A CLOUD BASED STUDENT SELF ASSESSMENT TOOL USING FUZZY LOGIC

P.Tamizhchelvi*

S.Muthukumar**

ABSTRACT

The Quality of candidates admitted into any higher Institution affects the level of research & training in the Institution & thereby affects the overall development of the country as a whole since these candidates eventually become key players in the affairs of the country. Therefore, it would be appropriate if an analysis is made on the characteristics of students to identify their weaknesses at their entry into the institution itself & suggestions can be offered for their betterment in the academic activities. We have already designed a tool using fuzzy approach to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information. Here, we have taken steps to design a Cloud based tool using Fuzzy logic to predict the academic performances of any student in the world.

Keywords: Fuzzy, Cloud, Hadoop framework, Linguistic variable, Defuzzification.

*Associate Professor & Head, Department of computer science and Information technology Ayya Nadar Janaki Ammal College, Sivakasi

** Department of computer science and Information technology, Ayya Nadar Janaki Ammal College,Sivakasi

INTRODUCTION

In USA in 1966, the federal government directed the Coleman Commission to study the factors which determine educational results of students. They compared many schools around the USA and attempted to correlate student performance with all factors which are thought to improve education

Our focus is to mould our students to have high caliber when they leave the campus and become positive role models who are ready to face the challenges of traditional and non-traditional education programs that require a high degree of flexibility, Observation power, memory power, Attitude and Presence of mind and so on.

A fuzzy logic approach may help to predict the academic capabilities of the students, to appraise their weaknesses and to take measures in improving their overall personality.

About Fuzzy Logic

Most of our traditional tools for formal modeling, reasoning and computing are crisp, deterministic, and precise in character. By crisp we mean dichotomous, that is yes-or-no type rather than more or less type. In Conventional dual logic, a statement can be true or false and nothing in between. In set theory an element can either belong to a set or not .For factual models or modeling languages, two major complications arise

     1. Real situations are very often not crisp and deterministic and they cannot be described precisely.

     2. The complete description of a real system would require far more detailed data than a human being could ever recognize simultaneously, process and understand.

Real Situations are very often uncertain or vague in number of ways. We call these types of uncertainty or vagueness as Fuzziness.

Fuzzy Set Theory

A fuzzy set is any set that allows its members to have different grades of membership (membership function) in the interval [0, 1].It provides a strict mathematical framework in which vague conceptual phenomena can be precisely and rigorously studied. It can also be considered as a modeling language well suited for situations in which fuzzy relations, criteria, and phenomena exist.

A Fuzzy set is defined as follows

 If X is a universe of Discourse and x is particular element of X,then a fuzzy set A defined on X may be written as a collection of ordered pairs.

$$A=\{(x,\mu\tilde{A}(x)),x\in X\}$$

Where each pair (x, μÃ(x)) is called a singleton.
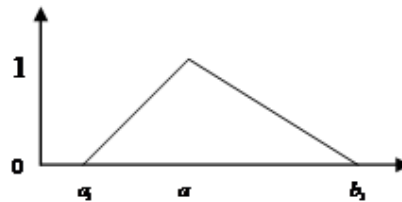
Membership functions

The membership function is a graphical representation of the magnitude of participation of each input. It associates a weighting with each of the inputs that are processed, define functional overlap between inputs, and ultimately determines an output response. The rules use the input membership values as weighting factors to determine their influence on the fuzzy output sets of the final output conclusion. Once the functions are inferred, Ůnscaled, and combined, they are defuzzified into a crisp output which drives the system. There are different memberships functions associated with each input and output response.

Types

There are several types in membership functions. Some of them are

- Triangular membership functions
- Bell membership functions
- Trapezoidal membership functions
- Triangular membership function

The simplest membership function is the triangular membership function. It's nothing more than a collection of three points forming a triangle.


