Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
`http://siba-ese.unisalento.it/index.php/ejasa/index`
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v12n3p657

Voting-based approach in consensus clustering through *q*-fold cross-validation

By Shamsuddin, Mahat

Electronic Journal of Applied Statistical Analysis
Vol. 12, Issue 03, November 2019, 657-673
DOI: 10.1285/i20705948v12n3p657

Voting-based approach in consensus clustering through q-fold cross-validation

Norin Rahayu Shamsuddin[a] and Nor Idayu Mahat[b]

[a] *Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA, Malaysia*
[a,b] *School of Quantitative Sciences, Universiti Utara Malaysia, Malaysia*

Over the past 50 years, extensive research have been carried out to understand how clustering work in classifying data into meaningful groups. Various clustering algorithms and cluster validity indexes have been proposed and improvised to obtain the *best* clustering result. However, there is no clustering method that is able to give consistent results on similar structure of a dataset. An alternative mechanism to control the variation of results and improved the quality of traditional clustering is through consensus clustering. In this paper, we generate multiple partitions of consensus clustering through a resampling method by employing q-fold cross-validation approach. q-fold cross-validation approach is able to speed-up the consensus partitions procedure with q^{th} iterations. To encounter with different number of cluster labels occur in the partitions, we employed voting-based method in the second stage of consensus clustering to obtain optimal consensus partition. The performance of optimal consensus partitions is evaluated from Silhouette plot.

keywords: consensus clustering, resampling, k-medoids, optimal consensus partition, voting-based method).

1 Introduction

Clustering goal is to discover structure of dataset based on observations' similarity to form a group. The structure of clustering is represented as a subset $C = C_1, C_2, \ldots, C_k$ of $X = x_1, x_2, \ldots, x_n$, in which $X \bigcup_{i=1}^{k} C_i$ and $C_i \cap C_j = \emptyset$ for $i \neq j$. This traditional clustering approach requires data to be in one and only one cluster. One of the traditional clustering approach is known as centroid-based clustering such as k-means, k-medoids,

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

k-modes, k-median well as k-prototypes. The objective function of centroid-based clustering is to minimize the intra-cluster distance based on iterative relocation between objects and its cluster center (know as centroid in k-means and medoids in k-medoids).

A considerable amount of literature has been published on centroid-based clustering as seen in Park and Jun (2009); Arbin et al. (2015); Xie et al. (2011); Madhuri et al. (2014); Anderson et al. (2006); de Assis and de Souza (2011); Zhong et al. (2017) and not limited to it. However, the results from clustering indicate variations in the output. It has been observed that traditional clustering methods suffers few major shortcomings;

- effect of random initialization center, and
- number of cluster has to be determine in prior.

The drawbacks had engaged researchers to carried out extensive research on these matter (Bubeck et al., 2009; Celebi et al., 2013; Liu et al., 2016; Raykov et al., 2016). Yet, the result is not promising.

Over the past ten years, a number of researchers have sought for alternatives to overcome the drawback of traditional clustering and further asses the stability of clustering. Consensus clustering gives beyond than a single clustering can provide to researchers. The basic concepts is to combine different run across algorithm (input) to obtain single output which is more *reliable* (Goder and Filkov, 2008; Yang, 2016).

An alternative to this issue is to conduct multiple clustering on the same algorithm or different algorithm on the same dataset to obtain optimal single results. This approach fomerly known as consensus clustering is able to control (or minimize) the diversity of results from clustering algorithms regardless of the structure of the clusters (such as in k-means). The clustering results tend to be more robust and consistent (Fred and Jain, 2005; Nguyen and Caruana, 2007; Șenbabaoğlu et al., 2014; Lancichinetti and Fortunato, 2012) and improve the quality of clustering.

According to Minaei-Bidgoli et al. (2004); Fred and Jain (2002b) the construction of multiple partition in consensus clustering can be achieved by:

1. combining result of different clustering algorithm,
2. resampling data to produce different results,
3. run clustering algorithm multiple times with different parameter or initializations of a clustering algorithm, and
4. using different clustering algorithms.

The issue here is how to obtain the optimal consensus partition when there exist different results of labeling, $\mathfrak{L}$ in each partition?

In this paper, we are interested in determining the $\mathfrak{L}$ for the objects in a dataset for the optimal consensus partitions. Here, we employed resampling approach using q-fold cross-validation to generate multiple consensus partitions. Through voting-based mechanism, we able to encounter with different clustering formation which associate with different allocation of objects in different partitions. More importantly with q-fold cross-validation, this procedure may to speed-up the consensus partitions procedure with q^{th} iterations. Moreover, our study may guide in indicate misclassifcation of objects in cluster.

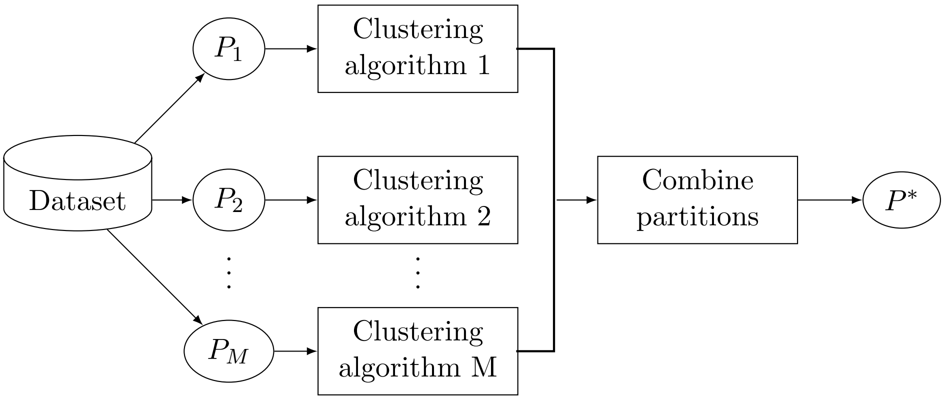
2 Some Insight of Motivational Approach

A number of clustering algorithms been developed, yet there is no single algorithm capable in finding the best clustering results. Even the structure of clusters changes in each run for the same dataset. Therefore, by introducing different results through multiple run and combine them to generate single optimal result as in consensus clustering able to produce a stable and quality clustering algorithm. This approach is a powerfull tool which give beyond than a single clustering can provide to users.

