Optimization based clustering algorithms in Multicast group Hierarchies

Long Jia [1], A. Bagirov, I. Ouveysi, A. M. Rubinov

School of Information Technology and Mathematical Sciences, University of Ballarat, Victoria, 3353, Australia

Abstract– **In this paper we propose the use of optimization based clustering algorithms to determine hierarchical multicast trees. This problem is formulated as an optimization problem with a non-smooth, non-convex objective function. Different algorithms are examined for solving this problem. Results of numerical experiments using some artificial and real-world databases are reported. We compare several optimization based clustering methods and their combinations with the k-means method. The results demonstrate the effectiveness of these algorithms.**

Index terms– optimization, clustering algorithms, multicast, hierarchies.

I. INTRODUCTION

The main objective of multicast communication is to supply various group communication services with required QoS (Quality of Service)while reducing the cost of data transfer (i.e. minimizing number of data copies are sent for a group). However, two problems have to be solved when implementing multicast over multiple networks: Firstly, an efficient routing mechanism has to be developed; secondly, the cost of maintaining multicast routing has to be reduced in supporting the scalability of the multicast services.

One of the muticast routing approaches is to treat the problem as computing optimal Steiner Trees in graphs. The cost function can incorporate the QoS constraints. The optimal Steiner Tree problem has been shown to be NP-Complete. As one of the possible methods, hierarchical multicast trees offer cost-effectiveness, much more flexibility and scalability, as local trees are built and maintained independently in each cluster. Hierarchical routing protocols have been introduced in the early 80's [8]. Hierarchical clustering for multicast routing is discussed by many authors([3], [5], [10], [13], [6], [12], [4]). The key of this problem here is how to do the hierarchical clustering.

In this paper we emphasize the using of optimization based clustering algorithms to this problem. To simplify the problem we only consider a two-level hierarchical multicast tree, see Fig. 1. We further only consider the geographical locations of the nodes: all nodes are equally weighted and the distance covered by the tree is the total cost of our tree topology network. However, the approach also allows us to consider different types of cost functions.

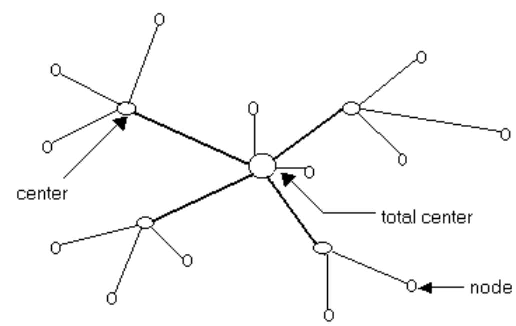

Fig. 1. A two-level Hierarchical Multicast Tree

II. THE SETTING OF THE PROBLEM AND ALGORITHMS FOR ITS SOLUTION

Assume that a set A of m nodes on the plane is given:

$$A = \{a^1, \dots, a^m\}, \text{ where } a^i = (a_1^i, a_2^i).$$

Let us decompose A into a given number k of disjoint subsets (clusters) $A^i, \ i = 1, \dots, k$. The choice of k is according to the scale of network and hops-constraint(hierarchical levels). We can consider different values of k and then choose the one which gives the best result. In each cluster we choose one node as a centre (x_i) and all other nodes will connect to this centre. Then we will choose one node (x_*) from all nodes as a total centre; all centres will connect to this total centre. These centres can be found by minimizing the so-called cost function. Assume that clusters $A^1, \dots, A^k$, their centres $x_1, \dots, x_k$ and the total centre x_* are given. Then the total cost C of this tree can be calculated as

$$C(A^1, \dots, A^k, x_1, \dots, x_k, x_*)$$

[1] Corresponding author. E-mail address: ljia@students.ballarat.edu.au.

$$= \sum_{i=1}^{k} \sum_{a \in A^i, a \neq x_*} \|a - x_i\| + \sum_{i=1}^{k} \|x_i - x_*\|.$$

We include the condition $a \neq x_*$, since when we do clustering the total centre node belongs to one of the clusters. Our goal is to solve the following problem (P): find clusters $\bar{A}^i$, their centres $\bar{x}_i$ and the total centre $\bar{x}_*$ such that

$$C(\bar{A}^1, \ldots \bar{A}^k, \bar{x}_1, \ldots, \bar{x}_k, \bar{x}_*)$$
$$\leq C(A^1, \ldots, A^k, x_1, \ldots, x_k, x_*)$$

for each collection of clusters A^i, their centres x_i and the total centre x_*.

This is a global optimization problem, however this formulation is not suitable for the direct application of optimization techniques. Thus we reformulate the problem (P) to the problem in which it can be solved by methods of non-smooth optimization.

