CASH LOGISTICS PROBLEM AS A SPECIAL CASE OF TRAVEL SALESMAN PROBLEM

S. Ivliev,
Prognoz Risk Lab, Perm
V. Sabirov,
Ph.D. student , Perm State National Research University
B. Singh,
Senior Bachelors Student, Indian Institute of Technology Delhi

Introduction

The standard traveling salesman problem (TSP) involves finding an optimal route which minimizes the total travel distance in a group of cities such that each city is visited only once [1,2]. Several applications exist of the TSP in the field of routing and logistics [4], as well as graphic information processing [3]. Relatively lesser studied are the applications of the clustered traveling salesman [2,6]. In this study we modify the clustered TSP to solve a problem of cash logistics for a multiple collecting vehicles.

In recent times there has been a large increase in the intensity of cash flows especially in emerging countries mainly because of the voluminous growth of the FMCG industry and retail. Money supply (M0) in Russia has increased more than 10 times between 2001 and 2011, with only 25% of this increase explained by inflation. Consequently the demand of delivery and encashment of money and treasures has also escalated. For taking advantage of this process top banks are attempting to develop their encashment services. Since large amounts of money are being transported it is important for the banks not just to optimize the routes for their vehicles, but also to consider the risk of loss of transported values.

In this study we investigate Cash Logistics Problem (CLP) as an extension of the TSP by considering a values logistics with additional criteria of optimality and multiple vehicles routing. The model is applied for the raw data of the city of Perm. First we provide the general formulation of the CLP, after that we attempt to develop the optimal routing procedure for some of the optimality criteria.

Mathematical statement of the CLP

In the general statement of the CLP bank considers a set of objects, P, which should be serviced by the money collector (cash collection or delivery).

Each object has a geo-position information (address, longitude, latitude). The typical set for a large city contains several hundreds of objects (see example for city of Perm's case on Fig. 1).

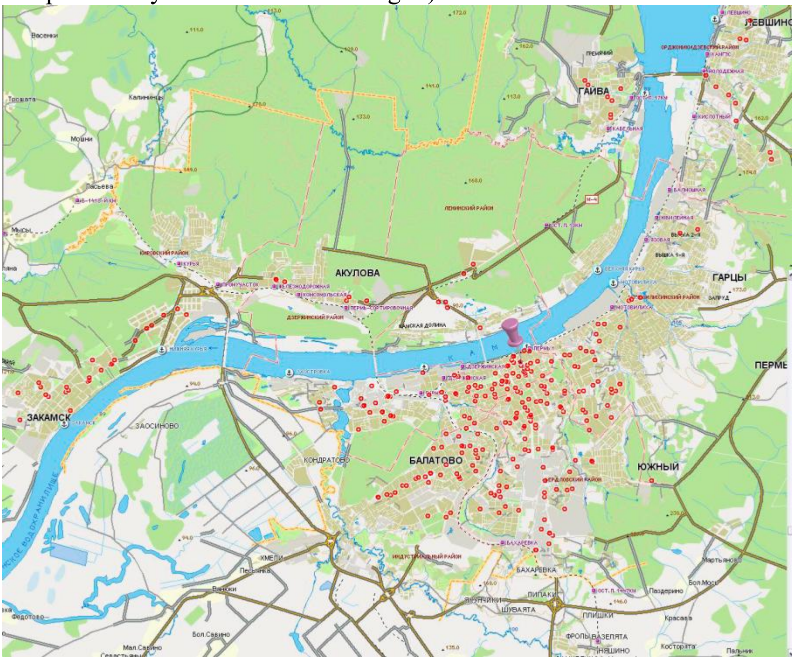

Fig. 1. Map of Perm with the geo-positions of points.

