Mean Shift Algorithm with Heterogeneous Node Weights*

Ji Won Yoon and Simon P. Wilson
School of Computer Science and Statistics, Trinity College Dublin, Ireland
yoonj@tcd.ie, swilson@tcd.ie
January 20, 2010

Abstract
The conventional mean shift algorithm has been known to be sensitive to selecting a bandwidth. We present a robust mean shift algorithm with heterogeneous node weights that come from a geometric structure of a given data set. Before running MS procedure, we reconstruct un-normalized weights (a rough surface of data points) from the Delaunay Triangulation. The un-normalized weights help MS to avoid the problem of failing of misled mean shift vectors. As a result, we can obtain a more robust clustering result compared to the conventional mean shift algorithm. We also propose an alternative way to assign weights for large size datasets and noisy datasets.

1 Introduction
The mean shift (MS) algorithm [5, 6, 3] has been used for finding stationary points. Given a starting point, the MS procedure is iteratively implemented based on a MS vector that is calculated using a gradient estimate. It has been shown that the procedure is guaranteed to converge to a stationary point. A region that converges to the same stationary point defines the basin of attraction, in which the data points form one cluster. In this sense, the MS algorithm is a non-parametric statistical clustering method. It does not require prior knowledge of the number of clusters nor does it constrain the shape of the clusters. It has been widely used in image processing and computer vision applications.

A key to implementing MS is to find an optimal bandwidth for the kernel. MS may lead to misdirected mean vectors or undesirable local optima in a sparse region when an improper bandwidth is used. To reduce such difficulties, we therefore propose a new preprocessing scheme which introduces another type of weight which is named a heterogeneous node weight for each node. While the conventional weights of MS

*This version is a technical report at School of Computer Science and Statistics, Trinity College Dublin, Ireland, 2010
represents the function of the distance from the old mean, the node weights stand for the prior geometric information which each data point has over the whole data points. With this preprocessing, we can speed up the conventional MS and correct potentially misled MS vectors.

2 Mean Shift Algorithm

The mean shift (MS) algorithm, based on the Parzen window technique [8], is a non-parametric statistical method, widely applied to image processing and computer vision. Given \( n \) data points \( \{x_i\}_{i=1}^n \) in the \( d \)-dimensional space \( \mathbb{R}^d \), the multivariate kernel density estimator with a kernel \( K(\cdot) \) and a symmetric positive definite \( d \times d \) bandwidth matrix \( H \) is given by

\[
\hat{f}_{H,K}(x) = \frac{1}{n} \sum_{i=1}^n K_H(x-x_i),
\]

where the kernel is defined as \( K_H(x) = \frac{1}{|H|^{1/2}} K(H^{-1/2}x) \). If we assume independence and isotropy between dimensions, we have \( H = hI_d \). With the multivariate normal profile, we have the density estimator [2] by

\[
\hat{f}_{h,K}(x) = c_{k,d} n h^d \sum_{i=1}^n k\left(\frac{x-x_i}{h}\right)^2.
\]

A natural estimator of the gradient of \( f \) is the gradient of \( \hat{f}_{h,K}(x) \)

\[
\nabla f_{h,K}(x) = \frac{2c_{k,d}}{nh^d} \sum_{i=1}^n \left[ \sum_{j=1}^n g\left(\frac{x-x_i}{h}\right) \right] \times \left[ \sum_{i=1}^n x_i g\left(\frac{x-x_i}{h}\right) - x \right],
\]

where we denoted \( g(x) = -k'(x) \) and \( \sum_{i=1}^n g\left(\frac{\|x-x_i\|^2}{h}\right) \) is assumed to be positive.

