Hill-Down strategy based DENsity CLUstEring and its application to medical image data (Work-in-Progress)

Conghua XIE
Depart. of Computer Science and Engineering Changshu Institute of Technology, China 215500
x7c8h5@yahoo.com

Jinyi CHANG
Depart. of Computer Science and Engineering Changshu Institute of Technology, China 215500
cjy@cslg.edu.cn

Yuqing SONG
School of Computer Science and Engineering, Jiangsu University, China 212013
yqsong@ujs.edu.cn

ABSTRACT
In order to overcome problems of DENCLUE and improve it for medical image segmentation, we propose HD-DENCLUE algorithm. It takes the optimization step length to find density attractor, designs a hill-down strategy to give different density thresholds for different clusters and stops at hill-down data i.e. the edge data of this cluster. Experiments show that HD- DENCLUE can decrease time overhead, have better clustering effects than DENCLUE and get the edges of each cluster.

Categories and Subject Descriptors
H.2.8 [Database Applications] *Image databases*

General Terms
Algorithms

Keywords
Hill-down strategy; density clustering, hill-climbing

1. INTRODUCTION
Clustering has strong inner relationships with segmentation of image. Increasing attentions have been paid to methods of clustering for image segmentation.

Hinneburg [1] offered DENCLUE to cluster large multimedia database and can deal with database containing many noise data. Li Cun-hua[2] improved its approximation density computation based on data grids technology and Gan Wen-yan[3] offered a method of getting the two key parameters of DENCLUE.

However, with a close looking into DENCLUE, we can still find some problems: hill climbing strategy of DENCLUE cannot find density attractors as fast as possible because of constant step length; it is not reasonable that only one density threshold is used to judge many clusters; it is not easy to get the boundary of each cluster. To address those problems, we present HD-DENCLUE, a new clustering implementation strategy on the basis of DENCLUE.

Infoscale 2007, June 6-8, 2007, Suzhou, China

2. HD-DENCLUE: HILL DOWN TRATEGY BASED DENSITY CLUSTERING
DENCLUE algorithm: firstly map data into grids i.e. all data are mapped into proper grids and all the populated grids are numbered depending on their relative positions from a given origin. Secondly, the populated grids are classified into highly populated grids and sparse ones. Only highly populated grids are clustered according to density function and cluster definitions [1], while points in sparse girds are ignored as outliers. Finally, DENCLUE uses hill-climbing strategy to extract clusters.

The basic idea of hill-climbing strategy: stochastically choose a data as current data, climb the hill of density function with a step length, search the next data along its gradient direction and compare density values of current data and next data; if the density value of next data is larger, then replaces the next data as new current data and continue searches the next data with the step length along gradient direction, until the density value of next data is smaller than current data. If density value of current data is larger than threshold, then all those data visited by hill climbing are in one cluster; else those data are outliers.

The key idea of our approach is to take hill-down strategy to cluster data and optimization gradient method to search density attractors on the basis of density function constructed by DENCLUE.

2.1 Find Attractors of Density Function With Variant Step Length
Firstly, local approximation density function is constructed by [1-3]. Secondly, an optimization gradient method offered by Cauchy to search the local maximum data [4].

Theorem 1. Let dataset $D=\{X_1,...,X_n\} \subset R^d$, $f(X)$ be density function, λ_i be step length of i[th] step, $X_{i+1} = X_i + \lambda_i \cdot \frac{\nabla f(X_i)}{\| \nabla f(X_i) \|}$ and unit vector of $f(X)$ at

X_i be $S_i = \frac{\nabla f(X_i)}{\| \nabla f(X_i) \|}$, then $\lambda_i = \frac{-\nabla^T f(X_i) S_i}{S_i^T \nabla^2 f(X_i) S_i}$ is the optimization step length for the i^{th} step.

2.2 HD-DENCLUE Algorithm

According to theorem1, hill-climbing strategy of DENCLUE can be improved with variable optimization step length to find the density attractors. Then, walk down hill from attractor to extract its cluster data. In order to implement downhill clustering idea, there are three problems need solving.

(1) If current location is X_k, then how do we find the downhill direction p of $f(X)$ at X_k?