The following variables and constraints are specific to cash logistics problem:

- Each object has its average money amount which should be collected or delivered, $\forall p_i (i = 1..n) \in P \, \exists c_i$. If this amount is collected, then $c_i > 0$, if delivered $c_i < 0$.
- Each object has its average service time, $\forall p_i (i = 1..n) \in P \, \exists s_i > 0$.
- Each object has a time interval for service, $\forall p_i (i = 1..n) \in P \, \exists ts_i, tf_i : tf_i > ts_i ; ts_i, tf_i \in [Ts, Tf]$. Default values are: $T_s = 9$, $T_f = 18$.
- Cost matrix, $A = (a_{ij})$, where a_{ij} is the cost (distance/time) between the i^{th} and j^{th} points, $a_{ij} > 0$, $a_{ij} \neq a_{ji}$, $a_{ii} = 0$.
- m – number of routes in the routing network.

The task is to determine the optimal routing network for set of objects *P* consisting of *m* sub-networks with regard to optimality criteria (minimal time, minimal costs, minimal risk of loss of values, etc.). The solution *W* is a set of *m* vectors. Each vector is a sequence of serviced objects. Each object can be assigned to only one cluster. Thus all the *n* points are covered and all clusters are mutually exclusive.
In this study we take the simplified version of task without constraints for the service time. We also considered two criteria of optimality: min distance and min time.

Procedure for solution
We employ a four step scheme for solving the above stated problem.

Step 1: Clustering
We use cluster analysis to partition the set P of n available points into 'k' subsets such that each subset is handled by one salesman (expert truck driver). These *n* points can be grouped together on the basis of existing similarities or dissimilarities amongst them. Some possible examples of the similarities relevant to our example are:

- Physical proximity of the points in a cluster in terms of travel distance or travel time.
- Nature of service provided (cash delivery, encashment, ATM servicing).

We choose the only the first criteria for the clustering and obtain two independent sets of *k* clusters each. For this we use the standard k-means method. The method provides data separation into k mutually exclusive clusters, where k needs to be known beforehand. Then using an iterative scheme, the distances of all observations within a cluster are minimized from its centroid, ultimately terminating when the distances cannot be minimized further. The distance (time) to be minimized which we use is the road distance (road time) and not the Euclidean distance. Thus a good clustering will be one in which objects within a single cluster are as far as possible from another cluster and as close as possible within a cluster. We dwell on this postulate to find the optimal number of clusters later.

Step 2: Choosing an optimal number of clusters
An optimal number of cluster, *m*, will be one which,

- Has the number of trucks vehicles not too large or too small. A large number of trucks imply a large expenditure on the maintenance and operation as well and a small number would imply a larger time commitment from each driver and a corresponding waste of inventory.

- Has most of the points from separate clusters to be 'well separated' and also being close to each other within a cluster.

For the realization of this we use the silhouette metric which weighs the distances of each point within one cluster to points in the other clusters [16]. This was defined as,

$$S(i) = \frac{\min((b(i,:),2) - a(i))}{\max(a(i), \min(b(i,:))} \tag{1}$$

where *a(i)* is the average distance from the i^{th} point to the other points in its cluster, and, *b(i,k)* is the average distance from the i^{th} point to points in another cluster *k*.

Thus this value lies between +1 and -1; with +1 indicating a perfectly well separated cluster with no points overlapping with other clusters and -1 implying a totally imperfect cluster in which points which should have been in one cluster are wrongly assigned to another cluster. We solve the clustering stage for a set of increasing (and common sense) values of *k*. For each value we create these silhouettes and find the corresponding average values of the complete metric. The maximum of these average values is then chosen as the optimal number *m*.

Step 3: Deciding the first point of each cluster

This step is used for decision on the starting point from where the salesman (cash collecting vehicle) should begin his route. We touched upon this in Section 3. Our salesman starts from this point (home) and visits all the remaining $(n_i - 1)$ points $\forall i \in [2,3...m]$ in this particular cluster before returning to this first point again. Each cluster is independent of the other and handled by separate salesmen; thus theoretically the home could be chosen either, 1) randomly or, 2) such that the total travel path is minimized within each cluster for each salesman. However in our physical scenario we have one fixed point in S which is the 'base' camp. Physically this is the point where the entire money is collected/deposited. Thus we seek the salesman's first point in each cluster to be the point closest to this base camp.

For this we compare the road distances of each point of the m clusters to the base camp and choose the one with the minimum.