Eq. (3) has two significant terms [3]. The first term of the product in Eq. (3) is proportional to the density estimate at \( x \) computed with a profile \( g(\cdot) \). The second term is the mean shift, the difference between the weighted mean and \( x \), and is defined as

\[
m_{h,G}(x) = \frac{\sum_{i=1}^n x_i g\left(\frac{\|x-x_i\|^2}{h}\right)}{\sum_{i=1}^n g\left(\frac{\|x-x_i\|^2}{h}\right)} - x.
\]

3 Node Weight embedded MS (WMS)

Performance of the MS algorithm depends on the selection of a bandwidth for the kernel. Too small a bandwidth is sensitive to noise, while too large a bandwidth leads to
over-smoothing. Much attention has been paid to optimal or adaptive bandwidth selection [5, 9, 4, 7]. In this paper, we propose a node weight embedded MS to compensate for the limitation of the MS by assigning the node weights. Consider a simple example in Fig. 1. The conventional MS is likely to have incorrect local optima as an MS vector but we can correct the MS vector by adding the weights of the nodes (data points) which are obtained by using geometric information of datasets in preprocessing time. In this paper, we use ”weighted MS” (WMS) instead of ”node weight embedded MS” for simplicity.

![Figure 1: The motivation of node weight embedded MS](image)

3.1 Delaunay Triangulation based Weight Mean Shift (DT-WMS)

Before applying Mean Shift (MS) procedure, we first build a Delaunay Triangulation (DT) to find the neighbours of data points (nodes) as shown in Fig. 2-(a). Next step is to assign the weights into nodes. In other words, we assign heterogeneous weights into individual nodes by computing the inverse of the mean of the distances between the node’s neighbours. Suppose $w_i$ is the weight of a node $i$ and it is calculated as

$$w_i = \left\{ \frac{\sum_{j \in ne(i)} l(i,j)}{|ne(i)|} \right\}^{-1}$$

where $ne(i)$ denotes the neighbours of node $i$ and $l(i,j)$ is the length between a node $i$ and a node $j$. Therefore, Eq. (4) for MS is rewritten with weights:

$$m_{h,G}(x) = \frac{\sum_{i=1}^{n} x_i w_i g \left( \frac{|x-x_i|}{h} \right)}{\sum_{i=1}^{n} w_i g \left( \frac{\|x-x_i\|}{h} \right)} - x.$$

In this paper, we name this modified MS algorithm as Delaunay Triangulation based Weight Mean Shift (DT-WMS).
3.2 K nearest neighbours based Weight Mean Shift (K-WMS)

If there are a large number of data points, then the Delaunay Triangulation (DT) can be a time consuming task due to its running time complexity, $O(n \log n)$ for $n$ data points. As an alternative to DT, we can assign the weights by considering $K$ nearest neighbours of each node for the neighbours. (Note that this operation also has $O(n \log n)$ time but it is much simpler to implement.) We name this approach by $K$ nearest neighbours based Weight Mean Shift (K-WMS).

4 Simulation Results

4.1 Synthetic Data Sets

We tested our proposed approaches with two open synthetic datasets ($A$ and $B$), which have 15 clusters with 5000 data points [1]. Fig. 3 -(a) and (c) show two dimensional views of raw data. Fig. 3 -(b) and (d) are their weighted graphs via Delaunay Triangulation respectively. As we can see in Fig. 3 -(b) and (d), denser areas have more weights (larger red spots) than sparse areas.

We compare the performance of the three methods with variant bandwidths ($h$) and variant number of nearest neighbours ($K$) as shown in Tables 1 for generalized evaluation. From this result, we find several important factors: (1) Weighted MS (WMS) is superior to conventional MS (MS), (2) DT-WMS efficiently works compared to other two approaches in case clusters are well separated, (3) K-WMS works better than conventional MS and DT-WMS in case clusters are overlapped a lot and (4) K-WMS is rather sensitive to the selection of $K$.

4.2 Robustness

We also test the robustness of our improved Mean Shift algorithms against the noise. Let $\text{Var}(x) = \mathbf{V} \Lambda \mathbf{V}^T$ be the eigendecomposition of $\text{Var}(x)$. With variant $\tau$ we added noise by

$$x^* = x + \tau \mathbf{V} \Lambda^{1/2} \mathbf{z}$$

(7)
Figure 3: Synthetic datasets (15 clusters)

