ITERATIVE HEURISTIC ALGORITHM FOR OPTIMUM LAYOUT OF CLUSTERS

Rekha Bhowmik
Department of Computer Science
Sam Houston State University
Huntsville, TX 77340
rbb002@shsu.edu

Sultan Aljahdali
Department of Computer Science
Taif University
Taif, Saudi Arabia

Abstract

This paper presents an Iterative Heuristic Algorithm for optimal layout of clusters for three-dimensional layout planning. The use of cluster analysis is proposed for grouping highly related objects. The vertical location problem for locating clusters on different levels is formulated. Intergroup Adjacency Matrix is developed by grouping closely related objects. Clusters are located on different levels based on the order of importance of clusters and also on the area restriction on each level.

Keywords

Iterative Heuristic Algorithm, Cluster Analysis, Intergroup Adjacency Matrix, Travel Cost, Three-Dimensional Layout Planning

1. Introduction

Unlike classification and prediction, clustering analyzes data objects without consulting a known class label. The objects are clustered or grouped based on the principle of maximizing the intra-class similarity and minimizing the interclass similarity. That is, clusters of objects are formed so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters.

Data clustering is under tremendous development [5,11]. Research areas include data mining, statistics, marketing, spatial database technology, machine learning and biology. Owing to the huge amounts of data collected in databases, cluster analysis has been an active topic in data mining research. As a branch of statistics, cluster analysis has been studied extensively over the years, focusing on distance-based cluster analysis. This paper presents clustering algorithm based on Euclidean-distance measure for location of clusters on different levels. There are a number of clustering algorithms available but we developed a program to use the hierarchical clustering method for grouping data objects into a tree of clusters. Our aim is to locate clusters on different levels, so that closely related objects do not split over levels. We discuss the use of clustering technique for identifying the groups of highly inter-related objects.

In this paper the emphasis will be on identifying the groups of highly interrelated objects, locating the clusters on different levels and using Iterative Heuristic Algorithm.

2. Cluster Analysis Approach

In layout planning, the location order calculations are not designed to isolate groups of closely related objects and that the location process is only able to optimize the location of an object with respect to the locations of pre-located objects. Also the processing time is very high for large-sized problems and that the objects tend to split between levels which may not be acceptable [6, 9, 10].

3-dimensional layout problem can be written as:

$$\text{Minimize Z} = \frac{1}{2}\sum_{l=1}^{f}\sum_{k=1}^{f}\sum_{i=1}^{n}\sum_{j=1}^{n}\big(x_{ik}\,x_{jl}\,t_{ij}(h_{kl}w + d'_{ij}\,\delta_{kl}\big) + (1-\delta_{kl})(d'_{ic} + d'_{jc}))$$

.....(1)

subject to

$$\sum_{i=1}^{n}(x_{il}\,a_i \le A_l) \qquad l = 1,2,\ldots,f$$

$$\sum_{l=1}^{f}(x_{il} = 1) \qquad i = 1,2,\ldots,n$$

where h_{kl} = absolute value of the vertical distance(in units) between k^{th} and l^{th} levels

$$\delta_{kl} = 0 \qquad for\ k \neq l$$
$$\delta_{kl} = 1 \qquad for\ k = l$$

d'_{ij} = horizontal distance between the location of objects i and j when both the objects are on the same level.

d'_{ic} and d'_{jc} = horizontal distance from object i and j to the circulation point of objects.

This is a mixed integer nonlinear programming problem of great complexity. The first term of the distance expression represents the weighted vertical travel, the second term represents the horizontal travel when the objects are located on the same level. The variables for the problem are:

$$x_{ik} = 0 \text{ or } 1 \qquad i = 1,2,\ldots,n \quad k = 1,2,\ldots,\mathrm{f}$$

$$(x^{(i)}, y^{(i)}) \in R$$

where R is the prescribed boundary of the level layout.

Rewriting the objective function(1) into three parts:

$$\mathrm{Z} = \frac{1}{2}\sum_{l=1}^{f}\sum_{k=1}^{f}\sum_{i=1}^{n}\sum_{j=1}^{n}\big(x_{ik}\,x_{jl}\,t_{ij}\,h_{kl}\mathrm{w}\big)$$

$$+ \quad \frac{1}{2}\sum_{l=1}^{f}\sum_{k=1}^{f}\sum_{i=1}^{n}\sum_{j=1}^{n}\big(x_{ik}\,x_{jl}\,t_{ij}(1-\delta_{kl})\ (d'_{ic} + d'_{jc})\big)$$

$$+ \quad \frac{1}{2}\sum_{l=1}^{f}\sum_{k=1}^{f}\sum_{i=1}^{n}\sum_{j=1}^{n}\big(x_{ik}\,x_{jl}\,t_{ij}d'_{ij}\,\delta_{kl}\big)$$

Here, the third term of the objective function represents the sum of intra-level travel costs. The first term represents the inter-level vertical travel cost while the second term represents the inter-level horizontal travel cost.

