Neighborhood-Based Smoothing of External Cluster Validity Measures

Ken-ichi Fukui and Masayuki Numao

The Institute of Scientific and Industrial Research (ISIR),
Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka, Japan
fukui@ai.sanken.osaka-u.ac.jp

Abstract. This paper proposes a methodology for introducing a neighborhood relation of clusters to the conventional cluster validity measures using external criteria, that is, class information. The extended measure evaluates the cluster validity together with connectivity of class distribution based on a neighborhood relation of clusters. A weighting function is introduced for smoothing the basic statistics to set-based measures and to pairwise-based measures. Our method can extend any cluster validity measure based on a set or pairwise of data points. In the experiment, we examined the neighbor component of the extended measure and revealed an appropriate neighborhood radius and some properties using synthetic and real-world data.

Keywords: cluster validity, neighborhood relation, weighting function.

1 Introduction

Clustering is a basic data mining task that discovers similar groups from given multi-variate data. Validation of a clustering result is a fundamental but difficult issue, since clustering is an unsupervised learning and is essentially to find latent clusters in the observed data[3,7,14]. Up until now, various validity measures have been proposed from different aspects, and they are mainly separated into two types whether based on internal or external criteria[7,10,8]:

- **Internal criteria** evaluate compactness and separability[3] of the clusters based only on distance between objects in the data space, that is *learning perspective*. As such measures, older methods of Dunn-index[4], DB-index[2], and recent CDbw[5] are well known. Surveys and comparisons of internal cluster validity measures are [3,9].
- **External criteria** evaluate how accurately the correct/desired clusters are formed in the clusters, that is *user's perspective*. External criteria normally uses class/category label together with cluster assignment. Purity, entropy, F-measure, and mutual information are typical measures[10,12,14].

This paper focuses on using external criteria, that is provided by human interpretation of data. It is more beneficial to use external criteria when class labels are available.

P.-N. Tan et al. (Eds.): PAKDD 2012, Part I, LNAI 7301, pp. 354–365, 2012.

In order to understand obtained clusters better, this work introduces a neighborhood relation among clusters. A neighborhood relation is useful especially in case of micro-clusters or, i.e., cluster number is larger than class number. Global structure of clusters, which means not only individual (local) clusters, can be evaluated with neighborhood relation of classes within each cluster.

The basic policies of introducing the neighborhood relation is as follows:

1. A data object which belongs to the same class should be in neighbor over clusters. To evaluate this property, we introduce a weighting function based on *inter-cluster* distance. The inter-cluster distance can be computed based on either topology-based or Euclidean distance in the data space.
2. A weighting function is introduced into basic statistics that are commonly used in the conventional measures. Therefore, our approach is generic, any conventional cluster validity measure that uses these statistics can also be extended in the same way.

Above mentioned conventional indices do not consider neighboring clusters, while very few works introduce inter-cluster connectivity for prototype based clustering[11]. The inter-cluster connectivity is introduced by the first and the second best matching units, but this work is based on internal criterion. The contribution of this work is to introduce *neighborhood relation* over clusters into conventional external cluster validity indices.

The reason why we assume the situation to evaluate an unsupervised learning by class labels is as follows. The fundamental difficulty of unsupervised learning is that the features and the distance metric are derived from observation and assumption, there is no information from human interpretation of data. On the other hand, it is often the case that a small number of samples, or data from the same domain, or simulated samples are available with class labels. In such cases, an external validity measure works as a preliminary evaluation instead of evaluating unlabeled target data.

This paper presents how to introduce the weighting function to smooth the conventional clustering validity measures. In the experiment, we revealed the optimal smoothing radius and also examined several parameters, prototype (micro-cluster) number, and class overlapping degree. We revealed the properties of our extended measure and showed potential to validate a clustering result considering neighborhood relation of clusters.

