Mutual Fund 수익률의 비정상 함수형 시그널을 위한 다해상도 클러스터 계층구조*

†김대룡** · ††정 욱**

Multi-scale Cluster Hierarchy for Non-stationary Functional Signals of Mutual Fund Returns*

†Dae-Lyong Kim** · ††Uk Jung**

Abstract

Many Applications of scientific research have coupled with functional data signal clustering techniques to discover novel characteristics that can be used for the diagnoses of several issues. In this article we present an interpretable multi-scale cluster hierarchy framework for clustering functional data using its multi-aspect frequency information. The suggested method focuses on how to effectively select transformed features/variables in unsupervised manner so that finally reduce the data dimension and achieve the multi-purposed clustering. Specially, we apply our suggested method to mutual fund returns and make superior-performing funds group based on different aspects such as global patterns, seasonal variations, levels of noise, and their combinations. To promise our method producing a quality cluster hierarchy, we give some empirical results under the simulation study and a set of real life data. This research will contribute to financial market analysis and flexibly fit to other research fields with clustering purposes.

Keywords : Mutual Fund Returns, Unsupervised Clustering, Non-stationary Functional Data, Wavelet, Multi-resolution Analysis

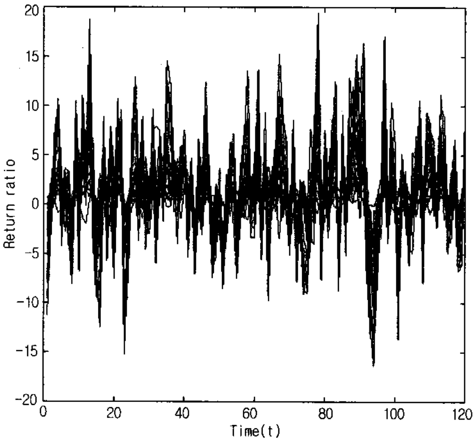
논문접수일 : 2007년 04월 19일 논문게재확정일 : 2007년 09월 10일

* This work is supported by the Dongguk University research fund of 2005(DRIMS 2005-2027-0).

** 동국대학교 경영학과
† 주저자
†† 교신저자

1. Introduction

The research motivation is from mutual fund returns which is one of financial observations from the real market. On the financial view points, the trends of return and risk of mutual fund are very important to evaluate fund performance and to predict future performance. Mutual funds with the trend of high return and low risk are treated as superior funds. Therefore, financial researchers and performance evaluators of real market continuously try to create several groups of mutual funds to find superior performing funds. It is, however, not an easy task to cluster mutual funds under traditional clustering techniques because the observed mutual fund returns are time-serial high-dimensional observations. For example, <Figure 1> represents a set of twenty nine mutual fund monthly returns during 10-year time period.

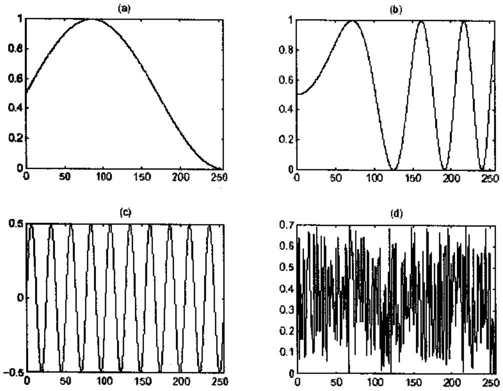


<Figure 1> Mutual Fund Returns in original time domain ; The number of funds=29 ; Length of time=120

Due to the complexity of analyzing high-dimensional signals, most researcher often only use very simple and basic descriptive statistics to characterize the signals and perform monitoring studies in industrial practice. For example, the maximum magnitude and the average value of the signal are the most commonly used statistics [7]. In these methods, a large portion of the information contained in the signals is not fully explored. Therefore, monitoring systems based on these simple statistics often suffer from high false-alarm rates and/or poor detection rates for various types of problematic conditions.

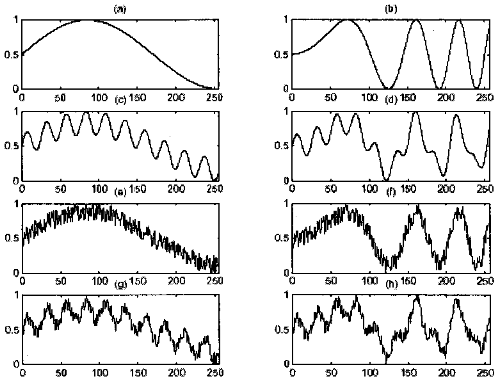
A solution to overcome this type of problems is to identify internal structure in the data and to use the corresponding prior knowledge to simplify data analysis. A very general internal structure model can be obtained by assuming that a high dimensional vector is in fact a discretized function. This model covers for instance time series, spectrometric data, etc. Functional Data Analysis (FDA) is an extension of traditional data analysis methods to this kind of functional data[20]. In FDA, each individual is characterized by one or more real valued functions, rather than by a vector of R^p.

For the clustering of a high-dimensional dataset, the first step is often to reduce the data dimension. Several dimension reduction techniques have been developed in recent years including those of Carreira-Perpinan[4]. As reviewed by Jeong et al. [12], however, there are limited studies dealing with the analysis of high-dimensional functional data. In dealing with complicated data patterns, a priori knowledge is commonly used to guide data preservation or feature extraction methods for selecting representative data in smaller size for subsequent analyses [13]. When a priori knowledge is limited, many studies have used wavelet-based data denoising techniques [5] for data reduction purposes. Jung et al. [14] proposed a vertical-energy thresholding (VET) procedure for locating representative wavelet coefficients in order to reconstruct multiple data curves effectively and efficiently. However, these are not for unsupervised clustering which can bring interpretable cluster hierarchy. In other data dimension reduction methods, such as principal component analysis and the self organizing map method, the feature selection techniques focus more on the faithful representation of the original data, instead of clustering [6]. Jung et al. [15] proposed a vertical group-wise threshold (VGWT) procedure for the reduction of multiple high dimensional functional data containing the cluster membership information. Although it was successful to reduce the dimensionality and enhance the cluster separability, it requires a priori knowledge of class information, which leads to supervised learning scheme. There is an extensive literature on variable selection in multiple regression and supervised classification [8]. However, few results have been presented on feature selection in unsupervised clustering analysis of functional data signals.



