階層的クラスタリングにおける最適解と局所最適解に関する一考察

3M－1

リチャード S. ワレス　末永 康仁

A look at global and local minima in hierarchical clustering

Richard S. Wallace and Yasuhito Suenaga
NTT Human Interface Laboratories
Nippon Telegraph and Telephone Corporation
1-2356 Take, Yokosuka 238-03 Japan

May 25, 1990

要約
コンピュータビジョン、パタン認識の手法の中でクラスタ解析は最も応用範囲の広い手法のひとつであり、カラー画像のセグメンテーション、知覚における群化、Hough変換などに応用されている。しかし最適クラスタリングを求める問題はNP-hardとなりうる。このため、我々は準最適な手法を検討せねばならない。本稿では、ある種の階層的クラスタリング問題における2種類の準最適な手法によって得られた解と最適解との違いを調べる実験を行なったので報告する。まず、少数の点の集合に対して、しらみつぶしにより全てのクラスタリングを求め、真の最適解を求めた。次に、準最適化手法により局所的最適解を求め、これと真の最適解を比較した。この結果、通常、局所的最適解は目的関数の値においては真の最適解に極めて近いが、求められるクラスタ木のトポロジカルな構造は極めて異なる場合が多いことがわかった。

1. Introduction.

Hierarchical clustering (HC) amounts to finding a cluster tree that minimizes an objective function defined recursively over the tree. In the case of a binary tree t for example, HC seeks to minimize $e(t)$ where $e(t) = 0$ when $t = NIL$ and $e(t) = f(t) + e(l) + e(r)$ when t has child subtrees l and r. The value of $e(t)$ is called *energy*. The question of whether a particular minimization procedure finds a global versus a local optimum arises often in discussions of HC under general objective functions $f(t)$. When $f(t) = d(l, r)$ is the nearest-neighbor distance between the sets $\mathcal{S}(l)$ and $\mathcal{S}(r)$ (the leaf nodes of subtrees l and r respectively), the agglomerative algorithm (e.g. [1]) is known to yield the global optimum solution. Pitt and Reincke [2] showed general conditions under which agglomeration finds an optimal *level* clustering, and pointed out other cases in which finding the optimal clustering is *NP*-hard. This paper summarizes experiments designed to compare two suboptimal algorithms, agglomeration and *NIHC* (Numerical Iterative Hierarchical Clustering)[3], with an optimal clustering found by a brute-force procedure.

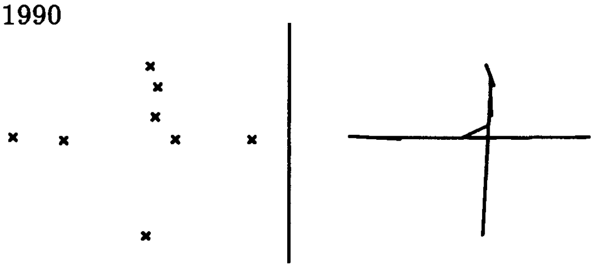


Figure 1: Random data and its optimal cluster tree. The tree is depicted by line segments connecting the mean values of the sets represented by each pair of sibling nodes.

The number of unique cluster tree for $n > 1$ points is

$$c(n) = \frac{1}{2}\sum_{k=1}^{n-1}\binom{n}{k}c(k)c(n-k) = \frac{1}{2^{n-1}}\frac{(2n-2)!}{(n-1)!}, \quad (1)$$

a number larger than $n!$ when $n > 6$. We wrote a combinatoric procedure that actually generates all $c(n)$ cluster trees and measures their energies. This program exhausts memory on a Sun-4 when $n = 9$ and in any case is very slow. But at least for the smallest set sizes n, it became possible to pinpoint the global minimum cluster tree. Thus, we proceeded to compare empirically the locally optimal trees found by efficient heuristic procedures to the true global optimum.

Every internal node u of a cluster tree represents a set having covariance $\mathbf{C}(u)$. When the clustering objective function is the *Gaussian entropy* function $f(t) = \log|\mathbf{C}(u)|$, where $|\cdot|$ is matrix determinant, we observe that the global minimum energy cluster tree separates clusters even in difficult cases like the "cross" pattern in figure 1. But since the agglomerative algorithm does not necessarily find this minimum, we experimented with using *NIHC* to improve the result. The input to *NIHC* is a cluster tree, such as a *k-d* tree or the output of agglomeration. *NIHC* then iteratively rearranges subtrees to reduce the tree energy. Sometimes the results of agglomeration can be improved by *NIHC*, without changing the order of computational complexity.

