Engineering Applications of Artificial Intelligence 45 (2015) 482–487

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

# Feature selection based on the center of gravity of BSWFMs using NEWFM

Sang-Hong Lee

*Department of Computer Science & Engineering, Anyang University, Anyang-si, Republic of Korea*

ARTICLE INFO

*Article history:*
Received 18 May 2015
Received in revised form
28 July 2015
Accepted 6 August 2015
Available online 29 August 2015

*Keywords:*
Feature selection
Euclidean distance
Fuzzy neural networks
Center of gravity

ABSTRACT

Feature selection has commonly been used to remove irrelevant features and improve classification performance. Some of features are irrelevant to the learning process; therefore to remove these irrelevant features not only decreases training and testing times, but can also improve learning accuracy. This study proposes a novel supervised feature selection method based on the bounded sum of weighted fuzzy membership functions (BSWFM) and Euclidean distances between their centers of gravity for decreasing the computational load and improving accuracy by removing irrelevant features. This study compares the performance of a neural network with a weighted fuzzy membership function (NEWFM) without and with the proposed feature selection method. The superiority of the NEWFM with feature selection over NEWFM without feature selection was demonstrated using three experimental datasets from the UCI Machine Learning Repository: Statlog Heart, Parkinsons and Ionosphere. 13 features, 22 features, and 34 features were used as inputs for the NEWFM without feature selection and these resulted in performance accuracies of 85.6%, 86.2% and 91.2%, respectively, using Statlog Heart, Parkinsons and Ionosphere datasets. 10 minimum features, 4 minimum features and 25 minimum features were used as inputs for the NEWFM with feature selection and these resulted in performance accuracies of 87.4%, 88.2%, and 92.6%, respectively, using Statlog Heart, Parkinsons and Ionosphere datasets. The results show that NEWFM with feature selection performed better than NEWFM without feature selection.



## 1. Introduction

Recently, the amount of data typically used to perform data mining or machine learning has rapidly increased in all areas (Pradipta and Partha, 2013). As a result, the noise, redundancy and complexity in data have also increased (Lazar et al., 2012). In general, additional data and input variables are thought to help classify or determine certain facts, but too much data or input variables may trigger inefficiency with respect to memory and time (Zhou and Gan, 2007). Furthermore, data that is irrelevant to other data may lead to incorrect outcomes. Therefore, feature selection is necessary to remove the irrelevant input features (Guyon and Elisseeff, 2003). Feature selection is an important and difficult problem in data mining and machine learning (Banka and Dara, 2015). It not only decreases the computational load, but can also improve learning accuracy and enhance output comprehensibility (Freeman et al., 2015). It also removes features that overlap or have no relevance, improves classification performance, and uses minimal features, reducing operational cost and thereby enhancing classification performance (Freeman et al., 2015; Fakhraei et al., 2014). To select features in important and difficult problems, fuzzy neural networks, particle swarm optimization, rough set approaches, and support vector machines (SVMs) have all been proposed (Wu et al., 2014; Otadi, 2014; Xue et al., 2014; Hu, 2013; Wu et al., 2008).

There are various feature selection and classification approaches based on the particle swarm optimization algorithm (Zhang et al., 2009) combined with the Hamming distance or Gaussian sampling (Banka and Dara, 2015; Xue et al., 2014; Huang et al., 2008). Modified SVM approaches have also been proposed to achieve better fit and predictive accuracy than traditional SVMs (Zhou and Gan, 2007). In the fuzzy set theory literature, a fuzzy rough set approach that combines rough sets and fuzzy set theory has been proposed for pattern recognition (Derrac et al., 2012). It is proposed to use fuzzy entropy-based feature selection combined with similarity classifier (Luukka, 2011) and hybrid system for feature selection based on a binary improved gravitational search algorithm and k-NN method (Xiang et al., 2015).

There have been many studies concerning the application of a fuzzy neural network (FNN). The FNNs that combine neural network (Taormina et al., 2015; Cheng et al., 2005; Chau 2007) and fuzzy set theory have also been proposed for learning, adaptation and rule extraction (Wu et al., 2014; Otadi, 2014). The FNN

*E-mail addresses:* shleedosa@anyang.ac.kr, shleedosa@gmail.com

http://dx.doi.org/10.1016/j.engappai.2015.08.003
0952-1976/

introduce not only the computational power of neural networks into fuzzy systems but also the human-like thinking and reasoning of fuzzy systems into neural networks. A neural network with a weighted fuzzy membership function (NEWFM) is a supervised classification neuro-fuzzy system that uses the bounded sum of weighted fuzzy membership functions (BSWFM). After the training process of the NEWFM is completed, all features are interpretably formed into weighted fuzzy membership functions preserving the disjunctive fuzzy information and characteristics and all feature differences are illustrated by the graphical characteristics of all BSWFMs.

In this study, we propose a feature selection combined BSWFM with the Euclidean distance for decreasing the computational load and improving accuracy by removing irrelevant features. This enables the selection of minimal features with the highest classification performance. The steps to select these features are as follows. All BSWFMs generated through the learning process are normalized and their centers of gravity are determined. The greater the distance between the BSWFM centers of gravity, the more the features are differentiated into good features with good ranking. Therefore, features with short distances between their centers are removed one-by-one to improve classification performance. As a result, a minimal set of features with the highest classification performance is selected. In this study, we compare the classification performances of NEWFM without and with feature selection based on the BSWFM and Euclidean distance. In this study, McNemar's test is employed to determine whether the difference between the performances of the two classification algorithms, without feature selection and with feature selection, is statistically significant. However, it shows that the performances obtained by minimum features with feature selection are not significantly different from that obtained by initial features without feature selection on McNemar's test.

