

Shuichi Iwata, Yukio Ohsawa, Shusaku Tsumoto, Ning Zhong, Yong Shi
and Lorenzo Magnani (Eds.)

Communications and Discoveries from Multidisciplinary Data

Studies in Computational Intelligence, Volume 123

Editor-in-chief

Prof. Janusz Kacprzyk

Systems Research Institute

Polish Academy of Sciences

ul. Newelska 6

01-447 Warsaw

Poland

E-mail: kacprzyk@ibspan.waw.pl

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

Vol. 102. Carlos Cotta, Simeon Reich, Robert Schaefer and Antoni Ligeza (Eds.)

Knowledge-Driven Computing, 2008

ISBN 978-3-540-77474-7

Vol. 103. Devendra K. Chaturvedi

Soft Computing Techniques and its Applications in Electrical Engineering, 2008

ISBN 978-3-540-77480-8

Vol. 104. Maria Virvou and Lakhmi C. Jain (Eds.)

Intelligent Interactive Systems in Knowledge-Based Environment, 2008

ISBN 978-3-540-77470-9

Vol. 105. Wolfgang Guenther

Enhancing Cognitive Assistance Systems with Inertial Measurement Units, 2008

ISBN 978-3-540-76996-5

Vol. 106. Jacqueline Jarvis, Dennis Jarvis, Ralph Rönquist and Lakhmi C. Jain (Eds.)

Holonic Execution: A BDI Approach, 2008

ISBN 978-3-540-77478-5

Vol. 107. Margarita Sordo, Sachin Vaidya and Lakhmi C. Jain (Eds.)

Advanced Computational Intelligence Paradigms in Healthcare - 3, 2008

ISBN 978-3-540-77661-1

Vol. 108. Vito Trianni

Evolutionary Swarm Robotics, 2008

ISBN 978-3-540-77611-6

Vol. 109. Panagiotis Chountas, Ilias Petrounias and Janusz Kacprzyk (Eds.)

Intelligent Techniques and Tools for Novel System Architectures, 2008

ISBN 978-3-540-77621-5

Vol. 110. Makoto Yokoo, Takayuki Ito, Minjie Zhang, Juhnyoung Lee and Tokuro Matsuo (Eds.)

Electronic Commerce, 2008

ISBN 978-3-540-77808-0

Vol. 111. David Elmakias (Ed.)

New Computational Methods in Power System Reliability, 2008

ISBN 978-3-540-77810-3

Vol. 112. Edgar N. Sanchez, Alma Y. Alanís and Alexander G. Loukianov

Discrete-Time High Order Neural Control: Trained with Kalman Filtering, 2008

ISBN 978-3-540-78288-9

Vol. 113. Gemma Bel-Enguix, M. Dolores Jiménez-López and Carlos Martín-Vide (Eds.)

New Developments in Formal Languages and Applications, 2008

ISBN 978-3-540-78290-2

Vol. 114. Christian Blum, Maria José Blesa Aguilera, Andrea Roli and Michael Sampels (Eds.)

Hybrid Metaheuristics, 2008

ISBN 978-3-540-78294-0

Vol. 115. John Fulcher and Lakhmi C. Jain (Eds.)

Computational Intelligence: A Compendium, 2008

ISBN 978-3-540-78292-6

Vol. 116. Ying Liu, Aixin Sun, Han Tong Loh, Wen Feng Lu and Ee-Peng Lim (Eds.)

Advances of Computational Intelligence in Industrial Systems, 2008

ISBN 978-3-540-78296-4

Vol. 117. Da Ruan, Frank Hardeman and Klaas van der Meer (Eds.)

Intelligent Decision and Policy Making Support Systems, 2008

ISBN 978-3-540-78306-0

Vol. 118. Tsau Young Lin, Ying Xie, Anita Wasilewska and Churn-Jung Liao (Eds.)

Data Mining: Foundations and Practice, 2008

ISBN 978-3-540-78487-6

Vol. 119. Slawomir Wiak, Andrzej Krawczyk and Ivo Dolezel (Eds.)

Intelligent Computer Techniques in Applied Electromagnetics, 2008

ISBN 978-3-540-78489-0

Vol. 120. George A. Tsihrintzis and Lakhmi C. Jain (Eds.)

Multimedia Interactive Services in Intelligent Environments, 2008

ISBN 978-3-540-78491-3

Vol. 121. Nadia Nedjah, Leandro dos Santos Coelho and Luiza de Macedo Mourelle (Eds.)

Quantum Inspired Intelligent Systems, 2008

ISBN 978-3-540-78531-6

Vol. 122. Tomasz G. Smolinski, Mariofanna G. Milanova and Aboul-Ella Hassanien (Eds.)

Applications of Computational Intelligence in Biology, 2008

ISBN 978-3-540-78533-0

Vol. 123. Shuichi Iwata, Yukio Ohsawa, Shusaku Tsumoto, Ning Zhong, Yong Shi and Lorenzo Magnani (Eds.)

Communications and Discoveries from Multidisciplinary Data, 2008

ISBN 978-3-540-78732-7

Shuichi Iwata
Yukio Ohsawa
Shusaku Tsumoto
Ning Zhong
Yong Shi
Lorenzo Magnani
(Eds.)

