CHAPTER 3

SYSTEM FRAMEWORK

The generic framework of the temporal data mining system used in this research work is shown in Figure 3.1. The framework comprises of two subsystems namely temporal data pre-processing and classification. Temporal data pre-processing aims at improving the quality of data for the classification process. Temporal classification builds the classification model for time series data. For experimentation this work uses the following clinical time series datasets: Hepatitis, Thrombosis and Parkinson's disease. Detailed descriptions of these datasets are provided in the chapter 4.

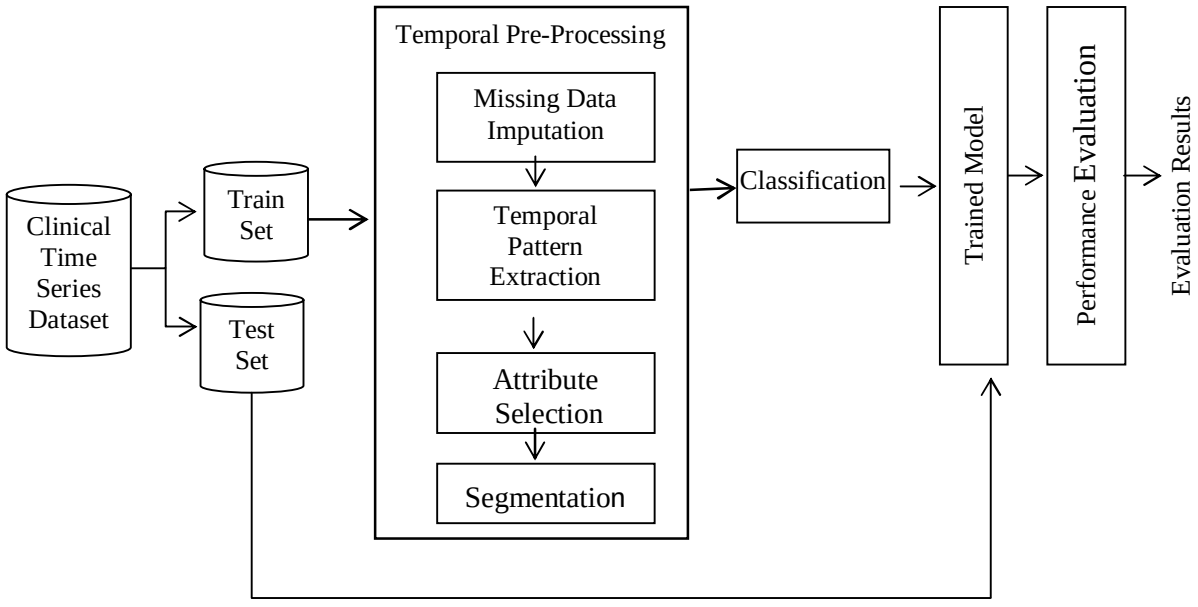

Figure 3.1 Generic Framework of the Temporal Data Mining System for Clinical Time Series Data

3.1 TEMPORAL DATA PRE-PROCESSING

The activities performed by temporal data pre-processing subsystem are missing value imputation, temporal pattern extraction, attribute selection and segmentation.

3.1.1 Missing Data Imputation

Two common ways of handling missing values in any datasets are either ignoring the missing records or imputing the missing values. Missing value imputation in clinical time series data aims at replacing a missing value with a computed value, without introducing bias. The following two approaches have been used in this research work for handling missing values: First, forecast model approach and Second, an enhanced interpolation method. In the first approach, the concept of double exponential smoothing is adopted and enhanced to construct a forecasting model. This model is used to extract the forecast value for imputing an unobserved data. In the second approach, the concept of inverse distance weight interpolation is adopted and enhanced to interpolate the missing value.