(2) After find the downhill direction p, how to assign step length to move from X_k to $X_{k+1} = X_k + \lambda p$ along p?

(3) How to judge current data point X_k is at hill-foot or not?

The first problem can be solved according to theorem 2.

Theorem 2 If differential coefficient of $f(X)$ at X_k with downhill direction p is satisfactory with $\nabla^T f(X_k) p < 0$, then p is a downhill direction of $f(X)$ at X_k.

For the second problem, the step length of hill down can use the same step length of hill climbing for each step according to theorem 1.

The third problem can be solved according to theorem 3.

Theorem 3 Let $\{X \in R^n \mid f(X) \le f(X_0)\}$ be a limitary set, if downhill direction p of each iterative data X_k is satisfactory with: $\frac{\nabla^T f(X_k)}{\| \nabla f(X_k) \|} . \frac{p}{\| p \|} < -\rho$, where ρ is a constant. Then, series data $\{X_k\}_{k=1}^{\infty}$ produced by downhill algorithm have limitary sub-series, and each converges to points of zero gradients.

HD-DENLUE: hill down based density clustering
Input: smooth parameter σ, data set D
Output: cluster and edge set

(1) For each $X \in D$, compute its density function by

$$f_{Gauss}^D(X) = \sum_{j \in Near(X)} e^{-\frac{d(X, X_j)^2}{2\sigma^2}} \qquad \text{// by references[1-3]}$$

(2) Find density each attractor X_0; // by theory 1.

(3) For each density attractor X_0

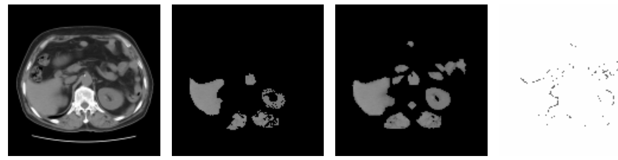
(4) Walk down hill from X_0 and compute all downhill directions p at point X_i; // by theory 2.

(5) Find next node $X_{i+1} = X_i + \lambda_i p$;// by theory 1.

(6) if $(\| \nabla f(X_{i+1}) \| > \varepsilon)$,all the data along downhill append cluster, else walk downhill stop and this data append edge;
// by theory 3.

3. EXPERIMENT

To investigate the practical relevance of our approach, we performed experiments on medical image data to cluster by DENCLUE and HD-DENCLUE. Experiment data are abdomen CT images (512×512 gray images) copied from a hospital. Both DENCLUE and HD-DENCLUE cluster medical image data can achieve image segmentation effects. Data of the same organs can be clustered into one group. Experiments results are like Figure1.



(a) (b) (c) (d)

Fig.1 Clusters results on CT image

(a.) Abdomen CT image (b) Clusters by DENCLUE with one threshold (c) Clusters by HD-DENCLUE with multi-thresholds (d) Image objects edge by HD-DENLUE

DENCLUE takes twice times time of HD-DENCLUE averagely from our experiments. From the above results, it is obvious that HD-DENCLUE has better segmentation quality than DENCLUE because of multi-threshold.

HD-DENCLUE has following several advantages: firstly, optimization-gradient variant step length reduces time overhead both hill climbing and down hill strategy; secondly, the hill-foot data are the multi-threshold for different clusters; thirdly, edges of each cluster can be gotten.

4. Acknowledgment

This paper is supported by Changshu Institution of Technology in China, Foundation for Young Teachers under Grant No. ky200657.

5. REFERENCES

[1] Alexander Hinneburg and Daniel A.Keim. A General Approach to Clustering in Large Databases with Noise. Knowledge and Information Systems(5):387-415,2003.

[2] Cunhua Li, Zhihui Sun, Yuqing Song. DENCLUE-M: boosting algorithm by mean approximation on grids. G.Dong et al. (Eds.): WAIM 2003.LNCS 2762, pp.202-213,2003.

[3] GAN Wen-yan; LI De-yi.Hierarchical Clustering based on Kernel Density Estimation. Acta Simulata Systematica Sinica.2004.2.(16): 302~305.

[4] Tasan Li, Weijie Peng, Zuoyi Zhou et al. Optimization compute theory and algorithm programming design. Housing of National University of Defense Technology, Changsha, China.2001.1.