Volume : 2 | Issue : 8 | Aug 2013 ISSN - 2250-1991

Research Paper Engineering

Fuzzy Metagraph Based Clustering Techniques

* A.Thirunavukarasu ** S.Uma Maheswari

* Teaching Fellow, Department of Computer Science and Engineering, Anna University, University college of Engineering- Ramanathapuram, Tamilnadu, India.

** Associate Professor, Department of Electronics and Communication Engineerng, Coimbatore Institute of Technology, Coimbatore, Tamilnadu, India.

ABSTRACT

Data analysis plays an obligatory role for understanding various phenomena like clustering and classification. Clustering is a process of grouping a set of objects into clusters based on similarity. Fuzzy metagraph is a talented technique used in the design of many information processing systems like transaction processing systems, decision support systems and workflow Systems. In this paper a new approach is proposed that allows the usage of fuzzy metagraph based clustering to represent the discovered knowledge. Some ideas on the application areas of fuzzy metagraph based clustering algorithms are given.

Keywords : metagraph, fuzzy metagraph, clustering, MST.

Introduction:
Graphs are general and powerful data structures that can be used to represent diverse kinds of objects. Clustering plays an important role in computer network, data mining and machine learning. It is considered powerful tool for model construction. Graph based clustering methods are normally based on some kind of connectivity of the nodes of a graph representing the data set. In the last three decades plenty of algorithms for graph based clustering using either optimal or approximate methods have been proposed with the specific aim of reducing computational complexity of the clustering process.

The clustering can be classified as hard clustering, soft clustering and fizzy clustering. Hard clustering is the same object can only belong to single cluster. *Hard clustering* algorithms allocate each object to a single cluster during their operation and in its output. *Soft clustering* is the same object can belong to different clusters. *Fuzzy clustering* methods assign degrees of membership in several clusters to each input pattern. Clustering and fuzzy logic together provide simple powerful techniques to model complex systems. Fuzzy logic is efficient theory to handle imprecision.

The clusters are classical (crisp) sets, a vertex is either included in a cluster or not, every vertex belongs to exactly one of the clusters, the clusters are fuzzy sets and a vertex is included in a cluster with a given grade of membership [0, 1]. Where 0 is not included, 1 is fully included, between 0 and 1 is partially included, clusters are treated as fuzzy sets, the additional constraint that the sum of the weights for each object must equal to 1. The fuzzy clustering approach is more likely to overcome the local minimum than hard clustering approach because it makes soft boundaries among clusters through the use of fuzzy membership values.

Related works
Ágnes Vathy Fogarassy and Balazs Feil have used graph theoretic divisive clustering algorithm is based on construction of the MST. They have presented a new splitting principle to improve the performance of the MST based clustering algorithm. The calculated similarities of the clusters can be used for the hierarchical clustering, in which information is useful for cluster merging and for the visualization of the clustering results [1]. Bill Andreopoulos and Michael Schroeder have presented a set of desirable clustering features that are used as evaluation criteria for clustering algorithms. They have compared different clustering algorithms of all approaches and data types on the biomedical applications [2].

Deepti Gaur and AdityaShastri proposed a model for metagraph data structure and meta network. They have used graph based data mining technique and they proposed fuzzy metagraph method of clustering to find the similar fuzzy nodes in a fuzzy metagraph. They have used Triangular Norms functions and join two or more T norms to cluster the fuzzy nodes [8-13]. Zheng-Hua proposed a Fuzzy Metagraph based knowledge representation in both graphic and algebraic format. They used introduced the concept of indexing, which was originally developed for information retrieval [30]. Thirunavukarasu and Uma Maheshwari have proposed a model for fuzzy metagraph based data structure and decision support system. They have developed an efficient algorithm for the Fuzzy metagraph ADT and have discussed about the vague metagraph based technique and their applications [22-26].

Jin-Tai Yan and Pei-Yung Hsiao have proposed fuzzy clustering based approach is implemented to obtain a better two-way partitioning with area-balanced constraints on circuit benchmarks [14]. Oleksandr Grygorash and Yan Zhou have proposed two MST based clustering algorithms. The MST clustering algorithm is to be capable of detecting clusters with irregular boundaries. The algorithm produces a k-partition of a set of points for any given k [17].

