Guest Editorial: Special Section on White Box Nonlinear Prediction Models

PREDICTIVE modeling aims at predicting the future using patterns learned on past data. Both classification and regression are popular predictive modeling tasks and have been intensively studied in the literature. For both tasks, a myriad of techniques have been introduced in books, journals, and conference proceedings, ranging from simple linear regression, to advanced nonlinear prediction methods. New techniques have been developed either by extending existing methods in, e.g., statistics, machine learning or artificial intelligence, or by introducing new learning paradigms (e.g., kernel methods, ensemble learning, and swarm intelligence). Most of these techniques have been implemented in various commercial software packages and hardware products, or in an open-source environment. The bulk of the current academic literature and research focusses far too often on developing new, complex algorithms capable of modeling nonlinear relationships by optimizing a context-specific statistical performance measure of interest such as (regularized) mean squared error and cross-entropy error. Although these nonlinear techniques typically provide the most accurate prediction models (e.g., based on a universal approximation property), they are often not suitable to be used in many practical application domains because of their lack of transparency and comprehensibility. Indeed, the estimated models often boil down to a complex nonlinear mathematical formula, relating an output measure of interest to a set of inputs in a black box and opaque way. In domains where validation of the underlying model is required (e.g., credit risk analysis and medical diagnosis), a clear insight into the reasoning made by the nonlinear prediction model is desired and necessary. The need for white box prediction models is further amplified by the fact that these models are often used to steer critical processes in various contexts such as engineering and economics. More and more human end users use the outcome of these models in their daily decision making and are typically reluctant to use complex, opaque, black box models for this purpose. Our aim with this special issue is twofold. First, we hope to increase the awareness of researchers and practitioners working on nonlinear prediction about the need to come up with new ways of creating white box nonlinear prediction models. Next, we hope the five papers included in this special issue provide some new ideas and insights about how to achieve this in a diversity of application settings.

A first way to come up with comprehensible nonlinear prediction models is by using rule extraction. Rule extraction typically starts from a complex nonlinear model (e.g., a neural network or support vector machine) and aims at extracting a set of If-Then rules mimicking the behavior of the black box as closely as possible. If-Then rules have typically been considered as interpretable and transparant as every classification made comes with a clear explanation, i.e., the rule antecedent. Three of the five papers of this special issue adopt this rather popular and effective approach to come up with white box nonlinear prediction models.

Neural-symbolic computation offers one way of implementing white box nonlinear prediction by means of rule extraction. Neural-symbolic systems open up the neural network black box by integrating nonlinear modeling with domain knowledge and rule extraction. In the first paper, Borges *et al.* introduce a novel neural-computational model capable of 1) representing temporal knowledge operators in recurrent neural networks; 2) adapting temporal knowledge models given a set of desirable system properties; 3) training the networks from examples of system behaviors; and 4) extracting a revised temporal knowledge from the trained networks. The new method has been implemented as part of a neural-symbolic toolkit and empirically evaluated using benchmark case studies.

Another approach toward rule extraction is by using neuro-fuzzy methods. Neuro-fuzzy classifiers combine concepts from neural networks and fuzzy systems to come up with white box fuzzy rule sets explaining the reasoning behavior of a neural network. Diago *et al.* propose a new neuro-fuzzy method for the quantification of qualitative judgments and evaluations of facial expressions. They suggest adding interpretability to a fuzzy-quantized holographic neural network by reducing the number of inputs, creating membership functions, and extracting fuzzy rules. Experimental results on a dataset of 20 facial images indicate that the method improves prediction accuracy while at the same time also assuring interpretability by means of the extracted fuzzy rules.

Chorowski and Zurada present a method for extracting rules by building reduced ordered decision diagrams that represent the logical relation learned by a network. The method first builds a diagram by analyzing network behavior on the training set. This is followed by generalization over the whole input space and minimizing the number of nodes in the diagram while preserving consistency with the network. An algorithm transforming decision diagrams into interpretable boolean expressions is also described. Experimental running times of rule extraction are proportional to the network's training time.

A second alternative to add interpretability to a nonlinear prediction model is by steering the training process using, e.g., prior information or improved weight initialization. This is the approach followed in the remaining two papers of this special issue. Qu and Hu present an approach to find the weights of RBF networks for regression problems in the presence of

Date of current version December 13, 2011.
Digital Object Identifier 10.1109/TNN.2011.2177735

linear equality and linear inequality priors. Examples of linear equality priors include interpolation points and invariance transformation such as periodicity. Linear inequality priors include ranking information, boundary condition information, and multiple output regression with output dependencies. The authors also differentiate between hard constraints which should always be respected and soft constraints in case prior information is not entirely accurate. Their method is illustrated using real-life data from the StatLib collection, namely pollution, Boston Housing, and California Housing datasets.

Song uses a white box approach to analyze the inner workings of recurrent neural networks for time series prediction. The resulting sensitivity and weight convergence analysis lead to insights into the tradeoff between training and testing errors, and are the basis of the novel algorithm that is proposed for the robust training of these recurrent neural networks with an output feedback loop. More specifically, to avoid slow learning and overfitting, novel weight initialization, adaptive learning, and dynamic hidden layer neurons selection schemes are suggested. Using several benchmark time series prediction datasets, it is shown that the algorithm indeed achieves superior generalization behavior.

