Journal of King Saud University – Computer and Information Sciences (2018) 30, 498–509

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

HRNeuro-fuzzy: Adapting neuro-fuzzy classifier for recurring concept drift of evolving data streams using rough set theory and holoentropy

Jagannath E. Nalavade [a,*], **T. Senthil Murugan** [b]

[a] *Research Scholar, Veltech Dr. RR & Dr. SR Technical University, Avadi, Chennai, India*
[b] *Department of Computer Science Engineering, Veltech Dr. RR & Dr. SR Technical University, Avadi, Chennai, India*

Received 26 March 2016; revised 1 November 2016; accepted 13 November 2016
Available online 21 November 2016

KEYWORDS

Data stream;
Neuro fuzzy;
Change of detection;
Rough set theory;
Holoentropy function

Abstract Data stream classification plays a vital role in data mining techniques which extracts the most important patterns from the real world database. Nowadays, many applications like sensor network, video surveillance and network traffic generate a huge amount of data streams. Due to the ambiguity in input data, imprecise input information and concept drift, some problems arise in classifying the data stream. To resolve these problems, we propose a HRNeuro fuzzy system in this paper based on rough set theory and holoentropy function. At first, the input database is given to the PCA algorithm to reduce the dimension of the data. An adaptive neuro fuzzy classifier is utilized where the designing of membership function and rule base are the two important aspects. Then, neuro-fuzzy system undergoes updating when the change of detection occurs between the data streams. Here, the updating behaviour of membership function and rules are performed using rough set theory and holoentropy function. The experimental results are evaluated for the datasets and the performance is analysed by some metrics and compared with the existing systems such as JIT adaptive K-NN and HRFuzzy system. From the result, it is concluded that our proposed fuzzy classifier attains the higher accuracy of 96% which proves the efficient performance of data stream classification algorithm.

1. Introduction

Data stream mining is defined as the extraction of structural information from the continuous data streams represented by models and patterns (Mena-Torres and Aguilar-Ruiz, 2014; Masud et al., 2011; Wang et al., 2003; Read and Bifet, 2012). Data stream classification is one of the significant tasks in data stream mining which has been utilized in spam filtering, real-time intrusion detection and malicious website monitoring

* Corresponding author.
E-mail address: jen20074u@gmail.com (J.E. Nalavade).
Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2016.11.005
1319-1578

(Zhang et al., 2015). Generally, data classification is the technique of the learning classifier which differentiates and describes the data classes. Then, the learning classifier could forecast the class of entities without knowing the value of class label (Ghosh and Biswas, 2014). Two approaches for classifying the data streams are single classifier based approach and ensemble based approach. Single classifier approach builds a model or pattern from the small set of data stream and then updates the model using machine learning incrementally. Some of the learning techniques are artificial neural network, Fuzzy, decision trees, instance based learning etc (Mena-Torres and Aguilar-Ruiz, 2014; Rutkowski et al., 2014). Ensemble based approach is the combination of individual base models in some manner to form an ensemble classifier of data stream. This ensemble classifier attains better accuracy and is easier to scale when compared with single classifier based approach (Bifet et al., 2009).

Nowadays, data mining is being challenged by the real time system which provides a huge amount of data at non-pattern rates. Examples of such data streams are telephone call records, surveillance video streams, credit card transaction flows and sensoring and network event logs (Ghosh and Biswas, 2014). Here, the two major problems are infinite length and concept drift in data stream classification algorithms. An infinite length is defined by the data stream which is the phenomenon fast and continuous (Masud et al., 2013). Then, the concept drift occurs when the underlying data stream may progress over time and its distribution may change subsequently. Examples of real life concept drifts are financial fraud detection, spam categorization, monitoring systems, evolving customer preferences and weather predictions (Brzezinski and Stefanowski, 2014). The imperfection of processed data, unavailability of data, imprecise and uncertainty in data is a major difficulty in any real applications. Thus, the streaming data classification algorithm is employed to deal with the concept drift and classifying the uncertain data stream by the machine learning methods (Han et al., 2015).

Data stream classification is done by the two step procedure. Firstly, a classifier is built by a predefined set of data classes and data concepts. During the training phase, the classifier is constructed using a classification algorithm and also by learning from the dataset tuples and their label attributes. This process is named as machine learning. Secondly, the classifier is used to analyse the performance of the dataset which is independent of the training predetermined data. This step is known as supervised learning (Patil et al., 2010). The fuzzy system is one of the most frequently used tasks for classification which is employed to describe the feature space and maintain fuzzy region with fuzzy rules. In classical set theory, the fuzzy set theory is easy to tackle the different aspects of uncertain data and imprecise data. The features are associated with the degree of membership function in the fuzzy classifiers which is determined using the fuzzy rule-based classification systems. The fuzzification approach is also utilized in this classification for mapping the fuzzy set by the input data in the data stream classification (Mitrakis et al., 2008; Nowak and Nowicki, 2014; Azar and Hassanien, 2014). Commonly, fuzzy based schemes are very effective as they are effectively applied for detecting the intrusion in networking environments (Zadeh, 1992; Snasel et al., 2010; Shojafar et al., 2016; Sh et al., 2014). Also, the neuro-fuzzy systems were effectively used for various real time scenarios. Accordingly, wind wake estimation was done in Shamshirbanda et al. (2014) and Nikolić et al. (2015) and anomaly detection was done in Singh et al. (2016), Moshtaghi et al. (2015) using neuro-fuzzy methods.

The main objective of this paper is to classify the data stream based on the class labels of previous data streams using the proposed adaptive neuro fuzzy classifier. The membership function and the fuzzy rule base are the two significant steps in the neuro fuzzy system. Initially, the dimension of the input database is mitigated using the Principal Component Analysis. Then, the database is divided into a number of data chunks which are used to build the neuro fuzzy system. Then, the fuzzy classifier classifies the first data stream based on the previous data stream by utilizing the rule base and membership function. Secondly, the previous data stream is requires updating when the concept change is detected between the current and previous data stream. Thus, the rough set theory is employed for updating the membership function and then the holoentropy based method is utilized for updating the fuzzy rules. After updating the fuzzy system, the fuzzy classifier classifies the current data stream based on the previous data stream.

The main contributions of this paper are:

- Due to the large amount of data streams, the Principal Component Analysis (PCA) is utilized in this paper for reducing the dimension of data.
- An adaptive neuro fuzzy classifier is proposed to classify the data streams based on the membership function and fuzzy rule base.

The paper is organized as follows. Section 2 presents the literature review of data stream classification by the machine learning method. Section 3 describes the motivation behind the approach by problem statement and challenges. Section 4 explains the proposed methodology of neuro fuzzy classifier to the data stream classification. Section 5 discusses the experimental results and the comparative performance analysis. Finally, this paper is concluded in Section 6.

2. Literature review

Mena-Torres and Aguilar-Ruiz (2014) introduced a technique, named Similarity-based Data Stream Classifier (SimC), which achieved good performance by introducing an insertion or removal policy that adapts quickly to the data tendency and maintained a representative, small set of examples and estimators that guaranteed good classification rates. This methodology was also able to detect novel classes or labels, during the running phase, and to remove useless ones that did not add any value to the classification process. Statistical tests were used to evaluate the model performance, from two points of view such as efficacy (classification rate) and efficiency (online response time). Five well-known techniques and sixteen data streams were compared, using Friedman's test. Also, to find out which schemes were significantly different, Nemenyi's, Holm's and Shaffer's tests were considered. The results showed that SimC was very competitive in terms of (absolute and streaming) accuracy, and classification or updating time, in comparison to most popular methods in the literature. It had the advantage of using Instance based learning techniques but had the drawback to refine the model to detect and deal with only abrupt concept changes.

