Volume 5, Issue 1, January 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Study on 'Poor Academic Performance of Students' using Fuzzy Clustering

A. Rajkumar*
Assistant Professor
Hindustan University
Chennai, India

Nikhil Nair
Student (Department of Computer Science)
Hindustan University
Chennai, India

N. Jose Parvin Praveena
Assistant Professor
K.C.G College of Technology
Chennai, India

Abstract*— Students play an integral part in deciding the future of a nation. Their academic learning trains them to visualise problems differently and come up with effective solutions. This paper deals with the analysis of problems that hinder this particular 'intellectual trait' in children to learn which in turn leads them to perform very poorly in academics. Fuzzy clustering is employed for this analysis. A total of twenty attributes have been chosen for this study and distributed along a 10-point rating scale based upon the level of distraction that causes students in the city of Chennai to perform poorly in their academics. The average rating scale is used to cluster the attributes with Fuzzy c-means clustering which in-turn classifies these attributes into – low, medium, and high levels of distraction. A proactive approach can then be taken to ensure that students perform well in their fields of study. The advantage of fuzzy clustering is that attributes can be a part of more than on cluster to a varying degree, thus giving a better result with respect to the student problem attributes.

***Keywords*— C-means Clustering, Degree, Mid-Value, Students, attributes**

I. INTRODUCTION

Nowadays students are the primary clients in the majority of all student affairs units. It is also true that in spite of having a set of agreed upon student development principles, student affairs units have been quite successful in providing effective and innovative programs and services. Students enter a college or university with varying patterns of personal, family, and academic characteristics and skills, including initial dispositions and intentions with respect to college attendance and personal growth. These intentions and commitments are subsequently modified and reformulated on a continuing basis through a longitudinal series of interactions between the individual and the structures and members of the academic and social systems of the institution. It varies from student to student. They tend to face many problems that affect their studies and give poor results and leads to a high level of distraction. From the survey of 100 Students, the most common problems are listed and are taken as the major attributes. This study aims to get the results of the students' problems about the varying levels of distraction using the attributes and classify these attributes based on fuzzy c-means clustering. Fuzzy clustering is more appropriate for perceptions based data since perceptions are always a matter of varying degree. The organization of this article is as follows: Choosing the attributes in methodology section, basic notion of clustering and fuzzy clustering in next section, Results and Discussion in the fourth section and finally we arrive at the conclusion and scope for our future study.

II. METHODOLOGY

In this methodology, the classification of problems that adversely affect the academic performance of students is performed based on the opinions and perceptions gathered from the students. The following twenty problem attributes are chosen by interviewing the Students of different age groups in Chennai.

1. Bad Thoughts
2. No vision about life
3. Lack of Concentration
4. Love affair and love failure
5. Depression
6. Health Problem
7. Bad Habits
8. Preparation time is not sufficient
9. Family Pressure
10. Lazy to Study
11. Loss of Memory
12. Loss of Self-confidence
13. Bad Companion
14. No Proper Teaching

15. Financial Problem
16. Long time travelling/Transportation problem
17. No proper food intake
18. Understanding capacity is very less
19. Social Network
20. Confusion

Three attributes which best define the characteristic of each segment have been selected to be rated by respondents. Fuzzy c-means clustering is done using the algorithm (4.1).

III. PRELEMINARIES

(a) HARD CLUSTERING

In **Hard Clustering** we make a hard partition of the data set **Z**. In other words, we divide them into $c \geq 2$ clusters. With a partition, we mean that

$$\bigcup_{i=1}^{c} A_i = Z$$

and $A_i \cap A_j = \phi$, $\forall\, i \neq j$ ..... (1)

Also, none of the sets, A_i may be empty. To indicate a portioning, we make use of **membership functions** $\mu_k(x)$. If $\mu_k(x) = 1$, then object x is in cluster k. Based on the membership functions, we can assemble the **Partition Matrix U**, of which $\mu_k(x)$ are the elements. Finally there is a rule that $\forall$x,

$$\sum_{i=1}^{c} \mu_k \ (x) = 1 \ \forall x \quad \ldots\ldots\ldots\ldots\ldots \ (2)$$

In other words, every object is only part of one cluster.