Step 4: Routing within each cluster

The final step is the development of the optimal route to be taken by the salesmen within each of the m clusters beginning from an assigned first point. We considered two optimality criteria:

- minimization of total road distance
- minimization of total time spent on the road

So now the problem is similar to solving the standard TSP with some of the abovementioned constraints separately in each of the clusters. Many algorithms exist for solving the TSP [18,19]. The emphasis in this study is not on the mathematical rigor but on achieving practical, swiftly computable yet close to the (global) optimum results; and results well-improved from the mind-calculations of the experts. Two well-known methods were employed here: a) the nearest neighbor method and b) the cross entropy method. Finally the results of these were compared with the routes which would have been employed by the experts given a particular cluster. This was done by handing out maps to them (with the service points present in a particular cluster) and asking them to prepare the routing for this cluster. Finally the total distance was computed.

The nearest neighbor method is an easy to conceptualize algorithm but may often give sub optimal results. Here from every point we choose the closest point which the salesman has not yet visited until all points are covered. However many paths may intersect amongst themselves and the final journey from the last point to the first point may be very large. As can be noted the total length of the journey depends on the first point.
The cross entropy method is a more 'efficient' method for solving the TSP. It is based on an iterative scheme where random data is generated in the first step which is then 'updated' to produce better samples in the next step [21]. Here the TSP is formulated originally as a minimization problem,

$$\min_{x\in \mathrm{X}} S\left(x\right)=\min_{x\in \mathrm{X}}\{\sum_{i=1}^{n-1}\left(c_{x_i,x_{i+1}}+c_{x_n,x_1}\right)\} \tag{4}$$

where, S(x) is the total distance/ time of tour $x \in X$ and X is the set of all possible tours, c_{x_i,x_j} - distance between the ith and jth node.

Thus first we need to know how to generate random tours and next how to update for newer and better tours. The random paths are generated by developing Markov chains for n points with a given transition matrix (distance or time matrix *A*) and parameter updating is done by using Rubenstein's updating formula.

Results and Discussions

The source data contains 237 points in Perm (see Fig.1) including some of the shops and other retail providers, as well as Sberbank branches and ATMs.
The time and distance matrices (237*237) were calculated based in GIS data. We solved the clustering stage for both of these choosing *k* values from 4 to 17. The silhouettes average values is shown on Fig.2. The maximum of silhouettes average was reached for *k*=14, which was selected as an optimal solution, *m*.

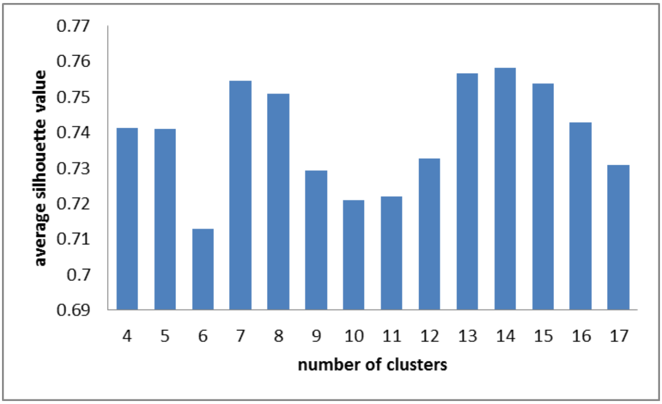

Fig.2. The plot of the silhouettes average values for different *k*

The shape of the silhouettes for optimal cluster partitioning is shown on Fig.3.