If the objects with high interactions are suitably clustered, it amounts to the vanishing of contributions from the first two terms. For moderate and small values of trips, t_{ij}, the objects are not expected to be located on the same level and the contribution from the third term vanishes. Further, for objects located on different levels with moderate values of t_{ij}, the first term will dominate the second term. Thus, if the objective function is approximated as the sum of the first and third terms, ignoring the second term, the problem splits into a partitioning problem in the domain of quadratic assignment problem.

The use of cluster analysis technique to maximize the adjacency within subsets of objects while minimizing the travel cost between clusters appears to be an extremely good approach. Thus after cluster formation, the layout problem constituting the first term of the objective function is minimized and later the level-wise layout problem is solved by minimizing the third term of objective function with appropriate set of constraints.

Thus, a three step procedure is used to solve the three-dimensional layout problem. They are: i) Use of clustering technique for identifying groups of highly interrelated objects, ii) Use of an exact or efficient algorithm for minimizing intergroup adjacency cost, and iii) Use of iterative heuristic algorithms for obtaining layouts of objects at each level.

In the following section, no discussion of step ii) is presented since the algorithm has been described elsewhere[2].

3. Procedure for Multi-Level Location Problem

The cluster analysis procedure is a four stage process. In the first stage, the cost of interaction(trips) is specified. Based on this we obtain a dendogram showing the successive fusion of objects, which culminates at the stage where all the objects are in one group. In the second stage, object areas are inserted which splits the single group into clusters of closely related objects, each of which is small enough to accommodate on a level. We determine the *Intergroup Adjacency Matrix* as discussed in section 3.1. In the third stage, vertical layout problem is carried out. This is discussed in section 4. The last stage of the process consists of locating objects on different levels using a two-dimensional layout procedure.

3.1 Clustering Objects and Determining Intergroup Adjacency Matrix

The procedure for clustering and determining inter-group adjacency matrix involves: i) Develop the *Adjacency Matrix* between pairs of objects. This is the number of trips between objects. ii) Find the largest number of interaction(trips) between pairs of objects from the adjacency matrix. This is the cluster level to start with the cluster analysis procedure. Choose some cluster level interval. The pairs of objects which fall in this cluster level form a cluster and is designated by some cluster name for the purpose of identification. Decrease the cluster level by the cluster level interval chosen. Find the objects which fall in this cluster level. We go in for the third and subsequent cluster levels by further reducing by the cluster level interval. In this way, the objects falling in a particular cluster level are searched and identified by cluster name. iii) Plot the dendogram, and iv) A search is made in the reverse direction to consider clusters of desired area in square units. If a cluster has an area less than the maximum permissible area per level, the identity and size of the cluster are stored in a table. A check is made for the non-repetition of an object. v) Construct an *Intergroup Adjacency Matrix* representing the interaction costs between clusters.

The element T_{ij} of the *Inter-group Adjacency Matrix* is given by:

$$T_{ij} = \sum_{k \in I} \sum_{l \in J} t_{kl}$$

where, C_i is obtained by grouping of objects belonging to the set I and C_j is another cluster representing the group of objects belonging to the set J.

3.2 Example

An example has been studied by using clustering algorithm for three-dimensional layout problems. An adjacency matrix is considered with 21 objects. These objects are to be clustered and the adjacency matrix containing the communication values between objects is known. The communication value, t_{ij} between objects i and j can be obtained from the adjacency matrix. At level 1 of the clustering procedure, objects 3 and 4 are fused to form a cluster, since t_{34} is the largest communication value in the adjacency matrix. The number of communication values between this and the remaining 19 objects are obtained. Next largest entry is 182, and so objects 12 and 13 are fused to form a second group. The next largest entry is

151 and so objects 10 and 11 are fused to form the third group. All the groups are designated by cluster names. Since the number of objects for this example is 21, the first group is named as 22. Finally, fusion of the groups takes place to form a single group containing all the 21 objects. The dendogram is thus created. Since the maximum permissible area per level is 19 sq units, the groups which have area less than or equal to 19 sq units are listed. Table 1 shows the object(s) forming the cluster and cluster area for three-dimensional layout problem.

Table 1- Identity and size of cluster on three levels

Number of clusters=12

Cluster Number	**Object(s) forming the Cluster**	**Cluster Area**
1	1, 2, 17 (group number=38)	8
2	3, 4, 20, 21 (group number=28)	4
3	9,10,11,14 (group number=27)	18
4	12, 13 (group number=23)	6
5	5	1
6	6	1
7	7	3
8	8	3
9	15	3
10	16	5
11	18	1
12	19	2

Table 2 shows the intergroup adjacency matrix representing the travel costs between clusters.

