2010:03

Fuzzy Clustering Analysis

Md. Ehsanul Karim
Feng Yun
Sri Phani Venkata Siva Krishna Madani

Thesis for the degree Master of Science (two years)
in Mathematical Modelling and Simulation
30 credit points (30 ECTS credits)

April 2010

Blekinge Institute of Technology
School of Engineering
Department of Mathematics and Science
Supervisor: Proff. Elisabeth Rakus Andersson

Abstract

The Objective of this thesis is to talk about the usage of Fuzzy Logic in pattern recognition. There are different fuzzy approaches to recognize the pattern and the structure in data. The fuzzy approach that we choose to process the data is completely depends on the type of data.

Pattern reorganization as we know involves various mathematical transforms so as to render the pattern or structure with the desired properties such as the identification of a probabilistic model which provides the explaination of the process generating the data clarity seen and so on and so forth. With this basic school of thought we plunge into the world of Fuzzy Logic for the process of pattern recognition.

Fuzzy Logic like any other mathematical field has its own set of principles, types, representations, usage so on and so forth. Hence our job primarily would focus to venture the ways in which Fuzzy Logic is applied to pattern recognition and knowledge of the results. That is what will be said in topics to follow.

Pattern recognition is the collection of all approaches that understand, represent and process the data as segments and features by using fuzzy sets. The representation and processing depend on the selected fuzzy technique and on the problem to be solved.

In the broadest sense, pattern recognition is any form of information processing for which both the input and output are different kind of data, medical records, aerial photos, market trends, library catalogs, galactic positions, fingerprints, psychological profiles, cash flows, chemical constituents, demographic features, stock options, military decisions.. Most pattern recognition techniques involve treating the data as a variable and applying standard processing techniques to it.

Table of Contents

List of Figures

List of Tables

Acknowledgement

It is an auspicious occasion for us as students of Mathematical Modelling and Simulation, one of the prestigious academic centres of the Blekinge Institute of Technology (BTH) to express our deep feelings of gratitude to the department especially to our supervisor, teachers and also to the departmental staffs.

We are immensely indebted to our supervisor and examiner, Professor Elisabeth Rakus Andersson. for her wonderful guidance, inspiration and continuous encouragement. We would like to extend our gratitude for her sincere review and correction of this dissertation work that could not be finalized without her astute supervision. We think ourselves fortunate enough to get such a wonderful person as my supervisor and the time that we spent with her for the thesis period will remain as an enjoyable experience since a long time ahead.

We pay profound regard to Raisa Khamitova for her very valuable directives and special attention. We also acknowledge to our friends who, through their interest and work, are our constant source of inspiration.

Finally, we would like to acknowledge our parents who always believe in us. Their unconditional support and continuous inspiration kept us alive in this frozen land. Our brothers and sisters who have truly been a source of inspiration are very much keen and hopeful for the best performance of the dissertation that we are going to submit. We wish we could fulfill their aspiration and we dedicate our thesis work to our family members.

Chapter 1: Technology & Trends

Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth - truth values between "completely true" and "completely false". It was introduced by Dr. Lofti Zadeh of U.C. Berkeley in the 1960's.

1.1 History

A pictorial object is a fuzzy set which is specified by some membership function defined on all picture points. From this point of view, each element participates in many memberships. Some of this uncertainty is due to degradation, but some of it is inherent…In fuzzy set terminology, making figure/ground distinctions is equivalent to transforming from membership functions to characteristic functions." 1970, J.M.B. Prewitt

On Fuzzy Logic

- Zadeh, L.A., "Fuzzy Sets," Information and Control, 8, 338-353, 1965.
- Zadeh, L.A., "Fuzzy Logic and Approximate Reasoning," Synthesis, 30, 407-428, 1975.

On Fuzzy Geometry

- Rosenfeld, A. (1984b): The Fuzzy Geometry of Image Subsets, Pattern Recognition Letters, Vol. 2, pp. 311-317.
- Rosenfeld, A., Pal, S. K. (1988): Image enhancement and Thresholding by optimization of fuzzy compactness, Pattern Recognition Letters, Vol. 7, pp. 77-86.

Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth - truth values between "completely true" and "completely false". Fuzzy Logic is aimed at precision of approximate reasoning.

The use of fuzzy logic for creating decision-support and expert systems has grown in popularity among management and financial decision-modeling experts. Still others are putting it to work in pattern recognition, economics, data analysis, and other areas that involve a high level of uncertainty, complexity, or nonlinearity.

1.2 Pattern Recognition

The term pattern recognition always invites debate. In a simple way it can be defined as the search for structure in data. Let us discuss the each of the major elements in this definition.

1.3 The Data

There are different types of data like qualitative, quantitative, numerical, pictorial, textural, and linguistic or in some cases can be indifferent combinations of these. Examples of the data sources are medical records, aerial photos, market trends, library catalogs, galactic positions, fingerprints, psychological profiles, cash flows, chemical constituents, demographic features, stock options, military decisions. The technique or method of search pattern is applicable to any of these data types and sources. The search option is used to know the techniques for data processing and usually the data set is denoted by X.

1.4 The Search

The most commonly used important method for data analysis is Eyeball technique. Many traditional data analyzers use subjective assessment of data patterns. Subjective pattern recognition is mostly used in Medical diagnosis. With the modern technology, more and more reliable forms of data analysers came into light.Statistics provides the identification of a probabilistic model which provides the explaination of the process generating the data. Statistical analysis is slow when done by hand, faster with the calculators and is very fast with the modern computers. So with the modern computers, we are encouraged to search for the data structures that are neither probabilistic or stochastic. The increased computational power also forces us to hypothesize the axiomatic structures, models and methods which enables the machines to send the findings to user readable forms and to be used by us. The structure,data and our models that we need to find decides the type of search to be performed.

1.5 The Structure

The manner in which the information, that is to be carried by data, can be organized so that relation between the variables in the process can be identified is called the Structure. Relationships might be causal or simply connective. Representations of the recognized structure depend on the data, the method of search, and the model used. Data contains the information, the search identifies it and the structure can represent it. Different models convey different amounts and types of information concerning the process they represent.

The preceding paragraphs discuss the following concepts in detail: Clustering analysis, classification, and feature selection. To explain them more accurately, let X be a data set of n items x_k where $k = 1,....,n,$ each of which is one observation from some physical process. The specific form of the observations is immaterial for the definitions below.

1.6 Cluster Analysis

Clustering in X has classically meant the identification of an integer $c, 2 \le c \le m$ and a partitioning of X by c mutually exclusive, collectively exhaustive subsets of X (the "clusters"). The member of each and every cluster have more similarity to one another than members of other clusters. It will be precisely applicable for the mathematical similarities between the x_k's in some operational sense. Cluster structure in X tells associations among individuals of a population.

1.7 Classifications

X has been drawn from the data space which is denoted by S, i.e., $X \subset S$. A classifer for S is a device where S itself is going to be partitioned into c decision regions. Clear representation of these regions rely on:

- The nature of S
- The way in which the regions are formed

- The model we choose such as on the data, the search, and the structure.

The above factors have an effect on the role played by a sample data set X from data space S in classifier design. X is commonly used to delineate the decision regions in S. It is possible to search for structure in an entire data space during the process of classification. The structure may enable us to classify subsequent observations rapidly and automatically. The main purpose of classification is construction of taxonomy of the physical processes involved. Classification attempts to discover associations between subclasses of a population.

1.8 Feature Selection

In feature selection we are going to search for the internal structure in data items. The goal is to enhance their usefulness as components of data sets for clustering or classification. We wish to notice the best data space by looking for structure within its components. Feature selection always looks for the associations between the characteristics of individuals in the sample and population.

The main issues are explained briefly as follows:

Feature selection: the search for structure in data items, or observations $x_k \in X$

Cluster analysis: The searches for structure in data sets, or samples $X \subset S$

Classification: The search for structure in data spaces, or populations S.

We can consider Medical diagnosis which is an excellent example for the above issues. A doctor wants to classify patients according to their actual diseases. He has to prescribe correct treatment by means of the classification.

The following steps need to be done in order to prescribe the correct treatment:

- In the first, step data has to collected.
- Symptoms are noted.
- Tests performed.
- Each patient generates a record x_k and records are clustered.

This group is observed to have hypertension due to primary aldosteronism, that group due to endovascular constriction. Eventually a classifier is a designer where a new patient is noticed with the doctors decision rule and has to be treated accordingly. Here we can observe all components of the pattern recognition such as data, search and structure.

1.9 Pattern Recognition Approach

There is a multiplicity of pattern recognition approach and no definite consensus on how to categorize them. The objective of a pattern recognition system is to perform a mapping between the representation space and the interpretation space. This kind of mapping is also called a hypothesis.

There are two distinct ways such hypothesis can be obtained:

1.9.1 Supervised, Concept driven or inductive hypothesis

Find in the representation space or hypothesis corresponding to the structure of the interpretation space. In order to be useful any hypothesis found to approximate the target values in the training set must also approximate unobserved patterns in a similar way.

1.9.2 Unsupervised, data driven or deductive hypothesis

Finds a structure in the interpretation space corresponding to the structure in the representation space. The unsupervised approach attempts to find a useful hypothesis based only on the similarity relations in the representation space.

Chapter 2: Classical Clustering Analysis

Clustering models are generally interpreted geometrically by taking into consideration of their responce on two-or more three dimensional examples. Some difficulties are intrinsic in trying to elaborate the successful clustering criterion for a wide spectrum of data structures. The ideal cases compact,well-separated,equally proportioned clusters are encountered in the real data of the applications.

Data sets are mixture of shapes as spherical,elliptical and sizes are intensities, unequal number of observations, and geometries are linear, angular and curved.

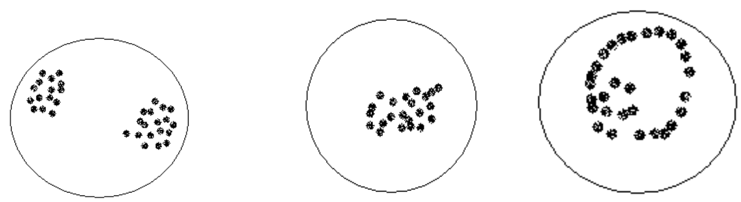


Figure2.0 Figure2.1 Figure2.2

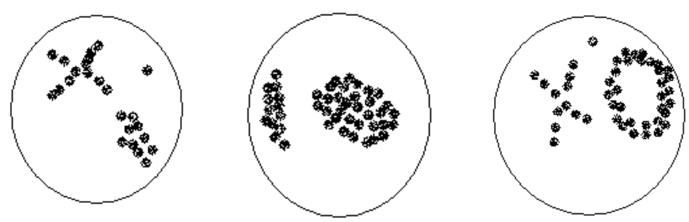


Figure2.3 Figure2.4 Figure2.5

Figure 2-1: Different forms of data sets

Typical clusters have been obtained by the processing of the two or more data sets with the methods as follows:

- distance based objective function algorithm and
- distance based graphic-theoretic method.

This kind of typical behavior could be occured, even though both algorithms use the same distance, it means the equal measure of the similarity between the points in X, but with the different clustering phenomenos in the process. We empasize the similarity measures by building blocks for clustering criteria, and in the simplest models the measure of similarity can be served as a criterion of validity, but more generally the same measure can be used with various criteria to yield different models and results for the same data set.

Figure 2.6

Figure 2.7

Figure 2.8

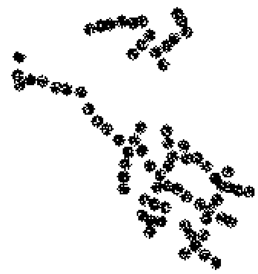


Figure 2.9

Figure 2-2: Different forms of clusters

2.1 Hierarchical Method

In agglomiration and splitting techniques new clusters are formed by reallocation of membership of one point at a time, which depends on the some measure of similarity. Therefore a hierarchy of nested clusters one for each cluster has been generated.

2.2 Graph-Theoretic Methods

X should be regarded as the node set in this group. Edge weights between pairs of the nodes could be depended on the measure of similarity between pairs of nodes. The criterion for clustering is commonly some measure of connectivity or bonding between the groups of nodes-breaking edges in minimal spanning tree to form subgraphs. This is often used the clustering strategy. These sought of techniques are well acceptable to data with chains or pseudolinear structure.

Example: The single linkage technique is well suited to data of figures in the above examples.

Some difficulty may cause to the pure graph theoretic methods due to mixed data structures. Data with hyperelliptical clusters, noise, and bridges are usually badly distorted by graph-theoretic models because of their chaining tendencies. This failure is explained in the above mentioned example.

The clustering analysis concerns to the partition of the data into the equallant subsets also known as clusters. The data in each equallant subset has the common behavior such as distance, density based algorithm. Density based algorithm is a major equallant subset which explains the number of unknown cluster in advance. It progressively increases the total strength or connectivity of the cluster set (equallant subset) by cumulative attraction of the nodes between the clusters. Major clust is the most pronounce and successful algorithm of unsupervised document clustering. Graph theory depends on the algorithm assigned to each document to that cluster the majority of its neighbors belongs to.

The node neighborhood is going to be calculated by using a particular similarity measure which has been assumed to be the weight of each edge between the nodes of the graph.

Major cluster tells the number of clusters and assigns each target document to precisely one cluster. A different approach should be required in order to determine how the document has to be assigned to more one than one category. The traditional major clust algorithm deals with the crisp data, the another version called F-major clust deals with fuzzy dat.

2.3 Example for Fuzzy Set

- As an example, we can regard the linguistic variable color which helps us to explain the conception of a fuzzy set. It can be described by a list of terms yellow, orange, red, violet and blue , where each term is a name of the corresponding fuzzy set.

  Color = {yellow, orange, red, violet, blue}

- The non-crisp boundaries between the colors can be represented much better. A soft computing becomes possible (see Figure below).

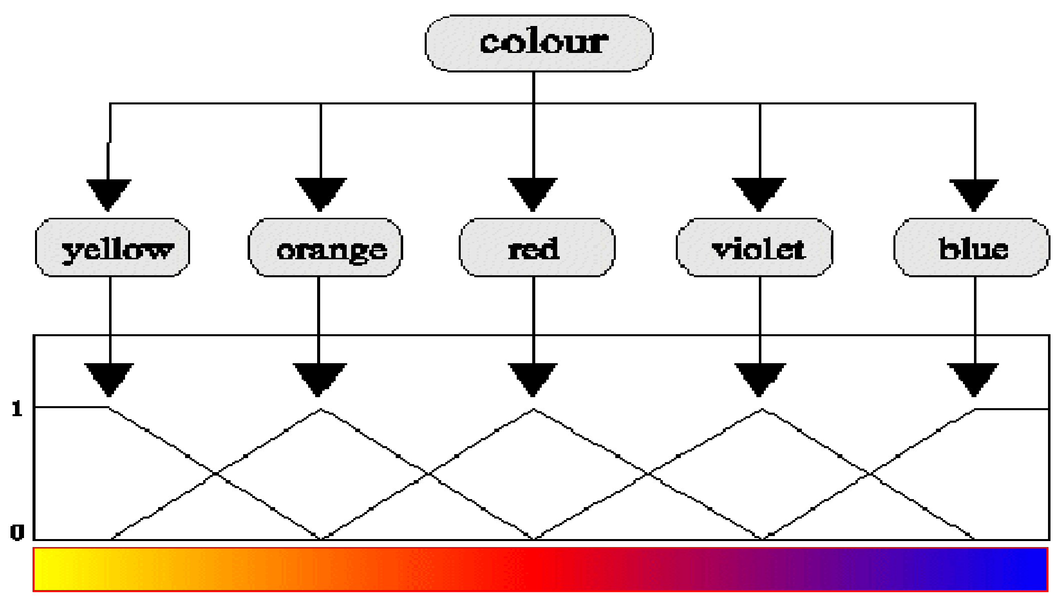
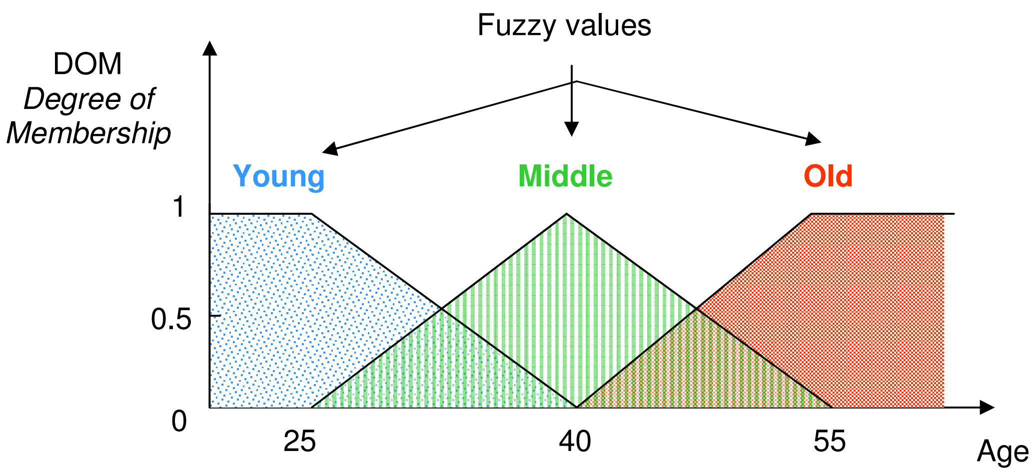


Figure 2-3: Example of Fuzzy sets from the list “color”



Elements of "age" have associated degrees of membership in the different set.

