Hybridization of Neighbourhood Search Metaheuristic with Data Mining Technique to Solve p-median Problem

D.Srinivas Reddy[1], A.Govardhan[2], Ssvn Sarma[3]

[1]Assoc.Professor, Dept of CSE, Vaageswari College of Engineering, Karimnagar,A.P,, India.
[2]Professor, Department of CSE, JNTUH College of Engineering, Hyderabsd, A.P., India.
[3]Professor, Department of CSE, Vaagdevi College of Engineering, Warangal, A.P., India.

Abstract:
Combinatorial optimization is the most panoptic area in current research paper. The p-median problem which is a combinatorial optimization problem is NP-Hard in nature that realizes facilitators which serves the maximum locations. The p-median problem will be practical in several applications areas such as escalating marketing strategies in the sphere of Management Sciences and in locating server positions in computer networks. In the proposed work the Metaheuristic based on Neighbourhood Search (NS) is hybridized with Data Mining Technique (HDMNS) with Frequent Mining to provide a solution to p-median problem. The resulting local optimal solution from NS method serves as a basis for identification of feasible solution space that holds different possible solutions of similar size and by the application of frequent mining technique on it results in identification of frequent items. Basing on the support count, most feasible solution is identified.

Keywords: Data Mining, Frequent item, GRASP, HDMNS procedure, Neighbourhood Search, NS Approach, USS

1. Introduction

The p-median problem can be represented with the mapping d: C x F → R, where F denotes the set of facilities, C represents the set of Customers and R the set of real numbers. The distance between the customer and the facility is represented by the mapping d, which is also termed as the distance function. The p-median problem ascertain a R facilities such that R⊆F and |R| = p, for any positive integer p and number of facilities n, where p ≤ n, such that the sum of the distances from each customer to its adjacent facility is minimized. Here every customer location is assumed as a facility i.e. F = C, and also for giving equal importance to each location it is considered that $w_i = 1$. The p-median problem can be represented mathematically as [24].

$$Minimize\ \ f(d,x)=\sum_{i=1}^{n}\sum_{j=1}^{n} w_i d_{ij} x_{ij} \qquad (1)$$

$$subject\ to \sum_{j=1}^{n} x_{ij} = 1 \quad \forall i \qquad (2)$$

$$x_{ij} \le y_j \ \ \forall i, j \qquad (3)$$

$$\sum_{j=1}^{n} y_j = p \qquad (4)$$

$$x_{ij} = 0\, or\, 1 \ \ \forall i, j \qquad (5)$$

$$y_i = 0\ or\, 1 \ \ \forall j \qquad (6)$$

Where,
n = number of locations
x_{ij} = 1 if a location i is assigned to facility located at j,
= 0 other wise
y_i = 1 if j th location is a facility
= 0 other wise
d_{ij} = distance measured from location i to location j j
p = preferred number of locations as facilities

The paper is structured as follows: In section 2 enlighten the existing GRASP and Neighbourhood Search based Metaheuristic approaches. Section 3 focuses on the proposed work, hybridization of Neighbourhood search method with data mining technique. Section 4 deals with experimental results and comparisons and Section 5 imparts the conclusions.

2. The Grasp And Neighbourhood Search Metaheuristics

Metaheuristics like genetic algorithms, GRASP, Neighbourhood Search Approach and others have been suggested and are functional to real-world problems in numerous areas of science [13] for p-median problem. GRASP's searching mechanism is iterative and each iteration consists of two phases: construction phase, to provide feasible solution and enhancement phase, to identify optimal solution [14][25][26]. The NS Approach mechanism is illustrated in Figure 1, Figure 2 and Figure 3.It is also having two phases- Then the construction phase and NB Search phase.

```
procedure NS Approach(list, p)

optml_sol ← ∅
sol ← Construction(data points);
best_sol ← NBSearch(sol);
if cost(sol) > cost(optml_sol)
 optml_sol ← sol;
end if
until Termination criterion;
return optml_sol;
```

