

Chapter 3

Learning Using Soft Computing Techniques

V. Susheela Devi
Indian Institute of Science, India

ABSTRACT

This chapter focuses on a few key applications in the field of classification and clustering. Techniques of soft computing have been used to solve these applications. The first application finds a new similarity measure for time series data, combining some available similarity measures. The weight to be given to these similarity measures is found using a genetic algorithm. The other applications discussed are for pattern clustering. A Particle Swarm Optimization (PSO) algorithm has been used for clustering. A modification of the PSO using genetic operators has been suggested. In addition, simultaneous clustering and feature selection and simultaneous clustering and feature weighting has been discussed. Results have been given for all the techniques showing the improvement achieved using these techniques.

1. INTRODUCTION

Soft Computing techniques generally refer to the techniques which use some method of search to solve the problem. It differs from conventional computing in that it is tolerant of imprecision, uncertainty, partial truth and approximation. It also refers to techniques which mimic the methods used by human beings and animals.

Some of the soft computing techniques are evolutionary computation, neural networks, particle

swarm optimization, ant colony optimization, and fuzzy computing. This chapter talks about specific applications where the techniques mentioned above have been used. The applications are in the fields of supervised and unsupervised learning namely classification and clustering.

There are a number of classification methods like nearest neighbour, decision tree, support vector machines, rule based systems etc. It is also possible to use soft computing techniques like neural networks, evolutionary computations,

particle swarm optimization etc. for carrying out classification. Even techniques for feature selection, feature extraction and prototype selection can be carried out using these techniques.

When the class labels of patterns are fuzzy, it means that each pattern belongs to each class with a membership value. When we use fuzzy classification techniques, we get the fuzzy membership value of each test pattern to every class which has to be defuzzified to get the crisp class label.

A rough set consists of two parts, the upper approximation which consists of objects which definitely belong to the set and the lower approximation which consists of objects which might belong to the set. In the context of pattern classification, each class is a rough set which has some patterns which belong to its upper approximation and some patterns which belong to its lower approximation. From the point of view of a pattern, it can belong to the upper approximation of one class or the lower approximation of more than one class.

Clustering can also be carried out using soft techniques like genetic algorithms, simulated annealing, particle swarm optimization etc.

The chapter will discuss the following topics:

1. Use of genetic algorithms for time series classification (Dohare & Devi, 2011).
2. Use of PSO for classification and clustering.
3. Modified PSO with genetic operators for clustering (Swetha & Devi, 2012).
4. Use of PSO for feature selection and clustering simultaneously (Swetha & Devi, 2012).
5. Use of PSO for feature weighting and clustering simultaneously (Swetha & Devi, 2012)

2. GENETIC ALGORITHMS FOR TIME SERIES CLASSIFICATION

Distance-based time series classification can be done using a distance measure. There are a number of distance measures each of them have

their own strengths and weaknesses. The distance measure used is crucial to the performance of the classification algorithm. The method discussed in this section, combines a number of distance measures by assigning weights to each one. The weights assigned are found by using a genetic algorithm.

Time series data consists of multivariate data where the values of one or more variable is in the form of a sequence. Time series data are found everywhere, for example, weather data, stocks and share prices, energy consumption at a place etc. These consist of data which vary in time. Generally, a time series

$$t = [t_p, \dots, t_r],$$

is an ordered set of r data points which are typically measured at successive points of time spaced at uniform time intervals. A time series could be a pattern which belongs to a particular class. The problem of time series classification is to learn a classifier C , which is a function that maps a time series t to a class label l , that is, $C(t) = l$ where $l \in L$, the set of class labels.

There are three basic methods of carrying out time series classification:

- **Distance Based Classification:** Distance based methods computes the distance or similarity between pairs of time series. The distance function, which measures the similarity between sequences, determines the quality of the classification significantly (Ding et.al, 2008) (Keogh & Kasetty, 2002) (Xi et.al, 2006).
- **Feature Based Classification:** Feature based methods transform time series data into feature vectors and then apply conventional classification methods. Feature selection plays an important role in these methods as it decreases the dimensionality of the data (Lesh, Zaki & Ogihara, 1999; Chuzhanova, Jones & Margetts, 1998).

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/learning-using-soft-computing-techniques/97053

Related Content

Computerized Detection of Lung Nodules on Chest Radiographs: Application of Bone Suppression Imaging by Means of Multiple Massive-Training ANNs

Sheng Chen and Kenji Suzuki (2012). *Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis* (pp. 122-144).

www.irma-international.org/chapter/computerized-detection-lung-nodules-chest/62227

Using MLP Neural Networks to Detect Late Blight in Brazilian Tomato Crops

Sergio Manuel Serra Cruz and Gizelle Kupac Vianna (2020). *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications* (pp. 1087-1108).

www.irma-international.org/chapter/using-mlp-neural-networks-to-detect-late-blight-in-brazilian-tomato-crops/237923

The Formal Design Model of a Lift Dispatching System (LDS)

Yingxu Wang, Cyprian F. Ngolah, Hadi Ahmadi, Philip Sheu and Shi Ying (2009). *International Journal of Software Science and Computational Intelligence* (pp. 111-137).

www.irma-international.org/article/formal-design-model-lift-dispatching/37491

Analyze Physical Design Process Using Big Data Tool: Hidden Patterns, Performance Measures, Predictive Analysis and Classifying Logs

Waseem Ahmed and Lisa Fan (2015). *International Journal of Software Science and Computational Intelligence* (pp. 31-49).

www.irma-international.org/article/analyze-physical-design-process-using-big-data-tool/141240

Sentiment Analysis of COVID-19 Tweets Using Adaptive Neuro-Fuzzy Inference System Models

Sabri Sabri Mohammed, Brahami Menaouer, Abid Faten Fatima Zohra and Matta Nada (2022). *International Journal of Software Science and Computational Intelligence* (pp. 1-20).

www.irma-international.org/article/sentiment-analysis-of-covid-19-tweets-using-adaptive-neuro-fuzzy-inference-system-models/300361