Figure 2-4: Membership functions of fuzzy sets forming the linguistic variable “age”

Membership functions representation

Ann is 28,	0.8 in set “Young”
Bob is 35,	0.1 in set “Young”
Charlie is 23,	1.0 in set “Young”
Don is 54,	0.0 in set “Young”

2.4 Fuzzy and probability concepts:

The relationship of fuzzy sets and probability has been intensively discussed since zadeh introduced fuzzy sets in 1965. Recently it has become accepted that they can be considered as independent and complementary to each other. Fuzzy sets are indicators for similarities or neighbourhood relations while probability is related to probabilistic uncertainity. Fuzzy set is defined in terms of the membership function which is a mapping from the universal set U to the interval [0,1]. Membership function could be defined as the graphical representation of the magnitude of participation of each input. It associates a weighting with each of the inputs that are processed, defines functional overlap between inputs, and ultimately determines an output response. Fuzziness is also considered as one form of the uncertainity and it is related to the linguistic variables. Linguistic variables are described as membership functions. To avoid this confusing this with fuzzy uncertainty, we explicitly refer to probabilistic uncertainity when we are in the field of probability theory.

2.4.1 Example

Fuzzy Concept:

Lets consider an example, a bank wants to classify their customers into two grups as rich and poor customers. Obviously there is no crisp separation between rich and poor in a way that customers own less than 2 million SEK are poor while they own a fortune of 2million SEK or more are rich. So according to the above explanation a person with a fortune of 2.1 million SEK is considered as reasonably rich but still a little bit poor. The indicator for similarity in fuzzy sets is called membership degree $\mu = \{0,.....,1\}$. A membership degree μ =1 indicates that an object completely belongs to set and a membership degree μ =0 shows a total dissimilarity between an object and a set. In our example, the customer with 2.1million SEK may have membership degrees of $\mu_{rich}(2.1millionSEK) = 0.65$ to the set rich and $\mu_{poor}(2.1millionSEK) = 0.35$ to the set poor. This indicates that the customer is rich but not extremely wealthy. However a customer having 30billion SEK would surely have memberships of $\mu_{rich}(1billionSEK) = 1.0$ and $\mu_{poor}(1billionSEK) = 0.0$. There is probabilistic uncertainity neither about the fortune of the customer who is having 2.1million SEK nor about the rules for how to classify him into one or other of the two sets which are determined by the functions given in figure2.2.

So the membership degrees do not indicate any probability of belonging to the sets, but similarities of values to those sets.

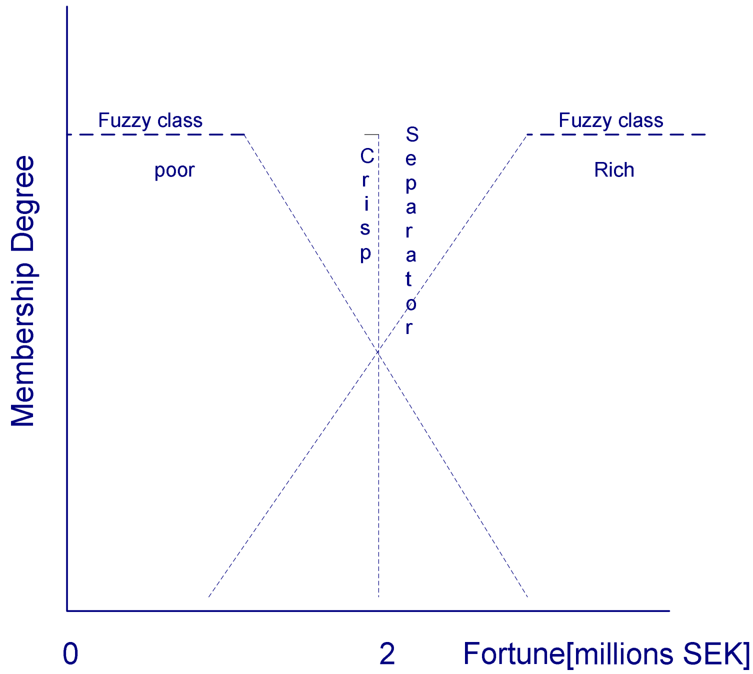


Figure 2-5: Fuzzy Concept

Probability Concept:

The same bank may face the probabilistic uncertainity about the wealth of the customer. For example, a new customer driving up with an old bicycle might be considered of having a fortune of say, 5 SEK while it might be assumed that a customer comes to the bank in a big limousine could have a millions of swedish kronars. However these are only the guesses of the bank employees. The vehicles of the customers are indicators for the wealth but no proof. So the bank clerks have to act under probabilistic uncertainity. The biker could be a billionaire while other customer who comes to the bank in a limousine might be a debt-ridden comman.

The biker (BI) might be having the fortune of 5 SEK with a probability of P_{BI}(5 SEK)=0.9 and a fortune of one million SEK with a probability of P_{BI}(1MILLION SEK)=0.1 while the limousine customer (LI) has the probabilities as follows: P_{LI}(5sek)=0.2 and P_{LI}(1million sek)=0.8.

In this example only probabilistic uncertainty is taken into consideration. In contrast to the fuzzy concept as shown in the previous section, the amounts of money (5SEK and 1 million SEK) are not examined with respect to their similarity to the sets poor and rich.

Combined Fuzzy and Probability Concept:

The fuzzy and probability concepts are independent and they can be combined. Lets consider the example of the bike rider. First the bank clerks estimate the fortune of the new customes: the bike might be having the fortune of 5 SEK with a probability of P_{BI}(5 SEK)=0.9 and a fortune of one million SEK with a probability of P_{BI}(1MILLION SEK)=0.1. Now the given amounts of money are examined with respect to their similarity to the sets rich and poor. 5 SEK may be classified with the following membership degrees: $\mu_{POOR}(5SEK) = 0.95$ and $\mu_{RICH}(5SEK) = 0.05$. For 1 million SEK we may get the membership degree as follows: $\mu_{POOR}(1MILLIONSEK) = 0.02$ and $\mu_{RICH}(1millionsek) = 0.98$.

Combining the probability and fuzziness we finally get: The biker does belong with a probability of P_{BI}=0.9 and to a membership degree of $\mu_{POOR} = 0.95$ to the set poor as well as to the set rich with $\mu_{RICH} = 0.05$. With a probability of $P_{BI} = 0.1$ and with a membership degree of $\mu_{RICH} = 0.98$ he belongs to the set rich and as well as to the set poor with $\mu_{POOR} = 0.02$.

2.4.2 Features of the function

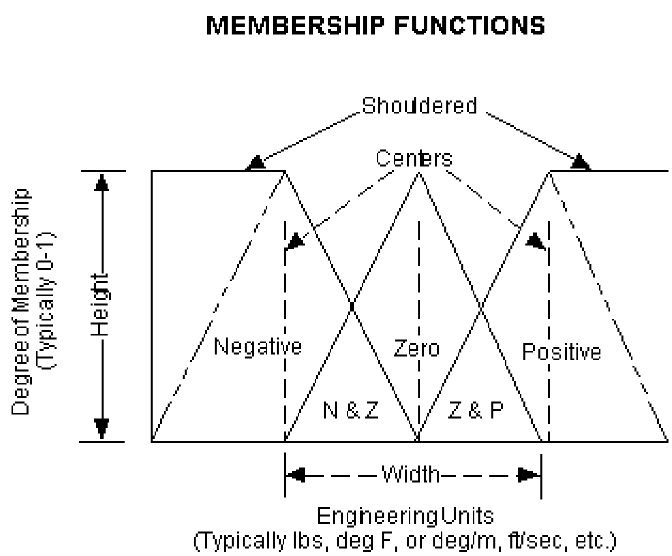


Figure 2-6: Triangular mebership functions

The above figure explains the features of the triangular membership function.

2.4.3 Butterfly Effect

In butterfly effect in classical clustering few nodes of the dataset belongs to more than one cluster. In this case the fuzzy algorithm works better than the crisp one. Figure2.4 and Figure2.5 represents an example when the classical majorclust algorithms found two clusters and then the clustering of eight nodes obtain the score of 21 by using the equation no2.1 In Figure2.5 when the classical major clust algorithm has found three major clusters for the same input data providing a total score of 18. In Figure2.6 and 2.7 we observe how the fuzzy algorithms work when few nodes can be adapted to many different clusters. We can assume that the algorithm uses the equation no 2.0 , where for the simplicity we take $\gamma_{jk}(C_k) = 0$; Two variants are presented in the figure 2.6 and we need to consider the case when the membership values are shared equally between two clusters with the membership value of 0.5, then the score we can get is 21. We have noticed that the value of objective function here is the same as in case of figure 2.4. If the nodes (documents) are highly applicable to the both databases with the membership function values 1 then the fuzzy algorithm results in better score equal 24 which is presented in figure 2.7.

The equations used for the butterfly effect are as follows:

$$\Lambda(C) = \sum_{k=1}^{K} \left(\sum_{j=1}^{|C_k|} \mu_{i(j),k} \lambda_k + \sum_{j=1}^{n} \gamma_{jk} C_k \right) \text{------------------> } (2.0)$$

$$\Lambda(C) = \sum_{k=1}^{K} |C_k| \lambda_k \text{ ----------------------------------> } (2.1)$$

$\Lambda(C)$: Total cluster connectivity

C: Decomposition of the given graph G into number of clusters C_1 ,C_2 ,C_3 ,....,C_k are clusters in the decomposition C.

λ_k: It designates the edge connectivity of the cluster $G(C_k)$ that is the minimum number of edges that must be removed to make graph $G(C_k)$ disconnected.

$\mu_{ij,k}$: membership degree of $arci(j)$ containing node j in cluster k

$\gamma_{jk} C_k$: fitness of node j to cluster k.

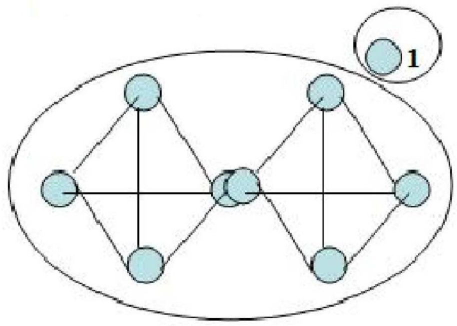


Figure 2.4

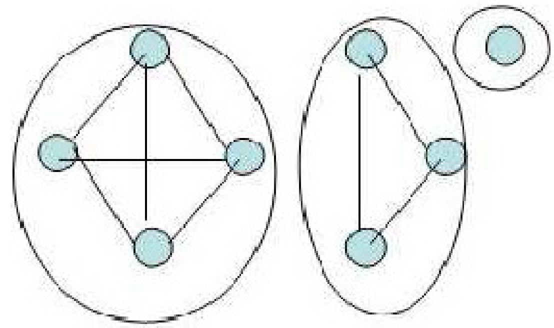


Figure 2.5

Figure 2-7: Butterfly effect in fuzzy clustering use classical major clust

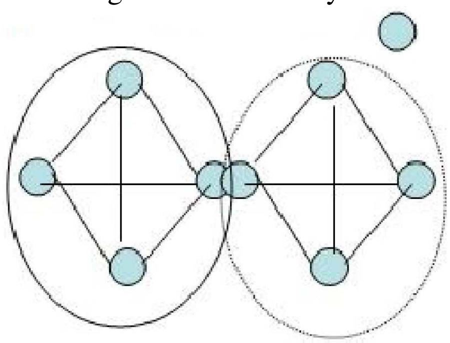


Figure2.6

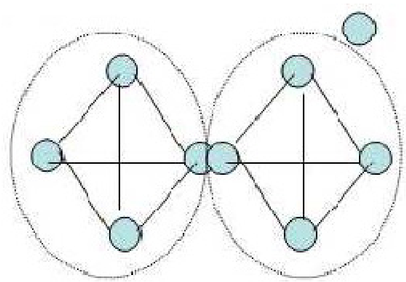


Figure2.7

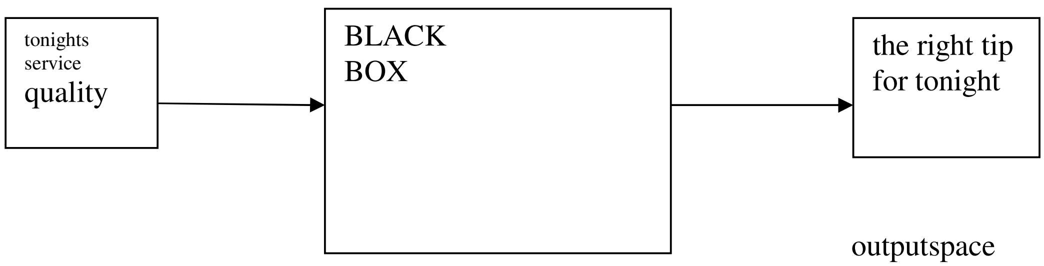
Figure 2-8: Butterfly effect in fuzzy clustering use fuzzy-majorclust approach

Butterfly effect in fuzzy clustering ,figure 2.4 and 2.5 use classical major clust, where as figure 2.6 and 2.7 use fuzzy-majorclust approach.

Chapter 3: Fuzzy Logic and Fuzzy Clustering

3.1 Fuzzy logic:

Fuzzy logic is the convenient way to map an input space to an output space.
Example: How good your service was at your resturant, and we will decide what the tip should be.



Input space

Figure 3-1: An input-output map for the tipping problem

It is all the matter of mapping inputs to the appropriate outputs. We have a black box between input and the output and the black box contains any number of things such as fuzzy systems, linear systems, expert systems,neural networks, differential equations,interpolated multi-dimensional lookup tables and etc... There are many ways to make the black box work, it turns out the fuzzy is often the best way.

3.2 An Introductory example for fuzzy approach using fuzzy logic tool box in MATLAB:

3.2.1 Fuzzy approach:

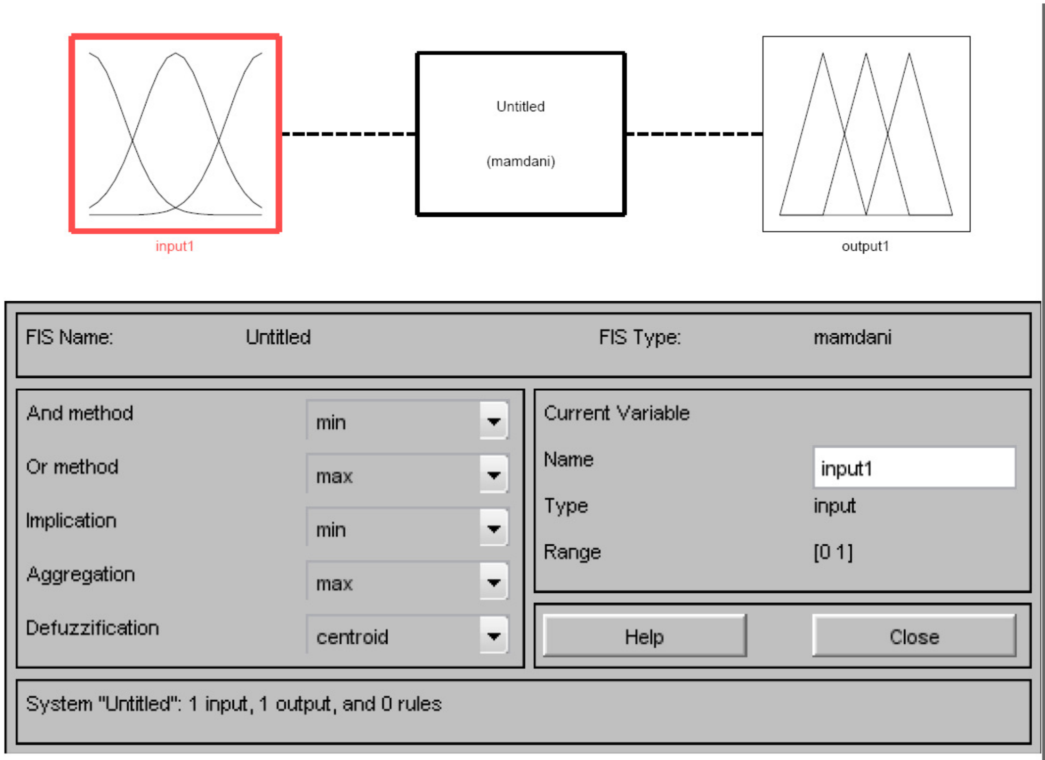
The basic tipping problem: Given a number between 0 and 10 that represents the quality of service at a restaurant (where 10 is excellent), what should the tip be?

