing Applications

Mo Jamshidi

In this lecture we present a series of what the speaker considers excellent application opportunities for soft computing tools with national and international implications. Economy and security are two of the most important issue of not just today, but also for tomorrow and years to come. These opportunities are intelligent multi-agent systems (e.g. V-Lab®, presented the day before) and autonomous agents monitoring, recognizance and border security. Another opportunity, not unrelated to national security, is the application of soft computing for energy efficiency of the so-called "Industries of the Future", where SC tools can be used to save energy, reduce waste or can help avoid production of defective products, reduce energy requirements for automation and remote operations, increase efficiency of existing processes via sensing and information technology systems where demand and automated production modeling can be integrated, etc.

Another opportunity exists in the chemical process systems, where relatively little effort has been made to use SC tools. Such problems as water treatment plants, oil extraction, etc. are candidate applications. Finally, two other opportunities are diagnostics/prognostics of national defense systems and image processing applications and remote sensing for the study of earth.

Fuzzy Logic for Genetic Networking

Paul P. Wang

This decade will probably be remembered as the "genome decade". Almost a dozen microorganism sequences, including bacteria, and genomic sequences for man, mouse, etc. have already been completed. However, the post genome sequencing era has just begun. The focus now is on uncovering the functional organization of cells. The systemic global studies of gene expression and DNA-protein interactions in different conditions is a topic of considerable interest.

There are numerous modeling approaches of genetic networks already being proposed, but the Boolean network model discovered about three decades ago remains quite strong among researchers in the field. In this paper, the “fuzzy logic network” FLN, is proposed as a viable model for genetic networking studies. The basic properties and characteristics of the FLN are presented in this paper. The Boolean network model can be shown to be a special case of the FLN. As pointed out by numerous researchers, the Boolean network is only a coarse-grained and symbolic modeling despite the faithful modeling of the gene self-regulating process. FLN not only provides more qualitative information, but also provides some essential quantitative information about the self-regulating mechanism of the genes in the network.

One important aspect of the improvement is the possible control of the genetic network. A node may be added to a graph which represents the genetic network. Mathematically, this node represents a biological stimulus where a stimulus is any relevant physical or chemical factor which influences the network and is itself not a gene or gene product. In system and control literatures, this node is called an input function or excitation function.

Theoretically speaking, we are able to show that this control node, under certain constraints, is capable of regulating the genetic network, which is already self-regulating. To say the very least, this theoretical result may prove to be useful in experimental design. This control strategy may potentially be useful in obtaining desired “target network states”. Achieving these preferred “target network states” may have significance in cell transformation, cell repair or drug target design.

Much more remains to be explored and we hope the fuzzy logic network will be eventually be accepted as a viable model in the genomic research community. This possibility certainly exists due to FLN’s better fit with the reality in biology as well as its simplicity in computations.

Performing Region Detection Using Fuzzy Agents

Yonathan Asfaw, Martine De Cock and Masoud Nikraves

Fuzzy Reasoning can be applied in work that requires decision-making. The use of simple agents is also gaining ground in computer science. We joint the two concepts to employ simple agents that have minimal interaction with one another to attempt to carry out region detection in images using fuzzy reasoning.

Panel Discussion: TUESDAY, DEC 16

Computation Intelligence

This Panel is concerned with the application of the methods of computational intelligence to the analysis, modeling and knowledge discovery in Internet and bioinformatics and to the development of new methods and systems for solving various challenging problems in the specified area. The methods of computational intelligence include methods of artificial intelligence, neural networks, fuzzy logic, evolutionary algorithms, methods for data mining, statistical methods and chaos analysis, algebraic methods and finite automata, and combination of them.

Panelists:

R. Bajcsy, M. Jamshidi, E. Ruspini, A. Sheth, H. Simon, and L. A. Zadeh

Moderator:

Masoud Nikravesh; University of California, Berkeley

WEDNESDAY, DECEMBER 17

BIOINFORMATICS

Abstracts of WEDNESDAY, DECEMBER 17:

Bioinformatics

Knowledge Based Neurocomputing for Data Analysis and Knowledge Discovery in Bioinformatics

Nik Kasabov

Problems in bioinformatics and medical decision support are defined and solved with the use of various soft-computing methods and results are compared. The problems include: gene expression data analysis for cancer and other disease profiling; promoter recognition; modeling gene regulatory networks; integrating gene and clinical data; cardio-vascular risk prognosis; renal function evaluation; survival prognosis; gene regulatory networks for brain function analysis. There are five main phases of information processing and problem solving in most bioinformatics systems that are covered for each of the above problems, namely: data pre-processing and filtering; feature evaluation and feature selection; model creation and evaluation; knowledge extraction; adaptation to new data.

Principal Component and Self-aggregation Clustering of Gene Expressions and Protein Interactions

Chris Ding

Cluster analysis partitions data points into disjoint groups to quickly gain first order knowledge from data. Recently we developed several principal component based clustering algorithms with well defined clustering objective functions. For the widely used K-means clustering we proved that the continuous solution of the cluster membership indicator vector is the principal component[1], leading to effective implementations of K-means clustering. For spectral graph clustering, we proved that scaled principal components form a dynamic process of self-aggregation in which data objects move towards each other to form clusters, revealing the inherent pattern of similarity[2]. The MinMaxCut spectral method follows a min-max clustering principle that the between-cluster associations are minimized, while the within-cluster associations are maximized.

DNA microarrays can simultaneously monitor expression levels of thousands of genes. Expression profiles of patients tissue samples provide molecular rather than morphological signature of cancer. We apply data clustering to expression profiles of human Lymphoma and demonstrate that data clustering can effectively discover different pathological stages of B-cell lymphoma thus providing an effective diagnostics methodology [3].

Proteins carry out many cellular functions such as metabolism, communication, growth. Systematic identification of these protein complexes provide essential knowledge linking proteome dynamics to cellular function and phenotype. We describe a unified representation of protein complex dataset based on bipartite graphs, from which protein-protein interactions and protein-complex - protein-complex associations are deduced naturally. Apply data clustering to the protein-protein network we obtain statistically significant protein clusters. Apply data clustering to the protein-complex - protein-

complex association network, we obtain protein supercomplexes that provide systems-level understanding of cellular processes through the concepts from Gene Ontology [4].

- [1] C. Ding and X. He. Principal component analysis and K-means clustering. LBNL Tech Report 52983. June 6, 2003.
 - [2] C. Ding, X. He, H. Zha, and H. Simon. Unsupervised learning: self-aggregation in scaled principal component space. Proc. of Principles of Data Mining and Knowledge Discovery (PKDD 2002), pp. 112-124. Aug. 2002.
 - [3] C. Ding. Analysis of gene expression profiles: class discovery and leaf ordering. Proc. of Conf. on Research in Comp.Mol.Bio (RECOMB 2002), pp.127-136. April 2002.
 - [4] C. Ding, X. He, R. Meraz, S. Holbrook. A unified representation for multi-protein complex data for modeling protein interactions. LBNL Tech Report 52984. To appear in Proteins: Structure, Function, and Genetics.
-

Mining Imprecise Discriminative Molecular Fragments

Michael Berthold

We present an algorithm to find imprecise fragments in a set of molecules that help to discriminate between different classes of, for instance, activity in a drug discovery context. Instead of carrying out a brute-force search, our method generates fragments by embedding them in all appropriate molecules in parallel and prunes the search tree based on a local order of the atoms and bonds, which results in substantially faster search by eliminating the need for frequent, computationally expensive reembeddings and by suppressing redundant search. An extension of the search algorithm allows finding fragments that also incorporate chemical expert knowledge about structural similarities that can be tolerated. We demonstrate the usefulness of our algorithm by demonstrating the discovery of activity-related groups of chemical compounds in the National Cancer Institute's HIV and Cancer-screening datasets.

A Decentralization Method For Modeling The Multiple Levels Of Organization And Function Within The Liver

C. Anthony Hunt and Glen E. P. Ropella

Development of a biological model is driven by the experimental context in which it will be used. Hence, computer models are often overfitted to a single, unique, experimental context and fail to be useful in other situations. So doing severely limits the model's usefulness, effectively blocking inferential extensions to somewhat different conditions such as a hypothesized new treatment intervention. To solve this problem, multiple, separate models of a biological system at different levels of organization are required to understand and adequately represent their behaviors. In this poster, we present the basics of a new modeling method, FURM (Functional Unit Representation Method) that attempts to address this problem. Here we focus on the primary functional unit of the liver within an in silico isolated perfused rat liver (IS-IPRL). FURM [<http://biosystems.ucsf.edu/Research/furm/index.html>] decouples the various aspects of functional units. It uses a middle-out model design strategy that enables and encourages selection of different models. It is an example of a new class of generative biological simulation models whose components are easily joined and disconnected and are replaceable and reusable. It works by decentralizing the modeling process without requiring that all of the data be of a specific type. FURM does not require any particular formalism. Rather the experimental framework is formulated using Partially Ordered Sets. We follow four fundamental guidelines: 1) standardize interfaces to

multiparadigm, multimode, and trans-domain models, 2) use discrete interactions, 3) enable knowledge discovery by designing for an extended model life cycle, and 4) define observables that will submit to a similarity measure. A data model represents the biological system. An established in vitro liver perfusion protocol is the source of our experimental data in [J Pharmacokin Biopharm27:343-82, 1999]. Two in silico system models are implemented. *RefModel* is the accepted, reference mathematical model [JPET297:780-89, 2001]. *ArtModel* is our functional unit model that assumes that liver function, as a whole, is an aggregate of lobule function, that sinusoids are primarily vascular objects, and that transit time for perfusate is governed by stochastic interactions between various agents inside the vascular structures in combination with the perfusion pressure at the inlet catheter and lobule portal vein. The IS-IPRL strives to replicate the experimental procedure that has provided the experimental data. The four primary assumptions are: (1) Outflow profiles alone are lossy projections of liver behavior. Thus, physiologically accurate models are necessary to begin fully exploring the liver behavior space. (2) Hepatic vascular structure and the arrangement of lobules within a lobe can be represented by a directed graph. (3) The primary functional unit is the lobule. (4) Outflow for sucrose, but not metabolized or transported solutes, is solely a function of the extracellular (vascular cavity) space and its geometry. Within lobules agents representing sinusoidal segments (SS) are located at each graph node. Agents within each SS represent functionalities within cellular and subcellular spaces. The explicit hypothesis being tested is that the selected parameter vectors cause the model to generate output that is experimentally indistinguishable from that seen in the in vitro data. A similarity measure is used to automate the evaluation of the solution sets put forth by the models. Those results make possible automatic searches of the parameter space for regions that solve the problem by matching the in vitro data. To date different parameterizations of a mathematical model have been needed to account for outflow data for two different solutes. One IS-IPRL model now accounts for the hepatic outflow profiles of both sucrose and diltiazem and is being extended to account simultaneously for five additional drugs and to be able to shift to represent either of two disease states.

