International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Implementation of Fuzzy Possibilistic Product Partition C-Means and Modified Fuzzy Possibilistic C-Means Clustering to Pick the Low Performers using R-Tool

T.Thilagaraj, N.Sengottaiyan

Abstract**:** ***The different techniques like clustering, classification, association rule and regression are available in data mining to deal with a huge number of datasets that are available in the education field. The main purpose of educational data mining is to extract useful information that will create a good impact on educational institutions. The identification of risk students, improving the graduation rates and placement opportunities will assess the institutional performance. The clustering is one of the famous techniques to deal with noisy and disjoint groups. The clustering technique is used to measure the distance between data objects of a similar group and also it finds the different cluster center in each iteration. The placement creates the opportunity to learn specific skills on their subject or industry and improves their knowledge in various sectors. In this paper, we are going to discuss Fuzzy Possibilistic Product Partition C-Means (FPPPCM) and Modified Fuzzy Possibilistic C-Means Clustering (MFPCM) performance while dealing with the student placement performance details. The improvement of the educational system will depend on reducing the low performing students rate. The main aim of this paper to pick the low performers by using FPPPCM and MFPCM algorithms. This will helps academia to identify the low performers and provide proper training to them in an early stage. And also the efficiency of FPPPCM and MFPCM is going to analyze with different parameters.***

Index Terms: Data mining, Fuzzy Clustering, Fuzzy Possibilistic Product Partition C-Means, Modified Fuzzy Possibilistic C-Means, Placement.

I. INTRODUCTION

The facts from a huge repository have been pulled out to make needy predictions are the main goal of data mining [1]. The data mining methods like association, classification and clustering are majorly used in the field of education [2]. The better accuracy will found in the prediction process by using classifiers which are constructed through data mining tools [3]. To find the cluster's center is the common approach made in clustering techniques to represent in each cluster [4]. In data mining k-means, neural networks, decision trees and fuzzy are the techniques that can easily extract the data in the repository related to the educational field [5]. The prime objective of data mining is a prediction that helps to take an early decision in all fields [6]. One way to improve the placement performance is to pick the low performers and it will provide the opportunity for academia to concentrate on them [7]. The clustering is a process of dividing data into a different level of groups and it shows the required perception with some form of statistics [8]. The variety of factors are available to analyze the student's performance in various aspects to understand their learning aspects [9].

To deal with the noisy data the fuzzy models will produce better performance and this will produce smarter results on a large variety of data [10]. The assignment of elements in a particular group itself is mentioned as hard clustering. The element of one group is also present is some other groups are named as fuzzy clustering [11]. In fuzzy clustering, the sharp boundaries will not found for clusters on many occasions. The crisp assignments are not made in fuzzy clustering instead of this zero and one is used [12].

The R-Tool is a statistical programming language that is open source mainly used to make analysis and to visualize the expected factors [13]. The packages and functions are available in huge numbers with different varieties and it is supported by operating systems like UNIX and windows [14]. The graphical packages will help the researchers to represent their results with a clear view [15].

II. LITERATURE REVIEW

Hu, et al. [16] Here dissimilarities measures have been considered in the categorical data by using fuzzy clustering. The Modified Possibilistic Fuzzy C-Means Model was used and found several advantages which are compared with Fuzzy Clustering models. A simple data set was analyzed from a different point of view to solve the data mining tasks. Pal, et al. [17] Here the Fuzzy C-Means model has been implemented and it generated typicality and membership values. The proposed model of the Possibilistic Fuzzy C-Means model produced the possibilities and membership values simultaneously. The comparison was made between Fuzzy Possibilistic C-Means and Possibilistic Fuzzy C-Means in numerical data.

Revised Manuscript Received on 30 July 2019.

***** Correspondence Author

T.Thilagaraj* – Department of Computer Applications, Kongu Arts and Science College, Erode -638107, Tamil Nadu, India.

Dr.N.Sengottaiyan, Director, Sri Shanmugha College of Engineering and Technology, Sankari - 637304 , Tamil Nadu, India,

Sreenivasarao and Vidyavathi [18] The Fuzzy C-Means clustering and Modified Fuzzy Possibilistic C-Means clustering models are compared with different datasets. The time complexity has been measured and also discovered functional dependencies. And also few aspects were analyzed about noisy data.

