IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 3, Issue 10, May 2015)

A Novel Method for Cluster Analysis in Data Mining using Improved Fuzzy C-Means Algorithm

Swapnil Jain
M.Tech. Scholar
Department of Computer Science & Engineering
Sushila Devi Bansal College of Technology, Indore
(India)

Mrs. Shraddha Kumar
Asst. Professor
Department of Computer Science & Engineering
Sushila Devi Bansal College of Technology, Indore
(India)

Abstract –**Clustering is a data mining technique of grouping set of data objects into multiple groups or clusters so that objects within the cluster have high similarity, but are very dissimilar to the objects in the other clusters. Fuzzy C-Means is the most widely used method where an element may have partial membership grades in more than one fuzzy cluster. This paper makes use of MATLAB language to produce a fuzzy clustering algorithm for classifying the batting statistics of Indian Premier League (IPL) T-20 cricket tournament into several numbers of clusters. The clusters as well as the membership function has been implemented using MATLAB. The results obtained from Indian premier league batting statistics dataset detect n-clusters to handle the imprecise and ambiguous result. Proposed research work provide an Improved Fuzzy C-Means clustering technique which provide sufficient and accurate data analysis in the field of data mining and the minimal distance between clusters is determined by using the Hermitian distance method.**

Keywords – **Clustering, Fuzzy C-Means, Hermitian distance, Indian Premier League.**

I. INTRODUCTION

Clustering is an important technique in exploratory data analysis, with applications in image processing, object classification, target recognition, data mining etc. In clustering, no data are tagged before being fed to a function. Clustering is designated as undirected knowledge discovery or unsupervised learning. In cluster detection the groups of similar data sets are formed and the cluster detection algorithm searches for groups or clusters of data elements that are similar to one another. Each implementation of the cluster detection algorithm adopts a method of comparing the values of the variables in individual records with those in the cluster centroids. Then the clustering algorithm uses these comparisons to calculate the distance of individual data sets from the centroids. After calculating the distances, the algorithm redraws the cluster boundaries.

The objective is focus on maximization of intra cluster similarity while minimization of inter subject similarity. This research work proposes and produce a fuzzy clustering algorithm for classifying the batting statistics of Indian Premier League (IPL) T-20 cricket tournament into several numbers of clusters. In previous research work [1], they used Euclidean distance to calculate minimal distance between clusters, though the Euclidean distance is efficient only with continuous variables datasets and the proposed research work uses ordinal datasets so we have used Hermitian distance method for the same.

II. DATA MINING

Data mining refers to extracting or 'mining' interesting knowledge from large amounts of data. It provides a means of extracting previously unknown, predictive information from the base of accessible data in data warehouses. Against this background, a great interest is being shown in the field of data mining or KDD (knowledge discovery in databases). KDD is like mining, where enormous quantities of data have to be removed before diamonds or gold can be found. Similarly, with a computer, one can automatically find the one "information-diamond" among the tons of data in one's database.

It was proposed at the first international KDD conference in Montreal in 1995 that the term 'KDD' be employed to describe the whole process of extraction of knowledge from data, which is a multi-disciplinary field of research where the knowledge here means relationships and patterns between data elements, data mining is used exclusively for the discovery stage of the KDD process [2].

Knowledge discovery as a process consists of the following steps:

- Data cleaning (to remove noise and inconsistent data).
- Data Integration (where multiple data sources may be combined).
- Data Selection (Relevant data from database are retrieved for analysis).

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 3, Issue 10, May 2015)

- Data transformation (where data are transformed or consolidated into forms appropriate for mining).
- Data mining (Process of intelligent methods to extract data patterns).
- Pattern evaluation (Identifies interesting patterns representing knowledge).
- Knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user).