The starting point is to write down the three important rules of tipping, based on the years of personal experience in the restaurant.

- If the service is poor then tip is cheap
- If the service is good then the tip is average
- If the service is excellent then the tip is generous

We assume that the average tip is 15%, a generous tip is 25% and a cheap tip is 5%. By using the bove three rules we begin working with the GUI tools.

Figure 3-2: The FIS editor



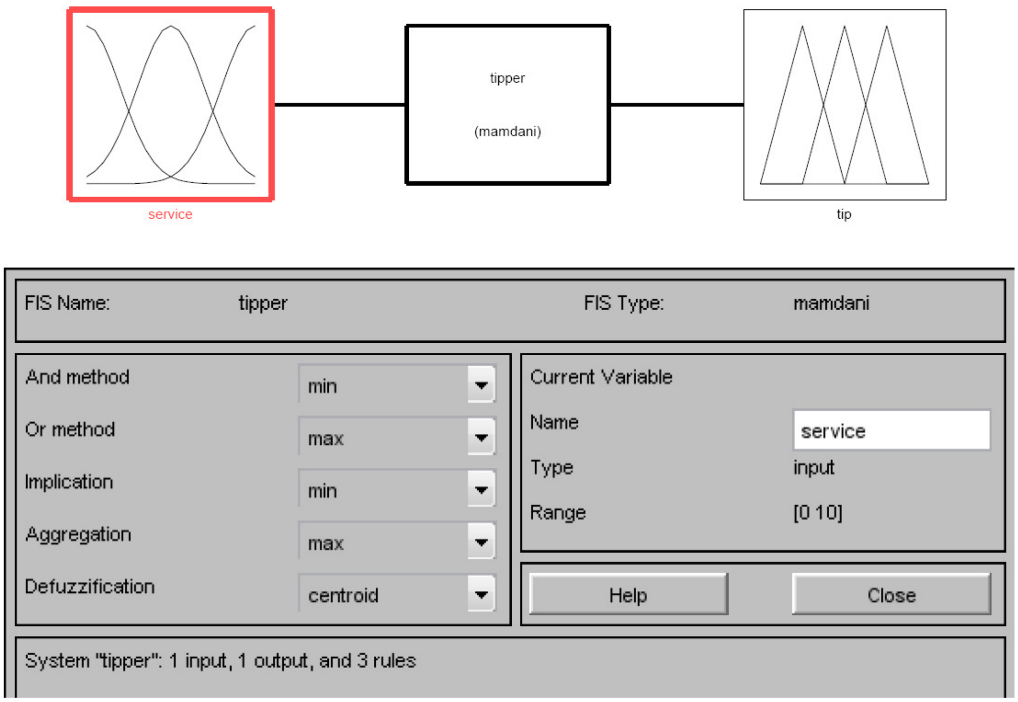
When creating the fuzzy inference system, the place to start is the FIS editor and to do that type fuzzy. Then this will take you to the view of FIS editor as shown in the above figure. The diagram shows the inputs on the left side and outputs on the right side. The system that is displayed is a default start-up system since we did not specify any particular system.We are going to build a new system from scratch.By typing fuzzy tipper1 we can load our required system for the basic tiping problem. This will take us to the fuzzy inference system associated with the file tipper1. The FIS edior displays general information of a fuzzy inference system. There is a simple diagram at the top that shows the names of each input variable and each output variable. Below the diagram is the name of the system and the type of the inference is used. The default Mamdani-style inference is used in our example. The default system already has one input and one output. Since our one input is service and our one output is tip. To change the names we have to do the following steps:

- Click once on the left-hand (yellow) box marked input1.
- In the white edit field on the right, change input1 to service and press return.
- Click once on the right hand (blue) box marked output1.
- In the white edit field on the right, change output1 to tip.

- From the file menu select *save work space as*, then enter the variable name tipper and click OK.

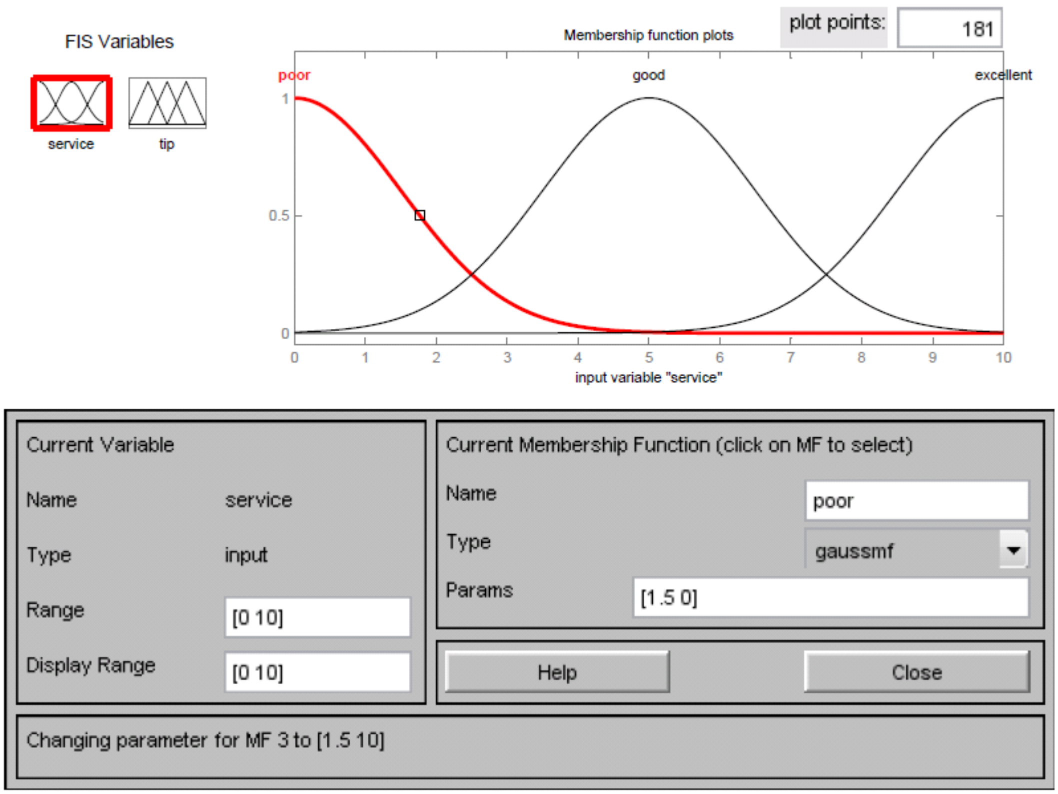
Now there is a new varaible in the workspace called tipper, So we have to rename the entire system with the same name.Now the window should lok something like this.

Figure 3-3: FIS EDITOR: TIPPER



The next thing is to define the membership functions associated with all the variables. To do this we need to open up the membership function editor by pulling down the view menu iten and selecting Edit membership functions.......

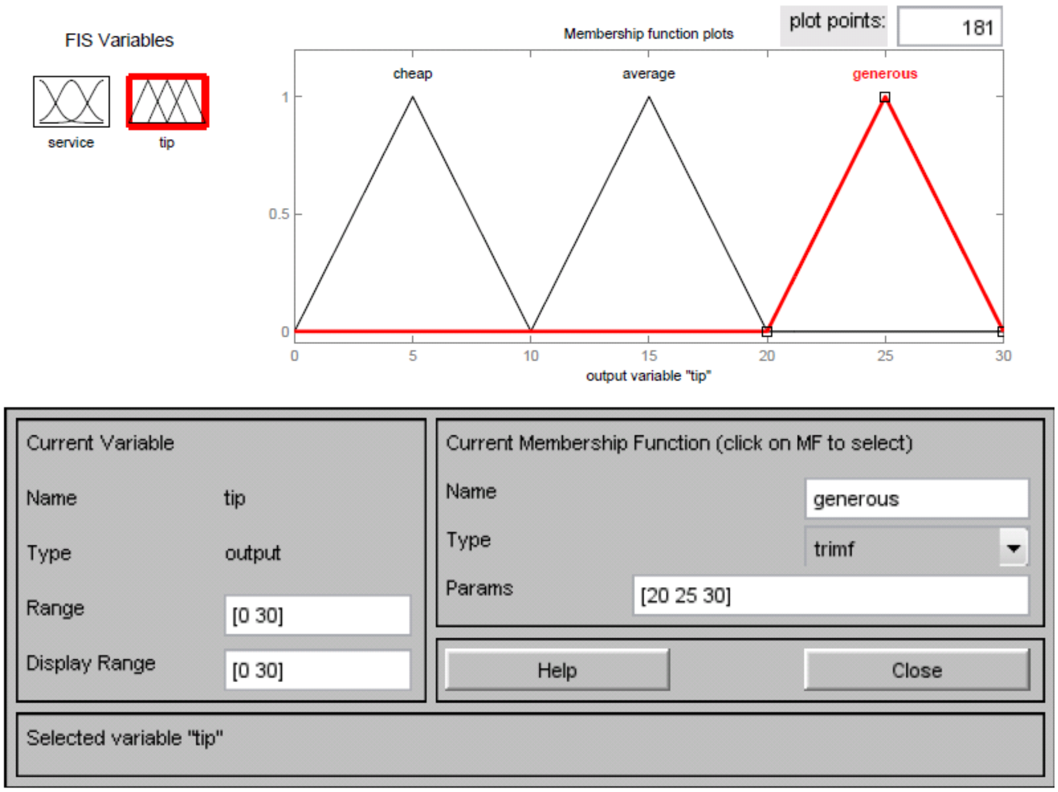
Figure 3-4: Membership function editor



The Membership Function Editor is the tool that lets you display and edit all of the membership functions for the entire fuzzy inference system, including both input and output variables. On the left side of the graph area is a variable pallet that lets you set the current variable.

Make sure that input variable is selected in the variable palette. Set the range to vector (0 10). Change the name of the left most curve to poor and change the parameters listing to (1.5 0). Name the middle curve good and right most curve to excellent and change the first parameter is to 1.5. Now we need to create the membership functions for the output variable,tip. The names of these membership function are cheap, average, and generous. Use the variable palette on the left, to display the output variable membership functions. The input range is a rating scale of 0 to 10 but the output scale is going to be a tip in between 5 and 25%. We are going to use the triangular membership function type for the output. The cheap membership function has the parameters (0 5 10), the average membership function has the parameters (10 15 20) and the generous membership function has the parameters (20 25 30).
So now our system looks like as follows:

Figure 3-5: Triangular membership function

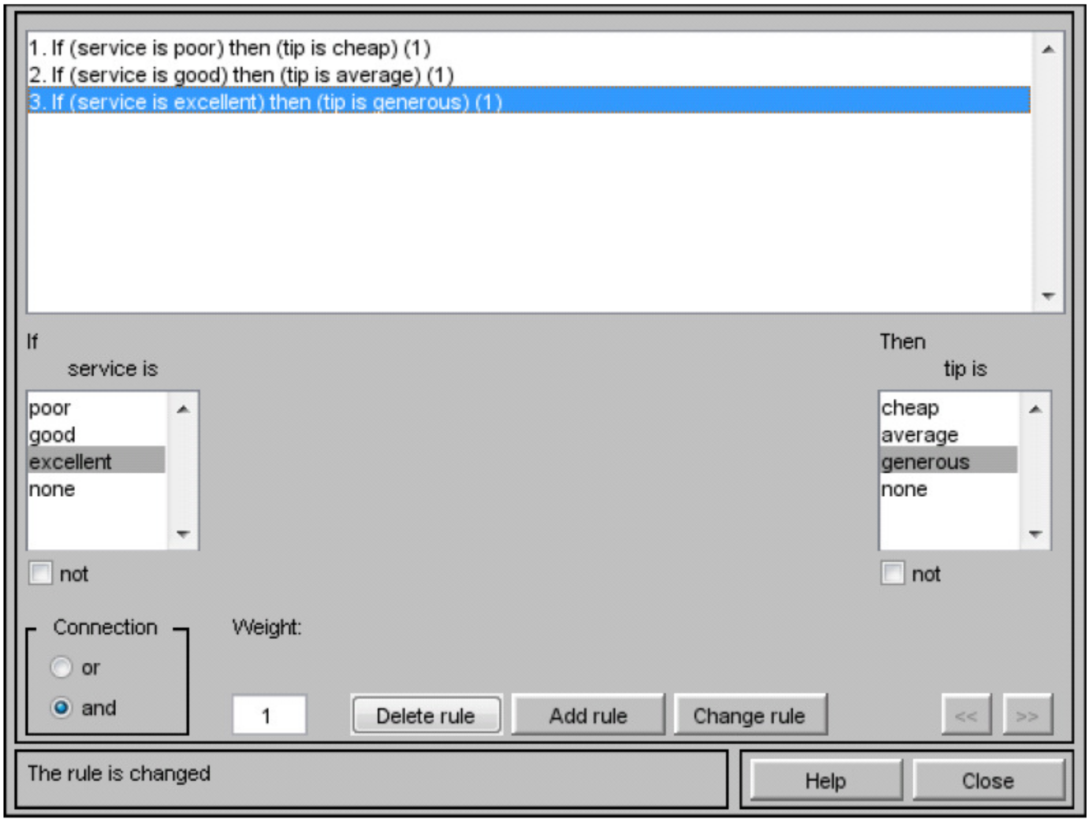


Now we are going to write down the rules by using the rule editor. The rule editor contains a large editable text field for displaying and editing rules. In the white text are write down the following rules:

1. If (service is poor) then (tip is cheap) (1)
2. If (service is good) then (tip is average) (1)
3. If (service is excellent) then (tip is generous) (1)

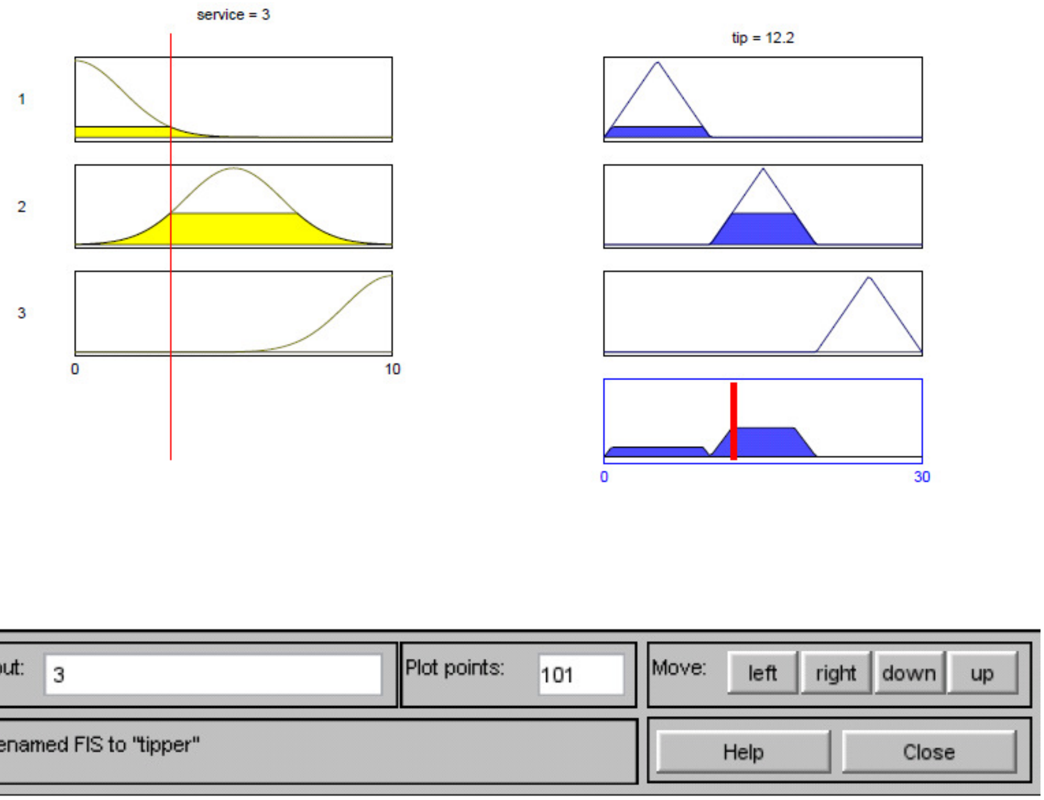
The numbers in the parentheses represents weights that can be applied to each rule if desired. If we are not specify them they are assumed to be one.

