

## Research Article

# A Semi-Supervised Framework for MMMs-Induced Fuzzy Co-Clustering with Virtual Samples

**Daiji Tanaka, Katsuhiko Honda, Seiki Ubukata, and Akira Notsu**

*Graduate School of Engineering, Osaka Prefecture University, Sakai, Osaka 599-8531, Japan*

Correspondence should be addressed to Katsuhiko Honda; [honda@cs.osakafu-u.ac.jp](mailto:honda@cs.osakafu-u.ac.jp)

Received 9 February 2016; Accepted 25 May 2016

Academic Editor: Kemal Kilic

Copyright © 2016 Daiji Tanaka et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Although the goal of clustering is to reveal structural information from unlabeled datasets, in cases with partial structural supervisions, semi-supervised clustering is expected to improve partition quality. However, in many real applications, it may cause additional costs to provide an enough amount of supervised objects with class labels. A virtual sample approach is a practical technique for improving classification quality in semi-supervised learning, in which additional virtual samples are generated from supervised objects. In this research, the virtual sample approach is adopted in semi-supervised fuzzy co-clustering, where the goal is to reveal object-item pairwise cluster structures from cooccurrence information among them. Several experimental results demonstrate the characteristics of the proposed approach.

## 1. Introduction

Clustering or cluster analysis is a basic technique for unsupervised classification, whose goal is to reveal intrinsic substructures varied in large scale unlabeled datasets. In some applications, however, it is possible to utilize partial knowledge on the substructures [1], such as *must-link* and *cannot-link* among some objects [2–4], class labels on a part of objects [5–7], or predefined fuzzy membership degrees [8, 9], and it is expected that partition quality can be significantly improved by utilizing such partial knowledge. In this study, a situation of having semi-supervision is considered, in which we have some supervising objects in conjunction with their class labels. Semi-supervised clustering [5, 10] is a practical approach for utilizing partial supervised information with the goal of improving the partition quality in unsupervised classification. Such partial knowledge can be utilized in two phases of *supervised initialization* and *supervised membership assignment* [6, 7].

Although partition quality is expected to be improved with sufficient semi-supervisions, in many real world applications, it is often difficult to utilize an enough amount of supervised objects. For example, in many web data analyses, various open data are available while most of them are unlabeled and their class labels cannot be provided without

heavy costs. A promising approach to improve classification quality in semi-supervised learning is the virtual sample approach, in which additional virtual samples are artificially generated from several supervised objects [11–13].

In this paper, the virtual sample approach is adopted in semi-supervised fuzzy co-clustering while the conventional model [13] was designed for semi-supervised classification in multidimensional data space. Fuzzy co-clustering is a fundamental technique for summarizing mutual cooccurrence information among objects and items such as document-keyword frequencies in document analysis [14] and customer-products preferences in purchase history analysis [15]. The task of co-clustering is achieved by simultaneously estimating memberships of both objects and items from a cooccurrence information matrix. Fuzzy co-clustering induced by MMMs concept (FCCMM) [16] is a fuzzy co-clustering algorithm induced by multinomial mixture models (MMM) [17], where statistical mixture models are interpreted as *k*-means-type classification models with regularized objective functions. The iterative algorithm is composed of updating two types of memberships of items and objects. In semi-supervised fuzzy co-clustering [18], partial knowledge was utilized in two phases of *supervised initialization* and *supervised object membership assignment* in a similar manner to the framework of [6, 7]. In *supervised initialization*, initial

item memberships are generated using labeled objects only. Then, in *supervised object membership assignment*, memberships of supervised objects are prefixed while others are updated in the iterative algorithm.

The remaining parts of this paper are organized as follows: Section 2 provides a brief review of the MMMs-induced fuzzy co-clustering, which is then combined with a semi-supervised framework. An artificial process for generating virtual samples is introduced in Section 3. Section 4 presents experimental results and Section 5 gives summary conclusions.

