FCM-type Cluster Validation in Fuzzy Co-Clustering and Collaborative Filtering Applicability

Katsuhiro Honda †, Mai Muranishi †, Akira Notsu †, and Hidetomo Ichihashi †, Osaka Prefecture University, Sakai, Osaka, JAPAN

Summary
FCM-type cluster validation is a technique for searching for the optimal fuzzy partition, in which the number of clusters is evaluated by considering the degree of overlapping of fuzzy memberships, cluster compactness or cluster separation. In this paper, a new approach for FCM-type cluster validation in fuzzy co-clustering is proposed. Because fuzzy co-clustering does not use cluster prototypes, cluster separation is evaluated without using the distances between cluster prototypes. In numerical experiment, the applicability of the new validity measure to collaborative filtering task is studied using a purchase history data.

Key words: Fuzzy clustering, Co-clustering, Cluster validation, Collaborative filtering.

1. Introduction

Fuzzy c-Means (FCM) [1] is a basic model of unsupervised classification, in which each cluster is represented by a point-type cluster prototype and a predefined number of clusters are estimated by an iterative algorithm. In order to search for the optimal fuzzy partition, the FCM algorithm is performed many times using different numbers of clusters and the best one is selected considering a cluster validity measure. Cluster validation has been widely studied in the FCM clustering context. First, the quality of fuzzy partitions was measured by the degree of crispness of fuzzy memberships such as Partition Coefficient (PC) [1] and Partition Entropy (PE) [2]. Second, the geometric feature of fuzzy partitions was employed for measuring the partition quality of prototype-based fuzzy clustering. Xie and Beni [3] considered the compactness and separateness of FCM-type clusters and proposed a cluster validity measure of the ratio of compactness and separateness. Several modifications of Xie-Beni index have been proposed [4-6]. Third, the cluster overlapping degree such as inter-cluster proximity [7] was adopted, which can be applied to unbalanced cluster densities by using fuzzy memberships only. We have still many other validation indices [8].

Co-clustering (or Bi-clustering) is the technique for clustering of two-way data sets such as co-occurrence matrix. For example, such data sets are common in document-keyword co-occurrence information in document clustering, user-item purchase history in personalized recommendation problems, machining tools-products relation in factory automation, and so on. Fuzzy clustering for categorical multivariate data (FCCM) [9] is an FCM-type co-clustering model, in which fuzzy partition of both users and items is estimated based on the FCM-like concept. The clustering criterion is given by the degree of aggregation to be maximized while different constraints are forced to the two memberships of users and items. Object memberships are forced to be exclusive in a similar manner with FCM, in which the sum of memberships w.r.t. clusters are 1 for each user. On the other hand, the sum of item memberships w.r.t. items are forced to be 1 in each cluster. So, item memberships only play a role for evaluating the relative responsibility of items in each cluster. This type of co-clustering model was proved to be useful in collaborative filtering tasks [10-12], in which several popular items can be shared by multiple clusters [13,14].

In this paper, the cluster validation problem is studied in the FCM-type co-clustering context. First, the degree of crispness of fuzzy memberships such as PC and PE is directly applied to fuzzy co-cluster memberships. Second, a new compactness/separateness index is proposed so as to evaluate the pseudo-geometric feature of prototype-less fuzzy co-cluster partitions. The applicability of the new index is demonstrated with several artificial data sets followed by a numerical experiment of a collaborative filtering task.

The remainder of this paper is organized as follows: Section 2 gives a brief review on the conventional FCM-type cluster validity indices and co-clustering model. Section 3 discusses the applicability of the FCM-type validity indices and proposes a new validity index for FCCM. Section 4 presents several experimental results to demonstrate the characteristic features of the proposed index. Section 5 summarizes the conclusions of this paper.
2. FCM-type Cluster Validity Indices and Fuzzy Co-clustering Model

2.1 FCM and Cluster Validity Indices

FCM [1] tries to partition \( n \) samples (objects) \( x_i \), \( i = 1, \ldots, n \) into \( C \) fuzzy clusters, in which each instance is presented by multi-dimensional observation vector. Clusters are represented by prototypical centroids \( b_c \), \( c = 1, \ldots, C \) and the clustering criterion is given by the distance between samples and centroids in the multi-dimensional space:

\[
L_{fcm} = \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci}^\theta \|x_i - b_c\|^2 ,
\]

where \( u_{ci} \) is the fuzzy membership of sample \( i \) in cluster \( c \) and \( \theta \) is an exponential weight for fuzzification. The sum of \( u_{ci} \) w.r.t. \( c \) is constrained to be 1. Other fuzzification approaches such as entropy-based [15] and K-L information-based [16,17] have been also proposed.

