Chapter 6

Preserving Global and Local Knowledge in Multi-Channel Temporal Data Mining using Wavelet Transform

Most research in Knowledge Mining deal with the basic models like clustering, classification, regression, association rule mining and so on. In the process of quest for knowledge most of the knowledge mining algorithms end up in generating global knowledge while losing focus on the local knowledge. This happens often due to two reasons. First reason is due to dimensionality reduction. The problem of dimensionality reduction has been viewed as the reduction of features to the maximum extent possible while being able to retain the information conveyed by the data set. But most of the dimensionality reduction techniques reduce the dimensions keeping only the retention of global knowledge in mind while compromising with the loss of local knowledge. Second reason is due to optimized feature selection for making a global classification while not being bothered about the intra class relationship. In this chapter, methodologies using wavelet transform for overcoming the loss of local knowledge along the process of mining is proposed. First a discrete wavelet transform based multi resolution approach is proposed to capture the local knowledge along the process of dimensionality reduction. The approach proposes a strategy to represent the global knowledge with few number of wavelet coefficients. A comparison with PCA has also been made here, which strongly supports the proposed technique using discrete wavelet transform. Through the proposed technique more number of local knowledge and no/minimum misleading knowledge is generated. Secondly, continuous wavelet transform based multiresolution approach is proposed for knowledge mining using histogram features. Both these approaches are demonstrated from a Multi-Channel Temporal Data Perspective. In the former
approach a methodology for generating/mining knowledge packets from individual channels and integrating this knowledge for generating a comprehensive knowledge base is proposed. In the later strategy, summing the distances between the objects when observed from different sources does the integration and then the knowledge is mined from overall distance matrix to obtain comprehensive knowledge base. A novel measure by name 'stability factor' is introduced to derive stable clusters from agglomerative clusters. These foundational techniques are illustrated with a multi-channel 80X data set and their performance is demonstrated through fish contour SQUID database.

6.1. Introduction

The curse of dimensionality refers to the exponential growth of hyper volumes as a function of dimensionality. All problems become harder as the dimensionality increases. In other words "more the amount of information or details that exist about the samples poorer will be the performance of the classifier". So it is always required to extract optimum amount of details that is necessary for mining or extracting the knowledge. A straightforward method like principal component analysis is commonly used for dimensionality reduction. Most research on dimensionality reduction deal with it only from the point of memory reduction while being able to preserve the global knowledge. Extraction of global knowledge may be appreciable by the non-critical pattern recognition applications. But when it comes to critical pattern recognition applications like signature verification in a bank cheque, the extraction of local knowledge may become highly crucial. Critical PR applications would need the knowledge extracted from both the finer details and also from the coarser details. The finer details would express the local knowledge and the coarser details would express the global knowledge. Thus dimensionality reduction should not be achieved by totally turning a blind eye towards the higher dimensions and an open eye only to the knowledge extracted in the lower dimensions. Dimensionality reduction has to be a process through which we are able to extract the knowledge all along from the higher dimensions to the lower dimensions rather than reduction of dimensions to the maximum extent and looking for some left out global knowledge.
In the first part of the chapter (Section 6.3) a methodology for extracting local and global knowledge from multi-channel temporal data set along the process of dimensionality reduction using discrete wavelet transform is proposed. The localization property of wavelet transform has the capability of extracting the finer details from the temporal signals. These finer detail features are processed and approximated thus extracting the knowledge at all levels of wavelet decomposition of the temporal signals. Dimensionality reduction can be achieved by representing the details with few number of wavelet coefficients by preserving the global knowledge. A two stage strategy is proposed for knowledge mining (1) Extraction of knowledge packets from individual channels along the process of dimensionality reduction through wavelet transform (2) Fusion of knowledge packets from multiple channels for comprehensive knowledge mining. For experimentations, the multi-channel signal is obtained by sampling the 80X characters along four direction as illustrated in section 2.2.5 of chapter 2. It is viewed as four signals being recorded by four sensors or four signals obtained from four different sources. The knowledge packets are extracted from the clusters obtained by a linkage algorithm applied on the data set generated at different wavelet decomposition level along each channel. Then these interesting knowledge packets obtained along multiple-channels or from multiple sources are fused to generate comprehensive knowledge base, which gives valid, novel, and interesting information.

Knowledge mining refers to the overall process of extracting high-level knowledge from low-level data in the context of huge multi dimensional databases. In the process of quest for some knowledge, most of the knowledge mining algorithms end up in generating global knowledge while losing focus on the local knowledge. These local knowledges may be very important in critical applications like medicine, signature verification, defence and so on. By the phrase 'local knowledge' we like to emphasize that the finer information of a sample/object/record in a dataset/database can bring in close association between a set of objects, which may turn to be very interesting. Figure 6.1 gives a better illustration of our statement on local knowledge.
In figure 6.1(a) we have extracted only 8 features whereas in figure 6.1(b) we collect the features along the horizontal axis moving from bottom to top along the vertical axis obtaining one set of 200 features for a 200 x 200 image. Likewise the features along the four directions (left, top, right and bottom) can be extracted resulting in around 800 features around the object. As mentioned earlier the data set thus extracted can be viewed as a multi-channel temporal data with 4 channels. Bellman Ford’s theory on ‘Curse of Dimensionality’ emphasizes that all problems become harder as the dimensionality increases. Thus most feature extraction techniques always choose a set of optimal features like the one shown in figure 6.1(a) to classify the samples. But in this process of choosing an optimal set of features, the finer information about the object is lost and thus the mining procedure ends up resulting only with few global knowledge. Consider a simple problem of generating association between a set of samples representing numeral ‘8’ and ‘0’ but of different fonts. With the kind of features extracted as shown in figure 6.1(a) it is possible to classify the samples into two groups with one cluster containing all ‘8’s and another cluster containing all ‘0’s. But when the question regarding the association of these samples within a cluster arises, it lacks finer information about the fonts in the feature set which restricts us from inferring finer/local knowledge in the sample set. For this, the feature extraction of the kind shown in figure 6.1(b) is required where the feature set carries information about the entire object. But processing the whole set of features (nearly 800 features) would be a non-trivial process. This is where a procedure to process this entire feature set by keeping in mind the available computational power is required. It is here, the second part of this chapter would make a novel proposal to perform knowledge mining by keeping the finer information of the samples intact throughout the mining process. The temporal
signals are subjected to continuous wavelet transformation. Histogram features are extracted from continuous wavelet transform coefficients and subsequently the knowledge is mined using the AB distance measure and the stability factor.