Figure 3-6: The rule editor



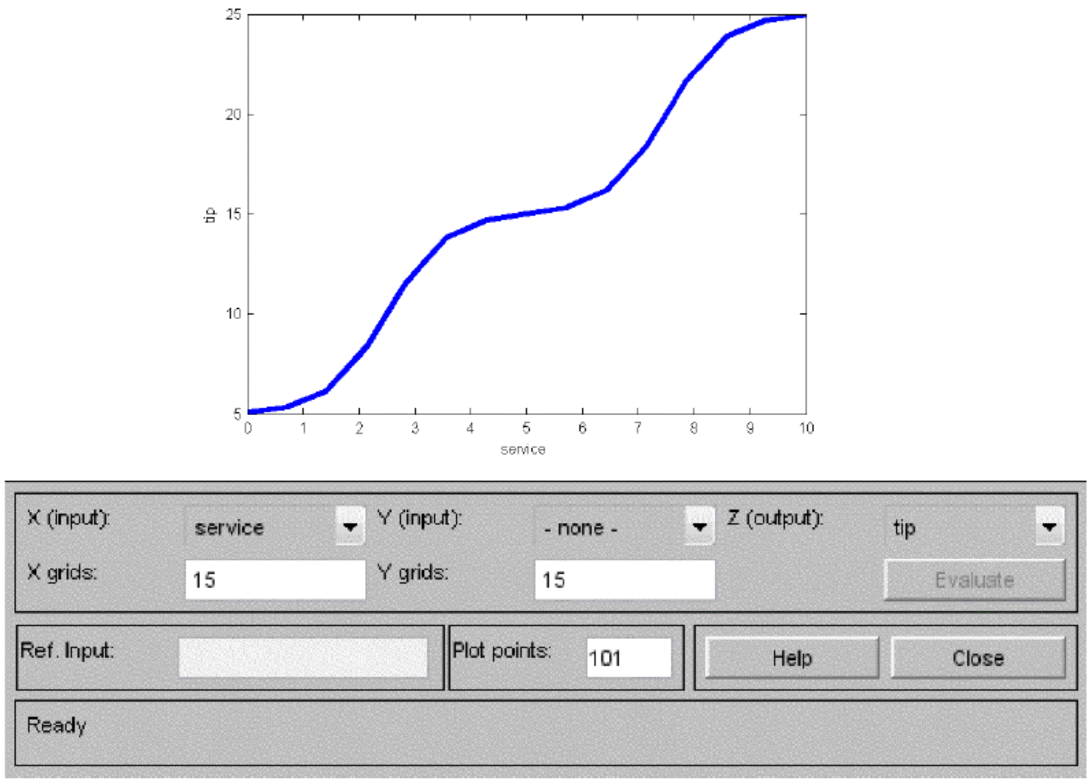
Now we are going to shift to the rule viewer where it can verify that everything is behaving the way we think it should be. The rule viewer displays the roadmap of the entire fuzzy inference process. We can observe a single figure window with seven small plots grouped in it. The yellow plots show how the input variable is used in the rules.The blue plots show how the output variable is used in the rules.

Figure 3-7: The rule viewer



The rule viewer presents a sort of micro view of the fuzzy inference system. If we would like to see the entire output surface of the system(entire span of the output set based on the entire span of the input set) we need to open up surface viewer. This is the last of our five basic GUI tools in the fuzzy logic tool box.

Figure 3-8: The surface viewer



By opening the surface viewer we have got the two dimensional curve which represents the mapping from service quality to tip amount.

3.3 K-Means Clustering- Numerical Example and Algorithm:

What is K-means clustering?

It is an algorithm to classify or to group the objects depending on attributes/features into K number of group. K is a positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus, the purpose of K-mean clustering is to classify the data.

Example:
Let us consider we have 4 objects as the training data point and each object has 2 attributes. Each attribute represents coordinate of the object.

Object	Attribute 1(X) :Weight index	Attribute 2(Y): pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

These objects belong to two groups of medicine(cluster 1 and cluster 2). Now we have to determine which medicines belong to cluster 1 and which medicines belong to the other cluster. Each medicine represents one point with two coordinates.The attribute 1(X) and atrribute 2(Y) values are assigned to the objects which give the cluster identification (which medicine belongs to cluster 1 and which medicine belongs to cluster 2).

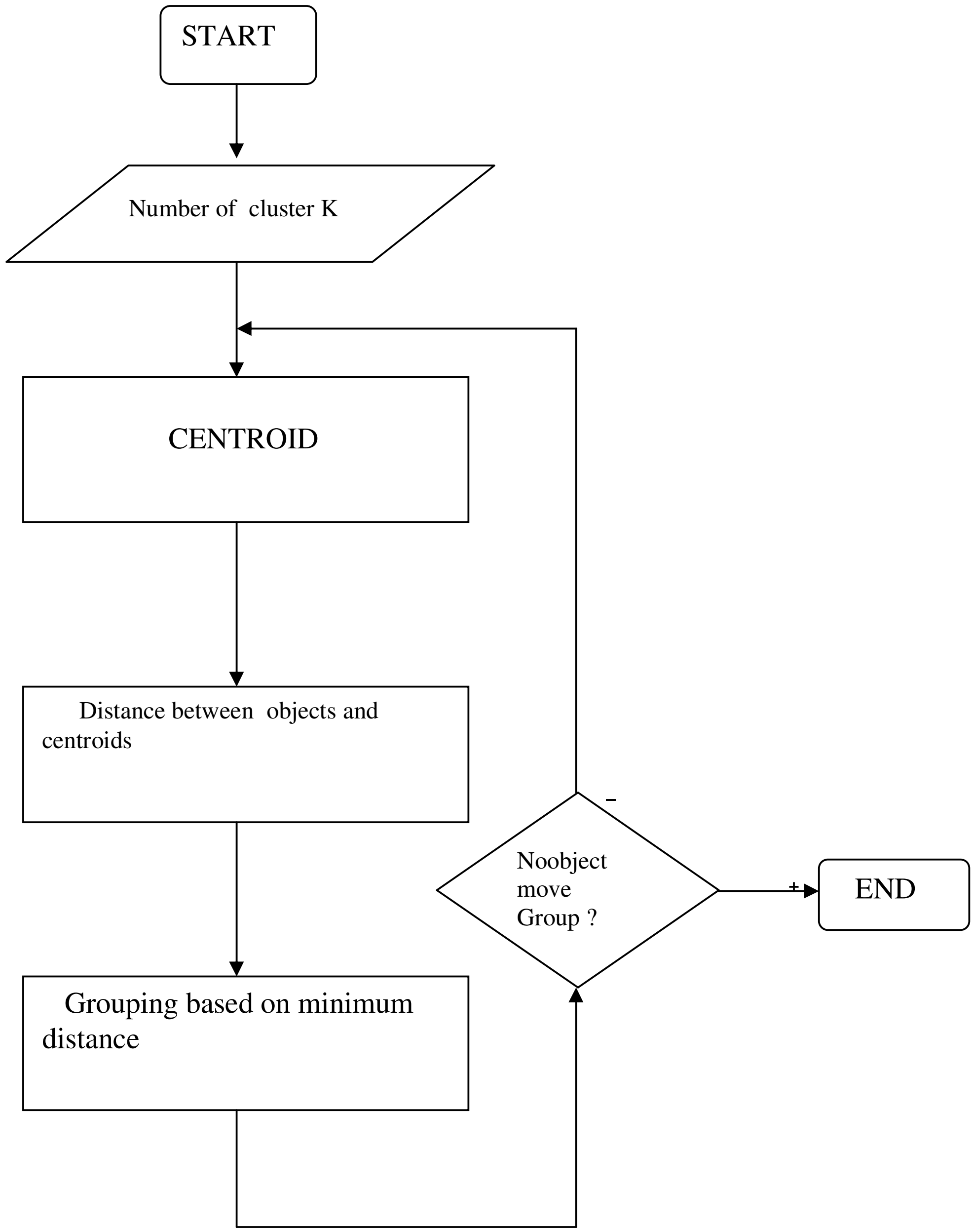
Numerical Example:

In the begining of the K-means clustering , we determine a number of clusters K and we assume the existence of the centroids or centers of these clusters. We can take any ramdom objects as the initial centroids or the first K objects can also serve as the initial centroids.

Then the K-means algorithm will do the three steps:
Iterate until stable (No object move group):

- Determine the centroid coordinates
- Determine the distance of each object to the centroids
- Group the objects based on minimum distance (we have to find the closest centroid)

Figure 3-9: K-Means Algorithm Flow Chart



The following numerical example is given to understand this simple iteration
Suppose we have several objects (4 types of medicines) and each object have two attributes or features as shown in the table below. Our aim is to group these objects into K=2 group of medicine depending on the two features (pH and weight index).

Object	Feature 1 (X): weight index	Feature 2(Y): pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

Each medicine represents one point with two features as (X , Y).

Initial value of the centroids:

In this we use medicine A and B as the first centroids. Let c_1 and c_2 are the coordinate of the centroid, then c_1= (1,1) and c_2=(2,1).

Objects-centroids distance:

We calculate the distance between cluster centroid and each object. We use Euclidean diatance, then the distance matrix at iteration 0 is:

$$D^o = \begin{bmatrix} 0 & 1 & 3.61 & 5 \\ 1 & 0 & 2.83 & 4.24 \end{bmatrix} \begin{matrix} c_1 = (1,1)\, group1 \\ c_2 = (2,1)\, group2 \end{matrix}$$

$$\begin{array}{c} \\ X \\ Y \end{array} \begin{array}{c} \begin{matrix} A & B & C & D \end{matrix} \\ \begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 1 & 3 & 4 \end{bmatrix} \end{array}$$

Each column in the distance matrix represents the object. The first row of the distance matrix corresponds to the distance between each object and the first centroid. The second row is the distance between each object and the second centroid. For example the distance between the third object C(4,3) and the first centroid is 3.61 and its distance to the second centroid is 2.83.

Object Clustering:

We assign each object to cluster 1 or cluster 2 respectively when basing on the minimum distance. So, medicine A is assigned to group 1, medicine B is assigned to group 2, medicine C to group 2 and medicine D to group 2. The element of the group matrix below is 1 if and only if the object is assigned to that group

$$G^O = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{matrix} group & 1 \\ group & 2 \end{matrix}$$

$$\begin{matrix} A & B & C & D \end{matrix}$$

Iteration 1, determine centroids:

Group 1 has only one member so the centroid remains in c_1= (1,1). Group 2 now has three members, so the centroid is the average coordinate among the three members C_2= (11/3, 8/3).

Iteration 1, Objects-Centroids distances:
The next step is to find out the distance of all objects to the new centroids.

$$D^1 = \begin{bmatrix} 0 & 1 & 3.61 & 5 \\ 3.14 & 2.36 & 0.47 & 1.87 \end{bmatrix} \begin{matrix} c_1 = (1,1) group1 \\ c_2 = (11/3, 8/3) group2 \end{matrix}$$

$$\begin{array}{cc} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} X \\ Y \end{matrix} & \begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 1 & 3 & 4 \end{bmatrix} \end{array}$$

Iteration 1, Objects clustering:
Similar to step 3, we assign each object based on the minimum distance and the group matrix is shown below.

$$G^1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{matrix} group1 \\ group2 \end{matrix}$$

A B C D

Iteration 2, determine centroids:
Group 1 and Group 2 both have two members, so the new centroids are c_1=(3/2 , 1) and c_2=(9/2 , 7/2).

Iteration 2, Objects-Centroids distances:
We have new distance matrix at iteration 2 as

$$D^2 = \begin{bmatrix} 0.5 & 0.5 & 3.20 & 4.61 \\ 4.30 & 3.54 & 0.71 & 0.71 \end{bmatrix} \begin{matrix} c_1 = (3/2, 1) group1 \\ c_2 = (9/, 7/2) group2 \end{matrix}$$

$$\begin{array}{cc} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} X \\ Y \end{matrix} & \begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 1 & 3 & 4 \end{bmatrix} \end{array}$$

Iteration 2, Objects-Clustering:
We assign each object bsed on minimum distance

$$G^2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

A B C D

We obtained the result as G^2=G^1 from the grouping of previous iteration and this iteration we conclude that the objects do not move group anymore. So the computation of the K-means clustering has reached its stability and no more iteration is needed. We conclude the final grouping as results.

Object	Feature 1 (X): weight index	Feature 2(Y): pH	Group (result)
Medicine A	1	1	1
Medicine B	2	1	1
Medicine C	4	3	2
Medicine D	5	4	2

3.4 K-Means Algorithm- MATLAB Example and Clustering Plot:

3.4.1 MATLAB Program:

```
clear all
close all
clc

load task4
X1=X(1,:);
X2=X(2,:);
plot(X1,X2,'o');
hold on
k=3;
mu = kmeans1(X,k);
m1=mu(:,1);
m2=mu(:,2);
m3=mu(:,3);
plot(m1(1),m1(2),'r*')
plot(m2(1),m2(2),'r*')
plot(m3(1),m3(2),'r*')
```

3.4.2 K-Means Algorithm:

The K-means Clustering Algorithm

The K-means algorithm can be considered the workhorse of clustering algorithms. It is a popular clustering
method that minimizes the clustering error criterium. Here follows the K-means algorithm.

1. **Initialization:** Choose K vectors from the training vector set $\mathbf{X}$ at random. These vectors will be the
initial centroids μk.

2. **Recursion:** For each vector $\mathbf{x}n$ in the training set, let every vector belong to a cluster k. This is done by
choosing the cluster centroid μk closest to the training vector $\mathbf{x}n$.

kn = argmin
k

_
$(\mathbf{x}n - \mu k)T\, (\mathbf{x}n - \mu k)$ (3)

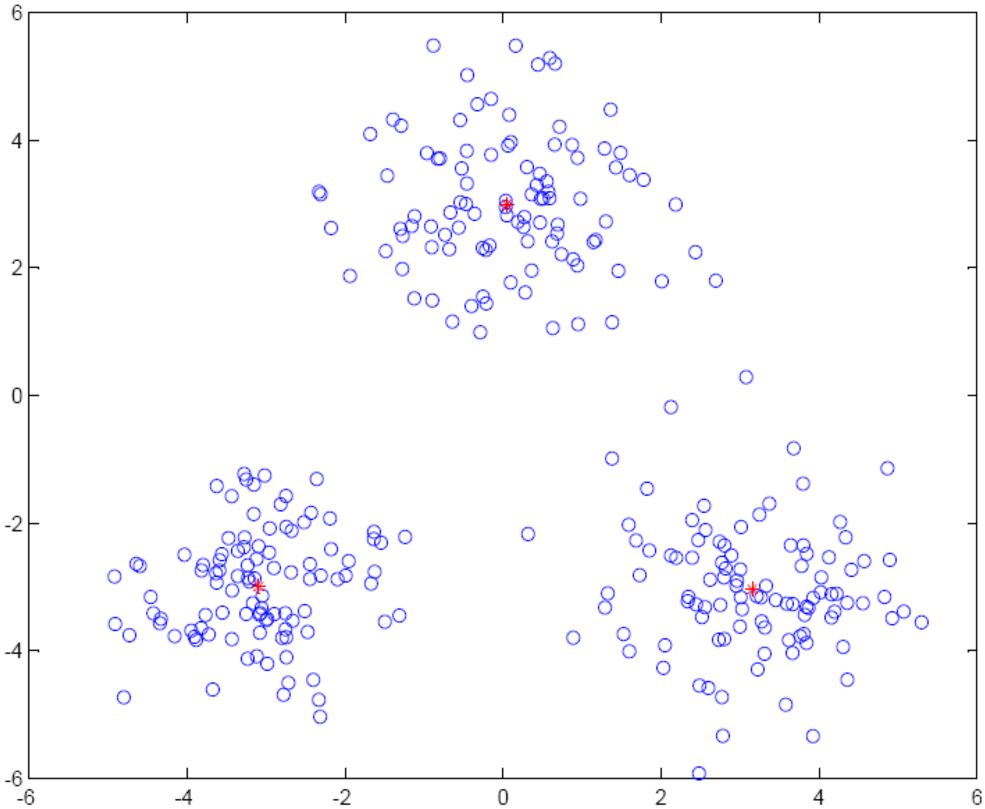
The function chooses the cluster which minimizes the Euclidean distance between the centroid and the
training vector.

