Clustering Strategy to Euclidean TSP

Hamilton Path Role in Tour Construction

Abdulah Fajar, Nur Azman Abu, Nanna Suryana Herman
Faculty of Information and Communication Technology
Universiti Teknikal Malaysia Melaka
abd.fajar@gmail.com, {nura,nsuryana}@utem.edu.my

***Abstract* — TSP arises in many applications such as transportation, manufacturing and various logistics and scheduling. This problem has caught much attention of mathematicians and computer scientists. A clustering strategy will decompose TSP into subgraphs and form clusters, so it may reduce problem size into smaller problem. The primary objective of this research is to produce a better clustering strategy that fit into Euclidean TSP. Hamilton path play important role to construct Euclidean TSP tour in this approach. Hamilton will be applied within clusters or inter clusters. Hamilton path construction will be proceed after clustering process, followed by producing inter cluster connection algorithm to find global tour. This approach is capable of producing fast solution result with error less than 10% compare to best known solution in TSPLib. This paper proposes an improvement to a hierarchical clustering algorithm in searching for Euclidean TSP solution.**

Keywords: Hierarchical Clustering Strategy, Hamilton Path, Euclidean TSP, Adjacency List

I. INTRODUCTION

Human beings possess the natural ability of clustering objects. This capability to organize the data a person perceived everyday form them into clusters, so that he/she may quickly draw important conclusions. However making a computer to do clustering is a quite difficult task and demanding the attentions of computer scientist and engineers all over the world till now [1].

The task of computerized data clustering problem has been approached from diverse domains of knowledge like graph theory, statistics, artificial neural network and so on. The most popular approach in this direction has been the formulations of clustering as an optimization problem. There has been growing interest in studying combinatorial optimization problems by clustering strategy, with a special emphasis on the traveling salesman problem (TSP) [2][3][4]. TSP naturally arises as a sub problem in much transportation, manufacturing and logistics application, this problem has caught much attention of mathematicians and computer scientists.

The rest of this paper will describe about motivation of this study regarding Euclidean TSP in section II. The objectives of this study will be described in section III. Literature background hierarchical clustering will be given in section IV and section V describes the approach done in this study. Hamilton path construction will be further elaborated in section VI. Preliminary result is presented in section VII. Section VIII concludes the paper.

II. MOTIVATION

Traveling salesman problem (TSP) is a combinatorial optimization task of finding the shortest tour of n cities given the intercity costs. In a more formal way the goal is to find the least weight Hamiltonian cycle in a complete graph G. The TSP problem is a '*NP-hard*' problem. An *NP-hard* problem is extremely unlikely to have a polynomial algorithm to solve it optimally. Although it is not proven, the worst case exponential running-time solution is some cases unavoidable. Good approximation algorithms can produce solutions that are only a few percent longer than an optimal solution and the time of solving the problem is a low-order polynomial function of the number of cities n. Some approximating algorithms produce tours whose lengths are close to that of the shortest tour, but the time complexity is substantially higher than linear running-time.

III. OBJECTIVE OF STUDY

A clustering process is expected to discover the natural grouping that exists in a set of pattern. Each cluster must be identified which has an Hamilton path that acts as both local and global optimal solution for Euclidean TSP. Finding process for a Hamilton path must be handle with care so that the path is not trapped into local optimal only. After an Hamilton path is found for each small cluster, the next step is to determine a tour by combining all Hamilton paths to form Euclidean TSP solution. There are several critical factors in solving Euclidean TSP have been identified. There should not be edge crossing in the solution. An algorithm to delete edge crossing is needed. Also hierarchical clustering algorithm that is suitable for this Euclidean TSP must be developed besides optimal Hamilton path algorithm both locally and globally for each cluster. Lastly, global tour construction as an approximation solution shall be compared to best known solution

The primary objective of this research is to produce a better hierarchical clustering strategy that is suitable for Euclidean TSP. In this study, hierarchical clustering is adopted to solve Euclidean TSP by producing an algorithm with the running time near $O(n^2)$. This approach is expected to performed well at most near $O(n^3)$. On average, final TSP solution costs not more than 10% compare to the best known solution. The other objectives are: produce non edge crossing

978-0-7695-3941-6/10 $26.00 © 2010 IEEE
DOI 10.1109/ICCMS.2010.476

algorithm in planar graph; produce an algorithm to generate clusters and able to handle large size cluster; produce an algorithm to determine incoming and outgoing vertices to form Hamilton path; produce Hamilton path algorithm by involved incoming and outgoing vertices as starting and ending points of Hamilton path and produce inter cluster connecting edge algorithm to find global tour.