Rajul Anand and Chandan K. Reddy have presented a novel constrained graph-based clustering method based on the chameleon algorithm. They proposed a new framework for embedding constraints into the graph-based clustering algorithm to obtain talented results [20]. Satu elisa schaeffe have given an overview of some of the essential definitions and techniques of graph clustering and measures of cluster quality. They have presented global algorithms for producing a clustering for the entire vertex set of an input graph, the task of identifying a cluster for a specific start vertex by local computation [21].

M.S. Yang has taken survey of fuzzy clustering based on three categories like fuzzy relation, objective function and nonparametric classifier [27]. Young-Seol Lee and Sung-Bae Cho have proposed an efficient genetic algorithm to solve graph coloring problem in short time. The method decrease the evaluation cost by reducing the number of evaluation using fuzzy clustering [28].

3. Graph-based Clustering Methodology

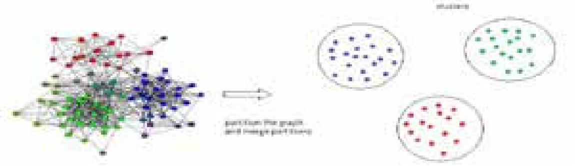

Fig .1. Graph based clusters is constructed from a graph

Figure 1 shows an example of graph based clustering method, the graph may be direct or indirect, if graph contains cycle or loop then we can merge the loop of vertices into single node or metagraph vertex. So that graph is transfer to metagraph with help of clustering method. Figure 2 shows an example of a graph is transfer to metagraph.

A graph is a triple $G=(V,E,W)$ where $V=\{v_1,v_2,v_3,\ldots,v_n\}$ is a set of vertices, $E \subseteq V \times V$ is a set of edges, and $W=(W_{ij})$ is called adjacency matrix in which each element indicates a non-negative weight $W_{ij} \geq 0$ between two vertices v_i and v_j. let $X=\{x_1, x_2, \ldots, x_n\}$ be a set of data points, $S = S_{ij}$ be the similarity matrix in which each element indicates the similarity $S_{ij} \geq 0$ between two data points x_i and x_j. S is also generalized adjacency matrix, Degree of node $d_i = \sum_j S_{ij}$.

The general methodologies of graph-based clustering are Hypothesis, Modeling, Measure, Algorithm and Evaluation

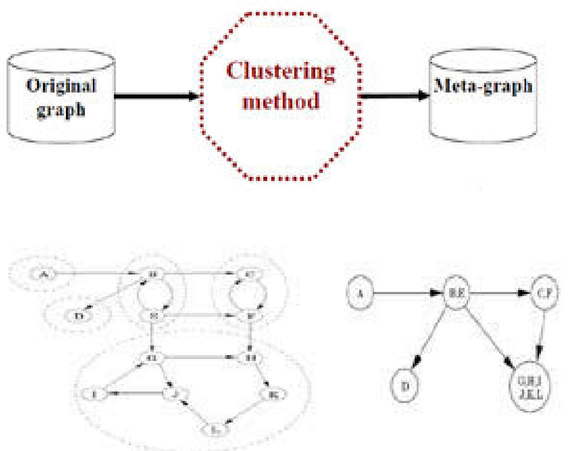

Fig .2. Example of a Graph is Transfer to Metagraph

.