Augmented Fuzzy Cognitive Maps based on Case Based Reasoning for Decisions in Medical Informatics

Voula C. Georgopoulos, Chrysostomos D. Stylios and Glen E. P. Ropella

This paper presents the Soft Computing technique of Fuzzy Cognitive Maps (FCMs), which is a knowledge-based methodology suitable to describe and model complex systems and handle information from an abstract point of view [1]. Soft computing techniques such as FCMs have been successfully used to model complex systems that involve discipline factors, states, variables, input, output, events and trends. These modeling techniques can integrate and include in the decision-making process the partial influence of controversial factors, can take under consideration causal effect among factors and evaluate the influence from different sources, factors and other characteristics using fuzzy logic reasoning. Thus, such methods are ideal for the development of Decision Support systems in Medical Informatics, because in this application area humans use decision approaches, such as in differential diagnosis, for example, mainly based on fuzzy factors. Some of these factors are complementary, others similar and others conflicting, which are all taken into consideration when a decision is reached [2, 3, 4]. The involved factors have different degrees of importance in determining (or influencing) the decision, and moreover the fact that they can be additive and/or conflicting increases the complexity of the problem and the vagueness of the decision.

Fuzzy Cognitive Maps develop a behavioral model of the system by exploiting the experience and knowledge of experts due to the way that they are developed. Fuzzy Cognitive Maps applicability in modeling complex systems has been successfully introduced [5]. FCMs originated from the synergism of

Fuzzy Logic and Neural Networks, taking advantage of both theories. An FCM is a signed fuzzy graph with feedback, consisting of concepts-nodes and weighted interconnections. Nodes of the graph stand for concepts that are used to describe main behavioral characteristics of the modeled system. Nodes are connected by signed and fuzzy weighted arcs representing the cause and effect relationship existing among concepts. Thus, an FCM is a fuzzy-graph structure, which allows systematic causal propagation, in particular forward and backward chaining [6]. Fuzzy Cognitive Maps have been successfully used to develop a Decision Support System for differential diagnosis [7], to determine the success of radiation therapy process estimating the final dose delivered to the target volume [8] and many other application areas.

When a FCM is used as part of a decision-making system, some of the concepts are considered output nodes, which correspond to each of the possible decisions. In general, the FCM could converge to a fixed point, limit cycle, or chaotic attractor, but when FCMs are used for the decision making process, it is desirable to converge to a region corresponding to the selection of one decision. Therefore, the output nodes must "compete" against each other in order for only one of them to dominate and be considered the decision. Here a new idea is proposed for achieving this "competition" between concepts. The interaction of each of these output nodes with the other output nodes should be strongly inhibitory. This implies that the higher the value of a given node, this should lead to a lowering of the value of competing nodes. In such a case the FCM, referred to as a Competitive Fuzzy Cognitive Map (CFCM) will always converge to a fixed appropriate region.

In this research work we further develop CFCMs by combining CFCMs with methods and approaches that have been used for Case-Based Reasoning (CBR). This is a successful methodology for managing implicit knowledge, which has also been used in medical informatics. CBRs embed a considerable amount of previous solved instances of problems (called cases). The problem solving experience is explicitly taken into account by storing past cases in database (case base), and by suitably retrieving them when a new problem has to be tackled [9]. It simply solves new cases by similarity of the solution of the old cases stored in its case-base rather than using some derivative representation, as is done for example in adaptive-type methodologies. But, if the new case has no match with the stored cases, CBR has no solution. Similarly, to FCMs and CFCMs, CBRs have been applied in medical diagnosis and patient treatment outcomes. Despite the limitations of CBRs they are usually assumed to have a certain degree of richness of stored knowledge, and a certain *degree of complexity* due to the way they are organized.

Even though successful medical Decision Support CFCMs have been developed [7], [10] there are situations where the patient data to be input into the system presents a very rare configuration of symptoms where most of the nodes of the CFCM would not be active. In other words, for example, although the CFCM-Model of a Medical Decision Support System has been designed to include all possible symptoms and causative factors (nodes-concepts) and the relationship between them (weights) for some medical condition, in a particular situation very few symptoms are available and are taken into consideration. Thus, in such a diagnosis or prognosis model Decision Support FCM, the decision would be made only using a very small subset of the concepts of the entire system. Such a system could lead to either an erroneous decision or difficulty in reaching stability since the weighting of the active nodes reflects only a small amount of the experts' stored knowledge. Using a CBR-Augmented CFCM Decision support system, in such situations, the decisions support system would draw upon cases that are maximally similar according to distance measures and would use the CBR subsystem to generate a sub-CFCM emphasizing the nodes activated by the patient data and thus redistributing the causal weightings between the concept-nodes.

The advantage of CBR-augmented CFCMs lies in the ability to represent rare occurrences of medical conditions/symptoms, which may not be adequately represented in an FCM due to its design methodology, which is dependent on human experts and learning algorithms.

In this paper, we describe the algorithm that generates the CBR-augmented CFCMs and present their use and applicability for a Medical Decision System for the Speech and Language Impaired.

Specifically the CBR-augmented CFCMs is proposed for Differential Diagnosis of Specific Language Impairment from Dyslexia and Autism, as an extension of an CFCM that has been developed [7], [10].

References

- [1] B. Kosko, "Fuzzy Cognitive Maps," *International J. of Man-Machine studies*, vol.24, pp. 65-75 (1986).
- [2] Kasabov N. Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering, Cambridge, Massachusetts: MIT Press (1996).
- [3] Kasabov, N. Decision support systems and expert systems, in: M. Arbib (ed) Handbook of brain study and neural networks, Cambridge, Massachusetts: MIT Press (2002).
- [4] J. Zelezniokow and J. Nolan, "Using Soft computing to build real world intelligent decision support systems in uncertain domains," *Decision Support Systems*, Vol. 31, pp. 263-285 (2001).
- [5] C. D. Stylios and P.P. Groumpos, V. C Georgopoulos "An Fuzzy Cognitive Maps Approach to Process Control Systems" *J.Advanced Computational Intelligence*, Vol. 3, No5, pp.1-9, (1999)
- [6] C.D. Stylios and P.P. Groumpos, "Fuzzy Cognitive Maps in Modeling Supervisory Control Systems" *Journal of Intelligent & Fuzzy Systems*, Vol 8. No 2. pp. 83-98 (2000).
- [7] V.C. Georgopoulos, G.A. Malandraki and C. D. Stylios 'A Fuzzy Cognitive Map Approach To Differential Diagnosis of Specific Language Impairment' *Journal of Artificial Intelligence in Medicine*, Vol. 29, No.3, pp. 261-278, (2003).
- [8] E. Papageorgiou, C. Stylios and P. Groumpos "An Integrated Two-level Hierarchical System for Decision Making in Radiation Therapy using Fuzzy Cognitive" *IEEE Transactions on Biomedical Engineering* vol. 50, No 12 (December 2003)
- [9] J.B. Noh, K.C Lee, J.K. Kim, J.K. Lee and S.H. Kim, "A case-based reasoning approach to cognitive map-driven tacit knowledge management" *Experts Systems with Applications*, Vol 19, pp. 249-259 (2000).
- [10] V.C. Georgopoulos and C. D. Stylios, "Competitive Fuzzy Cognitive Maps for Decision-Making Systems," presented at Comcon 8, The 8th International Conference on Advances in Communication and Control: Telecommunications/Signal Processing, Crete, Greece, June 25-29 (2001).

Soft Computing Methods for Drug Dosing in Chronic Anemia

Adam Gaweda

Drug delivery in chronic therapy is widely recognized as a complex control problem. The non-stationary character of the human body, as a result of tolerance to the drug and other pharmacokinetic parameters changing over time, requires intermittent therapy adjustment. Furthermore, it is virtually impossible to regularly access all the relevant variables to accurately represent a system as complex as human body. Hence, the development of a parametric model becomes difficult and the need to deal with imprecision arises. This presentation demonstrates how these two aspects are addressed by employing Soft Computing methods in long-term dosing of erythropoietin, a drug which is used for the treatment of patients with chronic anemia. Artificial Neural Network-based adaptive controller for erythropoietin dose adjustment will be presented and evaluated with respect to the currently used protocol. A modified Fuzzy Rule-Based System, capable of handling imprecise input and output information, will be introduced and its application to predict patients' response to erythropoietin will be demonstrated.

NeuCom – A Neuro Computing Environment for Evolving Intelligence

Nikola Kasabov and Liang Goh

A comprehensive software environment *NeuCom* (www.theneucom.com) for data mining, modeling and discovery will be demonstrated and will be made available to participants. NeuCom includes various

methods for data visualization, feature selection, data analysis, modeling, rule extraction and adaptation. As a special method for adaptive learning of data and knowledge integration, evolving connectionist systems (ECOS) are included [1]. ECOS are multi-modular connectionist architectures that facilitate modeling of evolving processes and knowledge discovery. It is a collection of neural networks that operate continuously in time and adapt their structure and functionality through a continuous interaction with the environment and with other systems. As many processes in biology are dynamically evolving, their modeling requires evolving methods and systems, i.e. requires *evolving intelligence*. In the area of bioinformatics for example, new data is being made available with a tremendous speed that would also require models to be continuously adaptive. The talk adds to the traditional statistical and artificial intelligence methods the new methods of evolving systems, so that participants can compare the various approaches and apply them to their problems.

[1] N.Kasabov, *Evolving Connectionist Systems: Methods and Applications in Bioinformatics, Brain Study and Intelligent Machines*, Springer Verlag, 2002 (www.springer.de)

Hybrid Fuzzy Neural Networks for Protein Secondary Structure Prediction

Yuchun Tang, Preeti Singh, Yanqing Zhang, Chun-Dar Lu and Irene Weber

This paper presents a new fuzzy neural network method for predicting protein secondary structures. Three protein secondary structures (i.e., alpha helix, beta sheet and coil) are used for protein classification. The traditional orthogonal encoding scheme on the other hand takes an inordinate amount of convergence time to train neural networks. To solve this problem, the new encoding scheme for the various amino acids is proposed to map 20 amino acid symbols to numerical values in [0, 1] based on relevance degrees among the 20 amino acids. For comparison, the conventional multi-layer neural network with the Back Propagation (BP) learning is also used to do protein secondary structure prediction. The training data sets and testing data sets are generated differently from public protein data in order to verify prediction performance objectively.

The hybrid neural network system consists of (1) three independent neural networks for predicting alpha helix, beta sheet and coil respectively, and (2) the final decision-making system that classify the current protein sequence into one class with the maximum prediction output of the three prediction outputs generated by the three independent neural networks.

The hybrid conventional neural network has three independent neural networks using BP learning. The accuracy for predicting alpha helix and beta sheet can reach 70% and 60%, respectively for a training data set of 200 and more. The prediction accuracy for coil can reach 77.7% for a training data set of 70 sequences.

