Journal of Universal Computer Science, vol. 22, no. 6 (2016), 735-741
submitted: 12/5/16, accepted: 27/5/16, appeared: 1/6/16 © *J.UCS*

Computational Intelligence Tools for Processing Collective Data

J.UCS Special Issue

Ngoc Thanh Nguyen
(Wroclaw University of Technology, Poland
Ngoc-Thanh.Nguyen@pwr.edu.pl)

Ireneusz Czarnowski
(Gdynia Maritime University, Gdynia, Poland
irek@am.gdynia.pl)

Dosam Hwang
(Yeungnam University, Gyeonsan, Korea
dshwang@yu.ac.kr)

1 Introduction

This special issue on Computational Intelligence Tools for Processing Collective Data deals with the important role of the computational intelligence paradigms in creation, maintenance and use of information systems. Collective data mean data originating from different and autonomous sources including text, numeric, images, video and other multimedia data. In most cases, such data are complex and their effective processing requires implementation of the advanced computational intelligence tools and techniques. Implementation of the above tools leads to creation of the intelligent information systems.

The issue focuses on various methods and techniques of computational intelligence. Application of these methods and techniques is today considered as a key success factor in a majority of the real-life information systems. These tools and techniques are used to enhance, support or replace traditional approaches to decision making in intelligent data mining, complex data analysis and managing data processing within information systems. Several real-life applications of these tools and techniques are also presented in the issue.

This special issue of the Journal of Universal Computer Science contains several carefully selected and extended versions of the invited and regular papers presented at the 2nd IEEE International Conference on Cybernetics - CYBCONF 2015, held in Gdynia (Poland) on 24-26 June 2015.

2 Contributions Included in the Special Issue

This special issue is a collection of research works written by specialists focusing on promising approaches in computational intelligence and consists of eight papers covering examples of application of novel methods and techniques.

2.1 Automatic Generation of Interactive Cooking Video with Semantic Annotation

Since video has become a popular format type for the interactive contents, we can observe growing interest in research on semantic annotation methods, which aim at searching and representing objects included in video data [Andrews 12]. The research on methods for detection and recognition of objects, events and actions in videos is receiving an increasing attention from the scientific community. This research has relevance for many applications, from semantic video indexing to intelligent video surveillance systems and advanced human-computer interaction interfaces [Ballan, 11].

The paper of Kyeong-Jin Oh, Myung-Duk Hong, Ui-Nyoung Yoon and Geun-Sik Jo proposes an interactive video system to generate automatically interactive videos. The domain of interest of the authors is to propose a system that makes it possible for people to interact with video to find a specific part or to obtain relevant information from it. However, such solutions need to be implemented based on dedicated and intelligent user interfaces and videos must be transformed to interactive videos. Semantic video annotation process is proposed to assure the interaction. The discussed process consists of adequate synchronization between elements and objects of the video and the users, alignment of the video parts with respect to the expectation of users, information extraction according to the expectation of users and semantic interconnection between elements and objects of video and users.

The authors present their approach using the example of an interactive video related with cooking. Cooking videos is a kind of how-to videos and many people use knowledge embedded in cooking videos to cook their foods. There is an example, where people want to find a specific part of a cooking video and they want to obtain more information on objects shown in a cooking video, such as ingredients, cooking tools and methods for ingredient preparation. The considered interactive cooking video system is an example of the intelligent information system, which is based on mining multimedia data, and an example of the system dedicated to working with data generated currently by different multimedia resources.

2.2 A Proposal for Recommendation of Feature Selection Algorithm based on Data Set Characteristics

Feature selection is an explicit part of most knowledge mining approaches. The aim of the feature selection process is to select most relevant features in data set, thus, forming such feature set that can be especially informative and support the knowledge discovery process. The feature selection process can bring benefits such as improving performance of mining tools, less processing, lower complexity, smaller structure of the system or improve the comprehensibility of the resulting knowledge models

[Guyon, 06]. Although several examples of feature selection approaches have been recently proposed, it is still an active research area in pattern recognition, statistics, and data mining communities [Shah, 12], [Singh, 15].

The paper written by Saptarsi Goswami, Amlan Chakrabarti and Basabi Chakraborty deals with the feature selection problem. The aim of the research work presented in the paper was to find a suitable representation of the dataset features and to recommend an appropriate feature selection algorithm to be used for particular datasets. The conviction of the authors is that the choice of suitable algorithm should result from the characteristics of the dataset, which is intended to be processed. The choice of the feature selection algorithm without the knowledge about the dataset can result in poor performance of the next steps of mining process. It means that the characteristics of the dataset influences the behaviour of the pattern classification algorithm.

The authors propose to characterise datasets using several parameters, including the correlation structure. It is claimed that this structure is a main parameter characterizing datasets. Finally a framework of recommendation regarding the choice of the most appropriate feature selection algorithm has been proposed.

2.3 TwiSNER: Semi-supervised Method for Named Entity Recognition from Text Streams on Twitter

The data from Social Network Services are recently a source of information, perhaps in some case very important. These data has become an interesting source of information enabling to carry-out many different analyses, such as sentiments, behaviours, relations, opinions of network community and many others. The data are processed and the information discovery is carried out using different and dedicated tools. However, the process is neither easy nor trivial, and standard tools can have difficulty obtaining satisfactory results [Ting, 11], [Can, 14].