In order to select the optimal fuzzy partitions with plausible cluster numbers, many cluster validity indices have been proposed. Partition Coefficient (PC) [1] and Partition Entropy (PE) [2] measure the crispness of fuzzy memberships:

\[
PC = \frac{1}{n} \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci}^2 ,
\]

\[
PE = -\frac{1}{n} \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log u_{ci} .
\]

PC (PE) becomes smaller (larger) as fuzzy memberships become very fuzzy (ambiguous). So, the optimal partition can be the one having large PC (or small PE). Although the indices can directly measure the partition quality, the selected partition sometimes does not suit our human sense because they lack geometric feelings.

Considering the geometric features, Xie and Beni [3] proposed a measure for evaluating both the compactness and separateness of fuzzy clusters:

\[
V_{AB} = \frac{\sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci}^2 \|x_i - b_c\|^2}{n \min_{k,j} \|b_k - b_j\|^2} ,
\]

where the numerator is the FCM objective function with \( \theta = 2 \) and measures the compactness of clusters while the denominator measures the separateness of clusters. The optimal compact/separate cluster can be found by minimizing the Xie-Beni index.

2.2 FCM-type Co-clustering

Co-clustering is a technique for capturing the intrinsic cluster structures from co-occurrence information among objects and items. Assume that we have a similarity (co-occurrence) matrix \( R = \{r_{ij}\} \) on objects \( i = 1, \ldots, n \) and items \( j = 1, \ldots, m \) and each element \( r_{ij} \in [0,1] \) shows the similarity degree among user \( i \) and item \( j \). Oh et al. [9] proposed FCCM, in which the objective function is defined by considering the aggregation degree of each cluster:

\[
L_{fcm} = \sum_{c=1}^{C} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ci} w_{cj} r_{ij} + \lambda_c \sum_{i=1}^{n} u_{ci} \log u_{ci}
\]

\[
+ \lambda_w \sum_{j=1}^{m} \sum_{i=1}^{n} w_{cj} \log w_{cj} .
\]

\( U = \{u_{ci}\} \) and \( W = \{w_{cj}\} \) are the fuzzy memberships of object \( i \) and item \( j \) to cluster \( c \), respectively. The entropy terms are the fuzzification penalty in the entropy-based fuzzification approach [15]. In order to extract co-clusters having high aggregation degrees, \( u_{ci} \) and \( w_{cj} \) are iteratively optimized so that \( u_{ci} \) and \( w_{cj} \) become large if object \( i \) and item \( j \) are highly relevant.

The sum of \( u_{ci} \) is constrained to be 1 in a similar manner to FCM. On the other hand, \( w_{cj} \) is estimated under a different constraint of \( \sum_{j=1}^{m} w_{cj} = 1 \) in order to avoid trivial solutions where all objects and items are assigned to a solo cluster. So, \( w_{cj} \) represents the relative responsibility of item \( j \) in cluster \( c \) and item assignment is not necessarily exclusive, i.e., items can be shared (or rejected) by multiple (all) clusters.

3. Cluster Validation in FCM-type Co-clustering

First, the applicability of the conventional FCM-type validity indices is discussed in the context of prototype-less fuzzy co-clustering. In order to evaluate the partition quality of fuzzy partitions, PC and PE can be applied with two different types of fuzzy memberships. The conventional PC and PE of Eqs.(2) and (3) can be directly used for measuring the degree of crispness of user memberships \( u_{ci} \) while they should be slightly modified for
measuring the quality of item partitions. Equations (2) or
(3) are monotonically increasing or decreasing as
\( n \) becomes larger because the sum of \( u_{ij} \) is 1 for every \( C \).
On the other hand, \( w_{ij} \) have the sum-to-one condition in
each cluster and the total sum of \( w_{ij} \) is equal to \( C \). In this
sense, \( PC \) or \( PE \) for \( w_{ij} \) should be normalized by \( C \)
instead of \( n \) in the conventional ones.
\[
PC_u = \frac{1}{n} \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci}^2, \quad PC_w = \frac{1}{C} \sum_{c=1}^{C} \sum_{j=1}^{m} w_{cj}^2, \quad \text{(6)}
\]
\[
PE_u = -\frac{1}{n} \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log u_{ci}, PE_w = -\frac{1}{C} \sum_{c=1}^{C} \sum_{j=1}^{m} w_{cj} \log w_{cj}. \quad \text{(7)}
\]
In the numerical experiments of the next section, the above
formulations are adopted in a hybrid manner.
Because FCCM is a prototype-less clustering method, the
geometric feature of prototypes is not available, i.e., Xie-
Beni index or its variants cannot directly be adopted. So, in
this paper, a new concept of measuring the intra-cluster
compactness and the inter-cluster similarity is proposed.
Although the clustering criterion of Eq.(5) measures the
aggregation degree for capturing a dense mass of users and
items, it does not necessarily measure the compactness of
the cluster. A compact cluster should reject the user-item
pairs having small \( r_{ij} \). So, a compactness measure is
defined as:
\[
\text{compact} = \frac{1}{C(C-1)} \sum_{k=1}^{C} \sum_{l \neq k}^{C} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ki} w_{lj} r_{ij} \left( 1 - r_{ij} \right). \quad \text{(8)}
\]
Next, two clusters \( k \) and \( l \) are separated if users of \( k \) are
not familiar to cluster \( l \), i.e., the users are not familiar to
items of \( l \). So, the total separateness is measured by
\[
\text{separate} = \frac{1}{C(C-1)} \sum_{k=1}^{C} \sum_{l \neq k}^{C} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ki} w_{lj} r_{ij}. \quad \text{(9)}
\]
Then, a Xie-Beni-like validation measure for fuzzy co-
cluster partition \( V_{co} \) is defined as:
\[
V_{co} = \frac{\text{compact}}{\text{separate}} = \frac{(C-1) \sum_{k=1}^{C} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ki} w_{lj} (2 r_{ij} - 1)}{\sum_{k=1}^{C} \sum_{l \neq k}^{C} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ki} w_{lj}}. \quad \text{(10)}
\]

