Fuzzy Rule Based Feature Extraction and Classification of Time Series Signal

Sandya H. B., Hemanth Kumar P., Himanshi Bhudiraja, Susham K. Rao

Abstract: - Time series signal is a continuous signal which varies continuously with respect to time. These signals involve a great deal of useful information, the information content in these signals can be used for Feature Extraction and Classification. The purpose of Feature Extraction is to reduce the dimension of feature space and achieving better performances. The Features are extracted based on the mathematical calculations like Average, Maximum, Minimum, Standard Deviation and Variance. The Classification of extracted features is carried out by Fuzzy Rule based Selection System. Fuzzy Systems (FS) are evaluated for accuracy, multiplexity, flexibility and transparency for simple and complex systems. In this paper mandani based Fuzzy System is used to achieve accurate results. Based on feature extracted data the Fuzzy System generates a fuzzy score and the Classifier Algorithm classify the feature extracted signals as Good, Bad and Best signals.

Key words: - Fuzzy, Feature Extraction, Classification, Time series signal

I. INTRODUCTION

A Time series is a collection of observations of well-defined data items obtained through repeated measurements over time. Example, measuring the value of retail sales each month will have sales revenue of time series which are measured at every instant of time. Data which are measured randomly are not defined as time series. Therefore, a time series is a sequence number of data collected at regular intervals over a period of time. In statistics, signal processing, econometrics and mathematical finance, a time series is a sequence of data points, measured typically at successive time instants spaced at uniform time intervals. Examples of time series are the daily closing value of the Dow Jones index and the annual flow volume of the Nile River at Aswan. Time series data is analyzed to extract meaningful statistics and other characteristics of the data. The feature values are predicted based on previously observed values using Time series forecasting model. Time series can be decomposed into three components: the trend (long term direction), the seasonal (systematic, calendar related movements) and the irregular (unsystematic, short term fluctuations) [4]. This paper describes Feature Extraction and classification of Time series signals. The Feature Extraction process results in a much smaller and richer set of attributes and can greatly reduce the dimension of feature space without degrading the performance of classifier system [13].

II. FUZZY SYSTEM

Fuzzy is Probabilistic in nature, an Uncertain and an imprecise. Mathematical concepts are very simple in fuzzy reasoning logic system. Fuzzy logic (FL) is a convenient way to map an input space to an output space. FL is flexible, tolerant of imprecise data and is based on natural language and human communication

Fuzzy Inference System (FIS) is a system used to solve new problems. FIS maps an input features to output classes using FL. FIS can be created by hand, using graphical tools or command line functions, or automatically generated using either clustering or adaptive neuro fuzzy techniques. Fuzzy logic is easy to modify by including or excluding rules, there is no need to start a new FIS from the beginning.

FIS is used to solve “Decision Problems” and act accordingly. The structure of Fuzzy Inference System is shown in Fig 1. FIS consists of four modules, Fuzzification module, Knowledge base module, Inference engine module and Defuzzification module. Fuzzy inference methods are classified as direct methods and indirect methods. Direct methods, such as Mamdani’s and Sugeno’s, are the most commonly used. Indirect methods are more complex. Mamdani method is most commonly used fuzzy inference technique. Mamdani model is a knowledge driven predictive model, it works with inputs of crisp data and also with intervals and or linguistic terms. The major advantage of this model is it provides a measure of confidence for predicting future value when the actual value is unknown. The important domain of its application is WEB shopping.

Fig.1. Structure of fuzzy inference system

A. Fuzzy Inference System modules

1) Fuzzification module: The system inputs, which are crisp numbers, are transformed into fuzzy sets. This is done by applying a fuzzification function.

2) Knowledge base module: Stores IF-THEN rules provided by experts.

3) Inference engine module: Using fuzzy inference on the inputs and IF-THEN rules simulates the human reasoning process.

4) Defuzzification module: The fuzzy set obtained by the inference engine transforms into a crisp value.

Manuscript received May, 2013.

Sandya H.B., PG Scholar, Deptt. of ECE, AMC Engineering College, VTU, Bangalore, India.