The new hybrid fuzzy neural network has three independent fuzzy neural networks using the new fuzzy learning. The initial simulation results have been generated by the hybrid fuzzy neural network. The accuracy for predicting alpha helix and beta sheet can reach 80.0% and 70.0%, respectively for a training data set of 200 and more. The prediction accuracy for coil can reach 88.8% for a training data set of 70 sequences.

In summary, the new hybrid fuzzy neural network is more effective than the hybrid conventional neural network in terms of prediction accuracy and learning speed. In the future, seven protein secondary structures will be used in prediction. The hybrid fuzzy neural network will be improved by adding new intelligent techniques such as genetic algorithms and kernel-based learning.

Multi-Agent System and Soft Computing Technologies as a Fruitful Symbiosis for Reliable Medical Diagnoses

Rainer Unland

Multi-agent systems are one of the most exciting research areas in computer science at the moment. In the last years there has been a growing interest in the application of agent-based systems in health care and especially in medical diagnosis. This stems mainly from the fact that in today's global world a fast and reliable medical diagnosis generation is of eminent importance as can be seen, for example, from the recent problems with SARS. Such highly contagious and lethal diseases can threaten the population if they are not fought immediately and with high efficiency. However, medical diagnosis can be a complex process with a lot of uncertainty and fuzziness. Thus, the integration of multi-agent system technology with soft computing technologies seems to be very promising, especially, since soft computing is tolerant of imprecision, uncertainty and partial truth. In this talk we will discuss whether and how multi-agent system technology can be used to improve and support the medical diagnosis process, especially, in situations when there is high uncertainty about the right diagnosis. Moreover, it will be shown in what way soft computing technologies can contribute to a further significant improvement of the diagnosis finding process and everything else related to it.

Soft Computing Tools for Gene Similarity Measures in Bioinformatics

Mihail Popescu, James M. Keller and Joyce Mitchell

One of the most important objects in bioinformatics is a gene product (a protein or an RNA). In clustering and subsequent knowledge discovery on unknown gene products, the principal features are the gene sequence and expression values found following a microarray experiment. The question arises as to what is the function of this gene product and is it similar in function or structure to other up-regulated or down-regulated gene products. Many (dis)similarity measures have been proposed to measure closeness of sequences. However, for many gene products, additional functional information comes from the set of Gene Ontology (GO) annotations and the set of journal abstracts related to the gene product. For these genes, it is reasonable to include similarity measures based on the terms found in the GO and/or the index term sets of the related documents (MeSH annotations). In both cases we deal with comparing two sets of terms arranged in a taxonomy (GO or MeSH.). Some measures have been constructed to assess closeness of terms in a taxonomy, including shortest path length between terms and information theory-related values where node probabilities are estimated using a corpus of relevant documents. Utilizing such factors in addition to sequence and expression should aid in the process of knowledge discovery. It will be easier to annotate clusters, for example, when they share common descriptive terms. When an unknown gene product joins the group via sequence and expression, it is reasonable to conjecture that this gene will also share the cluster annotations (at least partially).

In this paper we propose a fuzzy measure-based similarity (FMS) for computing the similarity of two sets of terms found in a taxonomy (and hence, the two gene products annotated with terms from the taxonomy). The advantage of FMS is that it takes into consideration the context of the whole set when computing the similarity. For the case when the objects are annotated to an ontology, we propose a method of dealing with the $G_1 \cap G_2 = \emptyset$ case, since it is likely that two gene products will not be described by identical ontology terms. In addition, we propose a modification of the Resnick similarity measure such that the similarity between two ontology terms produces a number in $[0,1]$. In dealing with large groups of documents describing the objects under consideration, not only do we determine the similarity between the document pairs, but, by introducing the Choquet integral to the scenario, we can

fuse this partial agreement function on pairs of documents into a single value relating the gene products. The measures for the final integral fusion can be tailored to produce order weighted average (OWA) operators (e.g., "at least two documents must support the connection") or can be based on assessments of the "worth" of individual and subsets of documents towards building the strength of connection. We present examples of FMS calculation for specific situations where two genes are described by a set of terms from the Gene Ontology, for two abstracts related to those genes, and for multiple abstract fusion. We also compare our measures to others from the literature.

Applications of Variable and Feature Selection in Bioinformatics

Isabelle Guyon

Variable and feature selection have become the focus of much research in areas of application for which data sets with tens or hundreds of thousands of variables are available, particularly in text processing and medical diagnosis with genomics and proteomics data. The problem consists in removing the input variables that are either not informative or redundant. The objectives include improving the performance of predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying process that generated the data.

We will present a historical perspective and outline the advantages and limitations of various approaches. We will show small examples that illustrate particular difficulties of the problem. Our examples will include DNA microarray analysis and protein profiling with mass spectrometry. We will explain the design of the NIPS2003 feature selection challenge and analyze the results of the benchmark.

Adaptive Clustering in Medical Image Segmentation as a Diagnostic Tool

Sunanda Mitra and Shuyu Yang

Image analysis techniques have been broadly used in computer-aided medical analysis and diagnosis in recent years. Computer aided image analysis is an increasingly popular tool in medical research and practice, especially with the increase of medical images in modality, amount, size and dimension. Image segmentation, a process that aims at identifying and separating regions of interests from an image, is crucial in many medical applications such as localizing pathological regions, providing objective quantitative assessment and monitoring of the onset and progression of the diseases, as well as analysis of anatomical structures. For clinical applications of segmentation, a compromise between the accuracy and computational speed of segmentation techniques is needed. Optimal segmentation processes based on statistical and adaptive approaches and their applicability to clinical settings will be addressed using diverse modalities of images. Current drawbacks of automated segmentation methodologies stem mostly from non-uniform illumination, inhomogeneous structures, and the presence of noise in acquired images. The effect of preprocessing on the accuracy of segmentation will be discussed. The superior performance of advanced clustering algorithms based on statistical and adaptive approaches over traditional algorithms in medical image segmentation will be presented.

Generally speaking, segmentation techniques are application specific and non-universal. There exists no approach that works best for all types of images. Numerous segmentation techniques have been developed for gray scale images [1-5], while color image segmentation techniques have been created much later than its gray level counterpart because of the computational complexity involved with the latter. However, the availability of fast digital processors in recent times allows easy implementations of

such complex algorithms. Most of the segmentation techniques applied to gray level images can also be extended to color images [6].

Clustering is a pattern recognition technique that has been frequently used in image segmentation. Similar to the variety of approaches in image segmentation, there are numerous clustering techniques with approaches based on statistics, fuzzy logic, neural network, or an integration of the above concepts [5-10]. Two recently developed advanced clustering algorithms, namely, deterministic annealing (DA) [11] and adaptive fuzzy leader clustering (AFLC) [10] have been applied and compared with the performances of other standard well-known algorithms in efficient segmentation of medical images. DA is designed using a statistical framework, while AFLC has a neural network structure embedded with fuzzy optimization. The performances of these two algorithms have been compared with classical clustering techniques such as k-means, and Fuzzy C-means (FCM) [12] in segmenting a few diverse types of medical images.

The major advantage of using clustering for medical image segmentation is that these unsupervised techniques for data partitioning do not require a training set which is not easy to find in most clinical data sets. The two clustering techniques, namely AFLC, and DA used in our study to investigate the effectiveness and accuracy of these techniques in medical image segmentation, can be considered as optimization processes. Both AFLC, and DA do not require an initial guess of the actual number of clusters present in a dataset and thus do not suffer from the instability inherent to traditional and well-known clustering algorithms such as k-means.

Several types of medical images are selected and used as examples of application of clustering in image segmentation. Both k-means and FCM, the well-known clustering algorithms, suffer from the initialization and local minimum problems. Cluster initialization is crucial in yielding satisfactory results. When not initialized properly, a clustering algorithm might be trapped in a local minimum, failing to proceed to the correct cluster. Our experimentations show that with random initialization, both k-means and FCM fail to generate the lesions by segmentation of magnetic resonance images of the brain of a patient affected by multiple sclerosis. AFLC is an automated and adaptive improvement over k-means and FCM by incorporating neural leader clustering and FCM. The performance is improved; however, similar problems are still encountered. Initialization is eliminated by selecting the first incoming sample as the initial centroid, therefore, the outcome is sample-order dependent. DA is the best candidate for medical image segmentation by an advanced clustering technique. It is not sensitive to parameter tuning, and initialization problem, and is noise tolerant and guaranteed to converge.

Advanced clustering techniques can provide general solutions for effective segmentation of a broad range of medical images. All segmentation examples presented here use image intensity as the single feature to demonstrate the efficiency of the clustering algorithms. In real applications, local property or connectivity of adjacent pixel can be embedded into the clustering process to achieve more accurate segmentation [13].

As an example we show 3-D segmentation of the optic disc/cup in the stereo images of the retina of a glaucoma patient, such segmentation is critical in early detection of glaucoma. The onset and progression of glaucoma can usually be found or monitored through the measurement of changes in the optic disc and optic cup area. It can be expressed as the cup-to-disc ratio in diameter (2D) or volume (3-D), for which segmentation of the optic cup/disc in 2-D or 3-D is necessary. The cup-to-disc ratios obtained from 3-D visualization of the optic cup/disc has been found to match closely with those provided by physicians [14]. Semi-automated methods for finding the contours of the optical nerve head (ONH) by digital image analysis attempt to find the disparities of pixels between the fundus stereo pairs in a region including the ONH. Recent studies [14-16] describe in detail the algorithms developed for feature extraction, registration, correlation, and dynamic programming leading to computing disparities based on a non-convergent stereo imaging system. However, fully automated methods of finding precise disparities between a pair of stereoscopic images may still be problematic due to the presence of noise, occlusions, distortions and lack of knowledge of the stereo imaging parameters. Under certain constraints, 3-D surface recovery of the optic cup/disc is possible from a pyramidal surface-matching algorithm based on the recovery of the optimum surface within a 3-D cross-correlation coefficient volume

via a two stage dynamic programming technique. The accuracy of the disparity map algorithm leading to 3-D surface recovery is highly dependent on the initial feature extraction process and the stereo imaging parameters.

The stereo images of a real world scene are taken from two different perspectives. The coordinate associated with depth of this scene can be extracted by triangulation of corresponding points in the stereoscopic images assuming a non-convergent imaging system. However, the geometrical information of the viewing angles when fundus images are taken in a clinical set-up are not documented, thus introducing certain ambiguities in the disparity mapping of the corresponding points in the stereo pairs. The overall process for 3-D visualization of retinal structures from stereo image pairs is complex and includes different matching strategies, area or feature based or a combination of both. Several preprocessing steps are also followed for feature extraction and registration prior to coarse to fine disparity search. Figures 1 and 2 demonstrate the effectiveness of computer-generated segmentation of the optic disc/cup from the stereo disparity maps.