3. **Test:** Recompute the centroids μk by taking the mean of the vectors that belong to this cluster. This is
done for all K centroids. If no vector belongs to μk, create a new centroid μk by assigning it a random
vector from the training set. If none of the centroids μk changed from previous iteration, the algorithm
terminates. Otherwise, go back to step 2.
4. **Termination:** From the clustering, the following parameters are found.
• The cluster centroids μk.
• The index kn that indicates which centroid training vector $\mathbf{x}n$ belongs to.
Task 4-A: Write a Matlab function for K-means clustering, kmeans. Use the function prototype
listed below.
Task 4-B: Write a Matlab function to calculate the smallest distance from a feature vector to a
cluster centroid, kmeansd. Use the function prototype listed below.
Task 4-C: In the file task4.mat there is a matrix of feature vectors, X. Find the centroids using the
K-means clustering algorithm, using $K = 3$ clusters. The feature vectors are two-dimensional and
the clustering can therefore be visualized and verified easily.

```
function mu = kmeans (X, K)
%function mu = kmeans (X, K)
%
%INPUT :
%X - feature vectors
%K - number of clusters to estimate
%
%OUTPUT :
%mu - K cluster centers
%
%INFO:
%Estimates K clusters in the feature vectors X.
function [dmin , kmin] = kmeansd (X, mu)
%function [dmin , kmin] = kmeansd (X, mu)
%
%INPUT :
%X - feature vectors
%mu - cluster vectors
%
%OUTPUT :
%dmin - dmin(n) is the minimum distance from vector n to any cluster
%kmin - kmin(n) is the cluster index to which the distance is minimum for vector n
%INFO:
%Calculates the minimum Euclidian distance from the feature
%vectors in X to any cluster vector in mu.
```

5

Figure 3-10: Clustering Plot



The data is taken from the brain hemarise patients in karolinska institute.

3.5 The FUZZY C-means Algorithm

Clustering is the process of grouping feature vectors into classes in the self organized mode. Choosing the cluster centers is crucial to the clustering. Fuzzy clustering plays an important role in solving problems in the areas of pattern recognition and fuzzy model identification. The FCM algorithm is more suited to data that is more or less evenly distributed around the cluster centers. The FCM algorithm lumps the two clusters with natural shapes but close boundaries into a large cluster.

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method (developed by Dunn in 1973 and improved by Bezdek in 1981) is frequently used in pattern recognition. It is based on minimization of the following objective function.

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{i,j}^m \left\| x_i - c_j \right\|^2, 1 \le m \le \infty \text{-----------------------------} > (3.1)$$

where m is any real number greater than 1, $u_{i,j}$ is the degree of membership of x_i in the cluster j, x_i is the i th of d-dimensional measured data, c_j is the d-dimensional center of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the center.

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership $u_{i,j}$ and the cluster centers c_j by:

$$u_{i,j} = 1/\sum_{k=1}^{C}(\|x_i - c_j\|/\|x_i - c_k\|)^{2/(m-1)}, c_j = \sum_{i=1}^{N}(u_{i,j}^m . x_i)/\sum_{i=1}^{N} u_{i,j}^m \text{ ------------------->(3.2)}$$

This iteration will stop when $\max_{i,j} \{|u_{i,j}(k+1) - u_{i,j}(k)|\} < \varepsilon,$ where ε is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m. A saddle point is a point in the domain of a function which is a stationary point but not a local extremum.

The algorithm is described with the following steps:

1. Initialize $U = [u_{i,j}]$ matrix, $U^{(0)}$
2. At k-step: calculae vectors $C^k = [c_j]$ with $U^{(k)}$

$$c_j = \sum_{i=1}^{N}(u_{i,j}^m . x_i)/\sum_{i=1}^{N} u_{i,j}^m$$

3. Update

$$u_{i,j} = 1/\sum_{k=1}^{C}(\|x_i - c_j\|/\|x_i - c_k\|)^{2/(m-1)}$$

4. If $\|U^{(k+1)} - U^k\| < \varepsilon$ then stop; otherwise return to step 2.

3.6 Example-C Means Algorithm-Implementations

Clustering analysis is the process of grouping the unity according with the similar characters in the data concentration. It is widely applied in the field of data $U^{(k)}, U^{(K+1)}$ drawing, image partition, model identification and signal compression. The problem of first class clustering can be treated as restriction optimal problem. Its target is to search the optimal classification of sample set, to make clustering function based on error of intra class and inter-class. (The main difference between intra class and inter-class is the data are pooled to estimate the mean and variance)

C-Mean Algorithm (CMA) is the common way to solve this kind of problem. This algorithm is easy and its convergence speed is very fast but it is sensitive to the initialization condition and it has different clustering result for different initialization value. Clustering method based on genetics algorithm can solve the problems of initialization sensitivity of CMA and has a lot of chance to get the optimal solution. This example in virtue of optimum mechanism vertebrate immune system combining CMA puts forward one kind of hybrid clustering algorithm which keeps the mechanism of individual variety. And also put forward new immune selection strategy. Using this strategy we can overcome the immature convergence phenomenon.

3.7 Hard C-Means and fuzzy C-means algorithm

Hard C- Means (HCM) and Fuzzy C- Mean (FCM) algorithm is Clustering in the data sets the number of known cases, to find the best number of it has been divided, making clustering for optimal performance. The former, each data will belong to one of the only cluster centre; the later, each of the respective data from the cluster centre to determine the membership function.

Consider the n samples of data set X = ($x_1, x_2, x_3, \ldots\ldots x_n$). The number of category c, the HCM and FCM membership function of the sub-matrix denote for

$$M_{cn} = \{u \in R^{c \times n} \mid \sum_{i=1}^{c} u_{ik} = 1, 0 < \sum_{k=1}^{n} u_{ik} < n, \quad u_{ik} \in \{0,1\}; 1 \le i \le c, 1 \le k \le n\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.3)$$

$$M_{fcn} = \{u \in R^{c \times n} \mid \sum_{i=1}^{c} u_{ik} = 1, 0 < \sum_{k=1}^{n} u_{ik} < n, \quad u_{ik} \in [0,1]; 1 \le i \le c, 1 \le k \le n\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.4)$$

And type (3.3) the corresponding function for the clustering performance

$$J_1(u,v) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} D_{ik}{}^2 (v_{i,} x_k) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.5)$$

$$R_1(v) = \sum_{k=1}^{n} \min\{D_{1k}, D_{2k} \ldots\ldots, D_{ck}\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.6)$$

Here: $D_{ik}(v_i, x_k)$ of x_k is the data in the first cluster centre Distance, the general norm for Euclid. And the type (3.4) the corresponding poly type of performance index function for

$$J_m(u,v) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m D_{ik}^2 (v_i, x_k) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.7)$$

Or

$$R_m(v) = \sum_{k=1}^{n} (\sum_{i=1}^{c} D_{ik}^{1/(1-m)})^{1-m} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.8)$$

Here: m is the fuzzy membership weighted index. The value m for greater ambiguity should be divided into the more intense and in general m = 2.

3.8 Base of artificial immune algorithm

Biological immune system of clone selection theory describes the immune system to antigen stimulation shows the basic characteristics of immune response. In principle based on immune algorithm of clone selection the antigen corresponding to the objective function, the antibody corresponding to the optimal solution of objective function. According to fitness of antibodies we can select and estimate the solution and then through the memory cell to retain the local optimal solution in order to maintain the diversity of solutions. Again using similar affinity of antibody to gradually improve the process of optimization and ultimately get the global optimal solution of the problem.

The fitness of antigen makes evaluation based on the affinity of antibody and antigen. Affinity between the antibody, if affinity between certain antibody and antigen is bigger and the affinity between other antibodies and antigen is smaller; the fitness of antibodies is greater. This fitness evaluation way that can maintain the diversity of individuals, improve the search efficient in the local solution space, and effectively get rid of the advantages of part, but the expression of fitness function is difficult to be determined, often through repeatedly testing. This example proposes a new immune selection strategy that is first of all according to antigen affinity to sort descending order of antibody populations and then based on the concentration of each individual to determine the probability of individual selection.

When the concentration is higher than setting threshold value the selection probability $P_s < 1$ otherwise $P_s = 1$.

3.9 C-Means algorithm for artificial immune

Suppose that the number of sample in data set is n, and the dimension of sample is l and the number of clustering is c, then the description of C-mean algorithm of artificial immune as follows:

Entries in the algorithm apply float coding based on clustering and each antibody S is composed of c clustering center. It stands for the float string code.

The function of accommodation value for entries can be defined:

$$f = \frac{1}{1+e} \quad \text{.......................... (3.9)}$$

In this formula e stands for the function of error of mean square. For hard clustering, the definition of e is equal to R_1 in the formula (3,6). For fuzzy clustering, the definition of e is equal to the R_m in formula (3.8).

Entity in colon increment can be defined:

$$N_c = \begin{cases} F(k_1 f(i)), c(i) > T_s \\ F(k_2 f(i)), \end{cases} \quad \text{.................... (3.10)}$$

In the above formula $k_1 = 0.2 k_2$. F denotes the function for getting integer. $f(i)$ is the accommodation value function of entity. $C(i)$ denotes the density of entity, the function is:

$$C(i) = \frac{N_n}{N} \quad \text{........................ (3.11)}$$

N_n denotes the number of entity in domain of entity x_i, domain $\delta(i) = \{x_j \mid s(x_i, x_j) \leq T_d\}, s$ is measurement function of similarity in each entity. T_d is the threshold of similarity.

N denotes the scale of population. The meaning of this definition is that if the density of antibody is lower than the enactment threshold, the multiplication of scale will be large, or else multiplication will be smaller.

Each clone multiplication produces the new entity and each one produces the random mutation with large mutative probability. Adopting random number of average distribution in the range of mutation to replace the original value the range of mutation and fitness of entity are inverse ratio.

Based on (3.9) and (3.10) we calculate the fitness and concentration of new entity then select the certain number of new memory cell according to selection plan of immune and add this cell into the original memory population.

The update method of antibody population is similar with the preservative plan and new got memory population. The stochastic new entity with certain number and the entity will eliminate new individual with low fitness original population according to the certain proportion and create a new generation population.

The steps of artificial immune C-mean clustering algorithm describe the following:

Step1: select clustering method, the measurement function D_{ik} of distance between samples, weight exponent m of fuzzy subjection and other algorithm parameter.

Appoint clustering number c and produce initialization population P by random. Set the expiration condition of algorithm.

Step2: carry out one step of operator of CMA for every entity in population P in order to get the new population P_1.

Step3: select the entity of high fitness according to the certain proportion from population P_1, then cope and produce new entity with the certain number based on (3.10) to get the new population P_2.

Step4: Implement super mutation operator for each individual in population P_2. Adopt the mesmerism operator of high mutation efficiency to get population P_3.

Step5: calculate the fitness and concentration of each individual of P_3, and use the immune selection plan to get new memory cell population M .

Step6: population M and the stochastic generating new individual replace the entity with certain number and low fitness in population P to generate the new population.

Step7: if meet the demand of the expiration condition, algorithm cannot stop, or else one needs to go step 2 and do the iterative circulation.

Simulation experiment

Simulation data includes two synthetic data sets and three actual data sets. The first synthetic data set is planar dimension distribution data set, divided into 5 types, each type having 20 data, and this data adopts 5 groups of stochastic and dependent generation with normal distribution parameter. The second synthetic data set is a single attribute data set which contains 50 data and is divided into 6 types. This data comes from nonlinear function output $y = \left(1 + x_1^{-2} + x_2^{-1.5}\right)^2, x_1 \geq 1, x_2 \leq 5$

Three actual data sets are multiple sclerosis (MS), iris and glass data set separately. MS is the multiple facial sclerosis diagnosis data including 98 data, divided into 2 types, and each data contains 5 attribute values, which are five kinds of guild line measurement of patients. Iris data set includes 150 data, divided into 3 types, and each type has 50 data, and each data has 4 attribute. Glass data set includes 214 data, divided into 6 types and each data contain 9 attribute.

3.9.1 Experiment result

The parameter set of the algorithm is as follows: $k_1 = 1$, $k_2 = 5$, $T_s = 0.1$, $T_d = 0.5$ mutation rate $P_m = 0.4$, selection proportion $P_s = 0.3$, the memory cell occupies 40% in population, and the elimination rate of population is 50%.

The parameter of genetic guideline algorithm is: league choice operators with its scale equal to 2. Cross rate is 0.9 and mutation rate is 0.005 and adopt the preserving strategy with its number of 2. Due to 2 kinds of algorithm are stochastic optimal algorithm, compared the algorithm capability through 30 times' experiment to get the average value. The concluding condition of algorithm is that if fitness of optimal individual in population cannot get improvement after consistent 50 generation, the algorithm ends.

For glass data set, because convergence time of GGA is very long, in order to increase the comparability, compare the GCM from the combination of GGA and CMA with textual algorithm and treat the CMA as the one search operator in GCM. GGA compares with the convergence speed and convergence accuracy in fuzzy clustering of this textual algorithm, which is shown in table 3-1. The comparation of convergence speed and convergence accuracy of hard clustering is shown in table 3-2.

The capability comparation of fuzzy clustering can be seen from table 3-1, when the population scale are the same, the convergence precision of textual algorithm is better than GGA obviously, and convergence speed is fast. It can be seen that GCM and convergence time of textual algorithm is very long for large data set of the glass with this scale, and both average convergence time is very close, but the convergence accuracy of the algorithm in this paper is high that GCM.

Table 3-1: the comparison of convergence speed and convergence accuracy of fuzzy clustering between GGA and textual algorithm.

data set	clustering algorithm	population scale	average convergence time/sec	J2	10^4*mean square	The optimum of objective function
synthetic	GGA	20	122.27	127.186	1.600	127.186
	textual algorithm	20	21.43	127.186	1.280	127.186
single feat	GGM	20	29.23	0.758	600.000	0.731
	textual algorithm	20	4.64	0.731	1.690	0.731
MS	GGM	20	118.26	65766.535	180.000	65766.523
	textual algorithm	20	12.89	65766.453	3.100	65766.453
Iris	GGM	20	126.65	60.582	20.000	60.576
	textual	20	13.73	60.576	1.290	60.576

	algorit hm					
Glass	GGM	20	217.26	154.146	1.770	154.146
	textual algorit hm	20	225.70	154.146	0.808	154.146

Table 3-2: the comparation of convergence speed and convergence accuracy of fuzzy clustering between GGA and textual algorithm.

Data set	clustering algorithm	population scale	average convergence time/s	J2	10^4*mean square	the optimum of objective function
Synthetic	GGA	30	172.87	206.960	0.482	206.459
	Textual algorithm	30	11.46	206.459	0.000	206.459
single feat	GGA	3 0	19.31	0.952	0.066	0.935
	textual algorithm	30	2.07	0.952	0.000	0.935
MS	GGA	30	126.92	82502.999	15.090	82494.655
	Textual algorithm	30	3.23	82494.574	0.000	82494.574
Iris	GGA	30	194.38	78.943	0.002	78.941
	Textual algorithm	30	7.36	78.941	.000	78.941
Glass	GGA	30	68.25	340.280	8.790	336.061
	Textual algorithm	30	72.27	336.968	0.026	336.061

The capability comparation of hard clustering, From Table 3-2 it can be seen that the algorithm in our studies the convergence speed is better than GGA obviously for the first 4 data sets. Due to the algorithm according to our discussion having the character of hard clustering algorithm, it can be guaranteed that the most optimal result can be gotten in every experiment (the standard error is 0). For the fifth data set, the convergence speed of GCM is a bit faster than textual algorithm, but accuracy of textual algorithm is still better.