IV. LITERATURE BACKGROUND

A clustering algorithm is expected to discover the natural grouping that exists in a set of pattern. In hierarchical clustering, data are not partitioned into a particular cluster in a single step. Hierarchical clustering may be represented by a two dimensional diagram known as a dendrogram which illustrates the fusions or divisions made at each successive stage of analysis.

Among the most popular hierarchical clustering algorithms, BIRCH[6] can typically find a good clustering formation in a single scan of the data and improve the quality further in a few additional scans. It is also the first clustering algorithm that handles noise effectively. CURE[7] represents each cluster by a certain number of points selected from a well-scattered sample and then shrinking them toward the cluster centroid by a specified fraction. It uses a combination of random sampling and partition clustering to handle large databases. ROCK[8] is a robust clustering algorithm for boolean and categorical data. It introduces two new concepts: a point's neighbors and links, and to measure the similarity/proximity between a pair of data points.

Haxhimusa *et. al* formulated a pyramid algorithm[1] for E-TSP motivated by the failure to identify an existing algorithm that could provide a good fit to the subjects data. More recently, hierarchical (pyramid) algorithms have been used to model mental mechanisms involved in other types of visual problems. The main aspects of the models are (multi resolution) pyramid architecture and a coarse to fine process of successive tour approximations. This algorithm was motivated by the properties of the human visual system has main idea to use agglomerative (bottom up) processes to reduce the size of the input by clustering the nodes into small group and top-down refinement to find an approximate solution, as same as Graham do in [3]. The size of the input (number of vertices in the subgraph) is reduced so that an optimal solution can be found by the combinatorial search. A pyramid is used to reduce the size of the input in the bottom-up processes.

In bottom-up clustering, cities as nodes are close neighbors are put into the same cluster using greedy approach. These clustered cities are considered as a single city at the reduced resolution. Graph pyramid strategy use minimum spanning tree(MST) using Bouruvka algorithm. MST is used as natural lower bound for TSP solution. In the case TSP with triangle inequality which in the case for the Euclidean TSP, MST can be used as to prove the upper bound.

Graph pyramid solution gives a good performance compared to TSPLIB. Solution error for this solution tends to give linear costs around 12%. Running time performance is quadratic for this solution especially while number of cities n more than 600.

V. METHODOLOGY

There are several algorithm to be produced in this study such as: non edge crossing algorithm in planar graph; an algorithm to generate clusters which is able to handle large size cluster; produce an algorithm to determine incoming and outgoing vertices to form Hamilton path; produce Hamilton path algorithm by involved incoming and outgoing vertices as starting and ending nodes of Hamilton path and produce inter cluster connection algorithm to find global tour.

Producing those algorithms above will be make use two basic approaches. First is general methodology for TSP solution and second is more on specific methodology for clustering. General steps for TSP solution as drawn in Fig. 1 begins with on planar graph construction. Planar graph is then filtered by its ascending adjacency list to form cluster using threshold R, where

$$R = \sqrt{\frac{\text{Area of Cluster } A}{\text{The number of Vertices } n}} \quad (1)$$

Every cluster will be evaluated if there is any clusters overlap or large size. If any, then respective cluster must be split to avoid overlap and making it smaller. Inter cluster connection is then performed to define which vertices will act as incoming or outgoing vertices for each clusters. Convex hull and cheapest insertion are applied to inter cluster connection using cluster center coordinate at every clusters. This step result an initial tour that will be used to construct Hamilton path within clusters.