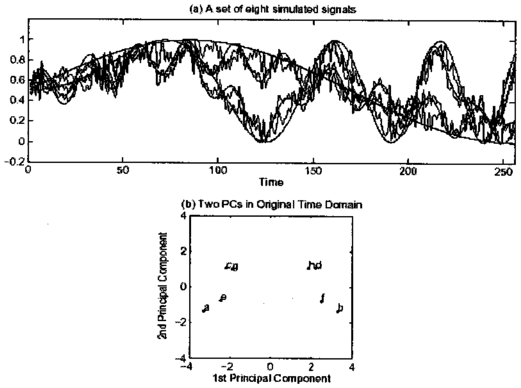
<Figure 2> Signal Components, $1 \le t \le 256$

For more detailed explanations of clustering problem on financial observations, some simulated signals were generated as follows. Most financial researchers and performance evaluators in real market generally assume that mutual funds returns may have internal trends which are interpretable as global patterns, seasonal variations, noise and so on. Therefore, without loss of financial generality, we may generate four simulated internal components of financial signals. <Figure 2> (a) and <Figure 2> (b) represent two different global patterns ($f_1(t)$and $f_2(t)$, respectively). <Figure 2>(c) is for a sample of seasonal variations ($s(t)$) and <Figure 2>(d) for a sample of noise ($n(t)$). All the components have 256 time positions ($1 \le t \le 256$). Using these four components, eight signals were generated as $f_a(t)=f_1(t)$, $f_b(t)=f_2(t)$, $f_c(t)=f_1(t)+s(t)$, $f_d(t)=f_2(t)+s(t)$, $f_e(t)=f_1(t)+n(t)$, $f_f(t)=f_2(t)+n(t)$, $f_g(t)=f_1(t)+s(t)+n(t)$, and $f_h(t)=f_2(t)+s(t)+n(t)$. The plot forms of simulated stationary and non-stationary financial signals are shown in <Figure 3> (a), (b), (c), (d), (e), (f), (g), and (h), for $f_i(t)$, $i=a,b,c,d,e,f,g,$ and h, respectively.



<Figure 3> Simulated signals, $1 \le t \le 256$

With these simulated stationary and non-stationary financial signals, we may consider to cluster for making objected groups using the traditional clustering technique. <Figure 4> (a) shows a set of eight simulated signals which were shown in-



〈Figure 4〉 First two principal components of eight simulated signals in original time domain

dividually in <Figure 3>. In <Figure 4> (b), the visual representation of signal locations in first two principal components in original time domain is shown. It appears that there seem exist two distinct groups $\{a, c, e, g\}$ and $\{b, d, f, h\}$ (in more detailed aspect, four groups $\{a, e\}$, $\{b, h\}, \{c, g\}$, and $\{d, h\}$). However, if we cluster to two or four groups based on its empirical appearance, we may loose some interpretable internal information. In other words, we do not know yet what interpretable reason bring the result of separation. Until we plot those grouped signals individually again, it just appears not the financially interpretable difference among several signals, but the dissimilarity among several signals in first two principal component space under mathematical perspectives. Therefore, we focus how effectively we can separate these signals to distinct groups in terms of a certain aspect of interest such as top-down hierarchy of several variation support. By the reason above, to create objected groups based on its non-stationary functional signals, we propose a multi-scale cluster hierarchy framework using its multi-aspect frequency information.

The suggested method is a four step procedure. At the first step, the functional data signals are modeled as wavelet structure to take advantage of multi-resolution analysis. It brings us the manageability of multiple sets of variables that can be used for different purposes of clustering and better understanding of cluster hierarchy structure. In the second step, Principal Component Analysis (PCA) is used for each resolution to find a major direction of different aspect of signal variations and to obtain visual representation of clusters. Hence, the dimension of the problem is significantly reduced. In the third step, a clustering algorithm is applied to the selected principal components to group the functional data signals. Finally, based on several aspects of clustering using multi-resolution analysis, we generate the cluster hierarchy which is interpretable. This suggested technique can automatically find the clusters in a set of functional data signals in an unsupervised manner. It is different from well-known hierarchical clustering generating *dendrogram* in the sense that the multi-scale (resolution) cluster hierarchy will give us additive meaning of cluster levels, not simply the meaning of distances among signals. To promise our framework producing a quality cluster hierarchy, we give some empirical results under the simulation study and a set of real life data. This research will contribute to financial market analysis and flexibly fit to other related research fields with clustering purposes. The proposed technique may also be considered as an important data pre-processing technique for data dimension reduction in the development of monitoring and diagnostic systems using functional data signals.

This article is organized as follows. In Section 2, we review the background of wavelet. In Section 3, we introduce our proposed method. Section 4 presents both a numerical simulation and a case study of mutual fund returns from real-word financial market to illustrate the effectiveness of the suggested method. The conclusions and future researches are presented in Section 5.

2. Brief Review of Wavelet Transformation

In order to introduce the new clustering method which guarantee interpretability of different aspect of clusters for non-stationary functional data, the Wavelet transformation is briefly reviewed below.

A wavelet is a function $\psi(t) \in L^2(R)$ with the following basic properties

$$\int_R \psi(t)\,dt = 0 \text{ and } \int_R \psi^2(t)\,dt = 1$$

where $L^2(R)$ is the space of square integrable real functions defined on the real line R. Wavelets can be to create a family of time-frequency atoms, $\psi_{s,u}(t) = s^{1/2}\psi(st-u)$, via the dilation factor s and the translation u. We also require a scaling function $\phi(t) \in L^2(R)$ that satisfies

$$\int_R \phi(t)\,dt \neq 0 \text{ and } \int_R \phi^2(t)\,dt \neq 1.$$

Selecting the scaling and wavelet functions as $\{\phi_{L,K}(t) = 2^{L/2}\phi(2^L t - k);\, k \in Z\}$, $\{\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k);\, j \geq L, k \in Z\}$, respectively, on can form an orthonormal basis to represent a signal function $f(t) \in L^2(R)$ as follows.

$$f(t) = \sum_{k \in Z} c_{L,k}\phi_{L,k}(t) + \sum_{j \geq L}\sum_{k \in Z} d_{j,k}\psi_{j,k}(t)$$

where Z denote the set of all integers $\{0, \pm 1, \pm 2, \cdots\}$ and the coefficients $c_{L,K} = \int_R f(t)\phi_{L,K}(t)\,dt$ are considered to be the coarser-level coefficients characterizing smoother data patterns, and $d_{j,k} = \int_R f(t)\psi_{j,k}(t)\,dt$ are viewed as the finer-level coefficients describing (local) details of data patterns. In practice, the following finite version of the wavelet series approximation is used :

$$\tilde{f}(t) = \sum_{k \in Z} c_{L,K}\phi_{L,K}(t) + \sum_{j=L}^{J}\sum_{k \in Z} d_{j,k}\psi_{j,k}(t) \quad (1)$$

where $J > L$ and L correspond to the coarsest resolution level. Consider a sequence of data $\mathrm{y} = (y(t_1), \cdots, y(t_N))$ taken from $f(t)$ or obtained as a realization of $y(t) = f(t) + \epsilon_t$ at equally spaced discrete time points where ϵ_{t_i}'s are independent and identically distributed (i.i.d.) noises. The superscript T represents the transpose operator. The discrete wavelet transform (DWT) of y is defined as