2 Preliminaries

Definition 1. *(Clustering) Given a set of v-dimensional objects $\mathbf{S} = \{\boldsymbol{x}_i\}_{i=1}^{N} \in \mathbb{R}^v$, a clustering produces a cluster set $\mathbf{C} = \{C_i\}_{i=1}^{K}$ with a cluster assignment $c(i) \in \mathbf{C}$ for each object $\boldsymbol{x}_i$.*

Definition 2. *(Class) Let a class set be $\mathbf{T} = \{T_i\}_{i=1}^{L}$, and $t(i) \in \boldsymbol{T}$ denotes a class assignment for $\boldsymbol{x}_i$. Classes are provided independent from a clustering.*

Definition 3. *(Inter-cluster distance)* $d(C_i, C_j) \in \mathbb{R}$ *is defined as inter-cluster distance between clusters that can be computed either Euclidean-based or topology-based distance.*

Ex) Inter-cluster distance. Euclidean-based distances can be given by single linkage, complete linkage, and other methods commonly used in an aggregative hierarchical clustering. While, topology-based distance by the number of hops in a neighbor graph. The neighbor graph can be obtained by such as a threshold on Euclidean-based distance or by k-nearest neighbor.

Note that though a neighbor graph is normally obtained independent from a clustering process, some method produces cluster (vector quantization) with topology preservation such as Self-Organizing Map(SOM)[6], which is also used in this experiment.

The objective of this work is to evaluate density of class $\mathbf{T}$ within intra-cluster $\mathbf{C}$ together with the neighbor relation based on inter-cluster distance $d(C_i, C_j)$.

3 Neighborhood-Based Smoothing of Validity Measures

There are two types of cluster validity measures, namely set-based and pairwise-based measures[1]. These two types of measures can be extended in different manners.

3.1 Extension of Set-Based Cluster Validity Measures

First, the way to extend set-based cluster validity measures[10,12] such as cluster purity and entropy are described in this section. The properties of each measure were studied in the literature[1].

By considering neighborhood relation of clusters, the neighbor class distribution should be taken into account to the degree of certain class contained in a cluster, that is, the data points of the same class in the neighbor clusters should have a high weight, while those of distant clusters should have a low weight based on the inter-cluster distance as the diagram is shown in Fig. 1.

Let $f(\boldsymbol{u}; l)$ be a density distribution of class label $l \in \mathbf{T}$ at $\boldsymbol{u} \in \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ denotes a data space, and $h(\boldsymbol{u}, \boldsymbol{v}) : \boldsymbol{\Omega} \times \boldsymbol{\Omega} \mapsto \mathbb{R}$ be a weighting function based on the neighborhood relation. Based on the above concept, a class density distribution $f(\boldsymbol{u}; l)$, a data density distribution $f(\boldsymbol{u})$, and a total volume of data N are smoothed by the weighting function $h(\boldsymbol{u}, \boldsymbol{v})$ as follows:

$$\hat{f}(\boldsymbol{u}; l) = \int_{\boldsymbol{\Omega}} h(\boldsymbol{u}, \boldsymbol{v}) f(\boldsymbol{v}; l) dv, \tag{1}$$

$$\hat{f}(\boldsymbol{u}) = \sum_{l \in \mathbf{T}} \hat{f}(\boldsymbol{u}; l) = \sum_{l \in \mathbf{T}} \int_{\boldsymbol{\Omega}} h(\boldsymbol{u}, \boldsymbol{v}) f(\boldsymbol{v}; l) dv, \tag{2}$$

[1] This work introduces the smoothing function into several cluster validity measures, in actual use, a measure should be selected according to the target application and the aspects the user wants to evaluate.

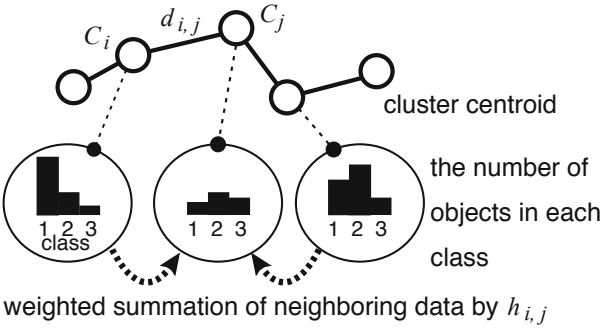


Fig. 1. Extension of a set-based clustering measure. The basic statistics are weighted by the neighborhood relation based on inter-cluster distance $d_{i,j}$. This example shows topology-based distance.