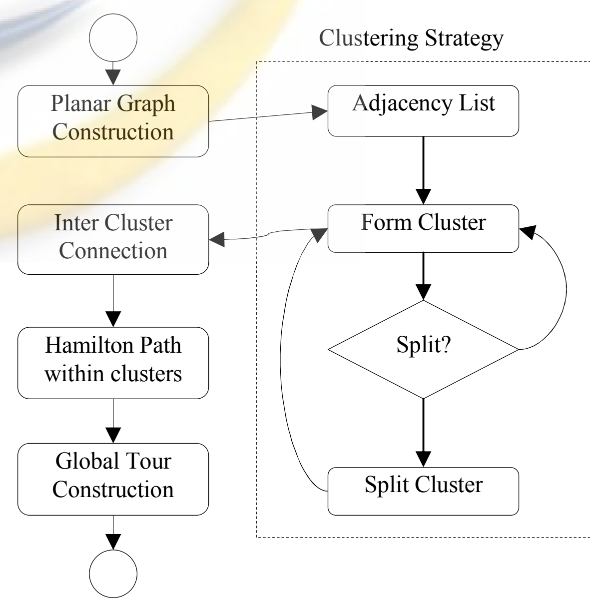

Figure 1. TSP Solution Methodology

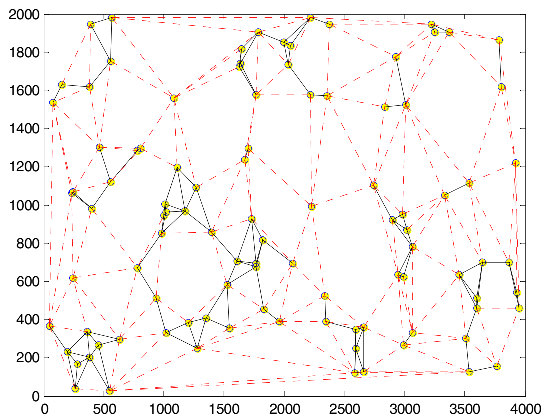

Figure 2. Adjacency list after being filtered and become clusters

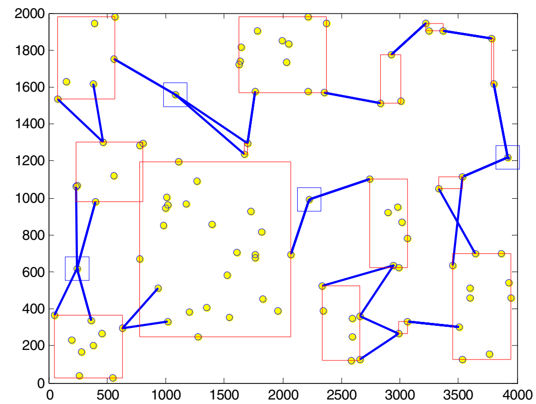

Figure 3. Inter Clusters Connection

The basic idea to construct Hamilton path within each cluster comes in handy with convex hull and cheapest insertion method. A good Hamilton path will be achieved from each cluster if an outgoing node of prior cluster and an incoming node of next cluster are chosen as listed in initial global tour. Convex hull will initialize Hamilton cycle among clusters and then break at outgoing and incoming nodes. The nodes which exclude when broken cycle is will reinserted using cheapest insertion.

Final step in this method will perform global tour construction. Every Hamilton path within cluster will be integrated to form global tour according to initial global tour of inter-cluster connection. TSP solution approach usually consist two main stages, namely, an initial tour construction and tour improvement. Integration Hamilton path is a final step in tour construction, after which tour shall be improved in order to get better result.

VI. HAMILTON PATH CONSTRUCTION

Hamilton path is subset of TSP tour, and play important role to construct global solution. This approach has determined that each cluster has two end points (incoming and outgoing nodes). These end points will be an initial path to construct Hamilton path. The initial solution coming from convex hull and cheapest insertion (CHCI) methods will be improved using k-opt exchange move.