Table 2- Intergroup adjacency matrix representing the travel costs between clusters

Number of clusters=12

Cluster No.	Cluster units	1	2	3	4	5	6	7	8	9	10	11	12
1	8	0											
2	4	118	0										
3	18	72	90	0									
4	6	60	265	515	0								
5	1	6	25	0	11	0							
6	1	3	2	2	4	3	0						
7	3	21	18	58	165	0	2	0					
8	3	3	44	135	64	0	3	56	0				
9	3	3	44	135	64	0	3	8	56	0			
10	5	21	18	58	165	0	2	62	8	56	0		
11	1	38	37	3	59	1	0	3	0	0	3	0	
12	2	37	116	20	56	0	0	2	7	7	2	2	0

4. Vertical Layout Problem

The vertical layout optimization problem[1] can be written as

$$C = \frac{1}{2}\sum_{l=1}^{f}\sum_{k=1}^{f}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ik}\, x_{jl}\, t_{ij}\, d_{kl})$$

where x_{ik} = 1 if the i^{th} object is located on k^{th} level

= 0 otherwise

x_{jl} = 1 if the j^{th} object is located on l^{th} level

= 0 otherwise

t_{ij} = number of interactive trips between the objects i and j

d_{kl} = vertical distance between the k^{th} and l^{th} level

$= |k - l|$

n = number of objects

f = number of levels

The constraints are:

$$\sum_{i=1}^{n}(t_{il} \cdot a_i \le A_l) \qquad l = 1, 2, \dots, f \qquad \text{......(2)}$$

$$\sum_{i=1}^{f}(x_{il} = 1) \qquad i = 1, 2, \dots, n \qquad \text{......(3)}$$

where a_i is the area required for object i and A_l is the available space on level l. Constaint (2) represents the restriction of available space on each level while the constraint (3) models the condition that a particular object must be located on any one of the levels[8].

The above problem is a quadratic assignment problem. There exists no reliable exact algorithms which can solve quadratic assignment problem where the number of objects is greater than 12. However, since the number of levels for medium sized problems is usually small and the number of clusters is much smaller than the number of objects, it is possible to attempt an exact solution if a suitable algorithm can be developed. An iterative heuristic algorithm is discussed in section 4.1 which is simpler and easier to implement.

4.1 Iterative Heuristic Algorithm for Multilevel Problem

We discuss below an Iterative Heuristic Algorithm for locating clusters to minimize vertical communication costs. In each iteration a hierarchical procedure for two-dimensional layout problem is made use of. We do not discuss the two-dimensional layout problem in this paper.

The iterative algorithm is described below.

Step 1. Construct an Intergroup Adjacency Matrix representing the communication costs between clusters. Set $i = 1$. Set all d_{ij}'s to unity.(d_{ij} is the distance between the pre-located cluster and the next cluster to be located)

Step 2. Compute the travel cost matrix(T_{ij}) wherein each element of the matrix is given by:

$$t_{ij} * d_{ij} = T_{ij}$$

Step 3. Rank the clusters for location on the basis of travel cost. The cluster having the maximum travel cost with other clusters should be ranked first. The cluster having the largest travel cost with the previously located cluster should be ranked next for location. Thus, at any step, the cluster having the maximum sum of travel costs with all the previously located clusters will be ranked next. Hence a complete ordering of clusters can be established.

Step 4. Locate the first cluster at the middle level(m^{th} level). Consider the next cluster in the priority list. Determine the optimal location of this cluster from cost considerations. It must be ensured that in a particular level sufficient area is available for locating the entire cluster. If sufficient area is not available, optimal location is attempted in the $(m-1)^{th}$ level having unfilled level areas. The procedure is repeated for all clusters until the entire level is filled. It should be noted that the problem is a constrained one on account of the limitation on the number of levels and available area on each level.

Step 5. As in Step 4, complete the $(m+1)^{th}$ level.

Step 6. Repeat Steps 4 and 5 until all the clusters are located.

Step 7. Calculate the vertical distance representing the weighted distance between clusters(or between levels) taking the middle level as the centre point. Compute the communication cost and print the layout.

Step 8. Perform Steps 2 to 7 until a sufficient number of alternate layouts are available.

Step 9. Select the least cost layout.

4.2. Results

A program is developed for the Iterative Heuristic Algorithm for multilevel layout problem. The input consists of number of levels, adjacency matrix containing number of travel trips between clusters, area(in square units) of each cluster, and maximum area permitted per level. The program requires to determine pairs of objects for forming a cluster, check if a cluster has been considered for a particular level, and also allocate objects on different levels. An optimal layout design is thus obtained.