We can conclude that present day state-of-the-art nonlinear prediction modeling is a lively research area with many new, sophisticated algorithms and techniques being developed and investigated. However, important challenges remain when putting those new scientific contributions to work in practical settings and application domains, where besides having statistically accurate prediction models, also interpretability of these models is a key concern. We hope this special issue offers some interesting new insights and attracts further research and developments in the field.

ACKNOWLEDGMENT

The Guest Editors would like to thank all authors for their contributions and the reviewers for their careful reading, timely feedback, and critical comments. Also, we are grateful to Derong Liu, the Editor-in Chief of this journal, for his invaluable support and advice in putting together this Special Issue.

BART BAESENS, *Guest Editor*
Department of Decision Sciences
and Information Management
Faculty of Business and Economics
K. U. Leuven
Leuven B-3000, Belgium
Bart.Baesens@econ.kuleuven.be

DAVID MARTENS, *Guest Editor*
Faculty of Applied Economics
University of Antwerp
Antwerp 2000, Belgium

RUDY SETIONO, *Guest Editor*
School of Computing
National University of Singapore
117417, Singapore

JACEK M. ZURADA, *Guest Editor*
Electrical and Computer Engineering
Department
University of Louisville
Louisville, KY 40292 USA

Bart Baesens is an Associate Professor with K. U. Leuven, Leuven, Belgium, and a Lecturer with the University of Southampton, Southampton, U.K. He has done extensive research on white box nonlinear prediction modeling, predictive analytics, data mining, customer relationship management, fraud detection, and credit risk management. His findings have been published in well-known international journals *Machine Learning, Management Science*, the IEEE TRANSACTIONS ON NEURAL NETWORKS, the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, and the *Journal of Machine Learning Research* and presented at top international conferences. He is the co-author of the book *Credit Risk Management: Basic Concepts*, in 2008. He regularly tutors, advises, and provides consulting support to international firms with respect to their data mining, predictive analytics, and credit risk management policy.

David Martens is an Assistant Professor with the Faculty of Applied Economics, University of Antwerp, Antwerp, Belgium. His research has been published in high impact international journals and is mainly focused on learning from social network data and the development of comprehensible data mining techniques, using support vector machines, rule extraction, and swarm intelligence. Applications of his work can be found in the banking, telecommunication, and marketing domains.

Rudy Setiono received the Bachelors degree in computer science from Eastern Michigan University, Ypsilanti, and the M.Sc. and Ph.D. degrees in computer science from the University of Wisconsin-Madison, Madison, in 1984, 1986, and 1990, respectively.

He has been with the National University of Singapore, Singapore, since 1990, and he is currently an Associate Professor. His current research interests include linear programming, nonlinear optimization, and neural networks.

Dr. Setiono served as the Vice Dean of Undergraduate Affairs from November 2001 to July 2005 at the School of Computing. He was an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS from 2000 to 2005.

Jacek M. Zurada (M'82–SM'83–F'96) received the M.S. and Ph.D. degrees (with distinction) in electrical engineering from the Technical University of Gdansk, Gdansk, Poland, in 1968 and 1975, respectively.

He currently serves as a University Scholar and a Professor with the Electrical and Computer Engineering Department, University of Louisville, Louisville, KY. He was the Department Chair from 2004 to 2006. He has published 360 journal and conference papers in the areas of neural networks, computational intelligence, data mining, image processing, and very large scale integration circuits. He has authored or co-authored three books and co-edited a number of volumes in *Springer Lecture Notes on Computer Science*.

Dr. Zurada has held visiting appointments at Princeton, Northeastern, Auburn, and foreign universities in Australia, Chile, China, France, Germany, Hong Kong, Italy, Japan, Poland, Singapore, Spain, South Africa, and Taiwan. He was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, Pt. I and Pt. II, and served on the Editorial Board of the Proceedings of IEEE. From 1998 to 2003, he was the Editor-in-Chief of the IEEE TRANSACTIONS ON NEURAL NETWORKS. He is an Associate Editor of *Neurocomputing*, *Schedae Informaticae*, and the *International Journal of Applied Mathematics, and Computer Science*, an Advisory Editor of the *International Journal of Information Technology and Intelligent Computing*, and an Editor of Springer natural computing book series. He has served the profession and IEEE in various elected capacities, including as a President of the IEEE Computational Intelligence Society (CIS) from 2004 to 2005. He has been a member and the Chair of various IEEE CIS and the IEEE Technical Activities Board (TAB) committees and the Chair of a number of IEEE symposiums and conferences. He is currently serves as the Chair of the IEEE TAB Periodicals Committee from 2010 to 2011 and will Chair the IEEE TAB Periodicals Review and Advisory Committee from 2012 to 2013. He has received a number of awards for distinction in research, teaching, and service including the Presidential Award for Research, Scholarship, and Creative Activity in 1993, the IEEE Circuits and Systems Society Golden Jubilee Medal in 1999, and the Presidential Distinguished Service Award for Service to the Profession in 2001. He is a Distinguished Speaker of IEEE CIS. In 2003, he was conferred the Title of National Professor by the President of Poland. In 2004, 2006, and 2010, he received three Honorary Professorships from Chinese universities. Since 2005, he has been a member of the Polish Academy of Sciences and has been appointed a Senior Fulbright Specialist from 2006 to 2012.