III. FUZZY POSSIBILISTIC PRODUCT PARTITION C-MEANS ALGORITHM

The Fuzzy Possibilistic Product Partition C-Means Clustering algorithm will partition a numeric vector or data set. It eliminates the Fuzzy and Possibilistic clustering algorithms outlier effects. The additive combination is not used but a multiplicative way is implemented. The main objectives of the algorithm will show in (1).

$$J_{FPPPCM}(P,Q,R,S) = \sum_{j=1}^{k}\sum_{i=1}^{n} r_{ij}^{m}\left[s_{ij}^{n} d^{2}(\vec{p_i},\vec{q_j}) + \Omega_j\left(1-s_{ij}\right)^{n}\right] \quad (1)$$

The updated Fuzzy membership degree of Fuzzy Possibilistic Product Partition C-Means Clustering algorithm is shown in (2).

$$s_{ij} = \frac{\left[s_{ij}^{n} d^{2}(\vec{p_i},\vec{q_j}) + \Omega_j\left(1-s_{ij}\right)^{-\frac{1}{(m-1)}}\right]}{\left[\sum_{j=1}^{k} s_{ij}^{n} d^{2}(\vec{p_i},\vec{q_j}) + \Omega_j\left(1-s_{ij}\right)^{-\frac{1}{(m-1)}}\right]}; \quad 1 \le i \le n, 1 \le j \le k \quad (2)$$

The Typicality degree of Fuzzy Possibilistic Product Partition C-Means Clustering algorithm is calculated in the Possibilistic part is shown in (3).

$$s_{ij} = \left[1 + \left(\frac{d^{2}(\vec{p_i},\vec{q_j})}{\Omega_j}\right)^{\frac{1}{(n-1)}}\right]; \quad 1 \le i \le n, 1 \le j \le k \quad (3)$$

The Fuzzifier m will specify clustering fuzziness $1 \le m \le \infty$. The usual assignment value chosen is 2. The Typicality exponent n is to specify the amount of Typicality in clustering. $1 \le n \le \infty$ Here also 2 is the default value. The Possibilistic penalΩty is used to control clusters variance. The prototypes of clusters are updated as follows in (4).

$$\vec{q_j} = \frac{\sum_{i=1}^{n} r_{ij}^{m} s_{ij}^{n} \vec{p_i}}{\sum_{i=1}^{n} r_{ij}^{m} s_{ij}^{n}}; \quad 1 \le j \le k \quad (4)$$

By passing the input data frame into FPPPCM function and by assigning of various parameters like the number of clusters, membership, fuzziness exponent, typicality exponent and the maximum number of iterations. The typicality degree of data objects, each level of cluster centers and membership degrees are found by using R-Tool. Table 1 implies the student's placement attributes which are aptitude, the interpersonal, academic and technical level of 17 students who are showed their interest in the software field. Table 2 shows the typicality degree of data objects and four levels of clusters were formed as low, medium, high and very high using the FPPPCM algorithm.

Table 1: Data Set of 17 Students for Placement Training

St.Id	Aptitude	Interpersonal	Academic	Technical
1	55	35	46	55
2	64	70	31	85
3	78	49	54	75
4	40	67	78	45
5	78	85	17	47
6	67	75	74	24
7	40	45	45	45
8	31	85	56	21
9	78	86	58	84
10	47	88	54	35
11	34	29	85	38
12	37	32	52	15
13	45	93	24	45
14	41	75	26	45
15	65	77	65	76
16	52	66	62	25
17	17	45	42	36

Table 2: The Fuzzy Typicality degrees of data objects by Fuzzy Possibilistic Product Partition C-Means Clustering Algorithm

Stu.Id	Low (Cluster 4)	Medium (Cluster 3)	High (Cluster 2)	Very High (Cluster 1)
1	0.3031447	0.3303345	0.3315009	0.4946162
2	0.3091421	0.386118	0.2754129	0.3385344
3	0.271963	0.4004044	0.295186	0.354504
4	0.2927512	0.3575571	0.4070859	0.3900272
5	0.3359948	0.3131568	0.2948528	0.3207729
6	0.2792294	0.3269012	0.4539903	0.3422052
7	0.345621	0.321441	0.3703511	0.906341
8	0.3296112	0.2956543	0.4019414	0.3570128
9	0.2682762	0.48963	0.2775367	0.3046177
10	0.3650684	0.3430576	0.4211664	0.3678374
11	0.2441675	0.2842247	0.3228815	0.3719734
12	0.275619	0.2662293	0.3537317	0.4146076
13	0.4634866	0.314858	0.3161974	0.3447928
14	0.8174288	0.3202163	0.3399335	0.4060164
15	0.2866177	0.9893594	0.3089796	0.3390948
16	1	0	0	0
17	0.3185026	0.2797918	0.3294146	0.461651