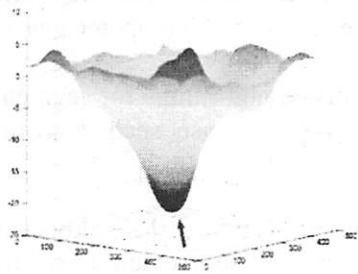
Further research is in progress to develop automated algorithms for segmentation of a variety of medical images, both in gray-scale and in color in order to use such segmentation techniques as diagnostic tools.

References:

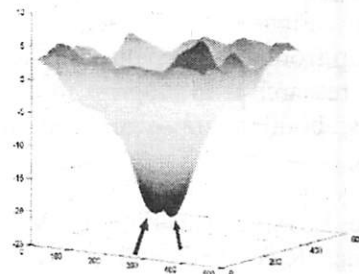
1. Clarke, L. P., Camacho, R. P., Velthuisen, M. A., Heine, J. J., Vaidyanathan, M., Hall, L. O., Thatcher, R. W. and Silbiger, M. L., Review of MRI segmentation: methods and applications, *Magn. Reso. Imaging*, Vol. 13, No. 3, pp. 343-368, 1995.
2. Pal N. R. and Pal, S. K., A review on image segmentation techniques, *Pattern Recognition*, Vol. 26, No. 9, pp. 1277-1294, 1993.
3. Haralick, R.M. and Shapiro, L. G., Image segmentation techniques, *Computer Vision, Graphics, and Image Processing*, Vol. 29, No. 1, pp. 100-132, Jan, 1985.
4. Fu, K. S. and Mui, J. K. A survey on image segmentation, *Pattern Recognition*, Vol. 13, pp. 3-16, 1981.
5. Skarbek, W. and Koschan, A., Colour image segmentation-A survey, Technical report 94-132, Technical University, Berlin, Oct. 1994.
6. L. A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Trans. On Syst. ,Man, Cybern.*, Vol. SMC-3, PP 28-44,1973.
7. Bezdek, J.C., Hall, L.O. and Clarke, L.P. Review of MR image segmentation techniques using pattern recognition, *Med. Phys.*, Vol. 20, pp. 1033-1048, 1993.
8. Carpenter, G. A. and Grossberg, S., A massively parallel architecture for a self-organizing neural pattern recognition machine, *Computer Vision, Graphics and Image Processing*, Vol. 37, pp. 54-115, 1987.
9. Newton, S. C., Pemmaraju, S. and Mitra, S., Adaptive fuzzy leader clustering of complex data sets in pattern recognition, *IEEE Trans. Neural Networks* , Vol. 3, pp.794-800, 1992.
10. Mitra, S., Castellanos, R., Yang, S.Y. and Pemmaraju, S., Adaptive clustering for image segmentation and vector quantization, book chapter in *Soft-computing for Image Processing*, Editors: Pal, S. K., Ghosh, A. and Kundu, M. K. New York: Springer-Verlag, 1999.
11. Rose, K. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems, *Proceedings of the IEEE*, Vol. 86, No. 11, Nov. 1998.
12. Bezdek, J., *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum Press, 1981.
13. Wang, W., Sun, C. and Chao, H., Color image segmentation and understanding through connected components, *Proc. Of 1997 IEEE Int'l Conf. on Systems, Man, and Cybernetics*, Vol. 2. pp. 1089-1093, Orlando, FL, Oct. 12-15, 1997.
14. Enrique Corona, Sunanda Mitra, Mark Wilson, Thomas Krile, Young H. Kwon, Peter Soliz, "Digital Stereo Image Analyzer for Generating Automated 3-D measures of Optic Disc Deformation in Glaucoma", *IEEE Transactions on Medical Imaging*, Special Issue on Image

Processing and Analysis in Drug Discovery and Clinical Trials, Vol.21, No.10, pp1244-1253, October, 2002.

15. Shuyu Yang and Sunanda Mitra, "Content Based Vector Coder for Efficient Information Retrieval", Invited Book Chapter in "New Directions in Enhancing the power of the Internet", Editors: Masoud Nikravesh, B. Azvine, R. Yagar, and Lotfi A. Zadeh. Physica-Verlag, Springer, November, 2003.
16. Shuyu Yang, Sunanda Mitra, Enrique Corona, Brian Nutter, and D. J. Lee, "Multi-level Wavelet Feature Statistics for Efficient Retrieval, Transmission, and Display of Medical Images by Hybrid Encoding", EURASIP Journal on Applied Signal Processing, Vol.5, pp 449-460, 2003.

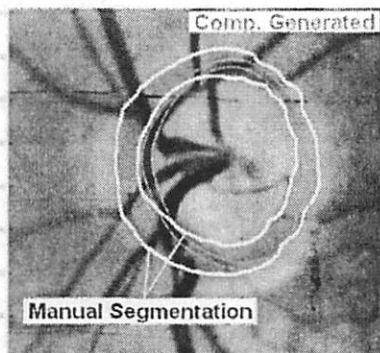


(a) Disparity map obtained from DA blood vessel segmentation.

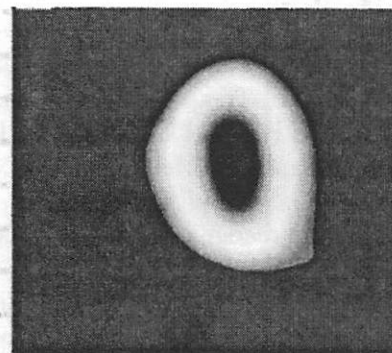


(b) Disparity map obtained from general edge detection for blood vessel segmentation.

Figure 1. Disparity maps generated with and without DA feature extraction



(a)



(b)

Figure 2. Segmentation of optic disc/cup from 3-D disparity map.

Panel Discussion: WEDNESDAY, DEC 17

Bioinformatics

The Panel's focus is to discuss advantage of the wealth of new information and develop new techniques and methodology for intelligent information analysis, knowledge management and knowledge discovery and rule-based reasoning in an environment that is imprecise, uncertain, complex and not known to our knowledge as of today. These tools include computational methods for the identification of functional elements in complex genomes and the identification of patterns in large datasets (for example, microarray data).

Panelists:

M. Berthold, V. C. Georgopoulos, I. Guyon, A. Hunt, N. Kasabov, Y. Zhang

Moderator:

Masoud Nikravesh; University of California, Berkeley

THURSDAY, DECEMBER 18

SEARCH ENGINE & INTERNET

Abstracts of THURSDAY, DECEMBER 18:
Search Engine & Internet

Google™

Peter Norvig

Peter Norvig is the Director of Search Quality at Google Inc.. He is a Fellow and Councilor of the American Association for Artificial Intelligence and co-author of Artificial Intelligence: A Modern Approach, the leading textbook in the field.

Previously he was head of the Computational Sciences Division at NASA Ames Research Center, where he oversaw a staff of 200 scientists performing NASA's research and development in autonomy and robotics, automated software engineering and data analysis, neuro-engineering, collaborative systems research, and simulation-based decision-making. Before that he was Chief Scientist at Jungle, where he helped develop one of the first Internet comparison shopping service; Chief designer at Harlequin Inc.; and Senior Scientist at Sun Microsystems Laboratories.

Dr. Norvig received a B.S. in Applied Mathematics from Brown University and a Ph.D. in Computer Science from the University of California at Berkeley. He has been a Professor at the University of Southern California and a Research Faculty Member at Berkeley. He has over fifty publications in various areas of Computer Science, concentrating on Artificial Intelligence, Natural Language Processing and Software Engineering including the books Paradigms of AI Programming: Case Studies in Common Lisp, Verbomobil: A Translation System for Face-to-Face Dialog, and Intelligent Help Systems for UNIX.

Research and Applications of Soft Computing at BT Exact

Marcus Thint

Advances in AI and Soft Computing are key to realizing the visions of the intelligent enterprise, iBusiness (intelligent e-Business), and web intelligence. BT Exact's Research and Venturing division is driven by applications and development of these key technologies for business solutions. This presentation describes projects in the Computational Intelligence Group, which focuses on software systems and tools for intelligent data analysis and information management. In particular, we discuss the use of soft computing for software agents and adaptive user profiling, data representation, information search and retrieval, document classification, text mining, and intelligent scheduling. Our goals are to increase automated reasoning, improve resource/information utilization, and assist human-machine interaction in the realm of natural language/unstructured data. The systems are targeted at both internal customers (e.g. Contact Centers), and external users.

The Role of User Modelling in Information Retrieval from the WWW

Robert I. John

Information Retrieval from the WWW through the use of search engines is known to be unable to capture effectively the information needs of users. The approach taken in our work is to add intelligence to

information retrieval from the WWW, by the modelling of users to improve the interaction between the user and information retrieval systems. In other words, to improve the performance of the user in retrieving information from the information source. To affect such an improvement, it is necessary that any retrieval system can somehow make inferences concerning the information the user might want. The system then can aid the user, for instance by giving suggestions or by adapting any query based on predictions furnished by the model.

We have taken two separate approaches. Firstly, by a combination of user modelling and fuzzy logic a prototype system has been developed (the Fuzzy Modelling Query Assistant (FMQA)) which modifies a user's query based on a fuzzy user model. The FMQA was tested via a user study which clearly indicated that, for the limited domain chosen, the modified queries are better than those that are left unmodified. As part of a European project, ELVIL, we have developed a Virtual Librarian. The European Legislative Virtual Library (ELVIL) is an Internet-based portal of information sources on European law and politics (<http://elvil.sub.su.se>) which provides software gateways with a uniform WWW-interface to national and European parliamentary databases, a searchable index to other sources of legal and political information on the WWW, and collections of learning resources on European law and politics. Information Retrieval from the WWW through the use of search engines is known to be very ineffective at meeting information needs of users. The approach taken in this work is to add intelligence to information retrieval from the WWW and online databases, by the modelling of users, using fuzzy logic, to improve the interaction between the user and the information retrieval system ELVIL. By a combination of user modelling and fuzzy logic a system has been developed (the Virtual Librarian (VL)) which takes a fuzzy user model and search information to choose the most appropriate databases or web sites within the ELVIL portal.

Evolution of the Laws that Deal with the Utilization of Information Networks

Babak Hodjat, Adam Cheyer and Russ Berg

Today, three Laws are used to explain how the value of a network, comprised of computational power and informational content, increases as the network expands: Moore's Law, Metcalf's Law, and Reed's Law. In order to achieve the exponential benefits promised by the Three Laws, nodes in the value network must possess a shared communication/information model, and there is a cost to acquiring this *knowledge*. In this presentation a new perspective is proposed that explores this effect, evolves the above three laws to take into account a knowledge-acquisition cost constant C , and concludes that technologies that bring this constant as close to zero as possible in real practice are the key to unlocking the exponential value inherent in information networks.