4. Metagraph and Fuzzy Metagraph

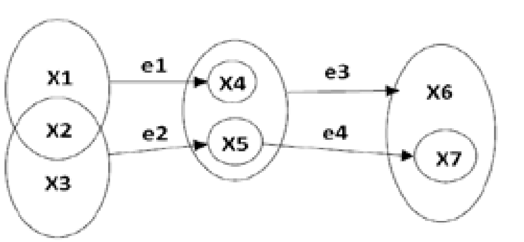

Fig. 3. Example of metagraph

A metagraph S = {X, E} is a graphical representation consisting of two tupelos X and E. Here X is its generating set and E is the set of edges defined on generating sets. Figure 3 shows an example of metagraph X = {x1, x2, x3, x4, x5, x6, x7} is the generating set and E = {e1, e2, e3, e4} is the set of edges. The edge set can be specified as E = {<{x1, x2}, {x4}>, <{x2, x3}, {x5}>, < {x4, x5}, {x6, x7}>, < {x5}, {x7}>}. In-vertex is a function having one argument which can find out the internal vertices from a given set. In-vertex (< {x4, x5}, {x6, x7}>) = {x4, x5}. Out-vertex is another function having one argument which can find out what are the out vertices from the given set. Out-vertex (< {x4, x5}, {x6, x7}>) = {x6, x7}. Generally the edges of the metagraph are labeled as: e1 = <{x1, x2}, {x4}>, e2 = <{x2, x3}, {x5}>, e3 = <{x4, x5}, {x6, x7}>, e4 = < {x5}, {x7}> [9, 24]. *Let* $G = (V, \sigma, \mu)$ be a fuzzy graph if V is the non-empty set of vertices or nodes.$\sigma : V \rightarrow [0,1]$, a fuzzy node set of G , $\mu : V \times V \rightarrow [0,1]$, a fuzzy edge set of G such that for all $x, y \in V$, $\mu(x, y) \leq \sigma(x) \wedge \sigma(y)$.

A fuzzy metagraph is a triple $\tilde{S}=\{X, \tilde{X}, \tilde{E}\}$ in which $\tilde{X}$ is a fuzzy set on X and $\tilde{E}$ is a fuzzy relation on X×X A fuzzy set X on X is completely characterized by its membership function μ:X→ [0, 1] for each x□ X, μ(x) is the truth value of the statement of "x belongs to $\tilde{X}$". $\tilde{E}$ is a fuzzy edge set $\{\tilde{e}m$,m=1, 2, 3,.. m}. Each component $\tilde{e}$ in $\tilde{E}$ is characterized by an ordered pair < $\tilde{V}_m, \tilde{W}_m$>. In the pair $\tilde{V}_m$ subset of $\tilde{X}$ is the in-vertex of $\tilde{e}_m$ and $\tilde{W}_m$ subset of $\tilde{W}$ is the out-vertex. The membership value of an edge is also called certainty factor (CF) of the edge.

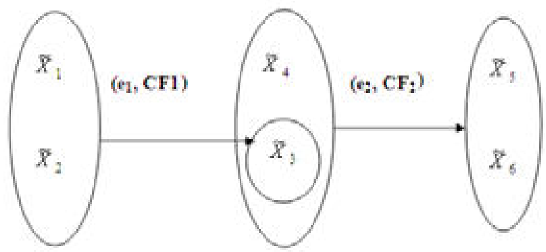

Fig. 4. Example of Fuzzy Metagraph

Figure 4 shows the fuzzy metagraph whose element set is X = { $\tilde{x}1$, $\tilde{x}2$,..., $\tilde{x}6$ } is known as fuzzy Meta Node and whose edge set consists of: $e1$ = <{ $\tilde{x}1$, $\tilde{X}2$ }, { $\tilde{X}3$ }> and $\tilde{e}2$ = <{ $\tilde{X}3$, $\tilde{x}4$ }, { $\tilde{x}5$, $\tilde{X}6$ }> , the in-vertex and out-vertex of $\tilde{e}$ 1 are {$\tilde{X}1$, $\tilde{X}2$} and {$\tilde{x}3$}.

5. Metagraph and Network Based Clustering

In computer networks, clustering may be used to identify relevant substructures and to analyse the connectivity for purposes of modelling or structural optimization. For networks with a dynamic topology, with frequent changes in the edge structure, local clustering methods prove useful, as the network nodes can make local decisions on how to modify the clustering to better reflect the current network topology. Extraordinary a cluster structure on a dynamic network eases the routing task .The Minimum Spanning Tree (MST) is subgraph of a graph, all the vertices must be there and edges may be or may not be there, it doesn't have cycle and also optimal, it always looks best at the instant [1,8,17].

