Annales Mathematicae et Informaticae
48 (2018) pp. 33–42
http://ami.uni-eszterhazy.hu

Efficiency Analysis of the Vertex Clustering in Solving the Traveling Salesman Problem

László Kovács, Anita Agárdi, Bálint Debreceni

University of Miskolc, Institute of Information Sciences
kovacs@iit.uni-miskolc.hu

Submitted March 5, 2018 — Accepted September 13, 2018

Abstract

The TSP is the problem to find the shortest path in a graph visiting every nodes exactly once and returning to the start node. Due to the high complexity of TSP, there exists no algorithm for global exact optimization with polynomial cost. In order to provide an acceptable solution for real life problems, the TSP are usually solved with some heuristic optimization problem. The paper proposes a multi layered optimization model, where the node set is partitioned into clusters or into hierarchy of clusters. Based on the test experiments the proposed method is superior to the single level optimization method for both the TSP and MTSP problems.

Keywords: Traveling Salesman Problem, Clustering

MSC: 90-08

1. Introduction

The Traveling Salesman Problem (TSP) is one of the most intensively investigated optimization problem in graph theory. The TSP is a problem to find the shortest path in a graph visiting every nodes exactly once and returning to the start node [5]. The solution path is a Hamiltonian cycle of the graph. The TSP is a NP-hard problem [1], that means it is at least as hard as the hardest problems in NP.

The formal model of TSP can be given with the following linear programming description:

- N: number of nodes in the graph

- u_i: the position of the i-th city in the solution path
- D: distance matrix; d_{ij} is the weight of the edge from the i-th node to the j-th node; the distance values are non-negative values
- X: adjacency matrix of the Hamilton cycle; $x_{ij} = 1$ if there is a directed edge in the path from the i-th node to the j-th node; otherwise $x_{ij} = 0$
- the objective function of the path optimization is: $\sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} x_{ij} \to min$
- every node has only one incoming edge: $\forall j \, (i \neq j) : \sum_{i=1}^{N} x_{ij} = 1$
- every node has only one outgoing edge: $\forall i \, (i \neq j) : \sum_{j=1}^{N} x_{ij} = 1$
- there is only a single tour covering all cities: $u_i - u_j + N x_{ij} \leq N - 1$.

In the case of multi-salesman traveling (MTSP) problem more than one cycles should be generated (each salesman has a separate cycle) and each node is visited only by one salesman. For MTSP, the formal model should be extended with the following elements:

- M: number of salesmen
- constraints on the depo node (start and stop): $\sum_{i=1}^{N} x_{i0} = M, \sum_{j=1}^{N} x_{0j} = M$.

Due to the high complexity of TSP, there exists no algorithm for global exact optimization with polynomial cost. For example, the Held-Karp algorithm solves the problem in $\mathcal{O}(n^2 2^n)$ complexity. In order to provide an acceptable solution for real life problems, the TSP is usually solved with some heuristic optimization method.

From the family of popular evolutionary algorithms, the following methods are used most widely to find the good approximation of the optimal Hamiltonian cycle:

- Genetic algorithm [6]: the parameter vectors are optimized using the selection, crossover and mutation operators;
- Particle swarm optimization [7]: more agents are generated which move randomly but the optimum found by them is reinforced by other members of the colony;
- Ant Colony Optimization [8]: more agents are generated which collaborate with their environment;
- Tabu search [9]: the local search phase is extended with prohibitions (tabu) rules to avoid unuseful position testings.

Considering the standard heuristic algorithms, we can highlight the following methods:

- Nearest neighbor algorithm [10]: it selects the nearest unvisited node as the next station of the route;
- Pairwise exchange of edges [10]: two disjoint edges are removed from the route and two new edges are involved into the route to reduce the total cost.

In our investigation, we have focused on the most widely used method, the application of Genetic Algorithm.

2. Solving TSP using Genetic Algorithm

In Genetic Algorithm, the chromosomes or individuals are the basic building blocks in search state representation. In the TSP problem, a state of the search space corresponds to a route in the graph, i.e. to a permutation of the nodes. In our implementation, the permutation is given with the sequence of node indexes in order of traversing. Thus the route $(n_1, n_4, n_3, n_5, n_2)$ is given with the sequence $(1, 4, 3, 5, 2)$.

