

Meanshift Clustering for DNA Microarray Analysis

Danny Barash
Department of Computer Science
Ben-Gurion University
84105 Beer-Sheva, Israel
dbarash@cs.bgu.ac.il

Dorin Comaniciu
Integrated Data Systems Department
Siemens Corporate Research
755 College Road East, Princeton, NJ 08450
dorin.comaniciu@scr.siemens.com

Abstract

Meanshift clustering is a well established algorithm that has been applied successfully in image processing and computer vision. Cluster centers are derived by local mode seeking identifying maxima in the normalized density of the data set. Recently, quantum clustering that highly resembles mean shift clustering has been proposed for analyzing microarray expression data. Quantum clustering is based on physical intuition derived from quantum mechanics. By an iterative process using a gradient descent procedure, the potential energy V belonging to the Hamiltonian of the time-independent Schrodinger equation develops minima that are identified with cluster centers. The analogies between the wavefunction in quantum clustering and the multivariate kernel density estimator in meanshift clustering are leading to closely related formulations. However, the approach towards the minima of the potential in quantum clustering needs to be performed unrelatedly to the formulation, by gradient descent steps. In contrast, in meanshift clustering the approach towards the maxima of the normalized density is performed by the meanshift vector that is derived by the formulation of the methodology. It points towards the direction of the maximum increase in the underlying density. Based on these observations, we propose implementing meanshift clustering to improve the efficiency of local mode seeking in analyzing expression data.

1. Background

Meanshift clustering offers a practical approach to non-parametric cluster analysis of large data sets. The meanshift procedure [1,2,3] has been developed for high-quality processing of low-level vision tasks. Recently, an analogous clustering method was suggested that is based on physical intuition derived from quantum mechanics, called quantum clustering [4]. Moreover, the quantum clustering method has been applied for the analysis of DNA microarray exper-

iments [5]. This suggests the capability of using the meanshift procedure for the analysis of gene expression data. Here, we present the analogy between mean-shift clustering and quantum clustering. The mathematical and physical approaches, formulated independently, are complementary and nicely illustrate the robustness of the method.

The quantum clustering method is described as follows. Given the Schrödinger equation

$$\left(-\frac{\sigma^2}{2}\nabla^2 + V(x)\right) = E\psi, \quad (1)$$

where

$$\psi(x) = \sum_i \exp\left(-\frac{(x-x_i)^2}{2\sigma^2}\right) \quad (2)$$

is the probability distribution for any set of data points x_i , one can solve for the potential energy V . In particular, requiring that $\min(V) = 0$, the energy E becomes:

$$E = -\min \frac{\sigma^2 \nabla^2 \psi}{\psi}, \quad (3)$$

where E can be shown to be positive, as well as being the lowest eigenvalue of the Hamiltonian, thus describing the ground state of the system. As explained in [4], $V(x)$ exhibits one or several local minima within the region of interest, identified as cluster centers. E has the role of setting the scale of which minima are observed. It is possible to locate the cluster centers by following a gradient descent procedure into potential minima, as shown in [4].

A mathematical derivation of the method was proposed in [1,2,3]. The approach belongs to the area of statistics and data analysis. Denoting by \mathbf{x}_i , $i = 1 \dots n$ a set of n data points in the d -dimensional space R^d , the multivariate kernel density estimator with normal kernel and a symmetric positive definite $d \times d$ bandwidth matrix \mathbf{H} , computed at the point \mathbf{x} is given by

$$\hat{f}(\mathbf{x}) = \frac{1}{n |2\pi\mathbf{H}|^{1/2}} \sum_{i=1}^n \exp\left(-\frac{1}{2}D^2(\mathbf{x}, \mathbf{x}_i, \mathbf{H})\right) \quad (4)$$

<i>Quantum Clustering</i>	<i>Mean Shift Clustering</i>
Schrödinger Equation	Currently Investigated
Wavefunction	Kernel Density Estimator
Energy (E)	Mean Shift Vector
Potential (V) Minima	Local Density Maxima