Alippi et al. (2013) considered a just-in-time strategy for adaptation; the sensing unit reacts exactly when needed, i.e., when concept drift was detected. Change detection tests (CDTs), designed to inspect structural changes in industrial and environmental data, were coupled here with adaptive k-nearest neighbour and support vector machine classifiers, and suitably retrained when the change was detected. Computational complexity and memory requirements of the CDT and the classifier, due to precious limited resources in embedded sensing, were taken into account in the application design. We showed that a hierarchical CDT coupled with an adaptive resource-aware classifier was a suitable tool for processing and classifying sequential streams of data. KNN and SVM detect the change drift exactly but it was more suitable for numerical data.

Zhang et al. (2015) proposed an Ensemble-tree (E-tree for short) indexing structure to organize all base classifiers in an ensemble for fast prediction. E-trees treated ensembles as spatial databases and employed an R-tree like height-balanced structure to reduce the expected prediction time from linear to sub-linear complexity. On the other hand, E-trees could be automatically updated by continuously integrating new classifiers and discarding outdated ones, well adapting to new trends and patterns underneath data streams. Theoretical analysis and empirical studies on both synthetic and real-world data streams demonstrated the performance of this approach. However, the E-tree requires high storage space and maintenance of tree.

Brzezinski and Stefanowski (2014) presented a data stream classifier, called the Accuracy Updated Ensemble (AUE2), which aimed at reacting equally well to different types of drift. AUE2 combined the accuracy-based weighting mechanisms known from block-based ensembles with the incremental nature of Hoeffding Trees. This algorithm is experimentally compared with 11 state-of-the-art stream methods, including single classifiers, block-based and online ensembles, and hybrid approaches in different drift scenarios. Out of all the compared algorithms, AUE2 provided best average classification accuracy while proving to be less memory consuming than other ensemble approaches. Experimental results showed that AUE2 could be considered suitable for scenarios, involving many types of drifts as well as static environments. The AUE2 classifier was easy to consider the periodic weighting mechanism but difficult to adapt the weight for different data space.

Mitrakis and Theocharis (2012) explained an efficient structure learning algorithm for the development of self-organizing neuro-fuzzy multilayered classifiers (SONeFMUC). These classifiers were hierarchical structures comprising small-scale fuzzy-neuron classifiers (FNCs), interconnected along multiple layers. The SONeFMUC structure was progressively expanded by generating layers based on the principles of the Group Method of Data Handling (GMDH) algorithm, which was appropriately adapted to handle classification tasks. Traditional GMDH proceeds blindly to the construction of all possible parent FNC pairs from the previous layer to obtain the individuals in the next layer without paying due attention to the diversity of the FNC combinations. Thus, a modified version of GMDH was devised for effective identification of the SONeFMUC structure. We incorporate the Proportion of Specific Agreement (Ps) to evaluate the diversity of the FNC pairs. In the devised method, only complementary FNCs were combined, i.e., FNCs which committed errors at different pattern subspaces, to construct a descendant FNC at the next layer. Accordingly, a computational reduction was achieved while high classification accuracy was maintained. The efficiency of the proposed structure learning was tested on a diverse set of benchmark datasets using land cover classification from multispectral images as a real-world application. It exploited the higher accuracy of ambiguous pattern and more complexity.

Nowak and Nowicki (2014) described the architecture of a neuro-fuzzy classifier with fuzzy rough sets which had been developed to process imprecise data. A raw output of such system was an interval which had to be interpreted in terms of classification afterward. To obtain a credible answer, the interval should be as narrow as possible; however, its width could not be zero as long as input values were imprecise. They discussed the determination of classifier parameters using the standard gradient learning technique. The effectiveness of the proposed method was confirmed by several simulation experiments. The neuro fuzzy classifier was used to mitigate the rate of improperly classified samples.

Snasel et al. (2010) proposed a genetic programming to evolve a fuzzy classifier in the form of a fuzzy search expression to predict product quality. They interpreted the data mining task as a fuzzy information retrieval problem and applied a successful information retrieval method for search query optimization to the fuzzy classifier evolution. They demonstrated the ability of the genetic programming to evolve useful fuzzy classifiers on two use cases in which they detected the faulty products of a product processing plant and discovered the intrusions in a computer network. The fuzzy classifier was used to describe the defective products only.

Azar and Hassanien (2014) presented a linguistic hedges neuro-fuzzy classifier with selected features (LHNFCSF) for dimensionality reduction, feature selection and classification. Four real-world data sets were provided to demonstrate the performance of the neuro-fuzzy classifier. This classifier was compared with the other classifiers for different classification problems. The results indicate that applying LHNFCSF not only reduced the dimensions of the problem, but also improved the classification performance by discarding the redundant, noise-corrupted, or unimportant features. The results strongly suggested that this method not only helps in reducing the dimensionality of large data sets but also could speed up the computation time of a learning algorithm and simplify the classification tasks. The recognition rate was decreased while increasing the number of fuzzy rules of each class.

3. Motivation

3.1. Problem statement

Generally, the learning of classifiers cannot be performed with full data stream to classify the data because the class labels of previous data stream only are known. Let us assume that the input database, I is partitioned into a chunk of data sample d_t having a size of s, $I = \{d_t; 0 \leqslant t \leqslant s\}$. At the current time, the data can be read out and we can perform the classification. Thus, each data contains v number of attribute vector a_j, $(d_t \in a_j; 1 \leqslant j \leqslant v$. The fuzzy classifier is used to handle the symbolic and qualitative data. Here, the main challenge is to classify the data in the current data stream using the neuro fuzzy classifier which handles also the numeric and quantitative data. Then, the classification is performed based on the known class labels of previous data stream.

3.2. Challenges

Normally, the data stream is defined as continuous and infinite and also the data are created with high rates. Thus, different challenges arise in these surroundings such as mining, querying and storage (Mena-Torres and Aguilar-Ruiz, 2014).

The challenge here is to determine the concept drift in the real world data stream classification because the data are generating continuously such as web logs, sensor networks, business transactions etc (Cao and Huang, 2013).

The other significant challenges in data stream are recurring changes in concept, far less attention, feature space evolution and integration of context information (Gomes et al., 2014). The evolving of data stream is more challengeable when the input data are imperfect, unavailable, uncertain and imprecise input information.

4. Proposed methodology: adapting neuro-fuzzy classifier for evolving data streams using rough set theory and holoentropy

The main objective of this paper is to classify the data streams based on the class labels of previous data stream by using the adaptive neuro fuzzy system. Fig. 1 represents the block diagram of the proposed methodology. Initially, the input database is fed into the Principal Component Analysis (PCA) system to reduce the dimension of the input. Then, the database is divided into chunks of data which are utilized for building the neuro fuzzy system. The membership function and fuzzy rule are the two significant aspects to build the fuzzy system for the first data stream. Then, the membership function and fuzzy rules are required to update in the neuro fuzzy system when the change of detection occurs between the current and previous data stream. The rough set theory is employed in this paper for updating the membership function and holoentropy is used to update the fuzzy rules in the neuro fuzzy system. Finally, updating of the neuro fuzzy system based on the previous data stream is used to classify the new data stream.

4.1. Principal Component analysis

Initially, the input database I consists of s number of datasets which is given as input to the Principal Component Analysis (PCA) algorithm for the dimension reduction. PCA (Ilin and Raiko, 2010) is an essential tool to mitigate the dimension of the data. PCA is utilized in this paper to extract the relevant information from the database and reduce the complex data set to a lower dimension. It has the advantage over other analyses in noise reduction, less memory requirement, lack of redundancy and low complexity. The PCA analysis is derived by,

(i) Each dataset I_i in the database I is computed by the mean value using Eq. (1).

$$M = \frac{1}{s}\sum_{i=0}^{s} I_i \tag{1}$$

(ii) Then, subtract the mean value from each dataset which is stored in the variable φ_i and is given by Eq. (2).

$$\phi_i = I_i - M \quad \text{where,} \quad 0 \leqslant i \leqslant s \tag{2}$$

(iii) Calculate the covariance matrix. Covariance calculations are used to find the relationships between the dimensions in high dimensional datasets. Thus, the matrix is expressed using PCA (Ilin and Raiko, 2010) algorithm and is given in Eq. (3).

$$C = \frac{1}{s}\sum_{i=0}^{s} \phi_i \phi_i^T \tag{3}$$

where, ϕ_i^T is the transpose of the matrix.