In the case of MTSP problem, the chromosome should represent the description of every cycles. There are four main representation forms for genotypes in the MSTP problem [11]:

- One chromosome technique (a special gene is used to separate the different cycle sections);
- Two chromosome technique: one chromosome is used for the description of the linked cycles, while the second contains the sections for the different agents;
- Multi chromosome technique: each agent has a specific chromosome to describe its route; the different chromosomes should be synchronized;
- Two-part chromosome technique: the chromosomes contain two parts. The first part is the linked list of the separate cycles, while the second part contains the cycle length values for the separate agents. For example, the chromosome $((1, 5, 6, 2, 3, 4)(4, 2))$ describe the following graph routes:

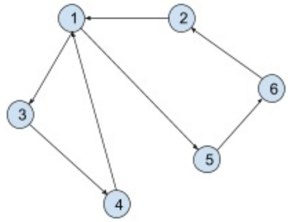

Figure 1

The implemented program applies the two-part chromosome technique to solve the MTSP problem. Regarding the genetic operators, the mutation of the route segment is performed with a swap operation. For example, the $(1, 5, 6, 2, 3, 4)$ route is modified to $(1, 5, 6, 4, 3, 2)$. In the case of two-part chromosome technique, the two segments are altered separately. In the agent segmentation part, the mutation is executed with selection of two positions where one of the values will be increased and the other will be decreased by one. For example, a mutation of $((1, 5, 6, 2, 3, 4)(4, 2))$ can be given with $((1, 3, 6, 2, 5, 4)(3, 3))$. For the crossover operation, the partially matched crossover method was implemented. In this method, some gene pairs are selected for substitution and the corresponding gene values are replaced with their substitution values. In the case of MTSP, only the route cycle part is altered with the crossover operation.

In our analysis, the base TSP and MTSP algorithms using genetic algorithms are used as baseline algorithms to be compared with the algorithm using node clustering.

3. Clustering Methods in Optimization

In the optimal route of the TSP, the edges usually connect the nearest nodes to each others. It would have no sense to make big jumps over and back among the nodes. This kind of behavior induces the heuristic rule of locality: near nodes in the route are near in the route graph too. Based on this heuristic assumption, it seems reasonable to group the near nodes into clusters and to perform a hierarchical optimization. There is a within-cluster optimization to determine the optimal route within the cluster and there is an intra-cluster optimization to determine the optimal path among the clusters. This approach implements the widely used divide and conquer concept, the problem of big complexity will be split into several subproblems of lower complexity.

In the literature, a big variety of clustering methods can be found. Most of the methods belong to the category of distance-based (discriminative) clustering [8] ,where the similarity between the objects are determined first and then the groups of similar objects are constructed. The main benefit of the distance-based methods is the simplicity and the descriptive power of the corresponding algorithm. In the case of model based clustering / generative clustering [7] approaches, a model type is specified a priori. A set of probabilistic generative models are defined, where each model corresponds to a group. The model for the cluster indexed by i is given with a parameter set λ_i. The method of expectation maximization is used to determine the assignments between the objects and clusters.

From another orthogonal viewpoint, the clustering methods can be categorized as hierarchical and partitional clustering. In the case of hierarchical method [12], a hierarchy of partitions are generated. At the top level, all objects are assigned to a single top-cluster, while the leaf nodes corresponds to the objects as singleton clusters. The most widely used hierarchical clustering method is the HAC [13] method, which uses an agglomerative clustering [14] algorithm, where existing groups are merged with similar objects / groups into new extended groups.

The partitional clustering [14] methods partition the objects into groups based on some optimization criteria. The objects are usually re-assigned from one group to some other group during the learning process. The k-means [15] clustering uses this partitional approach where clusters are represented by the centroids of the cluster members. This k-means method is the dominant method on the field of cluster analysis and also we have implemented this clustering as the baseline method in our investigation.

The input of the k-means method is the object set in a vector space and the initial set of the cluster centroids. The number of the required clusters and the initial positions of the centroids should be given as input parameters. The goal of the algorithm is to optimize the intra-cluster distances, i.e. the within-cluster sum of squares :

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} d^2(x, c_i)$$

In the formula, C_i denotes the i-th cluster with the centroid c_i. For the evaluation of the clustering, we have used the Silhouette measure [16]:

$$s_k = \frac{b_k - a_k}{max(a_k, b_k)}$$

where

a_k: the average distance of x_k to objects in cluster C_k

b_k: the average distance of x_k to objects in nearest cluster different from C_k.

4. Solving TSP using Clustered Vertices

In the literature, there are only few publications on application of clustering in TSP problems. In [2], the k-means clustering technique is implemented in the proposed two-levels optimization system. The work analyzes only the one-salesman problem and does not discuss the method of the intra-cluster optimization level. Only the one traveling salesman problem is investigated also in the work [4], where the HAC clustering method was applied. The publication [3] proposes a solution for the multi salesmen TSP domain. First the clusters of the nodes are generated and then each cluster is assigned to a single salesman.

In our investigation, which uses the k-means clustering method, we have extended the existing approaches with three novel elements:

- introduction of tree and multi level optimization
- adaption of k-means clustering to the MSTP problems.
- optimal selection of the k value

The three level optimization means that a hierarchy of k-means clustering is built up on the node set. In the single level clustering, the number of the nodes within a cluster was usually too high, thus the applied genetic algorithm could not achieve a good result. In order to reduce the complexity of the input data for within cluster TSP, the generated clusters can be clustered into subclusters in order to use the genetic algorithm with smaller node sets.