$$\mathrm{d} = \mathrm{Wy}$$

where W is the orthonormal $N \times N$ DWT-matrix. From (1), we can write $\mathrm{d} = (\mathrm{c_L}, \mathrm{d_L}, \mathrm{d_{L+1}}, \cdots, \mathrm{d_J})$, where $\mathbf{c}_{\mathrm{L}} = (c_{L,0}, \cdots, c_{L,2^{L-1}})$, $\mathrm{d_L} = (d_{L,0}, \cdots, d_{L,2^{L-1}})$, $\cdots$ $\mathrm{d_J} = (d_{J,0}, \cdots, d_{J,2^{J-1}})$ are called scales or subband. Using the inverse DWT, the $N \times 1$ vector y of the original signal curve can be reconstructed as $\mathrm{y} = \mathrm{W}^{\mathrm{T}}\mathrm{d}$. The process of transforming a data set via the DWT closely resembles the process of computing the Fast Fourier Transformation (FFT) of that data set. By applying the DWT to the data y_i's, $\mathrm{d} = \mathrm{W_y}$, we obtain the following model in the wavelet domain : $d_{j,k} = \theta_{j,k} + \eta_{j,k}$, for $j = L, \cdots, J$, $k = 0, 1, \cdots, 2^{j-1}$, and $c_{L,K} = \theta_{L,K} + \eta_{L,K}$, for $j = L, \cdots, J$, $k = 0, 1, \cdots, 2^{L-1}$, where $J = \log_2 N - 1$. The model can be represented in the vector format as follows.

$$\mathrm{d} = \theta + \eta$$

where d, θ and η represent the collection of all coefficients, parameters and errors, respectively. Since W is an orthonormal transform, $\eta_{j,k}$'s are still i.i.d. $N(0, \sigma^2)$ [25].

3. Proposed clustering method

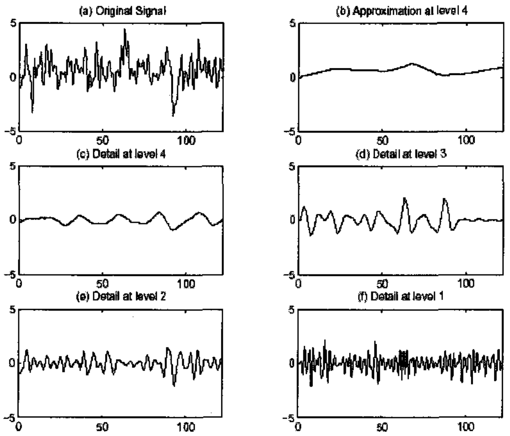
3.1 Multi-resolution Analysis

In this subsection, we introduce the subject of multi-resolution analysis. Those who need more for proofs and discussions is referred to Mallat [19]. A function or signal can be viewed as composed of a smooth background and fluctuations or details on top of it. The distinction between the smooth part and the details is determined by the resolution, that is, by the scale below which the details of a signal cannot be discerned. At a given resolution, a signal is approximated by ignoring all fluctuations below that scale. We can imagine progressively increasing the resolution ; at each stage of the increase in resolution finer details are added to the coarser description, providing a successively better approximation to the signal. Eventually when the resolution goes to infinity, we recover the exact signal.

The above intuitive description can be made more precise as follows. We label the resolution level by an integer j. The scale associated with the level $j=0$ is set to, say, unity and that with the level j is $1/2^j$. Let us consider a function $f(t)$.
At resolution level j it is approximated by $f_j(t)$. At the next level of resolution $j+1$, the details at that level denoted by $d_j(t)$ are included and we have the approximation to $f(t)$ at the new resolution level, $f_{j+1}(t)=f_j(t)+d_j(t)$. The original function $f(t)$ is recovered when we let the resolution go to infinity

$$f(t)=f_j(t)+\sum_{k=j}^{\infty}d_k(t)$$

The word multi-resolution refers to the simultaneous presence of different resolutions. The above equation represents one way of decomposing the function $f(t)$ into a smooth part plus details. Similarly, we can view the space of functions that are square integrable, $L^2(R)$, as composed of a sequence of subspaces W_k and V_j and the details $d_k(t)$ are in W_k.



<Figure 5> An example of reconstruction of Multi-resolution decomposition

In <Figure 5>, the details at several level ks, $d_k(t)$, are presented. In this article, we regard the approximation signal(such as <Figure 5> (b)) at level L as containing the intrinsic nature of global pattern in a signal, the low level detail signals (such as <Figure 5> (c) and <Figure 5> (d)) as the impact of seasonal variations at different support size, and the high level detail signals (such as <Figure 5> (e) and <Figure 5> (f)) as the impact of noise at different support size. Our proposed clustering method utilizes this multi-resolution property of wavelet in order to achieve the interpretable differences among functional signal clusters. That is, wavelet coefficients in several resolution levels in hierarchy structure will be used for clustering. This brings us to the subject of this subsection, multi-resolution analysis.

3.2. Cluster Hierarchy using Multi-resolution Principal components

Our wavelet coefficients obtained from multi-resolution analysis need to be clustered. We simplify the notation $\mathrm{d}=(\mathrm{d}_0,\cdots,\mathrm{d}_j,\cdots,\mathrm{d}_J)$ where $\mathrm{d}_j=(d_{j,0},\cdots, d_{j,2^j-1})$, instead of using $\mathrm{d}=(\mathrm{c}_L,\mathrm{d}_L,\mathrm{d}_{L+1},\cdots,\mathrm{d}_J)$ without confusing. Since the approximation signals may still have a high dimensionality of 2^L-1 and they will suffer the curse of dimensionality (i.e., the sample size needed to estimate the density function is proportional to the exponential of the number of dimensions [4], it is possibly very difficult to apply directly clustering algorithms to the raw dataset. In this paper, PCA is used to reduce the dimension of the dataset as well as the visual representation of clusters. PCA linearly transforms the raw dataset into a new set of variables, called Principal Components (PCs).

Since, at each resolution level j, the scheme to project signals to the PCs is identical, we simplify $\mathrm{d}_j=\mathrm{d}$. Given a dataset $D^{M\times p}$ with p variables ($p=2^L-1$) and M sample signals (i.e. $D=(\mathrm{d}^1,\cdots,\mathrm{d}^M)^T$, where $\mathrm{d}^i=(d_1^i,\cdots,d_p^i)$, d_p^i for a pth wavelet coefficient at a certain resolution level of signal i) and S is the $p\times p$ sample covariance matrix with eigenvalue-eigenvector pairs $(\lambda_1,\mathrm{e}_1),(\lambda_2,\mathrm{e}_2),\cdots,(\lambda_p,\mathrm{e}_p)$, the kth principal component is given by :

$$h_k=\mathrm{e}_k^T\mathrm{d}=e_{k1}d_1+e_{k2}d_2+\cdots e_{kp}d_p,\ k=1\cdots p$$

where $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_p\geq 0$ and d is a row vector of wavelet coefficients with size p of one sample signal. Also, the sample variance of h_k is λ_k, $k=1\cdots p$, and the sample covariance between h_k, and $h_{k'}$ is zero for $k\neq k'$. In addition, the total sample variance is $trace(S)$ that is equal to $\lambda_1+\lambda_2+\cdots+\lambda_p$, where $trace(S)$ is the summation of the diagonal elements of S. The sample variance explained by the kth principal component is given by $\lambda_k/trace(S)$.

