Progress in Intelligent Systems

The power of different intelligent techniques, such as K-means, evolutionary algorithms, artificial immune systems, fuzzy clustering algorithms, kernel regularized methods and others, have been demonstrated over the years by their successful use in many types of problems with different degrees of complexity and in different fields of application. However, the use of a machine learning solution involves choosing a particular model and properly tuning a set of parameters that, together, can effectively and efficiently solve the specific problem. Some difficulties arise in this process, such as the existence of several machine learning algorithms, the exponential number of combinations of parameter values, and also the need for a priori knowledge on the problem domain. Moreover, certain complex learning problems cannot be solved by a single intelligent technique and each intelligent technique has particular computation properties that can be complementary.

This special issue has a collection of papers on intelligent systems design. The papers have a combination of many different techniques including fuzzy systems, evolutionary computation, artificial immune systems and clustering algorithms.

Data clustering is a fundamental conceptual problem in data mining, which faces the challenger to extract relevant information from distributed data with good computational performance and scalability. For dealing with data distributed in separated repositories, Murilo Naldi and Ricardo Campello proposed in a previous work an algorithm based on Evolutionary k-means for distributed data. Two different distribution approaches of the algorithms were presented: the first obtains a model identical to the centralized version of the clustering algorithm (DF-EAC); the second generates and selects clusters for each distributed data subset and combines them afterwards (CDC). In the paper presented in this special issue, a novel comparison among DF-EAC and CDC algorithms using different validity indices was conducted based on two perspectives: the theoretical one, through asymptotic complexity analyses, and the experimental one, through a comparative evaluation of results obtained from a collection of experiments and statistical tests. Additionally, DF-EAC was revisited and two modifications were investigated: The first preserves data privacy among repositories and the second uses an alternative relative validation index to evaluate the resulting clusters. The modified DF-EAC variants are also compared in this study. The obtained results allows to conclude that the first approach is robust to different types of data distribution, and the use of privacy-preserving methods proposed in this study did not jeopardize the quality, speed, or amount of data transmitted by the algorithm. However, the algorithm demands considerable transmission speed if compared to the CDC, being recommended for systems with high transmission rates like local networks and high-speed metropolitan networks. The second approach reduces the data amount and the number of transmissions of the algorithm, causing a slight quality loss, and is recommended for systems with low transmission rates. Finally, the transmission costs of the algorithms do not depend on the number of objects in the data set, no matter what scenario is chosen. Thus, these algorithms are scalable regarding the number of data.

The second paper, “An immune-inspired algorithm for a scheduling problem with

unrelated parallel machines and sequence and machine dependent-setup times for makespan minimization", deals with a problem that frequently arise in production environments, the job scheduling. The paper proposes an immune algorithm to solve the unrelated parallel machines scheduling problem with sequence and machine dependent setup-times for makespan minimization. The algorithm is based on clonal selection process and its major features are: the initial population is generated through the construction phase of the Greedy Randomized Adaptive Search Procedure (GRASP); an evaluation function that helps the algorithm to escape from local optima; a Variable Neighborhood Descent (VND) local search heuristic; a somatic hyper mutation operator to accelerate the convergence of the algorithm and a re-selector operator, which strategically keeps good quality solutions with a high level of dispersion in the search space. A full comparison between the implemented algorithms including statistical analysis allows concluding that the proposed algorithm enables better results than those reported in recent literature studies. Moreover, the experiments showed the importance of each operator to the overall performance of the proposed algorithm. Furthermore, when all operators were incorporated into the genetic algorithm, its performance increased, but not enough to overcome the results obtained by the clonal algorithm.

Kernel methods are among the most flexible and powerful methods to tackle the problem of non-linear regression estimation. In paper 3, Improving the Kernel Regularized Least Squares Method for Small-Sample Regression, the authors Igor Braga and Maria Carolina Monard, decided to approach this problem by means of Kernel Regularizes Least Squares (KRLS), motivated by the good statistical and computational properties of the method. KRLS performance depends on proper selection of both a kernel function and a regularization parameter. The selection of these elements is very frequently conducted in a cross-validation setting. The most interesting property of KRLS is that it is relatively inexpensive when combined with cross-validation, compared to other learning methods. The authors of paper 3 state that, Gaussian RBF kernel has become a very common kernel choice in much of machine learning research because of their universal approximation properties. However, when combined with cross-validation and small training sets, RBF kernels have a great potential for overfitting. Recently, there has been a renewed interest in developing kernels with less potential for overfitting while retaining a good approximation property. Having that in mind, in this paper they investigate the use of splines as a safer choice to compose a multidimensional kernel function and proposed the use of additive spline kernels instead of multiplicative ones, employed in their previous work. They claim to have found experimental evidence that the additive version is more appropriate to regression estimation in small sample situations. In a second line of investigation, they consider alternative parameter selection methods that have been shown to perform well for other regression methods. Their study demonstrated that the parameter selection procedure Finite Prediction Error (FPE) is a competitive alternative to cross-validation when using the additive splines kernel.

