Editorial: Special issue HAIS20

The twelve papers included in this special issue represent a selection of extended contributions presented at the 15th International Conference on Hybrid Artificial Intelligent Systems, HAIS, 2020 held in Gijón, Spain, 11–13 September 2020, and organized by the BISITE group and the University of Oviedo.

The International Conference on Hybrid Artificial Intelligence Systems has become a unique, established and broad interdisciplinary forum for researchers and practitioners who are involved in developing and applying symbolic and sub-symbolic techniques aimed at the construction of highly robust and reliable problem-solving techniques to present the most relevant achievements in this field.

The twelve papers are organized as follows.

In the first contribution, by Clark *et al.*, incomplete data sets, or data sets with missing attribute values, have three interpretations: lost values, attribute-concept values and 'do not care' conditions. Additionally, the process of data mining is based on two types of probabilistic approximations, global and saturated. Authors present results of experiments on mining incomplete data sets using six approaches, combining three interpretations of missing attribute values with two types of probabilistic approximations. They compare six approaches, using the error rate computed as a result of ten-fold cross validation as a criterion of quality.

The following contribution, Díaz *et al.*, addresses a variant of the job shop scheduling problem with total tardiness minimization where task durations and due dates are uncertain. This uncertainty is modelled with intervals. Different ranking methods for intervals are considered and embedded into a genetic algorithm. A new robustness measure is proposed to compare the different ranking methods and assess their capacity to predict 'expected delays' of jobs. Experimental results show that dealing with uncertainty during the optimization process yields more robust solutions. A sensitivity analysis also shows that the robustness of the solutions given by the solving method increases when the uncertainty grows.

The next paper, by Jiménez-Herrera *et al.*, presents a novel approach to forecast streaming big time series based on nearest similar patterns. This approach combines a clustering algorithm with a classifier and the nearest neighbours algorithm. It presents two separate stages: offline and online. The offline phase is for training and finding the best models for clustering, classification and the nearest neighbours algorithm. The online phase is to predict big time series in real time. In the offline phase, data are divided into clusters and a forecasting model based on the nearest neighbours is trained for each cluster. In addition, a classifier is trained using the cluster assignments previously generated by the clustering algorithm. In the online phase, the classifier predicts the cluster label of an instance, and the proper nearest neighbours model according to the predicted cluster label is applied to obtain the final prediction using the similar patterns. The algorithm is able to be updated incrementally for online learning from data streams. Results are reported using electricity consumption with a granularity of 10 minutes for 4-hour–ahead forecasting and compared with well-known online benchmark learners, showing a remarkable improvement in prediction accuracy.

In the next contribution, by Carballido *et al*., an evolutionary method called PreCLAS is presented. Its main objective is to find a submatrix with fewer rows that exhibits some group structure. Three stages of experiments were performed. First, a benchmark data set was used to assess the correct functionality of the method for clustering purposes. Then, a microarray gene expression data matrix was used to analyse the method's performance in a simple classification scenario, where differential expression was carried out. Finally, several classification methods were compared in

Advance Access published 30 March 2023 https://doi.org/10.1093/jigpal/jzac014

terms of classification accuracy using an RNAseq gene expression data set. Experiments showed that the new evolutionary technique significantly reduces the number of rows in the matrix and intelligently.

The fifth paper, by Khodorchenko *et al*., proposes an approach that formalizes the learning strategy into a vector of parameters that can be solved with evolutionary approach. Authors also propose a surrogate-based modification that utilizes machine learning methods that makes the approach for parameters search time efficient. They investigate different optimization algorithms (evolutionary and Bayesian) and their modifications with surrogates in application to topic modelling optimization using the proposed learning strategy approach. An experimental study conducted in English and Russian data sets indicates that the proposed approaches are able to find high-quality parameter solutions for additively regularized topic models and substantially reduce the execution time of the search.

The subsequent contribution, by Simic *et al*., is focused on a new hybrid clustering system combining analytic hierarchy process and weighted fuzzy c-means clustering method for diagnosing children with primary headache disorder. The proposed three-stage hybrid diagnosing system is tested on data set collected from hospitalized children in the Clinical Centre of Vojvodina, Novi Sad, Serbia. This study analyses the most important International Classification of Headache Disorders (ICHD-3) attributes. Preliminary experimental results obtained an accuracy of the proposed diagnosing system of 77.27%.

The seventh research article, by Ferdous and Ichise, presents an embedding-based entity alignment method that finds entity alignment by measuring the similarities between entity embeddings. Existing methods mainly focus on the relational structures and attributes information for the alignment process. Such methods fail while the entities have a fewer number of attributes or when the relational structure could not capture the meaningful representation of the entities. To address this problem, the authors propose EASAE, an Entity Alignment method using Summary and Attribute Embeddings. They exploit the entity summary information available in KGs for entities' summary embedding. To learn the semantics of the entity summary, authors employ Bidirectional Encoder Representations from Transformers (BERT). Their model learns the representations of entities by using relational triples, attribute triples and summary as well. Experiments on real-world data sets are performed, and the results indicate that the proposed approach outperformed the state-of-the-art models for entity alignment.