Hemanth Kumar P., Asst. Prof., Deptt. of ECE, AMC Engineering College, VTU, Bangalore, India.

Himanshi Bhudiraja, Asst. Prof., Deptt. of ECE, AMC Engineering College, VTU, Bangalore, India.

Susham K. Rao, Asst. Prof. Deptt. of ECE, AMC Engineering College, VTU, Bangalore, India.
III. DATA DESCRIPTION

The required dataset for the extraction and classification is generated by using Matlab or Simulink with varying frequency in regular time interval. A constantly varying signal measured for uniform interval of time can be considered as a Time series signal. One best example for Time series signal is a simple sine wave whose frequency is varied with respect to time. The generated signals are similar to original real time signals (Ex: EEG or ECG) with small variations in frequency and amplitude. The data is extracted for all the time intervals for sine waves with different frequencies.

IV. METHODOLOGY

In Data Acquisition, the data samples are generated from time series signals and then Features are extracted based on parametric mathematical calculations like average, maximum, minimum, standard deviation and variance. Finally by using fuzzy Inference System, the output is classified as GOOD, BAD and BEST. The proposed methodology is shown in Fig.2.

A. Time Series Signal Generation

The Time Series Signals are generated using Matlab with frequency ranging from 8Hz to 17.5Hz with total time period of 2 seconds. Generating a time series signals at frequency of every 0.5Hz. A total 20 input data signals S100 to S119 are generated at 8Hz, 8.5Hz, 9Hz, 9.5Hz, 10Hz, 10.5Hz, 11Hz, 11.5Hz, 12Hz, 12.5Hz, 13Hz, 13.5Hz, 14Hz, 14.5Hz, 15Hz, 15.5Hz, 16Hz, 16.5Hz, 17Hz and 17.5Hz. The total time period of 2 seconds is divided into 200 samples by sampling the signal at a rate of 0.01 seconds. The signals S100 to S119 generate 4000 samples. One of the generated time series signals is shown in Fig.3.

B. Data Set Acquisition

Data Set is collected from the input signal as the signal is varying continuously. It has different value of amplitude at different interval of time. The data is collected for all the different input signals generated. A total of 4000 data set samples are collected. The collected data set is shown in Microsoft Excel Sheet in Fig.4.

\[ \text{Total number of samples} = \frac{T}{R} \times n \]

\(T = \text{Total time period} = 2 \text{ seconds} \)
\(R = \text{Sampling rate} = 0.01 \text{ seconds} \)
\(n = \text{Number of input signals} = 20\)

\[ \text{Total number of samples} = \frac{2}{0.01} \times 20 = 4000 \text{ samples.} \]

C. Feature Extraction

The purpose of Feature extraction is to reduce the dimension of feature space and achieving better performances. The features are extracted based on the parameters like Average,
Maximum, Minimum, Standard Deviation and Variance methods. Total 200 features are extracted. The Feature Extraction results as shown in Microsoft Excel Sheet in Fig.5.

1. **Average Method**

   Average is defined as the sum of ‘x’ data sets divided by ‘N’ total number of data sets. It is called the mean average.

   \[
   \text{Average} = \frac{\sum x}{N}
   \]

   Where, \( x \) is data

   \( N \) is total number of datasets

2. **Maximum Method**

   The maximum is defined as the greatest quantity or the greatest value attainable in a given data sets.

   \[
   \text{Max} = \text{maximum}(x_1, x_2, x_3, x_4, \ldots, x_N)
   \]

3. **Minimum Method**

   The Minimum is defined as the lowest quantity or the lowest value attainable in a given data sets.

   \[
   \text{Min} = \text{minimum}(x_1, x_2, x_3, x_4, \ldots, x_N)
   \]

4. **Standard Deviation (\( \sigma \))**

   Standard Deviation is defined as the how much variation or “dispersion” exists from the average (mean, or expected value). A low standard deviation indicates that the data points tend to be very close to the mean; high standard deviation indicates that the data points are spread out over a large range of values.