## 2. MMMs-Induced Fuzzy Co-Clustering and Its Variant for Semi-Supervised Clustering

### 2.1. Fuzzy Co-Clustering Model Induced by MMMs Concept.

Assume that we have a cooccurrence matrix  $R = \{r_{ij}\}$  on objects  $i = 1, \dots, n$  and items  $j = 1, \dots, m$ , in which  $r_{ij}$  represents the cooccurrence degree of object  $i$  and item  $j$ . For example,  $r_{ij}$  can be the frequency of keyword (item)  $j$  in document (object)  $i$  in document analysis. In many traditional results, it has been shown that co-cluster structure analysis is useful for summarizing contents such as intrinsic chapter or category information of many documents with their representative keywords. In fuzzy co-clustering context, the task is reduced to the problem of finding the pairwise clusters of mutually familiar objects and items, in which the goal is to estimate the fuzzy memberships of both objects and items such that mutually familiar object and item pairs have large memberships in the same cluster.

MMMs [17] is a probabilistic mixture model for co-clustering, in which each component density is multinomial distribution. Multinomial distribution is a multicategory extension of binomial distribution, where the probability of an object with item cooccurrence vector  $r_i = (r_{i1}, \dots, r_{im})^\top$  is defined as the joint probability of all items with their frequencies of  $r_{ij}$ . MMMs construct a mixture distribution by iteratively estimating the item occurrence probability and the a priori probability of generative distributions in conjunction with estimating object a posteriori probability to each generative model.

Following the soft clustering interpretation of probabilistic mixture models [19], Honda et al. [16] introduced a fuzzy co-clustering-based interpretation of MMMs, where the pseudo-log-likelihood function was decomposed into an object-item aggregation measure and K-L information-based fuzzification penalty. FCCMM [16] is an MMMs-induced fuzzy co-clustering model, in which the degree of object partition fuzziness can be tuned with an adjustable penalty weight on the K-L information-based penalty. For extracting  $C$  fuzzy co-clusters, the objective function to be maximized is defined as

$$L_{\text{FCCMM}} = \sum_{c=1}^C \sum_{i=1}^n \sum_{j=1}^m u_{ci} r_{ij} \log w_{cj} + \lambda_u \sum_{c=1}^C \sum_{i=1}^n u_{ci} \log \frac{\pi_c}{u_{ci}}, \quad (1)$$

where  $u_{ci}$  and  $w_{cj}$  are the fuzzy memberships of object  $i$  and item  $j$  to cluster  $c$ , respectively.  $\pi_c$  is the volume of cluster  $c$  such that  $\sum_{c=1}^C \pi_c = 1$ . From the MMMs viewpoint,  $u_{ci}$

is the a posteriori probability of class  $c$  given object  $i$  and should be constrained to  $\sum_{c=1}^C u_{ci} = 1$ . On the other hand,  $w_{cj}$  corresponds to the generative probability of item  $j$  in component  $c$  and should be  $\sum_{j=1}^m w_{cj} = 1$ . Then,  $u_{ci}$  is mainly responsible for exclusive object partition while  $w_{cj}$  just represents the typicality of item  $j$  within a generative model.

$\sum_{c=1}^C \sum_{i=1}^n \sum_{j=1}^m u_{ci} r_{ij} \log w_{cj}$  is the aggregation criterion of objects and items, which measures the degree of aggregation of familiar objects and items, and becomes larger when mutually familiar object-item pairs with large  $r_{ij}$  have large memberships  $u_{ci}$  and  $w_{cj}$  in a same cluster  $c$ . From the viewpoint of  $k$ -means-type clustering with respect to object memberships  $u_{ci}$ , the aggregation criterion is essentially a (linear) hard clustering measure, and the K-L information term [20, 21] is responsible for estimating soft partition of objects in the same manner with the soft partition nature of Gaussian mixture models [19].  $\lambda_u$  tunes the degree of fuzziness of object partition. When  $\lambda_u = 1$ , the objective function is reduced to the pseudo-log-likelihood function of MMMs with  $C$  component distributions. Then, in the case of  $\lambda_u > 1$ , FCCMM brings a fuzzier co-cluster partition than MMMs while object partition becomes more crisp with  $\lambda_u < 1$ . As an extremal case,  $\lambda_u \rightarrow 0$  implies a crisp co-clustering model. It was shown that a careful tuning of fuzziness degree can contribute to improvement of partition quality of MMMs by reducing initialization and noise sensitivities [16].