In the following contribution, by García-Victoria *et al*., an evaluation of configurations of the inception module and a mechanism for optimizing hyperparameters in deep learning architectures are stated. For this purpose, Population-based incremental learning as optimizer and MNIST as benchmark are used. The relevance of this task stems from the wide use of the Inception module in computer vision, where it holds large numbers of success cases.

Regarding the contributions of this proposal, it is aware of a low carbon footprint in the artificial intelligence area. Unacceptable deep learning architectures are no longer evaluated, thanks to the codification of these architectures. This allows saving CPU cycles and, therefore, reducing their carbon footprint. The analyses of results demonstrate that the optimized architecture of inception-A module using only three blocks, and therefore a low number of trainable parameters, achieves an excellent performance. This performance supports the use of evolutionary strategies for optimizing deep architectures while reducing the carbon footprint through the use of a single epoch during the parameters optimization. Furthermore, the best configuration is fully compatible with the usual configuration of Inception-A module.

The next research, by Basurto *et al*., proposes the implementation of a hybrid intelligent system in four steps to impute the missing values, by combining clustering and regression techniques, followed by balancing and classification tasks. This system applies regressions models to each one of the clusters built on the instances of data set. Subsequently, a variety of balancing techniques are applied to improve the classifier's ability to discern whether it is in an error or a normal state. These techniques support to obtain better classification ratios in which a robot is close to error and allow us to bring the behaviour back to a normal state. The experimentation is performed using a modern and public data set, which has been extracted from a component-based robotic system, in which different anomalies are induced by software in their components.

In the 10th contribution, Sokol *et al*. focus on increasing the efficiency of utilization of this part of cybersecurity. The main aim is to compare selected statistical models and models based on neural networks to find out which models are more suitable for network security situation awareness forecasting. Based on the analysis obtained, neural network methods prove a more accurate alternative than classical statistical prediction models in network security situation awareness forecasting. In addition, the selection criteria and suitability of time series are analysed, which do not only reflect information about the total number of security events but represent a category of security event (e.g. recon scanning), port or protocol.

The subsequent paper, by de la Cal *et al*., proposes a method based mainly on a low-power classification algorithm based on an ensemble of k-nearest neighbour (KNN) and KMeans (EKMeans) to identify FALLs and high-intensity ADLs (activities of daily living) like running, jogging and go upstairs, events based on triaxial accelerometer data gathered from wrist wearable devices. The proposal is validated on the Fall&ADL publicly available data sets UMAFAll, UCIFAll and FallAllD, considering two kinds of activity labelling: two-class and multi-class. An exhaustive comparative study between authors' proposal, and the baseline algorithms KNNs and a feed-forward neural network is deployed, where EKMeans clearly outperformed the Specificity (ADL classification) of the KNN and neural network for the three data sets. Finally, a comparative battery consumption study has been included deploying the analysed algorithms in a WearOS smartwatch, where EKMeans drains the battery from 100% to 0% in 27.45 hours, saving 5% and 21% concerning KNN and NN, respectively.

In the final contribution, by Zayas-Gato *et al.*, a novel two-step system for anomaly detection is presented and tested over several real datasets. In the first step, the novel exploratory projection pursuit Beta Hebbian Learning algorithm is applied, over each data set, to reduce the dimensionality of the original data set and to face to nonlinear data sets, by generating a new subspace of the original dataset with lower, or even higher, dimensionality selecting the right activation function. Finally, in the second step principal component analysis anomaly detection is applied of the new subspace to detect the anomalies and improve its classification capabilities. This new approach has been tested over several different real datasets, in terms of number of variables, number of samples and number of anomalies. In almost all cases, the novel approach obtained better results in terms of area under the curve with similar standard deviation values. In case of computational cost, this improvement is only remarkable when complexity of the dataset in terms of number of variables is high.

The guest editors wish to thank Professor Dov Gabbay, (Founding editor-in-chief of *Logic Journal of the IGPL*) for providing the opportunity to edit this special issue. We would also like to thank the referees who have critically evaluated the papers within the short time. Finally, we hope the reader will share our joy and find this special issue very useful.

Enrique Antonio de la Cal
Computer Science Department, Faculty of Geology
University of Oviedo, Oviedo, Spain
email: delacal@uniovi.es

José Ramón Villar Flecha
Computer Science Department, Faculty of Geology
University of Oviedo, Oviedo, Spain
email: villarjoseg@uniovi.es

Héctor Quintián
Department of Industrial Engineering, CTC
University of A Coruña, A Coruña, Spain
email: hector.quintian@udc.es

Emilio Corchado
Department of Computer Science and Automatic
University of Salamanca, Salamanca, Spain
email: escorchado@usal.es