

## ARTICLE

# NEW TECHNIQUE OF ACADEMIC DATA ANALYSIS BY FUZZY VARIANCE

Anil Kumar Tiwari<sup>1</sup>, G Ramakrishna<sup>2</sup>, Lokesh Kumar Sharma<sup>3</sup>, Sunil Kumar Kashyap<sup>4\*</sup>

<sup>1,2</sup>Department of Computer Science and Engineering, K L University, Vijayawada, INDIA

<sup>3</sup>National Institute of Occupational Health, Ahmadabad, Gujarat, INDIA

<sup>4</sup>Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology University, Vellore, Tamil Nadu, INDIA

## ABSTRACT

**Background:** The importance of academic data counts by the modern society. The analysis of the academic data is more important than previous statement. The objective of academic data analysis is to explore more information for future academic improvements. This paper deals with the fuzzy variance technique for the academic data analysis. The additional advantage of this technique is that it gives linguistic decision. Fuzzy variance interacts with data as the set and cluster as the subset. The subset say cluster of the set say data is optimized by the fuzzy variance and this is used for analyzing the performance of academic data of the students. **Methods:** Clustering the data is the mathematical method but its analysis is referred as the mining of data. The fuzzy variance is that mathematical method which enables the study into linguistic form also. **Results:** Thus, this paper proposes a new Data Mining Clustering Technique over the crisp data set of student's academic performances and then its interpretation by MATLAB. Its analysis over the fuzzy set is also presented. **Conclusions:** Hence, the optimized decision algorithm over the variance of Arithmetic-Fuzzy Variance is illustrated for the linguistic and numeric outcomes.

## INTRODUCTION

The information can be represented into various discrete forms e.g. text, number, codes etc. This is said to be a data if it maps with the certain information. Data is the source of infinite information. Data is the set with an element minimum and n number of variables maximum. The element of the data referred as the co-domain of the function where the domain is the certain objective and range is another data. Every data generates the new data if there be applied a particular rule. Data is the storage of information. The redefining the sequence of the information of the given data is the challenge. To search a map from the single information to multiple information is the major objective of data mining. The study of data refers data mining. Data mining is the process to define the new information from the current information. It is science and art also. The set of rules referred as science and the sequence of the elements of that set referred as art. The Several application has been existing since 1979, when Hartign et.al. [1] designed an algorithm. It was the first K-means clustering algorithm. After a decade this is presented another clustering overview in the platform of Artificial Neural Network. It is analyzed in the perspective of statistics. In 1993, Jang [2] proposed an adaptive network based fuzzy inference system in the domain of physical environment. Gau et.al. [3] discovered a new set theory as Vague Sets in the same year. Law [4] used fuzzy numbers for grading the educational system, in the year 1996. In the same year, Possey et.al. [5] drafted a student model, which is based on the Neural Network.

2006 become the year of real application of K-means clustering in students academic performances. It was a case study of pharmacy students. Its result and affect gave by Sansgiry [6]. Oyelade et.al. [7] presented an application of K-means clustering algorithm for prediction of the performances of the students, in the year 2010. Next year, Mankad et.al. [8] displayed an educational case study based on the genetic fuzzy algorithm. Next year, an improved academic performance system came in the existence. Defence University's data warehousing and data mining were the key operators of this system. In the same year, Choudhary et.al. developed a model based on soft computing for academic performances of teachers. It is also based on fuzzy logic. Upadhyaya [9] shown a result on fuzzy logic based evaluation of student's academic performances.

In this paper, an optimized fuzzy K-means clustering algorithm presents for the evaluation of student's academic performances. We analysed K-means, Fuzzy and Optimization Theory in the context to each other and classified the common structure.

The Intelligence is one of the qualities of any student, which becomes the key factor in the academic performances. Intelligence is one of the class or cluster. If we classify any class, then the intelligence plays an important role for grouping the class. It order is also an important phase of study. It can be studied as, either poor, average, intelligence or Intelligence, average, poor or, average, poor, intelligence or poor, intelligence, average or average, intelligence, poor or intelligence, poor, average. These can be formed as the elements of set but its conclusion always distinct with each other. For example, intelligence means score lies in the interval of 80-100, average means. Score lies in the interval of 60-80 and 40-60 for the poor. Suppose a student scores 40 marks in the quarterly examination, 60 marks in the half yearly examination and 90 marks in annual examination. Our conclusion will be as "his academic performance is improving". In other example, a student scored 90, 60, and 40 in the quarterly, half yearly and annual examination respectively. Then, definitely the conclusion will not as earlier. For this, we can say, the academic performance of student is reducing. The rate of learning is distinct in both the cases. Hence, one

Received: 11 Nov 2016  
Accepted: 9 Feb 2017  
Published: 8 March 2017

\*Corresponding Author  
Email:  
7sunilkumarkashyap@gmail.com  
Tel.: +91-94242-16777

objective is determining that the class homogeneity and academic performances can be functioned. The involvement of the linguistic variable, fuzzy logic interacts with this by the common application.