Table 3 describes each level of cluster centers along with the four levels and table 4 describes the maximum, minimum, mean, size, and median values of each cluster. The level of cluster 4 describes the low performer, cluster 3 describes the medium performer, cluster 2 describes the high-level performer and cluster 1 describes the very high-level performers in student placement data.

Table 3: The cluster centers of Low, Medium, High and Very High using Fuzzy Possibilistic Product Partition C-Means Clustering Algorithm

Levels	Aptitude	Interpersonal	Academic	Technical
Low (Cluster 4)	41.00314	75.00337	26.00431	44.99907
Medium (Cluster 3)	65	77	65	76
High (Cluster 2)	50.80052	71.08623	62.03092	33.13657
Very High (Cluster 1)	40.00006	45.00006	45.00007	44.99999

Table 4: Descriptive Statistics For The Membership Degrees By Clusters Using Fuzzy Possibilistic Product Partition C-Means Clustering Algorithm

Description	Low (Cluster 4)	Medium (Cluster 3)	High (Cluster 2)	Very High (Cluster 1)
Size	4	4	4	5
Min	0.3359948	0.386118	0.4019414	0.3719734
Q1	0.4316137	0.3968328	0.4057997	0.4146076
Mean	0.6542276	0.566378	0.421046	0.5298378
Median	0.6404577	0.4450172	0.4141261	0.461651
Q3	0.8630716	0.6145624	0.4293724	0.4946162
Max	1	0.9893594	0.4539903	0.906341

IV. MODIFIED FUZZY POSSIBILISTIC C-MEANS ALGORITHM

The Modified Fuzzy Possibilistic C-Means Clustering (MFPCM) algorithm is to incorporate the weight of the parameters and use every data object to calculate its weight for all clusters. The FPPPCM function is used by assigning of various parameters like input data frame, cluster size, initial membership degree, fuzzifier, typicality exponent and number of iterations. The fuzzy membership degree of data objects, levels of cluster centers and descriptive membership degrees are found. This helps to have better classification while dealing with noisy data. The objective function of MFPCM is as follows (5).

$$J_{MFPCM}(P,Q,R,S) = \sum_{i=1}^{n} r_{ij}^{m} t_{ij}^{m} d^{2m}(\vec{p_i},\vec{q_j}) + s_{ij}^{m} t_{ij}^{n} d^{2n}(\vec{p_i},\vec{q_j}) \quad (5)$$

$$t_{ij}^{m} = exp\left[-\frac{d^2(\vec{p_i},\vec{q_j})}{\sum_{i=1}^{n}\left(d^2(\vec{p_i},\vec{q_j})\frac{k}{n}\right)}\right] \quad (6)$$

The weight is calculated by using function (6). The update equations (7) are used to minimize the Modified Fuzzy Possibilistic C-Means Clustering (MFPCM) algorithm.

$$r_{ij}^{m} = \left[\sum_{j=1}^{k}\left(\frac{d^2(\vec{p_i},\vec{q_j})}{d^2(\vec{p_i},\vec{q_l})}\right)^{2m/(m-1)}\right]^{-1}; \quad 1 \le i \le n, 1 \le l \le k$$

$$s_{ij} = \left[\sum_{l=1}^{n}\left(\frac{d^2(\vec{p_i},\vec{q_j})}{d^2(\vec{p_i},\vec{q_l})}\right)^{2n/(n-1)}\right]^{-1}; \quad (7) \quad 1 \le i \le n, 1 \le j \le k$$

$$\vec{q_j} = \frac{\sum_{i=1}^{n}\left(r_{ij}^{m} t_{ij}^{m} + s_{ij}^{n} t_{ij}^{m}\right)\vec{p_i}}{\sum_{i=1}^{n}\left(r_{ij}^{m} t_{ij}^{m} + s_{ij}^{n}\right) t_{ij}^{m}}; \quad 1 \le j \le k$$

Table 5 shows the data objects membership degree by using the MFPCM algorithm the four levels of clusters were indicated as low, medium, high and very high performers. The different levels of clusters were described in Table 6. The aptitude, interpersonal, academic and technical is the student placement attributes used here. Table 7 shows the descriptive statistics of various membership degrees are size, minimum value, maximum, mean and median values are found.