6.2. Sample Set and Data Set

The 8OX sample set that is used for illustration is shown in figure 6.2. The corresponding multi-channel temporal signals obtained by sampling the characters along four directions are depicted in figure 6.3.

<table>
<thead>
<tr>
<th>Amer</th>
<th>Ariel</th>
<th>Arrus</th>
<th>Bandit</th>
<th>Bauhes</th>
<th>Tnr</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
</tbody>
</table>

Fig 6.2 Sample data set with font names and the index

6.3. Knowledge mining using Discrete Wavelet transforms

In this section a knowledge mining strategy for discovering local knowledge and global knowledge in a multi-channel temporal data set is proposed. The procedure is based on discrete wavelet transform. Section 6.3.1 briefs the interesting properties of DWT with respect to knowledge mining. Section 6.3.2 and section 6.3.3 describe the multi-resolution knowledge mining process. Section 6.3.4 discusses the various inferences.
6.3.1 DWT for Dimensionality Reduction

Consider the set of all $N$ dimensional, real valued vectors of the form $x = [x_1, x_2...x_N]$. This set forms an $N$-dimensional linear vector space and there exist $N$ linearly independent basis vectors $a_1, a_2, ..., a_N$ such that any vector in this space can be expressed as a linear combination of these basis vectors. In other words, there exist unique scalars $a_1, a_2, ..., a_N$ such that $x=a_1a_1+a_2a_2+...+a_Na_N$. Let us call this $N$-dimensional vector space $V_N$. We next look at approximating vectors in $N$-dimensional space by vectors in a subspace of a lower dimension. Suppose we generate all linear combination of just $N-1$ of the $N$ basis vectors, say $a_1, a_2, ..., a_{N-1}$. The set of all such linear combinations is also a vector space. It is of $N-1$ dimension and every vector in it also lies in $V_N$. It is thus a vector subspace of $V_N$. Let us call this $V_{N-1}$. By continuing to drop the last basis vector at each step, we can similarly construct subspaces $V_{N-2}, ..., V_1$ of dimension $N-2, ..., 1$ respectively. The
subspaces $V_1$ has as its basis the single vector $a_1$. These vector spaces form a nested sequence of subspaces, $V_1 \subset V_2 \subset \ldots \subset V_N$.

Suppose we want to approximate a vector $x$ in $V_N$ by a vector in $V_{N-1}$. The best possible approximation is the minimum mean squared error which is made by choosing that vector (say $\zeta_{N-1}$) in $V_{N-1}$ for which the length of the error vector $e_{N-1} = x - \zeta_{N-1}$ is minimized.

Thus $\zeta_{N-1}, \zeta_{N-2}, \ldots, \zeta_1$ are the approximation of the vector $x$ at different levels and $e_{N-1}, e_{N-2}, \ldots, e_1$ can be seen as the amount of detail that is lost in $x$ in going to its approximation $\zeta_{N-1}$ because $x - e_{N-1} = \zeta_{N-1}$. Thus the original vector $x$ can be obtained through the following equation

$$x = e_{N-1} + e_{N-2} + \ldots + e_1 + \zeta_1$$

where $\zeta_1$ is the coarsest of approximation.

In wavelet transform the approximation from $N$ dimensions to the lower dimension generates the approximation coefficients and the error vectors generate the corresponding detail coefficients. Thus the entire process of multi resolution knowledge mining involves analyzing the error vectors at different vector spaces $V_k$ where $k < N$ for being capable of representing a stable knowledge (a cluster of a set of objects) is one which gets generated and remains undisturbed/uninfluenced by other objects in the data space for sufficient period]. In the process of dimensionality reduction we search for the appropriate error vector $e_k$ in the lowest $K$ dimensional space where $K < N$ that would hold the optimum or enough detail for classifier and also extract the knowledge expressed by the error vectors at different dimensions.

6.3.2 Multiresolution Knowledge Miner

The motivation of Multiresolution analysis is to use a sequence of embedded subspaces to approximate $L^2(R)$ so that people can choose a proper subspace for a specific application task to get a balance between accuracy and efficiency. Higher subspaces can contribute accuracy or finer details but waste computing resources. On the other hand lower subspaces would provide approximate or global knowledge at lesser computational cost. Wavelet is related to Multi Resolution Analysis (MRA) because of the scaling function $\phi$ which easily generates a sequence which can
provide a simple multiresolution analysis. A direct application of multiresolution analysis is the fast discrete wavelet transform algorithm called the pyramidal algorithm. The idea is to progressively smooth the data using an iterative procedure and keep the detail along the way. After the first decomposition, the data are divided into two parts: one is of average information (projection in the scaling space $V_2$) and the other is of detail information (projection in the wavelet space $W_2$). Then similar decomposition on the data in $V_2$ is repeated to obtain the projected data in $V_3$ and $W_3$ etc. The detail information otherwise termed as the surprise information is used as the feature values for mining the knowledge. In the proposed method the knowledge is mined from the finest detail level to the coarsest detail level. Before we get on to the technique let the term *Stability factor* be defined.