The multi level optimization can be used also for solving the MSTP problems. In this case, the top level corresponds to the salesmen, the number of clusters is equal to the number of salesmen. The resulted clusters can be sub-clustered into smaller sets for the genetic algorithm to solve the separate TSP problems.

The desktop test program was developed in Java and it uses the WEKA API to perform the clustering operations.

5. Test Results

We have performed several tests for both the TSP and MTSP problems with clustering and without clustering. The parameter ranges and of the test runs is given in Table 1 .

	TSP no clustering	MTSP no clustering	TSP with 2-level clustering	TSP with 3-level clustering	MTSP with 2-level clustering
number of nodes	10-1000	10-1000	10-1000	10-1000	10-1000
number of iterations	1000	1000	1000	1000	1000
size of population	100	100	100	100	100

Table 1: The parameter range of the tests

Regarding the efficiency of the different methods, Table 2 summarizes the quality of the optimization (length of the shortest path found by the algorithm). It can be seen that the application of the clustering module improves significantly the efficiency of the optimization process.

number of nodes	TSP no clustering	TSP with 2-level clustering	TSP with 3-level clustering	MTSP no clustering	MTSP with 2-level clustering
50	763	523	550	1328	1043
200	4358	1106	1202	6473	2385
400	10938	1647	1791	16820	4177
1000	36156	5801	3528	41052	6097

Table 2: The length of the local optimum route

The time costs of the different optimization process is shown in Table 3 .

number of nodes	TSP no clustering	TSP with 2-level clustering	TSP with 3-level clustering	MTSP no clustering	MTSP with 2-level clustering
50	5.9472	5.4596	9.2239	3.1367	6.2551
200	50.6509	20.6353	25.492	46.3972	21.1015
400	90.4167	4.384	6.1536	66.3395	4.6927
1000	729.451	12.1869	12.7243	555.1245	11.3891

Table 3: The run time of optimization process (sec)

The Figure 2 shows the generated routes for a sample MSTP problem with a single source and destination node. In Figure 3 and Figure 4, the two main clustering methods are compared from the viewpoint of clustering efficiency. Our experiences show that both methods have the same efficiency, they can provide very similar results. Regarding the optimal route length, the average distance is about 4%. The two clustering methods show a larger difference from the viewpoint of execution costs. Here, the k-means methods provides a faster execution than the HAC method.

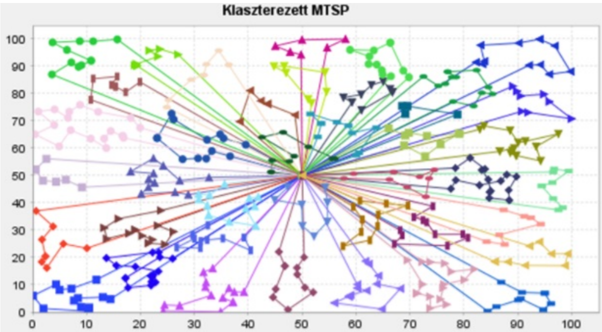

Figure 2: Optimum route of a sample MTSP problem

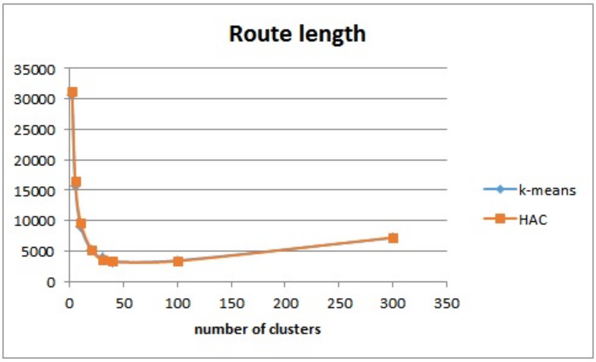

Figure 3: Optimum route with the different clustering methods ($N = 1000$)

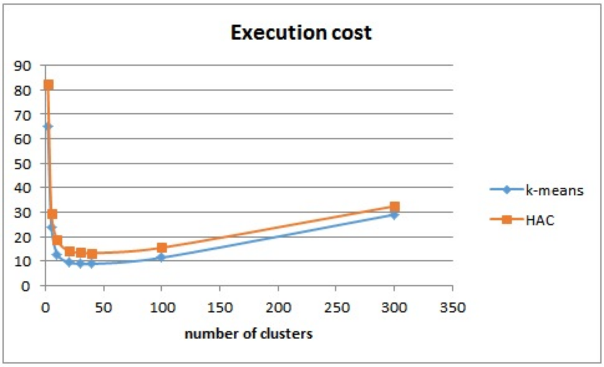

Figure 4: Execution cost of the route optimization ($N = 1000$)

6. Conclusion

For TSP problems with high number of nodes, the multi layered optimization is superior to the single level optimization. In the proposed multi layered optimization, the node set is partitioned into clusters or into hierarchy of clusters. The top cluster covers the whole input domain. For each cluster, a separate within cluster optimization is performed and then the local, cluster-level routes are merged into a global route. Based on the test experiments the proposed method is superior to the single level optimization method for both the TSP and MTSP problems.