3.1.2 Temporal Pattern Extraction

Temporal pattern describes the temporal behavior of each attribute over a period of observations. This work uses the trend pattern and state of each clinical attribute to interpret its temporal behavior. The trend refers to the overall growth rate of an attribute and state represents the range or mean value of an attribute. For example, if a patient's cholesterol level observed for five days is continuously increasing and is above the normal range then the trend pattern for that clinical examination is said to be increasing and the state is considered to be high. This work has adopted the concept of double exponential smoothing method to derive the growth rate (trend) pattern and mean value (state) for each clinical examination.

3.1.3 Attribute Selection

High dimensionality is one of the challenging characteristics of clinical time series data set which normally contains many insignificant attributes. The presence of insignificant attributes in the mining process increases the number of irrelevant rules thereby reducing the prediction accuracy of the decision making system. The attribute selection process identifies the significance of each clinical attribute and selects the relevant attribute set for mining process. This work adopts the concept of rough set and tolerance rough set for performing attribute selection. Rough Set concepts such as attribute selection, attribute extraction, data reduction, handling missing attribute value, pattern extraction and rule generation can be incorporated effectively in different phases of data analysis (Pradipta & ParthaGarai 2013; Pawlak 1982; Komorowski 1999; Zhong & Skowron 2001).

A detailed study about various attribute selection techniques proposed for classification tasks is discussed by Dash & Liu (1997). Attribute selection algorithms are classified into three categories namely filter approach, wrapper approach and embedded approach based on the evaluation procedure (Jensen et al. 2007). In filter based approach, the algorithm performs selection independently without any learning algorithm. In wrapper based approach a learning algorithm is used for attribute selection. Several research works that illustrates the importance of using rough set in attribute selection as pre-processing in knowledge discovery have been discussed in many literatures (Pawlak 1982; Komorowski 1999; Zhong & Skowron 2001; Dash & Liu 1997).

3.1.4 Segmentation

Clinical time series data often contains more than hundred observations for a single patient. Time series analysis on such data becomes

complex due to this huge set of observation. Segmentation process divides the longer time series into smaller sequences of segments. This process tends to reduce the complexity of time series analysis. Segmentation is considered as a pre-processing step in many time series analysis (Batal et al. 2013; Lin et al. 2007; Lovric et al. 2014). The following are the segmentation algorithmic strategies based on Piecewise Linear Representation (PLR): bottom-up, top-down and sliding window algorithm (Lovric et al. 2014; Keogh et al. 2004). This work performs segmentation using a bottom-up segmentation approach.

3.2 TEMPORAL CLASSIFICATION

Temporal classification process builds classification model with the significant attributes and its temporal patterns (Moskovitch & Shahar 2015b; Fu 2011). Clinical decision support system is normally recommended based upon the consistency and prediction accuracy of the constructed classification model. The following techniques are used in this research study for constructing a classification model: neural network, neuro-fuzzy, decision tree and time delay neural network. Neural network is a widely adopted classification technique that has been used in many real-world applications for pattern recognition, forecasting and prediction (Han & Kamber 2001). Neural networks represent a brain metaphor for information processing. These models are biologically inspired rather than an exact replica of how the brain actually functions (Hagan 1996, Jang 1996). An Artificial Neural Network (ANN) is composed of many artificial neurons that are linked together according to specific network architecture.

There are several types of neural network configurations namely Single-layer feed forward network, Multilayer Perceptrons (MLP), Radial Basis Function (RBF) network, wavelet neural network, self-organizing maps, neuro-fuzzy etc (Hagan 1996; Jang,1996). ANN is typically defined by three types of parameters: First, interconnection pattern between the different layers

of neurons; Second, the learning process for updating weights of the interconnections; Third, activation function that converts a neuron's weighted input to its output activation. Two main considerations are required before performing classification using neural network. First, choice of a network model depends on the data representation and the application. Second, learning algorithms are used to train the neural network. Hence selecting and tuning an algorithm for training on unseen data requires a significant amount of experimentation.