Chapter 4: Implementation

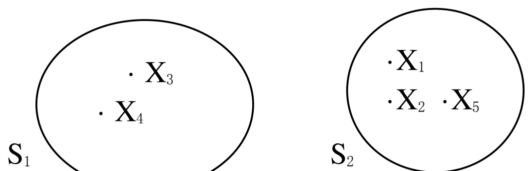
In this chapter we have discussed two basic examples along with two other practical examples. In the basic examples all the calculations are given so that readers can understand the practical examples properly.

4.1 Example 1

Let X={X1,X2,X3,X4,X5} , there are three elements divided in two subsets which are non empty by matrices.

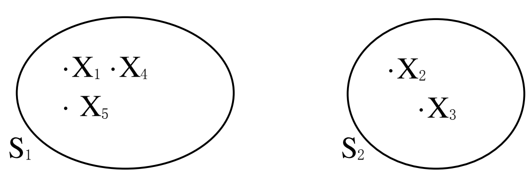
$$U1 = \begin{array}{c c} & \begin{array}{ccccc} X_1 & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{array}{c} S_1 \\ S_2 \end{array} & \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$$S_1 = \{X_3, X_4\},\ \ S_2 = \{X_1, X_2, X_5\}$$



$$U2 = \begin{array}{c c} & \begin{array}{ccccc} X_1 & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{array}{c} S_1 \\ S_2 \end{array} & \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

$$S_1 = \{X_1, X_4, X_5\},\ \ S_2 = \{X_2, X_3\}$$



Each element belongs only to one clustering

$$X = \{X_1, X_2, X_3, X_4, X_5\}$$

$$\tilde{u} = \begin{array}{c c} & \begin{array}{ccccc} X_1 & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{array}{c} S_1 \\ S_2 \end{array} & \begin{bmatrix} 0 & 0.25 & 0.5 & 0.75 & 1 \\ 1 & 0.75 & 0.5 & 0.25 & 0 \end{bmatrix} \\ \sum = & \begin{array}{ccccc} 1 & 1 & 1 & 1 & 1 \end{array} \end{array}$$

Where $\hat{S}_1 = 0/X_1 + 0.25/X_2 + 0.5/X_3 + 0.75/X_4 + 1/X_5$

$\hat{S}_2 = 1/X_1 + 0.75/X_2 + 0.5/X_3 + 0.25/X_4 + 0/X_5$

We use distance between two objects X_k and X_l as the function $d : X \times X \rightarrow R$ such that

$d(X_k, X_l) = d_{kl} \geq 0$

$d(X_k, X_l) = 0 \;\; if \;\; X_k = X_l$

$d(X_k, X_l) = d(X_l, X_k)$

Let $X = \{X_1, ..., X_n\}$, V_{cn} is the set of all real $c \times n$ matrices, where $l \leq c \leq n$ is an integer.

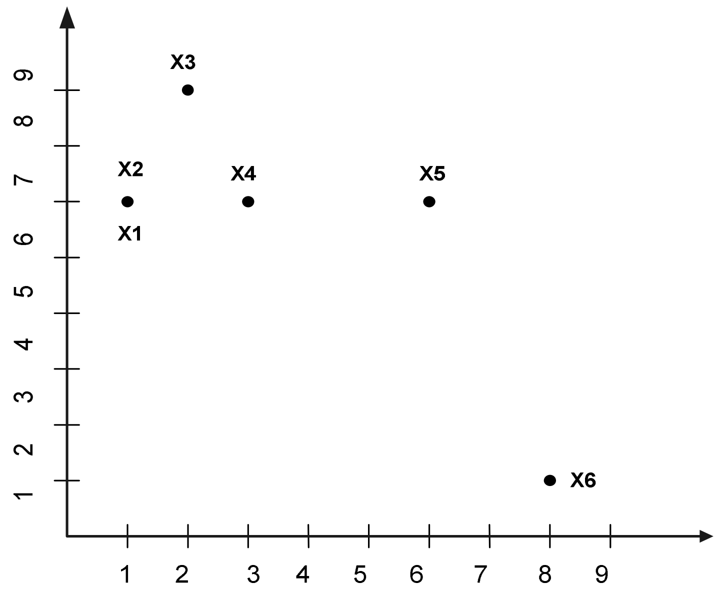
The matrix $u = (u_{ik}) \subset V_{cn}$ is called a crisp c-partition if

1. $u_{ik} \in \{0, 1\} \quad 1 \leq i \leq c$
2. $\sum_{i=1}^{c} u_{ik} = 1$
3. $0 < \sum_{k=1}^{n} u_{ik} < n$

4.2 Example 2

Here we have discussed the second basic example along with the whole algorithm. But in this example we have taken different parameters into account.

$X = \{(1, 6), (1, 6), (2, 8), (3, 6), (6,6), (8, 1)\}$



k=6, c=2, m=1

$$\begin{array}{cc} & \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \tilde{u} = \begin{array}{c} \tilde{S}_1 \\ \tilde{S}_2 \end{array} & \begin{bmatrix} 0.6 & 0.6 & 0.5 & 0.4 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.5 & 0.6 & 0.6 & 0.6 \end{bmatrix} \end{array}$$

$$V_1 = \frac{0.6*(1,6)+0.6*(1,6)+0.5*(2,8)+0.4*(3,6)+0.4*(6,6)+0.4*(8,1)}{0.6+0.6+0.5+0.4+0.4+0.4}$$

$$= \frac{(9,\ 16.4)}{2.9} = (3.103,\ 5.655)$$

$$V_2 = \frac{0.4*(1,6)+0.4*(1,6)+0.5*(2,8)+0.6*(3,6)+0.6*(6,6)+0.6*(8,1)}{0.6+0.6+0.5+0.4+0.4+0.4}$$

$$= \frac{(12,\ 16.6)}{2.9} = (3.871,\ 5.355)$$

$d_{11}=d(x_1, v_1)=\sqrt{(3.103-1)^2+(5.655-6)^2}=2.131$

$d_{12}=d(x_2, v_1)=\sqrt{(3.103-1)^2+(5.655-6)^2}=2.131$

$d_{13}=d(x_3, v_1)=\sqrt{(3.103-2)^2+(5.655-8)^2}=2.591$

$d_{14}=d(x_4, v_1)=\sqrt{(3.103-3)^2+(5.655-6)^2}=0.360$

$d_{15}=d(x_5, v_1)=\sqrt{(3.103-6)^2+(5.655-6)^2}=2.917$

$d_{16}=d(x_6, v_1)=\sqrt{(3.103-8)^2+(5.655-1)^2}=6.756$

$d_{21}=d(x_2, v_1)=\sqrt{(3.871-1)^2+(5.355-6)^2}=2.943$

$d_{22}=d(x_2, v_2)=\sqrt{(3.871-1)^2+(5.355-6)^2}=2.943$

$d_{23}=d(x_2, v_3)=\sqrt{(3.871-2)^2+(5.355-8)^2}=3.240$

$d_{24}=d(x_2, v_4)=\sqrt{(3.871-3)^2+(5.355-6)^2}=1.084$

$d_{25}=d(x_2, v_5)=\sqrt{(3.871-6)^2+(5.355-6)^2}=2.225$

$d_{26}=d(x_2, v_6)=\sqrt{(3.871-8)^2+(5.355-1)^2}=6.001$

$$\tilde{u}^1{}_{11}=\frac{\frac{1}{d(X1,V1)}}{\frac{1}{d(X1,V1)}+\frac{1}{d(X1,V2)}}=\frac{\frac{1}{2.131}}{\frac{1}{2.131}+\frac{1}{2.943}}=\frac{2.943}{5.074}=0.580$$

$$\tilde{u}^1{}_{12}=\frac{\frac{1}{d(X2,V1)}}{\frac{1}{d(X2,V1)}+\frac{1}{d(X2,V2)}}=\frac{\frac{1}{2.131}}{\frac{1}{2.131}+\frac{1}{2.943}}=\frac{2.943}{5.074}=0.580$$

$$\tilde{u}^1{}_{13}=\frac{\frac{1}{d(X3,V1)}}{\frac{1}{d(X3,V1)}+\frac{1}{d(X3,V2)}}=\frac{\frac{1}{2.591}}{\frac{1}{2.591}+\frac{1}{3.240}}=\frac{3.240}{5.831}=0.556$$

$$\tilde{u}^1{}_{14}=\frac{\frac{1}{d(X4,V1)}}{\frac{1}{d(X4,V1)}+\frac{1}{d(X4,V2)}}=\frac{\frac{1}{0,360}}{\frac{1}{0.360}+\frac{1}{1.084}}=\frac{1.084}{1.444}=0.751$$

$$\tilde{u}^1{}_{15}=\frac{\frac{1}{d(X5,V1)}}{\frac{1}{d(X5,V1)}+\frac{1}{d(X5,V2)}}=\frac{\frac{1}{2.917}}{\frac{1}{2.917}+\frac{1}{2.225}}=\frac{2.225}{5.142}=0.433$$

$$\tilde{u}^1{}_{16}=\frac{\frac{1}{d(X6,V1)}}{\frac{1}{d(X6,V1)}+\frac{1}{d(X6,V2)}}=\frac{\frac{1}{6.756}}{\frac{1}{6.756}+\frac{1}{6.001}}=\frac{6.001}{12.757}=0.470$$

$$\tilde{u}^1{}_{21}=\frac{\frac{1}{d(X1,V2)}}{\frac{1}{d(X1,V1)}+\frac{1}{d(X1,V2)}}=\frac{\frac{1}{2.943}}{\frac{1}{2.131}+\frac{1}{2.943}}=\frac{2.131}{5.074}=0.420$$

$$\tilde{u}^1_{22}=\frac{\frac{1}{d(X2,V2)}}{\frac{1}{d(X2,V1)}+\frac{1}{d(X2,V2)}}=\frac{\frac{1}{2.943}}{\frac{1}{2.131}+\frac{1}{2.943}}=\frac{2.131}{5.074}=0.420$$

$$\tilde{u}^1_{23}=\frac{\frac{1}{d(X3,V2)}}{\frac{1}{d(X3,V1)}+\frac{1}{d(X3,V2)}}=\frac{\frac{1}{3.240}}{\frac{1}{2.591}+\frac{1}{3.240}}=\frac{2.591}{5.831}=0.444$$

$$\tilde{u}^1_{24}=\frac{\frac{1}{d(X4,V2)}}{\frac{1}{d(X4,V1)}+\frac{1}{d(X4,V2)}}=\frac{\frac{1}{1.084}}{\frac{1}{0.360}+\frac{1}{1.084}}=\frac{0.360}{1.444}=0.249$$

$$\tilde{u}^1_{25}=\frac{\frac{1}{d(X5,V2)}}{\frac{1}{d(X5,V1)}+\frac{1}{d(X5,V2)}}=\frac{\frac{1}{2.225}}{\frac{1}{2.917}+\frac{1}{2.225}}=\frac{2.917}{5.142}=0.567$$

$$\tilde{u}^1_{26}=\frac{\frac{1}{d(X6,V2)}}{\frac{1}{d(X6,V1)}+\frac{1}{d(X6,V2)}}=\frac{\frac{1}{6.001}}{\frac{1}{6.756}+\frac{1}{6.001}}=\frac{6.756}{12.757}=0.530$$

$$\tilde{u}^1 = \begin{array}{c} \\ \tilde{S}_1 \\ \tilde{S}_2 \end{array}\begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \begin{bmatrix} 0.580 & 0.580 & 0.556 & 0.751 & 0.433 & 0.470 \\ 0.420 & 0.420 & 0.444 & 0.249 & 0.467 & 0.530 \end{bmatrix} \end{array}$$

$$V_1^1=\frac{0.580*(1,6)+0.580*(1,6)+0.556*(2,8)+0.751*(3,6)+0.433*(6,6)+0.47*(8,1)}{0.580+0.580+0.556+0.751+0.433+0.47}$$
$$=\frac{(10.883,\ 18.982)}{3.37}=(3.229,\ 5.633)$$

$$V_2^1=\frac{0.420*(1,6)+0.420*(1,6)+0.444*(2,8)+0.249*(3,6)+0.567*(6,6)+0.53*(8,1)}{0.420+0.420+0.444+0.249+0.567+0.53}$$
$$=\frac{(10.117,\ 14.081)}{2.63}=(3.847,\ 5.330)$$

$$d_{11}=d(x_1,\ v^1_1)=\sqrt{(3.229-1)^2+(5.633-6)^2}=2.259$$

$d_{12}=d(x_2, v^1{}_1)=\sqrt{(3.229-1)^2+(5.633-6)^2}$ =2.259

$d_{13}=d(x_3, v^1{}_1)=\sqrt{(3.229-2)^2+(5.633-8)^2}$ =2.667

$d_{14}=d(x_4, v^1{}_1)=\sqrt{(3.229-3)^2+(5.655-6)^2}$ =0.433

$d_{15}=d(x_5, v^1{}_1)=\sqrt{(3.229-6)^2+(5.633-6)^2}$ =2.795

$d_{16}=d(x_6, v^1{}_1)=\sqrt{(3.229-8)^2+(5.633-1)^2}$ =6.650

$d_{21}=d(x_1, v^1{}_2)=\sqrt{(3.847-1)^2+(5.330-6)^2}$ =2.925

$d_{22}=d(x_2, v^1{}_2)=\sqrt{(3.847-1)^2+(5.330-6)^2}$ =2.925

$d_{23}=d(x_3, v^1{}_2)=\sqrt{(3.847-2)^2+(5.330-8)^2}$ =3.247

$d_{24}=d(x_4, v^1{}_2)=\sqrt{(3.847-3)^2+(5.330-6)^2}$ =1.080

$d_{25}=d(x_5, v^1{}_2)=\sqrt{(3.847-6)^2+(5.330-6)^2}$ =2.255

$d_{26}=d(x_6, v^1{}_2)=\sqrt{(3.847-8)^2+(5.330-1)^2}$ =6.000

$$\tilde{u}^2{}_{11}=\frac{\frac{1}{d(X1,V^11)}}{\frac{1}{d(X1,V^11)}+\frac{1}{d(X1,V^12)}}=\frac{\frac{1}{2.259}}{\frac{1}{2.259}+\frac{1}{2.925}}=\frac{2.925}{5.184}=0.564$$

$$\tilde{u}^2{}_{12}=\frac{\frac{1}{d(X2,V^11)}}{\frac{1}{d(X2,V^11)}+\frac{1}{d(X2,V^12)}}=\frac{\frac{1}{2.259}}{\frac{1}{2.259}+\frac{1}{2.925}}=0.564$$

$$\tilde{u}^2{}_{13}=\frac{\frac{1}{d(X3,V^11)}}{\frac{1}{d(X3,V^11)}+\frac{1}{d(X3,V^12)}}=\frac{\frac{1}{2.667}}{\frac{1}{2.667}+\frac{1}{3.247}}=\frac{3.247}{5.914}=0.549$$

$$\tilde{u}^2{}_{14}=\frac{\frac{1}{d(X4,V^11)}}{\frac{1}{d(X4,V^11)}+\frac{1}{d(X4,V^12)}}=\frac{\frac{1}{0.433}}{\frac{1}{0.433}+\frac{1}{1.080}}=\frac{1.080}{1.513}=0.714$$

$$\tilde{u}^2{}_{15}=\frac{\frac{1}{d(X5,V^11)}}{\frac{1}{d(X5,V^11)}+\frac{1}{d(X5,V^12)}}=\frac{\frac{1}{2.795}}{\frac{1}{2.795}+\frac{1}{2.255}}=\frac{2.255}{5.05}=0.447$$

$$\tilde{u}^2_{16} = \frac{\frac{1}{d(X6,V^1 1)}}{\frac{1}{d(X6,V^1 1)} + \frac{1}{d(X6,V^1 2)}} = \frac{\frac{1}{6.650}}{\frac{1}{6.650} + \frac{1}{6.000}} = \frac{6.000}{12.650} = 0.474$$