The architecture of a typical data mining system has the following components:

- Database, data warehouse, or other information repository.
- Database or data warehouse server: It is responsible for fetching the relevant data, based on the user's data mining request.
- Knowledge base: This is the domain knowledge that is used to guide the search, or evaluate the interestingness of resulting patterns.
- Data mining engine: Consists of a set of functional modules for tasks such as characterization, association, classification, cluster analysis, and evolution and deviation analysis.
- Pattern evaluation module: It employs interestingness measures and interacts with the data mining modules in order to focus the search towards interesting patterns.
- Graphical user interface: This communicates between users and the data mining system, allowing the user to interact with the system by specifying a query or task, information to help focus the search and visualize the patterns in different forms.

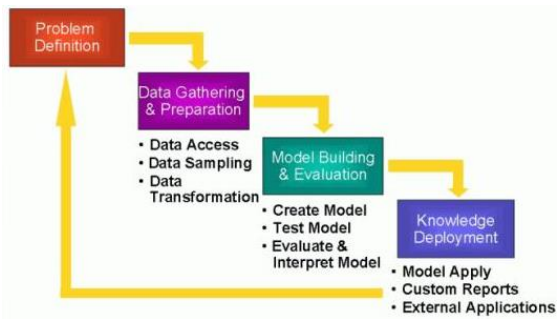


Figure 1: Structure of data mining

Figure 2: Process of data mining

III. PROPOSED METHOD

Improved Fuzzy C-means (FCM) algorithm, in which the data set are taken and cluster will be initialized, normally variable will be selected. The improved FCM is used to form the cluster group by taking some data set. The quality of cluster is calculated by finding the distance measurement.

The improved FCM method are applied in large volume of data and give the true prediction. The objective function is minimized in improved fuzzy c-mean algorithm. The overlapped function is less as compared to other by using the hermitition distance formula. The result are improved by hermitition distance. It give true prediction. Improved FCM is one of the algorithm which is used in data mining for clustering. The improved FCM algorithm by using two step, first is by decision tree approach with it which mine the data in accurate and sequential manner and second by creating the noise free data. Cluster validity function are often used to evaluate the performance of clustering in different index and even two different clustering method. Among the criteria there are important type of FCM in which fuzzy partition based sample set. The main idea of validity function based on fuzzy partition is that the less fuzziness of the partition is the better the performance.

An improved FCM algorithm is put forward & applied to deal with IPL data on the top of a traditional FCM algorithm. The proposed Algorithm improved the classical FCM Algorithm by adopting a novel strategy for selecting the initial cluster center to solve the problem of FCM. This algorithm is modified FCM called the improved fuzzy c-mean algorithm depend on the selection of the initial cluster center and the initial membership value. If a good initial cluster center that it is close to the actual final cluster center can be found. Hermitition Distance is an extension method & these technique are expected to have a better performance than conventional method using Euclidean method. The algorithm is an extension of the classical and k-mean clustering method in fuzzy-set domain. Consider the accuracy of IPL we successfully implement the FCM algorithm taking advantage of the basic IPL. The effectiveness of the FCM algorithm in term of computational rate is improved by modified the cluster and more accurate and precise format.it give a more Explanation and

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 3, Issue 10, May 2015)

detailed domain information. The major benefit of change in objective function and accurate result in term of DB index and DONE index. Flow chart of the improved fuzzy c-mean algorithm is shown in figure 3.

Improved FCM is used to form the cluster group by taking some dataset. The quality of clustering is calculated by finding the distance measure. The proposed FCM is used to form the cluster with less number of iteration .data sets are taken and cluster will be initialize, normally variable will be selected randomly but here tried to avoid selecting the variable randomly. Basically the Fuzzy clustering is useful to mine complex and multidimensional dataset. Where the member have partial or fuzzy relation. Among the variable developed technique, FCM algorithm is most popular one where a piece of data has partial membership with each of the pre-defined cluster center.