Trapezoidal membership function

The trapezoidal membership function has a flat top and it is just a truncated triangle curve. These straight line membership functions have the advantage of simplicity.
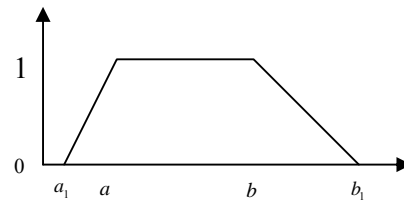
$$\mu_A(x) = \begin{cases} 0 & for \ x \leq a_1 \\ \frac{x-a_1}{a-a_1} & for \ a_1 \leq x \leq a \\ 1 & for \ a \leq x \leq b \\ \frac{b_1-x}{b_1-b} & for \ b \leq x \leq b_1 \\ 0 & for \ b_1 \leq x \end{cases}$$

Bell membership function

The bell membership function is specified by three parameters. Because of their smoothness and concise notation, bell membership functions are popular methods for specifying fuzzy sets. These curves have the advantage of being smooth and nonzero at all points.
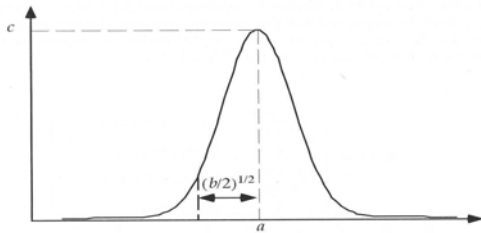

$$A(x) = c e^{-\frac{(x-a)^2}{b}}$$

Linguistic variables

A variable whose values are not numbers but words or sentences are called linguistic variables. Small, Medium, Large are examples of linguistic variables. The states of each linguistic variable are expressed by linguistic terms of a base variable. The values of the base variables are within specific range of real numbers. In a linguistic variable, linguistic terms representing approximate values of a base variable are captured by appropriate fuzzy numbers.

Operation of Fuzzy System
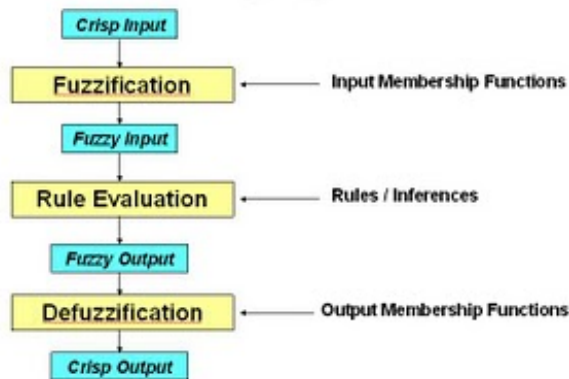
About Cloud Computing

Computing is being transformed to a model consisting of services that are commoditized and delivered in a manner similar to traditional utilities such as water, electricity, gas, and telephony. In such a model, users access services based on their requirements without regard to where the services are hosted or how they are delivered.

With the increasing popularity of cloud computing, Hadoop has become a widely used open source cloud computing framework for large scale data processing. In this paper, we are going to use the Hadoop cloud computing framework to develop our application.

Hadoop is an open source implementation of a subset of the Google system. The three Hadoop components that are analogous to Google's components are

1. Hadoop's MapReduce
2. Hadoop's Distributed File System
3. The HBase storage system for sparse structured data.

In this paper, we are going to use the HCatalog & Hadoop MapReduce framework programming model and execution environment for processing the large data sets in a reliable, fault-tolerant manner.

Solving our problem using Fuzzy logic on cloud framework

The problem to be solved in fuzzy logic should undergo the following steps.

- Collection of crisp input
- Mapping Inputs on Hadoop
- Fuzzification
- Applying Rule Base
- Defuzzification

A) Collection of Crisp Input

A conventional set, wherein the degree of membership of any object in the set is either 0 or 1. It is a term for the mathematical model that does not use _ fuzzy logic. Fuzzy logic is a mathematical or computer program that attempts to approximate the ambiguity inherent to human reasoning. Thus, a crisp set is any program that organizes data into neat categories where each data point can belong to one, and only one, category.

B) Mapping Inputs on Hadoop.

The Crisp input kept in the form of file is uploaded into Hadoop using HCatalog.

c) Fuzzification

The fuzzification comprises the process of transforming crisp values into grades of membership for linguistic terms of fuzzy sets. The membership function is used to associate a grade to each linguistic term.

D)Rule Base Engine

A rule-based system consists of if-then rules, a bunch of facts, and an interpreter controlling the application of the rules, given the facts. These if-then rule statements are used to formulate the conditional statements that comprise the complete knowledge base. A single if-then rule assumes the form 'if x is A then y is B' and the if-part of the rule 'x is A' is Called the antecedent or premise, while the then-part of the rule 'y is B' is called the consequent or conclusion. There are two broad kinds of inference engines used in rule-based systems: forward chaining and backward chaining systems. In a forward chaining system, the initial facts are processed first, and keep using the rules to draw new conclusions given those facts. In a backward chaining system, the hypothesis (or solution/goal) we are trying to reach is processed first, and keep looking for rules that would allow concluding that hypothesis.

