(IJAER) 2013, Vol. No. 5, Issue No. I, January **ISSN: 2231-5152**

ANALYSIS OF DIVIDE-AND CONQUER LOCAL SEARCH HEURISTICS WITH QUADRANT AND NEIGHBOR RESTRICTIONS

DAVENDRA KRISHAN

ABSTRACT

For DACQ, we restrict a point to the quadrant in which it is assigned initially. When a lattice point is divided into four new points, we consider only movements of currently assigned points to one of the four new points. Therefore, at each stage of the algorithm, only four lattice points are considered for local search.

For DACN, we restrict a point so that it can be assigned only to a neighbor of the lattice point to which it is currently assigned. A solution s' belongs to the neighborhood of a solution s if for any point i M with an assignment of k N in s, the assignment of i in s' is l, where l is any lattice point in N that is next to k, either horizontally, vertically, or diagonally (see Figure 4.3). Thus, at each step, a maximum of nine lattice points can be considered for local search. Note that i can stay where it is currently assigned.

Key words: *Quadrant, algorithm, neighborhood, diagonally.*

INTRODUCTION

Today's data sets are usually large and multidimensional, growing and changing with time; consequently, they are usually complex, dynamic, and difficult to visualize. Data visualization reveals the relationships and trends that are not evident from the raw multidimensional data sets by using mathematical techniques to reduce the number of dimensions while preserving the relevant inherent properties. Data visualization rests on the premise that a picture is worth a thousand words (Schiffman et al., 1981; Young, 1987). The practical value of data visualization is based on the fact that it is often easier and more informative to look at a picture of the data than to look at the data points themselves, particularly when the data set is large (Schiffman et al., 1981). Large and multidimensional data sets that require visualization are commonplace today and may be encountered in many disciplines ranging from the physical, biological, and behavioral sciences to product development, marketing, and advertising (Schiffman et al., 1981).

REVIEW OF LITERATURE

Popular techniques used to solve data visualization problems include multidimensional scaling (MDS) and Sammon maps (SM) (Borg and Groenen, 1997; Sammon, 1969; Schiffman et al., 1981; Young, 1987). These techniques solve the data visualization problem using nonlinear optimization techniques. A limitation of a nonlinear algorithm is the small number of vectors (data points) it can handle (Sammon, 1969). Even with today's fast computers, nonlinear optimization techniques are usually slow and inefficient for large data sets. Discrete optimization techniques may provide an efficient solution to the data visualization problem.

MATERIAL AND METHOD

Next, we suggest two refinements to DAC that reduce running time. We propose a divide-and-conquer local search heuristic with quadrant restrictions (DACQ) and a divide-and-conquer local search heuristic with neighbor restrictions (DACN). In these local search algorithms, there are neighborhood restrictions on the lattice points to which points can be assigned. Both algorithms follow the same steps used in DAC. However, the points in M are not assigned to all of the lattice points. The neighborhood is restricted as follows.

For DACQ, we restrict a point to the quadrant in which it is assigned initially. When a lattice point is divided into four new points, we consider only movements of currently assigned points to one of the four new points. Therefore, at each stage of the algorithm, only four lattice points are considered for local search.

For DACN, we restrict a point so that it can be assigned only to a neighbor of the lattice point to which it is currently assigned. A solution s' belongs to the neighborhood of a solution s if for any point i M with an assignment of k N in s, the assignment of i in s' is l, where l is any lattice point in N that is next to k, either horizontally, vertically, or diagonally (see Figure 4.3). Thus, at each step, a maximum of nine lattice points can be considered for local search. Note that i can stay where it is currently assigned.

We show the results for DACQ and DACN for the same problems used in the previous experiments. DACQ generates very poor results. It never finds the global optimum to any of the 50-point, 100-point, or 150-point problems. On the other hand, DACN produces very good results. In nine of 10 problems of size 50 and nine of the 10 problems of size 100, DACN finds the global optimum. DACN finds the global optimum in all 10 problems of size 150.

The average running times for DACQ are 1.22 seconds, 5.01 seconds, and 11.26 seconds for the 50-point problems, the 100-point problems, and the 150-point problems, respectively. For DACN, the average running times for the 50-point problems, the 100-point problems, and the 150-point problems are 4.07 seconds, 13.24 seconds, and 27.95 seconds, respectively. Both heuristics have much lower running times than DAC.