Communications and Discoveries from Multidisciplinary Data

With 121 Figures and 57 Tables

 Springer

Shuichi Iwata
Graduate School of Engineering
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-8656
Japan
iwata@k.u-tokyo.ac.jp

Yukio Ohsawa
Department of Systems Innovation
School of Engineering
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-8656
Japan
ohsawa@sys.t.u-tokyo.ac.jp

Shusaku Tsumoto
Department of Medical Informatics
School of Medicine
Shimane University
Enya-cho Izumo City
Shimane 693-8501
Japan
tsumoto@computer.org

Ning Zhong
Department of Information Engineering
Maebashi Institute of Technology
460-1, Kamisadori-Cho
Maebashi-City, 371-0816
Japan
zhong@maebashi-it.ac.jp
and Director of the WICI/BJUT

Yong Shi
Director of Research Center on Data
Technology and Knowledge Economy
Chinese Academy of Sciences
Beijing 100080, PR China
and
College of Information
Science and Technology
University of Nebraska at Omaha
NE 68182, USA
yshi@mail.unomaha.edu

Lorenzo Magnani
Department of Philosophy
University of Pavia
Piazza Botta 6, 27100 Pavia, Italy
lmagnani@unipv.it

ISBN 978-3-540-78732-7 e-ISBN 978-3-540-78733-4
DOI: 10.1007/978-3-540-78733-4

Studies in Computational Intelligence ISSN 1860-949X

Library of Congress Control Number: 2008923733

© 2008 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable to prosecution under the German Copyright Law.

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

Cover design: Deblik, Berlin, Germany

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

For make your community creative

Scientific “Agendum” of Data Science

Shuichi Iwata

The University of Tokyo, Kashiwa, Japan
iwata@k.u-tokyo.ac.jp

Data on nuclear cross section have integrated metals, ceramics, waters and other materials to form various nuclear reactors, which is a basic for nuclear engineering. Data on intrinsic properties of substances are used to link extrinsic properties of substances and structure-sensitive engineering properties of defects as fundamental constants have been defined in a network of elaborate experiments and models in science. These understandings have driven me to data activities since 1970 when I at first became to know CODATA by name, and it is a good time now for me review the committed works in these 37 years and write down my personal gendumh reflecting the re-view so as to have more collaborators for the next milestone.

As scientific agendum of CODATA, we have added an aspect gdata and societyh explicitly through such opportunities of WSIS(World Summit on the Information Society)s and lessons on tragedies, namely, suffering from natural disasters and epidemics. Not only by global warming problems pointed out clearly by IPCC (Intergovernmental Panel on Climate Change) but also by other global issues as known well partly by UN MDGs(the eight Millennium Development Goals), we have become to know the necessity of linking scientific, technical, economic, social and political agenda with proper missions and guidelines for the society. Here human-centered reorganizations of domain-differentiated sciences from natural sciences to social sciences, are requested to be carried out, where and when we need common data for the proper and holistic reorganization to reach right decisions and consensus of society. If we do not share common data on global warming effects, we cannot write down our remedies against inconvenient truths and establish flexible and steady roadmaps for the sustainable society. It is necessary for us experts to write down proper remedies together by linking such associated scientific domains as politics, laws, ethics, economics, environmental sciences, ecologies, civil engineering, manufacturing, waste management and so on and complimenting missing links there for better solutions. And as a consequence of such efforts we may come to create a new scientific domain gdata-driven sustainable scienceh to design and manage the society properly. It is really requiring us continuous big efforts with challenging spirits to start everything from facts and data, which is our raison dfetre and concerns our identity as data scientists.

However, the above agendum essentially has already been associated to our core activities. Data and knowledge corresponding missing links of domain specific sciences have been daily works of CODATA. For example,

- fundamental constants have been compiled reflecting advances of precise measurements and basic sciences, and also higher coherences of scientific models, which has resulted in creating new sciences like nano-sciences, spintronics and other specialized scientific domains, and also several key standards for the information society, namely, radio wave standards, current, voltage and so on,
- spectra data and diffraction data have guided us to get microstructural information on substances and materials, and also on life. Together with models and interatomic potentials derived from first principles calculation, new scientific and technical fields have been exploited rapidly, and application areas are spreading widely over drug design, defect theory, fracture dynamics, materials design, process design, earth sciences, and even in bioscience to see the origin of life and medical diagnostics.
- many exemplars in biosciences to get insight through using common data: arabidopsis thaliana data for evolving botany; RNA data to link DNA and protein, and consequently associated with disease and health; data on Tradescantia ohiensis and/or nude mice for irradiation effects, and by taking advantage of recent advances in genomics, proteomics biophysics and biochemistry, an breakthrough is going on to overcome limitations of epidemiological survey based on statistics.