E)Defuzzification

Defuzzification is a reverse process of fuzzification. Defuzzification is used to obtain a crisp output. Since technical process require crisp control actions,a procedure that generates crisp value out of one or more fuzzy output.The crisp value to be chosen should generally be an element of the supports of the fuzzy sets to be defuzzified.

 Several methods available in defuzzification are

- Centroid Method
- Centre of Sums Method
- Mean of Maxima

Centroid Method

The most popular one is the centroid technique. It finds the point where a vertical line would slice the aggregate set into two equal masses.Centroid defuzzification method finds a point representing the centre of gravity of the fuzzy set A, on the interval, on the interval, ab.A reasonable estimate can be obtained by calculating it over a sample of points.

Center of Sums

In the Center of Sums (CoS) defuzzification method, the fuzzy logic controller first calculates the geometric center of area for each membership function, as the following figure illustrates:
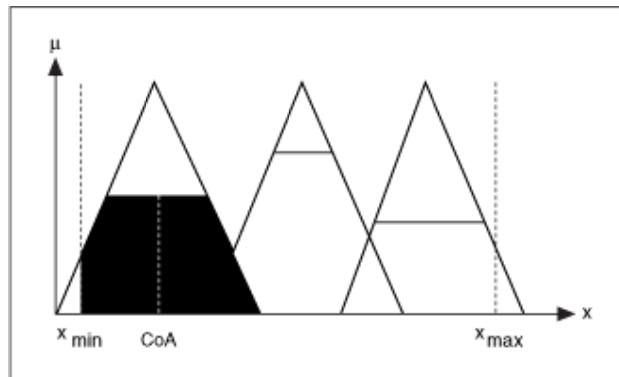
The fuzzy logic controller then uses the following equation to calculate a weighted average of the geometric center of area for all membership functions.

$$x_{final} = \frac{CoA_1 area_1 + CoA_2 area_2 + \ldots + CoA_n area_n}{area_1 + area_2 + \ldots + area_n}$$

where CoA_n is the geometric center of area of the scaled membership function n and $area_n$ is the area of the scaled membership function.

Mean of Maximum

In the Mean of Maximum (MoM) defuzzification method, the fuzzy logic controller first identifies the scaled membership function with the greatest degree of membership. The fuzzy logic controller then determines the typical numerical value for that membership function. The typical numerical value is the mean of the numerical values corresponding to the degree of membership at which the membership function was scaled.

The following figure illustrates the MoM defuzzification method:

Mean of Typical Values for
Maximum Membership Function
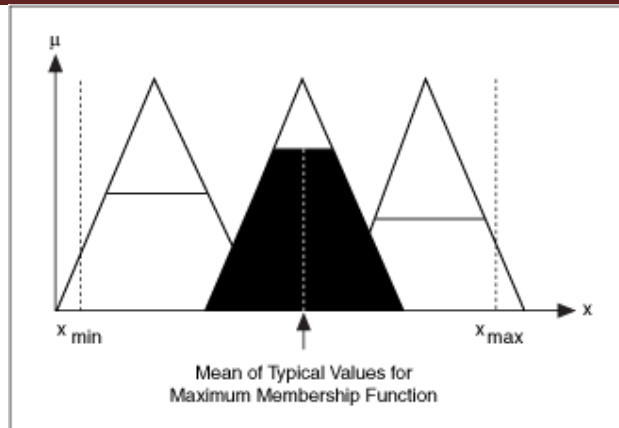
Problem Description

Sometimes the fresher student failed to perform well, even though his/her academics were good. To handle these types of situations, we have to predict the academic performance of the students. It is very much useful in identifying the students' skills and it gives suggestions to improve their performance. We are going to provide this via Cloud based framework.

Crisp Input

We have used 16 base variables which would mainly affect the students' performance. According to their importance in students' studies, we have assigned membership values to them.