   \[
   \text{SD}(\sigma) = \sqrt{\frac{\sum(x - \mu)^2}{N}}
   \]

   Where, \( \sigma \) = Symbol of Standard Deviation

   \( \mu \) = mean of all the values in the data set

   \( N \) = Total Number of values in data set

   \( x \) = each value in the data set

5. **Variance (\( \sigma^2 \))**

   Variance is defined as the square of standard deviation or the variance of a random variable or distribution is the expectation, or mean, of the squared deviation of that variable from its expected value or mean. Thus the variance is a measure of the amount of variation of the values of that variable, taking account of all possible values and their probabilities.

   \[
   \text{variance} (\sigma^2) = \text{SD}^2 = \frac{\sum(x - \mu)^2}{N}
   \]

D. **Fuzzy Rule Based Model**

The Extracted parameters Average, Max, Min, Standard Deviation and Variance is considered as input variables to the Fuzzy rule based selection process block. Inference System (FIS) maps an input features to output classes using FL. Fuzzy logic are easy to modify a FIS just by including or excluding rules. The fuzzy rules have written for Extracting Features to get results as Good, Bad and Best data sample values. Fuzzy Rule Based selection for five inputs and one output model is shown in Fig.6.

1. **Fuzzy Rules**

   Fuzzy rules are linguistic IF-THEN- constructions that have the general form "IF A THEN B" where A and B are propositions containing linguistic variables. A is called the premise and B is the consequence of the rule.

   Fuzzy Rule for five input variables and one output variable is defined as few example rules

   a) If Average is bad, Maximum is bad, Minimum is bad, Standard Deviation is bad and Variance is bad, the Fuzzy Classification is considered as “BAD”.

   b) If Average is good, Maximum is good, Minimum is good, Standard Deviation is good and Variance is good, the Fuzzy Classification is considered as “GOOD”.

   c) If Average is best, Maximum is best, Minimum is best, Standard Deviation is best and Variance is best, the Fuzzy Classification is considered as “BEST”.

   If any one of the Fuzzy Classifier output variable (Bad, Good and Best) is present more number of times in Feature Extracted parameters rule, the Classifier will assign that Fuzzy Classifier output variable to be the final output in the Fuzzy System. The Fuzzy Rule Design is shown in Fig.7.
Number of Fuzzy Rules is dependent on number of input variables and their membership functions. In Fuzzy Rule Based Selection model has 5 variables and 3 membership functions = \(3^5 = 243\) rules has shown in Fig 8.

2. **Fuzzy Constraints:**
- Input Variables = 5
- Output Variables = 3
- Membership Function = 3
- Connection = AND
- Defuzzification = Centroid
- Aggregation = Sum

3. **Membership Functions (MF)**
A Membership function is a curve that defines how each point in the input space is mapped to a membership value or degree of membership between 0 and 1. The only condition a membership function must really satisfy is that it must vary between 0 and 1. The Fuzzy Logic Toolbox includes several built in membership function types. The list of membership function types are linear function, Gaussian distribution function, sigmoid curve, straight lines, triangular, trapezoidal and quadratic and cubic polynomial curves. All membership functions have the letters mf at the end of their names.

The selected membership function types are:
- **a)** Triangle: The Triangular membership function name is trimf. It collects more than three points to form a triangle.
- **b)** Trapezoidal: The Trapezoidal membership function name is trapmf. It has a flat top and a truncated triangle curve.
- **c)** Sigmoid: The Sigmoid membership function name is sigmf. It provides asymmetric membership functions.
- **d)** Gaussian: The Gaussian membership function name is gaussmf. It provides smooth and nonzero curve at all points.

4. **Fuzzy Score**
The output will be in the form of Fuzzy Score, for all the inputs individual score will be generated by FIS. Fuzzy Score is the value that is calculated by FIS considering all fuzzy constraints and membership functions. Fuzzy Score is decided by fuzzy rules and input variables. The Fuzzy Score for this FIS is shown in Fig 10.