Consensus clustering also known as aggregation of clustering, has become an important tools for clustering to improve the quality of clustering algorithm (Monti, 2003; Topchy et al., 2005; Vega-Pons and Ruiz-Shulcloper, 2011). The idea of consensus clustering is to create the final clustering that encompassed all the information obtain by combining different clustering results. The two underlying processes of consensus clustering are

1. partition generation and
2. consensus function

as illustrated in Figure 1. The first stage is the generation of multiple partitions while in the second stage, the final optimal result from the combination results in the first stage is obtained.



Stage 1: Partition construction and clustering Stage 2: Function construction

Figure 1: The anatomy of consensus clustering

The main advantages of consensus clustering has been explained in Topchy et al. (2004, 2005); Fred and Jain (2005); Vega-Pons and Ruiz-Shulcloper (2011) in term of

- robustness - better average performance than single clustering.
- novelty - solution from consensus is unttainable by single clustering algorithm.
- consistency - result from combination produce as almost similar to all combined single clustering.
- stability - less sensitivity to noise and outliers.

For more than a decade,consensus clustering have been employed to improve the objective function of clustering (von Luxburg, 2010), clustering stability (Kuncheva and Vetrov, 2006; Ben-David et al., 2006; Ben-david et al., 2007; Kiselev et al., 2016) determine the number of clusters (Wang, 2010), validation of internal and/or external indices (Rendón et al., 2011; Novoselova and Tom, 2012).

2.1 Resampling approach

A considerable amount of literature has been publish on resampling data to generate multiple partition in consensus clustering. The construction of resampling approach has been documented by Milligan and Cooper (1987) in 1987. Since then, severeal studies acquire this approach to encounter the variation in clustering. According to Şenbabaoğlu et al. (2014) resampling approach is capable in assessing cluster stability.

Clest method has been proposed by Dudoit and Fridlyand (2002) to determining the number of clusters in microarray dataset. It is a predictive-based resampling technique that randomly assigned data into non-overlapping learning and test set to validate cluster assigments for individual samples. Dudoit and Fridlyand (2003) extended their clusters to further estimate accuracy of a partitioning clustering method by introducing new sampling techniques. Their ground is to assess the confidence of cluster assignments for individual observations.

Dresen et al. (2008) introduce a resampling method based on continuos weights. The methodology is similar to the boostrap, however the information regarding the whole observation is replace with random floating-point number larger than zero. The idea is to weight the observations by the random number to enable each of the observation represented in resampled dataset. Thus, the proposed method is used to evaluate the stability of clustering.

In obtaining the consistency in the number of clusters, Wang (2010) carried out an experiment through modified leave-many-out crossvalidation scheme. Their interest is to select the number of cluster that minimize the algorithm's instability, thus give robustness for clustering algorithm. The methodology involved splitting dataset into three parts - two parts of training set and left out a validation set.

Risso et al. (2018) has set up an R package for users to set up multiple pactition based on resampling approach known as *Resampling-based Sequential Ensemble Clustering* (RSEC) under `clusterExperiment` package. This package is developed to encounter clustering of of single-cell and other large gene expression datasets. It give a flexible to users *manipulate* the parameters, intergration of resampling and provide method for final consensus clustering.

2.2 Voting-based Algorithm

Dudoit and Fridlyand (2003) and Wang (2010) have incorporated resampling approaches with voting-based methods in their study. Dudoit and Fridlyand (2003) introduce *plurality* voting - assigning the label on individual observation based on majority votes. While, Wang (2010) overcome the varitions at partition generation through cross-validation with voting. The concepts is to select the algorithm that minimized the instability at generation partition based on mode of the number of time the algorithm is selected.

Fred and Jain (2002b,a, 2005) implement voting based mechanism in evidence accumulation clustering (EAC) to map data partitions into a new similarity matrix between patterns. The voting process determine the occurences of objects in pairs occur in the same cluster. The final partition is obtained by applying the single-link method on new similarity measures.

Meanwhile Saeed et al. (2012) explored the use of voting mechanism using different dissimilarity measures of chemical structures. They introduced cumulative voting-based aggregation algorithm (CVAA) with concerned to search an optimal label that represent the partition that summarised esemble partition. Nevertheless CVAA does not retained the same relabelling result since the final partition depend solely on arrangement of partition. To overcome this weakness, an addaptive cumulative voting-based aggregation algorithm (A-CVAA) has been proposed (Saeed et al., 2013).

3 Methodology

3.1 Proposed algorithm

Given an unlabelled dataset $Y = \{y_1, y_2, \dots, y_n\}$ with n objects in $\mathbb{d}$ dimensions space. $\mathbb{P} = \{P_1, P_2, \dots, P_M\}$ is a set of partitions where Y is breakdown into q^{th} fold of train and test sets. Each $P_q = \{\omega_1^q, \omega_2^q, \dots, \omega_{k_i}^q\}$ is a partition of the set of objects Y with k_i clusters or labels obtained through clustering algorithm with random initialization and different parameter k.