Let kroD100 be a complete graph $G(V, E, W)$ that consists of V as a set of vertices, E as a set of edges and W as the weight of edges. After clustering process, this sample problem is decomposed to sub graph/clusters such as Fig. 2 above.

Next step is to construct Hamilton path within clusters. Also there is a need to determine the incoming and outgoing vertices for each cluster. First, apply convex hull over center of cluster and insert the rest of cluster's center which is not included in convex hull using cheapest insertion to form initial global tour. The sequence of inter cluster connection will be used as guidance to determine incoming and outgoing vertices for each cluster.

.

Between two adjacent clusters subsequently, there will be incoming and outgoing vertices of each cluster to be connected to form inter clusters connection, as drawn on Fig. 3. Also those vertices will be marked as temporary nodes to find Hamilton path within clusters.

Let $C=\{c_1, c_2, \ldots, c_k\}$ be the set of clusters in G, and number of clusters in G is 15, or k=15. As shown in Fig. 3, each cluster may have more than a pair of end points. A sample result of Hamilton path for every cluster has been shown below in Fig. 4. The filtered adjacency list of the set of edges E provides a sketch of Hamilton path within a cluster. Since the sizes or the number of nodes for each cluster is small, there is a good chance to capture Hamilton path efficiently.

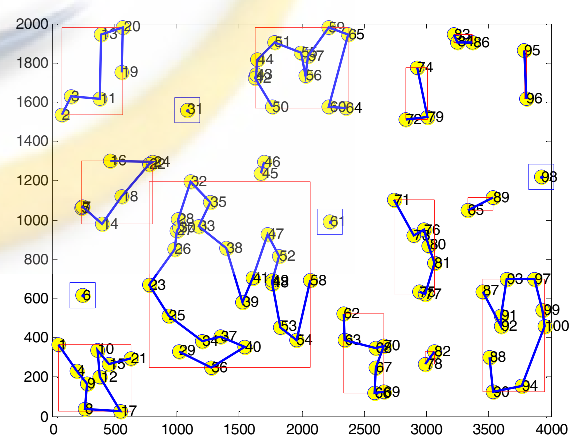

Figure 4. Hamilton paths within clusters

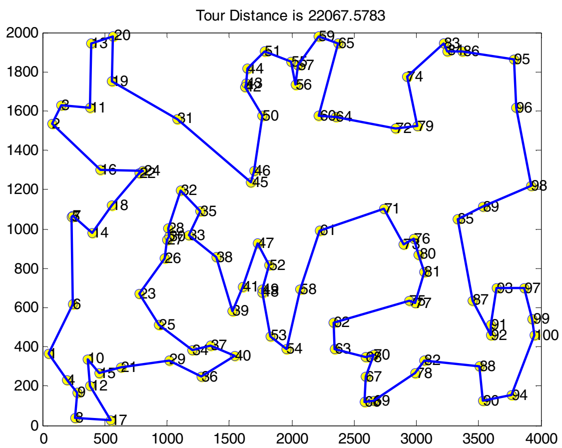

Figure 5. TSP Global solution for rd100 sample problem

The final step to form global tour is to integrate Hamilton paths within clusters by combining them via inter clusters connection. TSP solution approach usually consist two main stages there are tour construction and tour improvement. Integration of all Hamilton paths is a final step in tour construction stages, after that tour must be improved in order to get good result. This approach use 2-opt and 3-opt exchange move to improve the tour. The result is shown in Fig. 5 above.

VII. Discussion On Experimental Results

Compared to other methods[1][3], this approach gives promising result. MST pyramid method by Haxhimusa [1] and hierarchical model by Graham[3] has error up to 7% for number nodes up to 50. The error here refers to the costs of the ETSP tour higher than well known result in TSPLib. This method, however, produces error 3% maximum as shown in Table 1. Both methods carry k multi layer hierarchical approaches. Layer k has n/b_k nodes, n is number of nodes and b is reduction ratio. Increasing k potentially can lead a solution bias while this strategy strictly chooses which cluster must be split.