Table 5: The Fuzzy Memberships Degrees Of Data Objects By Modified Fuzzy Possibilistic C-Means Clustering

Stu.Id	Low (Cluster 2)	Medium (Cluster 3)	High (Cluster 4)	Very High (Cluster 1)
1	0.10534785	0.12633494	0.14917425	0.61914296
2	0.21429043	0.55981402	0.10036865	0.1255269
3	0.11807642	0.57781596	0.13613938	0.16796824
4	0.10721931	0.13107578	0.53685188	0.22485303
5	0.5304933	0.19563018	0.15147155	0.12240497
6	0.08699097	0.0874782	0.72504726	0.10048356
7	0.01798077	0.01176909	0.02780374	0.9424464
8	0.22815372	0.08075592	0.51891377	0.17217659
9	0.04396513	0.89158479	0.03609212	0.02835797
10	0.29418003	0.10256273	0.48327034	0.1199869
11	0.10334965	0.12079084	0.30151204	0.47434748
12	0.10011681	0.06426683	0.25302073	0.58259564
13	0.94760108	0.01615975	0.02025184	0.01598733
14	0.81137304	0.04231499	0.06674123	0.07957074
15	0.0300122	0.91148261	0.03306821	0.02543698
16	0.00762319	0.00487886	0.97390271	0.01359524
17	0.11730445	0.05803732	0.15063942	0.6740188

Table 6: The cluster centers of Low, Medium, High and Very High using Modified Fuzzy Possibilistic C-Means Clustering

Levels	Aptitude	Interpersonal	Academic	Technical
Low (Cluster 2)	47.05022	86.01999	25.43414	45.13136
Medium (Cluster 3)	70.68261	76.24679	57.61915	78.97859
High (Cluster 4)	51.11287	68.55943	64.24614	27.14108
Very High (Cluster 1)	37.87198	42.05762	48.4699	41.1104

Table 7: Descriptive statistics for the membership degrees by clusters using Modified Fuzzy Possibilistic C-Means Clustering

Description	Low (Cluster 2)	Medium (Cluster 3)	High (Cluster 4)	Very High (Cluster 1)
Size	5	4	3	5
Min	0.2628787	0.3784096	0.2811725	0.2642521
Q1	0.3148208	0.3788284	0.3046571	0.3285121
Mean	0.4659957	0.5423994	0.5164483	0.4468663
Median	0.3269696	0.5077757	0.3281418	0.3733872
Q3	0.5883089	0.6713468	0.6340862	0.3805213
Max	0.8370008	0.7756367	0.9400307	0.8876588

V. RESULTS AND FINDINGS

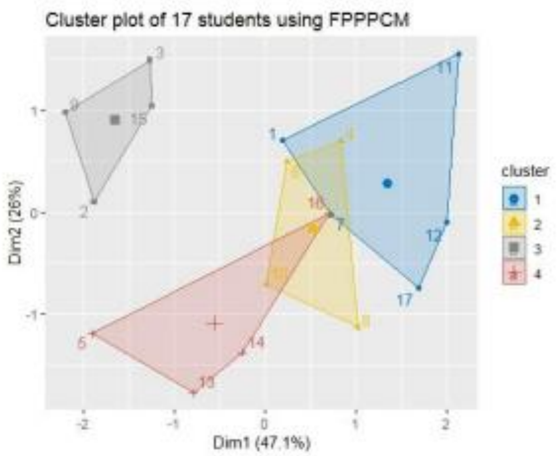


Figure 1: Result of the four clusters using the FPPPCM Algorithm.

Figure 1 shows the four levels of clusters that are low(Cluster 4), medium(Cluster 3), high(Cluster 2) and very high(Cluster 1) by implementing Fuzzy Possibilistic Product Partition C-Means Clustering Algorithm.