As mentioned in Section 2, the sum of the variances of the first few principle components account for a large portion of the total variance of approximation signals. Thus, we can characterize each wavelet coefficient vector of sample signal i, d^i, by a vector of the first few principle components $\mathrm{h}_i=(h_{i1},\cdots,h_{i\tilde{p}})$ where $\lambda_1\geq\cdots\geq\lambda_{\tilde{p}}\geq\cdots\geq\lambda_p\geq 0$ for the purpose of clustering. From PCA using approximation signals with first few PCs, we achieve a faithful representation of clusters disregarding outlying behavior of signals with high peaks and noise. See <Figure 6> (in this plot , $\tilde{p}=2$) and Section 4 for more explanation (The number that superimposed on each dot represents the index of 8 signals). However, the appropriate number of clusters and each signal cluster membership information is still required.

In the problem of obtaining cluster membership information, there are many clustering methods, ranging from heuristic approaches such as K-means [10] and linkage analysis [16] to more formal model-based procedure [1]. In this paper, K-means clustering which is one of the simplest unsupervised learning algorithms that solve the well known clustering problem, will be used. Consider a dataset composed of M elements in $R^{\tilde{p}}$ that contains a maximum of $K_{\max}$ clusters. Let $d(\mathrm{h}_i,\mathrm{h}_j)$ denote the distance between new representations of signals at a certain resolution level, h_i and h_j, $1\geq i,\ j\geq M$. Let us define the encoder $G(i)=k$, that assigns the ith observation, h_1, to the kth cluster. The K-means algorithm is one of the most popular iterative descent clustering methods. It is intended for situation in which all variables are of the quantitative type, and squared Euclidean distance

$$d(\mathrm{h}_i,\mathrm{h}_{i'}) = \sum_{j=1}^{p}(h_{ij}-h_{i'j})^2 = \| \mathrm{h}_i - \mathrm{h}_{i'} \|^2$$

is chosen as the dissimilarity measure. Note that weighted Euclidean distance can be used by redefining the h_{ij} values. The within-point scatter can be written as

$$W(G) = \frac{1}{2}\sum_{k=1}^{K}\sum_{G(i)=k}\sum_{G(i')=k} \| \mathrm{h}_i - \mathrm{h}_{i'} \|^2$$
$$= \sum_{k=1}^{K}\sum_{G(i)=k} \| \mathrm{h}_i - \overline{\mathrm{h}_k} \|^2$$

where $\overline{\mathrm{h}_k}$ is the mean vector associated with the kth cluster. Thus, the criterion is minimized by assigning the M observations to the K clusters in such a way that within each cluster the average dissimilarity of the observations from the cluster mean, as defined by the points in that cluster, is minimized. In all of the clustering methods, the data are clustered multiple times for various numbers of clusters. For each value of $K(1 \le K \le K_{\max})$, let $G_1, G_2, \cdots, G_k, \cdots, G_K$ be a set of clusters such that n_k is the number of h_i in G_k.

Most clustering procedures require one to choose the number of groups prior to fitting. This is one of the most difficult problems in cluster analysis. The earliest cluster number estimation approaches are based upon measures of within-cluster-homogeneity or between-cluster-heterogeneity, e.g., Calinski and Harabasz [3], Hartigan [9], Krzanowski and Lai[18], and Rousseeuw [21]. In all four methods, the data are clustered multiple times each time with a different number of clusters. For each number of clusters, a statistic measuring the quality of the clustering is computed. We utilize the approach suggested by Rousseeuw [21] in this paper. Rousseeuw introduces new functions to measure within-cluster-homogeneity and between-cluster-heterogeneity. For $\mathrm{h}_i \in G_k$

$$a(\mathrm{h}_i) = \frac{1}{\mathrm{n}_k - 1}\sum_{\mathrm{h}_j \in \mathrm{G}_k} \mathrm{d}(\mathrm{h}_i, \mathrm{h}_j)$$

and

$$b(\mathrm{h}_i) = \min_{s \ne k}\frac{1}{n_k}\sum_{\mathrm{h}_j \in \mathrm{G}_s} \mathrm{d}(\mathrm{h}_i, \mathrm{h}_j).$$

For each K, the quality of the clustering is summarized by the silhouette statistic :

$$s_K = \frac{1}{M}\sum_{i=1}^{M}\frac{b(\mathrm{h}_i) - a(\mathrm{h}_i)}{\max\{a(\mathrm{h}_i), b(\mathrm{h}_i)\}}$$

The number of cluster is estimated by

$$\mathrm{argmax}_{K=2,\cdots,K_{\max}}\ s_K$$

That is, to get an idea of how well-separated the resulting clusters are, we check a silhouette values. The silhouette value is a measure of how close each point in one cluster is to points in the neighboring clusters. This measure ranges from $+1$, indicating points that are very distant from neighboring clusters, through 0, indicating points that are not distinctly in one cluster or another, to -1, indicating points that are probably assigned to the wrong clusters. A good quantitative way to compare the solutions with several trial numbers of clusters is to look at the average silhouette values for each cases.

At this point, we need to address a major drawback of K-means algorithm from the fact that the clustering quality is greatly dependent on the choice of initial centers, $\overline{\mathrm{h}}_{\mathrm{k,initial}}$, so that K-means algorithm guarantees local, but not necessarily global optimization. Poor choice $\overline{\mathrm{h}}_{\mathrm{k,initial}}$, therefore, can degrade the quality of clustering solution. In order to avoid this drawback, our clustering method proposes a sufficient number of repeated center initialization at a certain number of clusters $k(1 \le k \le K)$.

In our case studies in Section 4, sixty repetitions were performed at each k. At each k, hence, the maximum average silhouette will be used to find the most appropriate number of clusters. See <Figure 7> (b) and <Figure 9> (d), <Figure 9> (e), and <Figure 9> (f) for comparison among several trials of ks. In the case of intrinsic unity (a single cluster, in other words, which is not suitable to divide a whole set of data to k clusters, $2 \leq k \leq K$, k is a positive integer) a certain threshold of the maximum average silhouette can be utilized (a threshold 0.7 is used in our case studies). That is, if a maximum average silhouette value for any trials of k is less than the threshold, we conclude the single cluster is most appropriate. The optimal threshold of maximum average silhouette will be remained as future research.

