DEVELOPMENT AND EVALUATION OF KERNEL-BASED CLUSTERING VALIDITY INDICES

Rui Fa[1], *Asoke K Nandi*[1,2] *and Basel Abu-Jamous*[1]

[1] Signal Processing and Communications Research Group, Department of Electrical Engineering and Electronics
The University of Liverpool, L69 3GJ, UK. {r.fa, a.nandi, basel88}@liverpool.ac.uk

[2] Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä, Finland.

ABSTRACT

In this paper, there are two objectives: on one hand, we extend four conventional validity indices, namely the *DI*, the *II*, the *CH*, and the *GI*, to four kernel-based validity indices, correspondingly, the *kDI*, the *kII*, the *kCH* and the *kGI*; on the other hand, we conduct a Monte-Carlo simulation to evaluate and compare these validity indices. The numerical results show that some kernel validity indices work significant better than conventional ones and some of the validity indices work poorly or do not work at all in our study.

1. INTRODUCTION

Clustering, also known as unsupervised learning, has been a useful exploratory technique for decades in many fields, such as image processing, data mining and artificial intelligence [1], and in recent years, has benefited microarray gene expression data analysis in genomic research [2]. The goal of the clustering analysis is to group individual objects or samples in a population within which the objects are more similar to each other than those in other clusters. Although there are many widely used clustering algorithms, for example, the k-means [3], the hierarchical clustering (HC) [3], the fuzzy c-means (FCM) [4] and so on, there is no existing guideline to guarantee that one clustering algorithm, which works well in one dataset, can perform also well in a different dataset. Even the same algorithm with different parameter settings or different initialization methods usually produce different clustering results. Thus, the task of assessing the clustering algorithms can be as important as the clustering algorithms themselves.

Since the unsupervised learning is conducted without teacher, it is more difficult to assess than a supervised approach. The procedure for evaluating the results of a clustering algorithm is known as *clustering validation* and the metric for clustering validation is known as *clustering validity index* [5, 6]. There are two objectives of clustering validation: firstly, clustering validity indices are used to assess the cluster results; secondly, clustering validity indices are also used as tools to determine the number of clusters in a given dataset. In general terms, clustering validation can be classified based on two different methodologies: based on the approaches how to investigate cluster validity, there are three classes, namely *external criteria*, *internal criteria* and *relative criteria* [5, 6]; based on the methods how the validity indices are calculated, clustering validation can be classified into three classes. Class one includes cost-function based indices, class two includes density-based indices and class three includes geometric approaches [9]. Among these validity approaches, we are more interested in the relative criterion and geometric indices because of their simplicity and low computational load. There are many validity indices, which belong to both relative criterion and geometric indices, proposed to assess clustering results, including the *Dunn's* index (*DI*) [7], the *I*-index (*II*) [8], the *Calinski Harabasz* (*CH*) index [10] and the geometrical index (*GI*) [9]. The basic principle behind these methods is to calculate the ratio of the intra-cluster scatter to the inter-cluster separation. However, none of these widely adopted methods can be claimed to work well for all types of data and there is no comprehensive evaluation and comparison study of these validity indices in the literature.

Recently, kernel-based clustering, which constructs a hyperplane to separate the linearly inseparable patterns, has attracted a lot of attention. These linearly inseparable patterns are nonlinearly transformed from a set of low-dimensional space into a higher-dimensional feature space to be linear separable [11]. At the core of the kernel-based clustering lies the difficulty of explicitly constructing the nonlinear mapping, which is sometime infeasible; but now it can be overcome by a kernel trick. The kernel trick is a way of mapping patterns from a input space into a feature space without having to compute the mapping explicitly, in the hope that the patterns will gain meaningful linear structure in the feature space. However, to our best knowledge, the kernel-based clustering validity indices have not been investigated. It motivates us to develop and evaluate the kernel-based clustering validity indices.

Thus, the objectives of this paper become two-fold: on one hand, we extend four conventional validity indices,

The project (Ref. NIHR-RP-PG-0310-1004-AN) is supported by National Institute for Health Research (NIHR), UK.

namely the *DI*, the *II*, the *CH*, and the *GI*, to four kernel-based validity indices, correspondingly, the *kDI*, the *kII*, the *kCH* and the *kGI*; on the other hand, we conduct a Monte-Carlo simulation using synthetic gene expression model [13]. Note that we develop these validity indices in order to make the literature complete and give readers a relatively comprehensive view of the kernel validity indices. In the numerical results, we will show that some of kernel validity indices work significantly better than conventional ones and some of validity indices work poorly or do not work at all.