(iv) The Eigen vectors and Eigen values of the covariance matrix are computed by Eq. (4).

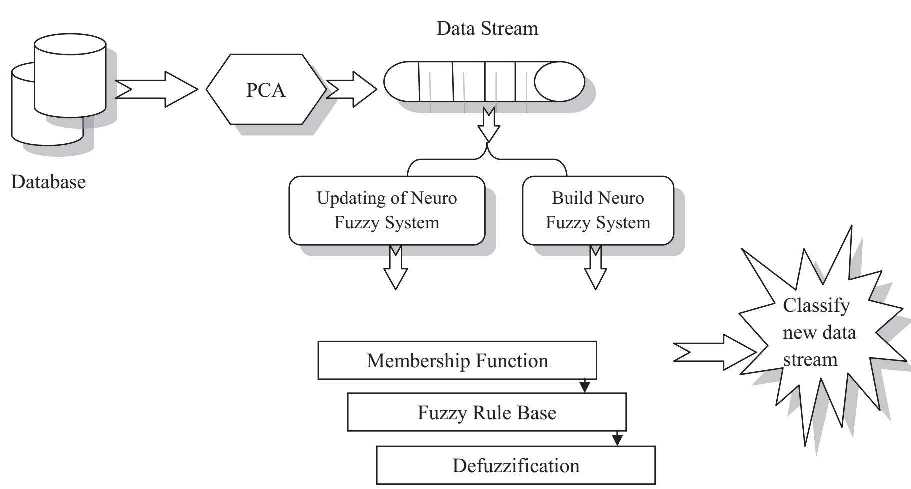


Figure 1 Block diagram of the proposed methodology.

$$u_i = \sum_{i=0}^{s} v_i \phi_i \tag{4}$$

where, u_i is the Eigen vector and v_i is the Eigen vector of ϕ_i^T. Then, the Eigen values w_i are represented with respect to the Eigen vector.

$$w_i = u_i^T (I_i - M) \tag{5}$$

From the above equations, we can observe that the Eigen values are quite different. Thus, to reduce the dimension, the Eigen vector with the highest Eigen value is considered as the principal component of the data set. Once Eigen vectors are formed, they are sorted by the Eigen values and the feature vector generated. The feature vector F is given by Eq. (6).

$$F = (w_1, w_2, \ldots, w_n) \tag{6}$$

where, n Eigen vectors correspond to the n largest Eigen values.

(v) Once the principal components are chosen, the final data or the new data are expressed by the multiplication of feature vector and mean value of the dataset.

$$D = F^T * M^T \tag{7}$$

Thus, the dimension reduction of database D is obtained by the PCA (Ilin and Raiko, 2010) analysis. Then, the database D is given as the input to the neuro fuzzy for building the fuzzy system.

4.2. Building neuro fuzzy system

The database D is obtained using PCA. The database D is composed of s number of data chunk which is defined by $D = d_t;\ 0 \leqslant t \leqslant s$. The fuzzy system is built by the current data stream based on the previous data stream. The classifier classifies the current data stream d_t based on the previous data stream d_{t-1} since we know only the class label of previous data stream in the current time. Subsequently, the fuzzy classifier is to be constructed with the help of membership function and the fuzzy rules.

4.2.1. Adaptive neuro-fuzzy inference system (ANFIS) architecture

The ANFIS is a class of adaptive networks which are functionally equivalent based on the concepts of fuzzy set theory, fuzzy if-then rules and fuzzy reasoning. The ANFIS (Jang, 1993) architecture represents both the Sugeno and Tsukamoto fuzzy models. The fuzzy inference system is constructed with the two attributes a and b of the first data stream d_0. The rule base contains the fuzzy if-then rules of Takagi and Sugeno's type fuzzy as follows. For a first order two rule Sugeno fuzzy inference system, the two rules may be stated as:

Rule 1: If a is P_1 and b is Q_1 then $c_1 = x_1a + y_1b + z_1$
Rule 2: If a is P_2 and b is Q_2 then $c_2 = x_2a + y_2b + z_2$

These are the two fuzzy rules which are employed in the adaptive neuro fuzzy system. Fig. 2 shows the architecture of ANFIS which contains five layers. Thus, the layers of adaptive neuro fuzzy system are described as below.

Layer 1: This layer is known as fuzzification layer which defines the membership grades for each attributes and depends on the chosen fuzzy membership function. Then, every node i in the fuzzification layer is constructed with the node function which is represented by Eq. (8)

$$L_i^1 = \delta P_i(a) \tag{8}$$

where, a is input attribute to node i, P_i is the linguistic variable associated with the node function and δP_i is the bell shaped membership function which is expressed by Eq. (9).

$$\delta P_i(a) = \frac{1}{1 + \left[\left(\frac{a - z_i}{x_i}\right)^2\right]^{y_i}} \tag{9}$$

where, $\{x_i, y_i, z_i\}$ is the premise parameter set. The parameter z determines the centre of the bell shaped membership function, the parameter x denotes the half-width and y (together with x) is defined to control the slopes at the different crossover points. The change of these parameters exhibits various forms of membership function by the linguistic variable. Thus, the generalized bell shaped membership function is employed with three membership function x, y, z to define the characteristics of every attribute.

In ANFIS (Jang, 1993) architecture, the membership function in utilized in this layer which can be indicated as, $\delta_{jk}^{(0)}$ where, j refers to the index of the attribute, k refers to the index of membership function. The bell shaped membership function of the jth attributes of ith membership function is defined in Eq. (10).

$$\delta_{jk}^{(0)} = mf(d_0^{ij}; x, y, z) \tag{10}$$

Then, the bell shaped membership function is derived by Eq. (11).

$$mf(d_0^{ij}; x, y, z) = \frac{1}{1 + \left(\frac{d_0^{ij} - z}{x}\right)^{2y}} \tag{11}$$

By modifying the values of premise parameter set x, y, z, we can obtain our desired membership function which provides more flexibility for classification.

Layer 2: In this layer, the node is fixed to determine the firing strength f_i of the rule. Thus, the output of each node in this layer is calculated by the product of membership function of attributes as given in Eq. (12).

$$L_i^2 = f_i = \delta P_i(a) \times \delta Q_i(a) \tag{12}$$

where, δP_i and δQ_i are the membership functions of two linguistic variables P and Q.

Layer 3: Each node in this layer is used to calculate the firing strength which is defined by the ratio of ith rule's firing strength to the sum of all rules firing strength by Eq. (13).

$$L_i^3 = \overline{f_i} = \frac{f_i}{f_1 + f_2} \tag{13}$$

where, $\overline{f_i}$ is defined as the normalized firing strengths in layer 3.

Layer 4: This layer is constructed with the node function as an adaptive function along with the inputs. The output from this layer is expressed by Eq. (14).

$$L_i^4 = \overline{f_i} c_i = \overline{f_i}(p_i a + q_i b + r_i) \tag{14}$$

where, $\{p_i, q_i, r_i\}$ is the parameter set. Parameters in this layer are termed as the consequent parameter set.