To cope with different number of labels produce at each partition, the final $\mathfrak{L}$ is obtained through voting-based mechanism. The optimal label is then assigned to each objects $\mathfrak{L} = \{y_1^{\mathfrak{L}}, y_2^{\mathfrak{L}}, \dots, y_n^{\mathfrak{L}}\}$ with respect to internal cluster indices. The voting mechanism concept applied in this paper by count the number of times the label is selected in each partition and select the one with the most count. The proposed method is is describe in the following Algorithm 1.

The $\mathfrak{L}$ for each objects in P_q is acquired from internal validity indices. For this paper we opt for three types of indices which are; (i) generalized Dunn index 33 (GDI33) (ii) Generalized Dunn index 53 (GDI53), and (iii) Silhouette index.

3.2 Labeling assessment

One of the most well know tools used to validate clustering strength and number of clusters is Silhoutte index (Rousseeuw, 1987) and the maximum value of the index is

```
Input:
    Dataset            dataset $Y$ of $n$ objects in $\mathbb{d}$ dimensions space
    Resampling         breakdown of $Y$ into $q$-fold partitions
    Clustering         $k$-medoids algorithm to determine label for objects
Output:
    Vote               optimal result for labeling through voting-based approach
    Misclassification  identify misclassification in optimal partitions
1: For $q = 1$ to $q_f - 1$:
     (i) Execute clustering algorithm for based partitions using $q$-fold
         cross-validation.
     (ii) Generate label from internal cluster index for Step (1)(i)
2: Determine majority voting as optimal partition for labeling process.
     (i) allocate the label to each objects in optimal partitions.
     (ii) employ Silhouette plot/index to identify misclassification of objects.
3: Determine the accuracy of labeling in $q - 1$ fold.
```

Algorithm 1: Voting -based mechanism in labeling data points

used to determine the label

$$Sil = \frac{\sum_{i=1}^{n} S(i)}{n}, \text{ Silhouette} \in [-1, 1] \tag{3.1}$$

where

- $S(i) = \dfrac{b(i) - a(i)}{max\{a(i); b(i)\}}$
- $a(i)$ = average dissimilarity of i to all other objects of A
- $d(i, C)$ = average dissimilarity of i to all other objects of C
- $b(i) = \min\limits_{C \neq A} d(i, C)$

GDI is an alternative for traditional Dunn Index introduce by Bezdek and Pal (1998) which is much influence by outliers. A standard GDIs' index measure the diameter of cluster with combination of dissimilarity of interclusters with six different measure of intercluster separation and three quantification of intracluster compactness. GDI is reprented as $\text{GDI}_{\delta,\Delta}$ where δ determine the intercluster distance and Δ estimate intracluster distance. The maximum value of the index determine the label for the objects.

Bezdek and Pal (1998) has carried out extensive experiments on each of intercluster and intracluster distance (6×3 variant) and they point out that δ_3 and δ_5 are more reliable measure for intercluster distance. Furthermore, from extensive experiment with the combination of δ, Δ, they point out that the combination of GDI_{33} and $GDI53$ give better results in determine the clusters.

δ_3 measure the average of all points between clusters and δ_5 is a combination of measurement of δ_3 and distance between the centroid clusters. $\Delta(A)$ is the standard diameter of set A and $\Delta_3(A)$ measure all points in the cluster that measure from the center (average) of diameter cluster (Bezdek and Pal, 1998). The GDI is define as:

$$GDI = \min_{1 \leqq a \leqq c} \left\{ \min_{1 \leqq b \leqq c} \left\{ \frac{\delta_i(Y_a, Y_b)}{\max\limits_{1 \leqq k \leqq c} \{\Delta_j(Y_k)\}} \right\} \right\} \tag{3.2}$$

where δ_i and $\Delta(A)$

- $\delta_3(A, B) = \dfrac{1}{|A||B|} \sum\limits_{\substack{x \in A \\ y \in B}} d(x, y); \quad |\cdot|$ is the total data in cluster A

- $\delta_5(A, B) = \dfrac{1}{|A| + |B|} (\sum\limits_{x \in A} d(x, g_B) + \sum\limits_{x \in B} d(y, g_A)); \quad g_A = \frac{1}{|A|} \sum\limits_{x \in A} x$ and $g_B = \frac{1}{|B|} \sum\limits_{y \in B} x$

- $\Delta_3(A) = \frac{2}{|A|} \sum d(x, \overline{g_A})$

3.3 Data sets

The simulation datasets were generated using `MixSim` package (Melnykov et al., 2012) with variation in datapoints, outliers, overlapped, and dimension. We generated five different condition of simulation data and information on the dataset is describe in Section 4. For dataset with outliers (SD3), it is important for us to understand how the structure and relabelling of objects is being determined with and without outliers. Since clustering is much infleunced by outliers, therefore we intend to create another dataset that removed the outliers through Mahalanobies distance for SD3 dataset. For real dataset, we acquired the data through UCI Machine Learning Repository (Dheeru and Karra Taniskidou, 2017).

4 Monte Carlo Study

We demonstrate the performance of the proposed method on the simulation datasets of equal mixing proportions of Gaussian distribution with different level overlap. The overlapped allow us to control the degree of interaction between mixture component. We set the simulation datasets to into few condition and he summary of the datasets as in Table 1:

(a) SD1 - mixture model with heterogenous components with ellipsoidal covariance matrices. No outliers are produced and there is almost no overlapped between mixture components;

(b) SD2 -difference from SD1 in Ω as well as $\mathbb{d}$ and maximum and average overlaps in dataset;

(c) SD3 - as SD1 except the Ω and $\mathbb{d}$ with maximum overlap and outliers/noise in dataset;

(d) SD4 - different from SD1 in term of Ω and $\mathbb{d}$ with average overlap in dataset; and

(e) SD5 - as SD1 except the Ω and $\mathbb{d}$ with maximum overlap and noise in dataset.