(b) FUZZY CLUSTERING

Hard clustering has a downside. When an object roughly falls between two clusters A_i and A_j, it has to be put into one of these clusters. Also, outliers have to be put in some cluster. This is undesirable. But it can be fixed by fuzzy clustering.

In **Fuzzy clustering**, we make a **Fuzzy partition** of the data. Now, the membership function $\mu_k(x)$ can be any value between 0 and 1. This means that an object z_k can be for 0.2 parts in A_i and for 0.8 parts in A_j. However, requirement (2) still applies. So, the sum of the membership functions still has to be 1. The set of all fuzzy partitions that can be formed in this way is denoted by M_{fc}. Fuzzy partitioning again has a downside. When we have an **outlier** in the data (being an object that doesn't really belong to any cluster), we still have to assign it to clusters. That is, the sum of its membership functions still must equal one.

(c) *FUZZY C-MEANS CLUSTERING*

In fuzzy clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster, may be in a cluster to a lesser degree than points in the center of cluster for each point x there is no coefficient giving the degree of belonging in the k^{th} cluster $\mu_k(x) = 1$. Usually, the sum of those coefficients is defined to be 1.

$\sum_{n=1}^{mm\ cluster} \mu_k \ = 1 \quad \forall x$.....(3)

With fuzzy c-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster

$$\text{Center}_k = \frac{\sum_x \mu_k(x)^m x}{\sum_x \mu_k(x)^m} \quad \ldots\ldots\ldots\ldots\ldots\ldots \ (4)$$

The degree of belonging is related to the inverse of the distance to the cluster

$$\mu_k(x) = \frac{1}{d(center_k, x)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

then the coefficients are normalized and fuzzy field with a real parameter $m > 1$ so that their sum is 1. So

$$\mu_k(x) = \frac{1}{\sum_j \left(\frac{d(center_k, x)}{d(center_j, x)} \right)^{2/(m-1)}} \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

For m equal to 2, this is equivalent to normalizing the coefficient linearly to make their sum 1. When m is close to 1, then cluster center closes to the point is given much more weight than the others, and the algorithm is similar to k-means.

IV. RESULTS AND DISCUSSIONS

We have interviewed 100 Students in Chennai city to find what are all the problems that affects their studies to give poor results. For this, twenty attributes have been chosen and the respondents had related the attributes of problems that engendered high levels distraction on a 10-point rating scale and the results of the average rating scale and the results of the average rating are shown in Fig.1 2^{nd} and 9^{th} attributes are rated highest by the students with an average rating of

8.5 and 8.3 respectively on a 10-point scale. This means that, when we compare other attributes, 2nd, 6th,7th,and 10th engendered high levels of distraction and the 13th attribute rating averaged at 2.5which engendered low levels distraction.

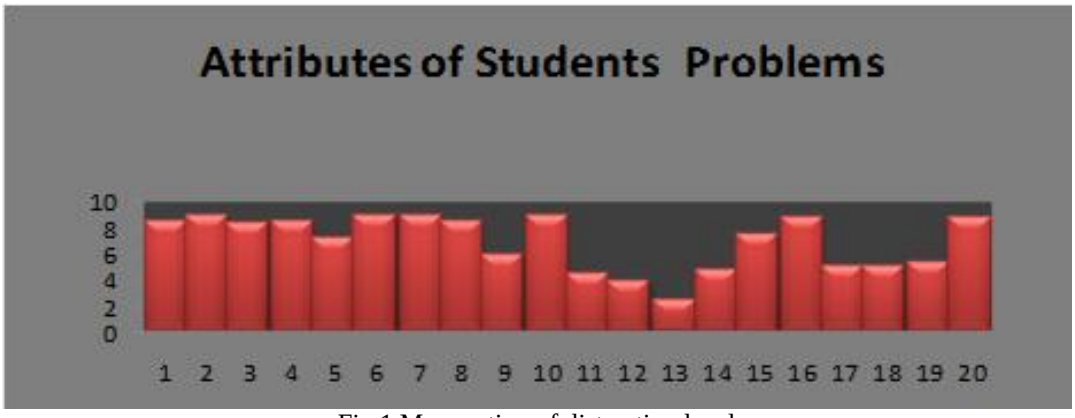

Fig.1 Mean rating of distraction level

The ratings and the Standard Deviation of the attributes of the problems that gives poor results of students engendered high level distraction have been subjected to fuzzy c-means clustering using the algorithm:1 and the following results shown in Table:1 have been obtained for a 3-cluster combination. The first cluster comprises of the attributes with average rating from 2.0 to 5.5 with a mid value 3.75. The second cluster range is from 3.0 to 8.0 with a mid valued 5.5 and the third cluster has a range of 7.0 to 10 with a mid value 8.5.