**Stability factor:**

The stability factor of a node introduced in the above algorithm is defined as the difference between the length of span taken for a group of samples to cluster and the length of span it remains stable without allowing any other samples to join the cluster. Consider the following simple dendrogram

![Simple Dendrogram](image)

**Fig 6.4 A simple dendrogram**

*Stability factor ($S_f$) of the group $\{a_1, a_2, a_3\}$

$S_f = l_s - l_t$
where \( L \) is the length of span/stretch taken for a group to cluster and \( L_s \) is the stretch/length of span for which the group is stable without allowing any other sample to join the cluster.

\[
\text{if } S_f > 0 \text{ then cluster } \{a_1, a_2, a_3\} \text{ is a stable knowledge} \\
\text{else cluster } \{a_1, a_2, a_3\} \text{ is an unstable knowledge}
\]

The algorithm for generating stable knowledge packet is given illustrated as below:

**Step 1:** The temporal signals \( S_1, S_2, S_3, \) and \( S_4 \) along all the four directions of the characters are extracted.

**Step 2:** For each of the signal \( S_x \) for \( x = 1 \) to 4 discrete wavelet transform is applied.

**Step 3:** For each decomposition level from 1 to maximum level, considering the detail wavelet coefficients as the features an agglomerative clustering algorithm (linkage algorithm) is applied.

**Step 4:** The stability factor is calculated for all the nodes except the leaf nodes in the dendrogram obtained on applying the linkage algorithm.

**Step 5:** If the stability factor is greater than 0 label the ‘group of samples’/cluster formed under the node as an interesting node else it is a non-interesting node.

**Step 6:** The knowledge generated along the four directions is fused to obtain the comprehensive knowledge about the sample set.

The stability factor of a knowledge packet is also the measure of interestingness of the knowledge packet. The interesting measure is denoted by \( \text{Int}_M \). Few of the dendrograms obtained after applying linkage clustering for the detail coefficients obtained from wavelet transform of the spatial signal are shown below. The temporal signal obtained along direction 1 and direction 2 are considered for mining the knowledge about ‘8OX’ dataset.

![Fig 6.5 D1, D2, D3, D4 represent the four directions along which the temporal signals were recorded](image)
Fig 6.6(a) Cluster formations with the first level decomposition detailed coefficients

Fig 6.6(b) Cluster formations with the second level decomposition detailed coefficients
Fig 6.6(c) Cluster formations with the eighth level decomposition coefficients

Table 6.1

<table>
<thead>
<tr>
<th>Set No</th>
<th>Knowledge Packets</th>
<th>Int_M</th>
<th>Level in which Knowledge found</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1</td>
<td>13, 18</td>
<td>9.19</td>
<td>1</td>
</tr>
<tr>
<td>K2</td>
<td>13, 18</td>
<td>43.83</td>
<td>2</td>
</tr>
<tr>
<td>K3</td>
<td>13, 18</td>
<td>42.15</td>
<td>3</td>
</tr>
<tr>
<td>K4</td>
<td>13, 18</td>
<td>35.81</td>
<td>4</td>
</tr>
<tr>
<td>K5</td>
<td>10, 11, 12</td>
<td>5.73</td>
<td>4</td>
</tr>
<tr>
<td>K6</td>
<td>2, 4</td>
<td>5.63</td>
<td>4</td>
</tr>
<tr>
<td>K7</td>
<td>10, 11, 12</td>
<td>0.44</td>
<td>5</td>
</tr>
<tr>
<td>K8</td>
<td>13, 14, 15, 16, 18</td>
<td>47.49</td>
<td>5</td>
</tr>
<tr>
<td>K9</td>
<td>7, 8, 9, 10, 11, 12</td>
<td>64.00</td>
<td>6</td>
</tr>
<tr>
<td>K10</td>
<td>1, 2, 3, 4, 5, 6</td>
<td>23.19</td>
<td>6</td>
</tr>
<tr>
<td>K11</td>
<td>13, 14, 15, 16, 17, 18</td>
<td>28.75</td>
<td>6</td>
</tr>
<tr>
<td>K12</td>
<td>8, 10</td>
<td>18.40</td>
<td>7</td>
</tr>
<tr>
<td>K13</td>
<td>11, 12</td>
<td>0.52</td>
<td>7</td>
</tr>
<tr>
<td>K14</td>
<td>16, 18</td>
<td>5.91</td>
<td>7</td>
</tr>
<tr>
<td>K15</td>
<td>7, 8, 9, 10, 11, 12</td>
<td>10.48</td>
<td>7</td>
</tr>
<tr>
<td>K16</td>
<td>6, 13, 14, 15, 16, 17, 18</td>
<td>3.14</td>
<td>7</td>
</tr>
<tr>
<td>K17</td>
<td>4, 5</td>
<td>40.1</td>
<td>8</td>
</tr>
<tr>
<td>K18</td>
<td>1, 2, 3</td>
<td>29.82</td>
<td>8</td>
</tr>
<tr>
<td>K19</td>
<td>17, 18</td>
<td>25.81</td>
<td>8</td>
</tr>
<tr>
<td>K20</td>
<td>7, 8, 9, 10, 11, 12</td>
<td>69.97</td>
<td>8</td>
</tr>
<tr>
<td>K21</td>
<td>1, 2, 3, 4, 5</td>
<td>23.91</td>
<td>8</td>
</tr>
<tr>
<td>K22</td>
<td>13, 14, 15, 16, 17, 18</td>
<td>348.59</td>
<td>8</td>
</tr>
<tr>
<td>K23</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12</td>
<td>242.60</td>
<td>8</td>
</tr>
</tbody>
</table>
6.3.3 Fusion of Knowledge from different sources

The knowledge obtained from the four channels are integrated by a set of procedures which involves tree building, node shedding, node collection and then Cross intersection. The following algorithm is executed for integrating the interesting details.