In their paper, Van Cuong Tran, Dosam Hwang and Jason J. Jung consider an approach for solving the entity identification problem, also known as Named Entity Recognition (NER) task. NER is the main task of natural language processing systems and is a subset of information extraction problem. The proposed method is called TwiSNER and it proposes how to classify of named entities in the Twitter data. A feature of the method is that it is based on a semi-supervised learning approach combined with the conditional random field model, hand-made rules, and the co-occurrence coefficient of the featured words surrounding. The approach is discussed in details and the experiment results obtained using the approach are shown and compared with results obtained using other systems. The proposed method can be considered as an alternative approach to others, especially for cases with small amount of labelled data, as it has been show based on the experiment results.

2.4 Using Soft Set Theory for Mining Maximal Association Rules in Text Data

The paper of Bay Vo, Tam Tran, Tzung-Pei Hong, Nguyen Le Minh is dedicated to the problem of association rules discovery and presents an approach to apply soft set theory for maximal association rule mining from transaction databases. The approach is an effective strategy for maximal rules mining based on soft set theory and consists

of the steps, where, at first the item tree is constructed, next, from the tree consisting in each node the maximal itemsets maximal association rules are generated. The experiment results show that the mining time of the proposed approach is faster than other methods.

2.5 A Quick Method for Querying Top-k Rules from Class Association Rule Set

Loan T.T. Nguyen, Ngoc-Thanh Nguyen and Bogdan Trawiński focus on the problem of finding class association rules. Typical methods of mining frequent items or association rules cannot be used for mining class association rules because of differences of both problems. In case of the association rule on the right-hand side is any frequent itemset, whereas in case of the class association rule, on the right-hand side there are class labels. Of course, mining class association rules is one of variations of mining rules including mining association rules [Van, 14]. Finding class association rules has recently become a focus of the research interest since it has numerous applications in many fields. The interest of the systems based on class association rules stems from the fact, that they provide an explicit knowledge on the problem in the form of rules, and these rules are determined directly from the data set. An example are systems for mining of medical datasets [Nguyen, 08].

In the paper the authors propose a method for mining top-k class association rules. From the set of mined class association rules that satisfy the minimum support and the minimum confidence thresholds, the top k-class association rules are selected based on a QuickSort-based method. The proposed approach is presented and the experiment results are discussed. In the conclusion the authors confirm that the proposed approach can be used to significantly enhance performance of discovery of class association rules.

2.6 Parameterized and Dynamic Generation of an Infinite Virtual Terrain with Various Biomes using Extended Voronoi Diagram

Kazimierz Choroś and Jacek Topolski deal with the problem of space virtualization and evaluate the method for generating an infinite terrain in 3D space.

The modelling and simulation in 3D virtual space has already become a very significant task in visualisations of different processes including industrial or medicine ones. The modelling and simulation can provide more information on the behaviour and performance of existing systems, as well as prediction of new situations and different conditions within the considered systems. As it is mentioned by the authors, which has been also pointed-out by [Kopácsi, 13], virtual systems provide faster planning, easier system integration, and more reliable operations and control of many real and dynamic processes.

The proposed approach is based on generation of various biomes in a virtual 3D space. The authors define the biomes, as shapes generated from different set of textures, to form shape in the landscape. The biomes are generated using Gaussian blur and Voronoi diagram algorithms. The dedicated tests have been performed by setting up a sample terrain and performing basic actions on this terrain, like moving or rotating, to gather frame times. The results showed that although the method demands much memory, it is efficient and suitable for the real-time processing.

2.7 PLA Based Strategy for Solving RCPSP by a Team of Agents

The paper written by Piotr Jędrzejowicz and Ewa Ratajczak-Ropel focuses on the application of the agent-based population learning algorithm, based on the A-Team architecture, for solving the Resource-Constrained Project Scheduling Problem (RCPSP), belonging to the class of NP-hard optimization problems. The authors give the general idea of the agent-based population learning algorithm and propose the so-called dynamic interaction strategy. The strategy supervises interactions between optimization agents and the common memory within an A-Team architecture. To validate the proposed approach computational experiment has been carried out. The computational experiment results shows that the proposed dedicated A-Team architecture and the interaction strategy is an effective and competitive tool for solving instances of the RCPSP.

2.8 Improving Performance of the Differential Evolution Algorithm Using Cyclic Decloning and Changeable Population Size

In the last paper, Piotr Jędrzejowicz and Aleksander Skakovski study a special case of the evolutionary algorithm, which is differential evolution algorithm. Differential evolution is an example of the stochastic direct search and global optimization algorithm, at first proposed in [Storn, 97].

The contribution of the paper is twofold. First, it proposes a decloning procedure, which aim is to cyclically replace genetically identical individuals (clones) with randomly generated ones. Second, it shows the extent to which performance of the considered differential evolution algorithm depends on such parameters as the population diversification rate, the size of the population, and the number of fitness function evaluations carried out by the algorithm to yield a solution to the problem.