$$\hat{N} = \int_{\mathbf{\Omega}} \hat{f}(\boldsymbol{u})du = \int_{\mathbf{\Omega}} \sum_{l \in \mathbf{T}} \int_{\mathbf{\Omega}} h(\boldsymbol{u}, \boldsymbol{v}) f(\boldsymbol{v}; l) dv du. \tag{3}$$

Discretizing eqs. (1) to (3), let $N_{l,i}$ be the number of objects with class l in the i^{th} cluster $C_i \in \mathbf{C}$; $N_{l,i} = \#\{\boldsymbol{x}_k | t(k) = l, c(k) = C_i\}$, where # denotes the number of elements. N_i denotes the number of objects in cluster C_i; $N_i = \#\{\boldsymbol{x}_k | c(k) = C_i\}$. Also N denotes the total number of objects; $N = \#\{\boldsymbol{x}_k | \boldsymbol{x}_k \in \mathbf{S}\}$. Eqs. (1) to (3) can be rewritten as follows:

$$N'_{l,i} = \sum_{C_j \in \mathbf{C}} h_{i,j} N_{l,j}, \tag{4}$$

$$N'_i = \sum_{l \in \mathbf{T}} N'_{t,i} = \sum_{l \in \mathbf{T}} \sum_{C_j \in \mathbf{C}} h_{i,j} N_{l,j}, \tag{5}$$

$$N' = \sum_{C_i \in \mathbf{C}} N'_i = \sum_{C_i \in \mathbf{C}} \sum_{l \in \mathbf{T}} \sum_{C_j \in \mathbf{C}} h_{i,j} N_{l,j}. \tag{6}$$

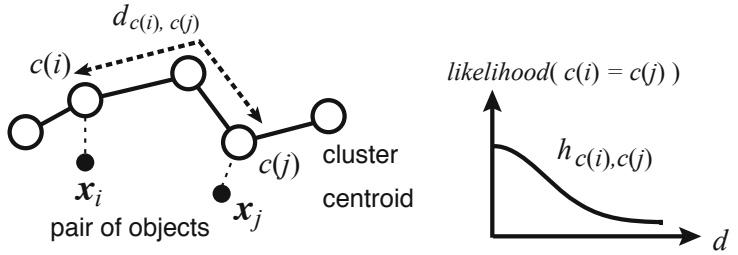
Here, $h_{i,j}$ can be used any monotonically decreasing function, for example, the often encountered Gaussian function: $h_{i,j} = \exp(-d_{i,j}/\sigma^2)$, where $d_{i,j}$ denotes inter-cluster distance and $\sigma(> 0)$ is a smoothing (neighborhood) radius.

Thus, weighted cluster purity and entropy, for example, are defined using the weighted statistics of eqs. (4), (5), and (6) as follows:

weighted Cluster Purity (wCP)

$$\mathrm{wCP}(\mathbf{C}) = \frac{1}{N'} \sum_{C_i \in \mathbf{C}} \max_{l \in \mathbf{T}} N'_{l,i}. \tag{7}$$

The original purity is an average of the ratio that a majority class occupies in each cluster, whereas in the weighted purity a majority class is determined by the neighbor class distribution $\{N'_{l,i}\}$.



(a) Distance of data pair on a graph (b) Likelihood function

Fig. 2. Extension of a pairwise-based clustering measure. A likelihood function is introduced to represent a degree that a data pair belongs to the same cluster.

weighted Entropy (wEP)

$$\mathrm{wEP}(\mathbf{C}) = \frac{1}{|\mathbf{C}|} \sum_{C_i \in \mathbf{C}} \mathit{Entropy}(C_i), \tag{8}$$

$$\mathit{Entropy}(C_i) = -\frac{1}{\log N'} \sum_{l \in \mathbf{T}} \frac{N'_{l,i}}{N'_i} \log \frac{N'_{l,i}}{N'_i}, \tag{9}$$

where $|\mathbf{C}|$ denotes a cluster number. The original entropy indicates the degree of unevenness of class distribution within a cluster, whereas the extended entropy includes unevenness of the neighboring clusters.