The rest of this paper is organized as follows: Sec. 2 reviews the validity indices and kernel method, Sec. 3 develops the kernel validity indices based on the conventional ones and Sec. 4 presents the method of the simulation and shows the result comparisons. Finally, Sec. 5 concludes the paper.

2. REVIEW FOR VALIDITY INDICES AND KERNEL METHOD

In this preliminary section, we review the validity indices and kernel method separately.

2.1. Validity Indices

We list five validity indices which we are going to investigate. All these validity indices belong to relative criteria category. The basic principle behind these methods is to calculate the ratio of the intra-cluster scatter to the inter-cluster separation.

***Dunn's* index (*DI*):** This index [7] is defined as a fraction, which is written as

$$DI(K) = \min_{1\le i\le K}\left\{\min_{1\le j\le K}\left\{\frac{\delta(\mathcal{C}_i,\mathcal{C}_j)}{\max_{1\le k\le K}\{\Delta(\mathcal{C}_k)\}}\right\}\right\}, \quad (1)$$

where $\delta(\mathcal{C}_i,\mathcal{C}_j)$ is the minimum distance between cluster i and cluster j, $\Delta(\mathcal{C}_k)$ is the largest intra-cluster separation of cluster k. Large values of *DI* are supposed to represent good clustering results and the K-cluster with maximum *DI* value is supposed to be the true number of clusters.

***I*-index (*II*):** The *II* [8] is written as

$$II(K) = \left(\frac{1}{K}\times\frac{E_1}{E_K}\times D_K\right)^P, \quad (2)$$

where $E_1 = \sum_j \|\boldsymbol{x}_j - \boldsymbol{u}\|_2$ where $\boldsymbol{u}$ is the centroid of the whole dataset, $E_K = \sum_{k=1}^K \sum_{j\in\mathcal{C}_k}\|\boldsymbol{x}_j - \boldsymbol{u}_k\|_2$, $D_K = \max_{i,j}^K \|\boldsymbol{u}_i - \boldsymbol{u}_j\|_2$ and power P is constant, which is 2 in our experiments. $\boldsymbol{u}_i$ is the centroid of cluster i. Similar to the *DI*, large values of the *II* are supposed to represent good clustering results and the K-cluster with maximum *II* value is supposed to be the true number of clusters.

Geometrical index (*GI*): The *GI* [9] is expressed as

$$GI(K) = \max_{1\le k\le K}\left\{\frac{(2\sum_{m=1}^M\sqrt{\lambda_{mk}})^2}{\min_{1\le j\le K}\|\boldsymbol{u}_k - \boldsymbol{u}_j\|_2}\right\}, \quad (3)$$

where M is the number of dimensions, λ_{mk} are the eigenvalues of the covariance matrix of the k-th cluster. Note that the closest *GI* value to zero suggests the best number of clusters. Different with the previous two indices, smaller value of *GI* is supposed to represent good clustering result and the K-cluster with minimum *GI* value is supposed to be the true number of clusters.

***Calinski Harabasz* (*CH*) index:** The *CH* [10] is given by

$$CH(K) = \frac{\text{trace}(B)/(K-1)}{\text{trace}(W)/(n-K)}, \quad (4)$$

where n_k is the number of memberships in the cluster k and n is the total number of the objects, and

$$\begin{aligned}\text{trace}(B) &= \sum_{k=1}^K n_k\|\boldsymbol{u}_k - \boldsymbol{u}\|^2\\ \text{trace}(W) &= \sum_{k=1}^K\sum_{i=1}^{n_k}\|\boldsymbol{x}_i - \boldsymbol{u}_k\|^2.\end{aligned}$$