Figure 1. NS Approach procedure

```
procedure NBSearch(sol)

for each s in sol
  Compute neighbourhood of s;
   For each point x in neighbourhood
    new_sol ← swap(s,x);
     improved sol ← Enhancement(sol);
compare cost of new_sol and
        improvedsol and update best_sol
end for
end for
return best_sol
```

Figure 2: NB Search algorithm

The updation of the solution space is described in Figure 5. Basing on the updated solution space, the Frequent item set (FIS) are generated which consists the set of all distinct items that are present in the updated solution space. It is described in Figure 6. Support count is calculated and updated for each item in the FIS which is elucidated in Figure 7. After updating support count, sort the frequent items in the decreasing order of support count and then the mined solution is constructed by deliberating the items with high support count until the size of the solution or the number of items exactly equals to p. It is exemplified in Figure 8. The final phase is used to derive the optimal solution. The optimal solution is updated using Enhancement phase described in NS Approach which examines the global optimal solution that optimizes the objective function of the given p-median problem.

3. Proposed Work

The hybridized Data Mining Neighbourhood Search Approach (HDMNS()) procedure given in Figure 4 consists of three phases. The first phase NS Approach(), computes the initial solution using Neighbourhood Search approach by considering the given list and user specified p. It is described in detail in Figure 1, Figure 2 and Figure 3. The second phase is the core of the proposed work .i.e. application of data mining technique to the basic feasible solution. The obtained solution in first phase is input for the second phase which is used to generate the solution space (USS) that consists of all the probable solutions generated using the result obtained in the NSApproach().

```
procedure Enhancement( sol )

imp_sol ← sol;
imp_cost ← cost_eval( sol);
repeat
no_improvements ← true;
for i = 1 to p
temp_best_sol ←∅;
temp_best_cost ← ∞;
for each element e in Pi close to ri
t_sol ← swap(best_sol, ri, e);
t_cost ← appcosteval(t_sol);
if approxcost < approxbestcost then
          temp_best_sol ← t_sol;
          temp_best_cost ← t_cost;
end if
end for
exactsolcost ← costeval(appbestsol);
if exactsolcost < bestcost then
          imp_sol ← appbestsol;
          imp_cost ← exactsolcost;
noimprovements ← false;
end if
end for;
until noimprovements;
return imp_sol;
```

Figure 3. Enhancement phase NB Search

```
procedure HDMNS()
Initialize sol, optimal_sol, Mined_sol ←Φ
Initialize sol_space←Φ
Initialize USS, FIS, UFIS←Φ
Read list, p
Sol←NSApproach(list, p)
USS←Update_sol_space(sol)
FIS←Generate_frequent_items(USS)
UFIS←Update_supportcount(FIS, USS)
Mined _ sol←Generate _ mined
_solution(UFIS)
Update optimal_sol
```

Figure 4. Hybridized data mining NS

```
procedure Generate_Frequent_items(USS)

    1.  Initialize FIS←Φ
    2.  Read all solutions in USS
    3.  For each solution in USS
        Update FIS with distinct elements
    4.  End for
    5.  Return FIS
```

Figure 6. Generate frequent items algorithm

```
procedure Update_sol_space(sol)

    1.  Initialize USS←Φ
    2.  Mod_list← list – sol
    3.  For each element in sol
        Exchange element with element in
        mod_list to get new solution
        Update USS
        End for
    4.  Return USS
```

Figure 5. Update solution space algorithm

```
procedure Generate_Mined_solution(UFIS)

    1.  Initialize Mined_sol←Φ
    2.  Sort UFIS in decreasing order of support count
    3.  Update Mined_sol with elements in UFIS with
        high support count
        Until Mined_sol size == p;
    4.  Returm Mined_sol
```

Figure 8. Generate mined solution algorithm

```
procedure Update_Support_count(FIS, USS)

    1.  Initialize UFIS←Φ
    2.  Initialize each element support
        count to 0
    3.  For each element(x) in FIS
        For each solution in USS
        Update support count of x
        End for
    4.  End for
    5.  Return UFIS
```

Figure 7. Update support count

4. Experimental Results

The experimental results obtained for GRASP, NS approach and Hybrid Data Mining Neighbourhood Search(HDMNS) are analyzed in this section, and the results are evaluated on the basis of quality of the solution against p. Experiments are carried out on data sets with 15, 25, 50 points. Results are tabularized and graphs are outlined. The origin of data sets under study is acquired from the web site of Professor Eric Taillarrd, Kent University of Applied Sciences of Western Switzerland. The associated website for p-median problem instances is http://mistic.heig-vd.ch/taillard/problemes.dir/location.html. In Figure 9 the optimal Objective function value i.e. cost for p-median problem is compared using algorithms GRASP, NS approach and HDMNS for the data set of size 50 with number of facility locations (p) incremented by 10. It is observed that HDMNS is working better than the other two techniques.