3.2 Extension of Pairwise-Based Cluster Validity Indices

This section describes an extension of pairwise-based cluster validity measures[1,14]. Table 1 shows a class and cluster confusion matrix of data pairs, where a, b, c, d are the number of data pairs where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ do or do not belong to the same class/cluster.

Table 1. Class and cluster confusion matrix of data pairs

	$t(i) = t(j)$	$t(i) \neq t(j)$
$c(i) = c(j)$	a	b
$c(i) \neq c(j)$	c	d

Here, we introduce $\mathit{likelihood}(c(i) = c(j))$ indicating a degree that a data pair $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ belongs to the same cluster instead of the actual number of data pairs. The likelihood is given by the inter-cluster distance of the data pair as shown in Fig. 2(a). The same weighting function as in sec. 3.1 is available for the

likelihood function (Fig. 2(b)); $likelihood(c(i) = c(j)) = h_{c(i),c(j)}$. Then, a, b, c, d are replaced by summation of the likelihoods as follows:

$$a' = \sum_{\{i,j|t(i)=t(j)\}} h_{c(i),c(j)}, \tag{10}$$

$$b' = \sum_{\{i,j|t(i)\neq t(j)\}} h_{c(i),c(j)}, \tag{11}$$

$$c' = \sum_{\{i,j|t(i)=t(j)\}} \left(1 - h_{c(i),c(j)}\right) = a + c - a', \tag{12}$$

$$d' = \sum_{\{i,j|t(i)\neq t(j)\}} \left(1 - h_{c(i),c(j)}\right) = b + d - b'. \tag{13}$$

With these extended a', b', c' and d', weighted pairwise accuracy and pairwise F-measure are defined as follows:

weighted Pairwise Accuracy (wPA)

$$\mathrm{wPA}(\mathbf{C}) = \frac{a' + d'}{a' + b' + c' + d'}. \tag{14}$$

The original pairwise accuracy is a ratio of the number of pairs in the same class belonging to the same cluster, or the number of pairs in different classes belonging to different clusters, against all pairs. The weighted PA is the degree to which pairs in the same class belong to the neighbor clusters or that pairs in different classes belong to distant clusters.

weighted Pairwise F-measure (wPF)

$$\mathrm{wPF}(\mathbf{C}) = \frac{2 \cdot P \cdot R}{P + R}, \tag{15}$$

where $P = a'/(a'+b')$ is precision, that is a measure of the same class among each cluster, and $R = a'/(a'+c')$ is recall that is a measure of the same cluster among each class. The original pairwise F-measure is a harmonic average of the precision and the recall. While, the weighted PF is based on a degree that the data pairs belong to the same cluster.

3.3 Weighting Function

For the weighting function for smoothing in the set-based and the likelihood in pairwise-based measures, any monotonically decreasing function $h_{i,j} \geq 0$ is feasible, including the Gaussian or a rectangle function. Note that the extended measures are exactly the same as the original measures when $h_{i,j} = \delta_{i,j}$ (δ is the Kronecker delta).

The neighborhood radius effects the degree of smoothing and likelihood. Fig. 3 illustrates that the measure evaluates individual clusters, that is the original

values, as the radius becomes zero ($\sigma \to 0$). On the other hand, as the radius becomes larger ($\sigma \to \infty$), the data space is smoothed by almost the same weights, and all micro-clusters are treated as one big cluster. The way to find the optimal radius is described in section 3.4.

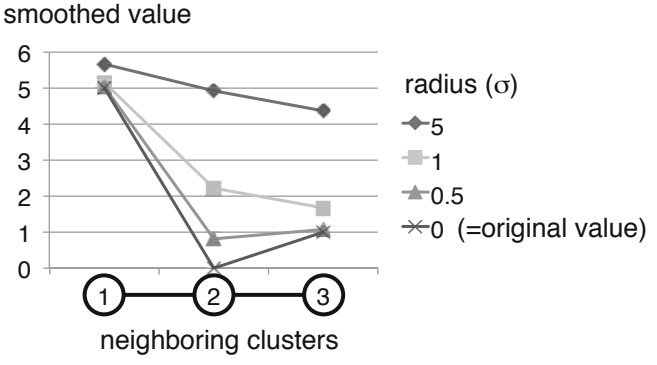


Fig. 3. Example of the effect of smoothing radius. Values over the neighborhood relation of the clusters become smoother as the radius increases.