Table 1: Description of datasets

	Simulation datasets informations				
Dataset	Ω	$\mathbb{d}$	Overlapped (B/M)	n	noise/ outliers
SD1	5	2	B = 0.00001	500	Null
SD2	7	6	M = 0.05, B=0.01	400	Null
SD3	3	6	M = 0.01	350	30
SD4	4	5	B = 0.001	300	Null
SD5	3	4	M = 0.001	200	10

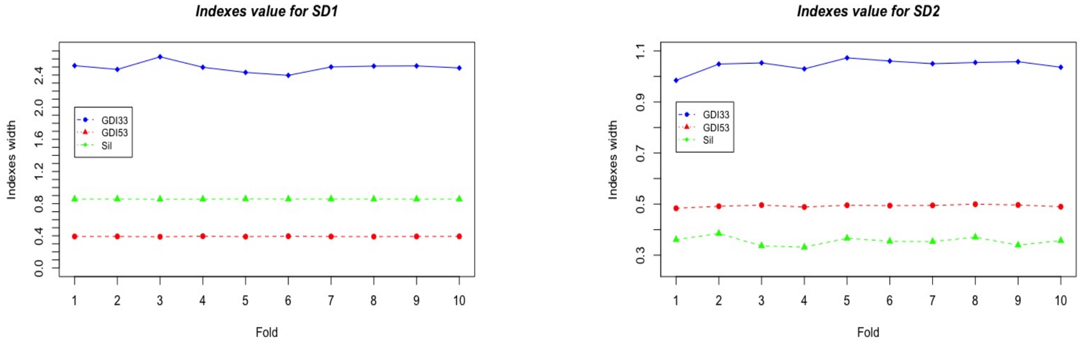
Ω = no. of component; $\mathbb{d}$ = dimension; B = average overlapped; M = maximum overlapped; n = sample size include noise/outliers.

For the clustering approach we employed k-medoids algorithm with Gower distance available in R. The random initilization was employed and the parameters were set to be between $k = 2$ and 10. The labeling value from voting-based mechanism as summarised in Table 2 for three indices namely GDI_{33}, GDI_{53} and Sil index. The trend of three indices based for selected $\mathfrak{L}$ is represented in Figure 2.

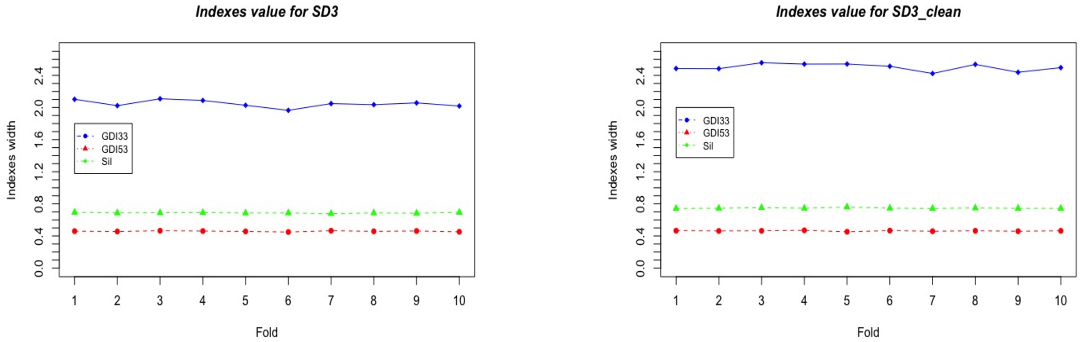
For the representation of object allocation in optimal partition, we opt for silhouette information since it can measure how well an individual belongs to its assigned cluster. We make comparison between original dataset as reference (von Luxburg, 2010) with clustering obtained from q-fold cross-validation. The idea is to look at improvement of performances obtain in consensus clustering.

4.1 Overlapping dataset

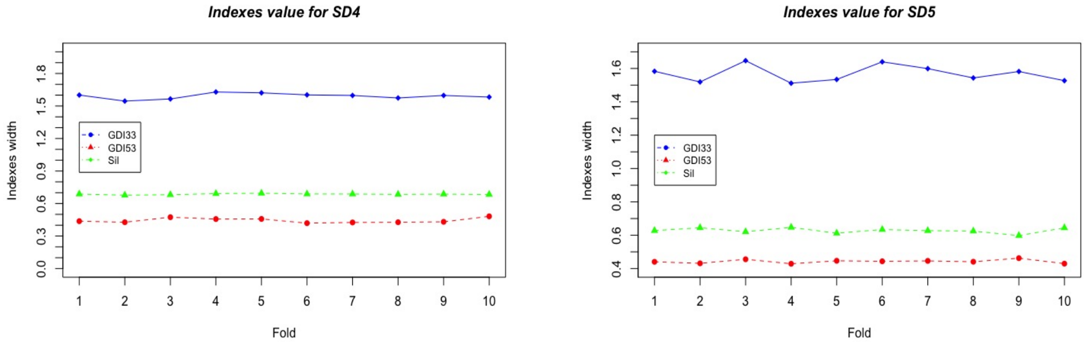
Overlapping dataset usually are unable to be divided clearly by k-medoids algorithm for SD2 and SD3. In Figure 3, there is a clear complexity of clustering in SD2 . From the Silhoutte plot, we identify a misclassification in labeling on quit number of objects (Silhouette value approaching -1) concurrent with overlapping of objects (Silhouette



(a) SD1 (b) SD2



(c) SD3 with outliers (d) SD3 without outliers



(e) SD4 (f) SD5

Figure 2: Indexes width of three indices GDI_{33}, GDI_{53} and Silhouette on simulation datasets based on selected $\mathfrak{L}$ for (a) SD1, (b) SD2, (c) SD13 with outliers, (d) SD3 without outliers, (e) SD4 and (f) SD5.