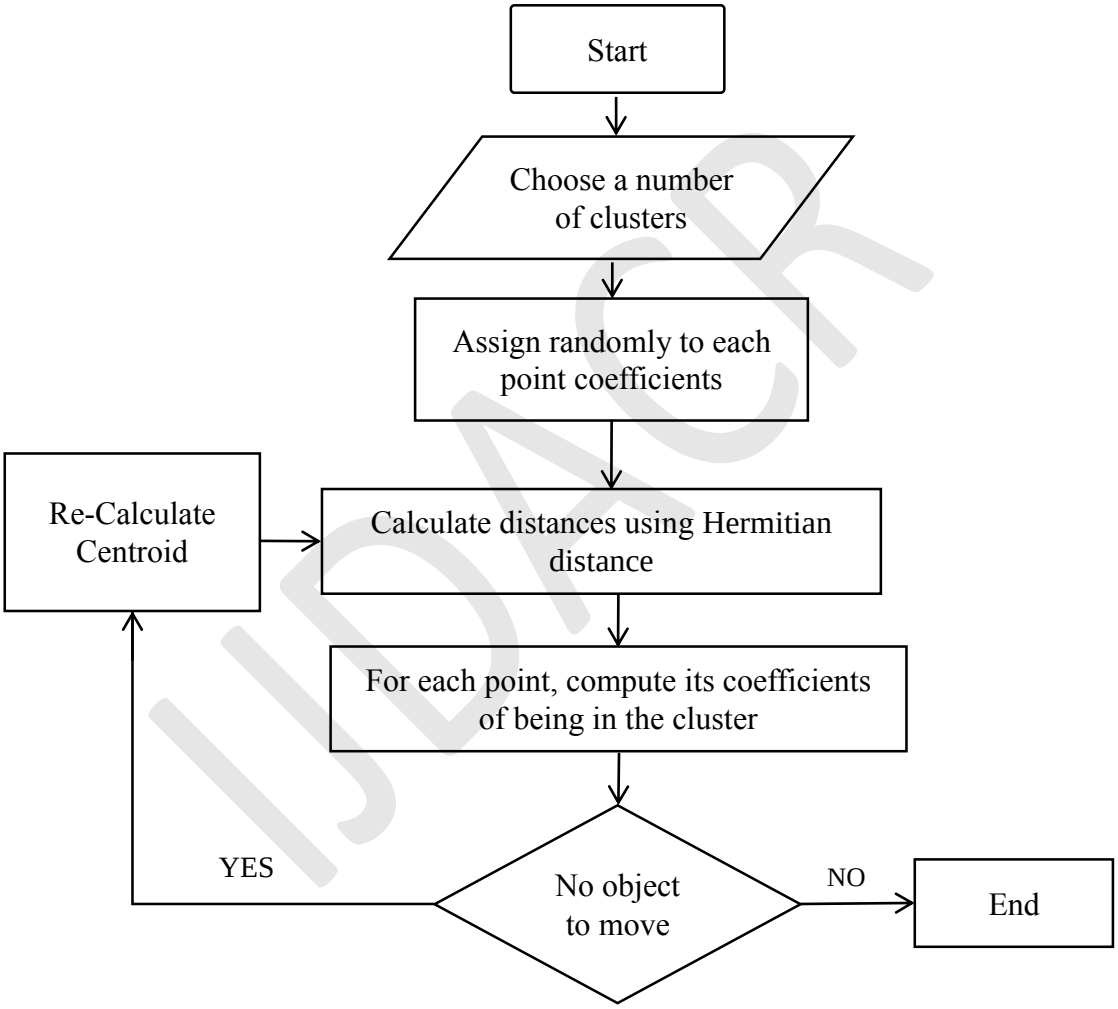


Figure 3: Flow chart of the improved fuzzy c-mean algorithm

Algorithm for Proposed Work

Improved fuzzy c-mean algorithm is put forward and applied to deal with IPL data on the top of a traditional fuzzy c-mean algorithm.

1. First the initial fuzzy partition matrix is generated and the initial fuzzy cluster center are calculated.
2. In each step of iteration the cluster center and the membership grade point are updated and the objective function is minimized to find the best location for the cluster.
3. Improved FCM is proposed cluster technique. It is used to solve the minimal

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 3, Issue 10, May 2015)

distance by using the Hermitian distance method formula:

$$X = [(x_1 - y_1), (x_2 - y_2), (x_3 - y_3) \dots \dots (x_n - y_n)] \quad (1)$$

$$D = (X * X^t)^2 \quad (2)$$

Where X is a matrix and X^t is a transpose of matrix X.