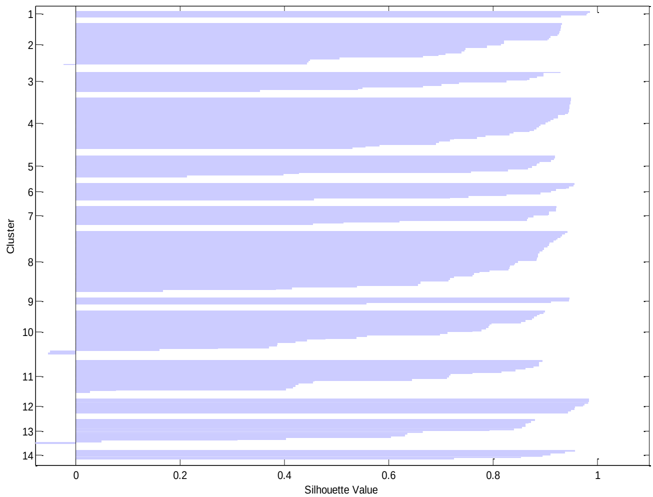

Fig.3. Complete silhouette for k=14

The breadth of each cluster gives us an estimate of the number of points in it (the complete vertical axis thus corresponds to 237).The length for each point (within a cluster) is an indicator of the quality of clustering with values lying between +1 and -1. For instance it can be seen that in cluster number 2, 10 and 13 some points have negative silhouette values which are examples of bad clustering. The area of each cluster (composed of a number of points) is averaged, which is then averaged again to give the values provided in Fig.2.

The routing was done using the nearest neighbor and the cross entropy methods. We also compared the results with the distances that would have been achieved by the experts. Some staff drivers of Prognoz company with large driving experience were asked to do routing on the map of Perm for a sample of the clusters. The route length comparison is provided on Fig.4.
As expected the cross entropy method has shown better routing than the nearest neighbor method and expert solutions.

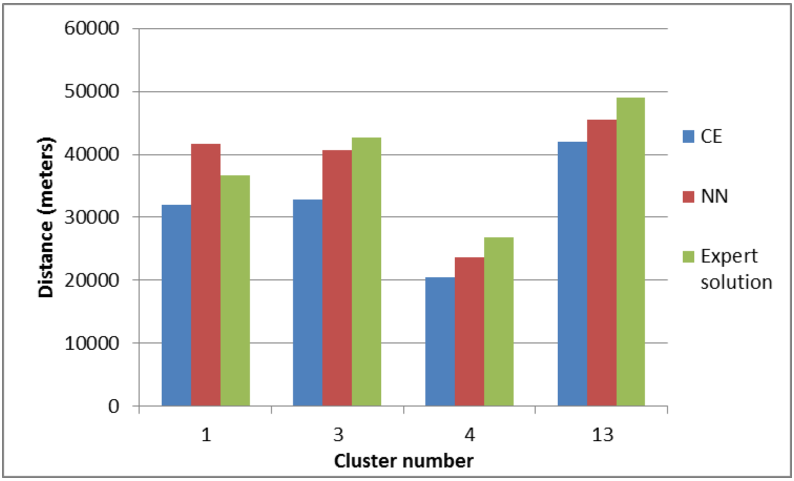

Fig. 4. Comparison of total route length for sample of the clusters

Conclusions and desired developments

The Cash Logistics Problem (CLP) could be considered as a special adaption of the standard Traveling Salesman Problem. The four step CLP solution algorithm which we have developed is applied and tested for the 237 point set of the city of Perm and can be extended or applied to other cities as well.

We gave a brief talk on our work at the Perm State University wherein several valuable suggestions and comments were given. We present some of them here for the inquisitive reader and hope to dwell on some of these for our further studies in this subject:

- Deciding the optimal number of clusters instead of using the average value of the silhouettes should be used. In this view we are attempting to not incorporate clusters with a high average (made possible by some very good clusters present in it) which also contains one (or more) bad clusters. However in our study, as discussed above, we chose that optimal value which averages to the maximum, despite containing some bad clusters (if any).
- Solving the complete problem, for each value of k, and then deciding on the optimal routing. However this would increase the running time of the problem enormously (especially for a larger city).
- Classify the problem’s complexity before solving it to decide which algorithm to choose.

We are currently working on incorporating the more general aspects of the CLP into our model. As also described above we plan to include new optimality criteria for the time, risk, penalty costs etc. Penalties can also be added for a delay in the service from the constrained opening/closing hours. Another desirable aspect of the model which we seek to include in our further studies (as suggested during our brief talk) is the large degree of randomness due to the presence of human factors.