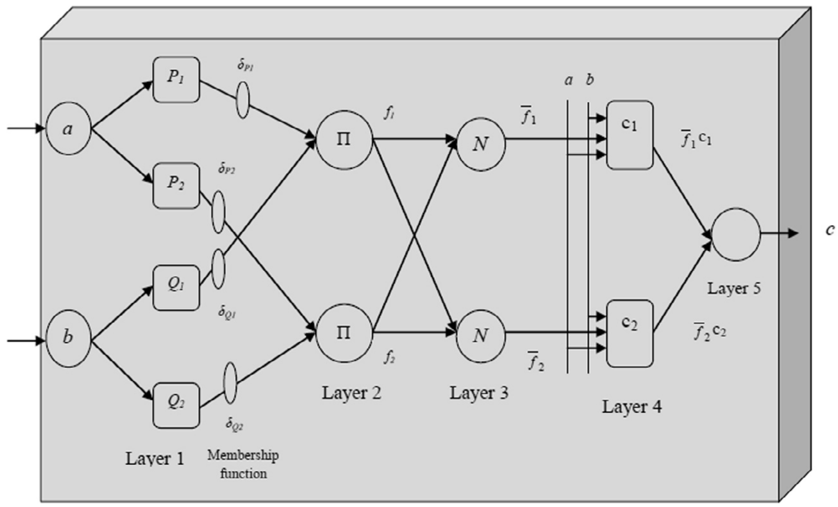


Figure 2 Architecture of ANFIS system.

Layer 5: The output layer consists of single node which computes the final output or overall output as the summation of all input and it is estimated by Eq. (15).

$$L_i^5 = \sum_i \bar{f}_i c_i = \frac{\sum_i f_i c_i}{\sum_i f_i} \tag{15}$$

Thus, the ANFIS system contains two parameter sets such as premise parameter and consequent parameter. During the learning process, these parameters are tuned until the desired response of this system is achieved. When the premise parameters are fixed, then the output of the ANFIS (Jang, 1993) can be expressed by the linear combination of consequent parameters.

$$c = (\bar{f}_1 a)p_1 + (\bar{f}_1 b)q_1 + (\bar{f}_1)r_1 + (\bar{f}_2 a)p_2 + (\bar{f}_2 b)q_2 + (\bar{f}_2)r_2 \tag{16}$$

4.2.2. Data classification using fuzzy system

The neuro fuzzy system is defined by the membership function and fuzzy rule base. Thus, the devised neuro fuzzy system can be represented as $F_{d_0}\{\delta_{jk}^{(i)}, R_0\}$. Then, the input data are fed into the devised neuro fuzzy system for classification. This fuzzy system generates a score value which can be then used for classifying the data labels.

4.3. Updating the neuro fuzzy system

After the neuro fuzzy system is built, the updating behaviour is required to classify the data in each data stream d_t based on the class labels of previous data stream d_{t-1}. Due to the evolving nature of data streaming information, the change of detection is an essential data mining process. Thus, the neuro fuzzy system undergoes updating of the previous data stream when the concept change occurs. Thus, the rough set theory and holoentropy function are employed to update the previous data streams.

4.3.1. Change of detection by rough set theory

The change of detection for the previous data stream is computed by the rough set theory. The rough set theory (Pawlak, 2002) is used to examine the data set and approximations. Normally, the data set consists of objects and attributes which are utilized by the rough set theory. In this theory, the set Y belongs to the object and A belongs to the attributes in the data set. Then, the approximation set is defined as the one which consists of equivalence relation of data objects. Then, the equivalence relation is referred as the elementary sets or granules g. Thus, the lower and upper approximations of the set Y are given below.

(i) The lower approximation set of a set Y, with regard to A is the union of all granules included in the set, which is defined in Eq. (17).

$$\underline{A}Y = \{g|\,[g]_A \leqslant \mathrm{Y}\} \tag{17}$$

The upper approximation set of a set Y, with regard to A is the union of all granules that have a non-empty intersection with the set, which is defined in Eq. (18).

$$\overline{A}Y = \{g|\,[g]_A \cap Y \neq \varphi\} \tag{18}$$

where, g is the granules or elementary sets.

(ii) The boundary region (Pawlak, 2002) A of the set Y is given by Eq. (19).

$$AN_A(Y) = \overline{A}Y - \underline{A}Y \tag{19}$$

If the boundary region of Y is the empty set, that is, $AN_A(Y) = \varphi$, then the set Y is crisp with respect to A. Whereas, when the boundary region is non-empty, i.e., $AN_A(Y) \neq \varphi$, then the set Y is rough with respect to A.

(iii) The accuracy of approximation is computed based on the lower and upper approximation set of the data which is expressed by Eq. (20)

$$\alpha_A(Y) = \frac{|\underline{A}Y|}{|\overline{A}Y|} \tag{20}$$

Then, the accuracy of approximation is compared with a threshold, called T. If the accuracy of approximation $\alpha_A(Y)$ is less than the threshold value, then there is a concept change which requires updating the membership function.

4.3.2. Updating the membership function

After computing the concept drift using the rough set theory, the parameters of membership function in the previous data stream needs to be updated. The previous data stream should be updated when there are following two limitations. i) Firstly, the concept change is detected between the current and previous data stream when the accuracy of approximation is less than the threshold value and ii) Secondly, the membership function has different values of parameters $x_{t-1}, y_{t-1}, z_{t-1}$ in the previous data stream d_{t-1}. Thus, the membership function for the previous data stream is expressed in Eq. (21),

$$\delta_{jk}^{t-1} = mf(d_{t-1}, x_{t-1}, y_{t-1}, z_{t-1}) \tag{21}$$

The values of x_{t-1}, y_{t-1} and z_{t-1} can be computed by comparing the previous and current variables of membership function.

4.3.3. Updating fuzzy rules by holoentropy function

Once the membership function is updated, the fuzzy rules are then updated in the neuro fuzzy system for the previous data stream d_{t-1} based on the class label. Thus, the previous data stream d_{t-1} is given to the subtractive clustering to update the fuzzy rules by using the holoentropy function. The holoentropy function is employed to update the fuzzy rule with the rules of previous data streams R_j^{t-1} and R_j^{t-2}. The weight of the rules and holoentropy are the prerequisites to update the fuzzy rule base.

The holoentropy combines all the rules and their weights are used to find which rules are to be replaced. After knowing the rules of previous data stream, the holoentropy function is applied to the previous data streams. Thus, the fuzzy rule is updated by the following two steps such as (i) evaluate the holoentropy for previous fuzzy rule R_j^{t-1} and (ii) estimate the conditional holoentropy for the previous fuzzy rules R_j^{t-1} and R_j^{t-2}.

(i) Firstly, the holoentropy is evaluated using the weights and entropy function for the fuzzy rule R_j^{t-1} which is given in Eq. (22).

$$h(R_j^{t-1}) = w \times e(R_j^{t-1}) \tag{22}$$

where, w is the weight of fuzzy rule R_j^{t-1} and e defines the entropy of the rule. Thus, the weight and entropy are calculated by Eqs. (23) and (24).

$$w = 2\left(1 - \frac{1}{1+\exp(-e(R_j^{t-1}))}\right) \tag{23}$$

$$e(R_j^{t-1}) = -\sum_{i=1}^{n(R_j^{t-1})} X_i \times \log X_i \tag{24}$$

where, $n(R_j^{t-1})$ is the number of unique values in the rule vector R_j^{t-1} and X_i is the probability

(ii) Secondly, the conditional holoentropy is evaluated by the probability and holoentropy of two previous fuzzy rules R_j^{t-1} and R_j^{t-2} which is expressed in Eq. (25).

$$Ch(R_j^{t-2}, R_j^{t-1}) = \sum_{i=0}^{n(R_j^{t-1})} X_i \times h(R_j^{t-2}, R_j^{t-1}) \tag{25}$$

where, h is the holoentropy function of both previous fuzzy rules and X_i is the probability. Thus, the holoentropy is defined by the weight and entropy of rules R_j^{t-1} and R_j^{t-2}.