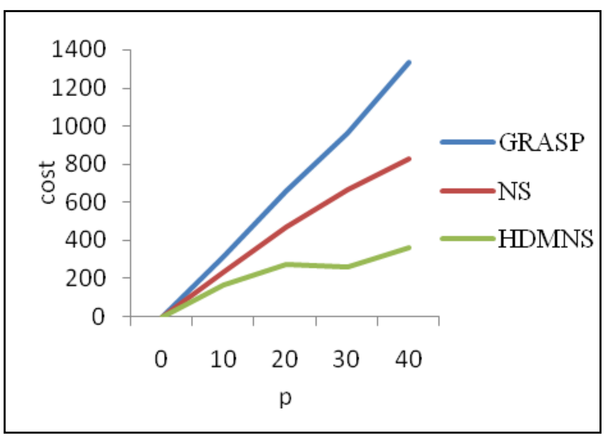


Figure 9. Cost comparison of GRASP, approach & HDMNS when N=50

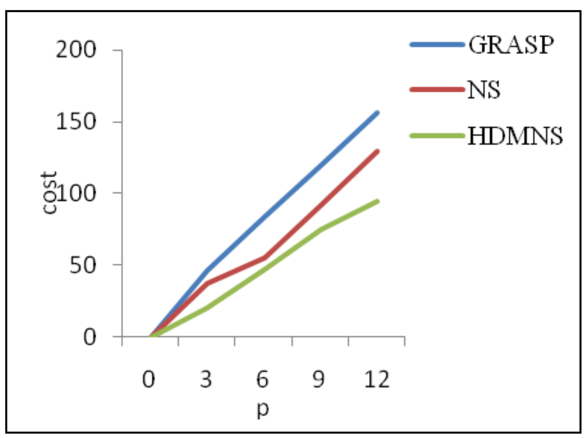
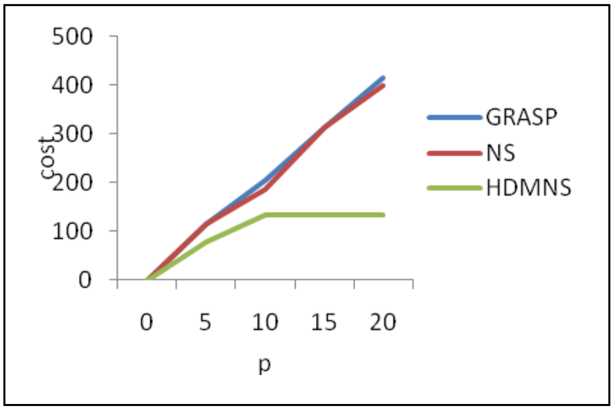


Figure 10. Cost comparison of GRASP, NS NS approach & HDMNS when N=15

In Figure 10 the optimal Objective function value i.e. cost for p-median problem is evaluated using algorithms GRASP, NS approach and HDMNS for the data set of size 15 with number of facility locations (p) raised by 3. It is perceived that HDMNS outperforms the other two techniques. Similarly Figure 11 is plotted with data set size 25 with p=5 and observed the same.