3.4 Optimal Smoothing Radius

Our smoothed measures include a neighborhood relation in the conventional cluster validity measures. In order to evaluate a *neighbor component* within the measure, we defined as:

Definition 4. *(neighbor component) The quantity within the smoothed cluster validity value (Eval) that are caused by the neighborhood relation.*

Here, *Eval* refers to the output value of wCP, wEP, wPA, or wPF in this paper.

Then, the neighbor component (NC) can be computed by comparing $Eval$s with randomized neighborhood relation.

$$NC(\sigma) = |Eval - \lim_{n\to\infty} Eval_{rnd(n)}| = |Eval(d_{i,j}) - Eval(\bar{d}_{i,j}))|, \tag{16}$$

where $Eval_{rnd(n)}$ denotes an average of $Eval$ when inter-cluster distances are n times shuffled, and when $n \to \infty$ this value converges to $Eval$ with the average of all inter-cluster distances $\bar{d}_{i,j}$. It is assumed that the smoothing radius that maximizes the neighbor component is the optimal one, i.e., $\sigma^* = \arg\max_\sigma NC(\sigma)$. Then, the optimal evaluation value can be $Eval^* = Eval(\sigma^*)$.

4 Evaluation of the Smoothed Validity Measures

This section describes the experiment to clarify the properties of the proposed smoothed validity measures.

4.1 Settings of Clustering and Neighborhood Relation

1. **kmc-knn**
   Typical k-means clustering was used to produce a clustering and mutual k-nearest neighbor (kmc-knn) was used to obtain the neighbor relation. With parameters of the prototype (micro-cluster) number $k1$ and of nearest neighbors $k2$, adjacent matrix $\mathbf{A} = (a_{i,j})$ can be given by:

$$a_{i,j} = \begin{cases} 1 & \text{if } C_j \in \mathbf{O}(C_i) \text{ and } C_i \in \mathbf{O}(C_j) \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

   where $\mathbf{O}(C_i)$ denotes a set of k-nearest neighbor clusters from C_i, where $d(C_i, C_j)$ is given by Euclidean distance. Then, distance matrix $\mathbf{D} = (d_{i,j})$ can be given by topological distance, in this experiment the shortest path between C_i and C_j is used, where the shortest path of all pairs are calculated by Warshall-Floyd Algorithm.
2. **SOM**
   Also the SOM[6] was used as an another type of producing micro-cluster prototypes with neighbor relation. In the SOM, the neurons of prototypes correspond to centroids of micro-clusters. The standard batch type SOM is used in this work. A distance matrix $\mathbf{D}$ is given by $d_{i,j} = ||\boldsymbol{r}_i - \boldsymbol{r}_j||$, where $\boldsymbol{r}$ is a coordinate of a neuron within the topology space of the SOM.