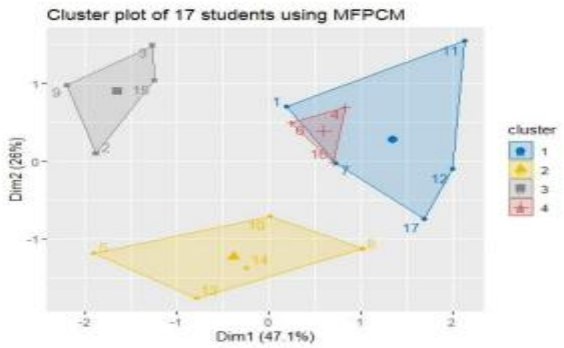
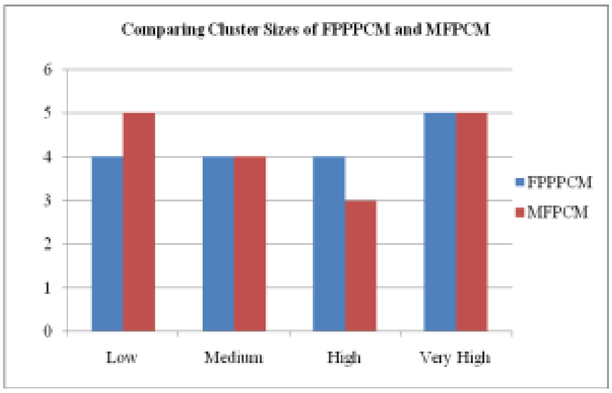


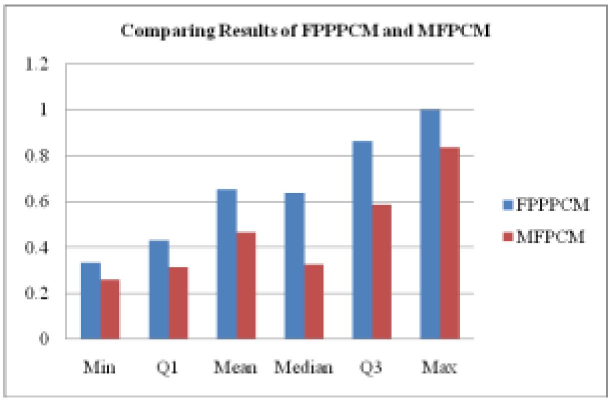
Figure 2: Result of the four clusters using the MFPCM Algorithm.

Figure 2 shows the four levels of clusters that are low(Cluster 2), medium(Cluster 3), high(Cluster 4) and very high(Cluster 1) by implementing Modified Fuzzy Possibilistic C-Means Clustering Algorithm.



Graph 1: Comparison of FPPPCM and MFPCM using clusters sizes.

Graph 1 represents the cluster sizes of FPPPCM and MFPCM to pick the low performer from student placement data. The comparison of the result shows that the MFPCM model will show better results than the FPPPCM model.



Graph 2: Comparison of various factors in FPPPCM and MFPCM models.

Retrieval Number: B3580078219/19©BEIESP
DOI: 10.35940/ijrte.B3580.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Graph 2 shows the minimum, q1, mean, median, q3 and max values of FPPPCM and MFPCM. While comparing both models the MFPCM model will show minimum levels in all categories than the FPPPCM model.

VI. CONCLUSION

Nowadays every educational institution is also focusing on a low range of performers to improve their skills for placement. Here the Fuzzy Possibilistic Product Partition C-Means and Modified Fuzzy Possibilistic C-Means clustering algorithms are used to identify the different levels of performers. The implementation of Modified Fuzzy Possibilistic C-Means Clustering Algorithm will identify the low performers with high accuracy while comparing Fuzzy Possibilistic Product Partition C-Means Clustering Algorithm. And also low performer's clusters size is high by using the Modified Fuzzy Possibilistic C-Means algorithm. Overall MFPCM algorithm will help academia to identify low performers and provide the essential training to get placed in the future.