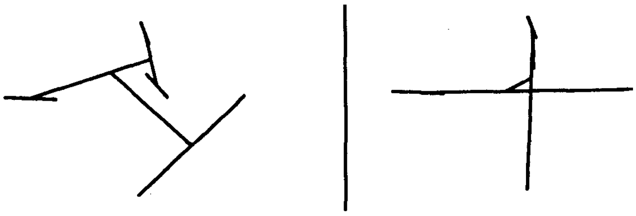
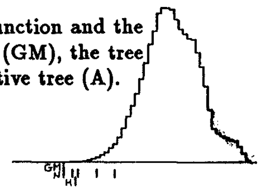
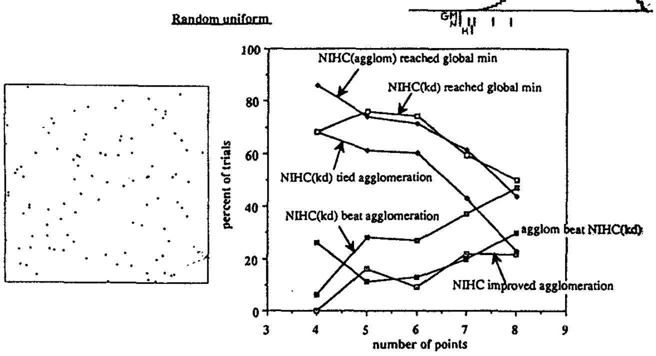
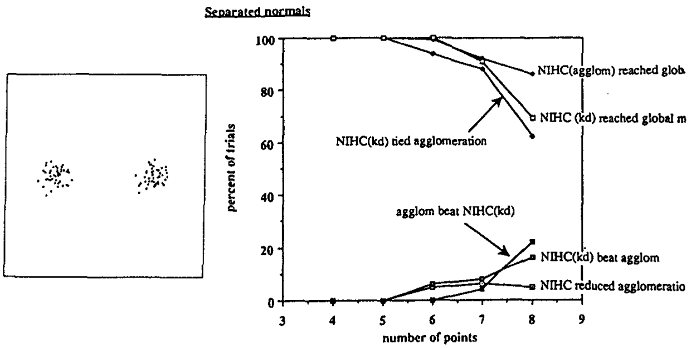


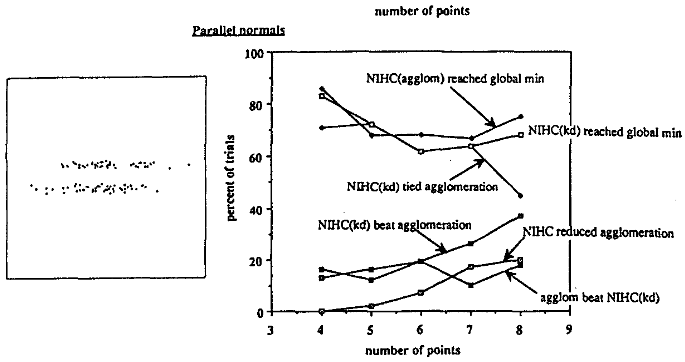
Figure 2: Agglomeration tree (left) and *NIHC* tree (right)

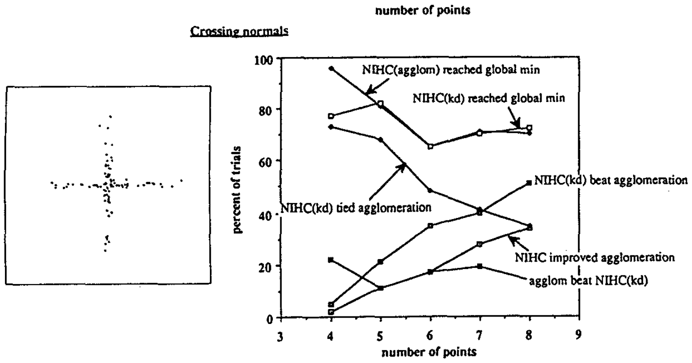
Figure 3: Histogram of tree objective function and the energies of global minimum energy tree (GM), the tree found by *NIHC* (N), and the agglomerative tree (A).











2. Experiments. We randomly generated $n \in \{4, \ldots, 8\}$ points from four types of two-dimensional distributions shown in figures. Figure 3 shows a histogram of tree energies for the set of trees constructed from the "crossing" point set in figure 1. Note that although the energy of this local optimum is close to the energy of the global optimum, the two trees are quite different topologically.

The experiments consist of running 100 random trials like the one illustrated in figures 1 through 3. The data are randomly selected from $[0, 0] \times [255, 255]$. Leaf nodes have $\sigma_{xx} = \sigma_{yy} = 1.0$ and $\sigma_{xy} = 0.0$. The data labeled "*NIHC* improved agglomeration" show the number of times that *NIHC* could reduce the energy of the tree found by agglomeration. The curve "*NIHC*(agglom) reached global min" is the number of times that either agglomeration alone combined with *NIHC* found the global minimum. The data called "*NIHC(kd)* beat agglomeration", "*NIHC(kd)* tied agglomeration", and "agglomeration beat *NIHC*" show the number of trials when *NIHC* started with the *k-d* tree and found a lower, the same, or higher energy tree than agglomeration, respectively. Finally, the curve "*NIHC(kd)* reached global min" shows the percentage of trials in which *NIHC* found the global minimum energy cluster tree when seeded with the *k-d* tree.

3. Conclusion. This research has shown that the energy of solutions obtained by both *NIHC* and agglomeration may be very close to the global minimum energy, and that for minimizing the Gaussian entropy objective function *NIHC* is marginally better than agglomeration. While inconclusive, the data reported here allows us to make some informed conjectures about finding the global minimum HC. *NIHC* finds the global optimum when starting with the *k-d* tree almost as often as when starting from the agglomerative tree. As the number of points increases, (1) the chance of *NIHC* terminating at the global minimum decreases. (But does it converge to zero?) (2) the chance of *NIHC* improving the result of agglomeration increases. (3) The chance that *NIHC* starting from the k-d tree is better than or equal to agglomeration increases.

Acknowledgement. Thanks to Ken Goldberg (CMU) for suggesting the experiments reported here.
本文をまとめるのあたり、
有益な御助言を頂いたCMUのKen Goldberg氏に感謝いたします。

References

[1] Duda, R. O. and Hart, P. E.: *Pattern Classification and Scene Analysis*, Wiley-Interscience (1973).

[2] Pitt, L. and Reinke, R.: Criteria for polynomial-time (conceptual) clustering, *Machine Learning*, 2 pp. 371-96 (1987).

[3] Wallace, R. S.: *Finding Natural Clusters through Entropy Minimization*, PhD thesis, Carnegie Mellon (1989).