According to the clustering scheme we discussed so far, we have multiple cluster memberships which depend on different resolution levels. Thus, we consider a sequence of cluster index for a certain signal $\mathbf{h}_i$ at successive resolution levels ; the cluster index of a signal h_i at any resolution level r is denoted by $I_r(i)$. A particular sequence of cluster index from every resolution levels for a signal $\mathbf{h}_i$ is denoted by

$$I(i) = \{I_{A(L)}(i), I_{D(L)}(i), I_{D(L-1)}(i), \cdots, I_{D(1)}(i)\}$$

where L = the coarsest resolution level ; $A(L)$ = approximation at level L ; $D(L)$ = details at level L ; $D(L-1)$ = details at level $L-1$, and so on.

Based on $I(i)$, the final cluster encoder

$$f(I(i); R_1; R_2) = k$$

will assign the ith observation, h_i, to the kth cluster. k is an integer which has a range, $1 \leq k \leq K$, where K is the number of different sequences of $I_r(i)$s where $r$is from R_1 to R_2 ($A(L) \geq R_1 \geq \cdots \geq R_2 \geq D(1)$; $a \geq b$ means a is a coarser level than b). $f(I(i); R_1; R_2)$ is a function of $I(i)$ to assign the final cluster index to each signals. In this notation, it is possible for two different signals h_i and h_j ($i \neq j$) to belong to different clusters as resolution level r even though $I_r(i) = I_r(j)$ still holds, due to the hierarchical structure of clusters. In other words, they are different branch from same stem if $I_{r^* > r}(i) \neq I_{r^* > r}(j)$ for any r^*. Thus, we notice the fact that the interpretation of cluster index $I_r(i)$ is dependent of $I_{r^* > r}(i)$ (let us say, the cluster index in coarser resolution level) in the sequence $I(i)$. The examples of the proposed cluster scheme is in <Table 1> and <Table 2> in Section 4. More details about both tables are given in Section 4.

<Table 1> List of Eight Simulated Signals and Cluster Hierarchy Notation Index

Signal i	$I(i)$	f_1	f_2	f_3	f_4	f_5
a	$I(a)$ = {1, 1, 1, 1, 1}	1	1	1	1	1
b	$I(b)$ = {2, 1, 1, 1, 1}	2	3	3	2	2
c	$I(c)$ = {1, 2, 2, 1, 1}	1	2	2	3	3
d	$I(d)$ = {2, 2, 2, 1, 1}	2	4	4	4	4
e	$I(e)$ = {1, 1, 1, 2, 2}	1	1	1	5	5
f	$I(f)$ = {2, 1, 1, 2, 2}	2	3	3	6	6
g	$I(g)$ = {1, 2, 2, 2, 2}	1	2	2	7	7
h	$I(h)$ = {2, 2, 2, 2, 2}	2	4	4	8	8

where $f_1 = f(I(i); A_{(4)}; A_{(4)})$, $f_2 = f(I(i); A_{(4)}; D_{(4)})$, $f_3 = f(I(i); A_{(4)}; D_{(3)})$, $f_4 = f(I(i); A_{(4)}; D_{(2)})$, $f_5 = f(I(i); A_{(4)}; D_{(1)})$

〈Table 2〉 List of twenty-nine mutual fund return signals and cluster hierarchy notation index

Signal Index i	Cluster Hierarchy Notation $I(i)^*$	Cluster Index $f(I(i);A_{(4)};D_{(1)})$
1	$I(1)=\{1, 1, 3, 1, 2\}$	3
2	$I(2)=\{1, 1, 1, 3, 3\}$	1
3	$I(3)=\{1, 1, 3, 1, 2\}$	3
4	$I(4)=\{1, 1, 3, 1, 2\}$	3
5	$I(5)=\{1, 1, 1, 3, 3\}$	1
6	$I(6)=\{1, 1, 3, 1, 2\}$	3
7	$I(7)=\{1, 1, 3, 1, 2\}$	3
8	$I(8)=\{1, 1, 2, 2, 1\}$	2
9	$I(9)=\{1, 1, 2, 2, 1\}$	2
10	$I(10)=\{1, 1, 3, 1, 2\}$	3
11	$I(11)=\{1, 1, 3, 1, 2\}$	3
12	$I(12)=\{1, 1, 2, 2, 1\}$	2
13	$I(13)=\{1, 1, 2, 2, 1\}$	2
14	$I(14)=\{1, 1, 1, 3, 3\}$	1
15	$I(15)=\{1, 1, 2, 2, 1\}$	2
16	$I(16)=\{1, 1, 3, 1, 2\}$	3
17	$I(17)=\{1, 1, 2, 2, 1\}$	2
18	$I(18)=\{1, 1, 1, 3, 3\}$	1
19	$I(19)=\{1, 1, 3, 1, 2\}$	3
20	$I(20)=\{1, 1, 2, 2, 1\}$	2
21	$I(21)=\{1, 1, 1, 3, 3\}$	1
22	$I(22)=\{1, 1, 2, 2, 1\}$	2
23	$I(23)=\{1, 1, 2, 2, 1\}$	2
24	$I(24)=\{1, 1, 1, 3, 3\}$	1
25	$I(25)=\{1, 1, 1, 3, 3\}$	1
26	$I(26)=\{1, 1, 2, 2, 1\}$	2
27	$I(27)=\{1, 1, 1, 3, 3\}$	1
28	$I(28)=\{1, 1, 2, 2, 1\}$	2
29	$I(29)=\{1, 1, 3, 1, 2\}$	3

where $I(i)=\{I_{A_{(4)}}(i), I_{D_{(4)}}(i), I_{D_{(3)}}(i), I_{D_{(2)}}(i), I_{D_{(1)}}(i)\}$

4. Case studies

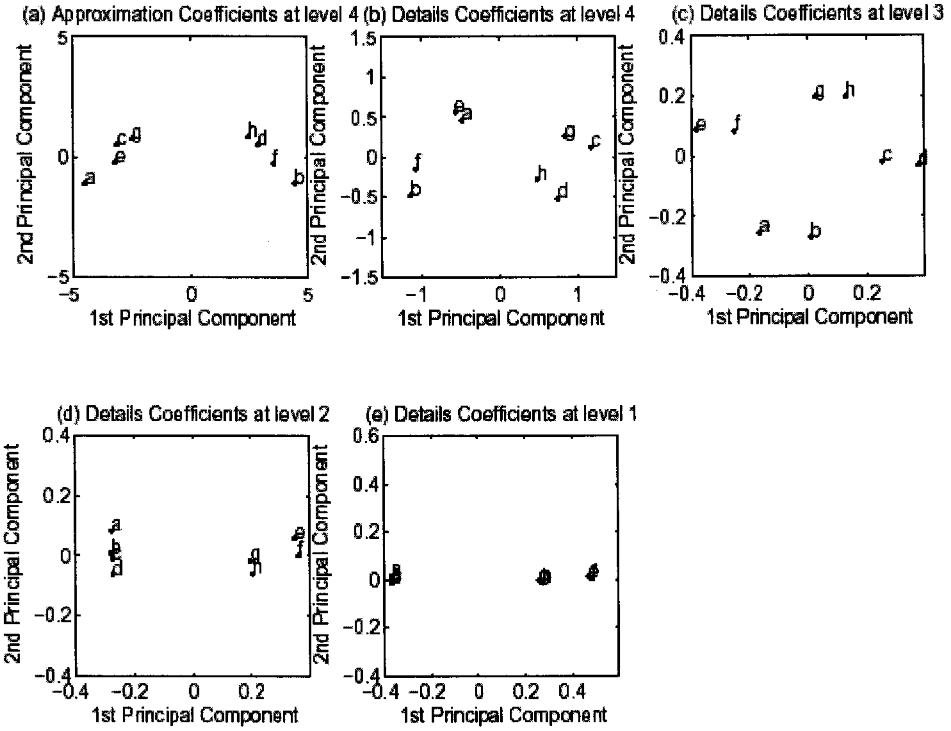
4.1 Simulation Data with Several Differences