The updating rules for these model parameters are given as

$$u_{ci} = \frac{\pi_c \prod_{j=1}^m (w_{cj})^{r_{ij}/\lambda_u}}{\sum_{\ell=1}^C \pi_\ell \prod_{j=1}^m (w_{\ell j})^{r_{ij}/\lambda_u}}, \quad (2)$$

$$\pi_c = \frac{1}{n} \sum_{i=1}^n u_{ci}, \quad (3)$$

$$w_{cj} = \frac{\sum_{i=1}^n r_{ij} u_{ci}}{\sum_{\ell=1}^m \sum_{i=1}^n r_{i\ell} u_{ci}}. \quad (4)$$

The clustering algorithm is the 3-step iterative process composed of these updating rules.

### Algorithm of FCCMM

- (1) Let  $C$  be the number of clusters. Choose the fuzzification weight  $\lambda_u$ .
- (2) *Initialization*: initialize object memberships  $u_{ci}$ ,  $\forall c, i$  with random values such that  $\sum_{c=1}^C u_{ci} = 1$ .
- (3) *Iterative Process*: iterate the following process until convergence of all  $u_{ci}$ .
  - (a) Update cluster volumes  $\pi_c$  using (3).
  - (b) Update item memberships  $w_{cj}$  using (4).
  - (c) Update object memberships  $u_{ci}$  using (2).

**2.2. Semi-Supervised Fuzzy Co-Clustering and Inductive Classification.** When some objects have their intrinsic class labels,

the partition quality of unsupervised clustering is expected to be improved with their supports [5, 10]. A possible semi-supervised framework for fuzzy co-clustering [18] utilized such partial supervision in two levels: *supervised initialization* and *supervised membership assignment*. Assume that we have an  $n \times m$  cooccurrence matrix, where only a part of objects, such that  $i \in O_s$  ( $|O_s| < n$ ), has class labels but others do not. A framework of semi-supervised fuzzy co-clustering is given as follows [18].

#### Semi-Supervised Fuzzy Co-Clustering Framework

- (1) Let  $C$  be the number of clusters, which is usually equivalent to the class number of the supervised objects. Choose the fuzzification weight  $\lambda_u$ .
- (2) *Initialization*: initialize the memberships of supervised objects  $i \in O_s$  such that  $u_{ci} = 1$  in the labeled class  $c$  and  $u_{ci} = 0$  in others. For unsupervised objects  $i \notin O_s$ ,  $u_{ci} = \varepsilon$ ,  $\forall c, i$ , where  $\varepsilon$  is a small positive. (The initial cluster volumes and item memberships are estimated by mainly reflecting supervised objects only.)
- (3) *Iterative Process*: iterate the following process until convergence of all  $u_{ci}$ .
  - (a) Update cluster volumes  $\pi_c$  using (3).
  - (b) Update item memberships  $w_{cj}$  using (4).
  - (c) Update memberships  $u_{ci}$  of unsupervised object  $i \notin O_s$  using (2). (The memberships of supervised objects are prefixed and unchanged.)

In the initialization level, a plausible initial co-clusters are estimated considering only the supervised objects. Here, this approach is available only if the amount of supervised objects is enough for estimating cluster-wise item preferences, and each cluster index strictly corresponds to the intrinsic class index. In case with insufficient supervision, this initialization step should be performed with the conventional procedure of random assignment for avoiding overfitting to a few supervised objects.

Next, in the iterative optimization level, a fixed crisp object membership is assigned to each supervised object reflecting its class label. The partial supervision can contribute to guiding other unlabeled objects to plausible co-clusters. This second level is expected to be useful even if we have only a few supervised objects.

Once we got a co-cluster structure, we can perform inductive classification of new (unlabeled) objects supported by the co-cluster information [18].

#### Inductive Classification Considering Fuzzy Co-Cluster Structure

- (1) Assume that we have estimated co-cluster model parameters of  $\pi_c$  and  $w_{cj}$ ,  $c = 1, \dots, C$ ,  $j = 1, \dots, m$ . The goal is to predict the class of a new test object  $i^*$ , which is associated with its cooccurrence information  $r_{i^*} = (r_{i^*1}, \dots, r_{i^*m})$ .