Finally, this paper process a new structure of fuzzy K-means and its application is presented as the algorithm for the efficient evaluation of student's academic performances. Mathematically, the presentation is based on the transformation from hypothetical to physical form as, let a set of  $n$ -data points in the space  $k$  is any integer. Then the objective is defined as "obtain  $k$ -points set in ". It calls centre point and hence, minimize the distance from each point to its nearest.

## MATERIALS AND METHODS

There are various method for the similar objectives. It can be classified into quantitative and qualitative. Our objective is to find the decision in linguistic and numerical both, thus the following existed method is essential to read:

### K-means clustering method

In section 2, it is defined, this section, it will be presented in operational way.

Let  $X$  be set and  $\{A_i\}$ , where  $i = 1, 2, \dots, C$  be a family set. Its partition will be represented as,

$$\begin{aligned} \bigcup_{i=1}^C A_i &= X, \\ A_i \cap A_j &= \phi; \forall i \neq j, \\ \phi &\subset A_i \subset X; \forall i, \end{aligned}$$

Where,  $X = \{x_1, \dots, x_n\}$  is a data sample and  $C$  is the number of clusters.

For  $C$ ;  $2 \leq C < n$ ;  $C = n$ .

Hence, the objective function be,

$$J(U, v) = \sum_{k=1}^n \sum_{i=1}^C \chi_{ik} (d_{ik})^2,$$

Where,  $U$  is the partition matrix,  $v$  is a vector of cluster centre and  $d_{ik}$  a Euclidean distance measure between  $x_k$  and  $v_i$ , it is represented as,

$$d_{ik} = d(x_k - v_i) = \|x_k - v_i\| = \sqrt{\sum_{j=1}^m (x_{kj} - v_{ij})^2}$$

### Fuzzy C-means clustering method

Its fuzzy based C-means clustering method for the multiple clusters. Its objective function will be,

$$J(U, V) = \sum_{i=1}^k \sum_{x_k \in X} (\mu_{C_i}(x_k))^m \|x_k - v_i\|^2$$

Where,  $U$  is a fuzzy position and  $m$  is a weight.

For the local minimum, the objective function becomes;

$$\begin{aligned} \mu_{C_i}(x) &= \frac{1}{\sum_{j=1}^k \left( \frac{\|x - v_i\|^2}{\|x - v_j\|^2} \right)^{\frac{1}{m-1}}}; 1 \leq i \leq k, x \in X \\ r_i &= \frac{\sum_{x \in X} (\mu_{C_i}(x))^m x}{\sum_{x \in X} (\mu_{C_i}(x))^m}; 1 \leq i \leq k, \\ \sum_{i=1}^C \|r_i^{previous} - v_i\| &\leq \epsilon. \end{aligned}$$

Fuzzy mean

As arithmetic mean, fuzzy arithmetic mean is the generalized form the first. Let a multiple be  $(x, \mu(x), [0,1], l_1, l_2)$ , where  $l_1, l_2$  are the two linguistic variables. Then the Fuzzy Arithmetic Mean can be represented as consequence matrix in below::

$$\bar{f} = \begin{bmatrix} f_1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & f_n \end{bmatrix},$$

Where  $f_1, \dots, f_n$  are the grading.

Trace of the fuzzy mean is equal to the sum of the diagonal element, i.e.  $Tr(\bar{f})$ .

Below, we propose a new method, which deals with the decision of numerical and linguistic both. Its explained by the sample data of a student, but it can be generalized for n variables.

RESULT

Proposed method

**Table 1:** An observation of a student's academic performance is presented

Test	I	II	III	IV
Marks (100)	20	50	80	?

There are the following natural questions:

- a. What will be the marks in Test IV?
- b. The average of marks is constant (50) from Test I to III and from III to I also, but what about the performance of student is improving or reducing?

This case study is mentioned earlier in this paper, but in this section, it will be analyzed. For searching the answer of these questions, the following methodology is essential:

**Table 2:** The classical set theory and modern fuzzy logic applied on the above observation as below

Test	I	II	III	IV
Marks (100)	20	50	80	?
Crisp Set	0	0	1	?
Fuzzy Set	0	0.5	1	?

The objective of given the above table is, "for identify the intelligent student". If the student scored more than or equal to 80 marks then it referred as "an intelligent student", thus crisp and fuzzy are existed here. Let a data set (Crisp Set) of the marks of student be  $X = \{x_1, \dots, x_n\}$  and another data set (Fuzzy Set) of the marks of student be  $F = \{(x_1, \mu_A(x_1)), \dots, (x_n, \mu_A(x_n))\}$ . The arithmetic mean of the crisp set and fuzzy set be  $\bar{x}$  and  $\bar{f}$  respectively. The simulated marks of the student be  $x_m$ . It can be presented mathematically as,

$$x_m = f(\bar{x} \pm a); a \in R.$$

either,  $x_m = f(\bar{x} + a)$   
or,  $x_m = f(\bar{x} - a)$ .

