An adaptive mean shift algorithm based on LSH

Yanbin Cui\textsuperscript{a}, Kui Cao\textsuperscript{b}, Gang Zheng\textsuperscript{c}, Fan Zhang\textsuperscript{ab}*

\textsuperscript{a} College of Computer and Information Engineering, Henan University, Kaifeng, Henan 475001, PR China
\textsuperscript{b} Institute of Image Processing and Pattern Recognition, Henan University, Kaifeng, Henan 475001, PR China
\textsuperscript{c} Kaifeng Electric Power Company, Kaifeng 475001, PR China

Abstract

The time complexity of the adaptive mean shift is related to the dimension of data and the number of iterations. The computational complexity will increase proportionally with the increase of the data dimension. An approximate neighborhood queries method is presented for the computation of high dimensional data, in which, the locality-sensitive hashing (LSH) is used to reduce the computational complexity of the adaptive mean shift algorithm. Experimental results show that the proposed algorithm can reduce the complexity of the adaptive mean shift algorithm and can produce a more accurate classification than the fixed bandwidth mean shift algorithm.

© 2011 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
Selection and/or peer-review under responsibility of [name organizer]

Keywords: Pattern Classification; Adaptive bandwidth; Adaptive mean-shift; Locality-sensitive hashing

1. Introduction

Mean shift-based clustering is a nonparametric kernel density estimation method, which is based on Parzen window technique, to find the local maximum of density estimation. Any point in the dataset can be used as the initial point in the mean shift procedure. The points which converge on the same modal point in the mean shift procedure are classified as the same pattern, automatically determine the number of the clusters, and describe the class boundaries. Fukunaga and Hostetler [1] proposed the mean shift procedure, and later Yizong Cheng [2] applied it to the clustering algorithm. Dorin Comaniciu [3] proposed the adaptive mean shift based on variable bandwidth density estimation, and applied it to image segmentation and clustering [4], but its speed is not ideal. In the iterations of the algorithm, the nearest neighbors of the prior approximate modal point are used to compute the new approximate modal point

* Corresponding author. Tel.: +86-0378-3883088; fax: +86-0378-3883088.
E-mail address: zhangfan@henu.edu.cn.
until the local maximum of density estimation is found. This means that the nearest neighbor search
directly determines the computing speed. Therefore, we present the adaptive mean shift-based clustering
algorithm based on locality-sensitive hashing, which store the points in the hash table and query the
neighbors quickly.

2. Adaptive Mean Shift

2.1. Mean shift procedure

Let \( X, i = 1, \ldots, n \) be a set of \( d \)-dimensional points in the space \( \mathbb{R}^d \). The density at point \( X \) can be
estimated by the Parzen window Kernel density estimator \([3]\)

\[
\hat{f}_K (x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^d} k \left( \frac{\| x - X_i \|}{h_i} \right)
\]

(1)

where function \( k \) is called the profile of the spherically symmetric kernel \( K \) with bounded support, that satisfies

\[
K(x) = c_{k,d} k \left( \|x\|^2 \right) > 0 \quad \|x\| \leq 1
\]

(2)

\( c_{k,d} \) is a normalization constant that makes \( K \) integrate to one. In (1), \( h \) is called the kernel bandwidth or
window size, and determines the range of influence of the kernel located in \( X_i \). In this work, the
multivariate Epanechnikov kernel is used:

\[
K(x) = \begin{cases} 
\frac{1}{2} c_{d}^{-1} (d + 2)(1 - x^T x) & \text{if } x^T x < 1 \\
0 & \text{otherwise}
\end{cases}
\]

(3)

Which is the Epanechnikov kernel that minimizes the mean integrated square error (MISE) between
the underlying probability density function of the data and the kernel density estimation?

The gradient of the density estimate are given by

\[
\nabla \hat{f}_K (x) = \nabla \hat{f}_K (x) = \frac{2}{n} \sum_{i=1}^{n} \frac{x - x_i}{h_i^d} k' \left( \frac{\| x - X_i \|}{h_i} \right)
\]

\[
= \frac{2}{n} \sum_{i=1}^{n} \frac{x - x}{h_i^d + 2} g \left( \frac{\| x - x_i \|}{h_i^d} \right)
\]

(4)

Where \( g(x) = -k'(x) \). The last part of (4) contains the adaptive mean shift vector

\[
M(x) = \left[ \frac{\sum_{i=1}^{n} x_i}{h_i^d + 2} g \left( \frac{\| x - x_i \|}{h_i} \right) \right] - x
\]

(5)

Suppose \( G(x) = c_{e,d} g(\|x\|) \). The mean shift vector in (4) can be written as,
\[ M(x) = C \frac{\hat{V}f_k(x)}{f_G(x)} \] (6)

From (6) we can see that the mean-shift vector is proportional to the normalized gradient of the density estimate computed for kernel. Therefore, the mean-shift vector points toward the direction of maximum density increase. This property is the basis of the mean-shift clustering algorithm.

2.2. Bandwidth selection

Data-driven bandwidth selection for multivariate data is a complex problem. The most often used method for local bandwidth adaptation takes the bandwidth proportional to the inverse of the square root of a first approximation of the local density [4].