Data-oriented statistical approaches are combined with scientific models and practical monitoring, and traditional established safety/risk/reliability standards are changing into proactive and dynamic adaptive standards. Safety/risk issues in medical services, nuclear reactors, aircrafts, company managements, energy resource security and so on can be dealt with in a similar way. Openness and transparency of many disciplines and scientific domains promoted by e-science projects and so called global information commons are prerequisites for a revolution of sciences by the 7 billion people. Devices for the revolution might be something creative of emerging wisdoms and welling up emotions in the internet, which may be more than such knowledge management approaches as ontology, metadata, object-oriented approach, semantic-web, common sense reasoning and so on.

Through evaluation of fundamental constants we are integrating quantum worlds, atomistic worlds and macroscopic worlds quantitatively, where scientific disciplines and domains are networked with a certain consistency. Through preparation of geometric data of parts with properties we can design an artifacts and assemble the available parts into the artifacts of integrity and cost-effective performances, where domain differentiated engineering disciplines are integrated to establish manufacturing industries. Design and maintenance of landscapes, cities, countries, regional environments and global climates are used to be carried out in a similar way, namely, sharing data by stakeholders and coordinating different views and opinions by the shared common data. Quality of data matters every time and everywhere. Gold in, gold out. Reasonable estimations of uncertainties on data may produce better results

and outcomes. The more the problem to be solved is uncertain, the more we should become flexible. Evidence-based deterministic approaches do not work effectively, and adaptive and heuristic approaches work better coupled with in situ data capture, evaluation, and quick decision and timely actions. Holistic creativity as a group is a key for a success of the group, where practical maintenance of data quality for proper decision is important. Time constants of data life cycle are becoming shorter, and diversities of stakeholders and complexities of data are increasing. New disciplines are to be continuously created by taking advantage of available data and devices so as to prepare solutions on time. Without proper managements of continuously produced big data, and without productivity of new disciplines based on data, we cannot solve important problems of the world. Data science may play an important role there.

Shuichi Iwata

Contents

Scientific “Agendum” of Data Science	
<i>Shuichi Iwata</i>	VII

Part I Thought, Communication, and Actions

1 Sharing Representations and Creating Chances through Cognitive Niche Construction. The Role of Affordances and Abduction	
<i>Lorenzo Magnani and Emanuele Bardone</i>	3
2 Discovering and Communicating through Multimodal Abduction	
<i>Lorenzo Magnani</i>	41

Part II Creative Community Working on Multidisciplinary Data

3 Augmented Analytical Exploitation of a Scientific Forum	
<i>Xijin Tang, Yijun Liu, and Wen Zhang</i>	65
4 Multi-Data Mining for Understanding Leadership Behavior	
<i>Naohiro Matsumura and Yoshihiro Sasaki</i>	81
5 Discussion Visualization on a Bulletin Board System	
<i>Wataru Sunayama</i>	95
6 Design of BBS with Visual Representation for Online Data Analysis	
<i>Yasufumi Takama and Yuta Seo</i>	111
7 A Study on Web Clustering with Respect to XiangShan Science Conference	
<i>Wen Zhang and Xijin Tang</i>	127

Part III Discoveries from Data and Application to Business

8 A Multilevel Integration Approach for E-Finance Portal Development
Jia Hu and Ning Zhong 139

9 Integrated Design Framework for Embedded GUI System
Noriyuki Kushiro and Yukio Ohsawa 157

10 A Unified Probabilistic Inference Model for Targeted Marketing
Jiajin Huang and Ning Zhong 171

**11 Computational Methods for Discoveries from Integrated Data -
Human-Interactive Annealing for Multilateral Observation**
Yoshiharu Maeno, Kenichi Horie, and Yukio Ohsawa 187

**12 Human-Interactive Annealing Process with Pictogram for Extracting
New Scenarios for Patent Technology**
Kenichi Horie, Yoshiharu Maeno, and Yukio Ohsawa 205

**13 Pharmaceutical Drug Design Using Dynamic Connectionist Ensemble
Networks**
Ajith Abraham, Crina Grosan, and Stefan Tigan 221

**14 A Framework of Knowledge Management Platform for Middle
and Small Business**
Xingsen Li, Jun Li, Yuejin Zhang, and Yong Shi 233

Part IV Mining Risks from Multidisciplinary Data

**15 Discovery of Clusters from Proximity Data: An Approach
Using Iterative Adjustment of Binary Classifications**
Shoji Hirano and Shusaku Tsumoto 251

**16 Evaluating Learning Algorithms to Support Human Rule Evaluation
with Predicting Interestingness Based on Objective Rule Evaluation
Indices**
Hidenao Abe, Shusaku Tsumoto, Miho Ohsaki, and Takahira Yamaguchi 269

17 Risk Mining for Infection Control
Shusaku Tsumoto, Kimiko Matsuoka, and Shigeki Yokoyama 283

**18 Evaluating the Error Risk of Email Filters Based on ROC Curve
Analysis**
Wenbin Li, Ning Zhong, and Chunnian Liu 299

19 Categorized and Integrated Data Mining of Medical Data
Akinori Abe, Norihiro Hagita, Michiko Furutani, Yoshiyuki Furutani,
*and Rumiko Matsuoka*315

20 Privacy-Preserving Data Mining for Medical Data: Application
of Data Partition Methods
*Yi Peng, Gang Kou, Yong Shi, and Zhengxin Chen*331