**Binary tree building for the set of knowledge obtained along each channel**

**Step 1:** Create a root node with all the samples present in the node initially. Let the set of elements be denoted as T.

**Step 2:** Sort the knowledge obtained based on interesting measure

**Step 3:** Create a left child node containing the elements with the knowledge of highest interesting measure. Assume the set of elements in left child node as A.

**Step 4:** Create a right child node with the remaining elements i.e. with the elements T – A.
Step 5: For each of the leaf nodes from left to right follow the steps 2-4 by considering the knowledge one after the other. If the subset of the leaf nodes cannot be created for any knowledge then stop the algorithm.

The tree building process is illustrated below with the knowledge obtained along direction 1

Step 1:
Create a node with all the elements.

\[ T = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18\} \]

Fig 6.7(a)

Step 2:
Table 1 is sorted based on their interesting measure.

Step 3:
Create a left child node containing the elements with the knowledge of highest interesting measure. As given in table 1 the knowledge with highest interesting measure is with the set of elements 13-18 with an interesting measure of 348.59.

\[ A = \{13, 14, 15, 16, 17, 18\} \]

Fig 6.7(b)

Step 4:
Create a right child node with the remaining elements i.e. with the elements \( T - A \).

Now the right child will consist of

\[ T - A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\} \]
Step 5:
For each of the leaf node repeat the steps 2-4 by considering the knowledge in ascending order. If the entire objects in a node comprise of knowledge then the left child of the node would carry the entire set of objects while the right child would carry a null set. The final tree obtained looks like as shown in figure 6.8. Nodes B, B1 and B2 of figure 6.8 illustrate the case where node B1 comprises the entire set of objects in B which is knowledge and node B2 is a null set.

Knowledge filtering
The Knowledge filtering phase involves node shedding and node collection. It can be observed that all the nodes forming the left child are knowledge obtained along a direction and the right nodes consist of the remaining elements. So we shed the right nodes and collect the left nodes. Thus the filtered knowledge obtained along channel-I (direction 1) is as illustrated below:

KD-I = \{\{13, 14, 15, 16, 17, 18\}, \{13, 14, 15, 16, 18\}, \{7, 8, 9, 10, 11, 12\}, \{13, 18\}, \{8, 10\}, \{4, 5\}, \{11, 12\}, \{1, 2, 3\}\};

Likewise the filtered knowledge obtained along channel II (direction-II) is:

KD-II = \{\{4, 5\}, \{7, 8, 9, 10, 11, 12\}, \{10, 11\}, \{1, 3\}, \{7, 9, 12\}, \{13, 14, 15, 16, 17, 18\}, \{13, 14\}, \{15, 18\}\};

T = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18\}
A= \{13, 14, 15, 16, 17, 18\}
B= T - A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}
\[ B_1 = \{1,2,3,4,5,6,7,8,9,10,11,12\} \]
\[ B_2 = B - B_1 = \{\} \]
\[ C = \{13, 14, 15, 16, 18\} \]
\[ D = A - C = \{17\} \]
\[ E = \{7, 8, 9, 10, 11, 12\} \]
\[ F = B - E = \{1, 2, 3, 4, 5, 6\} \]
\[ G = \{13, 18\} \]
\[ H = C - G = \{14, 15, 16\} \]
\[ I = \{8, 10\} \]
\[ J = E - I = \{7, 9, 11, 12\} \]
\[ K = \{4, 5\} \]
\[ L = E - I = \{1, 2, 3, 6\} \]
\[ M = \{11, 12\} \]
\[ N = J - M = \{7, 9\} \]
\[ O = \{1, 2, 3\} \]
\[ P = L - O = \{6\} \]

**Cross Intersection**

Cross Intersection of two sets A and B represented by CI (A, B) can be defined as follows

Given two sets A and B with a set of sets say \( A = \{\{a_1\}, \{a_2\}, \ldots, \{a_N\}\} \) and \( B = \{\{b_1\}, \{b_2\}, \ldots, \{b_N\}\} \) then their cross intersection is given by

\[
\text{CI}(A, B) = \{\{a_1 \cap b_1\}, \{a_1 \cap b_2\}, \ldots, \{a_1 \cap b_N\}, \{a_2 \cap b_1\}, \{a_2 \cap b_2\}, \ldots, \{a_2 \cap b_N\}, \ldots, \{a_N \cap b_1\}, \{a_N \cap b_2\}, \ldots, \{a_N \cap b_N\}\}
\]

The null sets in CI(A, B) can be eliminated. Thus the integration of the two sets of knowledge KD-I and KD-II are done using cross intersection

**Integrated / Comprehensive Knowledge**

\[
\text{IKD} = \text{CI}(\text{KD-I}, \text{KD-II})
\]