Considering both solution quality and running time, DACN appears to be the best heuristic, when compared to LS, DAC, and DACQ. It gives high-quality solutions in a reasonable amount of time. Increasing the size of n does not increase the size of the neighborhood, as in DAC, since a maximum of nine lattice points is considered at each stage of the algorithm. This results in a more gradual increase in the running time of DACN as a function of n.

RESULTS AND ANALYSIS FOR DACN

In this section, we apply DACN to several data sets with 50, 100, and 150 points with nonzero global optimal objective function values. The data sets were randomly generated from lattice sets in three, four, and five dimensions (e.g., for three dimensions, points were generated from a 16 × 16 ×16 lattice; for four dimensions, points were generated from a 16 × 16 ×16×16 lattice, and so on). For each combination of dimension and size, 10 different problems were generated. The nine problem sets (problem sets 1 to 9) are described in Table 1.

In all our experiments, we use q = 2. In cases where r = q, it is easy to compare the quality of the computational results. In this case, the optimal objective function value is known and equal to zero. However, for problems where q < r, the optimal value of the objective function is unknown and greater than zero. No local criterion exists for deciding how good a local optimal solution is as compared to a global one (Cela, 1998). Actually, from a complexity point of view, Cela (1998) states that even deciding whether a given local solution is a global optimal is an NP-hard problem.

Problem Set	Dimensions	Number of Points
1	3	50
2	3	100
3	3	150
4	4	50
5	4	100
6	4	150

7	5	50
8	5	100
9	5	150

Table.1 Characteristics of problem sets

In Table 1, we show the results for problem sets 1, 2, and 3. We do not know the global optimal value for these problems and so we cannot compare the results we obtained. We observed though that the frequencies for the best solutions are very low. The frequencies are all less than 10, expect for problem six of problem set 2, which has a frequency of 17. The average running times are 4.33 seconds, 15.73 seconds, and 36.65 seconds, for problem sets 1, 2, and 3, respectively. In Figure 4 we show a plot of the final result obtained by DACN for problem two of problem set 1. This plot and all other plots in this thesis are produced using Matlab 7.0 (Sigmon and Davis, 2002). In Figures 5 and 6, we show the plots for problem nine of problem set 2 and problem four of problem set 3, respectively.

10	147	133	145	76	97				
82	86	93	125	107	16	131	65	126	
121	17	143	74	127	89	91			
92	14	123	36	124	134	75			
135	84	113	29						
130	115	117	12	55	5	150	13	129	108
73	102	106	59	27	137	28	18	11	54
99	56	144	79	48	47	81			
80	53	22	136	43	23	148	67	61	
100	103	60	94	122	2	58	78		
26	25	51	45	139	24	141	96	38	142
62	112	7	104	77	35	37	57	9	
21	105	138	110	72	85	31	68		
90	114	66	140	33	116	63	146	109	
132	118	88	71	111	42				
40	119	20	95	149	30	83			

Figure 1 Plot for problem 5 from Problem Set 6.

	50-point problems			100-point problems			150-point problems		
	Running time						Running time		
	Best			Best			Best		
Problem	solution	Freq	(secs)	solution	Freq	(secs)	solution	Freq	(secs)
1	75247.8	1	4.57	325175	1	17.81	802908	1	38.22
2	64814.3	1	5.24	333587	1	18.68	841420	1	39.86
3	78566.4	1	4.45	348777	1	17.31	860219	1	41.67
4	75835.4	1	4.74	356028	1	17.25	812147	1	40.28
5	72817.3	1	4.39	329985	1	15.57	813606	2	40.68
6	81840.3	1	4.64	360491	1	18.31	856291	1	40.79
7	64342.7	1	4.46	297337	1	16.99	747779	1	37.02
8	71669.1	1	4.49	305608	1	17.46	753434	1	43.03
9	85219.9	1	4.76	368939	1	16.49	877125	1	39.62
10	78102.9	2	4.82	319882	1	18.24	784065	1	41.53
Average	4.66	17.41							

40.27

Table 2 Results for problem sets 7, 8, and 9 for DACN. These problem sets are originally in five dimensions.

35	131	36	49	96	44	111	134						
117	13	65	25	74	145								
72	24	90	22	42									
2	118	113	124	112	27	10	93	141					
31	147	30	106	143	85	99	116						
3	79	73	88	133	110	29	123	71					
127	69	58	101	84	136	107	76						
15	40	138	148	59	7	68	94	11					
75	33	144	4	135	14	86	77	5	100	9			
16	46	132	142	126	32	140	20	26	57				
149	39	87	61	47	45	64	17	115	28	48	56	83	70
89	43	41	53	108	146	109	125	38					
6	62	81	82	19	37	97	92	12	102				
60	150	129130	63	18	55								
105	104	80	50	122									
114	120 121	137	95	54	128								

Figure 2 Plot for problem 10 from Problem Set 9.