4. The process stops when the maximum number of iteration is reached or when the objective function improvement between two consecutive iteration is less than the minimum amount of data specified.

The result is in more detailed, precise and better optimized. We focus on the optimization. It will be precisely applicable for the mathematical similarity. The update in the iteration is done using the membership degree as well as the center of the cluster that is the two parameter changes as the steps are being repeated until a set point called the threshold is reached or the process stop when the maximum number of iteration is reached or when the objective function improvement between two consecutive iteration is less than the minimum amount of improvement specified.

Hermitian Distance Method

This method is based on the method of Kumar and yang. The objective of these method is to give a short and easy prove of the minimum distance of hermitition points. It is main worked with a précised format. It also provides the data classification. The result are accurate in term of DB index and DONE index.

Database Used

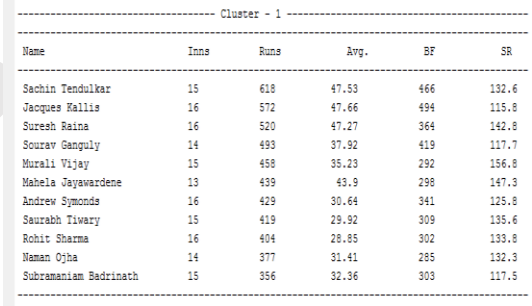
The concept of clustering has been considered in order to classify the IPL bating statistics of fuzzy data into appropriate clusters. A fuzzy database has been constructed by using MATLAB and the insertion of whole records comprises of large data which are collected from IPL website. The dataset consists of several attributes like player-id, player name, team name, innings, runs, average, balls and strike rates which are clearly shown in Table 1.

Table 1: Sample of Raw Data before Clustering

PLAYER	TEAM	INNS	RUN	AVG	BALL	SR
Shikhar Dhawan	RCB	10	191	19.1	170	112.35
Shane Watson	DC	5	185	37	114	162.28
Abhishek Jhunjhunwala	RCB	11	183	20.33	166	110.24
Adam Voges	DC	7	181	45.25	143	126.57
Mohnish Mishra	RR	11	166	15.09	144	115.27
Faiz Fazal	KKR	9	164	20.5	154	106.49
Mithun Manhas	RR	8	157	39.25	149	105.36
Chris Gayle	KKR	9	292	32.44	184	158.69
Virender Sehwag	DD	7	356	25.42	218	163.3
Ambati Rayudu	MI	4	356	27.38	246	144.71
Kumar Sangakkara	DC	13	357	29.75	257	138.91
MS Dhoni	CSK	11	287	31.88	210	136.66
Gautam Gambhir	KKR	10	277	30.77	217	127.64
Yuvraj Singh	KI	14	255	21.25	199	128.14
Manoj Tiwary	KKR	11	237	26.33	186	127.41
JP Duminy	CSK	7	157	31.4	131	119.84
Harbhajan Singh	MI	7	105	26.25	63	166.66
Parthiv Patel	KI	4	72	18	79	91.13

IV. Simulation And Results

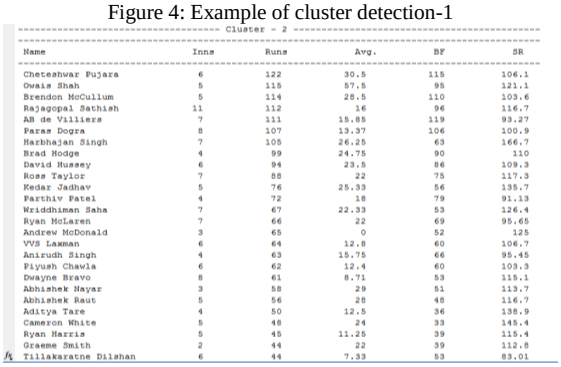
The performance of proposed algorithms has been studied by means of MATLAB simulation.



