Chapter 2
An Uncertainty-Based Model for Optimized Multi-Label Classification

J. Anuradha
VIT University, India

B. K. Tripathy
VIT University, India

ABSTRACT

The data used in the real world applications are uncertain and vague. Several models to handle such data efficiently have been put forth so far. It has been found that the individual models have some strong points and certain weak points. Efforts have been made to combine these models so that the hybrid models will cash upon the strong points of the constituent models. Dubois and Prade in 1990 combined rough set and fuzzy set together to develop two models of which rough fuzzy model is a popular one and is used in many fields to handle uncertainty-based data sets very well. Particle Swarm Optimization (PSO) further combined with the rough fuzzy model is expected to produce optimized solutions. Similarly, multi-label classification in the context of data mining deals with situations where an object or a set of objects can be assigned to multiple classes. In this chapter, the authors present a rough fuzzy PSO algorithm that performs classification of multi-label data sets, and through experimental analysis, its efficiency and superiority has been established.

INTRODUCTION

In the past few years, although information retrieval has become easier, retrieving relevant information from the repository has become a challenge as the data available in the repositories grow exponentially. Retrieval of meaningful and appropriate data is crucial for qualitative decision making. This can be achieved by an efficient knowledge discovery or data mining tool. Classification, clustering and feature selection are some of the popular mechanisms those help in analyzing data to identify hidden patterns. Classification is a supervised technique that classifies the pattern based on the sample data using standard algorithms.

Classification problem will become complex when there exist high possible combination of patterns. Rule generation by tree or other induction techniques face the difficulty in generating unambiguous optimized rule from the complex, vague data. Clustering is a popular unsupervised learning technique that partitions data into clusters of data having similar characteristics. The challenging job of clustering is that, the sharp identification of the dissimilarity and the degree of similarity that helps to classify the data into different groups. The dynamic and adaptable nature of algorithms is important for formation of good clustering. Feature selection is the problem of filtering out the essential data and discarding the irrelevant information from the given inputs. This is an important preprocessing step that the performance of knowledge discovery can be enhanced or reduced. The self-adaptability nature of evolutionary algorithm handles these problems in a simple and easy way to produce the best solution from a large data.

Modern optimization techniques have aroused great interest among the scientific and technical community in a wide variety of fields recently, because of their ability to solve problems with a non-linear and non-convex dependence of design parameters. Several new optimization techniques have emerged in the past two decades that mimic biological evolution, or the way biological entities communicate in nature. The most representative algorithms include Genetic Algorithms (GA), Particle Swarm Optimization (PSO) and the method of Differential Evolution (DE).

By analogy with natural selection and evolution, in classical GA the set of parameters to be optimized (genes) defines an individual or potential solution X (chromosome) and a set of individuals makes up the population, which is evolved by means of the selection, crossover and mutation genetic operators. The optimization process used by the GA follows the next steps. The genetic algorithm generates individuals (amplitude excitations and phase perturbations of the antenna elements). The individuals are encoded in a vector of real numbers, that represents the amplitudes, and a vector of real numbers restrained on the range $(0, 2\pi)$, that represents the phase perturbations of the antenna elements.

One of the main drawbacks of GA is their lack of memory, which limits the search and convergence ability of the algorithms. In GA, the concept of memory relies on elitism, but there is no stronger operator to propagate accurate solutions in a faster way. However, the PSO algorithm emerges as a powerful stochastic optimization method inspired by the social behavior of organisms such as bird flocking or fish schooling, in which individuals have memory and cooperate to move towards a region containing the global or a near optimal solution. PSO like any other evolutionary algorithm is an optimization technique that performs randomized search in the solution space for an optimal solution by updating generations.

Particle Swarm Optimization (PSO) like any other evolutionary algorithm it is an optimization technique that performs randomized search in the solution space for an optimal solution by updating generations. PSO has less number of computations compared to Genetic Algorithm (GA), as it does not have operations like cross over and mutation. PSO has a swarm of particles, where swarm represents a set of solutions (solution space) and a particle represents a potential solution. It was developed by Kennedy and Eberhart (Kennedy et al, 1995) in 1995. This model was developed based on the behavior of flock of birds flying coherently with a competition that leads to a flight with uniform speed and direction influenced by its positions, velocity and their previous state. PSO has swarm of particles searching for the solution in the solution space. Each particle keeps track of its coordinates in the problem space which are associated with the best solution (fitness) it has achieved so far. This value is called *pbest*. Another "best" value that is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the neighbors of the particle. This location is called l*best*. When a particle takes all the population as its topological neighbors, the best value is a global best and is called *gbest*. Different topologies of particles are used in the search space out of those, the most popular are

Related Content

Agent-Based Semantic Interoperability of Geo-Services
Iftikhar U. Sikderand Santosh K. Misra (2010). *Methodological Advancements in Intelligent Information Technologies: Evolutionary Trends (pp. 110-127).*
www.irma-international.org/chapter/agent-based-semantic-interoperability-geo/38524

A Novel Review on the Adaptation of Artificial Intelligence in Human Resources Management by Organizations in Gulf Countries
Astha Bhanot (2022). *Handbook of Research on Innovative Management Using AI in Industry 5.0 (pp. 19-38).*
www.irma-international.org/chapter/a-novel-review-on-the-adaptation-of-artificial-intelligence-in-human-resources-management-by-organizations-in-gulf-countries/291459

Opportunities of Public Transport Experience Enhancements with Mobile Services and Urban Screens
Marcus Foth, Ronald Schroeterand Jimmy Ti (2013). *International Journal of Ambient Computing and Intelligence (pp. 1-18).*
www.irma-international.org/article/opportunities-public-transport-experience-enhancements/75567

Comparative Analysis of Neural Network and Fuzzy Logic Techniques in Credit Risk Evaluation
Asogbon Mojisola Graceand Samuel Oluwarotimi Williams (2016). *International Journal of Intelligent Information Technologies (pp. 47-62).*
www.irma-international.org/article/comparative-analysis-of-neural-network-and-fuzzy-logic-techniques-in-credit-risk-evaluation/145777

ANN Development with EC Tools: An Overview
Daniel Riveroand Juan Rabuñal (2009). *Encyclopedia of Artificial Intelligence (pp. 125-130).*
www.irma-international.org/chapter/ann-development-tools/10236