III. FIRST APPROACH: THE USE OF ARTIFICIAL CENTRES

First, we assume that the centre of a cluster A^i ($i = 1, \ldots, k$) is not necessarily a real node, it can be an arbitrary point y_i on the plane. In such a case we call it an artificial centre. If the set $(y_1, \ldots, y_k)$ of artificial centres of clusters is known then the clusters themselves can be easily described; namely the cluster A^i consists of points $a \in A$ such that

$$\|y_i - a\| < \min_{i' \neq i} \|y_{i'} - a\|. \tag{1}$$

Here $\|x\|$ is the Euclidean norm of a point x.

We suggest two approaches to define the total centre y_*. First we define y_* as the centroid of the set $(y_1, \ldots, y_k)$:

$$y_* = (1/k)(y_1 + \ldots + y_k). \tag{2}$$

Assume that centres of clusters $y_1, \ldots, y_k$ are known. It follows from (1), and the definition of y_* that the total cost $\tilde{C}$ of this tree depends only on centres:

$$\tilde{C}(y_1, \ldots, y_k) = \sum_{a \in A} \min_{i=1,\ldots,k} \|y_i - a\| + \sum_{i=1}^{k} \|y_i - y_*\|, \tag{3}$$

where y_* is defined by (2).

Recall that the search for clusters can be characterized as the simultaneous minimization of the variation within clusters and the maximization of the variation between clusters. A formalization of this idea can be given as follows:

We say that a set $(y_1, \ldots, y_k)$ of artificial points is the set of centres of k-clusters of the set A if

$$\sum_{a \in A} \min_{i=1,\ldots,k} \|y_i - a\| = \min_{y'_1,\ldots,y'_k} \sum_{a \in A} \min_{i=1,\ldots,k} \|y'_i - a\|, \tag{4}$$

where the minimum is taken over all collections of points $y'_1, \ldots, y'_k$, where y'_i belongs to the plane. A detailed discussion of this definition can be found in [2]. Let

$$f_1(y_1, \ldots, y_k) = \sum_{a \in A} \min_{i=1,\ldots,k} \|y_i - a\|, \tag{5}$$

$$f_2(y_1, \ldots, y_k) = \sum_{i=1}^{k} \|y_i - y_*\|. \tag{6}$$

Then $\tilde{C} = f_1 + f_2$. Thus the minimization of cost does not coincide with the search for clusters. Indeed, the search for clusters can be characterized as the simultaneous minimization of the variation within clusters and the maximization of the variation between clusters. This can be done by the minimization of f_1. (See [2] for details.) However we are not interested in the maximization of the variation between clusters. This is the reason for involving the term f_2. Thus the problem of the finding best artificial centres is reduced to the following unconstrained optimization problem:

$$\text{minimize} \quad \tilde{C}(y_1, \ldots, y_k)$$
$$\text{subject to}$$
$$(y_1, \ldots, y_k) \in R^{2k} \tag{7}$$

Assume that we have found a collection of artificial centres $y_1, \ldots, y_k, y_*$ by the minimization of the cost function $\tilde{C}$. The most natural way to find real centres is the following: we keep the clusters that are found by the minimization; then we substitute y_i with x_i which is the node from the cluster A^i closest to y_i and we substitute y_* with the closest node x_* from all of the nodes.

We can use methods of constraint optimization in order to reduce the difference between the cost with artificial centres and the real situation. Consider the following optimization problem;

$$\text{minimize} \quad \tilde{C}(y) \;\; \text{s.t} \;\; h(y) = 0, \tag{8}$$

where $y = (y_1, \ldots, y_k)$ and

$$h(y) = \sum_{i=1}^{k} \min_{a \in A} \|y_i - a\|. \tag{9}$$

It is easy to check that $h(y) = 0$ is equivalent to the following: for each i there exists $a \in A$ such that $y_i = a$. This means that each y_i is a real node. We can convert the constrained problem (8) to an unconstrained problem, using the penalty approach. Namely, a solution of the following unconstrained problem

$$\text{minimize} \qquad \tilde{C}(y_1, \ldots, y_k) + \lambda h(y_1, \ldots, y_k) \quad (10)$$

is close to a solution of (9) for some λ. Thus we can automatically find centres of clusters $y_1, \ldots, y_k$ that are close to real nodes. However, the total centre is still artificial and we need to find a corresponding real node to replace this artificial center.

Numerical experiments show that this approach works well if the penalty coefficient λ is not too large.