Cluster - 1

Name	Inns	Runs	Avg.	BF	SR
Sachin Tendulkar	15	618	47.53	466	132.6
Jacques Kallis	16	572	47.66	494	115.8
Suresh Raina	16	520	47.27	364	142.8
Sourav Ganguly	14	493	37.92	419	117.7
Murali Vijay	15	458	35.23	292	156.8
Mahela Jayawardene	13	439	43.9	298	147.3
Andrew Symonds	16	429	30.64	341	125.8
Saurabh Tiwary	15	419	29.92	309	135.6
Rohit Sharma	16	404	28.85	302	133.8
Naman Ojha	14	377	31.41	285	132.3
Subramaniam Badrinath	15	356	32.36	303	117.5

Figure 4: Example of cluster detection-1



Cluster - 2

Name	Inns	Runs	Avg.	BF	SR
Cheteshwar Pujara	6	122	30.5	115	106.1
Owais Shah	5	115	57.5	95	121.1
Brendon McCullum	5	114	28.5	110	103.6
Rajagopal Sathish	11	112	16	96	116.7
AB de Villiers	7	111	15.85	119	93.27
Paras Dogra	8	107	13.37	106	100.9
Harbhajan Singh	7	105	26.25	63	166.7
Brad Hodge	4	99	24.75	90	110
David Hussey	6	94	23.5	86	109.3
Ross Taylor	7	88	22	75	117.3
Kedar Jadhav	5	76	25.33	56	135.7
Parthiv Patel	4	72	18	79	91.13
Wriddhiman Saha	7	67	22.33	53	126.4
Ryan McLaren	7	66	22	69	95.65
Andrew McDonald	3	65	0	52	125
VVS Laxman	6	64	12.8	60	106.7
Anirudh Singh	4	63	15.75	66	95.45
Piyush Chawla	6	62	12.4	60	103.3
Dwayne Bravo	8	61	8.71	53	115.1
Abhishek Nayar	3	58	29	51	113.7
Abhishek Raut	5	56	28	48	116.7
Aditya Tare	4	50	12.5	36	138.9
Cameron White	5	48	24	33	145.4
Ryan Harris	5	45	11.25	39	115.4
Graeme Smith	2	44	22	39	112.8
Tillakaratne Dilshan	6	44	7.33	53	83.01

Figure 5: Example of cluster detection-2

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 3, Issue 10, May 2015)

Cluster - 3

Name	Inns	Runs	Avg.	BF	SR
George Bailey	1	18	18	27	66.66
Dale Steyn	4	13	3.25	17	76.47
Pradeep Sangwan	2	10	10	14	71.42
Shane Warne	7	10	2	19	52.63
Morne Morkel	2	9	9	16	56.25
Rudra Pratap Singh	5	9	4.5	19	47.36
Jaskaran Singh	4	8	8	11	72.72
Shreevats Goswami	1	8	8	10	80
A Kumble	5	6	0	11	54.54
Ishant Sharma	2	6	6	17	35.29
Ramesh Powar	2	5	5	8	62.5
A Uniyal	2	4	4	7	57.14
AN Ahmed	1	4	0	7	57.14
Dirk Nannes	2	4	4	13	30.76
Mandeep Singh	2	4	2	11	36.36
Munaf Patel	2	4	4	5	80
Praveen Kumar	4	4	2	8	50
Arun Karthik	1	3	0	5	60
Pragyan Ojha	6	3	1	8	37.5
Sudeep Tyagi	1	3	0	4	75
Umesh Yadav	1	3	0	4	75
AA Bilakhia	1	2	2	4	50
Juan Theron	2	2	1	5	40
KP Appanna	2	2	2	3	66.66
Manpreet Gony	1	2	2	6	33.33