$$h(R_j^{t-2}, R_j^{t-1}) = w_h * e(R_j^{t-2}, R_j^{t-1}) \tag{26}$$

where, wh is the weight of the two fuzzy rules for holoentropy function and e is the entropy function. Thus, the weight and entropy are estimated by Eqs. (27) and (28).

$$w_h = 2\left(1 - \frac{1}{1+\exp(-e(R_j^{t-2}, R_j^{t-1}))}\right) \tag{27}$$

$$e(R_j^{t-2}, R_j^{t-1}) = \sum_{i=1}^{n(R_j^{t-2})} X(R_j^{t-2} = i, R_j^{t-1} = i) \times \log X(R_j^{t-2} = i, R_j^{t-1} = j) \tag{28}$$

Finally, the fuzzy rules are updated in the neuro fuzzy system based on the holoentropy and conditional holoentropy function. The holoentropy is utilized among all the combination of rules from two previous data streams. It is represented in Eq. (29).

$$Ih(R_j^{t-2}, R_j^{t-1}) = h(R_j^{t-1}) - Ch(R_j^{t-1}, R_j^{t-2}) \tag{29}$$

where, $Ih(R_j^{t-2}, R_j^{t-1})$ defines the informative entropy, $h(R_j^{t-1})$ is the entropy function and $Ch(R_j^{t-1}, R_j^{t-2})$ is the conditional entropy for the previous data streams.

Thus, the fuzzy rules are updated using informative holoentropy which utilizes holoentropy and conditional holoentropy function. The holoentropy is computed by weighting the entropy of the rules of previous data stream. Then, the conditional holoentropy is also a weighted property of conditional entropy which is computed by finding the conditional probability between the fuzzy rules R_j^{t-1} and R_j^{t-2}. Finally, the weight of the rules determines which rules are to be selected for updating the fuzzy rules.

$$R_j^{t-1} = \begin{cases} R_j^{t-2}; & w(R_j^{t-1}) < w(R_j^{t-2}) \\ R_j^{t-1}; & else \end{cases} \tag{30}$$

By updating the membership function and rules using rough set theory and holoentropy function, the neuro fuzzy system is modified in the current data stream for classification of data labels. Thus, the updated neuro fuzzy system is represented as $F\{\mu_{jk}^{t-1}, R^{t-1}\}$. The neuro fuzzy system should be updated for the every new data stream classification based on the previous data stream class labels. Table 1 shows the algorithmic description of HRNeuro-fuzzy.

5. Results and discussion

This section describes the experimental results of the proposed adaptive neuro fuzzy system and the performance is analysed and compared with the existing systems.

Table 1 Algorithmic description of HRNeuro-Fuzzy.

Algorithm: HRNeuro-Fuzzy
Initialization:
  d_t = Input data
  T = COD threshold
Output:
  $F_{d_{t-1}}(\delta^{t-1}, R^{t-1})$ = Updated membership function
Procedure
Begin
  Apply PCA into an input database
  Divide database into a chunk of data
    For t = 1:s
      Read d_t
      If(t = = 1)
        Build neuro fuzzy system $F_{d_0}(\delta_0, R_0)$ using d_0
      Endif
      If(t! = 1)
      Find accuracy of approximation $\alpha_A(Y)$ using rough set theory
      If $(\alpha_A(Y) < T)$
      Update neuro fuzzy system to $F_{d_{t-1}}(\delta^{t-1}, R^{t-1})$
    Endif
    End if
    Perform classification of d_t using $F_{d_{t-1}}(\delta^{t-1}, R^{t-1})$
    Return neuro fuzzy system for classification, $F_{d_{t-1}}(\delta^{t-1}, R^{t-1})$
  End for
End

5.1. Experimental results

The experimental results of our proposed neuro fuzzy system to evolving data stream classification are implemented using MATLAB software. The results are calculated and discussed below.

(a) *Database description:* We utilized three different databases such as Skin Segmentation Data Set and Localization Data for our experimentation which are taken from UC Irvine Machine Learning (Irvine Machi, 0000).

Skin Segmentation Data Set (db3): The skin dataset is collected by randomly sampling R,G,B values from face images of various age groups (young, middle, and old), race groups (white, black, and asian), and genders obtained from FERET database and PAL database. The total number of instances in these data is 245057; out of which 50859 are the skin samples and 194198 are non-skin samples.

Localization Data for Person Activity Data Set (db2): This database is collected from the people who were used for recording the data by wearing four tags (ankle left, ankle right, belt and chest). Each instance is a localization data for one of the tags. The tag can be identified by one of the attributes. The total number of instances is 164860.

Breast Cancer Data (db1): This database is generated by the chronological grouping of data. This dataset contains 286 instances; in these 201 instances corresponds to one class and 85 instances belong to another class. The instances are described by nine attributes which are either linear or nominal

(b) *Evaluation Metrics:* The performance of proposed adaptive neuro fuzzy classifier is analysed by the parameters such as accuracy, precision, recall, F-measure and computational time. Then, the analysed performance is compared with existing systems like HRFuzzy and JIT adaptive k-NN. The evaluation metrics are described as follows.

Accuracy: Accuracy is used to determine the degree of closeness of measurements of a quantity to that quantity true's value. Thus, the accuracy is expressed in Eq. (31).

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{31}$$

where, TP is true positive which is correctly identified, TN is true negative which is correctly rejected, FP determines incorrectly identified which is named as false positive and similarly FN is a false negative which denotes incorrectly rejected.

Precision: The precision is determined by the closeness of two or more measurements to each other. Precision is also defined as the positive predictive value. It is expressed in Eq. (32).

$$\text{Precision } (p) = \frac{TP}{TP + FP} \tag{32}$$

Recall: The recall is also known as the true positive rate or sensitivity which is used to evaluate the fraction of relevant instances that are retrieved. The recall is given in Eq. (33).

$$\text{Recall}(r) = \frac{TP}{TP + FN} \tag{33}$$

F measure: It is a measure of test's accuracy. It considers both the precision p and recall r of the test to compute the measure which is represented in Eq. (34).

$$Fmeasure = \frac{2pr}{p + r} \tag{34}$$

Computation time: The computation time is the length of time required to perform a computational process by calculating running time of the classifier. Hence, the computation time is measured using the 'tic' and 'toc' variables in the matlab.

(c) *Experimental database:* We created the three databases for our experimentation to classify the data stream. Let db_1 be the database which includes the collection of breast cancer data set and db_2 is defined as the collection of dataset which consists of localization data for person activity. The database db_3 is generated by the collection of skin segmentation data set.

(d) Classifiers taken for comparison:

JIT adaptive K-NN: The operational framework onto which JIT classifiers for stationary conditions operate is that of asymptotical learning and converges when the training set composed of number of independent and identically distributed pair increases. Then, the k -Nearest Neighbours is employed in JIT classification systems due to the absence of a proper training phase and an easy management of the classifier complexity. Thus, the k-NN based JIT can be made adaptive and automatic through a suitable management of its knowledge base.

HRFuzzy: The HRFuzzy system is used for classification of data stream which is based on the holoentropy function and rough set theory. In this classifier, we employed k-means clustering method to generate the fuzzy rules and updating the fuzzy system when concept change occurs between the current and previous data stream.

HRNeuro Fuzzy: An adaptive neuro fuzzy classifier is proposed to enhance the accuracy for better classification of data stream. The neuro fuzzy classifier is built by the bell shaped membership function and the fuzzy rules. Then, updating behaviour of neuro fuzzy system is required for classification based on the previous data stream which is done by holoentropy function and using the weight function of fuzzy rules.