Basic Principles of Information Monitoring Systems

Alexander Ryjov

This article describes main ideas of Information Monitoring Systems (IMS) and applications of IMS in real-world problems. Information monitoring systems relate to a class of hierarchical fuzzy discrete dynamic systems. The theoretical base of such class of systems is made by the fuzzy sets theory, discrete mathematics, and methods of the analysis of hierarchies which was developed independently in works of Zadeh, Messarovich, Saaty and others. IMS address to process uniformly diverse, multi-level, fragmentary, unreliable, and varying in time information about some problem/process. Based on this type

of information IMS allow perform monitoring of the problem's/process' evolution and work out strategic plans of problem/process development. These capabilities open a broad area of applications in business (marketing, management, strategic planning), socio-political problems (elections, control of bilateral and multilateral agreements, terrorism), etc. One of such applications is a system for monitoring and evaluation of state's nuclear activities (department of safeguards, IAEA) has been shortly described in the report.

BeMySearch: Concept-Based Information Retrieval and Search Engine **Tomohiro Takagi and Masoud Nikravesh**

Since a fuzzy set is defined by enumerating its elements and the degree of membership of each element, we can use it to express word ambiguity by enumerating all possible meanings of a word, then estimating the degrees of compatibilities between the word and the meanings.

Based on this approach, we have proposed using conceptual fuzzy sets (CFSs) to represent the various meanings of a concept that change dynamically depending on the context. A CFS (is realized as neural networks in which a node represents a concept and a link represents the strength of the relation between two (connected) concepts. The activation values agreeing with the grades of membership are determined through this associative memory. In a CFS, the meaning of a concept is represented by the distribution of the activation values of the other nodes. The distribution evolves from the activation of the node representing the concept of interest.

This presentation will start with my motivation to propose CFSs and algorithm to generate CFSs. It will describe how it works to represent the context dependent meaning of a word and to measure a conceptual distance between documents. Next, information filtering and image search (Google-Based Search Engine for Multimedia Data) will be introduced as its applications to information retrieval using capability of conceptual matching. Finally we will introduce our approach to enhancing CFSs based on brain architecture.

BeMySearch: Conceptual Search Engine and Navigation for Google™ and Yahoo!, Web Intelligence

Masoud Nikravesh and Tomohiro Takagi

In this presentation, first we will present the role of the fuzzy logic in the Internet. Then we will present an intelligent model that can mine the Internet to conceptually match and rank homepages based on predefined linguistic formulations and rules defined by experts or based on a set of known homepages. The FCM model will be used for intelligent information and knowledge retrieval through conceptual matching of both text and images (here defined as "Concept"). The FCM can also be used for constructing fuzzy ontology or terms related to the context of the query and search to resolve the ambiguity. This model can be used to calculate conceptually the degree of match to the object or query. We will also present the integration of our technology into commercial search engines such as Google™ and Yahoo! as a framework that can be used to integrate our model into any other commercial search engines, or development of the next generation of search engines.

ARSIN

Danis Yadegar

Danis Yadegar founded Arsin Corporation in 1989. Today, Arsin is an industry leader in developing, integrating, and testing systems, hardware, and software that enable companies to harness the power of the Internet for e-Business applications. As president and CEO, Danis is responsible for Arsin's long-term strategy and day-to-day operations. He has more than 20 years experience in all aspects of software design, development, and testing with special emphasis on integration, scalability testing, database, and user interface development.

Danis and Arsin Corporation have spearheaded major strategic software projects for industry leading corporations including: Bank of America, McKessonHBOC, Cisco Systems, Bechtel, Digital Equipment/Tandem, AT&T, Pacific Bell, and others. In addition, Danis recently served as chief systems architect for a Fortune 100 company's Y2K client/server conversion project.

Prior to Arsin Corporation, Danis worked as a software engineer and designer for ROLM/IBM, Micro Focus, HP, and Tandem Computer. He has provided consulting services to Ingres, Informix, and Tandem as well. Danis holds a bachelor's degree in Computer Science, and attended graduate school at the University of Illinois, Urbana-Champaign.

Context-directed information access

Charlie Ortiz

I will present work that represents a departure from the master-slave metaphor of human-system interaction --- in which a user first (precisely) communicates a request that is then followed by a response --- with one based on collaboration in which a user and an application work jointly toward some goal. Such collaboration leads to a somewhat different view of communication, particularly in information-gathering tasks such as those prevalent on the web. In such tasks, the fruits of collaborations take the form of the establishment of what we call an access perspective. An access perspective reflects the inherently multifaceted way in which information at one location may be accessed from another, depending on the intended use of that information and the required level of detail. Access perspectives are dynamic entities that are informed by the structure of a user's activity. Conventional, nonadaptive user interfaces characteristically provide a static access method. I will describe methods for representing a context tree which establishes a particular access perspective; as a user explores a particular branch of a context tree, the context becomes more and more constrained and the access perspective becomes, correspondingly, narrower.

Scalable Web Usage Mining and Soft Computing Approaches for High Performance Intelligent Web Recommender Systems

Olfa Nasraoui

Intelligent Web *personalization* aims at adapting a user's interaction with the Web information space based on information gathered about the user. A complete *automated* Web *personalization* system is

generally based on Web usage mining to *discover useful* knowledge about user access patterns, followed by a *recommendation* system to *act* on this knowledge in order to respond to the users' individual interest, in a manner *transparent* to the user, and while protecting the user's *privacy* and *anonymity*.

Web usage mining has recently attracted attention as a viable framework for extracting useful access pattern information, such as user profiles, from massive amounts of Web log data for the purpose of Web site personalization and organization. These efforts have relied mainly on clustering or association rule discovery as the enabling data mining technologies. Typically, data mining has to be completely re-applied periodically and offline on newly generated Web server logs in order to keep the discovered knowledge up to date. Also, to date, the clustering techniques that have been used for Web usage mining have rather scored low on scalability both in terms of time and memory requirements, and simply cannot be expected to keep up with the huge flux of Web clickstream data on today's busy websites. In addition to their lack of scalability and difficulty to adapt in the face of continuously evolving patterns, current clustering techniques, such as most KMeans variants, also suffer from one or more of the following limitations: requirement of the specification of the correct number of clusters/profiles in advance, sensitivity to initialization, sensitivity to the presence of noise and outliers in the data. Other techniques relying on association rule discovery suffer from their sensitivity to various thresholds such as support and confidence, as well as sparsity of the data. Hence, there is a crucial need for scalable, noise insensitive, initialization independent techniques that can continuously discover possibly changing/evolving Web user profiles without any stoppages or reconfigurations.

We overview some recent evolutionary computation techniques for Web mining. Then, we present a new scalable clustering methodology that gleams inspiration from the natural immune system to be able to continuously learn and adapt to new incoming patterns. The Web server plays the role of the human body, and the incoming requests play the role of foreign antigens/bacteria/viruses that need to be detected/recognized by the proposed immune based clustering technique. Hence, our clustering algorithm plays the role of the cognitive agent of an artificial immune system, whose goal is to continuously perform an intelligent organization of the incoming noisy data into clusters/patterns. Our approach exhibits superior learning abilities, while at the same time, requiring *modest* memory and computational costs. Like the natural immune system, the strongest advantage of immune based learning compared to current approaches is expected to be its ease of adaptation to the dynamic environment that characterizes several applications, particularly in mining data streams. We illustrate the ability of the proposed approach in detecting clusters in noisy data sets, and in mining user profiles from Web clickstream data in a single pass under different usage trend sequence scenarios. Finally, we present several soft computing approaches to high performance mass profile-based web recommender engines that we have recently developed.

Mining opinion

James G. Shanahan

Today, much of product feedback is provided by customers/critiques online through websites, discussion boards, mailing lists, and blogs. When trying to make strategic decisions (a product launch, a purchase), using a websearch will return many useful but heterogeneous and, increasingly, multilingual opinions on a product. Generally, the user will find it very difficult and time consuming to assimilate all available information and make an informed decision. To date most work in automating this process has focused on

a monolingual setting. Existing monolingual approaches will be overviewed. In addition, I will describe our preliminary work on mining product ratings in a multilingual setting. The proposed approaches are automatic, using a combination of techniques from classification and translation, thereby alleviating human-intensive construction and maintenance of linguistic resources.

Constructing Fuzzy Thesauri for the WWW: application to BeMySearch

Martine De Cock, Sergio Guadarrama and Masoud Nikravesh

The remarkable growth of the World Wide Web (WWW) since its origin in the 1990's calls for efficient and effective tools for information retrieval. Attempting to deal with the overwhelming amount of information provided on billions of webpages nowadays does not necessarily imply that we have to develop entirely new technologies from scratch. In the 1970's and 1980's initial research was performed on the retrieval of information from modest text collections, using fuzzy relations to represent dependencies between terms on one hand, and between terms and documents on the other. Since then the fuzzy mathematical machinery (i.e. fuzzy logical operators, fuzzy similarity measures, operations with fuzzy relations) has come of age. We revisit the "old" information retrieval strategies and brush them up with new insights to prepare them for a great challenge: search in the biggest collection of text known by mankind.

On Similarity Measures and Pruning Strategies for Fuzzy Ontologies

Andrey Kolobov, Daniel Kuo, Martine De Cock, Masoud Nikravesh

We critically examine the applicability and usefulness of mathematical models for the construction of fuzzy term-term relations based on fuzzy similarity and fuzzy inclusion measures. Much attention is paid to realistic pruning algorithms of generated fuzzy relations, both semantically as well as regarding execution time.

Genetic Fuzzy Neural Agents Using Type-2 Fuzzy Reasoning for Intelligent Web Information Search Task

Yuchun Tang and Yanqing Zhang

Finding the desired information on the World Wide Web is not an easy task because the information available on the WWW is inherently unordered, distributed, and heterogeneous. As a result, the ability to search and retrieve information from the Web efficiently and effectively is a key technology for realizing the Web's full potential.

Expressing a search request is the first thing we need to solve. At almost all search scenarios, the desired information is on some Web pages. For some similar search requests, the desired Web pages share some similar "content characteristics" and/or "structure characteristics". Traditional search methods let users to submit "keywords" that may be displayed on the desired Web pages to express their search

requests. So we call them "keyword-based search" or "content-based search". Obviously, they are not so powerful. There are 2 problems: 1 in many cases, a search request is inherently fuzzy and thus difficult or even impossible to be expressed by "crisp" keywords. Furthermore, many "partial related" Web pages can not be retrieved in a crisp way. And 2 in many, if not all, cases, some keywords provide some "structure characteristics" while others only provide "content characteristics". Unfortunately, traditional search methods do not differentiate the two kinds of keywords. In the following, "keywords" will be referred as those words that only provide content characteristics.