Table 2: Voting-based mechanism for labeling via ICV

		Labeling of ICV		
Simulation Datasets	n	GDI_{33}	GDI_{53}	Sil
SD1	500	5	2	3
SD2	400	2	2	3
SD3	350	2	2	2
SD3_clean	327	2	2	2
SD4	300	2	4	2
SD5	200	2	2	2

value almost 0) in two and four clusters as display in Figure 4. Since the overlapping of data is quiet large, it is no doubt we obtain such of results.

There is no much different between reference in clustering result between reference datatset and SD2 dataset in $\mathfrak{L} = 2$. However, some differences were observed for $\mathfrak{L} = 4$ in both reference and consensus clustering. Reference dataset identified quiet number of overlapping objects occur between clusters 3 and 4 and few objects are misclassified. While SD2 applied for consensus clustering identify one object misclassified in cluster 1 and 2 as well as a few objects in cluster 3. Quiet a number overlapping objects occur between clusters 3 and 4 and few objects are misclassified in consensus clustering.

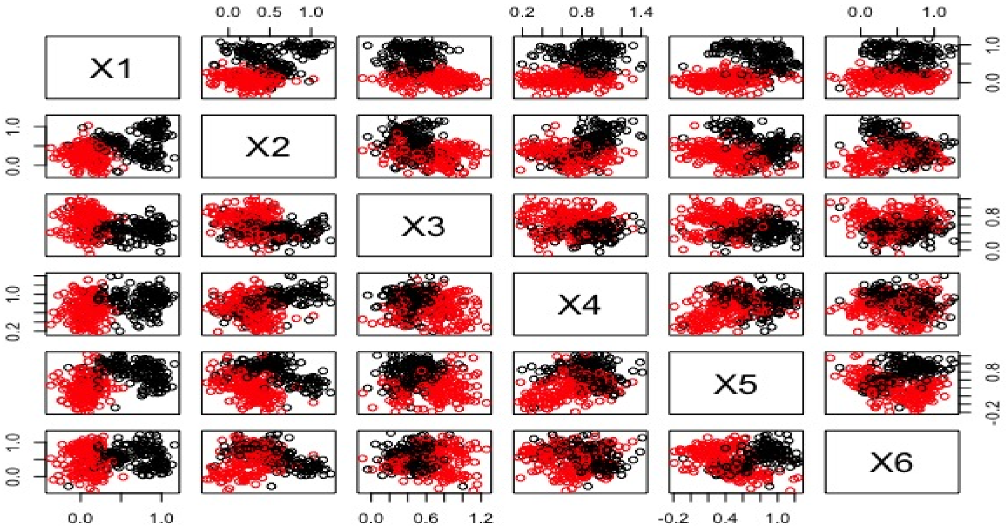
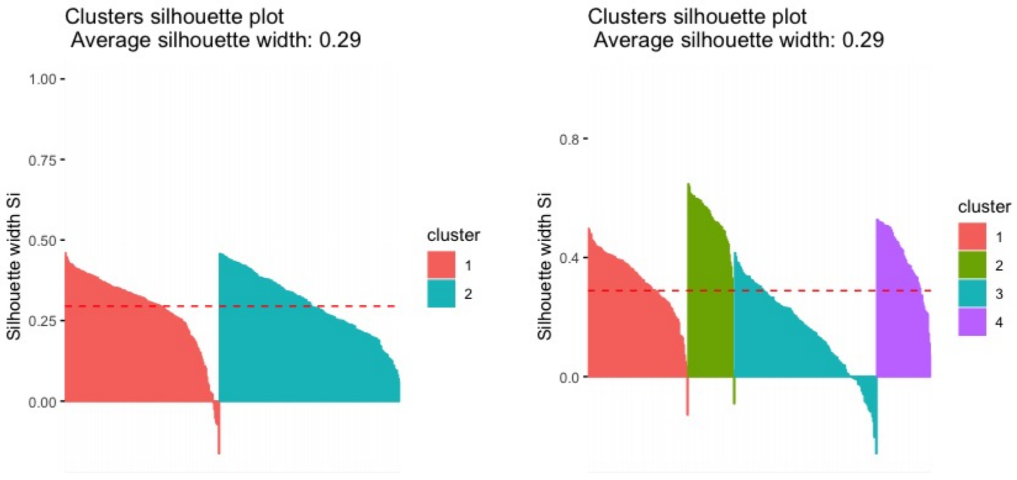


Figure 3: Redundancy of cluster partition



(a) SD2 with $\mathfrak{L} = 2$

(b) SD2 with $\mathfrak{L} = 4$

Figure 4: Silhouette plot identify misclassification of objects in a cluster in overlapping dataset.

Compared with SD3, the overlapping setting were set lower than SD2 (almost no overlapping setting). Still, smaller number of objects were identify being misclassified in dataset SD3 with outliers (result from Silhouette plot not include). Another set ot SD3 wihtout outliers give a better result than SD3 with outliers. Silhouette plot and index value indicate less than five objects were overlapped between clusters.