CONCLUSIONS

DACN provides an approximate solution to the data visualization problem in a small amount of computing time. For the problem sets originally in two-dimensions, DACN produces the global optimum in 28 of the 30 problems. For the other problem sets, the global optimal solutions are unknown. In the remaining chapters, we will use other algorithms on these problem sets and then compare these results to the solutions generated by DACN.

REFERENCES

E. Aarts and J. K. Lenstra (editors), Local Search in Combinatorial Optimization, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, West Sussex, England, 1997

W. P. Adams and T. Johnson. Improved linear programming based lower bounds for the quadratic assignment problem, in Quadratic Assignment and Related Problems,

P. Pardalos and H. Wolkowicz (editors), DIMACS Series in Discrete Mathematics and Theoretical Computer Science 16, 43 - 75, 1994.

R. K. Ahuja, J. B. Orlin, and A. Tiwari, A greedy genetic algorithm for the quadratic assignment problem, Computers & Operations Research 27, 917 - 934, 2000.

M. Berry and G. Linoff, Mastering Data Mining, John Wiley & Sons, New York, 2000.

M. Berry and G. Linoff, Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, John Wiley & Sons, New York, 2004.

M. S. Bazaraa and H. D. Sherali, Benders' partitioning scheme applied to a new formulation of the quadratic assignment problem, Naval Research Logistics Quarterly 27, 29 - 41,1980.

I. Borg and P. Groenen, Modern Multidimensional Scaling: Theory and Applications, Springer, New York, 1997.

R.E. Burkard and F. Rendl, A thermodynamically motivated simulation procedure for combinatorial optimization problems, European Journal of Operational Research 17, 169 - 174, 1984.

R.E. Burkard, Quadratic assignment problems, European Journal of Operational Research 15, 283 - 289, 1984.

P. Cabena, P. Hadjinian, R. Stadler, J. Verhees, and A. Zanasi, Discovering Data Mining: From Concept to Implementation, Prentice Hall, Upper Saddle River, New Jersey, 1997.

E. Cela, The Quadratic Assignment Problem: Theory and Algorithms, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

N. Christofides and E. Benavent, An exact algorithm for the quadratic assignment problem, Oper. Res. 37, 760 - 768, 1989.

E. Condon, B. Golden, S. Lele, S. Raghavan, and E. Wasil, "A visualization method based on adjacency data," Decision Support Systems 33, 349 - 362, 2002.

E. Condon, B. Golden, and E. Wasil, "Visualizing group decisions in the analytic hierarchy process," Computers & Operations Research 30, 1435 - 1445, 2003.

K.C. Cox, S. G. Eick, G. J. Wills, and R. J. Brachman, "Visual data mining: recognizing telephone calling fraud," Data Mining and Knowledge Discovery 1, 225 - 231, 1997.

L. Davis, Handbook of Genetic Algorithm, Van Nostrand Rienhold, New York, 1991.

G. Deboeck and T. Kohonen (editors), Visual Explorations in Finance, Springer, London, 1998.

U. Dorndorf and E. Pesch. Data Warehouses, in Handbook on Data Management in Information Systems, J. Blazewicz, W. Kubiak, T. Morzy, and M. Rusinkiewicz (editors),

International Handbooks on Information Systems, 387 - 430, Springer, New York, 2003.

J. Elder and D. Pregibon, A statistical perspective on knowledge discovery in databases, in Advances in Knowledge Discovery and Data Mining, U.M. Fayyad et al. (editors), 83 - 116, AAAI/MIT Press, California, 1996.

G. Finke, R.E. Burkard, and F. Rendl, Quadratic assignment problems, Annals of Discrete Mathematics 31, 61 - 82, 1987.

C. Fleurent and J.A. Ferland, Genetic hybrids for the quadratic assignment problem, in Quadratic Assignment and Related Problems, P. Pardalos and H. Wolkowicz (editors), DIMACS Series in Discrete Mathematics and Theoretical Computer Science 16, 173 - 187, 1994.

C. Fleurent and J.A. Ferland, Genetic and hybrid algorithms for graph coloring, Annals of Operations Research 63, 437 - 461, 1996.