Retrieving information effectively is another issue we need to think about. Current search engines are known for poor accuracy: they have both low recall and low precision. Furthermore, the most relevant/desired documents are not always displayed at the top of the query result list. It means the query result is not listed in a desired order.

We proposed an intelligent web information search and retrieval model called Web Information Search Task (WIST). WIST model has two goals: one is to make the interface of a search service more expressive and another is to make information retrieval more effective. Many search requests have different content characteristics but share similar structure characteristics. These structure characteristics are expressed by simple "structure rules". Basically, a structure rule is an IF-THEN formula defined on web pages' URL, Title, Text, Input Links, Output Links, or other related sections. If the IF part is satisfied by a web page, it may be desired, otherwise it may be not desired. These structure rules will function as an input fuzzifier so we can make fuzzy reasoning to derive the "desirability", the possibility whether a web page is desired and how much is the possibility, of a Web page. In this way, all search requests with similar (and usually fuzzy) structure characteristics can be "grouped" into a WIST, which is implemented as an intelligent software agent using FL, NN, GA, and other technologies to automatically find all relevant Web pages based on the relevance inferred from structure rules and user-submitted keywords, and rank them in a desired way. Essentially, a WIST agent uses a TSK-based Fuzzy Neural Network (FNN) to infer the desirability. The agent is "intelligent" because it can 1 learn to get better parameters of FNN, 2 learn to get better structure of FNN, and 3 learn to define structure characteristics by adding/modifying structure rules.

Panel Discussion: THURSDAY, DEC 18

Internet and Industry

At the dawn of the new millennium, we can expect dramatic increase in the use of intelligent systems in the internet applications, since we have to deal with an increasing amount of data, that is mainly unstructured, dynamic and designed for human access. Therefore, it is usually hard to extract relevant information automatically. These aspects will be reflected in the subjects treated at this panel discussion.

The main purpose of the discussion panel is to draw the attention of the academic community as well as the industrial community to the fundamental importance of specific Internet-related problems. This issue is critically significant about problems that center on search and deduction in large, unstructured and distributed knowledge bases, and the use of intelligent techniques in e-business and B2B applications. The panel will provide a unique opportunity for the academic and corporate communities to address new challenges, share solutions, and discuss research directions for the future.

Panelists:

B. Hodjat, P. Norvig, T. Takagi, M. Thint, J. Shanahan, D. Yadegar

Moderator:

Masoud Nikravesh; University of California, Berkeley

SPEAKERS

Speakers



Prof. Ruzena Bajcsy
CITRIS Director
284 HMMB, University of California at Berkeley, CA 94720-1776
email: bajcsy@eecs.berkeley.edu
www: <http://www.citris.berkeley.edu/>
phone: +1 (510) 642-9423



Dr. Michael R. Berthold
Director of Data Analysis, Tripos Inc.
President, NAFIPS
601 Gateway Blvd, Suite 720
South San Francisco, CA 94080, USA
email: berthold@tripos.com
www: <http://www.icsi.berkeley.edu/~berthold/>
phone: +1 (650) 553-8004, fax: +1 (650) 553-8005



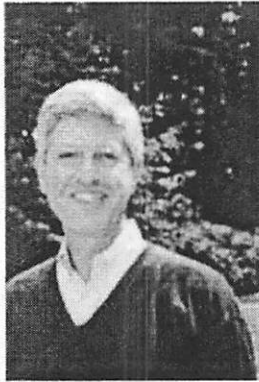
Dr. Miguel Delgado
Department of Computer Science and Artificial Intelligence,
University of Granada, 18071 Granada, Spain
email: mdelgado@ugr.es

Dr. Chris Ding
Staff Computer Scientist
NERSC Division, Building 50F, Rm 1608
Lawrence Berkeley National Lab
1 Cyclotron Rd, Berkeley, CA 94720
email: CHQDing@lbl.gov
www: <http://www.nersc.gov/~cding>
phone: 510-486-6901, fax: 510-486-5812

Prof. Adam Gaweda
Instructor of Medicine
Kidney Disease Program
Health Sciences Center, University of Louisville
615 South Preston
Louisville, Kentucky 40202
email: Adam.Gaweda@kdpnw.kdp-baptist.louisville.edu
www: <http://ci.uofl.edu/adam/>
phone: (502) 852-0766, fax: (502) 952-7643



Prof. Voula C. Georgopoulos
Professor of Informatics
Dept. of Speech and Language Therapy
Technological Educational Institute of Patras
Patras, Greece
and
Health Sector Head
Technological Research Center of Patras
email: voulag@otenet.gr
phone: +306972790344



Prof. William A. Gruver
Intelligent Robotics and Manufacturing Systems Laboratory
School of Engineering Science
Simon Fraser University
Burnaby, British Columbia, Canada V5A 1S6
email: gruver@cs.sfu.ca
www: <http://www.ensc.sfu.ca/research/irms/personnel/gruver.html>
phone: 604-291-4339, fax: 604-291-4951



Dr. Isabelle Guyon
ClopiNet
955 Creston road
Berkeley, CA 94708, USA
email: isabelle@clopinet.com
www: <http://clopinet.com/isabelle/>
<http://clopinet.com/isabelle/Projects/NIPS2003/#challenge>
phone: (510) 524-6211



Babak Hodjat

CTO

Dejima Inc. (www.dejima.com)

160 W Santa Clara St. #102

San Jose, CA 95113

email: Babak.Hodjat@dejima.com

phone: (408) 535 0850 x 4515, fax: (408) 535 0859



Prof. C. Anthony Hunt

Associate Professor

Department of Biopharmaceutical Sciences

School of Pharmacy

513 Parnassus Ave., S926

University of California

San Francisco, CA 941430446

email: hunt@itsa.ucsf.edu

www: <http://bioeng.berkeley.edu/graduate/cvs/Hunt.html>

phone: (415) 476-2455/476-2657, fax: (415) 514-2008



Prof. Mo Jamshidi

Regents Professor of Electrical and Computer Engineering

AT&T Professor of Manufacturing Engineering

Director, Autonomous Control Engineering (ACE) Center

The University of New Mexico

Albuquerque, NM 87131 USA

email: jamshidi@eece.unm.edu, moi@cybermesa.com

www: <http://ace.unm.edu/>



Dr. Robert I. John
Reader in Computer Science
Head of Division of Artificial Intelligence & Computer Modelling
De Montfort University
email: rij@dmu.ac.uk
www: <http://www.cse.dmu.ac.uk/~rij>
phone: (44) 116 2078491, fax: (44) 116 2541891



Prof. Nikola Kasabov
Director, Knowledge Engineering and Discovery Research Institute
Chair of Knowledge Engineering, School of Computer and Information Sciences
Auckland University of Technology (AUT), New Zealand
email: nkasabov@aut.ac.nz
www <http://www.kedri.info>
phone: +64 9 917 9506 ; fax: +64 9 917 9501



Prof. James Keller

Professor, Electrical and Computer Engineering Department
Editor-in-Chief, IEEE TFS
217 Engineering Building West
University of Missouri-Columbia
Columbia, MO 65211-2060
email: KellerJ@missouri.edu
www: <http://www.missouri.edu/~kellerj>
phone: (573) 882-7339, fax: (573) 882-0397



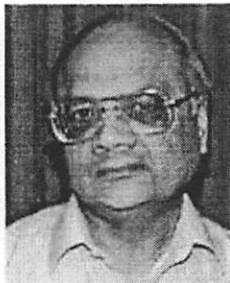
Prof. Mohammad Khoshnevisan
Griffith University
Gold Coast Campus
Queensland, Australia
email: m.khoshnevisan@griffith.edu.au
www: <http://www.gu.edu.au/text/school/acf/Staff/khoshnevisan.html>
phone: (61)7-555-2876



Dr. Galina Korotkich
Faculty of Informatics and Communication
Mackay Campus
Central Queensland University
Mackay QLD 4740, Australia
email: g.korotkich@cqu.edu.au



Prof. Victor Korotkich
Associate Professor
Faculty of Informatics and Communication
Mackay Campus
Central Queensland University
Mackay QLD 4740, Australia
email: v.korotkich@cqu.edu.au
phone: + 61 7 4940 7455, fax: + 61 7 4940 7407



Dr. Dilip B. Kotak
System Integration, Testing & Evaluation
National Research Council Canada Institute for Fuel Cell Innovation
3250 East Mall, Vancouver, B.C. V6T 1W5, Government of Canada.
email: Dilip.Kotak@nrc-cnrc.gc.ca
phone: 604-221-3045, fax: 604-221-3001
and Adjunct Professor
Intelligent Robotics and Manufacturing Systems Laboratory
School of Engineering Science, Simon Fraser University
Burnaby, BC V5A 1S6 Canada
www: <http://www.ensc.sfu.ca/research/irms/personnel/kotak.html>

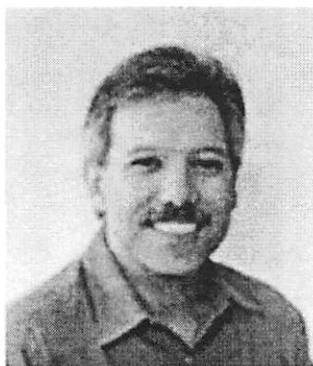


Prof. Lawrence J. Mazlack
Associate Professor

ECECS Department
University of Cincinnati
Cincinnati, Ohio
email: mazlack@uc.edu
www: <http://www.ececs.uc.edu/~mazlack/>



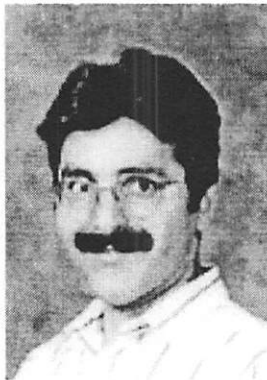
Dr. John A. Meech
Professor of Mineral Processing and Director of CERM3,
(The UBC Centre for Environmental Research in Minerals, Metals, and Materials),
Department of Mining Engineering,
The University of British Columbia,
6350 Stores Road,
Vancouver, B.C., V6T 1Z4, CANADA
Email: cerm3dir@mining.ubc.ca or jameech@dccnet.com
www: <http://www.mining.ubc.ca/faculty/meech/>
Phone: (604) 822-3984 Fax: (604) 822-5599



Dr. Juan C. Meza
Department Head
High Performance Computing Research
Lawrence Berkeley National Laboratory
One Cyclotron Road, MS: 50B-2239
Berkeley, CA 94720
email: JCMeza@lbl.gov
www: <http://crd.lbl.gov/~meza>
phone: 510 486-7684, fax: 510 495-2998



Prof. Sunanda Mitra
Professor and Director of Computer Vision and Image Analysis Laboratory
Department of Electrical and Computer Engineering
Mail Stop 3102
Texas Tech University
Lubbock, Texas 79409-3102
email: Sunanda.Mitra@coe.ttu.edu
www: <http://www.coe.ttu.edu/ee/mitra.htm>
phone: (806) 742-1381, fax : (806) 742-1245