Acknowledgments. The described article/presentation/study was carried out as part of the EFOP-3.6.1-16-00011 "Younger and Renewing University – Innovative Knowledge City – institutional development of the University of Miskolc aiming at intelligent specialisation" project implemented in the framework of the Szechenyi 2020 program. The realization of this project is supported by the European Union, co-financed by the European Social Fund."

References

[1] Held, M.; Karp, R. M., A Dynamic Programming Approach to Sequencing Problems, *Journal of the Society for Industrial and Applied Mathematics* Vol. 10 (1962), 196–210.

[2] Priyanka Yashwant Chavan, Ashwini Ravindra Gundale, Mrunali Shekhar Thorat, Ashwini Hirendra Jambhulkar, Two-Level Genetic Algorithm for Clustered Traveling Salesman Problem with Application in Large-Scale TSPs, *International Journal of Advance Research in Computer Science and Management Studies*, (2015), 335–340.

[3] R.Nallusamy, K.Duraiswamy, R.Dhanalaksmi, P. Parthiban, Optimization of Non-Linear Multiple Traveling Salesman Problem Using K-Means Clustering, Shrink Wrap Algorithm and Meta-Heuristics, *International Journal of Engineering Science and Technology* (2009), 129–135.

[4] S N S Kalyan Bharadwaj.B, és Krishna Kishore.G, Srinivasa Rao.V, Solving Traveling Salesman Problem Using Hierarchical Clustering and Genetic Algorithm, *(IJCSIT) International Journal of Computer Science and Information Technologies* (2011), 1096–1098.

[5] Arthur E. Carter , Cliff T. Ragsdale, A new approach to solving the multiple traveling salesperson problem using genetic algorithms, *European Journal of Operational Research 175* (2006), 246–257.

[6] Arthur E. Carter , Cliff T. Ragsdale, A new approach to solving the multiple traveling salesperson problem using genetic algorithms, *European Journal of Operational Research 175* (2006), 246–257.

[7] Elizabeth F. G. Goldbarg, Marco C. Goldbarg, Givanaldo R., Swarm Optimization Algorithm for the Traveling Salesman Problem, *EvoCOP 2006: Evolutionary Computation in Combinatorial Optimization* (2006), 99–110.

[8] Pan Junjie1, Wang Dingwei2, An Ant Colony Optimization Algorithm for Multiple Travelling Salesman Problem Proceedings of the First International Conference on Innovative Computing, *Information and Control (ICICIC'06)* (2006).

[9] Sumanta Basu, Swarm Optimization Algorithm for the Traveling Salesman Problem, *EvoCOP 2006: Evolutionary Computation in Combinatorial Optimization* (2006), 99–110.

[10] C. Nilsson., Heuristics for the traveling salesman problem., *Tech. Report, Linköping University, Sweden* (2003), http://www.ida.liu.se/~TDDB19/reports_2003/htsp.pdf

[11] Arthur E. Carter , Cliff T. Ragsdale, A new approach to solving the multiple traveling salesperson problem using genetic algorithms, *European Journal of Operational Research 175* (2006), 246–257.

[12] Anna Szymkowiak, Jan Larsen, Lars Kai Hansen, Hierarchical Clustering for Datamining, *in Proceedings of KES-2001 Fifth International Conference on Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies* (2001), 261–265.

[13] Maria-Florina Balcan Yingyu Liang y Pramod Gupta, Robust Hierarchical Clustering, *Journal of Machine Learning Research 15* (2014), 4011–4051.

[14] Anitha Elavarasi, J. Akilandeswari , B. Sathiyabhama, A SURVEY ON PARTITION CLUSTERING ALGORITHMS, *International Journal of Enterprise Computing and Business Systems (Online) (Online)* `http://www.ijecbs.com` Vol. 1 (1 January 2011),

[15] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, An Efficient k-Means Clustering Algorithm: Analysis and Implementation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2002), Vol. 24 881–892.

[16] Saitta S., Raphael B., Smith I.F.C.,A Bounded Index for Cluster Validity., *In: Perner P. (eds) Machine Learning and Data Mining in Pattern Recognition. MLDM 2007. Lecture Notes in Computer Science* (2007), Vol. 4571 174–187.