In this section we apply the proposed clustering method to a set of simulated signals in <Figure 4> (a). For comparison purpose, the result of applying PCA directly to original time domain was plotted in <Figure 4> (b). It was problematic to interpret as mentioned before in Section 1. In order to avoid this drawback, wavelet transformation is applied to the simulated signals. In this case, the simulated signals were decomposed at level 6 using Symmlet4 as a type of wavelets. Then the approximation was performed at level 4 ($L=4$) which contains details at level 6 and 5 in order to avoid over-smoothing. Then PCA was applied to the each approximation/details signals at each level and the result is shown in <Figure 6>.

Using those PC plots for K-means clustering with the number of cluster = 2 at each resolution level, the sequence of cluster index, $I(i)$, and the final cluster index function, $f(I(i);R_1;R_2)$, are listed in <Table 1>. According to the $I(i)$s, the two clusters at $A(4)$(which represents the global pattern) are {a, c, e, g} and {b, d, f, h} in accordance with our intention during the simulation of signal generation. Also, the two clusters at $D(4)$ and $D(3)$ (which represents the different support size of seasonal variations) are {a, b, e, f} and {c, d, g, h} and the two clusters at $D(2)$ and $D(1)$ (which represents the different support size of noise) are {a, b, c, d} and {e, f, g, h}. All the results are identical with the intrinsic difference from the signal generations. Also, the results of final clustering in regard of hierarchy structure are obtained by $f(I(i);R_1;R_2)$ (in the table, f_1, f_2, f_3, f_4, f_5). f_1 divided a set of clusters into two groups (cluster index 1 and 2) in terms of global pattern alone, f_2 and f_3 into four groups in terms of global pattern and seasonal variations, and f_4 and f_5 into eight groups in terms of global pattern, seasonal variations, and noise.

4.2 Real Life Data of Mutual Fund Return

As we mention in Section 1, financial researchers and evaluators of real financial market continuously try to group funds to find superior performing funds for predicting future performance. To make ob-

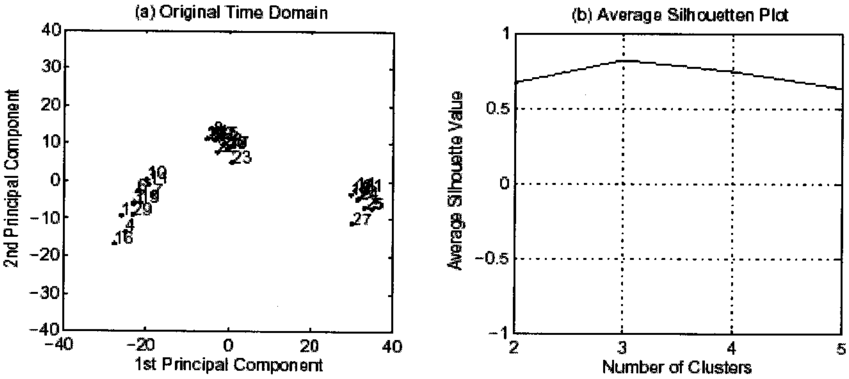


<Figure 6> First two principal components of simulated signals in multi-resolution wavelet domain

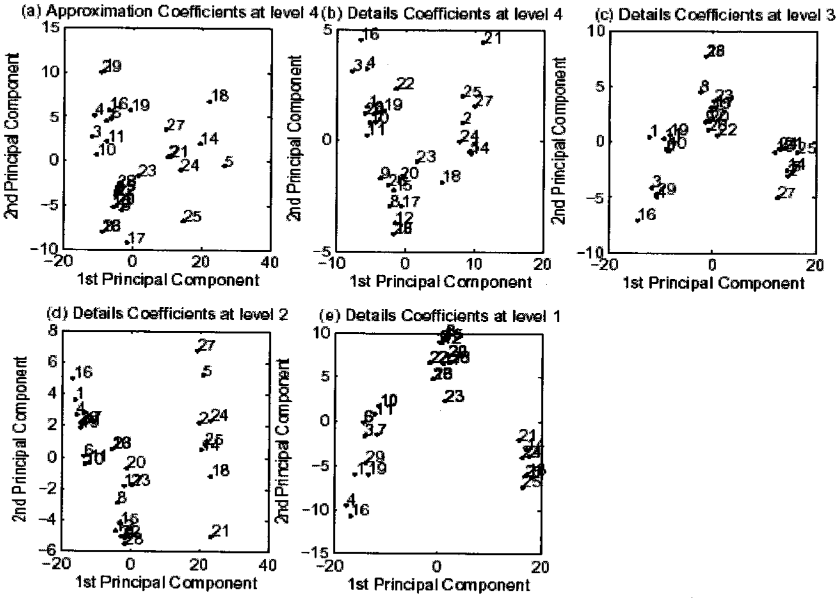
jected groups based on its non-stationary functional signals, we apply our multi-scale cluster hierarchy framework under its multi-aspect frequency information. In this part, we show the evaluation process for actual mutual fund returns using our suggested method. To promise our method to produce a quality cluster hierarchy, we divided the most actively traded 541 firms to several risk (variance) categories by percentile grouping and randomly chose total 29 firms in three categories. That is, each sample signal consists of 120 monthly rates of return data from June 1988 to May 1998 on each of the twenty-nine firms in the mutual funds databases which is maintained by Alexander Steels' Mutual Fund Expert. The original data of this database is provided by Standard and Poors Mircopal. <Figure 1> shows signals of mutual fund return ratio from twenty nine mutual funds. In this case, the number of signals (M) is 29 and the dimension (N) of each signal is 120.

<Figure 7> shows the results of direct PCA using first two PCs in the original time domain. Although the first two PC plot in <Figure 7> (a) and the average silhouette value plot in <Figure 7> (b) identically suggest three distinct clusters, the interpretation of the source of major difference is vague. Thus, we applied our proposed method to these signals as we did in the previous simulation study.