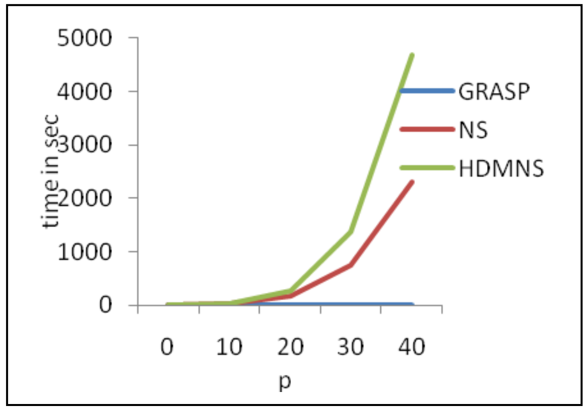
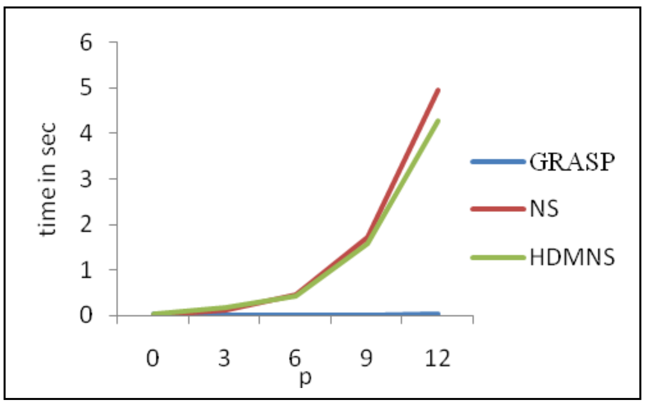


Figure 11. Cost comparison of GRASP, NS approach & HDMNS when N=25

Figure 12. Execution time comparison of GRASP, NS approach & HDMNS when N=50



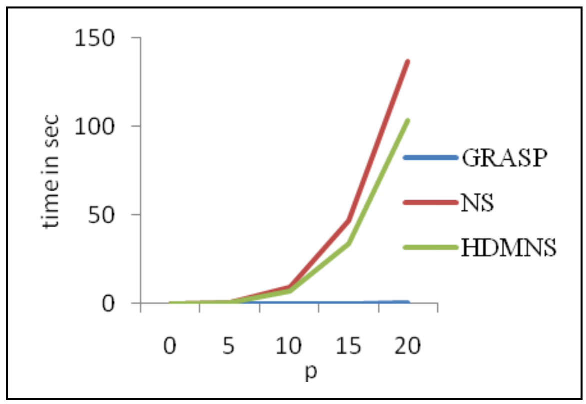


Figure 13. Execution time comparison of GRASP, NS approach & HDMNS when N=15

Figure 14. Execution time comparison Vs GRASP, NS approach & HDMNS, N=5

Execution times are compared for all the three algorithms for N=50, 15 and 25 with p increments 10, 3 and 25 respectively and identified that HDMNS works better than other two. They are represented in Figure 12, Figure 13 and Figure 14.

5. Conclusions

It is observed that in all the test cases, Hybrid Data Mining Neighbourhood Search (HDMNS) Metaheuristic performs much better when compared with the efficient existing methods like GRASP and Neighbourhood Search Metaheuristic. It is also observed that in most of the cases the though the HDMNS includes Mining technique in addition with NS approach it takes almost same execution time as NS approach and produces much more better results than NS approach.

References

[1] R. Agrwal and R. Srikanth. Fast algorithms for mining association rules, Proc. of the Very Large Data Bases Conference, pages 487-499, 1994.

[2] T. A. Feo and M. G. C. Resende. A probabilistic heuristic for a computationally difficult set covering problem, Operational Reseach Letters, 8:67-71, 1989.

[3] T. A. Feo and M. G. C. Resende. Greedy randomized adaptive search procedures, Journal of Global Optimization, 6:1609-1624, 1995 .

[4] T. A. Feo and M. G. C. Resende. GRASP: An annotated bibliography, Essays and Surveys in Metaheuristics, Kluwer Acadamic Publishers, 2002.

[5] M. D. H. Gamal and Salhi. A cellular heuristic for the multisource Weber Problem, computers & Operations Research, 30:1609-1624, 2003.

[6] B. Geothals and M. J. Zaki. Advances in Frequent Itemset Mining Implementations: Introduction to FIMI03, Proc. of the IEEEICDM workshop on Frequent Itemset Mining Implementations, 2003.