- (2) *Membership Calculation*: calculate fuzzy memberships  $u_{ci^*}$ ,  $c = 1, \dots, C$  by (2).
- (3) *Maximum Membership Assignment*: search for the largest  $u_{ci^*}$  and output its class label.

This inductive classification approach was shown to outperform the supervised classification with a small set of supervised objects only [18]. This implies that unlabeled objects can contribute to effectively estimating class-wise distributions rather than supervised learning with insufficient supervisions.

### 3. Generation of Virtual Samples for Semi-Supervised Fuzzy Co-Clustering

Although the performance of semi-supervised clustering is expected to be improved as the number of labeled objects becomes larger, it may cause a high cost to generate an enough amount of supervised objects in real applications. The virtual sample approach is a practical strategy for improving classification quality in semi-supervised learning without additional costs. In the remaining parts of this paper, the virtual sample approach is adopted in semi-supervised fuzzy co-clustering.

Sassano [13] proposed two methods for generating virtual samples for text classification based on the following assumption:

*The category of a document is unchanged even if a small number of words are added or deleted.*

Because the documents belonging to the same category usually contain several common keywords, deletion or addition of a small number of words is expected not to have a severe impact on classification quality. In [13], two strategies of *Deletion* and *Addition* were considered and were utilized in semi-supervised support vector machine learning of text documents. In *GenerateByDeletion*, virtual samples were generated by deleting some portions of the original supervised documents and were added to the class of the original ones. On the other hand, in *GenerateByAddition*, virtual samples were generated by adding a small number of words into the original supervised documents. The words to be added are taken from documents, the label of which is the same as that of the original document.

In this paper, the two generative strategies are introduced to semi-supervised fuzzy co-clustering tasks. Assume that  $r_i = (r_{i1}, \dots, r_{im})$  is the cooccurrence information vector on a supervised object  $i \in O_s$ , where  $r_{ij} \in \{0, 1\}$  just represents the appearance/absence of item  $j$  in object  $i$ . Its virtual copy  $r'_i = (r'_{i1}, \dots, r'_{im})$  is given as an additional virtual (supervised) sample to be added to partial supervision after slight revisions based on the two strategies.

#### *GenerateByDeletion*

- (1) Copy  $r_i$  to  $r'_i$ .
- (2) For each item  $j$  with  $r_{ij} = 1$ , if  $\text{rand}() \leq t$  then set  $r'_{ij} = 0$ , where  $\text{rand}()$  is a function which generates

TABLE 1: A toy example of supervised objects set.

Object ID: $i$	Cooccurrence information vector $r_i = (r_{i1}, \dots, r_{i10})$	Class label
1	(1, 1, 0, 1, 1, 0, 0, 0, 0, 1)	1
2	(1, 0, 1, 1, 0, 1, 0, 0, 0, 0)	1
3	(1, 1, 1, 0, 1, 0, 0, 0, 0, 1)	1
4	(0, 0, 0, 0, 0, 1, 1, 0, 1, 1)	2
5	(0, 0, 0, 0, 1, 0, 1, 1, 0, 0)	2
6	(0, 0, 0, 0, 0, 1, 0, 1, 1, 1)	2

a random number from 0 to 1, and  $t$  is a parameter to decide how many elements are deleted.

This procedure implies that each virtual object is a virtual copy of the original supervised object having a smaller number of appearances. Virtual object  $r'_i$  is almost equivalent to the original  $r_i$  with  $t \rightarrow 0$  while  $r'_i$  becomes sparse as  $t$  is larger.

Here, the detailed process is demonstrated with a toy example using a set of 6 supervised objects ( $|O_s| = 6$ ) shown in Table 1, where each cooccurrence information vector  $r_i$  is composed of ten items ( $m = 10$ ) and the objects belong to one of two supervised classes ( $C = 2$ ). Be noted that class 1 is mainly related to the first 5 items while class 2 is related to others.

Using *GenerateByDeletion*, class 1 object  $r_1$  can generate such virtual objects as (1, 1, 0, 1, 0, 0, 0, 0, 0, 1), (1, 1, 0, 0, 1, 0, 0, 0, 0, 1), or (1, 0, 0, 1, 1, 0, 0, 0, 0, 0) with label “1”. Another class 2 object  $r_5$  can generate such as (0, 0, 0, 0, 0, 0, 1, 1, 0, 0), (0, 0, 0, 0, 1, 0, 1, 0, 0, 0), or (0, 0, 0, 0, 0, 0, 0, 1, 0, 0) with label “2”.

