

Guest Editors' Introduction to the Special Issue on Computational Intelligence

Computational intelligence is branch of artificial intelligence that studies a wide range of tasks and problems especially difficult for traditional algorithms to deal with: those problems that include uncertainty, highly stochastic behavior, or fuzziness. Probably the majority of problems currently dealt with within the artificial intelligence framework, and a large share of problems addressed by modern computer science in general, fall in this category. Computational intelligence seeks to achieve solutions of such complex problems with the techniques that presumably human brain uses, and with the degree of quality equal, or superior, to the quality with which humans solve such problems.

Among the problems particularly difficult for computers and particularly easy for human brain we can mention everyday human activities such as intuition, learning, noticing patterns, dealing with language, vision, and spatial orientation. Their computer counterparts are the areas of machine learning, natural language processing, and image processing, among others. Of these areas, machine learning is a set of techniques deeply penetrating all other areas of artificial intelligence, while natural language processing, image processing and computer vision, and other research areas have more applied character and are oriented to model different specific abilities of the human brain.

For this special issue, we have selected a representative collection of fourteen papers presenting the latest advances in all these areas of research and practical applications.

The first two papers in this issue represent an important application of computational intelligence: recommender systems and evaluation scales. Recommender systems improve the quality of life of the consumers by helping them to make informed decision on buying products and services, basing on the experience of other users. Evaluation scales play a key role in correct functioning of recommender systems, as well as help the businesses to improve their products and services to better match the opinions of the users.

C. Ríos et al. from Argentina in their paper "Selecting and weighting users in collaborative filtering based POI recommendation" analyze a wide range of techniques that improve location awareness of the recommender systems. In many scenarios it is important for the user to obtain recommendations of products and services available in the geographic vicinity of that user, as well as rated highly by other users from the same geographical region. In addition, geographic awareness of the system helps in disambiguating the names of products and services, such as local restaurants or shops with common names. The authors show how to extract information on geographic location from the data available in various social networks.

I. Batyrshin *et al.* from Mexico and Russia in their paper “Bipolar Rating Scales: A Survey and Novel Correlation Measures Based on Nonlinear Bipolar Scoring Functions” give a comprehensive state of the art review of the theory and practices of the use of bipolar rating scales: evaluation scales in which people can express various degrees of positive or negative opinion on some product or subject. Such scales are ubiquitous in all kinds of evaluation and their applications, from opinion mining and recommender systems to healthcare, administration and politics. Basing on the analysis of the current state of the art, the authors propose novel, improved techniques for the analysis of opinions expressed in terms of such scales. The techniques proposed in the paper are based on the so-called non-linear bipolar scoring functions, which describes in the objective terms the degree of utility, or satisfaction, expressed in a given scale.

The next group of five papers are devoted to natural language processing, one of the key areas of research and application in computational intelligence and probably the most “human” one. In a wide sense, natural language processing is a research area devoted to the ways to enable the computers to deal with text, or speech, in ordinary human language, such as English or Hungarian, the way people do, or even better—given the ability of the computer to quickly process huge quantities of data. In more specific applications, natural language processing research results in a wide range of important technologies, from information retrieval and machine translation to opinion mining and sentiment analysis. The latter techniques, for example, are the basis for the development of business intelligent tools and recommender systems.

I. Markov *et al.* from Mexico and Portugal in their paper “Authorship Attribution in Portuguese Using Character N-grams” present a method for detecting the author of a given text out of a number of possible alternatives. They show that character n-grams are very good features for this task, and analyze the performance of a wide range of types of character n-grams. Authorship attribution has numerous applications in culture, education, forensics, and business intelligence. For example, in culture and education it helps fighting plagiarism, a dangerous phenomenon that has become threateningly common with the proliferation of Internet. In forensics, it allows for the determination of the author of texts related with, for example, a crime. In business intelligence, authorship attribution and author profiling methods improve the performance of opinion mining techniques.

J.-P. Posadas-Durán *et al.* from Mexico in their paper “Algorithm for Extraction of Subtrees of a Sentence Dependency Parse Tree” describe the procedure for enumerating the so-called syntactic n-grams present in a syntactic dependency tree of a sentence. Syntactic information present in a sentence is important for its interpretation. However, it is difficult to represent this information in a way useful for modern machine-learning methods, which are mostly suitable for data represented in the form of vectors and not trees or other graphs. Typically, for the use with such methods, the text is represented in the form of word n-grams, which are linear sequences of words located in the text next to each other, with the

syntactic information completely lost. The technique of syntactic n-grams allows to keep the syntactic information while still representing the text as a vector of features. Thus, extracting these features from the text becomes the basic task for any application of syntactic n-grams. The paper presents a detailed algorithm for extracting these features from the texts.