Figure 5 shows an example of the undirected based weighted metagraph can be transformed in to MST based weighted metagraph. This technique will helpful for QoS of shortest path method and clustering. From the above MST based metagraph, the total cost or total weight is 20, it is also optimal solution and follows greedy algorithm techniques or divides and conquer technique.

Divide and conquer technique for MST problem procedure

1. Divide the graph into N sub-graph by clustering.
2. Solve each sub-problem separately using MST algorithm and

3. Merge the sub-solutions and solve MST for the meta graph

The MST based metagraph- clustering is to defined a mapping, the output of this algorithm shows the cluster as a set of ordered pairs (t_i , j) where f(t_i) = K_j.

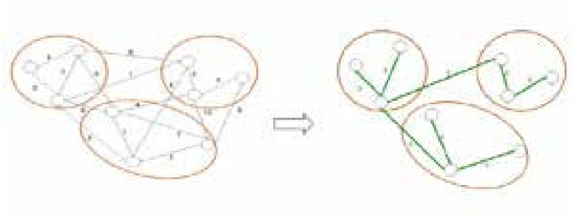

Fig .5. Example of a weighted Metagraph is Transfer to MST based Metagraph

Algorithm MST based metagraph
Input: D={t1,t2,t3,..,tn} // set of elements
A // adjacency matrix showing distance (cost) between elements of metagraph

K // Number of desired clusters

Output: f // mapping represented as a set of ordered pairs

Partitional MST algorithm

M= MST(A)

Identify inconsistent edges in M, remove k-1 inconsistent edges and Create output representation.

End
Metapath is a set of edges connecting a set of source elements to a set of target elements. The connectivity properties of metagraphs can be used to determine whether a specific collection of models is sufficient to calculate a set of target variables from a set of input variables, possibly under a set of assumptions. MST-based metagraph are working to lead the splitting and merging process. The cost of constructing a MST is *O (m log n)*, where *m* is the number of edges in the graph and *n* is the number of vertices. A Euclidean minimum spanning tree (EMST) is a spanning tree of a set of *n* points in a metric space (E^n), where the length of an edge is the Euclidean distance between a pair of points in the point set.

The QoS routing can be formally defined as the problem that consists in finding an optimal-cost path or set of feasible cost paths from source to destination(s) subject to one or more constraints on the path. Scalability and efficiency of the QoS routing may be achieved by in network clustering. In networks based clustering, cluster based control structures provide more efficient use of resources for large dynamic networks and clustering can be used for transmission management and routing efficiency.

According to Wireless Sensor Network (WSN), the sensor nodes are grouped into many clusters. The grouping of sensor nodes is known as clustering (figure 6). In cluster formation, every cluster has a leader which is known as cluster head (CH). A CH is also one of the sensor nodes which have advanced capabilities than other sensor nodes. The CH is selected by the sensor nodes in the relevant cluster and may also possible by the user to pre-assign the CHs [7, 21]. Figure 6 shows an example of the network or WSN based clustering, each cluster has a CH, one or more gateways and zero or more ordinary nodes. CH schedules transmission and allocates resources within its cluster. Gateways connect adjacent clusters. CHs are Responsible for resource allocation and maintain network topology, election of CH optimally is critical. It acts as routers like forwards packets from one node to another.

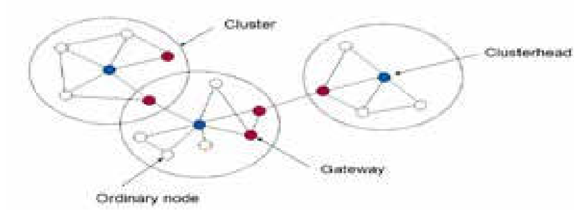

Fig .6. Example of network based clustering

In order to decrease the number of nodes and data transmission time, network clustering contribute in the routing process, nodes that share the same class of equivalence are grouped together and form a cluster. A cluster itself can be viewed as a logical node like metanode. The network topology of the metanodes may be a metagraph. Once the network of n nodes is clustered into k clusters, the routing overhead is reduced by the factor of n / k. For scalability, we need the number of clusters k to be small. In any communication network, metagraph based clustering serves as a tool for analysis, modelling and prediction of the function, usage and end of the network.