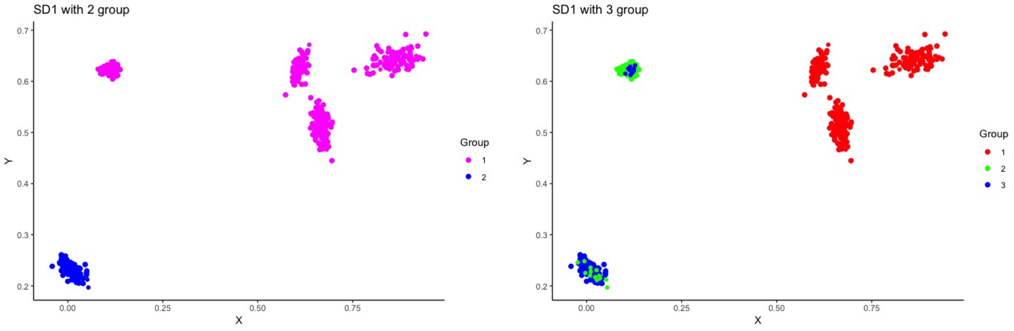
4.2 Non overlapping dataset

For dataset with non-overlapping and zero outliers, SD1 revealed three different results for labeling the objects based on three different indices - GDI$_{33}$, GDI$_{53}$ and Sil (see Table 2). The average silhoutte width estimate for $\mathfrak{L} = 2$, $\mathfrak{L} = 3$, $\mathfrak{L} = 5$ is 0.73, 0.77 and 0.84 respectively.

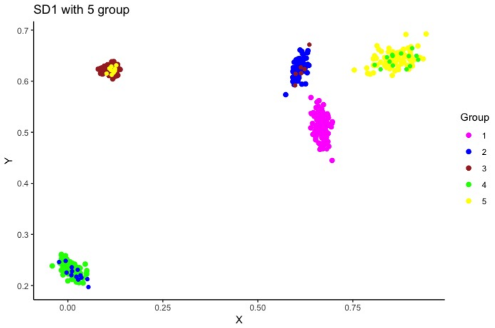
To our best judgement, the labelling for DS1 should be appointed to $k = 5$ as it really minimize intracluster distance while maximizing intercluster separation. GDI$_{33}$ display the characteristics of k-medoids algorithm is to assign each point to it nearest medoid (representative object) while maximizing the distance between clusters. Figure 5 presents clustering results in train and test set for SD1 dataset.

We identify a misclassification in $\mathfrak{L}$ with five groups in single algorithm for SD1 dataset which should be assign to group three. However this situation does not occur in cross-validation approach. q-fold cross-validation may divide the noise data point into its correct partition which does not applicable in reference group.

Closer inspection indicate no misclassifications capture in Silhouette plot for SD4 and SD5 display in Figure 6. GDI$_{53}$ give slightly a better performance in indicating the label of objects in SD4.



(a) SD1 with $\mathfrak{L} = 2$ (b) SD1 with $\mathfrak{L} = 3$



(c) SD1 with $\mathfrak{L} = 5$

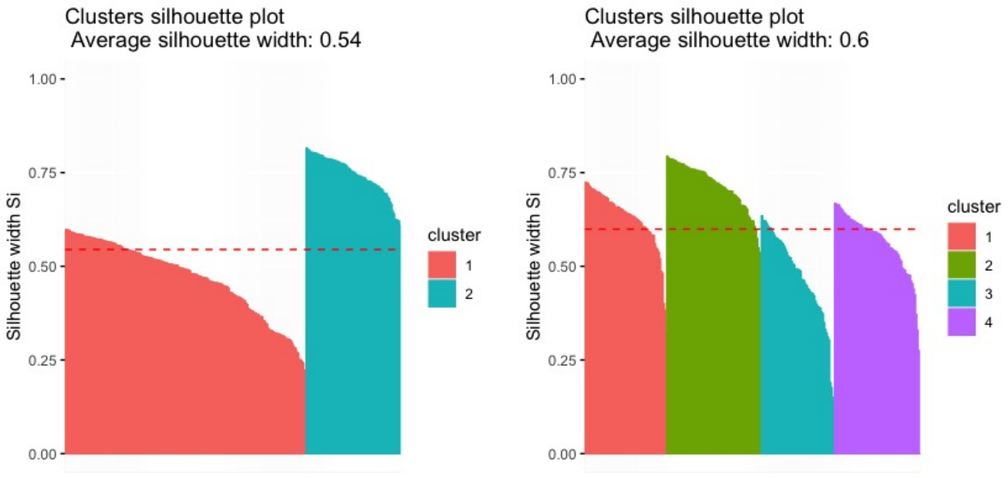
Figure 5: Different clustering result result in well-separated distribution based on ICV

Based on overall performance between test and training set, Sil index tend to select between 2 and 4 labels. As for GDI$_{33}$, we observed that it only give different results when the data is well-seperated as in SD1 dataset and perform better than the other two indices based on average silhouette witdh. Overall, the results for $\mathfrak{L}$ the objects provided by GDI_{33}, GDI_{53} and Sil index exhibit quit similar behaviour with slightly difference in SD2 and SD4 as reported in Table 2. We also observed that the test set for all dataset give similar clustering output.

5 Real Dataset Evaluation

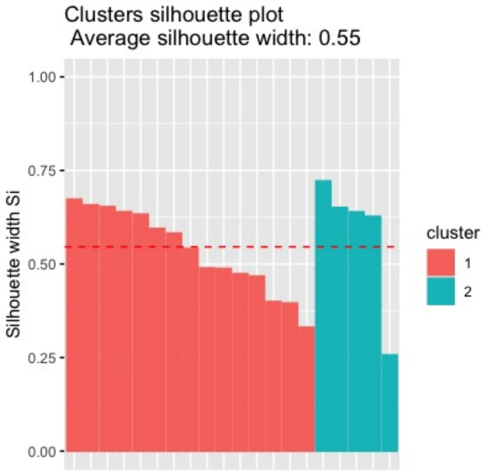
We performed the proposed methods on real dataset namely blood transfussion, Cleaveland heart disease and *E.*coli dataset which summarised in Table 3. The Cluster* refers to the class label that have been identify beforehand. In controlling the effect of outliers on clustering, we also performed Mahalanobis distance to identify the outliers and discard the extreme values. We are not intend to acquire the number of label as in classification dataset since the result may not reveal the same group in clustering (Guyon et al., 2009).