Back propagation Leven-Marquadt (LM) has been a widely used neural network learning algorithm. Backpropagation learns by iteratively processing a data set of training tuples, comparing the network's prediction for each tuple with the actual known target value. Neuro-fuzzy system is a hybridization technique that combines ANN and fuzzy logic. The structure of a neuro-fuzzy system has input and output layers and three hidden layers that represent membership functions and fuzzy rules (Zadeh 1965, Jang 1996). Time Delay Neural Networks (TDNNs) is a feed forward networks, the only difference is that the tap delay line associated with each input (Waibel 1989). The tap delay allows the network to have a dynamic finite response to time series data.

Decision tree is a classification technique that constructs a classification model based on the target output. Quinlan (1986) has proposed a decision tree induction algorithm ID3 which selects the test attributes at each node using information gain as splitting criteria. The main limitation of ID3 is that it does not handle missing values and numeric attributes. Later, limitation of ID3 was addressed in C4.5 proposed by Quinlan (1993) with gain ratio as splitting criteria. It handles numeric or missing values and performs error-based pruning. The usage of continuous attribute in C4.5 was

further enhanced by Quinlan (1996) using minimum description length principle to improve the efficiency of handling continuous attributes.

Classification and Regression Trees (CART) proposed by Breiman et al. (1984) uses index for constructing binary split for each attribute and cost-complexity pruning techniques for tree pruning. Chi-Squared-Automatic-Interaction-Detection (CHIAD) developed by statistical developers (Kass 1980) was initially designed to handle nominal attributes. It handles missing values but does not perform pruning. The Quick Unbiased Efficient Statistical Tree (QUEST) proposed by Loh & Shih (1997) generates binary decision tree using univariate and linear combination splits. The main difference among these algorithms is in the splitting criteria they choose for identifying test attribute during the decision tree construction.

3.3 PERFORMANCE EVALUATION

Finally, the performance of the constructed classification model was evaluated and the results are compared with the existing state of art methods. The K-fold cross validation method has been used to generate the training and test set. The obtained classification accuracy of the proposed systems proves its efficiency in clinical decision making. To evaluate the constructed classification model confusion matrix has been generated (Han & Kamber 2001). The following are the metrics considered for evaluating the performance of the classifier: True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), accuracy, error rate, sensitivity, specificity and precision. The Positive (P) class represents the class that is of interest and Negative (N) class refers to all other classes.

TP represents the records correctly classified as P by the classifier. TN represents the records correctly classified as N by the classifier. FP represents the records incorrectly classified as P by the classifier and FN

represents the records incorrectly classified as N. Accuracy also known as recognition rate refers to the percentage of the records correctly classified as P and N by the classifier. Accuracy is defined in the Equation (3.1).

$$\text{Accuracy} = (TP + TN)/(P + N) \tag{3.1}$$

Error Rate also known as misclassification rate refers to the percentage of the records incorrectly classified as P and N by the classifier. Error Rate is defined in the Equation (3.2).

$$\text{Error Rate} = (FP + FN)/(P + N) \tag{3.2}$$

Sensitivity also known as true positive rate (TPR) refers to the proportion of correctly classified positive records. It is also referred as recall. Equation (3.3) defines the specificity.

$$\text{Sensitivity} = TP/P \tag{3.3}$$

Specificity also known as true negative rate (TNR) refers to the proportion of correctly classified negative records. Equation (3.4) defines the sensitivity.

$$\text{Specificity} = TN/N \tag{3.4}$$

Precision refers to percentage of record actually classified as P. Equation (3.5) defines the precision.

$$\text{Precision} = TP/(TP + FP) \tag{3.5}$$

The following statistical performance metrics are considered for evaluating the performance of forecasting model and imputation process: Mean Absolute Deviation (MAD), Root Mean Squared Error (RMSE), Mean

Absolute Percentage Error (MAPE), FB (Fractional Bias Error), Index of Agreement (IA). The metrics (Cooray 2008) are defined in the equations (3.6) to (3.10). Let ES_i be the estimated value of i^{th} observation, n is the number of observations and AC_i be the actual value of i^{th} observation.