Cluster analysis is a very active research topic in the machine learning area. Roughly speaking, Clustering methods aim to organize a set of items into clusters such that items within each cluster have a high degree of similarity, whereas items belonging to different clusters have a high degree of dissimilarity. These methods have been widely applied in fields such as taxonomy, image processing, information retrieval, and data mining. There

is a huge amount of research work on this topic addressing several different aspects of the whole clustering process. In Paper 4, A Multi-View Relational Fuzzy c-Medoid Vectors Clustering Algorithm, the authors Francisco de Carvalho, Filipe de Melo and Yves Lechevallier explored the aspects of representation and multi-view data. Two usual representations of the objects upon which clustering can be based are feature data and relational data. When each object is described by a vector of quantitative or qualitative values, the set of vectors describing the objects is called a feature data. When each pair of objects is represented by a relationship, then we have relational data. In multi-view data, the objects are represented by several (feature or relational) data matrices. Multi-view data can be found in many domains such as bioinformatics, and marketing. The authors of Paper 4 emphasize that, while several clustering models and algorithms have been proposed aiming to cluster feature data, few clustering models have been proposed for relational data. They have also observed that several applications, such as content-based image retrieval, would be strongly benefited by clustering methods for relational data. In their paper, the authors have proposed a multi-view relational fuzzy c-medoid vectors clustering algorithm that is able to give a fuzzy partition of the objects taking into account simultaneously their relational descriptions given by multiple dissimilarity matrices. The method proposed is a modified and extended version of the algorithms they proposed before, namely a single-view fuzzy c-medoid algorithm and the fuzzy c-medoids clustering algorithms based on multiple dissimilarity matrices. The main idea is to obtain a collaborative role of the different dissimilarity matrices aiming to obtain a final consensus partition. These matrices could have been obtained using different sets of variables. This algorithm is designed to give a fuzzy partition and a vector of medoids for each fuzzy cluster as well as to learn a relevance weight for each dissimilarity matrix by optimizing an adequacy criterion that measures the fitting between the fuzzy clusters and their representatives. These relevance weights change at each iteration of the algorithm and are different from one cluster to another. Moreover, various tools for interpreting the fuzzy partition and fuzzy clusters provided by this algorithm are also presented. Several examples illustrate the performance and usefulness of the proposed algorithm.

This special issue includes four papers selected among the best contributions of the 2nd Brazilian Conference on Intelligent Systems (BRACIS 2013), which took place in Fortaleza, Brazil, from October 19 to October 24 2013. BRACIS is the most important scientific event in Brazil in Artificial Intelligence (AI) and Computational Intelligence (CI), which originated from the combination of the Brazilian Symposium on Artificial Intelligence – SBIA and the Brazilian Symposium on Neural Networks – SBRN. BRACIS is sponsored by the Brazilian Computer Society (SBC) and promotes research of international level and scientific exchange among researchers, practitioners, scientists, and engineers in related disciplines, congregated in Brazil. The conference is focused on original work addressing theories, methods and novel applications dealing mainly with the use and analysis of AI and CI techniques. BRACIS is an international conference with an international program committee, which includes well- established researchers from Brazil and overseas. The papers have been written in English and are published by the Conference Publishing Services (CPS), a division of the IEEE Computer Society press.

The BRACIS 2013 received 100 submissions from several different countries. Among these submissions, 43 full papers were accepted for oral presentation. All papers were reviewed by, at least, two independent specialized referees. The authors of the papers with the best reviews were invited to submit an extended and updated version for this special issue. The selection process emphasized three main aspects: originality, relevance and technical contribution. The papers selected from BRACIS 2013 are listed in the references [1-4]. The new versions were submitted to a rigorous peer review process conducted by international reviewers. Only the papers recommended by the reviewers were accepted for this special issue. We believe that this issue presents a set of very high quality papers. As a result, this edition will provide the readers a rich material of current research on Intelligent Systems and related issues.

We would like to thank all the authors for their effort to submit high quality papers and the referees for their meticulous and useful reviews with relevant comments and suggestions that surely improved the quality of this special issue. We would also like to thank the Neurocomputing Editor-in-Chief Tom Heskes, the Journal Editorial Board and Elsevier for the opportunity and for efficiently handling the publication procedure. Finally, we would further like to acknowledge Jacqueline Zhu, Vera Kamphuis and Chandini Emmanuel for their help and careful edition of this issue.

Aurora Pozo
Universidade Federal do Paraná, Brazil
E-mail address: aurora@inf.ufpr.br
Heloisa de Arruda Camargo
Universidade Federal de São Carlos, Brazil
E-mail address: heloisa@dc.ufscar.br
Teresa B. Ludermir
Universidade Federal de Pernambuco, Brazil
E-mail address: tbl@cin.ufpe.br

References

[1] Murilo Coelho Naldi and Ricardo José Gabrielli Barreto Campello, Distributed K-Means Clustering with Low Transmission Cost. In 2013 Brazilian Conference on Intelligent Systems (BRACIS 2013), pp. 70-75. IEEE, ISBN 978-0-7695-5092-3, DOI: 10.1109/BRACIS.2013.20.

[2] Rodney O.M. Diana, Moacir F. de F. Filho, Sergio R. de Souza, and Maria Amélia L. Silva, A Clonal Selection Algorithm for Makespan Minimization on Unrelated Parallel Machines with Sequence Dependent Setup Times. In 2013 Brazilian Conference on Intelligent Systems (BRACIS 2013), pp. 57-63. IEEE, ISBN 978-0-7695-5092-3, DOI: 10.1109/BRACIS.2013.18.

[3] Igor Braga and Maria Carolina Monard, Statistical and Heuristic Model Selection in Regularized Least-Squares. In 2013 Brazilian Conference on Intelligent Systems

(BRACIS 2013), pp. 231-236. IEEE, ISBN 978-0-7695-5092-3, DOI: 10.1109/BRACIS.2013.46.

[4] Francisco de A.T. de Carvalho, Filipe M. de Melo, and Yves Lechevallier, A Fuzzy C-Medoids Clustering Algorithm Based on Multiple Dissimilarity Matrices. In 2013 Brazilian Conference on Intelligent Systems (BRACIS 2013), pp. 107-112. IEEE, ISBN 978-0-7695-5092-3, DOI: 10.1109/BRACIS.2013.26.