### 4. Numerical Experiments

#### 4.1 Artificial Data Set 1: Dual Exclusive Situation

In artificial data set 1 shown in Table 1 (a), 12 objects and
10 items form three disjoint clusters, where each object
and item belongs to a solo cluster. The data set 1-(a) has a
noise-less boundary while the data set 1-(b) includes 10%
noise, in which randomly selected 10% elements of data
set 1-(a) was exchanged as “0 ⇔ 1”.

<table>
<thead>
<tr>
<th>Table 1. Artificial co-occurrence matrix 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{item} )</td>
</tr>
</tbody>
</table>
| \hline
| 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 |
| 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 |
| 3 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 |
| 4 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 |
| 5 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 |
| 6 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 |
| 7 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 |
| 8 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 |
| 9 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 |
| 10 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 |
| 11 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 |
| 12 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 |

<table>
<thead>
<tr>
<th>Table 2. Comparison of validation indices (data set 1: noise-less)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( PC_u )</td>
</tr>
</tbody>
</table>
|\hline
| 2 & 0.893 & 0.259 & 0.231 & 0.155 & 1.487 & 0.231 & 23.93 |
| 3 & 0.999 & 0.276 & 0.276 & 0.002 & 1.411 & 0.003 & 35.15 |
| 4 & 0.995 & 0.232 & 0.231 & 0.017 & 1.634 & 0.028 & 7.07 |
| 5 & 0.991 & 0.206 & 0.204 & 0.033 & 1.769 & 0.058 & 5.04 |
| 6 & 0.987 & 0.188 & 0.185 & 0.048 & 1.859 & 0.089 & 4.29 |

<table>
<thead>
<tr>
<th>Table 3. Comparison of validation indices (data set 1: 10% noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C )</td>
</tr>
</tbody>
</table>
|\hline
| 2 & 0.908 & 0.281 & 0.255 & 0.135 & 1.440 & 0.195 & 5.04 |
| 3 & 0.997 & 0.306 & 0.305 & 0.012 & 1.419 & 0.017 & 7.48 |
| 4 & 0.985 & 0.253 & 0.249 & 0.046 & 1.645 & 0.075 & 4.17 |
| 5 & 0.973 & 0.221 & 0.215 & 0.081 & 1.780 & 0.144 & 3.38 |
| 6 & 0.961 & 0.200 & 0.192 & 0.118 & 1.871 & 0.220 & 3.01 |

The FCCM algorithm was applied to the data set with
various cluster numbers \( C = \{ 2,3,4,5,6 \} \). The fuzzification
weights were set as \( \lambda_u = 0.1 \) and \( \lambda_w = 1.0 \). Tables 2 and 3
compare the values of several validity indices and the
selected cluster numbers whose index values are given by
bold and italic. In the tables, the result of the proposed
method is compared not only with each of PC and PE for
objects and items but also the products of them. When both objects and items are exclusively partitioned, not only the proposed method but also PC and PE for both objects and items could successfully select the optimal cluster number of $C = 3$. This result implies that dual exclusive situations can be handled by the conventional FCM-type validity approaches.

4.2 Artificial Data Set 2: Partially Sharing Situation

In artificial data set 2 shown in Table 4 (a), 12 objects form three disjoint clusters while some of 11 items were shared by multiple clusters. Item 4 and Item 7 were shared by two clusters and Item 11 was shared by all three clusters. This kind of sharing situation is common in collaborative filtering tasks, where some popular items are preferred by many persons having different preference characteristics. In the same manner with Sec. 4.1, the data set 2-(a) has a noise-less boundary while the data set 2-(b) includes 10% noise.