Mean Absolute Deviation (MAD)

Mean absolute deviation, or MAD is a common average forecast error measurement used in many applications. This measure represents the positive and negative deviations between the forecast and the actual demand. Mathematically, it is represented as,

$$MAD_n = \frac{1}{n}\sum_{i=1}^{n} |AC_i - ES_i| \tag{3.6}$$

Root Mean Squared Error (RMSE)

RMSE measures the average squares of the errors. Mathematically, it is represented as,

$$RMSE_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(AC_i - ES_i)^2} \tag{3.7}$$

Mean Absolute Percentage Error (MAPE)

MAPE is a measure that represents the magnitude of the error relative to the magnitude of the demand. The average ratio is multiplied by 100 to represent this relative measure as a percent. Mathematically, it is represented as,

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{|(AC_i - ES_i)|}{AC_i} * 100 \tag{3.8}$$

Index of Agreement (IA)

It is a measure of correlation between the observed and estimated value. It varies in the range of 0 and 1. A value of 1 indicates strong correlation and a value of 0 indicates no agreement.

$$IA = 1 - \left(\left(\sum_{i=1}^{n}\left(AC_i - ES_i\right)^2\right)/\sum_{i=1}^{n}\left(\left|\left(AC_i - \overline{AC}\right)\right| + \left|\left(ES_i - \overline{AC}\right)\right|\right)^2\right) \quad (3.9)$$

Fractional Bias (FB)

This measure is used to identify the underestimate and overestimate of the predicted value. This value lies in the range -2 to +2 and desired value is 0.

$$FB = \frac{1}{n}\sum_{i=1}^{n}\left(AC_i - ES_i\right)/\left(\left(AC_i + ES_i\right)/2\right) \quad (3.10)$$

The Wilcoxon rank sum test and paired t-test presented by Wilcoxon (1945) and Zimmerman et al. (1997) has been carried out with a significant ρ-value (generally taken as 0.05) to identify whether there was any significant improvement in the classification accuracy of proposed classifier with traditional classifiers such as NN, ID3, SVM and other state of art methods.

In this research work, the first contribution presents a TRiNF mining framework. TRiNF performs temporal prep-processing using a forecast model approach. An enhanced DES method presented by Wright (1986) was adopted to build a forecasting model. The forecasted results have been used in the process of missing value imputation and temporal pattern extraction. The relevant attributes are selected using a temporal pattern induced rough set. TRiNF uses a temporal pattern induced neuro-fuzzy classifier for classifying the unevenly spaced clinical time series data.

The second contribution presents a Q-BTDNN classifier that identifies the severity of Parkinson's disease based on gait disturbances. Q-Backpropagation (Q-BP) training algorithm was proposed to train the Q-BTDNN. Q-BP combines the reinforcement Q-learning approach and backpropagation strategy to adjust the network weights. The constructed classification model has been used to develop a CDMS that helps the physician in diagnosing the severity of PD. The third contribution focuses on imputing the missing values in unevenly spaced clinical time series data. This work has presented an enhancement to the Inverse Weight Distance (IDW) interpolation using rough set and PSO. The rough set and PSO concepts have effectively handled the limitations of IDW in selecting the relevant known data points and its influence factor for imputing the unknown data points.

The fourth contribution presents a STRiD mining framework. A Fuzzy Inference Double Exponential Smoothing (FIDES) method has been proposed to build a forecasting model. The forecasted results have been used in the process of missing value imputation and temporal pattern extraction. The relevant attributes are selected using a temporal pattern induced tolerance rough set, which differs from the first contribution in the generation of equivalence classes. The classification model is built using a decision tree classifier with temporal pattern induced gain ratio as splitting criteria. The fifth contribution focuses on reducing the dimension (size) complexity of clinical time series data using FeAB segmentation. FeAB segmentation adopts bottom-up segmentation strategy and uses a forecast error approximation to compute the merge cost. The segments are then given as input to the TDNN for building a classification model.

The experimental results of the research contributions are evaluated using metrics derived from confusion matrices and statistical performance metrics such as MAD, MAPE, RMSE, IA and FB. The significance tests were carried out using Wilcoxon rank sum test and paired t-test.