Case Study: Fuzzy based Classification technique for Time Series data

Introduction

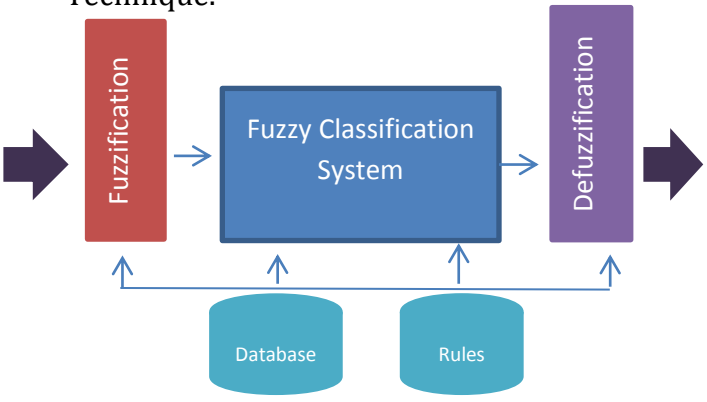
A Time series is a collection of observations of well-defined data items obtained through repeated measurements over time. Time series data have a natural temporal ordering.
Examples of time series are the daily closing value of the Dow Jones index and the annual flow volume of the Nile River at Aswan

Feature Extraction

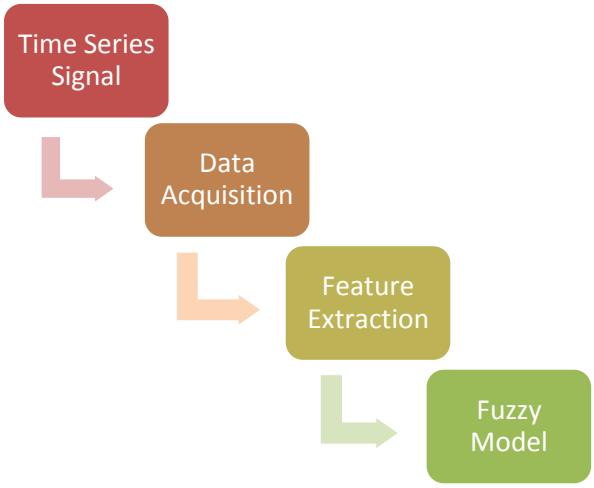
Feature extraction is a special form of dimensionality reduction which transforms the input data into a set of features.

Fuzzy Technique

A Fuzzy System is a combination of fuzzy logic and fuzzy set theory, provide a rich and meaningful addition to standard logic. Fuzzy Logic is flexible, tolerant of imprecise data and is based on natural language and human communication.
This case study deals with the feature extraction and classification using a Fuzzy Technique.



Stages of Implementation



Time Series Signal Data

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.

Data Acquisition

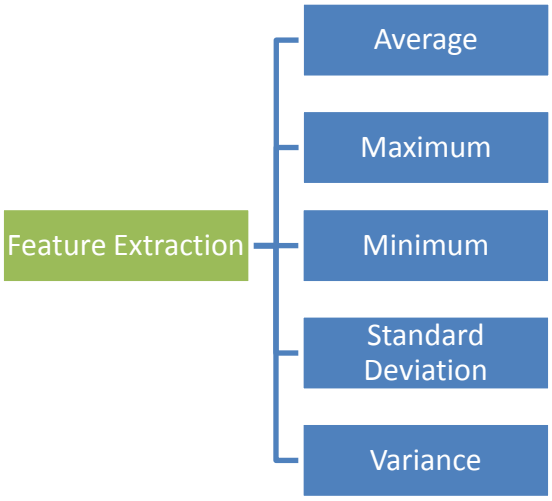
Data Set is collected from the input signal as the signal is varying continuously. It has different values of amplitude at different interval of time. A total of 4000 data set samples are collected.

$$\text{Total number of samples} = \frac{T}{R} \times 20$$

T = Total time period = 2 seconds
R = Sampling rate = 0.01 seconds

Feature Extraction

The features are extracted based on the parameters like Average, Maximum, Minimum, Standard Deviation and Variance methods. Total 100 features are extracted.



Average Method (A)
Average is defined as

$$Average = \frac{\sum x}{N}$$

x is data, N is total number of datasets

Maximum Method (M)
The maximum is defined
Max = maximum($x_1, x_2, x_3, x_4 \ldots\ldots x_N$)

Minimum Method (m)
The Minimum is defined as
Min = minimum($x_1, x_2, x_3, x_4 \ldots\ldots x_N$)

Standard Deviation (σ)

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.

