PREFACE

Hybrid and ensemble methods in machine learning have gained considerable attention of the scientific community over the last few years. Ensemble algorithms and hybrid methods of reasoning have found their application in various real word problems ranging from person recognition to medical diagnosis and text classification to financial forecasting.

The design of hybrid (intelligent) systems has become the core of soft computing and computational intelligence. In fact, the main reason for the popularity of these research areas during the last two decades is the synergy derived from their components. The integration of their constituent technologies into hybrid systems provides us with complementary reasoning and search methods able to combine domain knowledge and empirical data to solve complex problems. In particular, they have allowed us to develop a large variety of machine learning tasks with improved performance.

Meanwhile, in the last two decades classifier/model ensembles have been presented as very promising tools for improving the performance of single weak learners when dealing with complex, high dimensional classification and regression problems. The design of an ensemble is mainly based on two stages: the derivation of the component weak learners and the combination mechanism for the individual decisions provided by them into the global system output. Since an ensemble results from combining the outputs of a group of individually trained weak learners, the accuracy of the finally derived system relies on the performance and proper integration of these two tasks. The best possible situation for an ensemble is where the component classifiers/models are both accurate and diverse, in the sense that they explore different parts of the problem space.

On the one hand, the correct definition of a set of base classifiers/models is fundamental to the overall performance of ensemble methods. Different approaches have thus been proposed to succeed in generating diverse weak learners with uncorrelated errors such as data resampling techniques (mainly bagging and boosting), specific diversity induction mechanisms (feature selection, diversity measures, use of different learning models resulting in a hybrid ensemble, etc.), or combinations of the latter two families (data resampling and specific diversity induction mechanisms).

The research area of combination methods is also very active. It does not only consider the issue of aggregating the results provided by the entire initial set of component classifiers/models derived from the first learning stage to compute the final output, which is usually called classifier or model fusion. It could also involve the task of globally selecting the subgroup of weak learners which will be considered for every input pattern. Besides, hybrid strategies involving the two groups have also been introduced.

This special section encompasses seven papers devoted to both hybrid and ensemble methods as well as their application to classification and forecasting problems. The section originated from the special session on *Multiple Model Approach to Machine Learning (MMAML 2011)* organized by the guest editors at the *3rd Asian Conference on Intelligent Information and Database Systems (ACIIDS 2011)*, which was held in Daegu, Korea, in April 2011. In total, fifteen papers were nominated by the reviewers and finally selected for presentation at the special session. Afterwards, the authors of twelve papers were invited to submit substantially extended versions of their contributions to the current special section and eleven submissions were finally received. After a thorough review process, only seven of the submitted contributions were finally accepted by the editors to become a part of the section. Unfortunately, the remaining four contributions were rejected due to the high quality standards we wanted to impose for the special section.

The seven accepted contributions can be classified into three different groups within the wide area of the design of hybrid and ensemble methods for machine learning. Three of them are devoted to the design of hybrid systems and their application to real-world problems. Another three focus on the design of ensemble methods. Finally, the remaining manuscript is devoted to a statistical analysis of machine learning algorithms.

These seven contributions are briefly reviewed as follows. As said, the first three are related to hybrid systems. In the first paper, Li and Chiang focus on a real-world application area, that is, financial time series forecasting. They introduce a neuro-fuzzy system based on the use of complex fuzzy set theory. Complex fuzzy sets make use of complex-valued membership functions characterized within the unit disc of the complex plane. To deal with the estimation of the large number of parameters involved in the system, a hybrid learning method is proposed combining a multi-swarm particle swarm optimization method to learn the rule antecedents and a recursive least square estimation method to derive the parameters of the Takagi–Sugeno fuzzy rule consequents. The new proposal is tested on three different financial time series providing good performance.

Colomo-Palacios *et al.* are concerned with another real-world problem, the management of software development processes in software engineering. In their contribution they introduce a new hybrid approach to assist project managers to find the best resources configuration for each of the work packages in order to keep the team motivated and committed to the project and to improve productivity and quality. The proposed recommender system deals with Scrum team roles, with Scrum being one of the recently proposed agile methodologies to tackle software project work package definition. ReSySTER gives recommendations for team formation in Scrum driven projects based on the hybridization of fuzzy logic, rough sets, and semantic technologies. The system has been evaluated in a real scenario of development with the Scrum framework obtaining promising results.

Qin *et al.* deal with the data preprocessing machine learning task in their manuscript. To do so, they consider a methodology to handle uncertain data, soft set theory, and focus on the case when incomplete data arise. The authors propose a hybrid design method outperforming the existing ones based on evidence theory and Bayesian models. The new proposal considers data filling using association degrees between parameters when tackling classification problems. Results obtained on five UCI classification datasets show the accuracy improvement of the new approach with respect to the existing ones.