5.2. Sample fuzzy rules

The sample rules generated from the HRNeuro-fuzzy classifier are given in Table 2. Accordingly, the set of rules derived from db1 is: If (Age = = Low) OR (menopause = = Low) ⇒ (Decision = Low): If (Age = = High) AND (Tumour size = = HIGH) AND (menopause = = HIGH) ⇒ (Decision = High). These rules only played a significant role to achieve the better accuracy of the proposed system for db1 database. From the rules, we can identify that the proposed system makes use of only three attributes to find the presence of the disease. Other attributes are assumed here as "don't care' attributes. Similarly, fuzzy rules derived for other two databases are also given in Table 2.

5.3. Performance analysis

This section demonstrates the performance evaluation of the proposed HRNeuro-fuzzy classifier. The data in the data stream are classified based on the known class labels of previous data stream which is analysed by the metrics such as accuracy, precision, recall, F- measure and time. The performance is also compared with the existing systems such as JIT adaptive K-NN and HRFuzzy classifier.

Table 2 Sample fuzzy rules.

Database	Sample rules
Skin Segmentation Data Set (db3)	If (R value = = Low) AND (G value = = High) ⇒ (Decision = Low) If (G value = = Low) AND (B value = = Low) ⇒ (Decision = High)
Localization Data for Person Activity Data Set (db2):	If (X coordinate = = High) AND (Y coordinate = = High) ⇒ (Decision = Low) If (Y coordinate = = Low) OR (X coordinate = = Low) ⇒ (Decision = High)
Breast Cancer Data (db1)	If (Age = = Low) OR (menopause = = Low) ⇒ (Decision = Low) If (Age = = High) AND (Tumour size = = HIGH) AND (menopause = = HIGH) ⇒ (Decision = High)

5.3.1. Analysis using accuracy

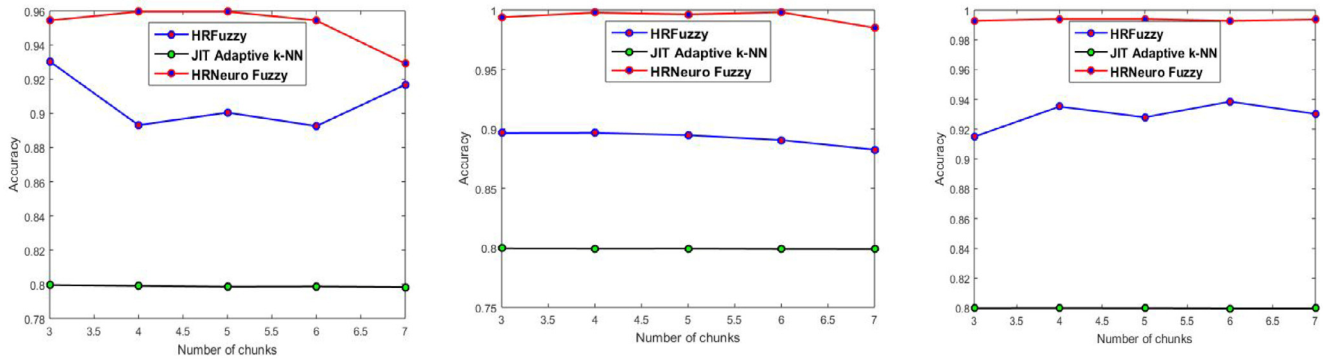
The performance analysis based on accuracy is shown in Fig. 3. The accuracy is defined as the measure of quality or state of being correct or precise. Here, the accuracy is analysed by the number of chunks which is obtained from the PCA. Fig. 3a shows the accuracy performance and db_1 is the database of breast cancer dataset. When the number of chunk is four, the existing system like JIT adaptive K-NN classifier achieves the accuracy of 80% and HRFuzzy classifier attains the 89%. But, our proposed HRNeuro fuzzy classifier enhances the accuracy up to 96% which is shown in Fig. 3a. Then, db_2 is the database of localization dataset for person activity. The existing JIT classifier has only the 80% of accuracy value even when the number of chunks is to be varied. Then, the performance is gradually mitigated while increasing the number of chunks in the existing HRFuzzy system. However, our proposed classifier attains the maximum accuracy when compared with the existing systems which is demonstrated in Fig. 3b. The dataset of skin segmentation is collected in the database db_3. The Fig. 3c depicts the performance analysis of database db_3. The accuracy is analysed while increasing the number of chunks. The proposed system attains the higher accuracy percentage when compared with the existing systems such as JIT adaptive K-NN and HRFuzzy system and is shown in Fig. 3c. Thus, the proposed classifier ensures better classification of data stream based on the previous data stream.

5.3.2. Analysis by precision

The performance analysis of proposed system using the precision parameter is depicted in Fig. 4. The Fig. 4a shows the precision analysis of the database db_1. The precision is greatly improved to 97% by the proposed HRNeuro fuzzy system rather than the existing systems such as JIT K-NN and HRFuzzy classifier. The number of chunks is increased which tends to increase the precision performance which is depicted in Fig. 4a. Then, the precision analysis for the localization dataset for person activity is shown in Fig. 4b. When the number of chunk is three, the K-NN method acquires 49% accuracy value. Similarly, existing HRFuzzy system also achieves the minimum accuracy of 69%. But, compared to existing systems, our proposed classifier attains the maximum accuracy for the localization dataset as shown in Fig. 4b. When the number of chunk is six, the proposed classifier achieves the maximum precision of 99.2% for the skin segmentation dataset db_3. Subsequently, the precision value of 99.35% is obtained for chunk size seven, which is shown in Fig. 4c. Thus, our proposed system attains the maximum accuracy and precision value for evolving the data stream.

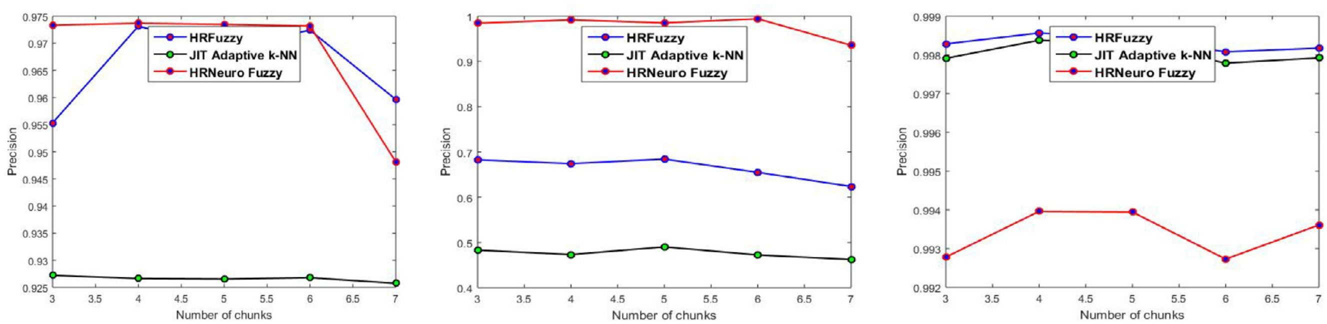
5.3.3. Analysis using recall

Fig. 5 demonstrates the performance analysis using the recall metric. The recall is also known as the sensitivity which defined as the ratio of number of true positives and the total number of elements that actually belong to the positive class. Fig. 5a shows the recall performance of the breast cancer database db1. 80% recall for JIT adaptive K-NN classifier and 88% for HRFuzzy system is obtained using four numbers of data chunk. Our proposed system achieves the higher recall value of 97% compared to the existing systems which is plotted in Fig. 5a. For the localization dataset db2, the recall value of