Fig 6.8 Knowledge Filter Tree
After the process, the comprehensive stable knowledge obtained is
IKD = \{\{13, 14, 15, 16, 17, 18\}, \{7, 8, 9, 10, 11, 12\}, \{4, 5\}, \{1, 3\}, \{10, 11\}, \{7, 9, 12\}, \{13, 14\}, \{15, 18\}, \{13, 14, 15, 16, 18\}, \{13, 18\}, \{8, 10\}\}

6.3.4 Discussions
Discussion on IKD
IKD(1) holds all the samples of character ‘X’.
IKD(2) holds all the samples of character ‘O’.
IKD(3) brings together samples 4 and 5 which is a local knowledge within IKD(1) because the shape of the character ‘8’ of these two samples look very much similar and are dissimilar from the other samples holding the character ‘8’.
Likewise IKD(4) are similar looking ‘8’, IKD(5) are similar looking ‘O’, IKD(6) are similar looking ‘O’, IKD(8) and IKD(10) are similar looking ‘X’, IKD(11) are similar looking ‘O’.
IKD(9) is also a very interesting knowledge because it holds all the samples holding the character ‘X’ except sample 17 which is of a different shaped ‘X’.
Apart from IKD there can also be some knowledge extracted from the knowledge tree by collecting the nodes with a single element. These single element nodes give very interesting and surprising information. If the tree drawn with the knowledge extracted along direction I is observed node ‘D’ and node ‘P’ consist of single sample elements 17 and 6 respectively. If the sample 6 is observed carefully it is the only sample with both the upper and the lower portion of ‘8’ of same size whereas in all other samples holding ‘8’ the upper part is slightly smaller than the lower part. Likewise sample 17 is the only ‘X’ with curved shape whereas all other samples holding ‘X’ maintains linearity along its shapes.
The above discussion establishes that the knowledge extracted by using a multi-resolution approach using wavelets are very interesting and are also powerful enough to extract both the global and local knowledge along the process of dimensionality reduction.
Comparison with PCA
Dimensionality reduction achieved through PCA is an irreversible process while the one we have proposed using wavelet transform can be made reversible. Also PCA is capable of retaining only the global knowledge while losing all components pertaining to the local knowledge. On achieving dimensionality reduction through PCA and then applying the clustering technique as before similar samples are never found to come together locally within a cluster and some of the local knowledge packets obtained are totally unacceptable due to misleading information generated. The stable knowledge obtained along direction I and direction II obtained from the first few principal component is given below along with the dendrograms shown in figure 6.9.

Fig 6.9 Cluster formations with the first three Principal components obtained from the data set recorded along direction-1
Fig 6.10 Cluster formations with the first three Principal components obtained from the data set recorded along direction-2

KD-I $\text{PCA}_1 = \{\{13, 14, 15, 16, 17, 18\}, \{13, 14\}, \{7, 8, 9, 10, 11, 12\}, \{7, 9\}, \{2, 3\}\}$

KD-II $\text{PCA}_1 = \{\{13, 14, 15, 16, 17, 18\}, \{15, 16, 17\}, \{16, 17\}, \{1, 3, 6, 7, 8, 9, 10, 11, 12\}, \{3, 9, 12, 10\}, \{4, 5\}\}$

The knowledge obtained along both the direction are integrated though cross intersection

IKD $\text{PCA} = \text{CI} (\text{KD-I PCA}_1, \text{KD-II PCA}_1)$

$= \{\{13, 14, 15, 16, 17, 18\}, \{13, 14\}, \{13, 14\}, \{7, 8, 9, 10, 11, 12\}, \{7, 9\}, \{15, 16, 17\}, \{4, 5\}\}$

**Discussion on IKD $\text{PCA}$**

IKD $\text{PCA} (1)$ is a global knowledge

IKD $\text{PCA} (2)$ is not an interesting pattern which can be observed from the shapes of character ‘8’ in samples 13, 14, 16. It is a misleading knowledge

IKD $\text{PCA} (3)$ is a local knowledge

IKD $\text{PCA} (4)$ is a global knowledge

IKD $\text{PCA} (5)$ is a local knowledge

IKD $\text{PCA} (6)$ is a misleading knowledge

IKD $\text{PCA} (7)$ is a local knowledge
Thus out of the seven knowledge obtained there are 2 global knowledge, 3 local knowledge and 2 misleading information. The local and global knowledge obtained through PCA is a subset of the knowledge in the set IKD obtained through multi-resolution approach. There were 2 misleading knowledge generated by PCA whereas not even a single misleading knowledge was generated though multi-resolution approach. Thus a total of 11 interesting knowledge were obtained through multi-resolution approach in which \{4, 5\}, \{1, 3\}, \{7, 9, 12\}, \{15, 18\}, \{13, 14, 15, 16, 18\}, \{13, 18\} are highly interesting.

**Discussion: DWT for Multiresolution Knowledge Mining**

In this section a technique based on multi-resolution approach using wavelet transform to mine the knowledge existing in a sample set with spatial features is proposed. An illustrative example with 18 samples and 200 feature points along four channels was considered for illustration. The results establish the strength of the multi-resolution techniques over PCA’s approach. This section emphasizes that dimensionality reduction should not be considered as a process of only reducing the features by retaining the global properties but essential efforts need to be taken to retain the local knowledge present in the data. Thus it is very much necessary to propose methodologies for dimensionality reduction where along the process of dimensionality reduction we collect the local knowledge which otherwise would be lost in the reduced feature set. This attempt is one such effort for capturing and retaining the stable local knowledge along the process of dimensionality reduction. This section also emphasizes that all knowledge may not be stable or reliable knowledge. Thus methodologies for generating stable knowledge and new methodologies to integrate/fuse the knowledge obtained from different directions or from multiple sensors can be very good contributions in future.