#### *GenerateByAddition*

- (1) Construct the set of all supervised objects, whose class labels are the same as that of  $r_i$ , and generate a temporal item set composed of all items that appeared in the supervised object set.
- (2) Copy  $r_i$  to  $r'_i$ .
- (3) For each item  $j$  with  $r_{ij} = 1$ , if  $\text{rand}() \leq t$  then randomly select item  $a$  from the temporal item set and set as  $r'_{ia} = 1$ .

This procedure implies that each virtual object is a virtual copy of the original supervised object having a larger number of appearances. Virtual object  $r'_i$  is almost equivalent to the original  $r_i$  with  $t \rightarrow 0$  while  $r'_i$  becomes dense as  $t$  is larger.

In the toy example of Table 1, *GenerateByAddition* first constructs temporal item sets as {1, 2, 4, 5, 10, 1, 3, 4, 6, 1, 2, 3, 5, 10} for class 1 and {6, 7, 9, 10, 5, 7, 8, 6, 8, 9, 10} for class 2. Then, a class 1 object  $r_1$  can generate such virtual objects as (1, 1, 1, 1, 1, 0, 0, 0, 0, 1), (1, 1, 0, 1, 1, 1, 0, 0, 0, 1), or (1, 1, 1, 1, 1, 1, 0, 0, 0, 1) with label “1”. Another class 2 object  $r_5$  can generate such as (0, 0, 0, 0, 1, 1, 1, 1, 0, 0), (0, 0, 0, 0, 1, 0, 1, 1, 1, 0), or (0, 0, 0, 0, 1, 1, 1, 1, 0, 1) with label

“2”. Here, it should be noted that  $r'_{ij}$  is always  $r'_{ij} = 0$  if item  $j$  does not appear in all supervised objects of the class.

The generated virtual objects are added to the set of supervised objects and are utilized as semi-supervision in semi-supervised learning.

## 4. Numerical Experiment

The classification quality of semi-supervised fuzzy co-clustering with virtual objects is investigated through numerical experiments in this section.

Two benchmark datasets *CiteSeer dataset* and *Cora dataset* were used, which are available from LINQS webpage of Statistical Relational Learning Group UMD (<http://linqs.cs.umd.edu/projects/index.shtml>). *CiteSeer dataset* consists of 3312 scientific publications (objects,  $n = 3312$ ), each of which is described by a 0/1-valued word vector ( $r_{ij} \in \{0, 1\}$ ) indicating the absence/presence of 3703 unique words (items,  $m = 3703$ ) and is classified into one of six classes ( $C = 6$ ). *Cora dataset* also consists of 2708 scientific publications (objects,  $n = 2708$ ) and 1433 unique words (items,  $m = 1433$ ) with  $r_{ij} \in \{0, 1\}$ , each of which is classified into one of seven classes ( $C = 7$ ).

The classification quality was investigated through 5-fold cross-validation scheme. In applying semi-supervised fuzzy co-clustering, the dataset was first partitioned into 5 disjoint subsets. Four subsets were utilized for the training set to be used in semi-supervised fuzzy co-clustering and the remaining one subset was used for the test set to validate the classification ability. This training/test trials were iterated 5 times rotating different test subsets. The number of clusters was set as the actual class number; that is,  $C = 6$  in *CiteSeer dataset* and  $C = 7$  in *Cora dataset*. The fuzzification weight was set as  $\lambda_u = 5.0$  for *CiteSeer dataset* and  $\lambda_u = 3.0$  for *Cora dataset*, respectively, such that a slightly fuzzier model rather than MMMs with  $\lambda_u = 1.0$  can contribute to better performances [16].