6. Fuzzy graph and fuzzy metagraph based clustering algorithm
The main advantage of the fuzzy graph concept is the very compact and easy to understand representation of a function. Graph based theoretic clustering methods are normally based on some kind of connectivity of the nodes of a graph representing the data set. Figure 7 illustrate a typical example of cases in which simply remove the *k*-1 longest edges does not necessarily output the desired fuzzy graph based cluster structure.

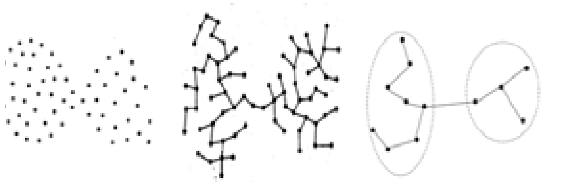

Fig .7.Clusters connected through a point of fuzzy graph

An un directed edge weighted graph defined as G = (V, E), Where V = { x 1, x2 ,..., xn } and E = { y 1, y2 ,..., ym}. Fuzzy clustering constraints on separated groups are formulated , the grades of membership and the fuzzy functions in clustering groups for the vertex set V are introduced as follows : Every function u_i : V→ [0,l] is said to assign to each x_k ∈V its grade of membership in the i^{th} fuzzy group. The fuzzy function u_i is called the i^{th} fuzzy set of V. In order to partition V by means of fuzzy sets, there exist some clustering constraints among the fuzzy membership of the vertex set V. Given a vertex set V and M2n is the set of real 2 x n matrices [14, 25, 28]. V = (xl, x2 ,..., xn), a fuzzy graph partition of V can be represented by a fuzzy matrix U∈M2n whose entries satisfy the following conditions.

1. Row i of U, say U_i = (u_{il}, u_{i2}) exhibits the i^{th} membership function (or i^{th} fuzzy set) of V.
2. Column j of U, say Uj = (u_{lj}, u_{2j}) exhibits the values of the 2 membership functions of the j^{th} datum in V.
3. u_{ik} shall be interpreted as $u_i(x_k)$, the value of the Membership function of the i^{th} fuzzy set for the k^{th} datum.
4. The sum of the membership values for each x_k is 1 (u_{l-} $+u_{2k}$=l for all k).
5. No fuzzy subset is empty and No fuzzy subset is all of V.

Let U ∈ M2n be a fuzzy graph partition of V and let v = (v1, v2) be the group centers, vi $\in$ V. The ith objective function J_i :

$$J_i(U, v_i) = \sum_{k=1}^{n} (u_{ik})^2 (d_{ik})^2$$

$J(U, v) = J_1(U, v_1) + J_2(U, v_2)$

Fuzzy graph clustering can be stated as an approach which attempts to find a solution for the following mathematical program: Minimize J (U, v)

$$= \sum_{i=1}^{2} \sum_{k=1}^{n} (u_{ik})^2 (d_{ik})^2$$

Subject to

$u_{lk} + u_{2k} = 1, u_{ik} \geq 0, 1 \leq k \leq n, l \leq i \leq 2,$

$x_i \in V$, $1 \leq i \leq n$ are vertices in the graph, $v_j \in V$, $1 \leq j \leq 2$ are unknown cluster centers,

U = { u_{ik} } is a 2 x n matrix, u_{ik} is referred to as the grade of membership of x_k in the ith row of the matrix U. According to necessary conditions of U and v, fuzzy graph clustering, via iterative optimization of J(U, v) on U and v, produces a feasible fuzzy graph partition of the vertices V = { xl, x2 ,..., xn). The basic steps of the Fuzzy graph clustering algorithm are given as follows:

Algorithm Fuzzy Graph Clustering
Input : An undirected edge-weighted graph;
Begin

1) Determine the clustering distance d_{ij} between x_i and x_j, $1 \leq i, j \leq n$,
2) Initialize an arbitrary partition and establish a fuzzy matrix U,
3) Calculate the centers $v = (v_1, v_2)$ using U as follows:

(1) Determine v_1 such that $J_1(U, v_1)$ is the least,
(2) Determine v_2 such that $J_2(U, v_2)$ is the least,
4) Calculate a new fuzzy matrix U' using $v = (v_1, v_2)$ as follows

for k=1 to n do

if ($x_k \neq v_1$ and $x_k \neq v_2$)then

$P = d^2_{1k} \times d^2_{2k}$, $Q = d^2_{ik} (d^2_{1k} + d^2_{2k})$

$u'_{ik} = \frac{P}{Q}$, $1 \leq i \leq 2$

else $u'_{ik} = 1$ if $x_k = v_i$

$u'_{ik} = 0$ if $x_k \neq v_i$, $1 \leq i \leq 2$

end if

5) Compare U and U';

$$\text{if } |u'ik - uik| < \in$$

for $1 \leq i \leq 2$, $1 \leq k \leq n$ then stop

otherwise, U = U', and Goto Step 3.

End

Conclusion:

Clustering plays an important role in computer network, data mining and machine learning. Recognition of groups so that similar objects belong to the same group, dissimilar objects belongs to different groups. Metagraph and fuzzy graph based clustering method may have compact time complexity and space complexity. In this paper, we have developed hybrid clustering algorithm to fuzzy metagraph based clustering algorithm with help of fuzzy graph clustering and MST. Network clustering and Fuzzy metagraph based Clustering algorithms may have used in the structural design, operation of Adhoc and sensor networks.