To validate the proposed method and to evaluate the effectiveness of the decloning procedure the computational experiment has been carried out. The computational experiment concerned the application of the differential evolution algorithm for solving discrete-continuous scheduling problem with continuous resource discretisation.

3 Program Committee

The submitted papers have been reviewed by at least three referees. We wish to thank all peer reviewers whose invaluable work, suggestions and detailed feedback have helped to improve the quality of the papers included in the special issue. Special thanks are due to:

- Aleksander Byrski, AGH University of Science and Technology, Poland
- David Camacho, Universidad Autonoma de Madrid, Spain
- Kazimierz Choroś, Wroclaw University of Technology, Poland
- Tzung-Pei Hong , National University of Kaohsiung, Taiwan
- Ahmad Jalal, KyungHee University, Suwon, South Korea
- Piotr Jędrzejowicz, Gdynia Maritime University, Poland

- Elzbieta Kukla, Wroclaw University of Technology, Poland
- Kazuhiro Kuwabara, Ritsumeikan University, Japan
- Mark Last, Ben-Gurion University of the Negev, Israel
- Rey-Long Liu, Tzu Chi University, Taiwan
- Edwin Lughofer, Johannes Kepler University Linz, Austria
- Bernadetta Maleszka, Wroclaw University of Technology, Poland
- Marcin Maleszka, Wroclaw University of Technology, Poland
- Antonio D. Masegosa, University of Granada, Spain
- Tamas Matuszka, Eotvos Lorand University, Hungary
- Joao Mendes-Moreira, University of Porto, Portugal
- Gerardo Mendez, Instituto Tecnologico de Nuevo Leon, Mexico
- Jacek Mercik, Wroclaw School of Banking, Poland
- Radoslaw Michalski, Wroclaw University of Technology, Poland
- Addi Ait-Mlouk, Cadi Ayyad University, Marrakech, Morocco
- Ghulam Mustafa, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
- Grzegorz J. Nalepa, AGH University of Science and Technology, Poland
- Le Minh Nguyen, Ton Duc Thang University, Ho Chi Minh, Vietnam
- Loan T.T. Nguyen, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam
- Ewa Ratajczak-Ropel, Gdynia Maritime University, Poland
- Hoai An Le Thi, Universite de Lorraine, France
- Marin Vukovic, University of Zagreb, Zagreb, Croatia
- Zbigniew Wesołowski, Military University of Technology, Warsaw, Poland
- Qiangfu Zhao, The Univerity of Aizu, Japan

4 Conclusions

The editors believe that the issue has been an important and timely initiative. It is hoped that the presented ideas and results will be of value to the research community working in field of artificial intelligence, analysis of complex and multimedia data, data mining, text mining, web mining, pattern recognition, knowledge discovery and project management.

We would like to take this opportunity to thank all authors for their valuable contributions. We also thank all whose invaluable work and suggestions have helped to improve quality of papers.

References

[Andrews 12] Andrews, P., Zaihrayeu, I., Pane. J.: A Classification of Semantic Annotation Systems. Semantic Web 3(3), 223-248, 2012.

[Ballan, 11] Ballan, L., Bertini, M., Bimbo, A.D., Seidenari, L., Serra, G.: Event detection and recognition for semantic annotation of video. Multimedia Tools and Applications 51(1), 279–302, 2011.

[Can, 14] Can, F., Ozyer, T., Polat, F. (Eds.): State of the Art Applications of Social Networks Analysis. Springer International Publishing Switzerland, 2014.

[Guyon, 06] Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L.A.: Feature Extraction. Foundations and Applications. Studies in Fuzziness and Soft Computing. Springer-Verlag Berlin Heidelberg. 2006.

[Kopácsi, 13] Kopácsi, S., Kovács, G. L., Nacsa, J.: Some aspects of dynamic 3D representation and control of industrial processes via the Internet. Computers in Industry 64(9), 1282–1289, 2013.

[Nguyen, 08] Nguyen, N.T.: Advanced Methods for Inconsistent Knowledge Management. Springer London, 2008.

[Shah, 12] Shah, M., Marchand, M., Corbeil, J.: Feature Selection with Conjunctions of Decision Stumps and Learning from Microarray Data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34, (1), 174-186, 2012.

[Singh, 15] Singh, R.K., Sivabalakrishnan, M.: Feature Selection of Gene Expression Data for Cancer Classification: A Review. Procedia Computer Science 50, 52-57, 2015.

[Storn, 97] Storn, R., Price, K.: Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization, 11(4), 341-359, 1997.

[Ting, 11] Ting, I-H., Hong, T-P., Wang. L. S-L. (Eds.): Social Network Mining, Analysis, and Research Trends: Techniques and Applications. IGI Global. 2011.

[Van, 14] Van, T.T, Vo, B., Le, B.: IMSR_PreTree: an improved algorithm for mining sequential rules based on the prefix-tree. Vietnam Journal of Computer Science 1(2), 97-105, 2014.

Ngoc Thanh Nguyen
Ireneusz Czarnowski
Dosam Hwang
Poland & Korea
June 2016