Prof. Saeid Nahavandi
School of Engineering and Technology
Deakin University,
Waurin Ponds Campus,
Geelong 3217, Australia
email: nahavand@deakin.edu.au
www: <http://www.deakin.edu.au/engineering/research/intelligentsa/>
phone: (61)3-522-71231, fax: 61 3 52272167

Prof. Olfa Nasraoui
Assistant Professor,
Department of Electrical & Computer Engineering
206 Engineering Science Bldg.
The University of Memphis
Memphis, TN 38152-3180
email: onasraou@memphis.edu
www: <http://archer.ee.memphis.edu/www.ee.memphis.edu/people/faculty/nasraoui/nasraoui.html>
phone: (901) 678-3251, fax (901) 678-5469



Dr. Peter Norvig
Director of Search Quality Google Inc.
email: pnorvig@google.com
www: <http://www.norvig.com/>
phone: (650) 623-4248, fax: (650) 618-1499

Dr. Frank Olken
Lawrence Berkeley National Laboratory
Computational Research Division
Mailstop 50B-3238
1 Cyclotron Road
Berkeley, CA 94720-8147
email: olken@lbl.gov
www: <http://www.lbl.gov/~olken>
phone: 510-486-5891, fax: 510-486-4004



Dr. Charles Ortiz
Program Manager
Teambotics: Collaborative software and robotic environments
Artificial Intelligence Center
SRI International
333 Ravenswood Ave, Menlo Park, CA 94025
e-mail: ortiz@AI.SRI.COM
www: <http://www.ai.sri.com/~ortiz/>
phone: (650) 859-4461, fax: (650) 859-3735



Prof. Sankar K. Pal
Fellow IEEE, Distinguished Scientist & Head
Machine Intelligence Unit, Indian Statistical Institute
203, B.T.Road, Kolkata 700 108, INDIA
email: sankar@isical.ac.in, skpal@ieee.org
www: <http://www.isical.ac.in/~sankar>
phone:(+91)(+33) 2578-3357(Direct), 2577-8085/8086/2088 Extn.3101 (O)
fax:(+91)(+33) 2578-3357, 2577-3035/6680/6925



Dr. Enrique H. Ruspini
Principal Scientist
SRI Institute Fellow
Artificial Intelligence Center
SRI International, 333 Ravenswood Avenue
Menlo Park, California 94025, USA
email: ruspini@ai.sri.com
www: <http://www.ai.sri.com/~ruspini>
phone: 650 859 2314, fax: 650 859 3735



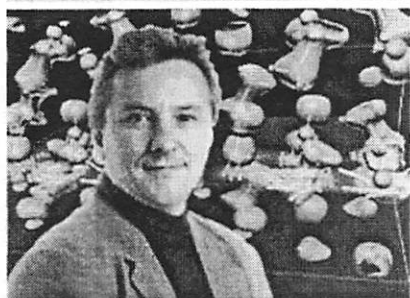
Prof. Alexander Ryjov
Chair of Mathematical Foundations of Intelligent Systems
Department of Mechanics and Mathematics
Lomonosov Moscow State University, Russia
email: ryjov@mech.math.msu.su



Dr. James G. Shanahan
Senior Research Scientist
Clairvoyance Corporation
5001 Baum Boulevard Suite 700
Pittsburgh, PA 15213-1854
email: J.shanahan@clairvoyancecorp.com
www: www.clairvoyancecorp.com
phone: 412-621-0570 x 310, fax: 412-621-0569



Prof. Amit Sheth
Large Scale Distr. Info. Sys. Lab
Department of Computer Science
University of Georgia
415 Graduate Studies Research Center
Athens GA 30602-7404
email: amit@cs.uga.edu
www: <http://lsdis.cs.uga.edu/~amit>
phone: (706) 542-2310, fax: (706) 542-2966



Dr. Horst D. Simon
Director, National Energy Research Scientific Computing (NERSC) Center
Director, Computational Research Division
Lawrence Berkeley National Laboratory

One Cyclotron Road, M/S: 50B-4230
Berkeley, California 94720
email: hdsimon@lbl.gov
www: <http://www.nersc.gov/~simon/>
phone: (510) 486-7377, fax: (510) 486-4300



Prof. Florentin Smarandache
Math Department, University of New Mexico,
200 College Road, Gallup, NM 87301, USA.
email: smarand@unm.edu
phone: (505) 863-7647, fax: (505) 863-7532



Prof. Thomas Sudkamp
Department of Computer Science
Wright State University
Dayton, OH 45435
email: tsudkamp@cs.wright.edu
www: <http://www.cs.wright.edu/~tsudkamp>
phone: 937-775-5118, fax: 937-775-5133



Yuchun Tang
Ph.D. Candidate
email: tyczjs@yahoo.com
www: <http://www.student.gsu/~ytang3>



Dr. Marcus Thint
BT Exact
9749 Vineyard Court
Boca Raton, FL 33428, USA
email: marcus.2.thint@bt.com
phone: 561-451-8081 , fax: 561-451-8081



Prof. S-M. Vincent Tseng
Assistant Professor
Department of Computer Science and Information Engineering
National Cheng Kung University
Tainan, Taiwan, R.O.C.
Co-Director, Database Lab, NCKU, Taiwan
Co-Director, BISC SIG on Data Mining, UC-Berkeley, USA
email: tsengsm@mail.ncku.edu.tw
www: <http://dbserver.csie.ncku.edu.tw/~tsengsm>
phone: 06-2757575 ext 62536, fax: 06-2747076

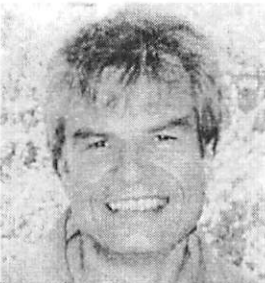


Prof. Burhan Turksen
Director-Knowledge/Intelligence Systems Laboratory
Department of Mechanical & Industrial Engineering
University of Toronto, Toronto, Ontario
M5S 3G8, Canada

email: turksen@mie.utoronto.ca
www: <http://www.mie.utoronto.ca/staff/profiles/turksen.html>
phone: (416) 978-1278, fax: (416) 946-7581



Dr. Mihaela Ulieru
Associate Professor
Department of Electrical and Computer Engineering
Faculty of Engineering
The University of Calgary
Calgary, Alberta, Canada T2N 1N4
email: ulieru@enel.ucalgary.ca
www: <http://www.ucalgary.ca/~ulieru>
phone: (403) 220-8616, fax: (403) 282-6855



Prof. Dr. Rainer Unland
University of Duisburg-Essen
Institute for Computer Science and Business Information Systems (ICB)
Practical Computer Science, especially Data Management Systems and Knowledge Representation
Schuetzenbahn 70, 45117 Essen, Germany
email: UnlandR@informatik.uni-essen.de
www: <http://davis.informatik.uni-essen.de/site/staff/unlandr/>
phone: (+49) 201-183 3421, fax: (+49) 201-183 4460



Prof. Paul P. Wang
ECE Department
Box 90291
Duke University
Durham, NC, 27708-0291
email: ppw@ee.duke.edu
phone: 919-660-5259, fax: 919-660-5293

Danis Yadegar
Founder, President, Chairman, and CEO
ARSIN Corporation
4800 Great America Parkway, Suite 425
Santa Clara, CA 95054-1228, USA
email: dyadegar@arsin.com
www: <http://www.arsin.com/>

Prof. Syozo Yasui, Dr. Eng.
Division of Higher Brain Functions,
Computational and Cognitive Brain Science,
Kyushu Institute of Technology,
Kitakyushu 808-0196, Japan
email: yasui@brain.kyutech.ac.jp
www: <http://www.brain.kyutech.ac.jp/~yasui>
phone: +81-93-695-6108



Dr. Yanqing Zhang
Assistant Professor
Dept. of Computer Science
Georgia State University, U.S.A.
email: yzhang@cs.gsu.edu
www: <http://www.cs.gsu.edu/~cscyzqz>
phone: 404 651 0682, fax: 404 463 9912



Prof. Lotfi A. Zadeh

Director, Berkeley Initiative in Soft Computing (BISC)

Professor in the Graduate School, Computer Science Division

Department of Electrical Engineering and Computer Sciences

University of California

Berkeley, CA 94720 -1776

email: zadeh@cs.berkeley.edu

www: <http://www.cs.berkeley.edu/People/Faculty/Homepages/zadeh.html>

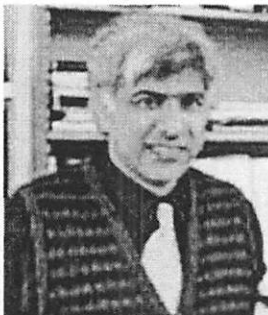
phone(office): (510) 642-4959, fax (office): (510) 642-1712

phone(home): (510) 526-2569, fax (home): (510) 526-2433, (510) 526-5181

BISC Homepage URLs:

URL: <http://www-bisc.cs.berkeley/>

URL: <http://zadeh.cs.berkeley.edu/>



Dr. Masoud Nikravesh

BISC Associate Director

BTEExact Technologies (British Telecom-BT); Senior Research Fellow

Chair: BISC-FLINT, ES, and RT, CIBI

Berkeley Initiative in Soft Computing (BISC)

Computer Science Division- Department of EECS

University of California, Berkeley, CA 94720

email: Nikravesh@cs.berkeley.edu

www: <http://www-bisc.cs.berkeley.edu/>

phone: (510) 643-4522; fax: (510) 642-5775

Staff Scientist

Lawrence Berkeley National Lab,

Imaging and Collaborative Computing Group

email: Masoud@media.lbl

www: <http://vision.lbl.gov/>



Prof. Tomohiro Takagi
Department of Computer Science,
Meiji University
1-1-1 Higashi-Mita, Tama-ku,
Kawasaki-shi, Kanagawa-ken 214-8571 Japan
email: Takagi@cs.meiji.ac.jp
www: <http://www.hi.cs.meiji.ac.jp/>
phone: +81-44-934-7469



Prof. Sema Alptekin
Berkeley Initiative in Soft Computing (BISC)
Computer Science Division- Department of EECS
University of California, Berkeley, CA 94720
email: alptekin@eecs.berkeley.edu
www: <http://www-bisc.cs.berkeley.edu/>
phone: (510) 642-5047; fax: (510) 642-5775



Dr. Souad Souafi-Bensafi
Berkeley Initiative in Soft Computing (BISC)
Computer Science Division- Department of EECS
University of California, Berkeley, CA 94720
email: sbensafi@eecs.berkeley.edu
www: <http://www.cs.berkeley.edu/~sbensafi/>
phone: (510) 643-4523; fax: (510) 642-5775