S. Miranda-Jiménez and E. Stamatatos from Mexico and Greece in their paper “Automatic Generation of Summary Obfuscation Corpus for Plagiarism Detection” continues the discussion of the topic of authorship attribution and plagiarism detection by describing a method they used for automatic generation of a corpus of plagiarized documents of a specific type: plagiarism obfuscated via summarization. The corpus was used as a dataset for the most prestigious international competition of plagiarism detection systems. Automatic plagiarism detection is nowadays extremely important for the normal functioning of our education system as well as academia, given, on the one hand, the ease of committing this kind of severe academic misconduct using huge information resources of the Internet and, on the other hand, the fact that our education system and academia completely depend on the evaluation of texts authored by the student or researcher for correct scoring and promotion of productive and honest researchers.

O. Pichardo-Lagunas et al. from Mexico in their paper “Automatic detection of semantic primitives with multi-objective bioinspired algorithms and weighting algorithms” address the topic of automatic semantic analysis of explicatory dictionaries and evaluation of their quality via detection of primitive concepts on which the description of all other words can be based—much as the description of all notions in school geometry is based on a few non-definable concepts such as point and line. For this purpose, they represent the dictionary as a directed graph, with words being the nodes, and an arc from one word to the other if the latter is a part of the description of the former. Using computational intelligence algorithms, the authors determine the optimal way of making this graph cycle-free; then the nodes only used in the definitions of other words represent the optimal defining vocabulary of the given dictionary.

J. Alvarado-Uribe et al. from Mexico in their paper “Semantic Approach for Discovery and Visualization of Academic Information Structured with OAI-PMH” discuss the ways for analysis of the complex network of interrelated information on published scientific papers available from open-source public repositories via indexing metadata. Open source publication model is considered by many researchers to be the best way of dissemination of scientific results. However, the great amount of published papers available via open source publishers and repositories requires efficient aggregation and search tools for its analysis and extraction of information relevant for a specific user and specific research topic. The authors present the tools they have developed for visualization of the structure of information available in open-source repositories and for

finding relevant materials by semantic queries using ontologies and other advanced analysis techniques.

The next group of three papers is devoted to another important research and application area of computational intelligence: image processing and computer vision, which is responsible for the ability of intelligent computers to “see” and interpret images as persons do, or even better. Among numerous applications of this research area are image retrieval, medical diagnostics, public security and forensics, agricultural monitoring, robotics, and autonomous vehicles, to name only a few. Image processing was the area where modern deep learning revolution began and from which it has spread to other areas of computational intelligence, such as natural language processing and general machine learning.

E. Moya-Albor *et al.* from Mexico in their paper “An Edge Detection Method using a Fuzzy Ensemble Approach” explain the use of fuzzy logic techniques for edge detection in image processing. Edge detection is one of the basic techniques in image processing, which is a preprocessing step for more advanced classification and image recognition methods. The task of edge detection is to determine the boundaries of the objects seen on an image, or the boundaries between different parts or faces of the object, such as different walls of a building (the one facing the camera and the ones facing sideways), the limits of a road in front of an autonomous vehicle, the contours of human figures, etc. The task is particularly interesting because it involves global analysis of the image and not only relations between neighboring pixels. The authors show that their method outperforms the known methods on a ground truth dataset for which the task has been previously solved manually.

H. Castillejos-Fernández *et al.* from Mexico in their paper “An intelligent system for the diagnosis of skin cancer on digital images taken with dermoscopy” apply edge detection techniques and fuzzy logic classification to the task of automatic skin cancer diagnosis. Skin cancer is a major health threat in all countries across the world, difficult to diagnose with traditional methods, which require highly qualified and experienced medical personnel. Such personnel is not available in many places, especially in highly populated regions of Asia, Africa, and Latin America with developing economy. On the other hand, early diagnosis of skin cancer is crucial to reduce mortality rate from this disease. The combination of these two factors makes computational methods of automatic diagnosis especially relevant. The authors present a detailed diagram of the system they have developed, and show that their method outperforms existing state-of-the-art classifiers on an available dataset of pre-classified medical images.