2.2. Kernel Method

Recently, kernel-based clustering, which constructs a hyperplane to separate the linearly inseparable patterns, has attracted a lot of attention. These linearly inseparable patterns are nonlinearly transformed from a set of low-dimensional space into a higher-dimensional feature space to be linear separable [11]. At the core of the kernel-based clustering lies the difficulty of explicitly constructing the nonlinear mapping, which is sometime infeasible; but now it can be overcome by a kernel trick. The kernel trick is a way of mapping patterns from a input space into a feature space without having to compute the mapping explicitly, in the hope that the patterns will gain meaningful linear structure in the feature space, mathematically expressed as

$$\boldsymbol{\kappa}(\boldsymbol{x}_i,\boldsymbol{x}_j) = \Phi(\boldsymbol{x}_i)^T\Phi(\boldsymbol{x}_j), \quad (5)$$

where $(\cdot)^T$ is the transpose operator. Thus, a straightforward way to transform the calculation of Euclidean distance in the feature space into the kernel version is to use the kernel trick as follows

$$\begin{aligned}&\mathcal{D}_E^\kappa(\Phi(\boldsymbol{x}_i),\Phi(\boldsymbol{x}_j)) = \|\Phi(\boldsymbol{x}_i) - \Phi(\boldsymbol{x}_j)\|^2 \quad (6)\\ =\ &\|\Phi(\boldsymbol{x}_i)\|^2 + \|\Phi(\boldsymbol{x}_j)\|^2 - 2\Phi(\boldsymbol{x}_i)^T\Phi(\boldsymbol{x}_j)\\ =\ &\boldsymbol{\kappa}(\boldsymbol{x}_i,\boldsymbol{x}_i) + \boldsymbol{\kappa}(\boldsymbol{x}_j,\boldsymbol{x}_j) - 2\boldsymbol{\kappa}(\boldsymbol{x}_i,\boldsymbol{x}_j),\end{aligned}$$

and the kernel version of modified Pearson correlation is given by [12]

$$\begin{aligned}&\mathcal{S}_P^\kappa(\Phi(\boldsymbol{x}_i),\Phi(\boldsymbol{x}_j))\\ =\ &\frac{\Phi(\boldsymbol{x}_i)^T\Phi(\boldsymbol{x}_j)}{\sqrt{\Phi(\boldsymbol{x}_i)^T\Phi(\boldsymbol{x}_i)}\sqrt{\Phi(\boldsymbol{x}_j)^T\Phi(\boldsymbol{x}_j)}} \quad (7)\\ =\ &\frac{\boldsymbol{\kappa}(\boldsymbol{x}_i,\boldsymbol{x}_j)}{\sqrt{\boldsymbol{\kappa}(\boldsymbol{x}_i,\boldsymbol{x}_i)}\sqrt{\boldsymbol{\kappa}(\boldsymbol{x}_j,\boldsymbol{x}_j)}}.\end{aligned}$$

3. KERNEL VALIDITY INDICES

In this section, we extend above validity indices to four kernel-based validity indices, namely the kernel *DI* (*kDI*), the kernel *II* (*kII*), the kernel *GI* (*kGI*) and the kernel *CH* (*kCH*).

Kernel *DI* (*kDI*): The *kDI* is given by

$$kDI(K) = \min_{1\le i\le K}\left\{\min_{1\le j\le K}\left\{\frac{\delta^{\kappa}(\mathcal{C}_i,\mathcal{C}_j)}{\max_{1\le k\le K}\{\Delta^{\kappa}(\mathcal{C}_k)\}}\right\}\right\}, \tag{8}$$

where $\Delta^{\kappa}(\mathcal{C}_k)$ is the largest intra-cluster separation of cluster k in the feature space, $\delta^{\kappa}(\mathcal{C}_i,\mathcal{C}_j) = \min \mathcal{D}_E^{\kappa}(\mathcal{C}_i,\mathcal{C}_j)$ is the minimum of kernel-based Euclidean distance between cluster i and cluster j in the feature space. The kernel-based Euclidean distance between cluster i and cluster j is given by

$$\begin{aligned}\mathcal{D}_E^{\kappa}(\boldsymbol{u}_i^{\Phi},\boldsymbol{u}_j^{\Phi}) &= \|\frac{1}{n_i}\sum_{i=1}^{n_i}\Phi(\boldsymbol{x}_i) - \frac{1}{n_j}\sum_{j=1}^{n_j}\Phi(\boldsymbol{x}_j)\|^2 \\ = \quad & \frac{1}{n_i^2}\sum_{i=1}^{n_i}\sum_{i'=1}^{n_i}\boldsymbol{\kappa}(\boldsymbol{x}_i,\boldsymbol{x}_{i'}) + \frac{1}{n_j^2}\sum_{j=1}^{n_j}\sum_{j'=1}^{n_j}\boldsymbol{\kappa}(\boldsymbol{x}_j,\boldsymbol{x}_{j'}) \\ & -\frac{2}{n_i n_j}\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}\boldsymbol{\kappa}(\boldsymbol{x}_i,\boldsymbol{x}_j).\end{aligned} \tag{9}$$