Table 1 below giving preliminary result, this study has run several experiments for ETSP problems from TSPLib with number of nodes up to 442. This simulation test with minimum tour improvement provides the costs of ETSP tour 6% above TSPLib optimal solution on average. Larger number of nodes n is not significantly influence the result, but above 200 nodes, result tend to increase the gap between this study compare to TSPLib. There is a need to look into the number of nodes allowed within each cluster.

Best solution is given by problem sample w29 and pr144, while worst solution shown by a280 and eil101. Both worst solutions may be influenced by their equally distance node distribution as shown in Fig. 6. This sample gives homogenous distribution and has the difficulty to form small clusters. The equally distance nodes which form a large cluster should be split into smaller clusters according to region. This shall be a part of future works.

TABLE I. Preliminary Experiment Result

Case	Size	TSPLIB	Cluster	%Error
w29	29	27,603	27,603	0.00%
att48	48	33,522	34,423	2.69%
eil51	51	426	447	4.93%
berlin52	52	7,542	8,206	8.80%
st70	70	675	746	10.52%
pr76	76	108,159	114,018	5.42%
eil76	76	538	591	9.85%
kroC100	100	20,749	21,180	2.08%
rd100	100	7,910	8,119	2.64%
kroD100	100	21,294	22,607	6.17%
kroE100	100	22,068	23,117	4.75%
eil101	101	629	694	10.33%
lin105	105	14,379	14,638	1.80%
pr124	124	59,030	60,086	1.79%
bier127	127	118,282	122,349	3.44%
pr136	136	96,772	102,943	6.38%
pr144	144	58,537	59,276	1.26%
kroB150	150	26,130	26,579	1.72%
pr152	152	73,682	77,107	4.65%
u159	159	42,080	45,581	8.32%
rat195	195	2,323	2,475	6.54%
d198	198	15,780	16,865	6.88%
kroB200	200	29,437	31,560	7.21%
pr226	226	80,369	89,264	11.07%
a280	280	2,579	2,914	12.99%
pr299	299	48,191	52,889	9.75%
lin318	318	42,029	44,272	5.34%
pcb442	442	50,778	56,232	10.74%

Running time result of this experiment is also promising compared to TSPLib. As shown in Table 2, below regarding Run time preliminary result, this study has more faster technique on average. Some sample result from TSPLib using Concorde did not show run time status.

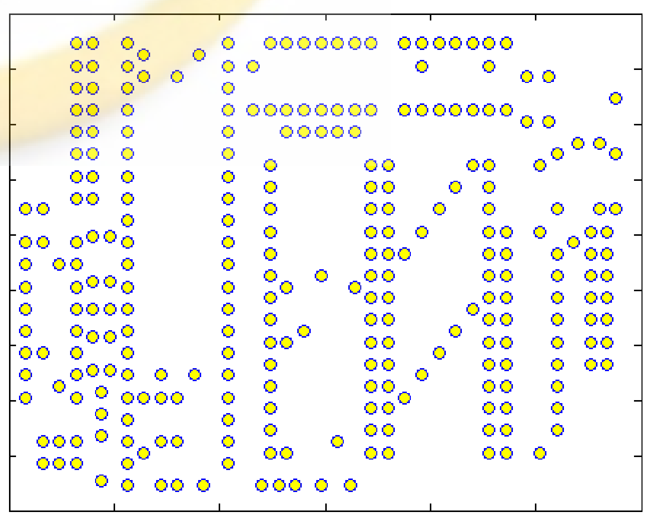

Figure 6. A280.tsp visually show the difficulty to form cluster

TABLE II. PRELLIMINARY RUN TIME RESULT (SECONDS)