The 16 Parameters used in our study are as follows:

- Memory Power
- Observing Power
- Numerical Ability
- Practical Knowledge
- Positive Attitude
- communication skills
- Confidence
- Hard Work
- Decision Making
- Creativity Power
- Technical Skills
- Analytical skills
- Logical thinking
- Team Skills

- Punctuality

etc.

The first main two attributes are memory power and observing power because, without these attributes a student cannot perform well.

We assign the membership values to these parameters.

Sno	Base Variables	Membership Values
1	Memory Power	0.99
2	Observing Power	0.98
3	Numeric Ability	0.91
4	Practical knowledge	0.95
5	Positive Attitude	0.97
6	communication skills	0.85
7	Confidence	0.90
8	Hard work	0.84
9	Presence of Mind	0.96
10	Decision Making	0.60
11	Creativity Power	0.65
12	Technical Skills	0.93
13	Analytical skills	0.61
14	Logical thinking	0.64

15	Team Skills	0.35
16	Punctuality	0.60

To test Memory power

Here, we kept a set of 20 items in one place. The students were asked to observe them for 10 seconds & later they were asked to list out them. According to the number of items listed out by them correctly, we gave values.

Testing Communication Skill

A good communication skill makes the student to convey his/her thoughts clearly to others. We predict this skill by conducting the Group Discussion. A positive attitude is the seed for positive results

B) Mapping Inputs on Hadoop.

Here, we are going to use the Hadoop cloud computing framework to deploy our application on it. Hadoop allows users to submit large data for processing on many nodes using MapReduce model.

Hadoop's MapReduce input is normally file – based. We need to change our crisp input to MapReduce input formats. Hcatalog helps to upload our file into HDFS. Finally, we process with those inputs and apply the following steps.

C) Fuzzification

We have to collect the data for the students, for these 16 attributes by conducting tests. To estimate few attributes, we have prepared questionnaires. As we use Hadoop, data size must be large. The students answered for the questionnaires were prepared by us. The maximum mark will vary according to the membership value of the attributes. The marks for the attributes depend on the number of correct answers. By this way, we evaluate the performance of the students.

D) Rule Base

Here,for every student,we compare the values of the attributes selected and group their performance in the ranges of Very Good,Good,Average,Poor,Very Poor etc.

If-then rules for Memory Power:

 If memorypower is greaterthan or equal to 0.87 then

      Very Good in Memory Power

Else if memorypower is greaterthan or equal to 0.75 and less than or equal to 0.86 then

   Good in Memory Power

Else if memorypower is greaterthan or equal to 0.50 and less than or equal to 0.74 then

   Avrage in Memory Power

Else if memorypower is greaterthan or equal to 0.35 and less than or equal to 0.49 then

   Poor in Memory Power

Else

   Very Poor in Memory Power

The possible results from the rule base are

Marks Obtained	Results
Above 45	Very Good
35-40	Good
40-45	Overlapping area for good and very good
25-30	Average
30-35	Overlapping area for good and average
15-20	Poor
20-25	Overlapping area for average and poor
Below 10	Very Poor
10-15	Overlapping area for poor and very poor

On predicting the performance of a student in the rule base process, a student may be classified into two or more categories. This is called overlapping.

E) Defuzzification

The Defuzzification is needed only if the overlapping occurs   in rule base process. To obtain crisp output, we have used Mean of Maxima method.

CONCLUSION

Here, we have tested only 16 parameters. If we use 50 parameters, then the academic performance of the students can be predicted still more accurately. If the student follows the

suggestions, the student would do well in his academic and he/she would be a key player in the affairs of the country. As this setup is implemented in the hadoop framework, this is more attractive because this can accessible from anywhere in the world.

REFERENCES:

[1] Amazon Elastic Compute Cloud: http//aws.amazon.com/ec2

[2] Apache Hadoop, http//hadoop.apache.org/

[3] Apache Hbase, http//hadoop.apache.org/hbase/

[4] Basic principles of Fuzzy Logic - http://access.feld.cvut.cz/view.php?cisloclanku=2012080002.

[5] Chennakesava R.Alavala., 2008, Fuzzy logic and neural networks Basic Concepts & Applications, First Edition

[6] Case Study of Scientific Data Processing on a Cloud Using Hadoop http://link.springer.com/chapter/10.1007%2F978-3-642-12659-8_29.

[7] GigaSpaces http://www.gigaspaces.com/.