The first cluster range indicates the problems that give poor results of students which engendered LOW level distraction. Second and Third clusters range shows that MODERATE and HIGH level of distractions respectively. There is Over-lapping ranges as in characteristic of a fuzzy based cluster.

Table: 1 3-Cluster Range of Level of distraction

	Cluster 1	**Cluster 2**	**Cluster 3**
Range	2.0-5.5	3.0-8.0	7.0-10
Mid- Value	3.75	5.5	8.5
Classification	LOW	MODERATE	HIGH

4.1 ALGORITHM TO FIND A MEMBERSHIP VALUES FOR THE ATTRIBUTES

Step 1: Start

Step 2: Fix, the values of 20 attributes on a 10-point rating scale in a set D (say)

Step 3: Fix the cluster, which is defined as Cluster 1 = LOW, whose range beginning with 2.0 (**bv1**) End with 5.5(**ev1**).Cluster 2 = MODERATE, whose range beginning with 3.0 (**bv2**) end with 8.0(**ev2**). Cluster 3 = HIGH, whose rangebeginning with 7.0 (**bv3**) end with 10 (**ev3**).

Step 4: Choose en element *x* in D

Step 5: If $x < ev1$, Go to *Step 6*, else Go to *Step8*

Step 6: If $x > bv2$, then ***x* lies in cluster 1 and cluster 2** whose membership value is defined as $\mu_k(x) = ev1-x: x-bv2$, Go to *Step 12*, else Go to *Step7*.

Step 7: Then ***x* lies in cluster 1** only, the membership value is $\mu_k(x) = 1$ Go to *Step 12*

Step 8: If $x < ev2$ Go to *Step 9*, else Go to *Step 11*

Step 9: If $x > bv3$, then ***x* lies in cluster 2 and cluster 3**, whose membership value is defined as $\mu_k(x) = ev2-x: x-bv3$, Go to step 12, else Go to *Step10*

Step 10: Then ***x* lies in cluster 2** only, the membership value is $\mu_k(x) = 1$ else
Go to *Step11*

Step 11: Then ***x* lies in cluster 3** only, the membership value is $\mu_k(x) = 1$

Step 12: Go to *Step 4*, until all the values in D have been checked

Step 13: Stop

Here **'bv'** denotes the beginning value and '**ev'** denotes the ending the value.

Degree of membership of the attributes of problems that gives poor results of students is found using the above algorithm is shown in Table: II. Attribute 13 with a mean rating 2.5 is entirely (100 percentage) with a membership value of 1in cluster 1.

(i.e.) Distraction level is LOW.

Attribute 9 with a mean rating 6.0 is entirely (100 %) with a membership value of 1in cluster 2. (i.e.) Distraction level is MODERATE.

Attribute 1,2,3,4,6,7,8,10,16,20 with a mean rating 8.5,9.0,8.4,8.5,9.0,9.0,8.5,9.08.9,8.8 is entirely (100 %) with a membership value of 1in cluster 3. (i.e.) Distraction level is HIGH.

Attributes 11,12,14,17,18,19 with a mean ratings 4.5,4.0,4.8,5.1,5.1,5.4 belongs to 40 % in cluster 1 and 60 % in cluster 2, 60 % in cluster 1 and 40% in cluster 2, 32 % in cluster 1 and 68% in cluster 2, 16 % in cluster 1 and 84 % in cluster 2, 16 % in cluster 1 and 84 % in cluster 2. 4 % in cluster 1 and 96 % in cluster 2. (i.e.) distraction level is between LOW and MODERATE.

Attributes 5,15with a mean ratings 7.2,7.6 belongs to 80 % in cluster 2 and 20 % in cluster 3, 40 % in cluster 2 and 60 % in cluster 3, (i.e.) distraction level is between MODERATE and HIGH.