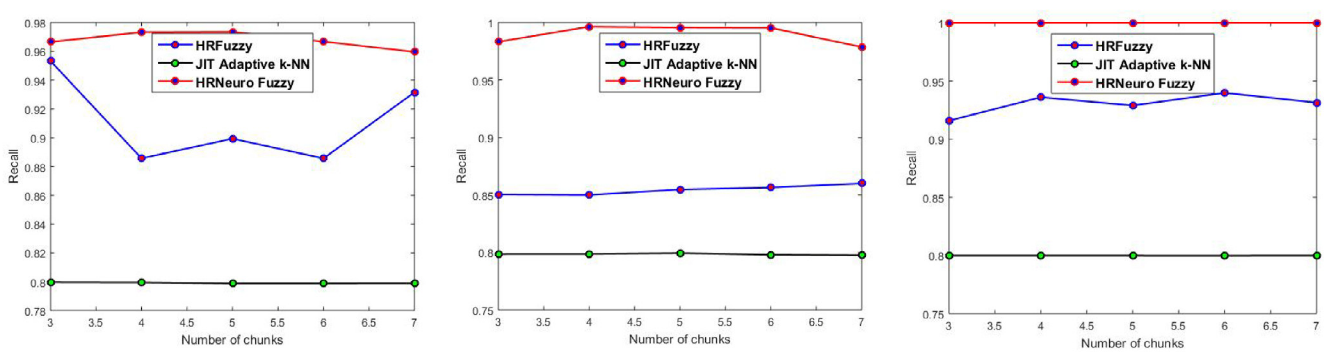
a. Accuracy-db_1 b. Accuracy-db_2 c. Accuracy-db_3

Figure 3 (a) Accuracy-db_1, (b) Accuracy-db_2, (c) Accuracy-db_3.



a Precision-db_1 b. Precision-db_2 c. Precision-db_3

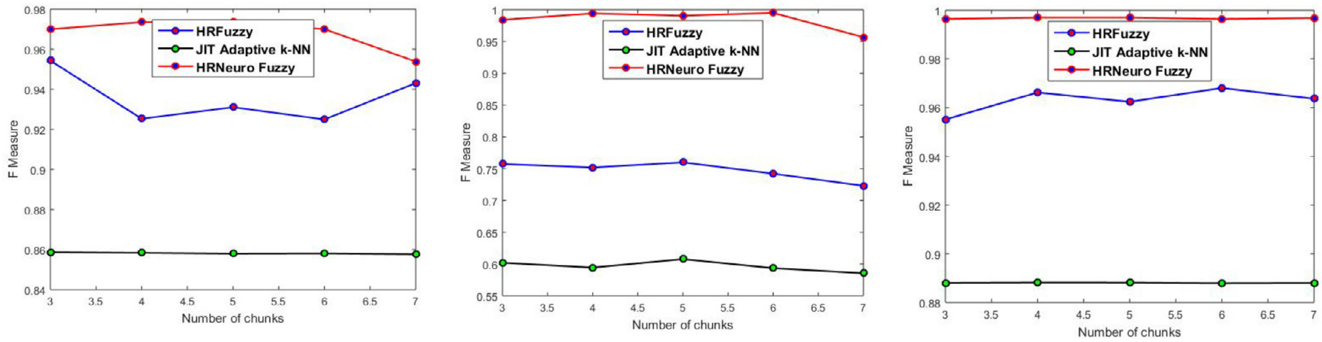
Figure 4 (a) Precision-db_1, (b) Precision-db_2, (c) Precision-db_3.



a Recall-db_1 b. Recall-db_2 c. Recall-db_3

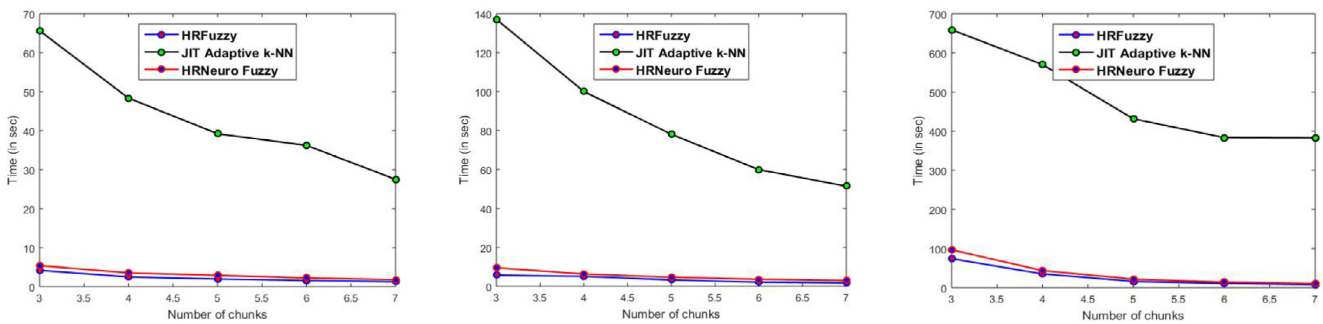
Figure 5 (a) Recall-db_1, (b) Recall-db_2, (c) Recall-db_3.

JIT adaptive K-NN classifier has the same value for all number of data chunks. Then, the HRFuzzy system classifier gradually increases the value to 86% by increasing the number of chunks. Thus, in Fig. 5b, our proposed system also enhances their performance moderately and attains the maximum value of 98%. Similarly, the performance of the skin segmentation data set is depicted in Fig. 5c. The proposed fuzzy classifier achieves the maximum recall value while compared with the existing classifiers such as JIT adaptive K-NN and HRFuzzy system. 80% of recall value is obtained by the existing K-NN system and also 93% of recall value is acquired by the HRFuzzy system. By comparing with the existing method, our proposed system attains the higher performance for data stream classification which is shown in Fig. 5c.



a F measure-db_1 b. F measure-db_2 c. F measure-db_3

Figure 6 (a) F measure-db_1, (b) F measure-db_2, (c) F measure-db_3.



a Computation time - db_1 b. Computation time - db_2 c. Computation time - db_3

Figure 7 (a) Computation time -db_1, (b) Computation time-db_2, (c) Computation time-db_3.

5.3.4. Analysis by F-measure

The performance analysis of F-measure for the three databases is shown in Fig. 6. Generally, the F-measure or F-score is the measurement of test's accuracy in statistical analysis which considers both the precision and recall. Thus, the F-measure is defined as the harmonic mean of precision and recall. Higher F-measure is achieved for the database db1 which is depicted in Fig. 6a. When the number of chunk is five, the K-NN system has 86% of F-measure value and the HRFuzzy system achieves 93% of F-measure. Simultaneously, the maximum F-measure value of 97% is obtained by the proposed neuro fuzzy classifier which is demonstrated in Fig. 6a. In Fig. 6b, using the existing K-NN classifier, the performance is gradually increased to 76%. When compared to this K-NN system, our neuro fuzzy classifier for the database db_2 attains the maximum value of 98%. The F-measure is analysed for the skin segmentation dataset and is shown in Fig. 6c. The F-measure for the proposed neuro fuzzy classifier is improved than the existing methods such as JIT adaptive K-NN and HRFuzzy classifier. i.e., 99% F-measure is obtained by the neuro fuzzy classifier and is shown in Fig. 6c

5.3.5. Analysis using computation time

The comparative analysis of the computation time is shown in Fig. 7. Computational time is employed to determine the length of running time. Fig. 7a shows the performance analysis of the breast cancer database db1. The existing system has the higher computation time for classifying the data stream by using the number of data chunk. But, the proposed classifier attains the minimum computation time for data stream classification rather than the existing systems which is depicted in Fig. 7a. The K-NN classifier has the computation time of 140sec which is gradually decreased to 50sec. While compared with the existing systems, we can infer from the Fig. 7b, that the proposed method of neuro fuzzy classifier attains the minimum value of two seconds for the localization dataset for person activity db2. The computation time for the skin segmentation dataset is shown in Fig. 7c. The proposed system has the computation time of 100sec to classify the data stream when the number of chunk is three. Then, by increasing the number of data chunk, the computation time for our proposed method is moderately reduced as shown in Fig. 7c. Thus, the neuro fuzzy system achieves the minimum computation time rather than the existing systems which ensures the less computational complexity of the proposed method for data stream classification.