Kernel *II* (*kII*): This index is expressed by

$$kII(K) = \left(\frac{1}{K}\times\frac{E_1^{\kappa}}{E_K^{\kappa}}\times D_K^{\kappa}\right)^P, \tag{10}$$

where $D_K = \min_{i,j}^{K}\|\boldsymbol{u}_i^{\Phi}-\boldsymbol{u}_j^{\Phi}\|_2$, $E_1^{\kappa} = \sum_j \|\Phi(\boldsymbol{x}_j)-\boldsymbol{u}^{\Phi}\|_2$ and $E_K^{\kappa} = \sum_{k=1}^{K}\sum_{j\in\mathcal{C}_k}\|\Phi(\boldsymbol{x}_j)-\boldsymbol{u}_k^{\Phi}\|_2$, where

$$\begin{aligned}\|\Phi(\boldsymbol{x}_j)-\boldsymbol{u}^{\Phi}\|_2 &= \boldsymbol{\kappa}(\boldsymbol{x}_j,\boldsymbol{x}_j) - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\kappa}(\boldsymbol{x}_j,\boldsymbol{x}_i) \\ &\quad + \frac{1}{N^2}\sum_{i}\sum_{i'}\boldsymbol{\kappa}(\boldsymbol{x}_{i=1}^{N},\boldsymbol{x}_{i'=1}^{N})\end{aligned} \tag{11}$$

Kernel *GI* (*kGI*): The *kGI* can be easily obtained by

$$kGI(K) = \max_{1\le k\le K}\left\{\frac{(2\sum_{m=1}^{M}\sqrt{\lambda_{mk}})^2}{\min_{1\le j\le K}\|\boldsymbol{u}_k^{\Phi}-\boldsymbol{u}_j^{\Phi}\|_2}\right\}. \tag{12}$$

Kernel *CH* (*kCH*): The *kCH* is given by

$$CH(K) = \frac{\text{trace}^{\kappa}(B)/(K-1)}{\text{trace}^{\kappa}(W)/(n-K)}, \tag{13}$$

where

$$\begin{aligned}\text{trace}^{\kappa}(B) &= \sum_{k=1}^{K} n_k\|\boldsymbol{u}_k^{\Phi}-\boldsymbol{u}^{\Phi}\|^2 \\ \text{trace}^{\kappa}(W) &= \sum_{k=1}^{K}\sum_{i=1}^{n_k}\|\Phi(\boldsymbol{x}_i)-\boldsymbol{u}_k^{\Phi}\|^2.\end{aligned}$$

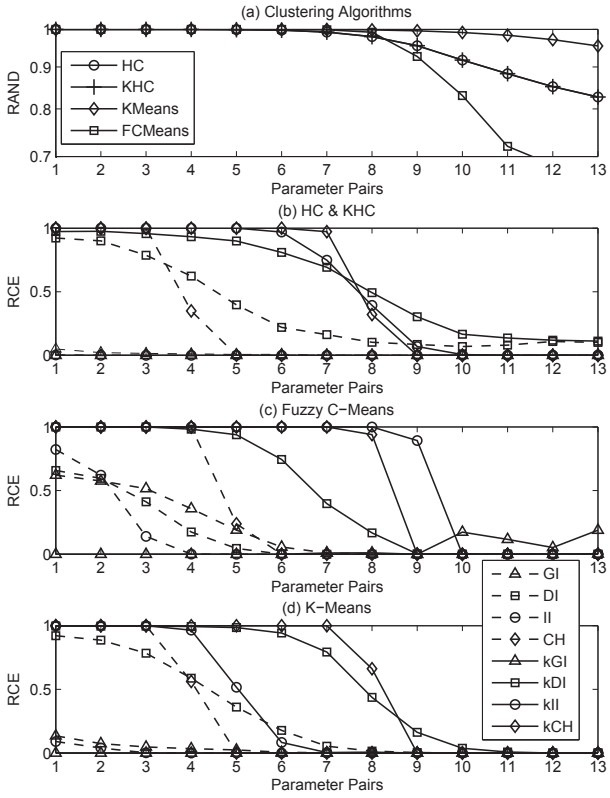

Fig. 1. (a) RAND index for four clustering algorithms. (b), (c) and (d) show the RCE of the number of clusters of all four conventional validity indices and their kernel counterparts for the clustering results of HC&KHC, FCM and k-means, respectively