The FCCM algorithm was applied with $C = \{2, 3, 4, 5, 6\}$. The fuzzification weights were set as $\lambda_u = 0.1$ and $\lambda_v = 1.0$. Tables 5 and 6 compare the values of several validity indices and the selected cluster numbers whose index values are given by bold and italic. The tables show that PC and PE could not find the optimal cluster number of $C = 3$ but selected $C = 2$ as the best one because of the influence of the shared items. On the other hand, the proposed validity measure of $V_{co}$ could still find the optimal one even if some items were shared by multiple clusters.

The reason why the proposed $V_{co}$ works well in such sharing situations is that the proposed compactness and separateness measures fairly calculate the partition quality even if fuzzy memberships for the shared items become more ambiguous than dual exclusive situations. This result implies that the proposed cluster validation index for fuzzy co-clustering is useful for selecting the optimal fuzzy co-cluster structures.

4.3 Applicability to Collaborative Filtering Task

Finally, the applicability of the proposed co-cluster validity measure to collaborative filtering task is studied. A purchase history data set collected by Nikkei Inc. in 2000 is used in a collaborative filtering task. The data set was used in the previous works of [13,14] and includes the purchase history of 996 users ($n = 996$) on 18 items ($m = 18$). The element $r_{ij}$ of $996 \times 18$ relational data matrix $R = (r_{ij})$ is 1 if user $i$ has item $j$ while otherwise 0. Randomly selected 1,000 elements were used as a test data set for validating the recommendation ability, and the applicability of the proposed model to collaborative filtering task was evaluated by predicting the test elements based on co-clustering results. The FCCM algorithm was applied with various cluster numbers of $C = \{2, 3, \ldots, 9, 10\}$ and the recommendation ability was evaluated in conjunction with the proposed fuzzy co-cluster validity measure.

In the co-clustering-based prediction procedure, each user cluster is first estimated by maximum membership assignment, and then, the membership of each item in the user cluster is drawn from co-clustering results. If the item has a large membership in the user cluster, the item is recommended to the user. The recommendation ability is assessed by ROC sensitivity [18]. The ROC curve is a true positive rate vs. false positive rate plots drawn by changing the threshold of the applicability level in recommendation.
and the lower area of the curve becomes large as the recommendation ability is higher.

Figure 1 compares the ROC curves for various cluster numbers and indicates that the recommendation model with $C = 5$ has the best recommendation ability.

<table>
<thead>
<tr>
<th>$C$</th>
<th>$V_{co}$</th>
<th>ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.001</td>
<td>0.821</td>
</tr>
<tr>
<td>3</td>
<td>1.090</td>
<td>0.820</td>
</tr>
<tr>
<td>4</td>
<td>1.110</td>
<td>0.821</td>
</tr>
<tr>
<td>5</td>
<td>1.112</td>
<td>0.835</td>
</tr>
<tr>
<td>6</td>
<td>1.109</td>
<td>0.829</td>
</tr>
<tr>
<td>7</td>
<td>1.100</td>
<td>0.822</td>
</tr>
<tr>
<td>8</td>
<td>1.089</td>
<td>0.825</td>
</tr>
<tr>
<td>9</td>
<td>1.077</td>
<td>0.825</td>
</tr>
<tr>
<td>10</td>
<td>1.063</td>
<td>0.822</td>
</tr>
</tbody>
</table>

5. Conclusions

In this paper, the applicability of the conventional FCM-type cluster validity indices to fuzzy co-clustering tasks was discussed. First, several modifications of the measures for partition quality such as Partition Coefficient and Partition Entropy were considered and several experimental results implied that such indices are only applicable for dual exclusive situations, where both each of objects and items is exclusively assigned to a solo cluster.

Second, a new fuzzy co-cluster validation index of $V_{co}$ was proposed, which is a co-clustering version of the conventional Xie-Beni index for FCM clustering. Compactness and Separateness considered in Xie-Beni index were newly re-defined in the context of fuzzy co-clustering, which is a prototype-less clustering model and geometrical features cannot constructed from prototypes. Several experimental results implied that the proposed index works well not only in dual exclusive situations but also in item sharing situations, which is common in collaborative filtering tasks.

Finally, the applicability of the proposed index was studied in a collaborative filtering problem. It was proved that the optimal fuzzy co-cluster partition selected by the proposed index is also useful in co-cluster-based recommendation tasks.

Possible future works include the development of the co-cluster version of other FCM-type validation indices and the comparative study in other co-cluster applications such as document analysis and product management.
Acknowledgments

This work was supported in part by the Japan Science and Technology Agency (JST) through Support for Accelerating Utilization of University IP, and by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under Grant-in-Aid for Scientific Research (23500283).

References