**6.4 Knowledge Mining using Continuous Wavelet Transform**

The CWT or continuous–time wavelet transform of f(t) with respect to a wavelet ψ(t) is defined as

\[
W(a,b) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{|a|}} \psi^* \left( \frac{t-b}{a} \right) dt
\]  

(6.1)
Wavelet is an astonishing transform having larger orientation towards knowledge mining because details on $f(t)$ are captured at multiple scales/resolutions.

### 6.4.1 CWT for Knowledge Mining

The perspective of CWT in knowledge mining has been already discussed in Chapter 4. Most of the time it is said that CWT domain has a redundant representation of $f(t)$. But redundant from what point of view is more important. CWT representation is redundant only when it comes to reconstruction of $f(t)$ i.e. a part of CWT coefficients are enough for an inverse transform to produce $f(t)$. But from knowledge mining point of view what are available in CWT domain are information rich coefficients which will allow us to extract huge amount of knowledge. In this section we discuss the extraction of knowledge packets at multiple scales or resolution in CWT domain in a multi-channel data set. When the coefficients of CWT ($f(t)$) at a particular scale are more in number the knowledge mining algorithms can become computationally intensive. So a methodology of aggregation is adopted by representing the set of coefficients at each scale for a particular sample by a histogram. The multi-channel signals shown in section 2 are used for further illustration.

### 6.4.2 Continuous Wavelet Histogram based Mining

The signals obtained along each channel are considered one by one. First for the temporal signals generated along direction $D_1$ continuous wavelet transform is applied to obtain the CWT coefficients at different scales. These coefficients are generated for all the samples by applying CWT on signals generated along channel 1 or direction $D_1$. As mentioned before these coefficients are information rich as they are obtained at multiple scales. Now the associations between these samples are analyzed at scales 2, 4, 8, 16, 32, 64, 128. The maximum scale is decided based on the length of the signal. The maximum scale is chosen in such a way that $2^n < \text{length of the spatial signal}$. The reason is that, the coefficients at the intermediate scales would be redundant (coefficients approximately equal to any one of these scales) i.e the intermediate scales would result in generating the same knowledge as these scales would do. For a particular scale $S_x$ the CWT coefficients of every sample is collected and a histogram corresponding to each of them is generated. This histogram is the
frequency distribution of the CWT coefficients binned at 10 centres. This decision for
obtaining a histogram with 10 centres helps in aggregating the entire set of 200
coefficients at a particular scale with just 10 values. As can be observed later in the
results, this aggregation never results in much loss of local associations.

The maximum and the minimum coefficient values at a particular scale for a sample
is recorded and the maximum and minimum among them are considered as the
spread for the histogram generated at that scale. This spread is binned at 10 equiwidth
positions. Thus for N number of samples given a direction Dx and a scale Sx we
generate N number of Histograms. Thus for four channels 4N Histograms is
generated for a scale Sx. For eight such scales we obtain 32N histograms. These
histograms are the aggregated "information rich symbolic objects (more details
provided in chapter 2)" obtained from the information wealthy wavelet coefficients.

There may be an argument why all of the wavelet coefficients are considered in
aggregated information as only the positive values of wavelet coefficients are going
to project the similarity. It should be understood that here wavelet is a template and
the coefficients generated are the measures of comparison between the spatial signal
and the wavelet at a particular scale. Here both the similarity and dissimilarity
between the signal and the wavelet template are going to generate information with
the template as reference. Now the knowledge is mined by estimating the distances
between these set of histograms.

The procedure can be summarized as follows

\{
  For temporal signals Sx obtained along channels x = 1,2,3,4
  \{
    Apply CWT at scale Sc = \{2, 4, 8, 16, 32, 64, 128\}
    \{
      Generate wavelet histograms corresponding to each signal at this scale;
      Compute distance between Histograms using AB regression based
      histogram distance measure;
      Generate the distance matrix along each channel;
      Add the four distance matrices obtained along 4 channels at this particular
      scale
      Generate knowledge packets at this particular scale
    }
  \}
  Obtain comprehensive knowledge
\}
For N number of samples/objects the following procedure is executed to compute the distance

\[
\text{For } \text{obj}1 = 1 \text{ to } N-1 \\
\{ \\
\text{For } \text{obj}2 = \text{obj}1+1 \text{ to } N \\
\{ \\
\text{For channels } C = 1 \text{ to } 4 \\
\{ \\
\text{Compute } AB \text{ Regression based distance between } \text{obj}1 \text{ & } \text{obj}2 \text{ along channel } C \\
\} \\
\text{Total Distance (obj}1,’\text{obj}2) \text{ or (obj}2,’\text{obj}1) = \text{Sum of distance computed along all four channels.} \\
\}
\}
\]

Thus Total distance is an N x N distance matrix, which is given as an input to the linkage clustering algorithm for knowledge extraction.