4. NUMERICAL RESULTS

In this section, we present the numerical comparison of both conventional and kernel validity indices to validate four different clustering algorithms, namely k-means, HC, fuzzy c-means and kernel HC (KHC). We employ complete linkage for both hierarchical algorithms. To conduct a monte-carlo simulation to obtain statistical steady results, we employ the method in [13] to generate a number of synthetic gene expression datasets with 500 synthetic genes in each dataset and 24 samples for each gene. These 500 genes locate in $K = 5$ clusters and each cluster has 100 members. The model of cyclic gene expression is given by

$$x_{ij} = r + [a+br](r + [a+br]\sin(2\pi j/8 - \omega_i + cr), \tag{14}$$

where x_{ij} is the expression value of the i-th gene at the j-th time point, each instant of r is an independent random number from the standard normal distribution $\mathcal{N}(0,1)$, a controls the magnitude of the sinusoid and it is fixed to three here, b controls the random component added to the magnitude, c controls the random component added to the

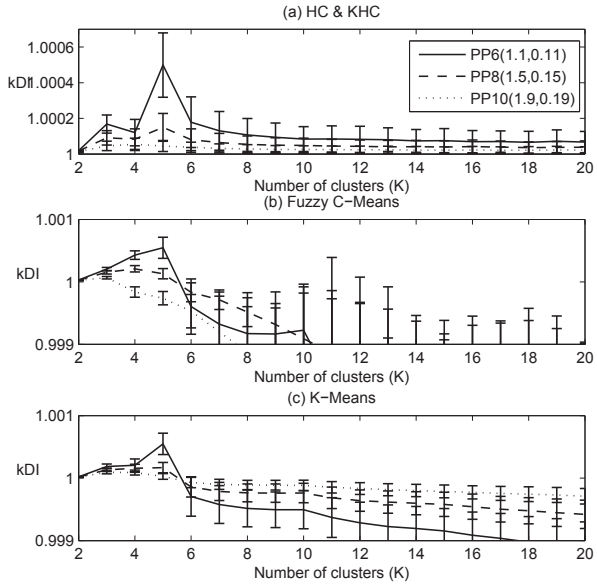

Fig. 2. Index values of the *kDI* against the number of clusters K in four noise levels, from low to high, corresponding to parameter pairs PP6, PP8 and PP10.

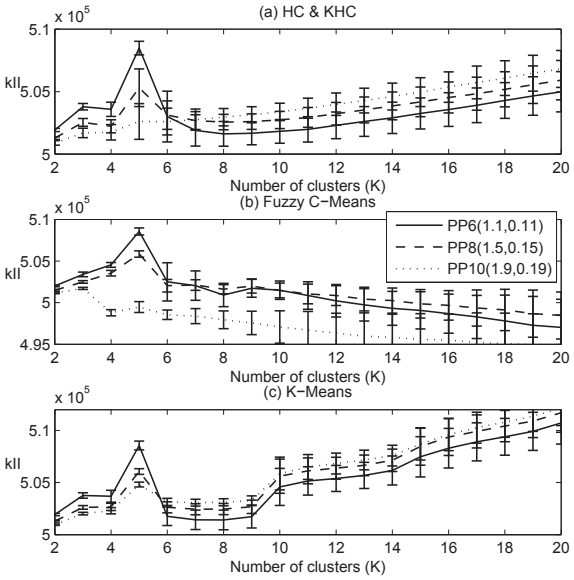

Fig. 3. Index values of the *kII* against the number of clusters K in four noise levels, from low to high, corresponding to parameter pairs PP6, PP8 and PP10.