IV. SECOND APPROACH: DIRECT CALCULATION OF CENTRES

The alternative approach is to directly calculate the centres as nodes from the data set A. We compute both cluster centres and the total centre as a solution to an optimization problem with the following objective function

$$f(x) = \sum_{i=1}^{m} \min_{1 \leq j \leq k} \|x^j - a^i\| +$$

$$\tau_1 \sum_{j=1}^{k} \min_{1 \leq i \leq m} \|x^j - a^i\| + \tau_2 \min_{1 \leq i \leq m} \sum_{j=1}^{k} \|a^i - x^j\| \quad (11)$$

The function f consists of three terms: first term is a cluster function, the second term presents a penalty in order to get cluster centres as nodes from the set A and third term is a penalty in order to get a total centre as a node from the set A. Here τ_1,$\tau_2 > 0$ are penalty coefficients.

V. ALGORITHMS

Both cost functions from (10) and (11) are nonsmooth and non-convex. Both problems (10) and (11) are global optimization problems. However, global optimization techniques fail to solve these problems, because the number of variables in this problems is large.

We use the derivative-free discrete gradient method [1] for local minimization of $\tilde{C}$. Numerical experiments confirm that the discrete gradient method escapes from saddle points and sometimes even from shallow local minima.

To compare with other clustering algorithm, k-means method is used here. The use of the k-means method in clustering for network design was proposed in ([9]). In our paper we use the k-means method for clustering for the search of artificial centres $(y_1, \ldots, y_k)$ of clusters. This method was used before in [6, 7]. Then each artificial centre y_i should be replaced by the closest to y_i real node x_i. The total artificial centre y_* should be also replaced by the closest real centre x_*. The search for centres of clusters by the k-means method is cheap. Unfortunately the k-means method is very sensitive to the choice of initial point. This is the main drawback of this method. Thus we need to apply the k-means method many times starting with different initial points, and then to choose a collection of real centres which provide the smallest cost C. In our numerical experiments for TC(k)(Total cost with number k of clusters)we apply k-means method with 20-30 different initial points and then choose the best result.

VI NUMERICAL EXPERIMENTS

In our first numeral experiment we use the set of the geographical locations of 51 North American cities. We got this data from the picture included in the paper[6]. In paper [6], the author chose $k = 6$ to maintain a maximum fan-out of six at each level by using the similar data. In our first numeral experiment we choose $k = 6$ within a two-level hierarchical tree topology in order to make our results comparable to theirs in someway. We designate the optimization based clustering algorithms described in section III with artificial centres as *op*1 and described in section IV with direct calculation of centres as *op*2.

Table 1: Cost comparisons: the combination of op2, op1 and k-means method

k-m	op2	*op*1	1-*km*	*km*-1
374	313	392	<u>308</u>	371
344	314	415	322	371
344	316	392	325	322
344	<u>308</u>	389	322	317
318	<u>308</u>	390	317	312
323	313	352	320	325
320	309	384	315	336
338	313	383	324	333

In Table 1 we compare the results from the *op*2 method, the *op*1 method, the k-means method and the combination of these methods, with the same initial centre nodes. Here 1-*km* means we choose the result of centres from the k-means method as the initial centre nodes for the $op1$; while km-1 means the opposite. Each row represents a different group of initial centre nodes; and the column represents the real total cost (TC)(in some units) of the tree. From our numeral experiment there are some conclusions: the k-means and op1 methods are influenced by the choice of initial centres; while the $op2$ is almost insensitive to the initial centre nodes; The combination of the k-means method with the op1 method($op1 - km$) can improve the results from both op1 and k-means method only. Using the $op1 - km$ method and $op2$ method we got the best result from all our experiments with this group of initial nodes : $TC = 308$. Thus the $op2$ allows us to improve the best result obtained by the k-means method (TC=318) by more than 2.5%.

Results of our experiments show that the values of $\tau_1 = 3$, a $\tau_2 = 1$, and $\lambda \in (2, 10)$ are best for τ_1,τ_2, and λ respectively. When the value of penalty parameter is too big the algorithm will choose the initial nodes as centres and for some initial nodes there will be many clusters with only one node, and one cluster with nearly all the other nodes.

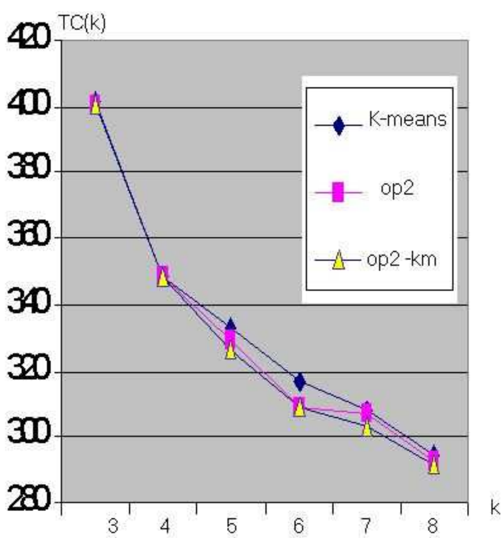

Fig1. TC(k) for data 51-node

The total cost (TC) with different k(where k is the number of clusters) for this data is shown at Fig. 1.We only consider the $op2$ method and its combination with k-means method here. This result shows that $op2$ method is better than k-means method. The combination $op2 - km$ can sometimes improve the results from k-means method and give the best result from all methods.

