ISSN 2007-9737

Thematic Section:

Advances in Pattern Recognition

Guest editors:

Osslan Osiris Vergara Villegas, UACJ, Mexico
Vianey Guadalupe Cruz Sánchez, UACJ, Mexico
Juan Humberto Sossa Azuela, CIC-IPN, Mexico
Jesús Ariel Carrasco Ochoa, INAOE, Mexico
José Francisco Martínez Trinidad, INAOE, Mexico
José Arturo Olvera López, BUAP, Mexico

This thematic section of *Computación y Sistemas* (CYS) contains a selection of seven papers presenting advances in the field of Pattern Recognition (PR). PR is a discipline concerned with the development of systems for helping the classification of objects into a particular class.

It can also be defined as a field that uses machine learning techniques to recognize patterns and regularities in data. The world's most advanced industries are using PR for intelligent decision-making. Moreover, industries are using PR to predict what will happen in the future with some degree of accuracy.

The guest editors carefully selected the seven papers of this thematic issue. Each paper was reviewed by at least three members of the scientific committee.

The features considered by the reviewers to accept a paper include originality, contribution to the field, soundness, and technical quality. In the following paragraphs, we provide an overview of the papers that made up this volume.

Valencia-Segura et al. from Mexico proposed a model for solving the task of automatic depression detection based on the characteristics of social network users. Two data sets related to depression on Reddit and Twitter were considered to evaluate the proposed model. First, a preprocessing stage was conducted to remove unnecessary information, such as special characters, numbers, URLs, and punctuation. Then, five features were extracted, including sentiment analysis, demographic information, polarity, personality, and thematic information. Next, the extracted features were fused using a Gated Multimodal Unit (GMU) based network. Finally, the GMU module was passed through a fully connected hidden layer, a dropout layer, and an output layer for final classification into depressed and non-depressed users. The proposed model outperforms the state-of-the-art works for both data sets.

Neri-Mendoza et al. from Mexico addressed the generic and updated text summarization tasks of a set of documents as a combinatorial optimization problem through a genetic algorithm and unsupervised textual features. The DUC01 and the TAC08 data sets were employed for testing. First, the documents of each data set were ordered chronologically to create a meta-document containing all sentences. Afterward, the text of the meta-document was separated into sentences. Then, a lexical analysis was conducted to split sentences into words. After preprocessing, the text modeling was performed. Then, the weighting and selection of sentences were implemented. The most important sentences were selected using a genetic algorithm. Finally, the authors calculated the text summarization heuristics for benchmarking. The

proposed method outperforms the state-of-the-art works for both data sets.

Chaparro-Amaro et al. from Mexico evaluated the Random Forest (RF) method for predicting the residues at the interface of Protein-Protein Interactions (PPI) that contribute most of the binding free energy (called hot spots and hot~regions). The proposal was tested on twenty-nine Bone Morphogenetic Proteins (BMP) complexes files from the Protein Data Bank (PDB). First, the protein data was preprocessed by order, standardization, and addition of hydrogen atoms. Then, features such as the B-factor, hydrophobicity index, prevalence score, Accessible Surface Area (ASA), conservation score, and the ground-state energy of the amino acids were calculated using the Density Functional Theory (DFT). Finally, the hot spots and hot regions were labeled using the ASM computational method. Random forest outperforms Support Vector Mchine (SVM) and Artificial Neural Networks (ANN) with 90% of correct classes assigned.

Ochoa-Montiel et al. from Mexico presented an evolutionary vision approach called brain programming for the automatic recognition of Acute Myelocytic Leukemia (AML) and Acute Lymphoblastic Leukemia (ALL) images. Brain programming was employed to find a set of operations through an optimization process. The operations were within a hierarchical process called the Artificial Visual Cortex (AVC), where Genetic Programming (GP) allows the discovery of a set of Visual Operators (VOs) within the AVC. Moreover, the AVC model and brain programming were employed to obtain the descriptors that guide the evolutionary process for the automatic recognition process. The data set included three types of ALL images and four kinds of AML images. For testing purposes, the random forest and the multilayer Perceptron were employed. The maximum accuracy of 83.63% was achieved using knowledge transfer compared with 80.09% obtained without knowledge transfer.

Ulloa-Poblete et al. from Chile proposed a modified Convolutional Neural Network (CNN) that assigns more importance to learn hard-to-classify voxels close to the boundaries of Multiple Sclerosis (MS) lesions. The proposal was tested with the longitudinal multiple sclerosis lesion segmentation data set. In the preprocessing stage, the correction of the intensity inhomogeneities, skull-stripping, dura mater stripping, and rigid-body registration was conducted. The CNN was trained with 650,000 patches generated from the available images. The network's input consists of three patches of size 33×33. The problem of class imbalance was addressed using the focal loss function. The model's accuracy was computed by comparing the predicted segmentation mask and the manually segmented by the expert. The segmentation results were comparable to the performance of human experts.

Rodríguez-Ramos et al. from Cuba presented a condition monitoring scheme for industrial plants characterized by its robustness versus noise and missing variables in the measurements. The proposal combines simple and effective imputation algorithms with a fuzzy classification kernel algorithm based on non-standard Pythagorean fuzzy sets. The proposal can execute online the imputation process for each observation obtained from the plant that has missing variables. The imputations methods have low computational complexity. The classification was implemented with the KPyFCM~algorithm. In the experiments, different noise levels and distinct quantities of missing variables in the measurements obtained from the plant were combined. The proposed scheme was validated using the DAMADICS test problem, obtaining excellent results.

Barajas-Montiel et al. from Mexico explored six different Multiview Learning (MVL) techniques for the classification of Electroencephalogram (EEG) signals to take advantage of complementary descriptive information from different representations of the same object. The problems of imagined speech and stress pattern recognition, areas of EEG signal classification, were addressed. The feature extraction stage combined two methods in the frequency domain and two in the time domain. Regarding imagined speech, experiments were performed using two views. The results show that accuracy can be improved by combining the information of the same object extracted from different domains. Furthermore, two feature sets were extracted from each domain regarding EEG stress signals.

For both imagined speech and stress, the multi-view technique that yielded the best results was the Majority Vote Co-Training.

We are convinced that this thematic issue will be of interest to students, researchers, and company managers working in PR and related areas.