phase and ω_i is the phase shift of the i-th gene. ω_i will determine which cluster the gene i will be in. Since the noise in this model is not additive, we have to couple b and c to be a pair and raise the both values to change the noise power. With the increasing of b and c, the noise power increases. The paired parameters are listed as $(b, c) \in \{(0.1, 0.01), (0.3, 0.03), (0.5, 0.05), (0.7, 0.07), (0.9, 0.09), (1.1, 0.11), (1.3, 0.13), (1.5, 0.15), (1.7, 0.17), (1.9, 0.19), (2.1, 0.21), (2.3, 0.23), (2.5, 0.25)\}$, thus, there are 13 parameter pairs (PPs) from PP1 to PP13 representing 13 noise levels from low to high. For each pair of parameters, we generate 1000 datasets, and subsequently, we get 1000 clustering results for each clustering algorithm.

Since the clustering validity indices have to work in an unsupervised situation, to "validate" these validity indices, we have to make use of the ground truth of the datasets, in this case, which is the nature clustering including the number of clusters and the membership of each cluster. In Fig. 1 (a), we calculate the RAND index [14] for four clustering algorithms based on the ground truth. It is worth noting that HC and KHC have exact same results in our simulation. Another fact worthy of note is that the k-means is the best algorithm out of the four that we evaluate. We will make use of this fact to "validate" the validity indices to show us which validity index would perform best in a statistical sense. It is logical to deduce that the best index will also work well in the similar type of dataset when the ground truth is not available. To illustrate the effectiveness of validity indices, we compare the validation results based on two experiments. On one hand, we compare the rate of correct estimation (RCE) of the number of clusters for all validity indices; on the other hand, to illustrate how good the clustering algorithms are, we compare the values of a given index for all clustering algorithms, the largest value indicates best clustering expect that the *GI* and the *kGI* are looking for smallest index values.

In Fig. 1 (b), (c) and (d), we show the RCE of the number of clusters of all four conventional validity indices and their kernel counterparts for the clustering results of the HC&KHC, the FCM and the k-means, respectively. Generally speaking, the kernel validity indices, which are shown with solid lines, have better estimation performance than the conventional ones, except the *kGI*. In this case, the *GI*, the *kGI* and the *II* are the most inferior three indices while *kDI*, *kII* and *kCH* are the most superior three indices. Let us look closer at the *kDI*, the *kII* and the *kCH*: for the *kDI*, the performance is moderate, not so good and not so bad. There are always some estimation errors for HC&KHC in some low noise cases where hundred percent data points are correctly clustered. We can analyse the results together with the results in Fig. 2, which depicts the error plots of index values of the *kDI* against the number of clusters K in three noise levels, from low to high, corresponding to parameter pairs PP6, PP8 and PP10. Fig. 2 (a), (b) and (c) illustrate both the means and the standard deviations of the index values of the HC&KHC, the FCM and the k-means, respectively. We can tell that the estimation errors result in the large standard deviations.

Similarly, for the *kII*, we analyse the results in Fig. 1 together with the results in Fig. 3, which depicts the error plots of index values of the *kII* against the number of clusters K in three noise levels. There is an interesting discovery in Fig. 1:

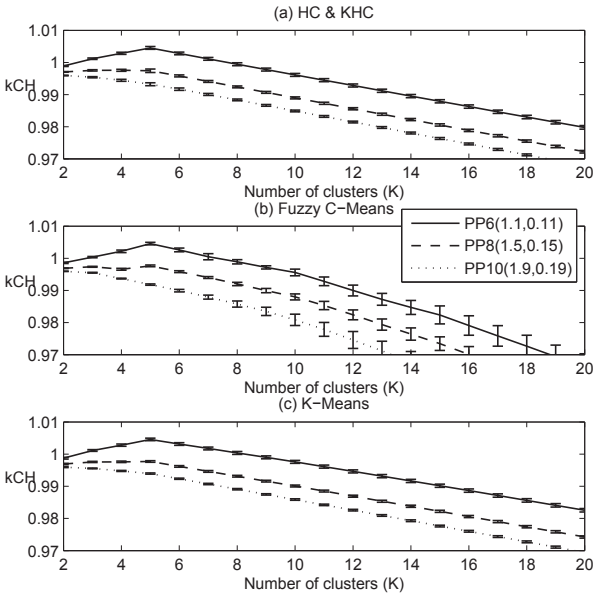

Fig. 4. Index values of the *kCH* against the number of clusters K in four noise levels, from low to high, corresponding to parameter pairs PP6, PP8 and PP10.