Dr. Martine De Cock
Berkeley Initiative in Soft Computing (BISC)
Computer Science Division- Department of EECS
University of California, Berkeley, CA 94720
email: mdecock@eecs.berkeley.edu
www: <http://www.cs.berkeley.edu/~mdecock/>
phone: (510) 642-3909; fax: (510) 642-5775



Sergio Guadarrama
Berkeley Initiative in Soft Computing (BISC)
Computer Science Division- Department of EECS
University of California, Berkeley, CA 94720
email: squada@eecs.berkeley.edu
www: <http://www.cs.berkeley.edu/~squada/>
phone: (510) 642-3909; fax: (510) 642-5775

Dr. Mahnaz Hassibi
Berkeley Initiative in Soft Computing (BISC)
Computer Science Division- Department of EECS
University of California, Berkeley, CA 94720
email: mhassibi@cs.berkeley.edu
www: <http://www-bisc.cs.berkeley.edu/>
phone: (510) 642-5047, fax: (510) 642-5775

Gamil Serag-Eldin
Berkeley Initiative in Soft Computing (BISC)
Computer Science Division- Department of EECS
University of California, Berkeley, CA 94720
email: gamil@eecs.berkeley.edu
www: <http://www-bisc.cs.berkeley.edu/>
phone: (510) 643-4524; fax: (510) 642-5775

F. Olcay Cirit
BISC - CS199 Student
email: olcay@uclink.berkeley.edu



Daniel Kuo
BISC - CS199 Student
email: dkuo@uclink.berkeley.edu

Andrey Kolobov
BISC - CS199 Student
email: kolobok83@hotmail.com

Yonathan Asfaw
BISC - CS199 Student
email: yonathan@uclink.berkeley.edu

Perception Based Information Processing and Retrieval Soft Computing and Computational Intelligence Technology A project funded by BText Technologies

 **Lotfi. A. Zadeh** is a Professor in the Graduate School, Computer Science Division, Department of EECS, University of California, Berkeley. In addition, he is serving as the Director of BISC (Berkeley Initiative in Soft Computing). His current research is focused on fuzzy logic, computing with words and soft computing, which is a coalition of fuzzy logic, neurocomputing, evolutionary computing, probabilistic computing and parts of machine learning. Lotfi Zadeh is a Fellow of the IEEE, AAAS, ACM, AAAI, and IFSA. He is a member of the National Academy of Engineering and a Foreign Member of the Russian Academy of Natural Sciences. He is a recipient of the IEEE Education Medal, the IEEE Richard W. Hamming Medal, the IEEE Medal of Honor, the ASME Rufus Oldenburger Medal, the B. Bolzano Medal of the Czech Academy of Sciences, the Kampe de Fariet Medal, the AACC Richard E. Bellman Central Heritage Award, the Grigore Moisil Prize, the Honda Prize, the Okawa Prize, the AIM Information Science Award, the IEEE-SMC J. P. Wohl Career Achievement Award, the SOFT Scientific Contribution Memorial Award of the Japan Society for Fuzzy Theory, the IEEE Millennium Medal, the ACM 2000 Allen Newell Award, and other awards and honorary doctorates. In 2003, Dr. Zadeh was elected as a foreign member of the Finnish Academy of Sciences, and received the Norbert Wiener Award of the IEEE Society of Systems, Man and Cybernetics "For pioneering contributions to the development of system theory, fuzzy logic and soft computing." Dr. Zadeh is a recipient of seventeen honorary doctorates from: Paul-Sabatier University, Toulouse, France; State University of New York, Binghamton, NY; University of Dortmund, Dortmund, Germany; University of Oviedo, Oviedo, Spain; University of Granada, Granada, Spain; Lakehead University, Canada; University of Louisville, KY; Baku State University, Azerbaijan; the Silesian Technical University, Gliwice, Poland; the University of Toronto, Toronto, Canada; the University of Ostrava, the Czech Republic; the University of Central Florida, Orlando, FL; the University of Hamburg, Hamburg, Germany; the University of Paris(6), Paris, France; Johannes Kepler University, Linz, Austria;; and the University of Aurel Vlaicu, Arad, Romania; and the University of Waterloo, Waterloo, Canada. Dr. Zadeh has single-authored over two hundred papers and serves on the editorial boards of over fifty journals. He is a member of the Advisory Committee, Center for Education and Research in Fuzzy Systems and Artificial Intelligence, Iasi, Romania; Senior Advisory Board, International Institute for General Systems Studies; the Board of Governors, International Neural Networks Society; and

is the Honorary President of the Biomedical Fuzzy Systems Association of Japan and the Spanish Association for Fuzzy Logic and Technologies. In addition, he is a member of the Advisory Board of the National Institute of Informatics, Tokyo; a member of the Governing Board, Knowledge Systems Institute, Skokie, IL; an honorary member of the Academic Council of NAISO-IAAC; and Honorary President of North European Society for Adaptive and Intelligent Systems (NSAIS).



Masoud Nikravesh received his B.Sc from Abadan Institute of Technology, MS and PhD in Chemical Engineering from the University of South Carolina (August 1993 and Dec.1994) and received full scholarship from University of California, Berkeley for his second Ph.D. in Material Sciences and Mineral Engineering Department for Fall 1994, when he decided to start his scientific career as Postdoc researcher in Spring 1995 at University of California-Berkeley and Lawrence Berkeley National Lab as joint appointment. Dr. Nikravesh is the BISC Deputy Director, BText technology Senior Fellow and BISC Program Manager in the Computer Science Division, Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley and Research Scientist in the Imaging and Informatics Group at NERSC (National Energy Research Scientific Computing Division, Lawrence Berkeley National Laboratory). In addition, he is serving as the Associate Director (Co-founder) of Zadeh Institute for Information Technology (Information Technology) and Chairs of BISC-Earth Sciences, BISC-Fuzzy Logic and Internet, and BISC-Recognition Technology Groups. His credentials have led to front-page news at Lawrence Berkeley National Laboratory News and headline news at the Electronics Engineering Times. Dr. Nikravesh is the LBNL-NERSC (National Energy Research Scientific Computing Division) representative to the DIMI Executive Committee, and Member of the Executive and Research Council Committee of UC Discovery Grants.

Cal **Souad Souafi-Bensafi** received her engineer degree in Computer Science from Université de Sidi-Bel-Abbès, Algeria in 1994. In 1997, she obtained a M.Sc. degree in Computer Science from Université Claude Bernard, Lyon, France. In 2002, she received her Ph.D degree in Computer Science from INSA de Lyon, France and Université Laval, Québec, Canada. She is now working as a postdoctoral researcher at BISC, EECS Department, UC Berkeley. Her research interests include document structure recognition, pattern recognition, evolutionary computation, machine learning, bayesian networks, fuzzy logic.

Cal **Martine De Cock** received a M.Sc. degree in Computer Science and a teaching degree at Ghent University (Belgium) in 1998. In 2002 she obtained a Ph.D. in Computer Science with a thesis on the representation of linguistic modifiers in fuzzy set theory. She works in the Fuzziness and Uncertainty Modeling Research Unit (Ghent University) as a postdoctoral researcher of the Fund for Scientific Research - Flanders (FWO). Currently she is a visitor at the BISC-group, University of California, Berkeley (USA). Her research interests include rough set theory, L-fuzzy set theory, and fuzzy relations.

Cal **Sergio Guadarrama** was born in Madrid-Spain in 1975. He graduated with B.S. degree in Computer Science at the Technical University of Madrid in 1999. He immediately obtain Research grant for 6 months from the Technical University of Madrid and European Union to work on the ADVICE research project. After that he earn a Ph.D. Fellowship by the Education Department of Spain. Since 2000, he is a graduate student at the Department of Artificial Intelligence of the Technical University of Madrid. He received his Masters degree in Artificial Intelligence in 2002. For his Master and Ph.D. thesis, Sergio Guadarrama has worked and works with Prof. Enric Trillas in the Fuzzy Group. Since August 2003 he is a visiting student researcher at BISC for 10 months. His research has been focused in the areas of Fuzzy Logic, Fuzzy Sets and Multi-Agent Systems. He is currently working in his Ph.D. thesis, which is focused on Consensus in Multi-Agent systems using Fuzzy Logic.

Cal **Sema Alptekin** is a Professor at Cal Poly – San Luis Obispo where she served as the chair of the Industrial and Manufacturing Engineering Department for the past six years. Prior to joining Cal Poly, she was a faculty member at the University of Missouri-Rolla, and worked at General Motors Technical Center. During academic year 2003-2004, she is on sabbatical from Cal Poly, conducting research as a member of BISC, UC Berkeley.

Cal **Mahnaz Hassibi** received her Ph.D. in Petroleum/Reservoir Engineering from University of Southern California in 1999. Her thesis focused on pattern classification, similarity analysis, and signal processing for geophysical data. Currently she is a Post Doctoral Research Engineer in the BISC program, EECS department, University of California Berkeley (USA). Her research interest includes statistical learning techniques, such as Support Vector Machines, and their applications in pattern classification problems.

Cal **Gamil Serag-Eldin** received his B.Sc. from Cairo University, M.Sc. from Carnegie-Mellon University, and to receive his Ph.D. from University of California, Berkeley (Dec 2003) in Architecture. His area of specialization are Computer Aided Design (CAD) & Design Methods and Theories. Dr. Serag is currently working as a researcher at the Berkeley Initiative in Soft Computing, Computer Science Division, EECS Department, UC Berkeley. His research interests include design methods and theories; computational design methods; computational model for negotiation and conflict management; machine learning; evaluation and prediction in design, decision analysis; search methods, shape grammars and case-based reasoning. He is member of ACM, IEEE, ARCADIA, and AIA.

BTexact **Nader Azarmi** is BTexact chief AI technologist and head of Intelligent Systems Lab. Nader has been involved in AI R&D within BT for over 11 years and has successfully applied AI to a number of key BT operations. He has been awarded a number of International and national awards for his work in application of AI in industry, notably INFORMS Franz Edelman medal. Prior to joining BT, he spent many years as an AI researcher and academic within the UK universities.

BTexact **Ben Azvine** is BTexact technical R&D manager and head of computational intelligence group. Ben joined BT in 1995. He leads an international R&D program focusing on soft computing and computational intelligence technology and its application to software agents, intelligent interfaces and intelligent data analysis. One of his projects called the Intelligent Personal Assistant won the British Computer Society's medal for IT in 1998.



**2003 BISC FLINT-CIBI INTERNATIONAL
JOINT WORKSHOP ON SOFT COMPUTING
FOR INTERNET AND BIOINFORMATICS**

Berkeley Initiative in Soft Computing (BISC)
Computer Science Division
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, CA 94720-1776
Tel: (510) 643-4522, FAX: 510-642-5775
Email: zadeh@cs.berkeley.edu
Email: nikravesh@cs.berkeley.edu
URL: <http://www-bisc.cs.berkeley.edu/>