The general steps used will be as follows:
1) Find a pilot estimate \( \hat{f}(x) \) that satisfies \( \hat{f}(x_i) > 0 \) for all \( i \).
2) Define bandwidth factor \( \lambda \) by
\[
\log \lambda = n^{-1} \sum \log \hat{f}(x_i) 
\] (7)
3) For each pixel \( x \), compute its adaptive bandwidth
\[
h(x) = h_0 \left[ \frac{\lambda}{\hat{f}(x)} \right]^{1/2} 
\] (8)

2.3. Adaptive mean shift-based clustering

Expression (6) shows that the mean-shift vector length is modulated by the inverse of the kernel density estimation with kernel \( G \). Therefore, as we move towards higher density regions along the mean-shift vector, its length gradually decreases. By repeating iteratively, the procedure progressively climbs to the nearest stationary point of the probability density function which is usually also one of its local maxima or ‘modes’ [5]. It means that in the stationary point \( M(x) \) approximately equal to 0. The stationary point is usually the local maximum of density estimation. The points which converge on the same stationary point in the mean shift procedure are classified as the same pattern, automatically determine the number of the clusters, and describe the class boundaries.

3. Locality-sensitive hashing

The main idea of locality-sensitive hashing (LSH) [6] is to create multiple hash tables with a particular constraint. In the hash tables, the points with greater similarity have little probability of conflict, and the points with little similarity have greater probability of conflict.

LSH algorithm can be divided into two parts: LSH hash table building and LSH query. The first step is to generate a hash table which holds the hash value and the corresponding data. LSH query is the process of searching the nearest neighbor in the hash table.

4. LSH based adaptive mean shift clustering

The bottleneck of mean shift in high dimensions is the need for a fast algorithm to perform neighborhood queries when computing (5). In expression (5), the mean shift iterations require a neighborhood query around the current location \( X \). The usual method is to scan the whole dataset and test
whether the kernel of the point \(x_i\) covers \(X\). Thus, for each mean computation the complexity is \(O(nd)\). Assuming that for every point in the dataset this operation is performed \(N\) times, the complexity of the mean shift algorithm is \(O(n^2dN)\) [7].

To improve the efficiency of the neighborhood queries, the following LSH data structure is constructed. In LSH structure, the neighbors are located in the same hash bucket. Therefore, the neighbors can be searched quickly [8]. LSH are represented by the following two procedures. Procedure 1 is used to partition the given data points into \(K/d + 1\) regions such that the points in each region have the same KD Boolean vector.

![Random points. MS Clustering. PE-AMS Clustering. LSH-PE-AMS Clustering.](image)

**Figure 1.** (a) Random points. (b) MS Clustering. (c) PE-AMS Clustering. (d) LSH-PE-AMS Clustering.

5. Experiment

Seven hundred and fifty random points in two-dimension space are used in the experiment, as shown in Figure 1 (a). The same dataset is clustered by the followed three algorithms: the fixed bandwidth mean shift (MS), the adaptive mean shift with bandwidth estimation (PE-AMS), the adaptive mean shift based on LSH with bandwidth estimation (LSH-PE-AMS). The results of the experiment are shown in Figure 1 (b), 1 (c), 1 (d). The bold points in the figures are the stationary points.

In the fixed-bandwidth mean shift algorithm, the fixed bandwidth \(h_0 = 200\). The points in the Figure 1 (b) are divided into 5 groups. As the bandwidth can’t be adjusted with the probability density, some points are classed into the isolated class. The next two algorithms avoid this mistake. The initial bandwidth for the bandwidth selection is 230. In the algorithm LSH-PE-AMS, \(L = 20, K = 15\). The
algorithm PE-AMS divides the points into three groups, and the algorithm LSH-PE-AMS divides the points into four groups. Experimental results show that the algorithm proposed by this article can provide more precise classification.

6. Conclusions

In this paper, the locality-sensitive hashing is used to query the neighborhood of the high dimensional data, and to reduce the computational complexity of the adaptive mean shift. Experimental results show that the proposed algorithm can reduce the complexity of the adaptive mean shift algorithm, and can produce a more accurate classification than the fixed bandwidth mean shift algorithm.

Acknowledgements

This research was supported by the National Natural Science Foundation of China grants No. 60873039, the Medical Science and Technology Research Project of Henan Province, grants No. 2011020114, and the Natural Science Foundation of Education Bureau of Henan Province, grants No. 2010B520003.

References

[1] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density functions with applications in pattern recognition. IEEE Trans. Information Theory, 1975, 21: 32-40.

[2] Y. Cheng. Mean Shift, Mode Seeking, and Clustering. IEEE Trans. Pattern Analysis and Machine Intelligence, 1995, 17: 790-799.

[3] D. Comaniciu, V. Ramesh, and P. Meer. The variable bandwidths mean shift and data-driven scale selection. IEEE Int. Conf. Computer Vision, 2001, 438–445.

[4] J. Jimenez-Alaniz, M. Pohl-Alfaro, V. Medina-Banuelos, and O. Yanez-Suarez. Segmenting brain MRI using adaptive mean shift. 28th International IEEE EMBS Annual Conference, 2006, 3114-3117.

[5] A. Mayer and H. Greenspan. An adaptive mean-shift framework for MRI brain segmentation. IEEE Transactions on Medical Imaging, 2009, 28: 1238–1250.

[6] B. Georgescu, I. Shimshoni, and P. Meer. Mean shift based clustering in high dimensions: A texture classification example. Proc. ICCV, 2003: 456–463.

[7] P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. STOC, 1998, 604–613.

[8] Fajie Li and Reinhard Klette. A Variant of Adaptive Mean Shift-Based Clustering. Lecture Notes in Computer Science, 2009, 1002–1009.