Similarly for the fuzzy set,

$$x_m = f(\bar{f} \pm a); a \in R.$$

$$\text{either, } x_m = f(\bar{f} + a)$$

$$\text{or, } x_m = f(\bar{f} - a).$$

A mapping  $a$  is defined from  $\bar{x}$  to  $\bar{f}$  as,  $a: \bar{x} \rightarrow \bar{f}$ , or,  $\bar{f} = a(\bar{x})$  and it variance be  $v_x, v_f$ .

Hence,  $x_m$  can be optimized as,

$$x_m = \max(v_x \cup v_f, \mu_{x_m}(v_x \cup v_f)),$$

s.t.,

$$v_x = \frac{\sum (d\bar{x})^2}{N}; \text{Population}$$

Or,

$$v_x = \frac{\sum (d\bar{x})^2}{N-1}; \text{Sample}$$

&

$$v_f = \frac{\sum (d\bar{f})^2}{N}; \text{Population}$$

Or,

$$v_f = \frac{\sum (d\bar{f})^2}{N-1}; \text{Sample}$$

&

Error: either,  $v_x, v_f \rightarrow <$ ,

or,  $v_x, v_f \rightarrow >$ .

Its algorithm is presented in below section.

### Algorithm

Input Set  $X$ : {(Student's Name, Marks)} =  $\{(s_1, t_1), \dots, (s_n, t_n)\}$ .

Input Set  $F$ : {(Student's Name, Membership Function of Marks)} =  $\{(s_1, \mu(t_1)), \dots, (s_n, \mu(t_n))\}$ .

Compute  $\bar{x}, \bar{f}$ .

Optimized by,

$$x_m = \max(v_x \cup v_f, \mu_{x_m}(v_x \cup v_f)),$$

s.t.,

$$v_x = \frac{\sum (d\bar{x}_n)^2}{N}; \text{Population}$$

Or,

$$v_x = \frac{\sum (d\bar{x}_n)^2}{N-1}; \text{Sample}$$

&

$$v_f = \frac{\sum (\mu d\bar{x}_n)^2}{N}; \text{Population}$$

Or,

$$v_f = \frac{\sum (\mu d\bar{x}_n)^2}{N-1}; \text{Sample}$$

&

Error:  $|ND - LD| \neq 0$ .

Where, ND and LD for Numeric and Linguistic Decisions respectively.

Output:

Numeric Decision:

$$x_m = f(\bar{x} \pm a); a \in R.$$

$$\text{either, } x_m = f(\bar{x} + a)$$

$$\text{or, } x_m = f(\bar{x} - a).$$

Linguistic Decision:

$$x_m = f(\bar{f} \pm a); a \in R.$$

$$\text{either, } x_m = f(\bar{f} + a)$$

$$\text{or, } x_m = f(\bar{f} - a).$$

Illustration is given below, which interacts with the above mentioned example.

Example:

Input Crisp Set  $X$  :

Input Fuzzy Set  $F$  :

## CONCLUSION

Individually, statistical means used as the common tool in general decision theory. Fuzzy mean is the newly launched method. Its application is presented in this paper for the better decision than existed. The study of the academic performance of students is equivalent to the study of discrete function of arithmetic and fuzzy means. Both is taken simultaneously is still challenging, but the paper is become the first step toward this under the error estimation.

## CONFLICT OF INTEREST

None

## ACKNOWLEDGEMENTS

To the Laboratory, KL University, Vijayawada, Andhra Pradesh, INDIA.

## FINANCIAL DISCLOSURE

This is an unfunded research.

## REFERENCES

- [1] Hartigan JA, Wong MA. [1979] A K-means clustering algorithm, Appl Stat, 28: 100-108.
- [2] Jang JSR. [1993] ANFIS: Adaptive-network based fuzzy inference system, IEEE Trans Sys, Man, Cybernetics, 23(3): 665-685.
- [3] Gau WL, Buehrer DJ. [1993] Vague sets, IEEE Transactions System, Man Cybernetics, 23(2): 610-614.
- [4] Law CK. [1996] Using fuzzy numbers in educational grading system, Fuzzy Sets Syst, 83: 311-323.
- [5] Posey CL, Hawkes LW. [1996] Neural networks applied in the student model, Intell Sys 88: 275-298.
- [6] Sansgiry SS, Bhosle M, Sail K. [2006] Factors that affect academic performances among pharmacy students, Am J Phar Edu, 70(5): 231-243.
- [7] Oyelade OJ, Oladipupo OO and Obagbua IC. [2010] Application of K-means clustering algorithm for prediction of students's academic performances, Int J Comp Sci Inf Sec, 7(1): 292-295.
- [8] Mankad K, Sajja PS, Akerkar R. [2008] Evolving rules using genetic algorithm: a hard problem for partition based approach, Int J Soft Comp Appl, 2: 6-15.
- [9] Upadhyaya MS. [2012] Fuzzy logic based on performances of students in college, J Comput Appl 5(1): 6-9.