Figure 6: Example of cluster detection-3

Cluster - 4

Name	Inns	Runs	Avg.	BF	SR
Robin Uthappa	14	374	31.16	218	171.6
Kumar Sangakkara	12	357	29.75	257	138.9
Ambati Rayudu	14	356	27.38	246	144.7
Virender Sehwag	14	356	25.42	218	163.3
ML Hayden	16	346	21.62	279	124
Yusuf Pathan	14	333	27.75	201	165.7
Tirumalasetti Suman	14	307	34.11	257	119.5
Virat Kohli	13	307	27.9	212	144.8
Chris Gayle	9	292	32.44	184	158.7
Adam Gilchrist	16	289	18.06	185	156.2
MS Dhoni	11	287	31.88	210	136.7
David Warner	11	282	28.2	191	147.6
Dinesh Karthik	14	278	21.38	237	117.3
Michael Lumb	11	278	25.27	192	144.8
Gautam Gambhir	10	277	30.77	217	127.6
Irfan Pathan	13	276	34.5	186	148.4
Kieron Pollard	14	273	22.75	147	185.7
Herschelle Gibbs	10	267	26.7	235	113.6
Rahul Dravid	11	256	28.44	199	128.6
Yuvraj Singh	14	255	21.25	199	128.1
Manish Pandey	13	249	20.75	226	110.2
Ravi Bopara	9	248	31	214	115.9
Manoj Tiwary	11	237	26.33	186	127.4

Figure 7: Example of cluster detection-4

Name	Inns	Runs	Avg.	BF	SR
Sachin Tendulkar	15	618	47.53	466	132.6
Jacques Kallis	16	572	47.66	494	115.8
Suresh Raina	16	520	47.27	364	142.8
Sourav Ganguly	14	493	37.92	419	117.7
Murali Vijay	15	458	35.23	292	156.8
Mahela Jayawardene	13	439	43.9	298	147.3
Andrew Symonds	16	429	30.64	341	125.8
Saurabh Tiwary	15	419	29.92	309	135.6
Rohit Sharma	16	404	28.85	302	133.8

Figure 8: Result on order basis on run

Name	Inns	Runs	Avg.	BF	SR
Jacques Kallis	16	572	47.66	494	115.8
Sachin Tendulkar	15	618	47.53	466	132.6
Suresh Raina	16	520	47.27	364	142.8
Mahela Jayawardene	13	439	43.9	298	147.3
Sourav Ganguly	14	493	37.92	419	117.7
Murali Vijay	15	458	35.23	292	156.8
Subramaniam Badrinath	15	356	32.36	303	117.5
Naman Ojha	14	377	31.41	285	132.3
Andrew Symonds	16	429	30.64	341	125.8
Saurabh Tiwary	15	419	29.92	309	135.6
Rohit Sharma	16	404	28.85	302	133.8

Figure 9: Result on order basis on average

Name	Inns	Runs	Avg.	BF	SR
Moises Henriques	1	11	11	5	220
RS Sodhi	1	4	0	2	200
Zaheer Khan	4	24	24	13	184.6
Harbhajan Singh	7	105	26.25	63	166.7
Srikkanth Anirudha	3	30	30	18	166.7
Stuart Binny	1	8	8	5	160
Aditya Dole	2	34	17	22	154.5
Cameron White	5	48	24	33	145.4
Siddharth Trivedi	4	20	10	14	142.8
Rajat Bhatia	3	24	8	17	141.2
Aditya Tare	4	50	12.5	36	138.9
Laxmi Ratan Shukla	2	36	18	26	138.5
Ajit Agarkar	4	40	40	29	137.9
Kedar Jadhav	5	76	25.33	56	135.7
MF Maharoof	4	31	10.33	23	134.8
MV Boucher	2	13	13	10	130
Bipul Sharma	2	9	0	7	128.6
Shanthakumaran Sreesanth	3	27	0	21	128.6
Wriddhiman Saha	7	67	22.33	53	126.4
Abhimanyu Mithun	1	5	5	4	125
Andrew McDonald	3	65	0	52	125
SA Asnodkar	2	5	2.5	4	125

Figure 10: Result on order basis on strike rate

Evaluation Parameters

Done Index: It is an evaluation parameter& quality index parameter which is the measurement of uniformity of clusters. It aims to identifying dense and well separated cluster. It is defined as the ratio between minimal inter-cluster distances to maximal intra-cluster distance for each partition. It specify the uniformity of cluster.