4.2 Datasets

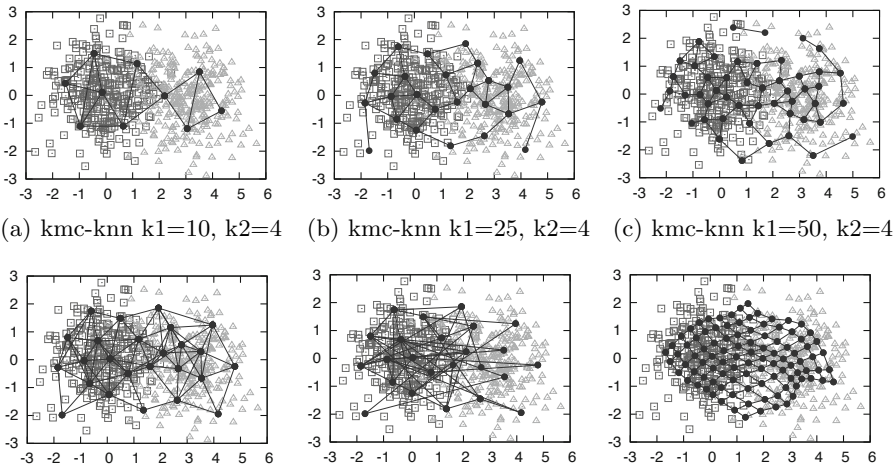
1. **Synthetic data**
   In order to evaluate the proposed measure, two classes of two-dimensional synthetic data were prepared, where 300 data points for each class were generated from different Gaussian distributions. The data distribution and examples of graphs are illustrated in Fig. 4.
2. **Real-world data**
   Well-known open datasets[2] were used as real-world data: Iris data (150 samples, 4 attributes, 3 classes), Wine data (178 samples, 13 attributes, 3 classes), and Glass Identification data (214 samples, 9 attributes, 6 classes).

4.3 Effect of Smoothing Radius - Finding the Optimal Radius

Fig. 5 shows the evaluation values of the smoothed validity measures for the synthetic data using kmc-knn. The larger value is the better except entropy. The values are average of 100 runs of randomized initial values.

Firstly, the total evaluation values ($Eval$) provides always better value than that of random topology ($Eval_{rnd}$) where neighborhood relation of the prototypes is destroyed. This means that the proposed measures evaluate both cluster validity and neighborhood relation of the clusters.

[2] http://archive.ics.uci.edu/ml/



(a) kmc-knn k1=10, k2=4 (b) kmc-knn k1=25, k2=4 (c) kmc-knn k1=50, k2=4

(d) kmc-knn k1=25, k2=8 (e) kmc-knn k1=25, k2=4, random topology (f) SOM 10x10

Fig. 4. Cluster prototypes (•) with topology-based neighbor relation on two dimensional synthetic data. The data points ($\square, \triangle$) were generated from two Gaussian distributions; $N(\mu_1, 1)$ and $N(\mu_2, 1)$, where $\mu_1 = (0, 0)$ and $\mu_2 = (3, 0)$.

Secondly, as the smoothing radius becomes close to zero ($\sigma \to 0$), the extended measure evaluates individual clusters without neighborhood relation. Whereas, as the radius becomes larger ($\sigma \to \infty$), the extended measure treats whole data as one big cluster as mentioned before. Therefore, the solid and the broken lines gradually become equal as the radius becomes close to zero or becomes much larger.

Thirdly, the neighbor component has a monomodality against the radius in all measures, since there exists an appropriate radius to the average class distribution. Since the smoothed measure is a composition of cluster validity and neighborhood relation, the radius that gives the maximum *Eval* does not always match with that of neighbor component, for instance, wCP, wEP, and wPF in Fig. 5. Therefore, the neighbor component should be examined to find the appropriate radius. Also the appropriate radius depends on function of the measure such as purity, F-measure, or entropy. This means that the user should use different radius for each measure.

These three trends appear also in SOM (omitted due to page limitation).

4.4 Effect of Prototype Number

The effect of prototype number is examined by changing $k1 = 10, 25, 50$ (Fig. 6). In wPF, $k1 = 25$ provides the highest neighbor component (0.116 at $\sigma = 1.4$) among three (Fig. 6(b)). wPF can suggest an optimal prototype number in terms of maximizing the neighbor component in the measure, which means neighbor

