Chapter 1
Recent Neuro-Fuzzy Approaches for Feature Selection and Classification

Heisnam Rohen Singh
NIT Silchar, India

Saroj Kr Biswas
NIT Silchar, India

Monali Bordoloi
NIT Silchar, India

ABSTRACT

Classification is the task of assigning objects to one of several predefined categories. However, developing a classification system is mostly hampered by the size of data. With the increase in the dimension of data, the chance of irrelevant, redundant, and noisy features or attributes also increases. Feature selection acts as a catalyst in reducing computation time and dimensionality, enhancing prediction performance or accuracy, and curtailing irrelevant or redundant data. The neuro-fuzzy approach is used for feature selection and classification with better insight by representing knowledge in symbolic forms. The neuro-fuzzy approach combines the merits of neural network and fuzzy logic to solve many complex machine learning problems. The objective of this article is to provide a generic introduction and a recent survey to neuro-fuzzy approaches for feature selection and classification in a wide area of machine learning problems. Some of the existing neuro-fuzzy models are also applied to standard datasets to demonstrate their applicability and performance.

1. INTRODUCTION

The focus of this era is not simply serving the purpose of a work but to optimize the process involved, in order to minimize time and space complexity. Machine learning algorithms in pattern recognition, image processing and data mining mainly ensure classification. These algorithms operate on a huge amount of data with multiple dimensions, from which knowledge is extracted. However, the entire dataset in hand does not always prove to be significant to each and every domain. An important concept that contributes extensively in classification and better understanding of the domain is feature selection (Kohavi and John, 1997). Feature selection is a process of selecting a subset of features from a set of features in a balanced manner, without losing most of the characteristics and identity of the original object. There are two factors that affect feature selection – irrelevant features and redundant features (Dash and Liu, 1997). Irrelevant features are those which provide no useful information in that context and redundant features are those which provide the same information as the currently selected features.

Selection of an optimal number of distinct features contributes substantially to the improvement of the performance of a classification system with lower computational effort, data visualization and improved understanding of computational models. Feature selection also reduces the running time of learning algorithm, the risk of data overfitting, dimensions of the problem and cost of future data acquisition (Guyon and Elisseeff, 2003). Thus, in order to cope up with the rapidly evolving data, many researchers have been proposing different feature selection techniques for classification tasks.

The main goals of feature selection are to select the smallest feature subset that yields the minimum generalization error, to reduce time complexity and to reduce memory and money for handling large datasets (Vergara and Este´vez, 2014). In most common scenarios, feature selection methods are used for solving classification problems or are a part of a classification problem. Many classical techniques exist for the purpose of feature selection such as Mutual Information (MI), decision tree, Bayesian network, genetic algorithm, Support Vector Machine (SVM), K-nearest neighbor (K-nn), Pearson correlation criteria, Linear Discriminant analysis (LDA), Artificial Neural Network (ANN), Fuzzy sets. The choice of using a specific algorithm is a critical step as no such best algorithm exists that fits for considering every scope and solving every problem of feature selection and classification.

The use of Mutual Information (MI) for feature selection can be found in many contributions by different researchers (Vergara and Estevez, 2014; Peng et al.2005; Grande et al., 2007; Chandrashekar and Sahin, 2014; Battiti, 1994). Mutual information provides the dependencies between variables in terms of their probabilistic density functions. However, if one among the two variables is continuous, a limited number of samples obtained after feature selection makes the computation of the integral in the continuous space a bit challenging (Peng et al. 2005). It has also been found that MI does not work efficiently in high-dimensional spaces and there exists no standard theory for MI normalization (Vergara and Estevez, 2014).

Decision tree is a tree-based knowledge representation methodology used to represent classification rule; where at each node one can select the most useful feature for classification using some appropriate estimation criteria. Decision tree methods like ID3 and CART, in general, split a node using only a single attribute. However, use of combinations of attributes makes more sense in some cases (Setiono and Liu, 1997) Also, their performance degrades if they face many features which are unnecessary for the desired output prediction (Kohavi and John, 1997). In decision tree interpretation, the constructed tree may contain sub-tree(s) containing irrelevant attributes even when the data is not noisy and which misleads the user's interpretation and thus lead to over-fitting and even data fragmentation.

Related Content

Topic Modeling Techniques for Text Mining Over a Large-Scale Scientific and Biomedical Text Corpus

Sandhya Avasthi, Ritu Chauhanand Debi Prasanna Acharjya (2022). *International Journal of Ambient Computing and Intelligence (pp. 1-18).*

www.irma-international.org/article/topic-modeling-techniques-for-text-mining-over-a-large-scale-scientific-and-biomedical-text-corpus/293137

Cech Fuzzy Soft Closure Spaces

Rasha Naser Majeed (2018). *International Journal of Fuzzy System Applications (pp. 62-74).*

www.irma-international.org/article/cech-fuzzy-soft-closure-spaces/201558

Generalized Fuzzy Divergence Measure, Pattern Recognition, and Inequalities

Ram Naresh Saraswatand Neha Khatod (2022). *International Journal of Fuzzy System Applications (pp. 1-22).*

www.irma-international.org/article/generalized-fuzzy-divergence-measure-pattern-recognition-and-inequalities/285983

Cybersecurity: Cybercrime Prevention in Higher Learning Institutions

Adel Ismail Al-Alawi, Arpita A. Mehrotraand Sara Abdulrahman Al-Bassam (2020). *Implementing Computational Intelligence Techniques for Security Systems Design (pp. 255-274).*

www.irma-international.org/chapter/cybersecurity/250616

A Secure and Privacy-Preserving Approach to Protect User Data across Cloud based Online Social Networks

 Neelu khareand Kumaran U. (2021). *Research Anthology on Artificial Intelligence Applications in Security (pp. 560-585).*

www.irma-international.org/chapter/a-secure-and-privacy-preserving-approach-to-protect-user-data-across-cloud-based-online-social-networks/270616