DB Index: It is an evaluation parameter whose value belongs to 0-1. It is also known as quality index parameter. It represent the similarity of data. The algorithm that produce cluster will low intra-cluster distance and high inter-cluster distance will have a low Davies–Bouldin index.

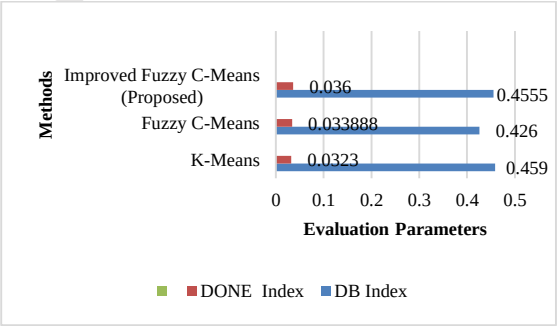


Figure 11: Comparative analysis of proposed work using graph method

Table 2: Comparative analysis of proposed research work

Method	DB Index	DONE Index
K-Means	0.459	0.0323
Fuzzy C-Means	0.426	0.033888
Improved Fuzzy C-Means (Proposed)	0.4555	0.036

V. CONCLUSION

Data Clustering plays a major role in grouping the similar type of data into a specific cluster. Cluster analysis aims at identifying groups of similar objects and, therefore helps to discover distribution of

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 3, Issue 10, May 2015)

patterns and interesting correlations in large data sets. Fuzzy clustering is an extension of the cluster analysis, which represents the affiliation of data points to clusters by memberships. In this research work, improved fuzzy c-means clustering has been adopted using fuzzy relational database to detect clusters on the IPL batting statistics dataset. The records in the database are partitioned in a manner such that similar records are in the same cluster. N-clusters have been detected from IPL batting statistics dataset. It was found that in Table 5.1 that the proposed method gives better results in terms of DB Index and Done Index. MATLAB has been used for the definition the membership function, threshold equation and detecting the several clusters.

The algorithm are developed in MATLAB for analysis and comparison. The result produced fairly higher accuracy and required less computation. The intention of the algorithm is to provide a categorization of some well–known clustering algorithm. Finally we show our experiment and prove that the improved FCM algorithm technique performance better than existing method.

REFERENCE

[1] Pabitra Kumar Dey, Gangotri Chakraborty, Purnendu Ruj, and Suvobrata Sarkar, "A Data Mining Approach on Cluster Analysis of IPL", International Journal of Machine Learning and Computing, Vol. 2, No. 4, August 2012.

[2] Adriaans, Pieter, and Dolf Zantinge, "Data mining", Addision-Wesley, Harlow, 1996.

[3] Aaron Ce., John F. R., David M.W. P., "CURIO: A Fast Outlier and Outlier Cluster Detection Algorithm for Large Datasets", AIDM'07 Proceedings of the 2nd international workshop on Integrating artificial intelligence and data mining - Vol. 84, pp: 39-47, 2007.

[4] Mrutyunjaya P., Manas R. P., "Some clustering algorithms to enhance the network performance of the network intrusion detection system", Journal of Theoretical and Applied Information Technology, Vol. 4, No: 9, pp: 795-801, 2008.

[5] Osama Abu A., "Comparison between data clustering algorithms", The International Arab Journal of Information Technology, Vol.5, No: 3, pp: 320-325, 2008.

[6] Yu Bao L. Jia R. C., Jian Y., "Clustering text data streams", Published in Journal of computer science and technology, Vol. 23 (1): pp: 113-128, Jan. 2008.

[7] M. D. Boomija," Comparison of partition based clustering algorithms", pub. in Journal of Computer Applications, Vol – 1, No.4, pp.: 18-21,Oct – Dec., 2008.