If we now turn to results obtain in real dataset, as expected, pattern acquired by Sil index for three real dataset is between 2 and 4. Whereas, GDI_{33} indicate$\mathfrak{L} = 4$ in *E.*coli



(a) SD4 with $\mathfrak{L} = 2$

(b) SD4 with $\mathfrak{L} = 4$



(c) SD5 with $\mathfrak{L} = 2$

Figure 6: Silhouette plot of non-overlapping dataset.

dataset while GDI_{53} identify $\mathfrak{L} = 3$ in Statlog heart disease. Different results obtain by Lord et al. (2017) where they identify three clusters in *E*.coli and bolld transfussion dataset based on their proposed method.

The unique about clustering is that, it is able to discover new structure of a dataset rather then reproduced the same structure (Guyon et al., 2009). This behaviour gives direct influences on internal cluster validity indices which then reflect in producing group labeling. As we look back of the definition in each ICV, the center of clusters play and important role in minimizing the objective function of clustering algorithm. Furthermore, random initialization and position of objects to minimized the distance in its

Table 3: Description of real dataset and labeling result

			Relabelling of ICV			Class*
Real datasets	n	p	GD133	GDI53	Sil	k
Blood transfussion	644	4	2	2	2	2
Statlog heart disease	267	13	2	3	2	2
E.coli	336	7	4	2	2	8

n = sample size ; p = no. of variables,

cluster in each partitions effect the dissimilarity/similarity characteristics between data point in each partitions to be differ.

6 Conclusion

In this work we presented the accuracy of labeling objects based on voting-based method through q-fold cross-validation approach. We carried out comparison between consensus and traditional clustering to look at performance between both approaches. However, we only reported the performance of one dataset in each overlapping and non-overlapping dataset. We carried out Silhoutte measurement index for performance evaluations.

It is observed that for each partitions, random initialization do produce different members in clusters. Voting-based method give great assistance in determine the *optimal* labeling for dataset. We also observed that the labeling of objects change in some of the generated partitions. There is a situation were we encountered indistinguishable voting results for selection of labeling. The option for this situation is to look at highest average Silhouette width produce for in both label and choose fewer number of clusters.

Beside, the application of q-fold cross-validation technique able to speed-up the computational time in inital stage of consensus clustering with q^{th} iterations. With the concept of selecting the most occurence of $\mathfrak{L}$ in internal validity indexes, voting-based method is less time consuming.

This work has addressed some questions and proposed suggestion for future work. Our method only emphasive of selection of labeling for objects based on voting-based approach and overall accuracy is determine based average Silhoutte width. We consider to improve our method by extending to stability of objects in a clusters. This include reallocation of objects that are misclassified and evaluation of its performance. In a contex of outliers and noise in dataset which give great influence in clustering, it will be interesting to look into.

Acknowledgment

The authors would like to thank the editor and reviewers for their insights which help to improve this paper. We gratefully acknowledge financial support by Univerisiti Teknologi

MARA (UiTM) and Ministry of Higher Education (MOHE) Malaysia.

References

Anderson, B. J., Gross, D. S., Musicant, D. R., Ritz, A. M., Smith, T. G., and Steinberg, L. E. (2006). Adapting k-medians to generate normalized cluster centers. In *Proceedings of the 2006 SIAM International Conference on Data Mining*, pages 165–175. SIAM.

Arbin, N., Suhaimi, N. S., Mokhtar, N. Z., and Othman, Z. (2015). Comparative analysis between k-means and k-medoids for statistical clustering. *Proceedings - AIMS 2015, 3rd International Conference on Artificial Intelligence, Modelling and Simulation*, pages 117–121.

Ben-David, S., Luxburg, U., and Pál, D. (2006). A sober look at clustering stability. *Learning Theory.*

Ben-david, S., Pál, D., and Simon, H. U. (2007). Stability of k-means clustering. In *Proceedings of the 20th Annual Conference on Learning Theory.*

Bezdek, J. C. and Pal, N. R. (1998). Some new indexes of cluster validity. *Part B (Cybernetics) IEEE Transactions on Systems, Man, and Cybernetics*, 28(3):301–315.

Bubeck, S., Meila, M., and von Luxburg, U. (2009). How the initialization affects the stability of the k-means algorithm. *arXiv preprint arXiv:0907.5494.*

Celebi, M. E., Kingravi, H. A., and Vela, P. A. (2013). A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems with Applications*, 40:200–210.

de Assis, E. C. and de Souza, R. M. (2011). A k-medoids clustering algorithm for mixed feature-type symbolic data. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, pages 527–531. IEEE.

Dheeru, D. and Karra Taniskidou, E. (2017). UCI machine learning repository.

Dresen, I. M. G., Boes, T., Huesing, J., Neuhaeuser, M., and Joeckel, K.-H. (2008). New resampling method for evaluating stability of clusters. *BMC bioinformatics*, 9(1):42.

Dudoit, S. and Fridlyand, J. (2002). A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome biology*, 3(7):research0036–1.