6. Conclusion

We proposed an adaptive neuro fuzzy system using rough set theory and holoentropy function for evolving the data streams. The HRNeuro Fuzzy system was utilized to classify the data streams based on the class labels of previous data stream. Initially, the input database was given to the Principal Component Analysis for dimension reduction. Then, the database was divided into data chunks which were employed to build the neuro fuzzy system by the bell shaped membership function. The concept change was detected using the accuracy of approximation which was compared with the threshold value. Thus, the rough set theory was used in this paper to estimate the accuracy of approximation of the data set. Then, the updating behaviour of neuro fuzzy system was required to update the membership function and fuzzy rules. Finally, the data stream was classified based on the previous data stream after updating the fuzzy system. The experimental results of the proposed system were evaluated for the breast cancer dataset, localization dataset and skin segmentation dataset. Then, the performance of these datasets were analysed using the metrics such as precision, recall, F-measure, computation time and accuracy. The higher accuracy of 96% was acquired by the proposed neuro fuzzy system for the better classification of data stream.

References

Alippi, Cesare, Liu, Derong, Zhao, Dongbin, Li, Bu, 2013. Detecting and reacting to changes in sensing units: the active classifier case. IEEE Trans. Syst. Man Cybern. Syst. 44 (3), 353–362.

Azar, Ahmad Taher, Hassanien, Aboul Ella, 2014. Dimensionality reduction of medical big data using neural-fuzzy classifier. Soft Comput.

Bifet, A., Holmes, G., Pfahringer, B., Kirkby, R., Gavalda, R., 2009. New ensemble methods for evolving data streams. In: Proc. 15th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD).

Brzezinski, Dariusz., Stefanowski, Jerzy., 2014. Reacting to different types of concept drift: the accuracy updated ensemble algorithm. IEEE Trans. Neural Networks Learn. Syst. 25 (1), 81–94.

Cao, Fuyuan, Huang, Joshua Zhexue, 2013. A concept-drifting detection algorithm for categorical evolving data. Comput. Intell. Chin. Inf. Process., 485–496

Ghosh, Soumadip., Biswas, Sushanta., Sarkar, Debasree, Sarkar, Partha Pratim, 2014. A novel Neuro-fuzzy classification technique for data mining. Egypt. Inf. J.

Gomes, Joao Bartolo, Gaber, Mohamed Medhat, Sousa, Pedro A.C., Menasalvas, Ernestina, 2014. Mining recurring concepts in a dynamic feature space. IEEE Trans. Neural Networks Learn. Syst. 25 (1), 95–110.

Han, Dong.-Hong., Zhang, Xin., Wang, Guo.-Ren., 2015. Classifying uncertain and evolving data streams with distributed extreme learning machine. J. Comput. Sci. Technol. 30, 874–887.

Ilin, Alexander, Raiko, Tapani, 2010. Practical approaches to principal component analysis in the presence of missing values. J. Mach. Learn. Res. 11, 1957–2000.

Jang, Jyh-Shing Roger, 1993. ANFIS: adaptive network based fuzzy inference system. IEEE Trans. Syst. Cybern. 23 (3).

Masud, M., Gao, J., Khan, L., Han, J., Thuraisingham, B., 2011. Classification and novel class detection in concept-drifting data streams under time constraints. IEEE Trans. Knowl. Data Eng. 23 (6), 859–874.

Masud, Mohammad M., Chen, Qing, Khan, Latifur, Aggarwal, Charu C., Gao, Jing, Han, Jiawei, Srivastava, Ashok, Oza, Nikunj C., 2013. Classification and adaptive novel class detection of feature-evolving data streams. IEEE Trans. Knowl. Data Eng. 25 (7), 1484–1497.

Mena-Torres, Dayrelis, Aguilar-Ruiz, Jesus S., 2014. A similarity-based approach for data stream classification. Expert Syst. Appl. 41, 4224–4234.

Mitrakis, N.E., Theocharis, J.B., 2012. A diversity-driven structure learning algorithm for building hierarchical neuro-fuzzy classifiers. Inf. Sci. 186, 40–58.

Mitrakis, N.E., Theocharis, J.B., Petridis, V., 2008. A multilayered neuro-fuzzy classifier with self-organizing properties. Fuzzy Sets Syst., 3132–3159

Moshtaghi, M., Bezdek, James C., Leckie, Christopher, 2015. Evolving fuzzy rules for anomaly detection in data streams. IEEE Trans. Fuzzy Syst. 23 (3), 688–700.

Nikolić, V., Shamshirband, Shahaboddin, Petković, Dalibor, Mohammadi, Kasra, Ćojbašić, Žarko, Altameem, Torki A., Gani, Abdullah, 2015. Wind wake influence estimation on energy production of wind farm by adaptive neuro-fuzzy methodology. Energy 80, 361–372.

Nowak, Bartosz A., Nowicki, Robert K, 2014. The learning of neuro-fuzzy classifier with fuzzy rough sets for imprecise datasets. ICAISC, 256–266.

Patil, Dipti D., Wadhai, V.M., Gokhale, J.A., 2010. Evaluation of decision tree pruning algorithms for complexity and classification accuracy. Int. J. Comput. Appl. 11 (2), 0975–8887.

Pawlak, Zdzisław, 2002. Rough set theory and its applications. J. Telecommun. Inf. Technol. 3.

Read, Jesse., Bifet, Albert., 2012. Scalable and efficient multi-label classification for evolving data streams. Mach. Learn. 88, 243–272.

Rutkowski, Leszek, Jaworski, Maciej, Pietruczuk, Lena, Duda, Piotr, 2014. Decision trees for mining data streams based on the gaussian approximation. IEEE Trans. Knowl. Data Eng. 26 (1), 108–119.

Sh, Shahaboddin, Patel, Ahmed, Anuar, Nor Badrul, Kiah, Miss Laiha Mat, Abraham, Ajith, 2014. Cooperative game theoretic approach using fuzzy Q-learning for detecting and preventing intrusions in wireless sensor networks. Eng. Appl. Artif. Intell. 32, 228–241.

Shamshirbanda, Shahaboddin, Petković, Dalibor, Hashimc, Roslan, Motamedid, Shervin, Anuare, Nor Badrul, 2014. An appraisal of wind turbine wake models by adaptive neuro-fuzzy methodology. Int. J. Electr. Power Energy Syst. 63, 618–624.

Shojafar, M., Nicola, Cordeschi, Enzo, Baccarelli, 2016. Energy-efficient adaptive resource management for real-time vehicular cloud services. IEEE Trans. Cloud Comput., 99

Singh, Heisnam Rohen, Biswas, S.O., Purkayastha, Biswajit, 2016. A neuro-fuzzy classification technique using dynamic clustering and GSS rule generation.

Snasel, Vaclav, Kromer, Pavel, Platos, Jan, 2010. The evolution of fuzzy classifier for data mining with applications. LNCB, 349–358.

UC Irvine Machine Learning Repository from http://archive.ics.uci.edu/ml/datasets.html.

Wang, H., Fan, W., Yu, P.S., Han, J., 2003. Mining concept-drifting data streams using ensemble classifiers. In: Proc. ACM SIGKDD Ninth Int'l Conf. Knowledge Discovery and Data Mining, pp. 226–235.

Zadeh, L., 1992. Fuzzy sets. In Fuzzy models for pattern recognition: Methods that search for structures in data. IEEE Press, NY.

Zhang, Peng., Zhou, Chuan., Wang, Peng., Gao, Byron J., Zhu, Xingquan., Guo, Li., 2015. E-Tree: an efficient indexing structure for ensemble models on data streams. IEEE Trans. Knowl. Data Eng. 27 (2), 461–474.