*4.1. Preliminary Experiment.* First, a preliminary experiment was performed with the goal of investigating the effects of the amount of supervised objects, where the semi-supervised framework was implemented *without* virtual objects. The amount of supervised objects, which are available in training data, was varied with various ratios of {0%, 1%, 5%, 10%, 20%, 50%, 100%} for two benchmark datasets but the class labels of the remaining objects were withheld to be unknown. “0%” corresponds to the conventional unsupervised model, which utilizes no supervisions. “100%” corresponds to the fully supervised model, where all training objects have their class information and  $C$  co-cluster models (cluster-wise item memberships) were independently estimated in each class. After co-cluster estimation, the classes of the unsupervised training objects were predicted with the largest memberships. The classes of the test objects were also predicted by the inductive classification scheme of Section 2.

The recognition rates are compared in Table 2 for *CiteSeer dataset* and Table 3 for *Cora dataset*. “100%” for training data are missing as shown with “—” because no unsupervised

TABLE 2: Comparison of recognition rates in preliminary experiment (*CiteSeer*).

	Proportion of supervision						
	0%	1%	5%	10%	20%	50%	100%
Training	0.528	0.542	0.677	0.694	0.701	0.719	—
Test	0.480	0.488	0.635	0.648	0.646	0.633	0.662

TABLE 3: Comparison of recognition rates in preliminary experiment (*Cora*).

	Proportion of supervision						
	0%	1%	5%	10%	20%	50%	100%
Training	0.465	0.528	0.664	0.675	0.703	0.735	—
Test	0.416	0.457	0.590	0.592	0.603	0.612	0.619

objects remained in training sets. In both *CiteSeer dataset* and *Cora dataset*, the recognition rates of unsupervised co-clustering (0% supervision) are inferior to those of semi-supervised co-clustering. These results imply that semi-supervised co-clustering is more effective than nonsupervised co-clustering when partial supervisions are available, and the partition quality becomes better as the amount of available supervisions is larger. And the recognition rates in 5% supervision or larger are almost comparative while the 1% supervision is significantly inferior to others. Then, if we have 1% supervision only, it is quite difficult to achieve sufficient recognition rates and we need additional information for improving the partition quality.

By the way, in these datasets, too many supervisions, such as more than 10%, could not contribute to further improvement of test evaluation. It may be because too many supervisions can bring the overfitting to training samples and cause poor generalization capability. Then, semi-supervised learning is expected to contribute to improving generalization capability rather than fully supervised learning.

In the following experiments, a situation with 1% partial supervision was designed for simulating an insufficient partial knowledge case, where the goal is to demonstrate the advantage of additional virtual partial supervision.

**4.2. Investigation of Effects of Virtual Objects.** Next, the effect of virtual objects in semi-supervised fuzzy co-clustering was investigated. In this experiment, the classification abilities of semi-supervised fuzzy co-clustering of the cases with/without virtual objects were compared. Following the result of the above preliminary experiment, 1% supervised objects were included in training sets for semi-supervised co-clustering as a simulation of the situation where an enough amount of supervised objects is not available. The virtual objects were generated from each of these 1% supervised objects by *GenerateByDeletion* and *GenerateByAddition*, and then they were added to training sets and utilized as additional supervision. Supported by the additional information of virtual objects, the partition quality is expected to be improved rather than the original 1% supervision case.

Additionally, the influences of the number of virtual objects were also investigated, where the number of virtual

TABLE 4: Comparison of recognition rates with and without virtual objects (*CiteSeer-training*).

Number of virtual objects per supervised object	$t$			
	0.01	0.05	0.1	0.2
No virtual objects	0.542			
1 (GBD)	<b>0.551</b>	<b>0.551</b>	<b>0.548</b>	<b>0.546</b>
1 (GBA)	<b>0.552</b>	<b>0.545</b>	<b>0.549</b>	<b>0.543</b>
2 (GBD)	<b>0.554</b>	<b>0.550</b>	<b>0.554</b>	<b>0.550</b>
2 (GBA)	<b>0.553</b>	<b>0.550</b>	<b>0.548</b>	<b>0.546</b>
2 (GBD+GBA)	<b>0.554</b>	<b>0.554</b>	<b>0.547</b>	<b>0.549</b>
4 (GBD+GBA)	0.541	<b>0.542</b>	<b>0.534</b>	<b>0.533</b>
5 (GBD)	0.534	0.536	0.532	<b>0.544</b>
5 (GBA)	0.504	0.530	0.527	0.532

TABLE 5: Comparison of recognition rates with and without virtual objects (*CiteSeer-test*).