[7] G. Grahne and J. Zhu. Efficiently using prefix-trees in mining frequent item-sets, Proc. of the IEEEICDM Workshop on Frequent Itemset Mining Implementations, 2003.

[8] J. Han, J. Pei and Y. Yin. Mining frequent patterns without candidate generation, Proc. of the ACMSIGMOD In. Conference on Management of Data, pages 1-12, 2000.

[9] J. Han and M. Kamber. Data Mining: Concepts and Techniques, 2^{nd} ed., Morgan Kaufman Publishers, 2006.

[10] O. Kariv and L. Hakimi. An algorithmic approach to network location problems, part II: the p-medians, SIAM Journal of Applied Mathematics, 37:539-560, 1979.

[11] N. Mladenovic, J. Brimberg, P. Hansen and Jose A. Moreno-Perez. The p-median problem: A survey of metaheuristic approaches, European Journal of Operational Research, 179:927-939, 2007..

[12] S. Orlando, P. Palmerimi and R. Perego. Adaptive and resource- aware mining of frequent sets, Proc. of the IEEE In. conference on Data Mining, pages 338-345, 2002.

[13] I. Osman and G. Laporte, Metaheuristics: A bibliography, Annals of Operations Research, 63:513-623, 1996.

[14] M. G. C. Resende and C. C. Ribeiro. Greedy randomized adaptive search procedures, Handbook of Metaheuristics, Kulwer Academic Publishers, 2003.

[15] M. H. F. Ribeiro, V. F. Trindade , A. Plastino and S. L. Martins. Hybridization of GRASP metaheuristic with data mining techniques, Proc. of the ECAI Workshop on Hybrid Metaheuristics, pages 69-78, 2004.

[16] M. H. F. Ribeiro, V. F. Trindade, A. Plastino and S. L. Martins. Hybridization of GRASP Metaheuristic with data mining techniques, Journal of Mathematical Modeling and Algorithms, 5:23-41, 2006.

[17] S. Salhi. Heuristic Search: The Science of Tomorrow, OR48 Keynote Papers, Operational Research Society, pages 38-58, 2006.

[18] L. F. Santos, M. H.F. Ribeiro, A. Plastinoand S. L. Martins. A hybrid GRASP with data mining for the maximum diversity problem, proc. of the Int. Workshop on Hybrid Metaheuristics, LNCS 3636, pages. 116-127, 2005.

[19] L.F. Santos, C.V.Albuquraque, S. L. Martins and A. Plastino. A hybrid GRASP with data mining for efficient server replication for reliable multicast, Proc. of the IEEE GLOBECOM Conference, 2006.

[20] L. F. Santos, S. L. Martins and A. Plastino. Applications of the DM-GRASP heuristic: A survey, In. Transanctions in Operational Research, 15:387-416, 2008.

[21] E. G. Talbi. A taxonomy of hybrid metaheuristics, Journal of Heuristics, 8:541-564, 2002.

[22] B. C. Tansel, R. L. Fransis, and T. J. Lowe. Location on networks: A survey, Management Science, 29:482-511, 1983.

[23] I. H.Witten and E.frank. DataMining: Practical Machine Learning Tools and Techniques.2^{nd} ed., Morgan Kaufmann Publishers, 2005.

[24] Moh'd Belal Al-Zoubi, Ahmed Sharieh, Nedal Al-Hanbali and Ali Al-Dahoud. A Hybrid Heuristic Algorithm for Solving the P-Median Problem. Journal of Computer Science (Special Issue) pages 80-83, Science Publications, 2005.

[25] Alexndre Plastino, Eric R Fonseca, Richard Fuchshuber, Simone de L Martins, Alex.A.Freitas, Martino Luis and Said Salhi. A Hybrid Datamining Metaheuristic for the p-median problem. Proc. of SIAM Journal (Data Mining), 2009.

[26] Sunil Nadella, Kiranmai M V S V, Dr Narsimha Gugulotu. A Hybrid K-Mean-Grasp For Partition Based Clustering of Two-Dimensional Data Space As An Application of P-Median Problem**,** In. Journal of Computer and Electronics Research 1:2278-5795, June 2012.