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

σ = Symbol of Standard Deviation

μ = mean of all the values in the data set

N = Total Number of values in data set
x = each value in the data set

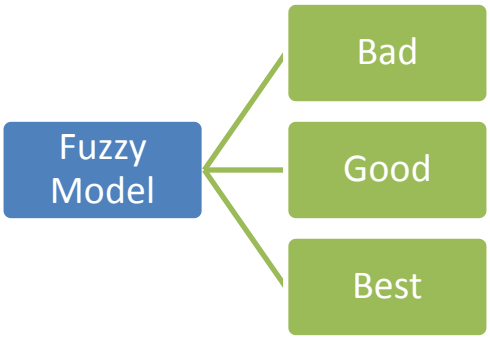
Variance (σ^2)
Variance is defined as the square of standard deviation or the variance of a random variable or distribution is the expectation, or mean, of the squared deviation of that variable from its expected value or mean.

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

Model

The Extracted parameters Average, Max, Min, Standard Deviation and Variance is considered as input variables to the Fuzzy rule based selection process block. The model maps an input features to output classes using FL. The fuzzy rules have written for Extracting Features to get results as Good, Bad and Best data sample values.

Case Study: Fuzzy based Classification technique for Time Series data



Fuzzy Constraint Table:

Input Variables	5
Output Variables	3
Membership Function	3
Connection	AND
Defuzzification	Centroid
Aggregation	Sum

Rules:

A	**M**	**m**	**σ**	**σ^2**	**O**
b	b	b	b	b	b
G	G	G	G	G	G
B	B	B	B	B	B

A: Average, M: Maximum, m: minimum, σ: Standard Deviation, σ^2: Variance, O: Output

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.

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.

Membership Functions:

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.

Classification:

Classifier assigns a class to the feature extracted signal as good, bad and best using fuzzy rules. 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.

Result:

The features are all classified as GOOD, BAD and BEST and can be used for further predictive analysis.

Conclusion:

Fuzzy system can be easily implemented to a time series signal that can be used to classify the signals using the feature selection methodology. The fuzzy system eases the comprehension of the signals which is a crucial step for building any model. Fuzzy systems help in the automated recognition of the different classes with minimal rate of misclassification.

Results

Feature Selection								
Sig	**Freq**	**Avg**	**Max**	**Min**	**SD**	**Var**	**Result**	**O**
S100	8	0.190836	0.99803	-0.9823	0.03377	0.02281	0.50	BEST
S101	8.5	0.1434	0.99951	-0.9956	0.0342	0.02339	0.50	BEST
S102	9	0.083236	0.99803	-0.9823	0.03562	0.02537	0.47	GOOD
S103	9.5	0.030215	0.99556	-0.998	0.03656	0.02673	0.47	GOOD
S104	10	0	0.95106	-0.9511	0.03627	0.02632	0.37	BAD
S105	10.5	-0.00075	0.99803	-0.9956	0.03539	0.02505	0.38	BAD
S106	11	0.024206	0.98229	-0.998	0.03509	0.02463	0.5	BEST
S107	11.5	0.062829	0.99556	-0.9877	0.03561	0.02536	0.47	GOOD
S108	12	0.099532	0.99803	-0.9823	0.03606	0.02601	0.47	GOOD
S109	12.5	0.120711	1	-1	0.03574	0.02555	0.50	BEST
S110	13	0.1192044	0.998027	-0.98229	0.035018	0.024526	0.50	BEST
S111	13.5	0.0962552	0.992115	-0.99951	0.034855	0.024297	0.47	GOOD
S112	14	0.0604868	0.998027	-0.95106	0.035534	0.025253	0.47	GOOD
S113	14.5	0.0244408	0.968583	-0.99951	0.036273	0.026315	0.50	BEST
S114	15	1.055E-16	1	-1	0.036274	0.026316	0.38	BAD
S115	15.5	-0.005677	0.998027	-0.98769	0.035661	0.025434	0.38	BAD
S116	16	0.0076459	0.998027	-0.99803	0.035268	0.024877	0.50	BEST
S117	16.5	0.0335974	0.929776	-0.94088	0.035563	0.025295	0.47	GOOD
S118	17	0.0617774	0.992115	-0.98229	0.036059	0.026004	0.47	GOOD
S119	17.5	0.0815926	0.987688	-1	0.036032	0.025549	0.47	GOOD