Table 1. Comparison between Mean Shift Clustering and Quantum Clustering

where

$$D^2(\mathbf{x}, \mathbf{x}_i, \mathbf{H}) \equiv (\mathbf{x} - \mathbf{x}_i)^\top \mathbf{H}^{-1} (\mathbf{x} - \mathbf{x}_i) \quad (5)$$

is the Mahalanobis distance from \mathbf{x} to \mathbf{x}_i ; the gradient of $\hat{f}(\mathbf{x})$ is calculated and the mean shift vector $\mathbf{m}(\mathbf{x})$ is obtained:

$$\mathbf{m}(\mathbf{x}) = \mathbf{H} \frac{\nabla \hat{f}(\mathbf{x})}{\hat{f}(\mathbf{x})}, \quad (6)$$

where it is defined by

$$\mathbf{m}(\mathbf{x}) \equiv \frac{\sum_{i=1}^n \mathbf{x}_i \exp\left(-\frac{1}{2} D^2(\mathbf{x}, \mathbf{x}_i, \mathbf{H})\right)}{\sum_{i=1}^n \exp\left(-\frac{1}{2} D^2(\mathbf{x}, \mathbf{x}_i, \mathbf{H})\right)} - \mathbf{x}. \quad (7)$$

The mean shift vector is an estimator of the normalized gradient of the underlying density. Since it points towards the direction of the maximum increase in the density, it can define a path leading to a local density maximum. The algorithm has been proposed in [6], generalized in [7], and achieves a high quality discontinuity-preserving filtering as demonstrated in [1,2,3]. The relationship between this algorithm and nonlinear diffusion filtering for image processing applications was analyzed in [8]. It is noted that the main difference between mean shift clustering [1] and quantum clustering [4] is that a gradient is used in the former, while a Laplacian is used in the latter as the competing force responsible for driving away the data distribution ψ from its minima. Using the mean-shift vector at each iteration for mode seeking is likely to improve the efficiency of the procedure. Table 1 summarizes the comparison between the two approaches.

2. Application in Biological Data

Quantum clustering was applied in [5] on data sets derived from three different microarray experiments in which the gene/sample classification is known. In addition to comparing the performance of mean shift clustering on these samples, the mean shift clustering can be tried on other biological data sets ranging from clustering after truncation [9,5] to other accumulated data sets. Figure 1 provides an experiment example (data set taken from

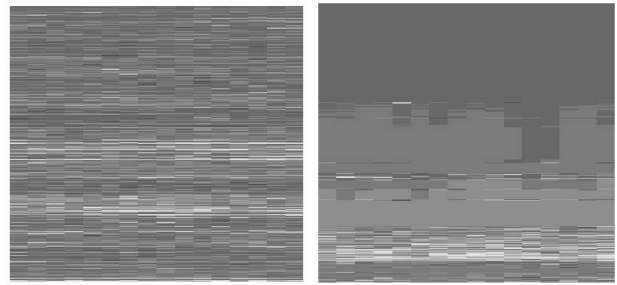


Figure 1. Mean shift clustering example of a data set taken from 24 tissues derived from a DNA microarray experiment.

<http://www.stat.berkeley.edu/users/terry/zarray>) as will be reported elsewhere.

3. Acknowledgements

We thank Dr. Bogdan Georgescu and Dr. Amit Chakraborty from Siemens Corporate Research for their assistance with the data set experiment. We also thank Drs. Vered Chalifa-Caspi and Avraham Melkman from Ben-Gurion University for helpful discussions.

References

- [1] D. Comaniciu and P. Meer, "Distribution Free Decomposition of Multivariate Data," *Patt. Anal. Appl.*, Vol. 2, pp. 22-30, 1999.
- [2] D. Comaniciu and P. Meer, "Mean Shift Analysis and Applications," *Proc. Seventh Int'l Conf. Computer Vision*, pp.1197-1203, 1999.
- [3] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis," *IEEE Trans Patt Anal Mach Intell*, Vol.24, pp.603-619, 2002.
- [4] D. Horn and A. Gottlieb, "Algorithm for Data Clustering in Pattern Recognition Problems Based on Quantum Mechanics", *Phys Rev Lett*, Vol.88, p. 018702, 2002.
- [5] D. Horn and I. Axel, "Novel Clustering Algorithm for Microarray Expression Data in a Truncated SVD Space," *Bioinformatics*, Vol.19, pp.1110-1115, 2003.
- [6] K. Fukunaga and L. Hostetler, "The Estimation of the Gradient of a Density Gradient, with Applications in Pattern Recognition," *IEEE Trans Inf Theory*, Vol.21, pp.32-40, 1975.
- [7] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," *IEEE Trans Patt Anal Mach Intell*, Vol.17, pp.790-799, 1995.
- [8] D. Barash and D. Comaniciu, "A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift," *Image Vision Comput.*, Vol.22, pp.73-81, 2004.
- [9] H. Ding, X. He, H. Zha, and H.D. Simon, "Adaptive dimension reduction for clustering high dimensional data," *Proc. Second IEEE Int'l Conf. Data Mining*, pp.1-7, 2002.