In our second numerical experiment we use a randomly generated database with 2,000 nodes on the plane. The result is presented in Fig.2. When k is less than 10 k-means method is better than $op2$. Then $op2$ method can offer more than 2.5 percents better result compared with the k-means.

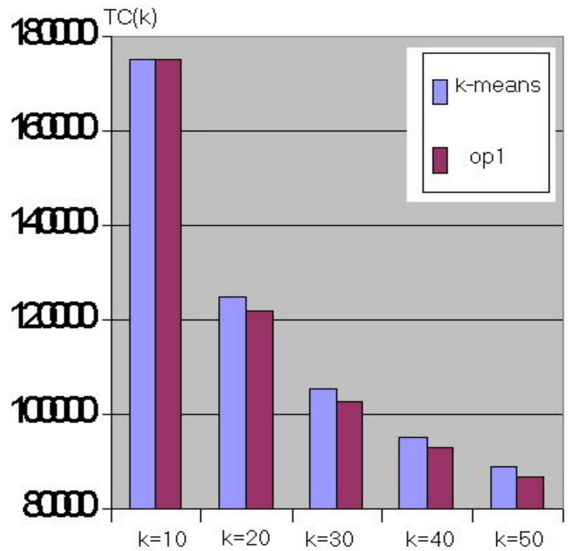

Fig2. TC(k) for data 2000-node

VII. CONCLUSIONS AND FURTHER RESEARCH

In this paper we examined two different optimization based clustering algorithms, k-means method and their combination to solve the multicast routing problem. From our numerical experiments we conclude: the optimization approaches are better than the k-means method.

Our techniques for organizing hierarchies using clustering can not only solve the scaling problem in multicast routing but also help evolve the network and other services such as Reliable Multicast Transport(RMT), Quality of Service(QoS). For example, our techniques can be used in multicast security when the clustering hierarchies are used [11].

The extension of the discussed algorithms for solving large scale network for Multicast group hierarchies problem is the subject of our further research.

1 ACKNOWLEDGEMENTS

We would like to thank Dr. Gill Waters for her kind offer her paper [7] and some useful discussion through emails. Our colleagues Nadejda Soukhoroukova, Jiapu Zhang provide their very helpful coding work as

well.

References

[1] A.M. Bagirov (1999) Derivative-free methods for unconstrained nonsmooth optimization and its numerical analysis, *Investigacao Operacional* **19**.

[2] A.M. Bagirov, A.M. Rubinov and J. Yearwood, A global optimization approach to classification, *Optimization and Engineering*, 3, 2002, 129-155.

[3] F. Baccelli, D.Kofman, J. L. Rougier, *Self organizing hierarchical multicast trees and their optimization*, Proceedings of IEEE Inforcom'99, Vol. 3, pp1081-1089, 1999.

[4] Duc A. Tran Kien A. Hua, Tai Do *ZIGZAG: An Efficient Peer-to-Peer Scheme for Media Streaming* http://www.cs.ucf.edu/dsg/ publication/2003/infocom03.pdf

[5] Hagouel Jacob, *Issues in routing for large and dynamic networks*, Ph.D thesis, Columbia University, 1983

[6] Gill Waters, *Hierarchies for network evolution*, Sixteenth UK Teletraffic Symposium on " Management of Quality of Service - the New Challenge", Harlow, UK, May 2000

[7] G. Waters,*Applying clustering algorithms to multicast group hierarchies*, private communication.

[8] Leonard Kleinrock and Farouk Kamoun, *Hierarchical routing for large networks_Performance evaluation and optimization*, Computer networks, pp155-174, 1977

[9] Robert S. Cahn, *Wide Area Network Design: concepts and tools for optimization* , Morgan Kanufmann, 1998

[10] Samir Chatterjee and Mostafa A. Bassiouni, *Hierarchical Message Dissemination in Very Large WAN's*,17th Conference on Local Computer Networks, IEEE computer Society Press, 1992

[11] Suman Banerjee,Bobby Bhattacharjee,*Scalable secure group communication over IP multicast* , IEEE Journal on selected areas in communication, Vol. 20, No.8, October 2002

[12] Suman Banerjee, Bobby Bhattacharjee, Christopher Kommareddy *Scalable Application Layer Multicast*,ACM Sigcomm, August 2002 (Also http://www.cs.umd.edu/projects/nice)

[13] R. Venkateswaran, C. S. Raghavendra, X. Chen, and V.P.Kumar, *Hierarchical Multicast Routing in ATM Networks*, IEEE International Conference on Communications, 1996