REFERENCES

[1] Ágnes Vathy-Fogarassy, Balázs Feil, and János Abonyi , " Minimal Spanning Tree based Fuzzy Clustering", World Academy of Science, Engineering and Technology, vol. 8,pp 499-504, 2007. | [2] Ajith Abraham , "Rule-Based Expert Systems", Handbook of Measuring System Design, Oklahoma State University, Stillwater, USA, pp. 909-919,2005. | [3] A. Basu and R.W. Blanning, "Metagraphs A Tool for Modeling Decision Support Systems," Management Science, vol. 40, pp. 1579-1600, 1994. | [4] A. Basu and R.W. Blanning, "Workflow Analyasis Using Attributed Metagraphs", Proc. 34th Hawaii Int'l Conf. System Sciences, pp.3735-3743, 2001. | [5] A. Basu and R.W. Blanning, "Synthesis and decomposition Of processes in organizations", Information Systems Research 14 (4), pp.337–355, 2003. | [6] Bill Andreopoulos, Aijun An, XiaogangWang and Michael Schroeder, "A roadmap of clustering algorithms: finding a match for a biomedical application", Briefings in bioinformatics, vol. 10, pp.297-314, February, 2009. | [7] S. R. Boselin Prabhu, S. Sophia, "A Review of Energy Efficient Clustering Algorithm for Connecting Wireless Sensor Network Fields", International Journal of Engineering Research & Technology, Vol. 2, pp. 477-481,April 2013. | [8] Deepti Gaur, AdityaShastri, Ranjit Biswas, " Metagraph a new hierarchical data structured as a decision tree", the journal of computer science and information technology, ISSN 0973-4872, vol. 6, no. 1, pp. 1-5, July -Dec. 2007, institute of technology and management. | [9] Deepti Gaur, AdityaShastri, Ranjit Biswas, "Metagraph: A new model of Data Structure", IEEE International Conference on Computer Science& Information Technology, pp 729-733, 2008. | [10] Deepti Gaur, AdityaShastri, Ranjit Biswas, "Metagraph-Based Substructure Pattern mining", IEEE International Conference on Advanced Computer Theory and Engineering, pp 865-869,2008. | [11] Deepti Gaur, AdityaShastri, Ranjit Biswas, "Meta-Network: A New Concept in Network Technology", International Journal of Maritime Information and Communication Sciences, Vol. 6, Number 4, pp. 470-474,Dec. 2008, Korea. | [12] Deepti Gaur, AdityaShastri, Ranjit Biswas, "Fuzzy Meta Node Fuzzy Metagraph and its Cluster Analysis",Journal of Computer Science, vol.4 , pp. 922-927, 2008, India. | [13] Deepti Gaur, AdityaShastri, Ranjit Biswas, "Vague Metagraph", International Journal of Computer Theory and Engineering, Vol.1, No.2, pp.126-130,2009, India. | [14] Jin-Tai Yan and Pei-Yung Hsiao, " A New Fuzzy-Clustering-Based Approach for Two-Way Circuit Partitioning",IEEE,8th International Conference on VLSI Design , pp. 359-364,January 1995. | [15] Jong-Shenge Cherng,Mei-Jung Lo , "A Hypergraph based clustering algorithm for spatial data sets", IEEE, pp.83-90, 2001. | [16] E. Mehdizadeh, S. Sadi-nezhad and R. Tavakkoli-moghaddam, "Optimization of fuzzy clustering criteria by a hybrid PSO and fuzzy c-means clustering algorithm", Iranian journal of fuzzy systems , Vol. 5, No. 3, pp. 1-14,2008. | [17] Oleksandr Grygorash, Yan Zhou, Zach Jorgensen," Minimum Spanning Tree Based Clustering Algorithms", School of Computer and Information Sciences University of South Alabama,pp.1-9, 2005. | [18] Pang Ning Tan, Vipin Kumar and Michael Steinbach, "Introduction to Data Mining", pearson, 2011. | [19] Pasquale Foggia, Gennaro Percannella, "A Graph-Based Clustering Method and Its Applications",Springer-Verlag Berlin Heidelberg, pp. 277 – 287, 2007. | [20] Rajul Anand and Chandan K. Reddy, " Graph-Based Clustering with Constraints", Springer-Verlag Berlin Heidelberg, Part II, LNAI 6635, pp. 51–62, 2011. | [21] Satu elisa schaeffe, "Survey Graph clustering", Elsevier, Computer science review 1 , pp. 27 – 64,2007. | [22] Thirunavukarasu and Uma Maheshwari, "Fuzzy metagraph based knowledge representation of Decision support system", international journal of computer engineering and Technology, ISSN 0976 – 6367(print), ISSN 0976 – 6375(online), vol.3, issue 2 , pp. 157-166, July- September 2012. | [23] Thirunavukarasu and Uma Maheshwari, "Fuzzy Metagraph and Vague Metagraph based Techniques and their Applications", International Journal of Computer Applications, vol. 56, pp. 6-11, October 2012. | [24] Thirunavukarasu and Uma Maheshwari, "Design and Analysis of Fuzzy Metagraph Based Data Structures", International Journal on Computer Science and Engineering, ISSN: 0975-3397, vol.4, No, pp. 1830-1835, 2012. | [25] Thirunavukarasu and Uma Maheshwari, "An Efficient algorithm for Fuzzy based Metagraph ADT", International Journal of Scientific and Research Publications, vol.2, Issue 12,ISSN 2250-3153,pp 1-7, December 2012. | [26] Thirunavukarasu and Uma Maheshwari, "Fuzzy Rule based Metagraph Data Structures and DSS", International Journal of Advances in Science and Technology, vol.5, ISSN 2229 5216, pp. 9-14, December 2012. | [27] M.S. Yang, " A Survey of Fuzzy Clustering", Math Comput Modelling, vol.18, no. 11, pp. 1-16, 1993. | [28] Young-Seol Lee and Sung-Bae Cho, "Solving Graph Coloring Problem by Fuzzy Clustering Based Genetic Algorithm", Springer-Verlag Berlin Heidelberg , LNCS 7673, pp. 351-360, 2012. | [29] Zheng Chen, Heng Ji, "Graph-based Clustering for Computational Linguistics: A Survey", Proceedings of the 2010 Workshop on Graph-based Methods for Natural | Language Processing, pp. 1–9, Uppsala, 2010. | [30] Zheng Hua Tan, Senior Member, "Fuzzy Metagraph and Its Combination with the Indexing Approach in Rule-Based Systems", IEEE transactions on knowledge and data Engineering, vol. 18, no. 6, pp. 829-841, June, 2006. | [31] Zhong, Duoqian Miao, Pasi Fränti , " Minimum spanning tree based split-and-merge: A hierarchical clustering method", Elsevier, Information Sciences 181, pp.3397–3410, 2011.