E. **Classification**
Classifier assigns a class to the feature extracted signal as good, bad and best using fuzzy rules. The Fuzzy Classification result is shown in Fig 10. The interval between 0 to 0.38 is classified as BAD, interval between 0.39 to 0.47 is classified as GOOD and the interval between 0.48 to 0.5 is classified as BEST. These results have plotted in Membership Function as shown in Fig 9.

V. **RESULT**
The Fuzzy rule based Classification was evaluated for the Feature Extracted data Set. FIS generates score for each input signal based on the fuzzy Constraints and Fuzzy Rules. Finally, the obtained Fuzzy Score is classified as GOOD, BAD and BEST by using algorithm.

\[
\text{if (Fuzzy Score} \leq 0.38) \\
\text{Result} = \text{BAD} \\
\text{else if (Fuzzy Score} \leq 0.47) \\
\text{Result} = \text{GOOD} \\
\text{else} \\
\text{Result} = \text{BEST}
\]

The Final Classification of Extracted Features is shown in Fig 10.
VI. CONCLUSION

The Time series signals of different frequency ranging from 8Hz to 17.5Hz, 20 input data signals S100 to S119 are generated with time period of 2 seconds and 0.01 sampling rate by using Matlab. The generation of input data signal is shown in Fig.3. Total 4000 samples are acquired in data set acquisition table is shown in Fig.4. The Features are extracted from data set acquisition table, based on the Mathematical calculations like Average, Minimum, Standard Deviation and Variance methods. Total 200 features are extracted as shown in Fig.5.

These Feature extracted five parameters are considered as input to the Fuzzy Rule Based Selection Block in Fuzzy System. The Fuzzy Rules have designed in Fuzzy Inference System as if sample value is less than or equal to 0.38 considered as best, if it is less than or equal to 0.47 considered as good and if it is less than or equal to 0.5 considered as best as shown in Fig.7. The Fuzzy Classifier Algorithm assigns classes to the Feature extracted outputs as GOOD, BAD and BEST. Finally, 7 BEST, 4 GOOD and 9 BAD Fuzzy results have obtained, as shown in Fig.10.

Fig.10. Feature Extraction and Classification Result

REFERENCES

Fuzzy Rule Based Feature Extraction and Classification of Time Series Signal


Sandya H.B is currently pursuing M.Tech, final year in the stream of Digital Electronics and Communication from AMC Engineering College, Bangalore, VTU, Belgaum. She had received her Bachelor's degree in Electronics and Communication Engineering from V.V.I.E.T College, Mysore, Karnataka, India. Her area of interest includes Digital and Analog Communication, Fuzzy logic, Digital Signal Processing etc.

Hemanth Kumar P. currently pursuing Ph.D. in the stream of Computer Science Engineering from VTU Resource Research Centre, Belgaum, Karnataka had received his Master's degree in Digital Communication from National Institute of Technology (MANIT) Bhopal, Madhya Pradesh, India in the year 2010 and Bachelor's Degree from Global Academy of Technology, Bangalore in 2008. Currently working as Assistant Professor in AMC Engineering College, Bangalore in Department of Electronics & Communication Engineering. His area of interest includes Soft Computing Techniques, Data Mining, Pattern Classification, Communication etc. He has have three years of Teaching Experience and guided six M.Tech projects successfully. He has published more than 10 International Journals.

Himanshi Budhiraja is currently working as Assistant Professor in AMC Engineering College, Bangalore in Department of Electronics & Communication Engineering. She had received Master's degree and Bachelor's Degree in branch ECE from Kurukshethra University, Kurukshetra, Haryana, India. She has Six years of Teaching Experience. Her specialization is in field of Optical Communication

Susham K Rao is currently working as Assistant Professor in AMC Engineering College, Bangalore in Department of Electronics & Communication Engineering. She had received Master's degree in branch VLSI and Embedded Systems from Reva Institute of Technology (RITM) Bangalore, Karnataka, India and Bachelor's Degree from KVG Engineering College, Sullia. She has seven years of Experience which includes Industry and Teaching. Her area of interests is VLSI, Signal Processing, Embedded systems etc.