$$\tilde{u}^2_{21} = \frac{\frac{1}{d(X1,V^1 2)}}{\frac{1}{d(X1,V^1 1)} + \frac{1}{d(X1,V^1 2)}} = \frac{\frac{1}{2.925}}{\frac{1}{2.259} + \frac{1}{2.925}} = \frac{2.259}{5.184} = 0.436$$

$$\tilde{u}^2_{22} = \frac{\frac{1}{d(X2,V^1 2)}}{\frac{1}{d(X2,V^1 1)} + \frac{1}{d(X2,V^1 2)}} = \frac{\frac{1}{2.925}}{\frac{1}{2.259} + \frac{1}{2.925}} = 0.436$$

$$\tilde{u}^2_{23} = \frac{\frac{1}{d(X3,V^1 2)}}{\frac{1}{d(X3,V^1 1)} + \frac{1}{d(X3,V^1 2)}} = \frac{\frac{1}{3.247}}{\frac{1}{2.667} + \frac{1}{3.247}} = \frac{2.667}{5.914} = 0.451$$

$$\tilde{u}^2_{24} = \frac{\frac{1}{d(X4,V^1 2)}}{\frac{1}{d(X4,V^1 1)} + \frac{1}{d(X4,V^1 2)}} = \frac{\frac{1}{1.080}}{\frac{1}{0.433} + \frac{1}{1.080}} = \frac{0,443}{1.513} = 0.286$$

$$\tilde{u}^2_{25} = \frac{\frac{1}{d(X5,V^1 2)}}{\frac{1}{d(X5,V^1 1)} + \frac{1}{d(X5,V^1 2)}} = \frac{\frac{1}{2.255}}{\frac{1}{2.795} + \frac{1}{2.255}} = \frac{2.795}{5.05} = 0.553$$

$$\tilde{u}^2_{26} = \frac{\frac{1}{d(X6,V^1 2)}}{\frac{1}{d(X6,V^1 1)} + \frac{1}{d(X6,V^1 2)}} = \frac{\frac{1}{6.000}}{\frac{1}{6.650} + \frac{1}{6.000}} = \frac{6.650}{12.650} = 0.526$$

$$\tilde{u}^2 = \begin{array}{c} \\ \tilde{S}_1 \\ \tilde{S}_2 \end{array} \begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \begin{bmatrix} 0.564 & 0.564 & 0.549 & 0.714 & 0.447 & 0.474 \\ 0.436 & 0.436 & 0.451 & 0.286 & 0.553 & 0.526 \end{bmatrix} \end{array}$$

$$V_1^2 = \frac{0.564*(1,6)+0.564*(1,6)+0.549*(2,8)+0.714*(3,6)+0.447*(6,6)+0.474*(8,1)}{0.564+0.564+0.549+0.714+0.447+0.474}$$

$$= \frac{(10.842,\ 18.6)}{3.312} = (3.274,\ 5.616)$$

$$V_2^2 = \frac{0.436*(1,6)+0.436*(1,6)+0.451*(2,8)+0.286*(3,6)+0.553*(6,6)+0.526*(8,1)}{0.436+0.436+0.451+0.286+0.553+0.526}$$

$$= \frac{(10.158,\ 14.4)}{2.688} = (3.779,\ 5.357)$$

$d_{11}=d(x_1, v^3{}_1)=\sqrt{(3.274-1)^2+(5.616-6)^2}$ =2.306

$d_{12}=d(x_2, v^3{}_1)=\sqrt{(3.274-1)^2+(5.616-6)^2}$ =2.306

$d_{13}=d(x_3, v^3{}_1)=\sqrt{(3.274-2)^2+(5.616-8)^2}$ =2.703

$d_{14}=d(x_4, v^3{}_1)=\sqrt{(3.274-3)^2+(5.616-6)^2}$ =0.472

$d_{15}=d(x_5, v^3{}_1)=\sqrt{(3.274-6)^2+(5.616-6)^2}$ =2.753

$d_{16}=d(x_6, v^3{}_1)=\sqrt{(3.229-8)^2+(5.616-1)^2}$ =6.606

$d_{21}=d(x_1, v^3{}_1)=\sqrt{(3.779-1)^2+(5.357-6)^2}$ =2.852

$d_{22}=d(x_2, v^2{}_2)=\sqrt{(3.779-1)^2+(5.357-6)^2}$ =2.852

$d_{23}=d(x_3, v^2{}_2)=\sqrt{(3.779-2)^2+(5.357-8)^2}$ =3.186

$d_{24}=d(x_4, v^2{}_2)=\sqrt{(3.779-3)^2+(5.357-6)^2}$ =1.01

$d_{25}=d(x_5, v^2{}_2)=\sqrt{(3.779-6)^2+(5.357-6)^2}$ =2.312

$d_{26}=d(x_6, v^2{}_2)=\sqrt{(3.779-8)^2+(5.357-1)^2}$ =6.066

$$\tilde{u}^3{}_{11}=\frac{\frac{1}{d(X1,V^21)}}{\frac{1}{d(X1,V^21)}+\frac{1}{d(X1,V^22)}}=\frac{\frac{1}{2.306}}{\frac{1}{2.306}+\frac{1}{2.852}}=\frac{2.852}{5.158}=0.553$$

$$\tilde{u}^3{}_{12}=\frac{\frac{1}{d(X2,V^21)}}{\frac{1}{d(X2,V^21)}+\frac{1}{d(X2,V^22)}}=\frac{\frac{1}{2.306}}{\frac{1}{2.306}+\frac{1}{2.852}}=0.553$$

$$\tilde{u}^3{}_{13}=\frac{\frac{1}{d(X3,V^21)}}{\frac{1}{d(X3,V^21)}+\frac{1}{d(X3,V^22)}}=\frac{\frac{1}{2.703}}{\frac{1}{2.703}+\frac{1}{3.186}}=\frac{3.186}{5.889}=0.541$$

$$\tilde{u}^3_{14}=\frac{\frac{1}{d(X4,V^21)}}{\frac{1}{d(X4,V^21)}+\frac{1}{d(X4,V^22)}}=\frac{\frac{1}{0.472}}{\frac{1}{0.472}+\frac{1}{1.01}}=0.682$$

$$\tilde{u}^3_{15}=\frac{\frac{1}{d(X5,V^21)}}{\frac{1}{d(X5,V^21)}+\frac{1}{d(X5,V^22)}}=\frac{\frac{1}{2.753}}{\frac{1}{2.753}+\frac{1}{2.312}}=0.456$$

$$\tilde{u}^3_{16}=\frac{\frac{1}{d(X6,V^21)}}{\frac{1}{d(X6,V^21)}+\frac{1}{d(X6,V^22)}}=\frac{\frac{1}{6.606}}{\frac{1}{6.606}+\frac{1}{6.066}}=0.479$$

$$\tilde{u}^3_{21}=\frac{\frac{1}{d(X1,V^22)}}{\frac{1}{d(X1,V^21)}+\frac{1}{d(X1,V^22)}}=\frac{\frac{1}{2.852}}{\frac{1}{2.306}+\frac{1}{2.852}}=0.447$$

$$\tilde{u}^3_{22}=\frac{\frac{1}{d(X2,V^22)}}{\frac{1}{d(X2,V^21)}+\frac{1}{d(X2,V^22)}}=\frac{\frac{1}{2.852}}{\frac{1}{2.306}+\frac{1}{2.852}}=0.447$$

$$\tilde{u}^3_{23}=\frac{\frac{1}{d(X3,V^22)}}{\frac{1}{d(X3,V^21)}+\frac{1}{d(X3,V^22)}}=\frac{\frac{1}{3.186}}{\frac{1}{2.703}+\frac{1}{3.186}}=\frac{2.703}{5.889}=0.459$$

$$\tilde{u}^3_{24}=\frac{\frac{1}{d(X4,V^22)}}{\frac{1}{d(X4,V^21)}+\frac{1}{d(X4,V^22)}}=\frac{\frac{1}{1.01}}{\frac{1}{0.472}+\frac{1}{1.01}}=\frac{0.472}{1.482}=0.318$$

$$\tilde{u}^3_{25}=\frac{\frac{1}{d(X5,V^22)}}{\frac{1}{d(X5,V^21)}+\frac{1}{d(X5,V^22)}}=\frac{\frac{1}{2.312}}{\frac{1}{2.312}+\frac{1}{2.753}}=\frac{2.753}{5.065}=0.544$$

$$\tilde{u}^3_{26}=\frac{\frac{1}{d(X6,V^22)}}{\frac{1}{d(X6,V^21)}+\frac{1}{d(X6,V^22)}}=\frac{\frac{1}{6.066}}{\frac{1}{6.066}+\frac{1}{6.606}}=\frac{6.606}{12.672}=0.521$$

$$\begin{array}{cc} & \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \tilde{u}^3 = \begin{array}{c} \tilde{S}_1 \\ \tilde{S}_2 \end{array} & \begin{bmatrix} 0.553 & 0.553 & 0.541 & 0.682 & 0.456 & 0.470 \\ 0.447 & 0.447 & 0.459 & 0.318 & 0.544 & 0.520 \end{bmatrix} \end{array}$$

4.3 Fuzzy clustering analysis and Fuzzy C-means algorithm-Implementations

The impact of hydrological forecasting results mainly performance uncertain factor such as hydrology, meteorology, geography and geology. These factors have a certain amount of random and fuzzy. How to make use of fuzzy analysis to such statistics these uncertain factors, combining with regional runoff model and complementing classification of real-time flood forecasting is the focus of this study. We have chosen Chao ER river of Lu Zhou station as a forecasting station and have tried to use fuzzy math and hydrological forecasting knowledge of history to extract the impact feature from the historical occurred floods and select the most appropriate impact factor to be fuzzy clustering. Fuzzy clustering shows the impact feature of same type of history flooding has same degree of influence to flooding. In real-time forecasting, the impact characteristics of real-time flooding and the characteristics of a sample historical flooding are analyzed by fuzzy clustering to get the membership of a real-time occurrence flooding relative to different categories of flood and do operations in real-time flood forecasting.

4.3.1 Fuzzy clustering analysis

Hydrological forecasting factor is very vague, as a result that the fuzzy cluster analysis is applied to classification of flood forecasting flood categories in order to make classification more realistic, the gotten result is more comprehensive, detailed, reasonable than the traditional classification but also provide the classification base for finial real-time forecasting.

The generally welcomed practical application is the fuzzy clustering method based on the objective function that is, clustering is boiled down into a constrained non-linear programming problem by optimizing solution to obtain the fuzzy partition and cluster of data sets. The method is designed simple a wide range of problem-solving and easy to achieve on the computer. In fuzzy clustering algorithm based on the objective function the algorithm theory of Fuzzy C-MEANS is better and applied more widely.

However, C-MEANS algorithm does not consider the problem of extent impact of the hydrological forecasting impact factor on forecasting results but in the actual application due to the calculation and theoretical reasons of prediction model the contribution of impact factor to the prediction result is non-uniform and in accordance with Principle calculation of the forecasting mode select impact factor one part is measured hydrologic data, the other part is the hydrological properties data. In order to reflect the extent of impact of numerical data on the forecasting results weighted optimization is complemented in sample feature with feature selection techniques Relief.

4.3.2 The fuzzy c-means clustering based on weighted feature

In order to consider each dimensional feature of vector sample of historical flood make contribution to pattern classification, to make use of feature selection technique, to put forward a C-MEAN fuzzy clustering algorithm based on the characteristic weighting, which makes use of feature selection technology features RELIEF algorithm to weighted select feature not only made the classification more effective than an ordinary C-MEAN fuzzy clustering algorithm and can also analyze the contributions and degree of impact of each dimensional characteristics to the classification results and provide a reference for the feature extraction of real-time forecasting.

When the characteristics weight will be directly applied to fuzzy C-MEAN clustering, the objective function fuzzy clustering can be expressed as:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{i,j}^{m} \left\| x_i - c_j \right\|^2 , 1 \le m \le \infty \text{-----------------(4.1)}$$

J_m (where m=1,……n) denotes the weight of each-dimensional characteristics of the sample. When the objective function J_m achieves the minimum, it gets the optimal result of fuzzy clustering, it is needed to point out that the samples of historical flood contain part of flooding property data needing artificial evaluation, a part of property data is not involved in the calculation of flood forecasting model which mainly works in fuzzy classification, therefore the weighted value of this part of attribute characteristics is defined as a relatively small number in order to reduce the impact made by artificial evaluation error of attribute data to forecasting result in real-time forecasting. At last relevance between reference characteristics factor and prediction objection make the posted-weight to all of the preferred characteristics.

Due to the limitations and particularity existing in the acquisition of historical flood samples in general, the big of basin flood is medium-sized floods, low incidence rate to the large flood. In the sample collection process, the difference between the quantities of various types of samples is very large, from the point of view of the calculation method of characteristics weight value. This will be the difference between the quantity will make a large effect on selection of sample characteristics. In order to minimize this impact as much as possible. In accordance with the distribution of all kinds of different samples, set number which is most close to points to be dynamic in RELIEF algorithm.

4.4 Example analysis calculation

In order to test the validity of fuzzy C-means algorithm -weighted based on the characteristics of weight, the proposed accuracy of historical flooding of "CAO ER river" basin is comparably analyzed around the proposed fuzzy by this method.

4.4.1 Basin basic situation

“CAO ER river” is the one of the main tributaries of “QIAN TAN river”, originating from “AN LIANG hill”. “CAO ER river” is fan-shaped river, its south low and high North, LU ZHOU above drainage area reaches 2400 square meters, LU ZHOU hydrological stations obtain the observations of water level, rainfall, evaporation and flow. Upper reaches have more than five times the rain volume sites and three large and medium-sized reservoirs. River basin belongs to sub-humid areas.

Annual precipitation is 1500mm. Flood is primarily caused by storm and rain. And the character of storm in rainy season is that it lasts for a short and has rain intensity

4.4.2 Flooding Feature Selection

The key of FCM algorithm lies the prior given cluster number of C, therefore, this review accords to the iterative self-organizing analysis technology, uses meaning of optimal classification to build function F of inter-class and extra-class distance of Fuzzy Cluster Analysis to achieve the C-MEAN fuzzy optimization algorithm together, to obtain the best cluster sample in this example for Category 5.

Since the prediction model applied the runoff model used in sub-humid areas, therefore firstly select rainfall intensity in greatest time, initial water content of soil, the average rainfall intensity and the initial flow to take part in model calculation which is characterized by basic numerical characteristics of fuzzy clustering. However, the hydrological element influencing formation of the impact of flooding has a strong randomness and fuzziness, and uncertainty predictor factor is very powerful, in order to fully reflect basin rainfall characteristics and distribution of rainfall. In accordance with distribution formation of pre-rainfall and its extend contribution to the flood peak, select three property impact characters such as a rainstorm center, rain trend and property on shape of rain type. In the calculation of fuzzy clustering analysis, it is necessary that the property characteristic numerical value of rainfall changes into a numerical value characteristics. Location of rainstorm center accords the five levels: the upper reaches, middle and upper reaches, middle reaches, the middle and lower reaches, evaluates from 1to 5 according to their degree of influence of flooding peak. Trend of rainfall accords from bottom to top, upstream, midstream, from top to bottom, lower reaches of the five-levels, gives evaluation from 1to 5 according to their degree of influence on the flooding speak. Shape of rain type interpolates evaluation in accordance with the figure 4-1.

The rain characteristics during the largest time based on forecasting calculating period as unity select the greatest periods of rainfall making impact on flood peak. Initial water content of soil refers to the water content of soil during the early period of prediction calculation. The average rainfall intensity is the peak-hour rainfall of rain peak. The initial flow refers to the flow value of early period of forecasting calculation in the forecast section. These characteristics compose series of characteristics samples of the basin historical flood, shown in Table 4- 14-1.