T. Katanyukul and J. Ponsawat from Thailand in their paper “Customer Analysis via Video Analytics: Customer Detection with Multiple Cues” present a computer vision application for video surveillance in closed circuit video system in a shop or supermarket. Identification and monitoring of specific customers provides important security and business intelligence information through the analysis of

customer behavior in a commercial establishment. However, such identification is a complex process that involves a number of cues of different nature, such as the semantic and spatial context, common-sense and domain knowledge, previous experience, etc. The authors present an integrated pipeline of feature extraction, classification, and integration of various sources of knowledge for detection of persons in the surveillance data. They report very significant improvement of more than 42% over the existing methods in terms of precision.

Finally, the last four papers sample practical useful applications of traditional computational intelligence methods such as forecasting, clustering, optimization, and intelligent control and show how these techniques can be used in diverse practical tasks from disaster prediction, bioinformatics, and economy to robotics and mechatronics.

Justin Parra et al. from the US in their paper “Use of Machine Learning to Analyze and – Hopefully – Predict Volcano Activity” analyze a wide specter of natural phenomena that precede known volcanic eruptions, with the ultimate goal of predicting new volcanic eruptions in time for the evacuation of people from the affected zone. The task is very important because, as the authors mention, hundreds of volcanoes in the state of unrest, ready to erupt at any moment, are located near large urban areas. As a case study, the authors consider a well-documented 1999 eruption of Redoubt volcano in Alaska, with rich geological and geophysical data relevant for this eruption publicly available through Smithsonian Institution Global Volcanology Project, as well as the Aerocom database. While not reporting a ready forecasting technique, the authors show that a suitable analysis of the precursors of eruption yields geophysically meaningful results, which makes such analysis promising for eventual development of algorithms for predicting dangerous volcanic activity.

I. Bonet et al. from Colombia in their paper “Clustering of Metagenomic Data by Combining Different Distance Functions” develop a novel clustering method based on the consensus of clustering results with different similarity measures and use this method for identification of species by genome sequences extracted from a natural mix of genetic material. In pure laboratory experiments one can isolate a species of microorganisms and obtain a clean sample of genomic material. However, in real-life environment often only a mix of genetic fragments belonging to many different organisms is available, with many of them not being identifiable using existing genomic databases. In such circumstances it is important to define automatically which genetic fragments belong to the same species and cluster the genetic sequences by this criterion. The authors solve this problem using a variant of k-means classifier with an ensemble of different distance functions.

J. G. Flores Muñiz et al. from Mexico, Russia, the US, and Ukraine in their paper “Gaussian and Cauchy Functions in the Filled Function Method – Why and What Next: On the Example of Optimizing Road Tolls” argue for computational

complexity to be an important criterion in selection the best smoothing function in filled function method of solving optimization problems. In optimization of functions with complex behavior, the key issue is to avoid local optima, since this prevents the algorithm from finding the globally optimal solution. The filled function method consists in approximating the original function by a smoother function with much simpler behavior the global optimum of which is easier to find; the global optimum of the original function is likely to be located near the global optimum of the smoothed function. The authors explain why two particular functions often used for smoothing are so efficient, and illustrate this with a case study of the economic problem of optimization of road tolls.

V. V. Chikovani *et al.* from Ukraine in their paper “External Disturbances Rejection by Differential Single-Mass Vibratory Gyroscope” introduce a novel operational mode of a vibratory gyroscope: a differential operation mode, which in their experiments shows high improvement in robustness as compared with the usual rate mode. Gyroscopes are essential for spatial orientation and stabilization of intelligent physical systems such as drones, robotic arms, and virtual reality devices, to name a few. Many of these intelligent devices operate in real-life environments where they are prone to mechanical stress such as vibrations and shocks, which prevent traditional gyroscopes from normal function. With their detailed analysis the authors show that in the new operational mode the sensitivity of the gyroscope to vibration and shocks is greatly reduced.

This special issue of *Acta Polytechnica Hungarica* devoted to diverse topics of computational intelligence theory and applications will be useful to researchers and students working in such areas of artificial intelligence as recommender systems, natural language processing, image processing and computer vision, forecasting, clustering, optimization, and intelligent control.

**Ildar Batyrshin****Alexander Gelbukh****Grigori Sidorov**

CIC, Instituto Politécnico Nacional, 07738 Mexico City, Mexico

nlp.cic.ipn.mx/~batyrshin

www.cic.ipn.mx/~gelbukh

www.cic.ipn.mx/~sidorov