Table 3 shows the allocation of clusters on different levels and the total travel cost and Table 4 shows the allocation of 12 clusters/objects on three levels.

To test cluster analysis program, layouts were obtained. The program was developed for clustering technique to generate lists of clusters for locating on different levels. Layouts were obtained for three levels with maximum area permitted per level as 19 square units., and number of clusters as 12. Tables 3, and 4 show the grouping of objects and also the total cost of locating clusters on three levels.

Table 3- Representation of object location on three levels[1]

		6	17	2	2		
Level 3	15	15	17	2	1		
	15	16	16	1	1		
		16	16	16			
		19	4	13	13	7	
Level 2		19	3	13	13	7	7
		21	20	12	12	8	8
				18		8	
Level 1		14	14	10	10	9	9
		14	14	10	11	9	9
		14	14	11	11	9	9
				5			

Number of objects= 21
Level 1 cost= 292.35 units
Level 2 cost= 2737.88 units
Level 3 cost= 1381.96 units

Table 4- Allocation of clusters/objects on three levels
Number of clusters = 12
Maximum area permitted / level = 19 square units
Total Cost= 2135 units

Level	Clusters	Objects
Third	1 , 6, 9, 10	1,2,17 , 6, 15, 16
Second	4, 2, 7, 8, 11, 12	12,13, 3, 4, 20, 21, 7, 8, 18, 19
First	3 , 5	9, 10, 11, 14, 5

Note: □ represents the clusters and the corresponding objects.

Thus, objects 1, 2, and 17 form a cluster and are on the same level. Also, objects 12, and 13 are on level 2 and objects 3, 4, 20, and 21 are clustered and located on second level. On level one, objects 9, 10, 11, and 14 are located.

5. Conclusion

It is interesting to note that the algorithm is extremely efficient and easy to implement. It is suited for solving reasonably large problems. The algorithm gives reasonably good results at low computing cost. It is postulated that the deviation of these solutions from the exact optimum in large problems will be marginal.

The cost of the layout would depend on the initial adjacency matrix containing the number of trips between objects. Given the location matrix, that is, an area for locating the clusters, the cost of layout would depend on where the first cluster is located. We can develop the layout by locating the first cluster in the center of the location matrix or at the extreme top-left corner of the layout matrix. It is found that layout generated by locating the first cluster in the center of the floor gives lower cost as compared to the one generated by locating at the top-left corner of the location matrix.

References

[1] Bhowmik, R.(2004), "*Allocating Clusters using Hierarchical Clustering Technique*", in *Hawaii*

International Conference on Computer Sciences, *Hawaii.*

[2] Bhowmik, R. (2007), "An Approach to the Facility Layout Design Optimization ", paper submitted to INFORMS Journal of Computing.

[3] Cinar, U.(1975), "Facilities Planning: A System Analysis and Space Allocation Approach", in C. M. Eastman(Ed.), Spatial Synthesis in Computer-Aided Building Design, Applied Science Publishers, 19-40.

[4] Carter, D. J. and Whitehead, B(1975), "The use of Cluster Analysis in Multistorey Layout Planning", Building Science, 10, 287-296.

[5] Han, J., Kamber, and M.(1999), "Data mining Concepts and Techniques", Morgan Kaufmann Publishers, pp 335-356.

[6] Jagielski, R., and Gero, J. S.(1997), "*A Genetic Programming approach to the Space Layout Planning Problem*", in R. Junge(Ed.), *CAAD Futures 1997*, Kluwer Academic Publishers, pp 875-884.

[7] Kaufman, L., Rousseeuw, P. J.(2005), "Finding Groups in Data: An Introduction to Cluster Analysis", New York, John Wiley InterScience, pp 44-48.

[8] Liggett, R.S(1981), "The Quadratic Assignment Problem: An Experimental Evaluation of Solution Strategies", Management Science, 27, 442-458.

[9] Mashford, J., Drogemuller, R. and Stuckey P. J. (2004), *"Building Design Optimization using Constraint Logic Programming", Eighteenth W(C)LP Workshop on Constraint Logic Programming.*

[10] Michalek, J. M., Choudhary, R. and Papalambros, P.Y.(2002), "Architectural Layout Design Optimization", Engineering Optimization, 34(5), 461-484.

[11] Zhang, T., Ramakrishnan, R., and Livny, M. (1996), "*BIRCH: An efficient data clustering method for very large databases*", *1996 ACM-SIGMOD International Conference Management Data(SIGMOD'96), Montreal, Canada*, *pp 103-114.*