the k-means has been endorsed by RAND to be the best in this case, while *kII* shows a contrary result that the RCE of number of clusters for the k-means is poor, while the RCE for the FCM is pretty good. We also notice that, at PP9, the RCE for the FCM is quite high while the RCE for the k-means is zero, but the fact is that the RAND of the k-means is higher than that of the FCM. Note that in Fig. 3 (b) and (c), which depict the index values of FCM and k-means, the index values at K=5 are similarly around 5.08×10^5 for both the FCM and the k-means, but the index values of the k-means at K=19 and 20 are higher than 5.08×10^5. It means that *kII* indicates that the clustering results of the k-means with the number of clusters 19 and 20 are better, which is obviously wrong. Based on this, we can conclude that the *kII* is not reliable.

The results shown in Fig. 1 (b), (c) and (d) indicate that the *kCH* has stable and superior performance in our simulation. It can achieve high estimation performance until the parameter pair PP7 corresponding to (b, c) of $(1.3, 0.13)$. In Fig. 4, it is worthy noting that the standard deviations are much smaller than other indices. Thus, the *kCH* is the most reliable and stable index out of the evaluated eight indices.

5. DISCUSSIONS AND CONCLUSIONS

In this paper, we developed and presented four kernel validity indices, namely the *kDI*, the *kII*, the *kCH* and the *kGI* from their conventional counterparts, corresponding to the *DI*, the *II*, the *CH* and the *GI*. We conducted a Monte-Carlo simulation using synthetic gene expression model to evaluate the validity indices and compare their results in order to find the best index which likely works well in the similar type of dataset when the ground truth is not available. In the numerical results, we showed that the *GI*, the *kGI* and the *II* are the most inferior three indices while the *kDI*, the *kII* and the *kCH* are the most superior three indices. Among the most superior three indices, the *kDI* has moderate performance, the *kII* is found not to be reliable, and most importantly, the *kCH* is the most reliable and stable index out of the evaluated eight indices in our study.

6. REFERENCES

[1] R. Xu and D. II Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Networks*, vol. 16, no. 3, pp. 645 – 678, 2005.

[2] D. X. Jiang, C. Tang, and A. D. Zhang, "Cluster analysis for gene expression data: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 11, pp. 1370–1386, 2004.

[3] A. K. Jain, R. C. Dubes, "Algorithms for Clustering Data," Prentice-Hall, Englewood Cliffs, NJ, 1988.

[4] F. Höppner, F. Klawonn, and R. Kruse, "Fuzzy Cluster Analysis: Methods for Classification, Data Analysis, and Image Recognition," Wiley, New York, 1999.

[5] M. Halkidi, Y. Batistakis and M. Vazirgiannis "Cluster validity methods: Part I," *SIGMOD*, Record 31(2) pp. 40-45, 2002.

[6] M. Halkidi, Y. Batistakis and M. Vazirgiannis "Cluster validity methods: Part II," *SIGMOD*, Record 31(3) pp. 19-27, 2002.

[7] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters," *J. Cyber.*, vol. 3, no. 3, pp. 32–57, 1973.

[8] U Maulik and S Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1650–1654, 2002.

[9] Benson S. Y. Lam and Hong Yan, "Assessment of microarray data clustering results based on a new geometrical index for cluster validity," *Soft Computing*, vol. 11, no. 4, pp. 341–348, 2007.

[10] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics - Theory and Methods*, vol. 3, no. 1, pp. 1–27, 1974.

[11] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Networks,* , vol. 12, no. 2, pp. 181 –201, 2001.

[12] Jie Qin, Darrin P. Lewis, and William Stafford Noble, "Kernel hierarchical gene clustering from microarray expression data," *Bioinformat.*, vol. 19, no. 16, pp. 2097–2104, 2003.

[13] L. P. Zhao, R. Presntice, and L. Breeden, "Statistical modelling of large microarray data sets to identify stimulus-response profiles," *PNAS*, vol. 98, no. 10, pp. 5631–5636, May 2001.

[14] W. M. Rand, "Objective criteria for the evaluation of clustering methods". *Journal of the American Statistical Association*, Vol. 66, No. 336, Dec., 1971