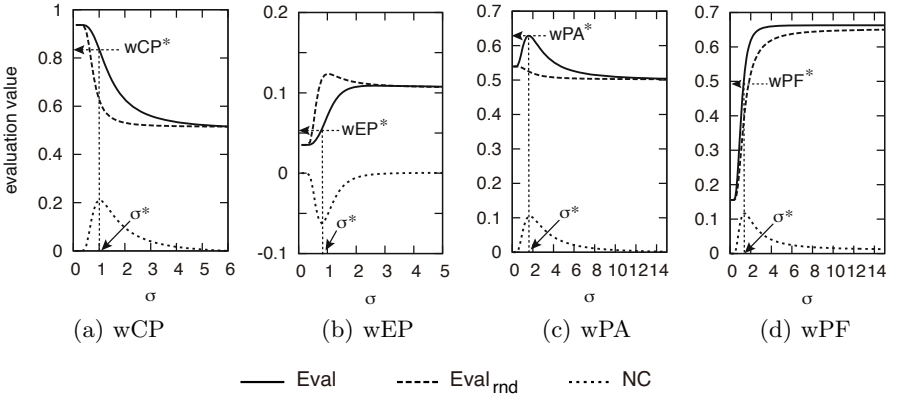


Fig. 5. The effect of smoothing radius (synthetic data, kmc-knn($k1 = 25, k2 = 4$)); total evaluation value ($Eval$), $Eval$ with random topology ($Eval_{rnd}$), neighbor component (NC)

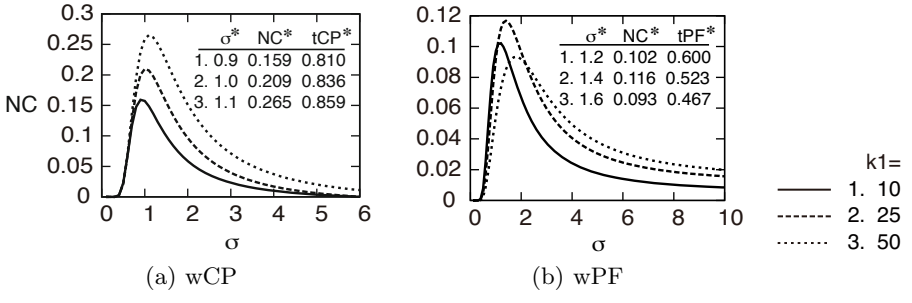
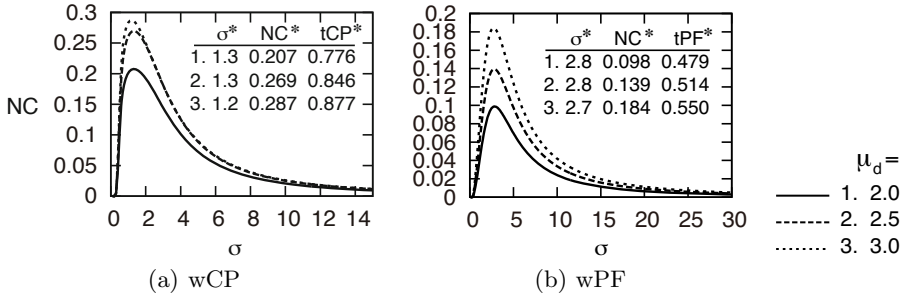


Fig. 6. The effect of prototype number (synthetic data, kmc-knn($k2 = 4$)). The maximum neighbor component (NC^*) and total values (wCP and wPF) are listed together in the table.

relation of class distribution is maximized. However, the larger $k1$ the better in wCP (Fig. 6(a)). This is because the function of cluster purity given by eq. (7), that is, the smaller number of elements in some cluster tends to give better purity.

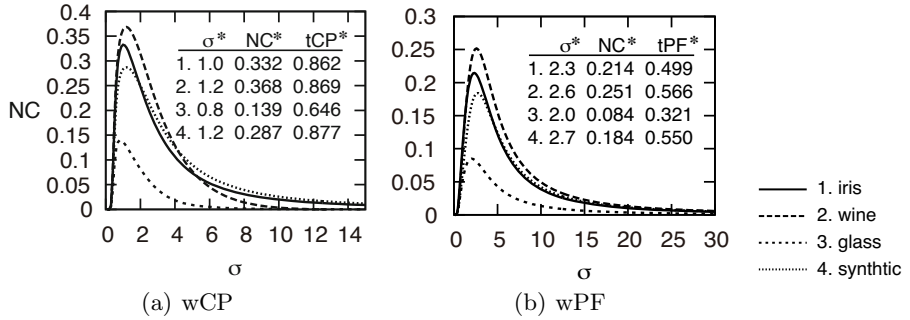
4.5 Effect of Class Overlap

The effect of class overlap is examined (Fig. 7) by changing distance between class centers $\mu_d = \mu_2^x - \mu_1^x$ from 2.0 to 3.0 in the synthetic data. Observing Fig. 7, the lower class overlap is, the better the neighbor component and the total values. However, the optimal radii are nearly the same even in different class overlap. This means that our measure can determine the optimal radius independent to class overlap, and can evaluate volume of overlap.