Dudoit, S. and Fridlyand, J. (2003). Bagging to improve the accuracy of a clustering procedure. *Bioinformatics*, 19:1090–1099.

Şenbabaoğlu, Y., Michailidis, G., and Li, J. Z. (2014). Critical limitations of consensus clustering in class discovery. *Scientific reports*, 4:6207.

Fred, A. and Jain, A. K. (2002a). Evidence accumulation clustering based on the k-means algorithm. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 442–451. Springer.

Fred, A. L. N. and Jain, A. K. (2002b). Data clustering using evidence accumulation.

In *Proc. Object recognition supported by user interaction for service robots*, volume 4, pages 276–280 vol.4.

Fred, A. L. N. and Jain, A. K. (2005). Combining multiple clusterings using evidence accumulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6):835–850.

Goder, A. and Filkov, V. (2008). Consensus clustering algorithms: Comparison and refinement. In *Proceedings of the Meeting on Algorithm Engineering & Expermiments*, pages 109–117. Society for Industrial and Applied Mathematics.

Guyon, I., Von Luxburg, U., and Williamson, R. C. (2009). Clustering: Science or art. In *NIPS 2009 workshop on clustering theory*, pages 1–11.

Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K. N., Reik, W., Barahona, M., Green, A. R., and Hemberg, M. (2016). Sc3 - consensus clustering of single-cell rna-seq data. *Nature methods*, 14(5):483.

Kuncheva, L. I. and Vetrov, D. P. (2006). Evaluation of stability of k-means cluster ensembles with respect to random initialization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1798–1808.

Lancichinetti, A. and Fortunato, S. (2012). Consensus clustering in complex networks. *Scientific reports*, 2:336.

Liu, A., Zou, S., Qiu, T., and Bai, X. (2016). Research on k-medoids clustering algorithm based on data density and its parallel processing based on mapreduce. *Journal of Residuals Science & Technology*, 13(7):e4015.

Lord, E., Willems, M., Lapointe, F.-J., and Makarenkov, V. (2017). Using the stability of objects to determine the number of clusters in datasets. *Information Sciences*, 393:29–46.

Madhuri, R., Murty, M. R., Murthy, J., Reddy, P. P., and Satapathy, S. C. (2014). Cluster analysis on different data sets using k-modes and k-prototype algorithms. In *ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India*, pages 137–144. Springer.

Melnykov, V., Chen, W.-C., and Maitra, R. (2012). Mixsim: An r package for simulating data to study performance of clustering algorithms. *Journal of Statistical Software*, 51(12):1.

Milligan, G. W. and Cooper, M. C. (1987). Methodology review: Clustering methods. *Applied Psychological Measurement*, 11:329–354.

Minaei-Bidgoli, B., Topchy, A., and Punch, W. F. (2004). Ensembles of partitions via data resampling. In *Proc. ITCC 2004. Int. Conf. Information Technology: Coding and Computing*, volume 2, pages 188–192 Vol.2.

Monti, S. (2003). Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, 52(1/2):91–118.

Nguyen, N. and Caruana, R. (2007). Consensus clusterings. In *Proc. Seventh IEEE Int. Conf. Data Mining (ICDM 2007)*, pages 607–612.

Novoselova, N. and Tom, I. (2012). Entropy-based cluster validation and estimation of the number of clusters in gene expression data. *Journal of bioinformatics and computational biology*, 10(5):1250011.

Park, H. S. and Jun, C. H. (2009). A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications*, 36:3336–3341.

Raykov, Y. P., Boukouvalas, A., Baig, F., and Little, M. A. (2016). What to do when k-means clustering fails: a simple yet principled alternative algorithm. *PLoS ONE*, 11(9):e0162259.

Rendón, E., Abundez, I., Arizmendi, A., and Quiroz, E. M. (2011). Internal versus external cluster validation indexes. *International Journal of computers and communications*, 5(1):27–34.

Risso, D., Purvis, L., Fletcher, R., Das, D., Ngai, J., Dudoit, S., and Purdom, E. (2018). clusterexperiment and rsec: A bioconductor package and framework for clustering of single-cell and other large gene expression datasets. *bioRxiv*, page 280545.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Saeed, F., Salim, N., and Abdo, A. (2012). Voting-based consensus clustering for combining multiple clusterings of chemical structures. *Journal of Cheminformatics*, 4(1):1.

Saeed, F., Salim, N., and Abdo, A. (2013). Consensus methods for combining multiple clusterings of chemical structures. *Journal of Chemical Information and Modeling*, 53(5):1026–1034.

Topchy, A., Jain, A. K., and Punch, W. (2004). A mixture model for clustering ensembles. In *Proceedings of the 2004 SIAM international conference on data mining*, pages 379–390. SIAM.

Topchy, A., Jain, A. K., and Punch, W. (2005). Clustering ensembles: Models of consensus and weak partitions. *IEEE transactions on pattern analysis and machine intelligence*, 27(12):1866–1881.

Vega-Pons, S. and Ruiz-Shulcloper, J. (2011). A survey of clustering ensemble algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(03):337–372.

von Luxburg, U. (2010). Clustering stability: An overview. *Foundations and Trends in Machine Learning*, 2(3):235–274.

Wang, J. (2010). Consistent selection of the number of clusters via crossvalidation. *Biometrika*, 97(4):893–904.

Xie, J., Jiang, S., Xie, W., and Gao, X. (2011). An efficient global K-means clustering algorithm. *Journal of Computers*, 6(2):271–279.

Yang, Y. (2016). *Temporal Data Mining Via Unsupervised Ensemble Learning.* Elsevier.

Zhong, X., Yu, T., and Xia, H. (2017). A new partition-based clustering algorithm for mixed data. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 1.