From <Figure 8>, we may notice the difference among possible clusters may exist in the seasonal variations and noise(from <Figure 8> (c), <Figure 8> (d), <Figure 8> (e)) rather than the global patterns (from <Figure 8> (a)). In order to assure



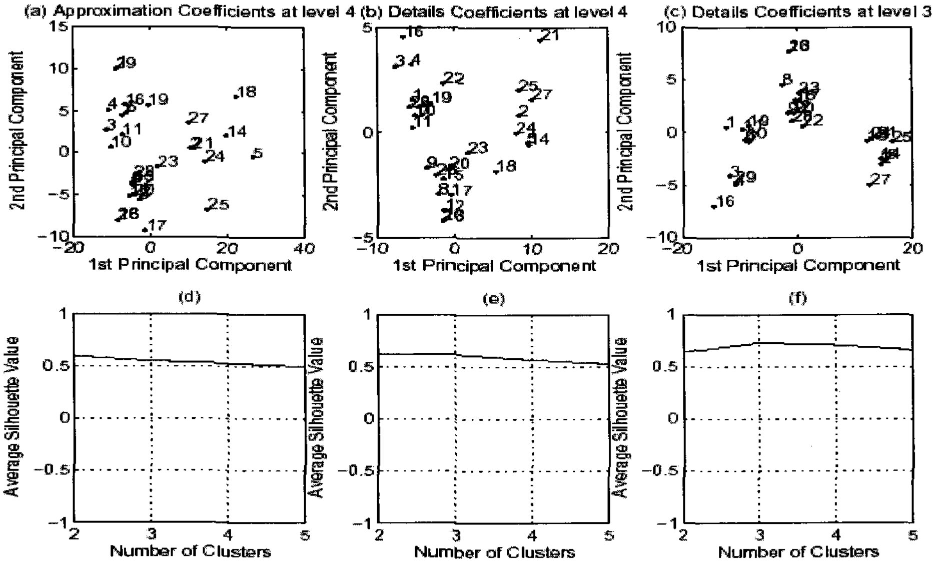
〈Figure 7〉 First two principal components in original time domain



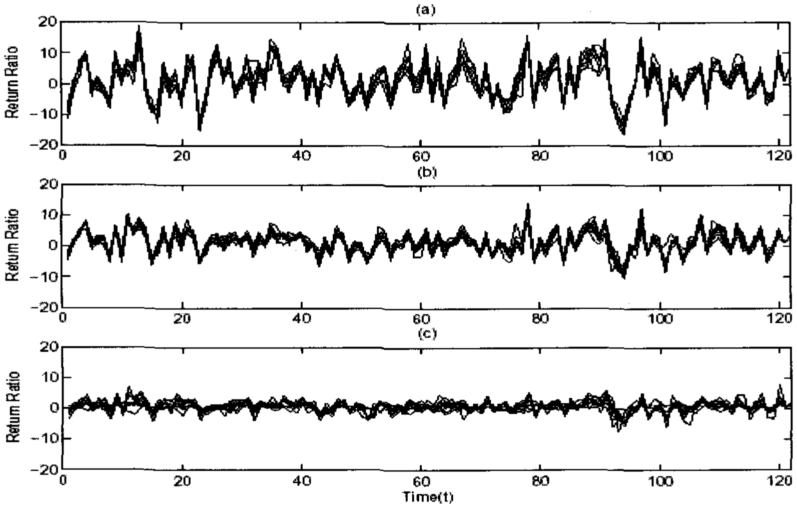
〈Figure 8〉 First two principal components in multi-resolution in wavelet domain

where the major differences exist among the possible clusters, we analyzed the average silhouette values like in <Figure 9>. From <Figure 9> (d) and <Figure 9> (e), the maximum of average silhouette values (0.5971 and 0.6262, respectively) are less than our thresholding value 0.7 so that we conclude there is an intrinsic unity in $A(4)$ and $D(4)$. In <Figure 9> (f), the maximum of average silhouette values (0.7273) is greater than the thresholding value 0.7 so that we conclude there are three intrinsic clusters in $D(3)$. In this way, we found all cluster membership information using K-means algorithm and the result is shown in <Table 2>.

Finally, based on the final cluster index, we can cluster twenty nine mutual funds by three performance groups on the size of variations and noise which are plotted the segmented signals in <Figure 10>. That is, it shows the global pattern is not the



〈Figuer 9〉 First two principal components in (a) Approximation Coefficients at Level 4($A(4)$) ; (b) Detail Coefficients at Level 4($D(4)$) ; (c) Detail Coefficients at Level 3($D(3)$) ; and their average silhouette plots of (d)$A(4)$; (e) $D(4)$; (f) $D(3)$



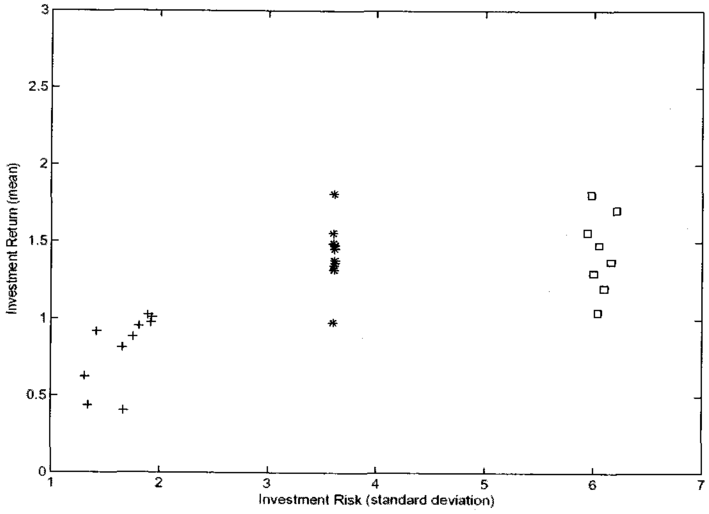
〈Figure 10〉 Segmented mutual fund return signals ; (a) Signals with $f(I(i)\,;A(4)\,;D(1))=1$; (b) Signals with $f(I(i)\,;A(4)\,;D(1))=2$; (c) Signals with $f(I(i)\,;A(4)\,;D(1))=3$

major difference among the clusters, but the size of variations and noise.

Currently, in the real financial market the set of superior performing mutual funds is typically built from percentile-based performance classification method [17] such as "risk-adjusted 5-star rating" which is produced by Morningstar [2]. This is a three stage evaluation procedure. In the first stage,

<Table 3> Average Values of Performance Measures of Each Clustering Group

	# of Funds	*(a)Sharpe*	*(b)Treynor*	*(c)Return*	*(d)Risk(SD)*	*(e)Beta*
Groupl	9	*0.750*	*6.339*	*1.435*	*6.063*	*0.931*
Group2	11	*1.553*	*21.094*	*1.416*	*3.609*	*0.908*
Group3	10	*1.486*	*29.624*	*0.803*	*1.671*	*0.451*



<Figure 11> Investment risk and return plot ; The symbol '□' indicates a signal with $f(I(i);A_{(4)};D_{(1)})=1$, '*' with $f(I(i);A_{(4)};D_{(1)})=2$, and '+' with $f(I(i);A_{(4)};D_{(1)})=3$.

historical return and risk are combined into a single numeric performance value by a performance measure such as Sharpe [22], Treynor [24] or Jensen's measure [11]. In the second stage, the evaluated numeric risk-adjusted performance value is converted to an ordinal performance ranking for indicating the relative performance position of each mutual fund. In the final stage, mutual funds were clustered based on preset number of performance groups and preset percentile as ranking criteria [23].