6.4.3 Discussions

Discussions on comprehensive knowledge

The 18 samples shown in figure 6.2 and 6.3 were considered for experimentation. Histograms were generated for each of these samples as per the procedure. The pairwise distances between these histograms were computed and given as input for the linkage-clustering algorithm. The results produced at various scales/resolutions are given in figure 6.11. Here too the stability factor illustrated in section 6.3.2 is used for mining stable knowledge at every scale.
Fig 6.11 (a)

Stable Knowledge at scale = 2:
\{1,6,3\} \{7,8,9,11,12\}

Fig 6.11 (b)

Stable Knowledge at scale = 4:
\{1,6,3\}
Fig 6.11 (c)
Stable Knowledge at scale = 8:
\{1,2,3,4,5,6,7,8,9,10,11,12\}
\{13,14,15,16,17,18\}

Fig 6.11 (d)
Stable Knowledge at scale = 16:
\{1,3,6\} \{2,4,5\} \{1,2,3,4,5,6\}
\{7,8,9,10,11,12\} \{13,14,15,16,17,18\}
Stable Knowledge at scale = 32:
{7,8,9,10,11,12}

Stable Knowledge at scale = 64:
{6,7,8,9,10,11,12} {2,4,5} {1,3,6}
{13,18} {13,18,15} {14,16,17}
Fig 6.11 (g)

Stable Knowledge at scale = 128:
{3,6} {4,5} {15,18} {7,8,9,10,11,12}
{13,14,15,16,17,18}

Fig 6.11 (h)

As the length of the signal is 200 we stop mining at a scale $n$ such that $n < 200$
As can be observed from figure 6.11 the comprehensive/integrated knowledge would consist of the following sets of stable knowledge

KN1 - \{1,3,6\}

KN2 - \{2,4,5\}

KN3 - \{1,2,3,4,5,6\}

KN4 - \{7,8,9,10,11,12\}

KN5 - \{13,14,15,16,17,18\}

KN6 - \{13,18\}

KN7 - \{13,18,15\}

KN8 - \{14,16,17\}

KN9 - \{3,6\}

KN10 - \{4,5\}

KN11 - \{1,2,3,4,5,6,7,8,9,10,11,12\}

There are 11 sets of knowledge out of which most of them are interesting and stable. KN1, KN2, KN6, KN7, KN8, KN9, KN10 are all interesting local knowledges or intra cluster groups while KN3, KN4, KN5 & KN11 are global knowledges. KN3 is the group of ‘8’s, KN4 is group of ‘O’s and KN5 is group of ‘X’s. These results emphasizes that the technique based on continuous wavelet transform and Histogram distance measure mine the global knowledge while keeping the intracluster association intact.

It is obvious when it comes to histogram analysis a few questions regarding flipped objects arises in the mind which raises doubts about this technique where usually the histogram of an object and its flipped version would be same. The wavelet histograms always have the capability to discriminate between flipped objects, as the histograms will also be flipped if the object is flipped. This is the advantage of wavelet histogram compared to conventional histograms. It is illustrated below

Consider a set of sample points

A1 = \{1 2 3 4 5 6 7 8 7 6 5 4\} and A2 = \{4 5 6 7 8 7 6 5 4 3 2 1\} which is the flipped version of A1. Now the computation of CWT coefficients for A1 and A2 at scale 2 results in

C1 = \{-0.7071 -0.7071 -0.7071 -0.7071 -0.7071 -0.7071 -0.7071 0.7071 0.7071 \}

202
C2 = {-0.7071, -0.7071, -0.7071, -0.7071, 0.7071, 0.7071, 0.7071, 0.7071, 0.7071, 0.7071, 0.7071, 0.7071}

The count of 0.7071 and -0.7071 are given below for C1 and C2

<table>
<thead>
<tr>
<th></th>
<th>-0.7071</th>
<th>0.7071</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>C2</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

which indicates a flip in the object results in flip in histogram. So wavelet histograms will also be useful in identifying the flipped objects.

**Discussions: CWT for Multiresolution Knowledge Mining**

A technique based on multi-resolution approach using continuous wavelet transform to mine the knowledge existing in a sample set with spatial features was proposed. An illustrative example with 18 samples and 200 feature points along four directions was considered for presentation. The results establish the strength of the multi-resolution techniques for mining comprehensive knowledge. The section emphasizes that consideration of only prominent features for mining application would tend to compromise with local knowledge. This local knowledge would be very important for critical knowledge mining applications. Thus it is very much necessary to propose methodologies for knowledge mining, which would be capable of extracting comprehensive knowledge. Methodologies for refining these comprehensive knowledge for filtering out interesting knowledge can also be future scope of interest. This section is one such effort for capturing and retaining the stable local knowledge along the process of knowledge mining.

The perspective of this methodology is to consider multiple sensors from many different channels recording the temporal signals. These spatial signals along each direction can be processed independently in a parallel environment to generate the distances between these samples/objects. Then knowledge can be mined along all directions and fused together as done in the first part of this chapter in section 6.3.3 using DWT or the distances can be added and could be given as input to a mining model (in this case clustering model) as done in the second part of this chapter in section 6.4.
6.5 Performance Analysis

The knowledge obtained from DWT approach discussed in section 3 and the CWT approach discussed in the section 4 is tabulated below.

<table>
<thead>
<tr>
<th>Knowledge obtained through DWT</th>
<th>Knowledge obtained through CWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>{13, 14, 15, 16, 17, 18}, {7, 8, 9, 10, 11, 12}, {4, 5}, {1, 3}, {10, 11}, {7, 9, 12}, {13, 14}, {15, 18}, {13, 14, 15, 16, 18}, {13, 18}, {8, 10}</td>
<td>{1,3,6}, {2,4,5}, {1,2,3,4,5,6}, {7,8,9,10,11,12}, {13,14,15,16,17,18}, {13,18}, {13,18,15}, {14,16,17}, {3,6}, {4,5}, {1,2,3,4,5,6,7,8,9,10,11,12}</td>
</tr>
</tbody>
</table>

It is very interesting to note that all of the local associations/intra-cluster associations among the samples are extracted in both the approaches. The knowledge packets obtained from both the approaches are non-misleading.