Table 1: The number of clusters with variant bandwidths

<table>
<thead>
<tr>
<th></th>
<th>h</th>
<th>MS</th>
<th>K-WMS K=10</th>
<th>K-WMS K=5</th>
<th>K-WMS K=1</th>
<th>DT -WMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4076</td>
<td>1192</td>
<td>1122</td>
<td>1097</td>
<td>1090</td>
<td>1075</td>
</tr>
<tr>
<td></td>
<td>11738</td>
<td>148</td>
<td>114</td>
<td>120</td>
<td>129</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>15771</td>
<td>52</td>
<td>43</td>
<td>45</td>
<td>49</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>19804</td>
<td>23</td>
<td>20</td>
<td>20</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>27869</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>B</td>
<td>4715</td>
<td>1274</td>
<td>1207</td>
<td>1186</td>
<td>1155</td>
<td>1149</td>
</tr>
<tr>
<td></td>
<td>13142</td>
<td>204</td>
<td>161</td>
<td>167</td>
<td>179</td>
<td>164</td>
</tr>
<tr>
<td></td>
<td>26449</td>
<td>32</td>
<td>28</td>
<td>26</td>
<td>29</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>30884</td>
<td>23</td>
<td>23</td>
<td>22</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>48626</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

where $z \sim N(\cdot; 0, 1)$ and $\tau \geq 0$. Fig. 4 shows the clustered results with $\tau = 0.15$. In this figure, we find that K-WMS has closer number of clusters to underlying ground.
truth. (Here, we also used $K_i$-WMS for "$K_{i}$".)

(a) Noisy data  
(b) MS  
(c) $K_{10}$-WMS  

(d) $K_5$-WMS  
(e) $K_1$-WMS  
(f) DT-WMS

Figure 4: Clustered data of synthetic data

Table 2 summaries the results of the Mean shift algorithms in noisy data with variant $\tau$. Here for simplicity, the bandwidth is automatically calculated by the mean ($\mu_d = E(d)$) and covariance ($\Sigma_d = V(d)$) of the shortest distances ($d$) of each data point between other points: $h = \mu_d + 4\sqrt{\Sigma_d}$.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>MS</th>
<th>$K$-WMS</th>
<th>DT-WMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$K=10$</td>
<td>$K=5$</td>
</tr>
<tr>
<td>0</td>
<td>99.2</td>
<td>99.4</td>
<td>99.2</td>
</tr>
<tr>
<td>0.05</td>
<td>99.1</td>
<td>99.1</td>
<td>99.2</td>
</tr>
<tr>
<td>0.1</td>
<td>98.2</td>
<td>98.3</td>
<td>98.2</td>
</tr>
<tr>
<td>0.15</td>
<td>95.5</td>
<td>95.6</td>
<td>95.1</td>
</tr>
<tr>
<td>0.2</td>
<td>78.5</td>
<td>82.7</td>
<td>80.6</td>
</tr>
</tbody>
</table>

4.3 Image Segmentation

Applying the MS and WMS algorithms to images requires a preprocessing step. In general the spaces $L * u * v$ and $L * a * b$ are used for image segmentation and filtering.
which are designed to best approximate perceptually uniform colour space. We used the $L^*u^*v^*$ colour spaces. In addition, the colour level or spectral information is represented in the spatial domain so the dimension of the each sample point $x_i$ for a colour image is five ($d = 5$): three for the range domain and two for the spatial domain.

The performances of the image segmentation are compared only with K-WMS ($K = 5$) and conventional MS since DT-WMS cannot be applied for large size dataset and we assumed the real image has many overlapped data points. As we can see in Fig. 5, K-WMS ($|C| = 115$) efficiently segments the image more than conventional MS ($|C| = 152$).

5 Conclusions

We showed Mean Shift algorithm is improved by adapting heterogeneous node weights into data points. This works because global prior information is provided to the mean shift mechanism by using geometric prior such as Delaunay Triangulation. This algorithm speeds up the conventional MS and corrects misled mean shift. We also showed a practical but alternative way to assign the weights by searching for $K$ nearest neighbours for large size datasets.

We are currently developing incremental weighted mean shift algorithms and computationally cheaper algorithms for large size datasets.

Acknowledgement

This work is supported by STATICA project which is funded by the Principal Investigator programme of Science Foundation Ireland, Grant number 08/IN.1/I1879.

References


Figure 5: Segmentation of an image