Table: II Degree of Membership of the Attributes

S.N0	MEAN	LOW	MODERATE	HIGH
1	8.5	0	0	1
2	9.0	0	0	1
3	8.4	0	0	1
4	8.5	0	0	1
5	7.2	0	0.80	0.20
6	9.0	0	0	1
7	9.0	0	0	1
8	8.5	0	0	1
9	6.0	0	1	0
10	9.0	0	0	1
11	4.5	0.40	0.60	0
12	4.0	0.60	0.40	0
13	2.5	1	0	0
14	4.8	0.32	0.68	0
15	7.6	0	0.40	0.60
16	8.9	0	0	1
17	5.1	0.16	0.84	0
18	5.1	0.16	0.84	0
19	5.4	0.04	0.96	0
20	8.8	0	0	1

V. CONCLUSION

In Crisp Clustering, where an attribute is a member of one cluster only, the fuzzy clustering process permits an attribute to be a member of more than one cluster although to a varying degree. This helps to find out where students get more distracted and what stimulate poor results in them so that remedies can be given to them. From this study, it is concluded that students were insisted on knowing where they were always; No vision about life, Lack of Concentration, Love affair and love failure, Depression, Loss of Self-confidence, Health Problem, Loss of Memory, Financial Problem which stimulate distraction thought in them very high. In this article, the effect of No vision about life, family pressure, loss of memory having high mean rating was observed and this will not only led to distraction but also health problems in them which will be the main scope of study in the future to analyze the causative agents for health problems.

.

ACKNOWLEDGEMENT

I would like to thank my professor A.Victor Devadoss for his valuable support and encouragement

REFERENCES

[1] A.J. Zolotor, A.C. Denham, A. Weil, "*Intimate partner Violence*", Prim Care 2009.
[2] A. Kaufmann, "*Introduction to the Theory of Fuzzy Subsets*", Academic Press, INC. (LONDON) LTD, 1975.
[3] A. Resolution/RES/48/104, *Declaration on the elimination of violence against women*, adopted by the UN General Assembly, 20 December 1993.
[4] B. Kosko, "*Hidden pattern in combined and Adaptive Knowledge Networks*" Proc. Of the First, *IEE International Conference on Neural Networks* (ICNN-86(1988) 377-398).
[5] B. Kosko, "*Neural Networks and Fuzzy systems: A Dynamical System Approach to Machine Intelligence*", Prentice Hall of India, 1997.
[6] C. Gacria-Moreno, HA. Jansen, M. Elsberg, L. Hesise, CH. Waltts, "*Definitions Questionnaire development In. Who multi-country study on women's health and domestic violence against women*", Geneva World Health Organization, 2005.
[7] D. Varma, PS. Chandra, T. Thomas, MP. Carey, *Intimate partner violence and sexual coercion among pregnant women in India: Relationship with depression and post traumatic stress disorder.* J Affect Disord 2007; 102:227-35.
[8] H.J. Zimmermann, "*Fuzzy Set Theory and its application*", Fourth Edition Springer 2011.
[9] J. Klir George/Bo Yuan, "*Fuzzy sets and Fuzzy Logic : Theory and Applications*", Prentice Hall of India.
[10] M.A. Strauss, "*Conflict Tactics Scales. In: NA. Jackson, editor Encyclopedia of Domestic violence*", 2007.
[11] S. Kumar, L. Jeyaseelan, S. Suresh, RC, Ahuja, "*Domestic violence and its mental health correlates in Indian women*", Br J Psychiatry 2005.
[12] Sa. Falsetti, "*Screening and responding to family and intimate partner violence in the primary care setting*", Prim Care 2007.
[13] J. Gunter, "Intimate Partner Violence. Obset Gynecol Clin North Am 2007; 34:367-88.
[14] WHO multi-country study on women's health and domestic violence against women: summary report of Indian results on prevalence, health outcomes and Women's responses, Geneva: World Health Organization; 2005, p-15-7.
[15] Fuzzy Clustering of the Domestic Violence for the Degree of Suicide thought based on Married Women Perception by A.Victor Devadoss, A.Felix, 2012, IJERA p: 1770-1773