<Table 3> shows average values of performance measures of each clustering groups that are from our suggested framework under its multi-aspect frequency information. Because there is statistically no difference of (a) Sharpe or (b) Treynor' performance measure between Group2 and Group3, total twenty nine mutual funds are clustered into two groups under the risk-adjusted performance measures. One group is Groupl and the other group is Group2 and Group3 by the percentile-based performance classification method which is typically used in the real market. However, considering financial risk such as (e) beta and (d) standard deviation, we can see that mutual funds in Group2 have more systematic and overall risk performance than those in Group3. Therefore, for finding superior performing mutual funds, risk-averse investors may have more information on the three group clustering than on the two group clustering. It is presented more clearly in <Figure 11> which is risk and return plot.

In the suggested multi-scale cluster hierarchy framework, the global pattern and the size of variations are treated as trend of return and risk, respectively. According to the our suggested method, we can evaluate more financially reasonable superior performing mutual funds based on the trend of return, risk, and may predict mutual funds future performance.

5. Conclusion and Future Research

Since mutual funds with the trend of high return and low risk are treated as superior funds, financial researchers and evaluators of real market need to cluster mutual funds to find superior performing funds. In this article, we apply our suggested framework to mutual fund returns and create superior performing fund group based on its non-stationary functional signals. We present an interpretable multi-scale cluster hierarchy framework for clustering functional data such as mutual fund returns under its multi-aspect frequency information. That is, assuming that mutual funds returns are composed of several internal factors, our suggested framework is to construct a cluster hierarchy that satisfies interpretability of different aspect of clusters, such as global patterns, seasonal variations, noise and so on. It is different from well-known hierarchical clustering generating dendrogram in the sense that the multi-scale (resolution) cluster hierarchy will give us additive meaning of cluster levels, not simply the meaning of distances among signals.

Based on the positive empirical results obtained from our study, the proposed method appears to have good potential in many real-world functional data analysis such as financial market analysis, econometric modeling, machine health monitoring, and bio-informatics applications.

Some future work would be to make the following advanced applications of this suggested clustering technique. First, we need to further study the optimal threshold of maximum average silhouette value to conclude whether there is a single cluster (intrinsic unity). Second, to explore a more rigorous framework to find the cluster hierarchy, we would better study our contents under the statistical distribution properties. If our suggested framework is valuable under the statistical distribution hypothesis, it will be more powerful clustering technique. Also, its statistical analysis will have a large contribution to the statistical data mining field.

Reference

[1] Banfield, J.D. and A.E. Raftery, "Model-based Gaussian and non-gaussian clustering," *Biometrics,* Vol.49(1993), pp.803-821.

[2] Benz, C., P.D. Teresa, and R. Kinnel, *Mutual Funds-Morningstar*, John Wiley & Sons, 2003.

[3] Calinski, T. and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics,* Vol.3, No.1(1974), pp.1-27.

[4] Carreira-Perpinan, M., "A review of dimension reduction techniques," Technical report CS-96-09(1997), Department of Computer Science, University of Sheffield, UK].

[5] Donoho, D.L. and I.M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of American Statistics Association*, Vol.90(1995), pp.1200-1224.

[6] Duda, R.O., P.E. Hart, and D.G. Stork, *Pattern Classification*, 2nd edn., Wiley, New York, NY, 2001.

[7] Grogan, R. "High speed stamping process

improvement thru force and displacement monitoring," Technical report(2002), Helm Instrument Company, Maumee, OH.

[8] Guyon, I. and A. Elissee, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, Vol.3(2003), pp.1157-1182.

[9] Hartigan, J.A. *Clustering Algorithms*, New York, USA : John Wiley and Sons, 1975.

[10] Hartigan, J.A. and Wong, M.A. "A K-means clustering algorithm," *Applied Statistics*, Vol.28, No.1(1978), pp.100-108.

[11] Jensen, M.C., "The performance of mutual funds in the period 1945~1964," *Journal of Finance* Vol.23, No.2(1968), pp.389-416.

[12] Jeong, M.K., J.C. Lu, X. Huo, B. Vidakovic, and D. Chen, "Wavelet-based data reduction techniques for fault detection and classification," *Technometrics*, Vol.48, No.1(2006), pp.26-40.

[13] Jin, J. and J. Shi, "Automatic feature extraction of waveform signals for in-process diagnostic performance improvement," *Journal of Intelligent Manufacturing*, Vol.12 (2001), pp.257-268.

[14] Jung, U., M.K. Jeong, and J.C. Lu, "A vertical-energy-thresholding procedure for data reduction with multiple complex curves," *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, Vol.36, No.5(2006a), pp.1128-1138.

[15] Jung, U., M.K. Jeong, and J.C. Lu, "Data reduction for multiple functional data with class information," *International Journal of Production Research*, Vol.44, No14, 15(2006b), pp.2695-2710.

[16] Kaufman, L. and P. Rousseeuw, *Finding Groups in Data : An Introduction to Cluster Analysis*, New York : Wiley, 1990.

[17] Kim, D.L., "An empirical study of performance rating and distribution of mutual funds under the finite mixtures of normal distribution hypothesis," Ph.D. dissertation, University of Nebraska, USA, 2003.

[18] Krzanowski, W.J. and Y.T. Lai, "A criterion for determining the number of groups in a data set using sum of squares clustering," *Biometrics*, Vol.44, No.1(1988), pp.23-34.

[19] Mallat, S.G., *A Wavelet Tour of Signal Processing*, Academic Press, San Diago, 1989.

[20] Ramsay, J., and B. Silverman, *Functional Data Analysis*, Springer Series in Statistics. Springer Verlag, 1997.

[21] Rousseeuw, P.J., "Silhouettes : a graphical aid to the interpretations and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, Vol.20(1987), pp. 53-65.

[22] Sharpe, W.F., "Mutual fund performance," *Journal of Business* Vol.39, No.1(1966), pp. 119-138.

[23] Sharpe, W.F., "Morning star's risk-adjusted ratings," *Financial Analysis Journal*, (1998), pp.21-33.

[24] Treynor, J.L., "How to rate management of investment funds," *Harvard Business Review*, Vol.43, No.1(1965), pp.63-75.

[25] Vidakovic, B., *Statistical Modeling by Wavelets*, John Wiley & Sons, 1999.