REFERENCES

1. B. M. Varghese, A. Unnikrishnan, G. Sciencist, N. Kochi, and C. Kochi, "Clustering student data to characterize performance patterns," Int. J. Adv. Comput. Sci. Appl, vol. 2, pp. 138-140, 2010.
2. S.H. Liao, P.H. Chu, and P.Y. Hsiao, "Data mining techniques and applications–A decade review from 2000 to 2011," Expert systems with applications, vol. 39, no. 12, pp. 11303-11311, 2012.
3. X. Wu et al., "Top 10 algorithms in data mining," Knowledge and information systems, vol. 14, no. 1, pp. 1-37, 2008.
4. A. Joshi and R. Kaur, "A review: Comparative study of various clustering techniques in data mining," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 3, 2013.
5. B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," arXiv preprint arXiv:1201.3417, 2012.
6. T. Silwattananusarn and K. Tuamsuk, "Data mining and its applications for knowledge management: a literature review from 2007 to 2012," arXiv preprint arXiv:1210.2872, 2012.
7. M. Gera and S. Goel, "A model for predicting the eligibility for placement of students using data mining technique," in International Conference on Computing, Communication & Automation, 2015: IEEE, pp. 114-117.
8. P. Berkhin, "A survey of clustering data mining techniques," in Grouping multidimensional data: Springer, 2006, pp. 25-71.
9. P. S. Saxena and M. Govil, "Prediction of Student's Academic Performance using Clustering," in Natl. Conf. Cloud Comput. Big Data, 2009.
10. M. Goebel and L. Gruenwald, "A survey of data mining and knowledge discovery software tools," ACM SIGKDD explorations newsletter, vol. 1, no. 1, pp. 20-33, 1999.
11. A. Gosain and S. Dahiya, "Performance analysis of various fuzzy clustering algorithms: a review," Procedia Computer Science, vol. 79, pp. 100-111, 2016.
12. D. Vanisri and C. Loganathan, "An efficient fuzzy clustering algorithm based on modified k-means," International Journal of Engineering Science and Technology, vol. 2, no. 10, pp. 5949-5958, 2010.
13. A. Malviya, A. Udhani, and S. Soni, "R-tool: Data analytic framework for big data," in 2016 Symposium on Colossal Data Analysis and Networking (CDAN), 2016: IEEE, pp. 1-5.
14. R. Pandey, "Elective Recommendation Support through K-Means Clustering Using R-Tool," in 2015 International Conference on Computational Intelligence and Communication Networks (CICN), 2015: IEEE, pp. 851-856.
15. R. Pandey, N. Srivastava, and S. Fatima, "Extending R Boxplot Analysis to Big Data in Education," in 2015 Fifth International Conference on Communication Systems and Network Technologies, 2015: IEEE, pp. 1030-1033.
16. Z. Hu, Y. V. Bodyanskiy, O. K. Tyshchenko, and V. O. Samitova, "Possibilistic fuzzy clustering for categorical data arrays based on frequency prototypes and dissimilarity measures," International Journal of Intelligent Systems and Applications, vol. 9, no. 5, p. 55, 2017.
17. N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," IEEE transactions on fuzzy systems, vol. 13, no. 4, pp. 517-530, 2005.
18. V. Sreenivasarao and D. S. Vidyavathi, "Comparative analysis of fuzzy C-mean and modified fuzzy possibilistic C-mean algorithms in data mining," IJCST, vol. 1, no. 1, pp. 104-106, 2010.

AUTHORS PROFILE

Mr. T.Thilagaraj completed Bachelor of Computer Science from Bharathiar University, Coimbatore in 2006, Master of Computer Applications from Bharathiar University in 2009 and Master of Philosophy from Bharathiar University, Coimbatore in 2014. He is currently pursuing his Ph.D. at Bharathiar University, Coimbatore, His main research work focuses on Data Mining. He has 10 years of teaching experience which includes 2 years of P.G Experience and 5 years of Research Experience. He has published 4 research papers in reputed international journals and it's also available online. He is currently working as an Assistant Professor in the Department of Computer Applications, Kongu Arts and Science College (Autonomous), Erode.

Dr. N.Sengottaiyan completed Bachelor of Engineering from Bangalore University, Bangalore in 1986, Master of Engineering from Annamalai University, Chidambaram in 2004 and Ph.D. from Anna University of Technology, Coimbatore in 2011. He is a Member in the Computer Society of India, Life Member in Indian Society for Technical Education and Institute of Engineers India, Associate Member in IEEE, India. He has published more than 20 research papers in reputed international journals including IEEE, Springer and it's also available online. He has published 3 books. His main research work focuses on Networks, Wireless Sensor Networks, Networks Security and Steganography, Communication Systems: Space-Time Block Coding, Artificial Neural Network and Ad-hoc Networks. He has 30 years of Teaching and Research Experience. He is currently working as Director, Professor in Sri Shanmugha College of Engineering and Technology, Sankari..