The remainder of this study is organized as follows. In Section 2, we review the experimental data used in this study. In Section 3, we describe the structure of the NEWFM, the learning process of NEWFM, and the BSWFM. In Section 4, we describe how the NEWFM selects minimum features using BSWFM and Euclidean distance. In Section 5, we analyze the experimental results of the feature selection algorithms proposed in this study. Finally, the discussion and conclusion are presented in Section 6.

## 2. Experimental data

In this study, the proposed feature selection algorithm was applied to data provided by the University of California Irving

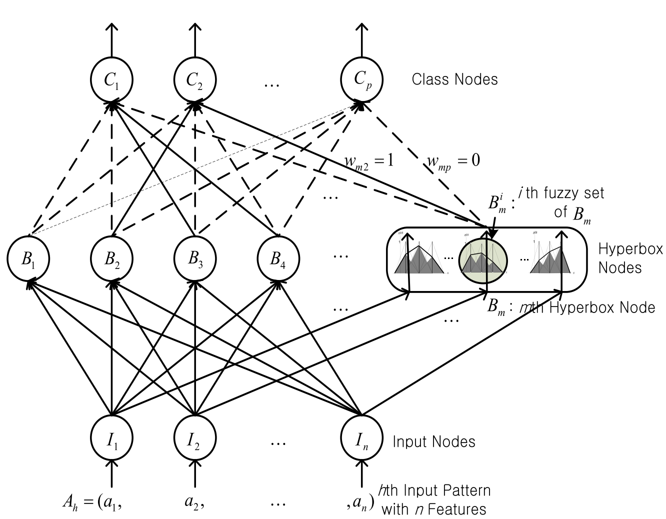


**Fig. 1.** NEWFM structure.

(UCI) Machine Learning Repository. To show the usefulness of the proposed feature selection algorithm, it was necessary to conduct extensive experiments with benchmark problems, such as UCI datasets. It is noteworthy that one advantage of the proposed feature selection algorithm is that it is simple enough to implement as a computer program without any statistical assumptions.

The datasets consisted of the “Statlog Heart”, “Parkinsons” and “Ionosphere” datasets that contain 270 instances of 13 features, 195 instances of 22 features and 351 instances of 34 features, respectively, with no missing values. The variable to be predicted in the Statlog Heart dataset is the absence or presence of heart disease. The Parkinsons dataset is composed of a range of biomedical voice measurements from 31 people, 23 of whom have Parkinson's disease (PD). Each column in the dataset is a particular voice measure and each row corresponds to a voice recording of an individual (identified in the “name” column). The main aim of the dataset is to discriminate healthy people from those with PD. The radar data in the Ionosphere dataset was collected by a system in Goose Bay, Labrador that consists of a phased array of 16 high-frequency antennas with a total transmitted power on the order of 6.4 kW. The targets were free electrons in the ionosphere. “Good” radar returns are those showing evidence of some type of structure in the ionosphere. “Bad” returns are those that do not; their signals pass directly through the ionosphere.

## 3. Neural network with weighted fuzzy membership functions (NEWFM)

NEWFM is a supervised classification neuro-fuzzy system that uses BSWFMs. The structure of the NEWFM, illustrated in Fig. 1, comprises three layers, namely, the input, hyperbox and class layers. The input layer contains $n$ input nodes for an $n$-featured input pattern. The hyperbox layer consists of $m$ hyperbox nodes. Each hyperbox node $B_l$ that is connected to a class node contains $n$ BSWFMs for $n$ input nodes. The output layer is composed of $p$ class nodes. Each class node is connected to one or more hyperbox nodes. The $h$th input pattern can be recorded as $I_h=\{A_h=(a_1, a_2, \ldots, a_n), class\}$, where *class* is the result of classification and $A_h$ comprises the $n$ features of an input pattern.

As shown in Fig. 2, the weight and center of each membership function are adjusted during the learning process, e.g., $W_1$, $W_2$, and $W_3$ are moved down, $v_1$ and $v_2$ are moved toward $a_i$, and $v_3$ remains in the same location. After learning, each of the $n$ fuzzy sets in hyperbox node $B_l$ contains three *weighted fuzzy membership functions* (*WFMs*), as indicated by the gray membership functions in Fig. 3. The *bounded sum* (a fuzzy set operation) of the WFMs (BSWFM) in the $i$th fuzzy set of$B_l^i(x)$, is denoted as $\mu_b^i(x)$ (indicated by the bold line in Fig. 3) and is defined as:

$$\mu_b^i(x) = \sum_{j=1}^{3} B_l^i(\mu_j(x)). \tag{1}$$

## 4. Proposed feature selection method for NEWFM

This study selects features using BSWFMs (an example is shown in Fig. 3) in order to improve classification performance using minimal features. Fig. 4 shows a BSWFM for two classes (A and B) that were generated during learning. The features shown in Fig. 4 have two BSWFMs, and they were obtained from the training process of the NEWFM. After the training process of the NEWFM is completed, all features are interpretably formed into weighted fuzzy membership functions preserving the disjunctive fuzzy information and characteristics. The two BSWFMs graphically illustrate the feature differences between class A and class B.