Interdisciplinary Data Analysis

Ana Carolina Lorena¹ · Anne Magaly de Paula Canuto²

Published online: 17 January 2018

© Ohmsha, Ltd. and Springer Japan KK, part of Springer Nature 2018

This Special Issue presents papers selected from the 5th Brazilian Conference on Intelligent Systems (BRACIS), which was held in Recife (Pernambuco), Brazil, from 09 to 12 October, 2016. BRACIS is sponsored by the Brazilian Computer Society (SBC) and it covers topics related to Artificial Neural Networks, Evolutionary Computation, Fuzzy Systems and other models of computational intelligence. The emphasis of BRACIS is on original theories and novel applications of these models, and the proceedings are traditionally published by the IEEE Computer Society Press. BRACIS has an international Program Committee, which includes well-established researchers from Brazil and abroad.

The papers submitted to BRACIS 2016 represented a broad range of research developed in Brazil and other countries. In 2016, a total of 176 submissions were received. After a rigorous review process, 76 papers have been accepted for publication in the IEEE proceedings. The papers with best reviews were then invited to submit an extended version for this Special Issue. The contents of each invited paper had to be substantially expanded, and the main focus of the special issue reviewers was in the originality, significance, and technical contribution of the extended papers. At the end of a rigorous reviewing process, four papers have been selected to be published in this Special Issue.

The scope of the four selected papers can be considered as interdisciplinary, ranging from Statistics and Computing to Engineering. In the first paper, we turn our

✉ Ana Carolina Lorena
aclorena@unifesp.br

Anne Magaly de Paula Canuto
anne@dimap.ufrn.br

¹ Universidade Federal de São Paulo, ICT/UNIFESP, Av Cesare Mansueto Giulio Lattes, 1201, Eugênio de Melo, São José dos Campos, SP 12247-014, Brazil

² Federal University of Rio Grande do Norte, 9, DIMAp/CCET/UFRN, Campus Universitario, Lagoa Nova, Natal, RN 59072-970, Brazil

attention to classification models applied to a different computational domain. The paper is entitled “Fault Detection in Hard Disk Drives Based on a Semi Parametric Model and Statistical Estimators”. This paper addresses the problem of fault detection on Hard Disk Drives (HDDs). The authors proposed the use of a Gaussian Mixture model to define the behavior of healthy HDDs. Then, based on this modeling, an anomaly can be detected when a statistical estimator computed over these dissimilarities exceeds a defined threshold. The proposed method, named fault detection of HDDs based on GMM and statistical estimators (FDGE), was compared to the state-of-the-art fault detection methods and achieved promising results.

The second paper, entitled “Semantic Analysis for Identifying Security Concerns in Software Procurement Edicts”, is related to semantic analysis in text mining. More specifically, this work presents an Automated Analyst of Edicts tool, which aids the analysis of a document by the automatic identification of absent relationships between its sentences and concepts related to software security risks or weaknesses. The main contribution is the use of this tool in the multi-label classification domain. In the empirical analysis, the proposed tool was compared to some software security experts, using five of the OWASP Top 10 risks. As a result of this analysis, it could be observed that a specificity of over 80% was achieved when analyzing individual sentences for multiple risks, and a 90% negative prediction probability result was obtained when applied to specific risk–sentence relationships.

In contrast to the first the papers previously described, which fall within the context of supervised learning, the next paper proposes a new semi-supervised learning algorithm. Semi-supervised learning has become popular due to the growing availability of unlabeled data from different domains. The idea is to enhance data classification by the use of both labeled and unlabeled data. One well-known semi-supervised learning algorithm is Co-training, which combines predictors based on multiple views of a data set. The paper “Fast Co-MLM: An efficient semi-supervised Co-Training method based on the Minimal Learning Machine”, presents a Co-training variant which employs Minimal Learning Machines (MLM) as base predictors (Co-MLM). Due to the high computational cost of MLM in both training and prediction stages, the paper proposes some modifications into the original Co-MLM formulation. First, a recursive least squares formulation is used to estimate the distance-mapping matrix in the training step, whilst a fast Nearest Neighbor MLM algorithm is used in classification. Fast Co-MLM had similar predictive performance to Co-MLM, with a significant computational cost reduction. It also achieved results competitive to those of other co-training-based methods.

The last paper in this Special Issue, “Centrality-based Group Profiling: A Comparative Study in Co-Authorship Networks”, deals with social network data analysis. Virtual social networks are noticeably a global phenomenon. Plenty of data is generated daily from these virtual interactions. One common analysis consists in the identification of communities of users in the networks, which correspond to groups of users that are densely connected. The presented paper addresses the relate problem of group profiling, in which the aim is to obtain descriptive profiles for the

communities. They adopt a novel relational approach which takes advantage of the network structure, called Centrality-Based Group Profiling. The algorithm makes use of network centrality measures to select a set of nodes for the characterization. Their method was able to obtain good profiles for one co-authorship network while reducing the input data to be processed by the group-profiling algorithm.

The references of the original BRACIS papers [1–4] are presented at the end of this editorial. We would like to express our thanks to the authors who accepted our invitation to submit papers for this Special Issue. We are also grateful to all the reviewers, including the members of the BRACIS Program Committee, who helped us in reviewing the submitted papers and guaranteeing the quality of the Conference and of this Special Issue. We are in debt also to the New Generation Computing editor, Masami Hagiya; the Journal Editorial Board; and Springer for the opportunity and for the efficient handling of the publication process, in special for Mio Sugino from Springer in Japan. In addition, we also would like to thank Prof. Myriam Regattieri de B. da S. Delgado, general chair of BRACIS 2016, for the invitation to organize this special issue.

References

1. Queiroz, L. P., Rodrigues, F. C. M., Gomes, J. P. P., Brito, F. T., Brito, I. C., Machado, J. C. (2016) Fault detection in hard disk drives based on mixture of Gaussians. In: Intelligent Systems (BRACIS), 2016 5th Brazilian Conference on, IEEE. p 145–150
2. Peclat, R. N., Ramos, G. N. (2016). Automatic identification of security risks in edicts for software procurement. In: Intelligent Systems (BRACIS), 2016 5th Brazilian Conference on, IEEE. p 271–276
3. Caldas, W. L., Gomes, J. P., Cacaís, M. G., Mesquita, D. P. (2016) Co-MLM: A SSL algorithm based on the minimal learning machine. In: Intelligent systems (BRACIS), 2016 5th Brazilian Conference on, IEEE. p 97–102
4. Gomes, J. E. A., Prudêncio, R. B., Nascimento, A. C. (2016) A comparative study of Group profiling techniques in co-authorship networks. In: Intelligent Systems (BRACIS), 2016 5th Brazilian Conference on, IEEE. p 373–378