The next three contributions are focused on the design of ensemble methods. They cover the different existing design stages and also include some real-world applications. The paper by Kajdanowicz and Kazienko explores the application of classifier ensembles to a hot topic in pattern recognition, multi-label classification. To do so, the authors assume the base multi-label classifiers are a noisy channel and apply Error Correcting Output Codes (ECOCs) methods in order to recover the classification errors made by individual classifiers. The proposal also focuses on the induction of diversity in the classifier ensemble design by making use of three different kinds of multi-label classifiers and four different ECOC methods. The experimental results show how some combinations of ECOCs and multi-label classification methods may be more accurate than a single multi-label classifier while other combinations are not able to improve the single classifier performance.

In the fifth paper, Sumi *et al.* introduce a multi-model system for forecasting rainfall in the Fukuoka city in Japan. The proposed ensemble design method considers the use of feature selection, principal component analysis, and four different learning methods to induce diversity among the weak learners. Then, model selection is performed by variable selection and leave-one-out cross-validation to obtain the final ensemble configuration. Finally, the outputs of the individual models are combined by a weighted average. The system designed is applied on data taken from six forecast stations located within a range of 48 km from Fukuoka to forecast one-step ahead rainfall for the rainy season and monthly rainfall in the city. The experimentation performed clearly show how the hybrid model ensemble outperforms the prediction accuracy obtained by the individual models.

Opposite to the two previous papers, Woźniak and Krawczyk's contribution is related to the last ensemble system design stage, the combination mechanism. In their manuscript the authors take their previous ensemble design method, Adaptive Splitting and Selection (AdaSS), originally based on combining the weak learner decisions by majority voting, and extends it considering a weighted combination of the individual classifier discriminant functions. The classifier fusion weights are not only defined at the classifier level but also at the class level, thus introducing a larger number of freedom degrees and consequently obtaining a higher accuracy. The obtained results are very promising as the proposed method outperforms similar methods based on the "clustering and selection" approach in the most of the datasets considered.

Finally, in the last contribution to the special section, Trawiński *et al.* introduce a new methodology to perform multiple comparisons of machine learning regression algorithms' performance. In their paper, they demonstrate how the typically used pairwise Wilcoxon test leads to overoptimistic conclusions when applied to multiple comparisons. Working on the results obtained from applying six neural methods to a test bed composed of twenty-nine regression datasets, the authors show that the obtained outputs do not satisfy the normality requirements. Taking this fact as a base, some guidelines for the proper application of nonparametric statistical tests and post hoc procedures devised to perform multiple comparisons of machine learning regression methods are presented.

Finally, as guest editors of this special section, we would like to thank all the authors for their high quality contributions and over thirty independent reviewers from some twenty countries for their outstanding cooperation, as well as for their interesting comments and suggestions that helped the authors to improve the final versions of their papers. Besides, we sincerely thank Professor Józef Korbicz, the Editor-in-Chief of the *International Journal of Applied Mathematics and Computer Science*, for providing us with the opportunity to edit this special section and for his close collaboration during the editorial process.

Oscar Cordón
Department Computer Science and Artificial Intelligence
University of Granada, Spain

Przemysław Kazienko
Institute of Informatics
Wrocław University of Technology, Poland

December 2012

Oscar Cordón is a professor at the University of Granada, Spain. From 2006 to 2011 he was one of the founding researchers of the European Centre for Soft Computing, Spain, where he is still contracted as a scientific consultant. He has published around 280 peer-reviewed scientific publications, including a research book on genetic fuzzy systems and 67 JCR/SCI-indexed journal papers (h-index = 23, WOK). He has also advised 13 Ph.D. dissertations, coordinated 20 research projects, and developed an international patent on an intelligent system for forensic identification. He is an associate editor for 11 international journals and has received 3 international awards (from IEEE CIS, IFSA, and EUSFLAT).

Przemysław Kazienko received his M.Sc. and Ph.D. degrees in computer science with honours, both from the Wrocław University of Technology, Poland, in 1991 and 2000, respectively. He obtained his habilitation degree from the Silesian University of Technology, Poland, in 2009. Presently, he serves as a professor of the Wrocław University of Technology at the Institute of Informatics. He was also a research fellow at the Intelligent Systems Research Centre, British Telecom, UK, in 2008. For several years, he held the position of the deputy director for development at the Institute of Applied Informatics. He was a co-chair of many international scientific events and a guest editor of several special issues in JCR-listed journals. He is a member of the editorial board of *Social Network Analysis and Mining, International Journal of Knowledge Society Research, International Journal of Human Capital and Information Technology Professionals*, as well as *Social Informatics*. He has authored over 130 peer-reviewed papers on a variety of topics related to multiple model classification, collective classification and relational learning, social and complex network analysis, knowledge management, collaborative systems, data mining, recommender systems, information retrieval, data security, and system integration. He also initialized and led over 25 projects chiefly in cooperation with commercial companies, including large international corporations.