(a) wCP (b) wPF

Fig. 7. The effect of class overlap (synthetic data, SOM(10×10))



(a) wCP (b) wPF

Fig. 8. The effect of dataset, SOM (10×10)

4.6 Real-World Data

Fig. 8 shows the result for real-world data using SOM. Though there exists an optimal radius, the optimal radii vary depending on dataset, i.e., the number of classes and the class distribution. This result indicates that depending on dataset and measure, a user should use different radius that gives the maximum volume of neighbor component.

5 Conclusion

This paper proposed a novel and generic smoothed cluster validity measures based on neighborhood relation of clusters with external criteria. The experiments revealed the existence of an optimal neighborhood radius which maximizes the neighbor component. A user should use an optimal radius depending on a function of measure and a dataset. Our measure can determine the optimal radius independent to class overlap, and can evaluate volume of class overlap. In addition, feature selection, metric learning[13,15], and a correlation index for multilabels to determine the most relevant class are promising future directions for this work.

Acknowledgment. This work was supported by KAKENHI (21700165).

References

1. Amigó, E., Gonzalo, J., Artiles, J., Verdejo, F.: A comparison of extrinsic clustering evaluation metrics based on formal constraints. Information Retrieval 699(12), 461–486 (2009)
2. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Transactions on Pattern Analsis and Machine Intelligence (TPAMI) 1(4), 224–227 (1979)
3. Deborah, L.J., Baskaran, R., Kannan, A.: A survey on internal validity measure for cluster validation. International Journal of Computer Science & Engineering Survey (IJCSES) 1(2), 85–102 (2010)
4. Dunn, J.C.: Well separated clusters and optimal fuzzy partitions. Journal of Cybernetics 4, 95–104 (1974)
5. Halkidi, M., Vazirgiannis, M.: Clustering validity assessment using multi representatives. In: Proc. 2nd Hellenic Conference on Artificial Intelligence, pp. 237–248 (2002)
6. Kohonen, T.: Self-Organizing Maps. Springer (1995)
7. Kovács, F., Legány, C., Babos, A.: Cluster validity measurement techniques. Engineering 2006, 388–393 (2006)
8. Kremer, H., Kranen, P., Jansen, T., Seidl, T., Bifet, A., Holmes, G., Pfahringer, B.: An effective evaluation measure for clustering on evolving data streams. In: Proc. the 17th SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2011), pp. 868–876 (2011)
9. Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J.: Understanding of internal clustering validation measures. In: Proc. IEEE International Conference on Data Mining (ICDM 2010), pp. 911–916 (2010)
10. Rendón, E., Abundez, I., Arizmendi, A., Quiroz, E.M.: Internal versus external cluster validation indexes. International Journal of Computers and Communications 5(1), 27–34 (2011)
11. Tasdemir, K., Merényi, E.: A new cluster validity index for prototype based clustering algorithms based on inter- and intra-cluster density. In: Proc. International Joint Conference on Neural Networks (IJCNN 2007), pp. 2205–2211 (2007)
12. Veenhuis, C., Koppen, M.: Data Swarm Clustering, ch. 10, pp. 221–241. Springer (2006)
13. Weinberger, K.Q., Blitzer, J., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research (JMLR) 10, 207–244 (2009)
14. Xu, R., Wunsch, D.: Cluster Validity. Computational Intelligence, ch. 10, pp. 263–278. IEEE Press (2008)
15. Zha, Z.J., Mei, T., Wang, M., Wang, Z., Hua, X.S.: Robust distance metric learning with auxiliary knowledge. In: Proc. International Joint Conference on Artificial Intelligence (IJCAI 2009), pp. 1327–1332 (2009)