Computational Complexity of DWT Technique is computed as below:
Let Number of sampled points in the spatial signal be = N. DWT of N sample points is order of O(N) complexity. Let Number of objects/samples be = K. Number of sources/directions from which the signals are generated/recorded = D. Therefore the complexity of DWT of D * K signals of N sample points each = O(N * K * D). Linkage algorithm with K number of objects will be of O(K^2) complexity. For D number of sources/directions it is O(D*K^2). Let the number of knowledge packets generated be P. The worst case complexity for Knowledge filtering would be O(P). For D directions it is O(P*D). Then for cross intersection the worst case complexity is O(P^2). Therefore the overall complexity of DWT technique for multiresolution knowledge mining is

\[ O(N\times K\times D) + O(D\times K^2) + O(P\times D) + O(P^2) \]

Computational Complexity of CWT Technique is computed as below:
Let Number of sampled points in the spatial signal be = N. CWT of N sample points for S number of scales for K number of object/samples for D directions is order of O(N*S*K*D) complexity. Histogram generation for each of these scales for K number of objects for D direction is O(N * S * D * K). Regression for b number of
bins would take $O(B)$ complexity. Distance matrix computation is of $O(K^2 \times S \times D)$. Clustering algorithm would consume $O(K^2)$ complexity. Thus the overall complexity of the technique is

$$O(N \times S \times K \times D) + O(K^2 \times S \times D)$$

**Experimental Analysis on Fish Contour**

The above foundational techniques were tested on the dataset obtained from SQUID database maintained by Mokhtar. This database comprises of fish contours and can be downloaded. The techniques were tested on these fish contours dataset where we followed the same approach as before by obtaining the spatial signals along the four directions and mining at different resolutions. A sample set of the experimental results has been presented here. The results obtained are very impressive and it can be observed that the association between the samples/objects pops up at different resolutions. It leads us to conclude that knowledge can be mined only by looking at the data samples at multiple resolutions. Trying to mine the knowledge at one particular resolution disallows us from capturing many hidden relationships among the samples (Even if it is done, it should be mined at optimal/best resolution). Depending on the criticality of the knowledge mining application different resolutions will have to be adopted. The results are presented in figure 6.13 and 6.14.

![Fish Contour Samples along with index](image)
Fig 6.13(a)
Knowledge packets at Scale = 2:
{15,22} {5,23} {12,21} {29,30}

Fig 6.13(b)
Knowledge packets at Scale = 4:
{10,22,15} {16,17} {8,11}
PRESERVING GLOBAL AND LOCAL KNOWLEDGE IN MULTI-CHANNEL TEMPORAL DATA MINING USING WAVELET TRANSFORM

Fig 6.13(c)
Knowledge packets at Scale = 8:
\{19,20\} \{1,26\} \{10,15,22\}

Fig 6.13(d)
Knowledge packets at Scale = 16:
\{16,17\} \{28,29,30\} \{19,20,23\}
\{10,22\} \{11,12\}

207
Knowledge packets at Scale = 32:
\{16,17\} \{28,29,30\} \{13,15\}
\{19,20\} \{4,6\} \{10,22\}

Knowledge packets at Scale = 64:
\{13,19\} \{16,17\} \{5,9\} \{3,14\}
\{28,29,30\}
Fig 6.13(g)

Knowledge packets at Scale = 128:
{6,26} {28,29,30} {3,25} {16,17,23} {13,15}

Note: Flipped Objects

Fig 6.14 Few Interesting Knowledge Packets
The comprehensive knowledge includes the following set of rich local knowledge that are extracted from the 30 input samples that of fish contours through multi-resolution mining:

\{15,22\} \{5,23\} \{12,21\} \{29,30\} \{10,22,15\} \{16,17\} \{8,11\} \{19,20\} \{1,26\}
\{28,29,30\} \{19,20,23\} \{10,22\} \{13,15\} \{4,6\} \{13,19\} \{5,9\} \{3,14\} \{6,26\} \{3,25\}
\{16,17,23\}

A few impressive knowledge packets are given in figure 6.14. Out of 20 knowledge packets 19 packets are impressive extracts.

### 6.6 Conclusion

In this chapter two techniques for multiresolution knowledge mining based on Discrete Wavelet Transform and Continuous Wavelet Transform was presented. Many novel ideas have been introduced in terms of knowledge filtering, knowledge integration through cross intersection, aggregation of details of a sample with wavelet histograms, and so on. The chapter also emphasizes the need for techniques where the intra cluster association between the samples or objects is retained even after dimensionality reduction or feature selection. The results indicate that knowledge packets pop up at multiple resolutions. If the data is observed at a single resolution the knowledge packets obtained will be only a fraction of knowledge that would have been embedded in the database. It basically brings out an approach to deal with distributed source or sensors recording data independently and how these data can be mined independently for obtaining knowledge packets and later integrated to obtain the comprehensive knowledge base.

This chapter acts as the initiator for multi-channel temporal data analysis. It emphasizes that, not much local knowledge is lost by aggregation of information contained in wavelet coefficients through histograms. Through multi-resolution analysis the temporal signals are in other words subjected to data digging before aggregating the information through histograms. The summarized results provided in section 6.5 clearly stand by this. The following chapter comprehends the methodologies proposed so far to propose wavelet networks for efficient multi-channel temporal data mining. Chapter 7 proposes two wavelet histogram networks based on multiple resolution mining and optimal resolution mining respectively.