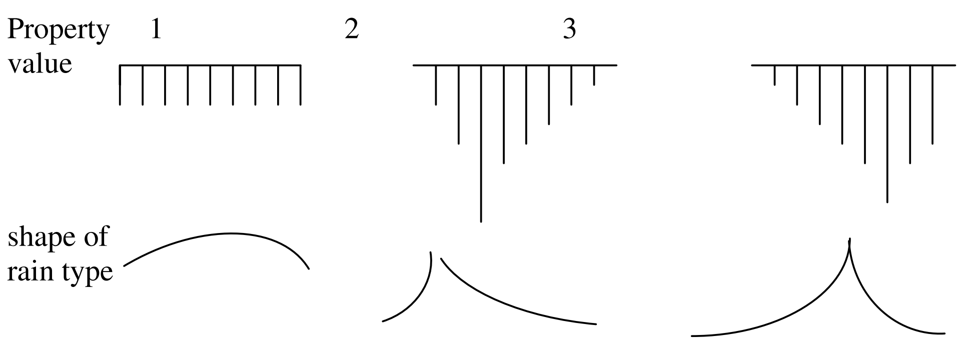


Figure 4-1: Sketch for evaluate property on the shape of rain type

Flood No.	No.	rainstorm centre	rainstorm trend	sharp of rain type	rainfall intensity in maximum period /$(mm*(3h)^{-1})$
860410	1	3	3	1.8	17.9
860708	2	2.8	3	2.9	17.9
860905	3	3	3	2	35.14
860919	4	3	3.2	2.8	12.4
870620	5	1.8	2.4	1.9	24.1
870722	6	3	3	2.5	10.9
870727	7	2.9	3	3	26.8
870908	8	3	3	2.4	18.8
880617	9	3	2	2	20.3
880807	10	3	3	2.9	47.83
880411	11	4	3	2.8	10.76
890522	12	3	3	2	14.27
890616	13	2.5	3	2.2	12.07
890630	14	3	3	2.5	32.28
890723	15	2	2.6	2.4	23.28
890726	16	2.8	3.2	3	10.8
890915	17	1.2	2.6	2.9	50.51
900421	18	2	3	2.5	22.05
900531	19	2	3	2.5	11.8
900830	20	2.8	3	1.5	25.2
900904	21	3	2.5	2.7	10.9
900909	22	1	2.6	2.5	26.2
900912	23	2	3	2.5	14.7
910417	24	3	3	2.9	24.83
910619	25	2	2.2	2.8	11.43
910812	26	3	3.5	2.8	21.86
920622	27	3	2	1.8	9.8
920624	28	2	2.5	2.7	14.2
920701	29	4	2.8	2.7	16.34
920829	30	3.6	4	2.9	23.9
920922	31	3.4	3.8	2.5	36.47
930501	32	2.4	3	2	21.4
930618	33	2.8	2.4	2.8	16.8
930630	34	3	4	1.6	19.63
950703	35	3	3	2	16.55
950428	36	2	3.2	2.2	16.48
950530	37	1.6	2	2.5	33.24
950625	38	1.2	1.2	2.5	18.16
950702	39	1.5	1.6	2.7	25.25
960630	40	3	4.2	1.7	19.8

Table 4- 1: Feature and flood peaks runoff of historical flood sample

Table 4-2: Sample data for rainfall intensity according to optimal classification

Flood No.	No.	initial water content /mm	average rainfall intensity/($mm*h^{-1}$)	initial flow/(m^3*s^{-1})	flow of flooding peak/(m^3*s^{-1})
860410	1	89.1	12.6	74.51	691
860708	2	85	15.7	60.98	532
860905	3	69.8	19.7	38.6	937
860919	4	68.6	7.2	33.37	630
870620	5	68.7	7.2	33.55	1220
870722	6	43.2	7	93.13	592
870727	7	90	16.6	194	2093
870908	8	59.1	13.2	89.4	1490
880617	9	74.3	14.9	48.23	1080
880807	10	28.5	32	71.2	1389
890411	11	51.8	5.3	12.7	660
890522	12	81.6	8.7	241	825
890616	13	42	12.9	29.24	1130
890630	14	90	17.8	162	2030
890723	15	38.3	18.4	52.7	653
890726	16	61.7	7.2	49.38	489
890915	17	66.2	23.6	217.68	2830
900421	18	83.1	8.4	68.2	673
900531	19	66.8	6.8	42	527
900830	20	31.1	19.4	12.7	2770
900904	21	90	6.2	128	997
900909	22	80.1	10.7	313	1420
900912	23	58.6	9.2	130	608
910417	24	89.8	18.5	102.33	1260
910619	25	70.8	10.8	51.67	451
910812	26	40.4	9.7	16.93	395
920622	27	54.7	5.7	28.55	470
920624	28	85	6.9	80.35	1090
920701	29	89.6	4.8	45.95	1120
920829	30	89.6	8.4	66.22	2270
920922	31	69.6	19.6	37.8	2170
930501	32	62.4	10.1	29.32	707
930618	33	90	9.9	68.3	1300
930630	34	90	9.6	50	1680
930703	35	85.2	11.8	215.5	965
950428	36	83.9	7.6	163	1212
950530	37	55.9	18.5	112.29	761
950625	38	80.5	17.9	189	863
950702	39	62.4	11.2	88.7	946
960630	40	60.3	13	25.32	837

Pre-classify samples in according to the optimal classification number of samples, then complement cycle calculation of RELIEF, the calculated results shows on Table 4-3 and Table 4-4

Table 4-3

Attribution feature		
Location of rainstorm	trend of rain	the sharp of rain type
0.02	0.04	0.78

Table 4-4

Value feature			
rainfall intensity in maximum period/(mm*(3h)$^{-1}$)	initial water content /mm	average rainfall intensity/(mm*h^{-1})	initial flow/(m^3*s^{-1})
0.57	0.17	0.31	0.10

According to the calculated result of Table 4-3 and Table4-4, it can be seen that the weight of rainstorm centre of property and rainfall trend is close to zero value. It means that contribution to peak flow is minimal, can be removed from the preferred characteristics. From the application of clustering, analyze the mechanism of the process of flood forecasting assignment, as a result of prediction model applying a distributed model, and model divides each forecasting based sites into a single sub-site to calculate runoff, therefore the characteristics of center of rainstorm and rain trend can reflect from calculation of rainfall by hour in each site. The two features actually including in the prediction basis sites on the mechanism of model, and it can be eliminated after their optimization but the spatial-temporal distribution of rainfall of the whole surface on basin has not fully shown out through hydrological numerical model. Taking consideration into differences between situation of basin flow convergence, it is hard to totally reflect all kinds of difference only responding on rainfall by period in each rainfall sites. Hereby this article by modifying the value of property characteristics, make characteristics more advantageous to the fuzzy clustering analysis of flooding feature after weight.

In the process of calculation of Relief weight, first of all, because of some error existing in setting the numerical characteristics, model parameters are adjustable parameters. Followed by pre-classification does not take consideration into the extent of impact of dimensional characteristics and redundancy features, pre-classification results have definite irrationality. Therefore the weight determination of this part characteristic should also be advised to the correlation coefficient of the forecasting objective results factor and computing mechanism in the prediction model, to determine distribution project of weight value.

Table 4-5: Optimized results of weight of flood feature

sharp of rain type	rainfall intensity in maximum/(mm*(3h)$^{-1}$)	initial water content/mm	average rainfall intensity/(mm*h^{-1})	initial flow/(m^3*s^{-1})
0.1	0.35	0.15	0.2	0.2

According to the best classification number of historical samples, use feature of the optimal selection above, as well as the weight on the sample object of "CAO ER river" basin historical flood to complement fuzzy c-means clustering analysis based on the characteristics weight, analyzed results as shown in Table 4-6.

Table 4-6: Results of fuzzy clustering analysis on weight feature

Number	the No. of diversity of samples
1	2,18,24,28,29,30,33,39
2	4,6,11,16,19,21,23,25,27
3	7,12,22,35,36,38
4	1,5,8,9,13,15,20,26,32,34,40
5	3,10,14,17,31,37

In order to further verify the reliability of the classification, comparing the difference between various types of similar sample with differences between heterogeneous samples, if the ratio between the two was significantly greater than 1.0, it means that these two categories should be merged, while around about 0.9 it means that classification is reasonable, the smaller the ratio is, the more reasonable classification results are. In view of the above classification results, calculate the ratio of the differences between class samples and the class, the calculated results shown in Table 4-7.

Table 4-7: Comparison of difference in a class and among classes

Class	the difference in class sample	the difference among class sample	difference ratio
1	0.53	1.17	0.45
2	0.53	1.46	0.36
3	0.59	1.09	0.54
4	0.71	1.73	0.41
5	0.85	1.16	0.73
Average	0.64	1.32	0.48

According to the results of clustering analysis in table 4, divides “CAO ER river” basin historical 40 types of flooding into five classes, to comprehensively compare calibrated situation of flood parameters with the calibrated situation before clustering, the results shown in Table 4-8. By comparing the analysis, the application of cluster analysis based on weighted feature can not only reflects forecasting effect of each hydrological impact factor of river basin, but also noticeably improves the relevance accuracy of flood forecasting.

Table 4-8: statistics of calibrated results after clustering historical flood samples

Average runoff	flow of flood peaks		average uncertainty coefficient		
	Class Number	error/%	good rate/%	error/%	good rate/%
After clustering					
1	5.32	85.7	7.45	85.7	0.91
2	4.24	100.0	3.92	100.0	0.91
3	4.10	100.0	5.08	100.0	0.94
4	3.78	100.0	3.12	100.0	0.94
5	5.97	83.3	4.29	100.0	0.91
Before clustering	5.86	92.5	5.92	85	

4.5 Conclusion

In this example, fuzzy clustering new algorithm based on weight complements weighted fuzzy C-MEAN cluster analysis for "CAO ER river" basin historical flooding, application of examples shows that the method can not only analyze the extent of contribution of dimensional feature to the classification, effectively select and extract feature of flooding impact factor, and the historical floods can operator fuzzy clustering analysis in accordance with the flooding impact factors, to make use of cluster analysis result for making the real-time forecasting operations. Through the fuzzy and stochastic analysis on flooding impact factor in flood forecasting, on the basis of model forecasting, the method of flood forecasting based on fuzzy clustering is proposed, and the accuracy of the model forecasts is also improved.

Chapter 5: Discussion

5.1 Where do we apply model

Physical process investigation produces models. There are so many examples for models. Few of them are welfare effect on civil disobedience, air plane wing stress and strain interaction, fluoridation's psychological effects on a community and stream pollution effect on population. This physical process generates lots of data depending on various functional factors of Models. By applying mathematics to this process we can represent generated data in terms of mathematical models. We can use these mathematical models to various applications.

5.2 How are model*s* altered

Models can be altered based on available parameters and model types. Altered models have effects on area of utilization. Models can be altered using different views externally and they can also be altered internally by finding new facts about existing models.

5.3 What about different model of the same process

Process can be model in different ways as suggested by various examples available. Different models of physical process provide ways to understand physical process. Using different models we can measure same parameters of process or we can use different models to support different properties of physical process.

5.4 Optimal model

Choosing optimal models always depends on area of application. We should consider different available models of a physical process which is under observation. Optimal model for any given application area are chosen based on existing standard ways. Different models provide different qualities of physical processes. So it's better to consider models depending on properties of Physical process under examination. Every model has different ways representing process using set parameter of that model. So we make use different available models to measure physical process properties.

5.5 Uncertainty

The most exciting and amazing analyses of modern scientific world is to get accurate or approximate results with the degree of uncertainty that effects towards mathematical models. There are several approaches of uncertainty so that our goal is to classify the difference between them and to choose the type of uncertainty that matches to the requirements of our model.

The major causes of uncertainty is due to

1. Inaccurate measurements

2.Random occurrence

3.Vague description

Interestingly there are so many other differentiable solutions, but these three provides a way to the demonstration of uncertainty and fuzzy models. The outcome from the process can be analysed according to the situations relevant to the model. For example the sample of bacteria say y (t), at the time t will obey the law of exponential growth.

y(t)=c(ekt) ----------------------------(5.1)

where c and t are the mathematical and physical parameters of the exponential model described above.

Approximate measurement can cause the uncertainty in models of the physical processes which can be analyzed. If the results of the physical process are in random pattern then we can experience another type of uncertainty. In this case there are certain elements in the process which are not affected by the environmental impression. The evolution of the sequence of identical and independent trails of this experiment is probable in nature so it can be predicted with certainty; these kinds of models are having more then one possibility.

For example a reasonable model of n bernoulli trails is the binomial distribution.

b(n,p,k)=(n:k)pk(1-p)n-k ---------------------------------(5.2)

Where n is the number of trails,k is the number of success in n trails, p is the probability of the success and q is the probability of failure equals to 1-p

We can analyze and observe with sequence of bernoullis trails afforded by binomial distribution are quite different from the inferences we can draw about the objective properties of bacteria growth through law of exponential growth.

Chapter 6: Conclusion

The C-means algorithm is treated as a new search operator so that proposed algorithm obtains the characters of strong ability of part searching and small operation in order to faster the part search optimum speed. Therefore, the textual algorithm can converge to the optimum faster and has higher accuracy. The algorithm can be extended to other clustering model whose objective function can be represented in terms of optimization of clustering centres.

From the features, concepts and properties discussed about pattern recognition it is evident that Fuzzy logic can indeed be part of this vast universe of data thereby finding itself in various applications and measures in relation with pattern recognition.

6.1 Future of technology

The future regarding pattern recognition still looks bright. As long as man's determination to solve every complex problem precisely and perfectly persists, one can never see the horizon beyond i.e. the philosophy that to solve complex problems approximate reasoning and imperfections are good enough and this is where fuzzy logic and its variants, in this case fuzzy data and different patterns, will remain prevalent.

It has been a wonderful experience to learn new vast universe in the form of Fuzzy Logic and having the knowledge of Fuzzy pattern recognition has indeed given a different dimension. The thesis presented has done quite a bit of justice in knowing about this topic right from introduction to fuzzy logic and pattern recognition, the reasons and its uses, to C means algorithms enhancement, segmentation using Fuzzy and finally its applications.

References

[1] K means clustering-algorithms A tutorial on clustering http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html

[2] Sheng W , TuckerA, Liu X. Clustering with Nichinggenetic K-means algorithm Proceedings of Genetic and Evolutionary Computation Conference

Berlin: Springer-Verlag, 2004.162-173.

[3] Maulik U, Bandyopadhyay S. Genetic algorithm-based Clustering technique, PatternRecognition. 1997.

[4] De Castro L N, von Zuben F J. Learning and optimization using the clonal selection principle, IEEE Transactionon Evolutionary Computation, 2002.

[5] ZHANG Honggang, GUO Shenglian. Recent advancement and prospect of hydrological forecasting uncertainty study. J of Shihezl University: Natural Science Edition 2006, 24(I):15-21.

[6] San jose state university, Fuzzy Logic, available at: http://www.sjsu.edu/faculty/watkins/fuzzysets.htm

[7] Fuzzy logic and introduction by Steven D.Kaehler, available at: http://www.seattlerobotics.org/Encoder/mar98/fuz/fl_part4.html#MEMBERSHIP FUNCTIONS

[8] Fuzzifying Clustering algorithms,The case study of major clust, Eugene levner,David pinto,availableat: http://users.dsic.upv.es/~prosso/resources/LevnerEtAl_MICAI07.pdf

[9] A tutorial on clustering algorithms, Fuzzy C-means Clustering, available at: http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html

[10] J. C. Bezdek (1981): "Pattern Recognition with Fuzzy Objective Function Algoritms", Plenum Press, New York

[11] *Pattern Recognition with Fuzzy Objective Function Algorithms* J.C Bezdek 1981 Kluwer Academic Publishers, Norwell, MA, 1981

[12] *Data Mining*. A *Knowledge Discovery Approach* Cios, K.J., Pedrycz, W., Swiniarski, R.W., Kurgan, L.A.. 2007

[13] *Pattern Recognition* S. *Theodoridis*, K. Koutroumbas

[14] Algorithms for Fuzzy Clustering methods in c-means clustering with applications Miyamoto S, Ichihashi H, Honda K.

[15] MATLAB Fuzzy Logic tool box, J.-S. Roger Jang Ned Gulley

[16] K-Means Clustering Tutorial By Kardi Teknomo,PhD
[17] Fuzziness and Probability Based on texts from Jim Bezdek, Didier Dubois, Bart Kosko, and Henri Prade, available at: http://www.neuronet.pitt.edu/~bogdan/research/fuzzy/fvsp/fvsp.html
[18] flexible completion of workflow activities, Georg Peters-Munich University of Applied Sciences, Department of Computer Science, Roger Tagg- University of South Australia, School of Computer and Information Science