Acknowledgements
The authors would like to thank the invaluable suggestions of the attendees of their small talk at Perm State University on the 22nd June 2011 for their comments and suggestions and Prof. Bela Myznikova for her advice regarding the solution. Bismark Singh would further like to thank Prognoz for the research grant provided.

References

1. *Lawler E.L., Lenstra J.K., Rinnooy Kan A.H.G., Shmoys D.B.* (eds.). The Traveling Salesman Problem. Wiley, New York.1985.
2. *Laporte G.* The Vehicle Routing Problem: An overview of exact and approximate algorithms. 1992.
3. *Bertin J.* Graphics and Graphic Information Processing. In SK Card, JD Mackinlay, B Shneiderman (eds.), \Readings in Information Visualization.1999.
4. *Lenstra J. K. and Rinnooy Kan A. H. G.* (1975). Some Simple Applications of the Travelling Salesman Problem, Operational Research Quarterly (1970-1977).1975. Vol. 26, N. 4. Part 1.
5. *Laporte G., Palekar U.* Some applications of the clustered travelling salesman problem. 2002.
6. *Dinga C., Cheng Y., Hea M.* Two-Level Genetic Algorithm for Clustered Traveling Salesman Problem with Application in Large-Scale TSPs.2007.
7. *Babaev A.A.* The formalization of the problem of the collector, Actual problems of economy and new technologies of teaching. Proceedings of the VIII International Scientific and Practical Conference.SPB.,2009. vol. 3.
8. The implementation of a parallel algorithm for solving the problem of conveying goods with the use of clustering, available at http://escience.ifmo.ru/files/hpc2009/2-2-bolgova.pdf accessed September 6, 2011. Original in Russian.

9. *Fozia Hanif Khan, Nasiruddin Khan, Syed Inayatullah, Shaikh Tajuddin Nizami*. Solving TSP problem by using genetic algorithm// International Journal of Basic & Applied Sciences IJBAS Vol. 9 N. 10.
10. *Suzdal V.A., Turukina E.A.*. Innovative methods of planning express freight transport company// poster at VII International Conference "Infocommunication technologies of the global information society" original in Russian : Суздальцев В.А., Турукина Е.А., Инновационные методы планирования экспресс-доставки грузов транспортной компанией.VRL: http://iktgio.mcrt.ru/rus/info.php?id=9921
11. Introduction to Data Mining by Pang-Ning Tan , Michael Steinbach and Vipin Kumar,(Chapter 8)
12. *MacQueen J.* Some methods for classification and analysis of multivariate observations.1967.
13. *Johnson S. C.*: Hierarchical Clustering Schemes// *Psychometrika*.1967.Vol. 2.P.241 – 254 .
14. *Dunn J. C.* : A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters// Journal of Cybernetics. 1973.Vol.3.P.32 – 57 .
15. *Bezdek J. C.* Pattern Recognition with Fuzzy Objective Function Algorithms// Plenum Press. New York,1981.
16. *Rousseeuw P. J.* Silhouettes: A graphical aid to the interpretation and validation of cluster analysis.1986.
17. *Kaufman L. and. Rousseeuw P. J.* Finding Groups in Data: An Introduction to Cluster Analysis. Hoboken, NJ: John Wiley & Sons, Inc., 1990.
18. *Bektas, T.* (2006) The multiple traveling salesman problem: an overview of formulations and solution procedures// Omega: The International Journal of Management Science.2006. Vol.34. (3).P. 209 - 219 . (doi:10.1016/j.omega.2004.10.004)
19. *Bezalel G. and Kizhanathan S.:* An optimal solution method for large-scale multiple traveling salesmen problems//*Operations Research*. 1986.Vol.34(5).P. 698 – 718 .
20. *Gutin G., Yeo A. and Zverovich A.*, Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP. Discrete Applied Mathematics. 2002.Vol.117.P.81 – 86 .
21. *De Boer P-T., Kroese D.P, Mannor S. and Rubinstein R.Y.* A Tutorial on the Cross-Entropy Method// *Annals of Operations Research*.2005.Vol. 134 (1).P. 19 – 67 .