[8] O. Georgieva, F. Klawonn, "Dynamic data assigning assessment clustering of stream data", pub. in ACM Journal of Soft computing, vol. 8, Iss.4, pp: 1305-1313, September, 2008.

[9] Hristo N. Djidjev, "A scalable multilevel algorithm for graph clustering and community structure detection", pub. in Springer-Verlag, Algorithms and Models for the Web-Graph Lecture Notes in Computer Science Vol. 4936, pp 117-128, 2008.

[10] Julio Ce. P. G., Felipe P. D, "Ant Colony Algorithm for Clustering through of Cliques", pub. in proceedings of 8th Mexican International Conference on Artificial Intelligence, (MICAI ,09), 2009.

[11] O. A. Mhd. Jafar, R. Sivakumar, "Ant-based Clustering Algorithms: A Brief Survey", pub. in International Journal of Computer Theory and Engineering, Vol. 2, No. 5, pp: 17938201, October, 2010.

[12] Xiankun Y., Weihong C., "A Novel Spatial Clustering Algorithm Based on Delaunay Triangulation ", pub. in J. Software Engineering & Applications, Vol. 3, pp: 141-149, 2010.

[13] Ashish G., Niladri S. M., Susmita G., "Fuzzy clustering algorithms for unsupervised change detection in remote sensing images", pub. in ELSEVIER, Journal of Information Sciences, 2010.

[14] Dr.N.Rajalingam, K.Ranjini, "Hierarchical Clustering Algorithm - A Comparative Study", pub. in International Journal of Computer Applications , Vol. 19– No.3, April 2011.

[15] Jiawei Han, Micheline K.,"Data Mining Concepts and Techniques", Elsevier Publication, 3rd Edition, 2011.

[16] Mamta M., A. K. Sharma, P. Singh, "A Proposed Quartile Clustering Algorithm to Detect Outliers for Large Data Sets", pub. in proceedings of 5th IEEE International Conference on Advanced Computing & Communication Technologies [ICACCT], Vol. 3, pp:286-290, 2011.

[17] Dewan Md. F., Nouria Ha., "Mining Network Data for Intrusion Detection through Naïve Bayesian with Clustering", pub. in World Academy of Science, Engineering and Technology, vol. 42, 2011.

[18] Yu Zong, Guandong Xu, "A Projective Clustering Algorithm based on Significant Local Dense Areas", pub. in proceedings of IEEE World Congress on Computational Intelligence WCCI, in Brisbane, Australia, pp.: 2182-2189, 2012.

[19] J. Huang, Heli Su., "SHRINK: A Structural Clustering Algorithm for Detecting Hierarchical Communities in Networks", CIKM '10 Proceedings of the 19th ACM international conference on Information and knowledge management, pp: 219-228, 2012.

[20] Mohammed A. H., Wesam M. A., "Cosine-Based Clustering Algorithm Approach", pub. in I.J. Intelligent Systems and Applications, Vol.-1, pp: 53-63, 2012.

[21] Quan Q., Tianhong W., Rui Z., "Relative Network Entropy based Clustering Algorithm for Intrusion Detection", pub. in International Journal of Network Security, Vol.15, No.1, PP.16-22, Jan. 2013.

[22] B. Borah, D.K. Bhattacharyya, "DDSC: A Density Differentiated Spatial Clustering Technique", pub. in JOURNAL OF COMPUTERS, VOL. 3, NO. 2, pp: 72-79, Feb. 2008.

[23] G. S. David, Bejoy J. T., "A New Procedure of Clustering Based on Multivariate Outlier Detection", Journal of Data Science, vol. 11, pp.: 69-84, 2013.

[24] Han, Jiawei, Micheline Kamber, and Jian Pei, "Data mining: concepts and techniques: concepts and techniques", Elsevier, 2011.

[25] Berson, Alex, and Stephen J. Smith, "Building data mining applications for CRM", McGraw-Hill, Inc., 2002.