Case	Size	TSPLIB	Cluster	%Time
w29	29	0.08	0.10	25%
att48	48	3.69	0.12	-97%
eil51	51	2.12	0.13	-94%
berlin52	52	0.06	0.14	133%
st70	70	3.24	0.21	-94%
pr76	76	1.96	0.25	-87%
eil76	76	2.66	0.21	-92%
kroC100	100	3.81	0.45	-88%
rd100	100	NA	0.38	NA
kroD100	100	3.35	0.42	-87%
kroE100	100	6.18	0.43	-93%
eil101	101	4.37	0.38	-91%
lin105	105	3.07	0.45	-85%
pr124	124	13.47	0.65	-95%
bier127	127	3.64	0.84	-77%
pr136	136	4.84	0.75	-85%
pr144	144	7.72	0.74	-90%
kroB150	150	NA	1.20	NA
pr152	152	5.10	1.31	-74%
u159	159	3.30	1.34	-59%
rat195	195	NA	2.70	NA
d198	198	NA	2.74	NA
kroB200	200	5.49	2.63	-52%
pr226	226	5.00	3.32	-34%

The results show that on average this technique is 50% faster for sample up to 226. This result must be re-evaluated against larger sample size problem.

This initial result has several finding to be followed through in future work, such as how to deal with larger size cluster, homogenous node distribution, or shape of clusters.

Fast algorithm which makes use of clustering technique is capable of producing near optimal solution. The number of nodes within each cluster should be limited. Smaller cluster size tends to provide near optimal Hamilton path closer within each cluster.

Several findings in preliminary result such as cluster shape overlapped or area of clusters shape near to area of problem should handle separately. In the case of large clusters, it has to be divided into further sub clusters. Prior explanation regarding this finding further will consider as future works.

VIII. CONCLUSION AND FUTURE WORKS

The initial result in this research shows a potential to produce a better clustering strategy that fit into Euclidean TSP. The current research in clustering to solve Euclidean TSP mostly producing their best algorithm with the running time near $O(n^2)$ while the initial run time result of this approach still have not been measured fully. However, this approach on average gives promising result with error less than 10% compare to the best known solution. Further research should test on larger size problem in the future works.

The other objectives of this research must be improved to support the primary objectives. Hamilton path play important role to construct global tour, it is important to take special attention to its algorithm.

Finally, this initial result open the door for the future works must be taken. Robust clustering techniques are needed to deal with large clusters or overlapped clusters. Hamilton paths both within clusters and inter clusters are already difficult problems. They are indeed crucial in maneuvering global tour construction.

ACKNOWLEDGMENT

This research funded by e-Science fund from Ministry of Science Technology and Innovation, Malaysia.

REFERENCES

[1] Y Haxhimusa , W. G Krospatch, Z Pizlo, A. Ion, "Approximate Graph Pyramid Solution of the E-TSP", *Image and Vision Computing Journal*, Elsevier, 2008,

[2] N.A Abu, S. Sahib, N. Suryana, "A Novel Natural Approach to Euclidean TSP", in *The 3rd International Conference on Mathematics & Statistics (ICoMS)*, Bogor, Indonesia 2008

[3] S.M. Graham, A. Joshi, Z. Pizlo, "The Traveling Salesman Problem: a Hierarchical Model", *Journal of Memory and Cognition,* vol 28 pp 1191–1204, 2000

[4] D. Vickers, M. Butavicius, M. Lee, A. "Medvedev, Human Performance on Visually Presented Traveling Salesman problems", *Journal of Psychological Research,* vol 65 pp 34–45, 2001

[5] A.P Punnen, The Traveling Salesman Problem: "Application, Formulations and Variations", in *The Traveling Salesman Problem and Its Variations* vol.12. G. Putin and A.P Punnen, Springer, 2007 pp 1- 28.

[6] T. Zhang, R. Ramakrishnan, and M. Livny. "BIRCH: an efficient data clustering method for very large databases". In *ACM SIGMOD Int. Conf. on Management of Data*, Montreal, Canada,(1996.

[7] S. Guha, R. Rastogi, and K. Shim. "CURE: An Efficient Clustering Algorithm for Large Databases". In *ACM SIGMOD Int. Conf. on Management of Data*, Seatle, WA, 1998.

[8] S. Guha , "ROCK: A Robust Clustering Algorithm for Categorical Attribute" s, *Journal of Information System,* vol. 25,pp:345–366, 2000

[9] R. Xu,"Survey of Clustering Algorithm", *IEEE Transaction On Neural Networks*, Vol 16,.645-678, 2005