Number of virtual objects per supervised object	$t$			
	0.01	0.05	0.1	0.2
No virtual objects	0.488			
1 (GBD)	<b>0.498</b>	<b>0.505</b>	<b>0.502</b>	<b>0.493</b>
1 (GBA)	<b>0.501</b>	<b>0.496</b>	<b>0.493</b>	<b>0.488</b>
2 (GBD)	<b>0.502</b>	<b>0.494</b>	<b>0.504</b>	<b>0.492</b>
2 (GBA)	<b>0.503</b>	<b>0.497</b>	<b>0.495</b>	<b>0.494</b>
2 (GBD+GBA)	<b>0.503</b>	<b>0.503</b>	<b>0.494</b>	<b>0.499</b>
4 (GBD+GBA)	0.487	0.486	0.475	0.479
5 (GBD)	0.475	0.477	0.470	<b>0.491</b>
5 (GBA)	0.474	0.469	0.464	0.472

objects was increased by iterating the generative procedures. Furthermore, a hybrid method was also applied using both *GenerateByDeletion* and *GenerateByAddition*, simultaneously. In this hybrid method, at least 2 virtual objects are generated from one supervised object.

Both the recognition rates of training and test objects were estimated in the same manner with the preliminary experiment. Tables 4 and 5 show the results with *CiteSeer dataset*. Tables 6 and 7 show the results with *Cora dataset*. Bold types indicate the improved quality rather than the original 1% supervision case (without virtual objects). “GBD” and “GBA” mean *GenerateByDeletion* and *GenerateByAddition*, respectively, and “GBD+GBA” is their hybrid method. In this experiment, the number of virtual objects per supervised object was varied in {1, 2, 5} adopting different random seeds in *GenerateByDeletion* and *GenerateByAddition* and varied in {2, 4} adopting different random seeds in the hybrid method. Parameter  $t$  of addition/deletion rate was varied with  $t = \{0.01, 0.05, 0.1, 0.2\}$ .

In both training and test cases, the recognition rates of semi-supervised co-clustering with virtual objects are superior to those without virtual objects when the number of virtual objects generated from a supervised objects is relatively small. These results indicate that virtual objects can contribute to improving

TABLE 6: Comparison of recognition rates with and without virtual objects (*Cora*-training).

Number of virtual objects per supervised object	$t$			
	0.01	0.05	0.1	0.2
No virtual objects	0.528			
1 (GBD)	<b>0.539</b>	<b>0.540</b>	<b>0.548</b>	<b>0.536</b>
1 (GBA)	<b>0.539</b>	<b>0.546</b>	<b>0.536</b>	<b>0.537</b>
2 (GBD)	<b>0.540</b>	<b>0.536</b>	<b>0.537</b>	<b>0.544</b>
2 (GBA)	<b>0.539</b>	<b>0.541</b>	<b>0.543</b>	<b>0.542</b>
2 (GBD+GBA)	<b>0.542</b>	<b>0.541</b>	<b>0.530</b>	<b>0.541</b>
4 (GBD+GBA)	<b>0.528</b>	<b>0.529</b>	<b>0.531</b>	<b>0.530</b>
5 (GBD)	0.525	0.512	<b>0.529</b>	<b>0.536</b>
5 (GBA)	0.525	0.521	0.503	0.504

TABLE 7: Comparison of recognition rates with and without virtual objects (*Cora*-test).

Number of virtual objects per supervised object	$t$			
	0.01	0.05	0.1	0.2
No virtual objects	0.457			
1 (GBD)	<b>0.478</b>	<b>0.479</b>	<b>0.489</b>	<b>0.476</b>
1 (GBA)	<b>0.483</b>	<b>0.487</b>	<b>0.474</b>	<b>0.474</b>
2 (GBD)	<b>0.475</b>	<b>0.475</b>	<b>0.475</b>	<b>0.487</b>
2 (GBA)	<b>0.476</b>	<b>0.482</b>	<b>0.476</b>	<b>0.480</b>
2 (GBD+GBA)	<b>0.479</b>	<b>0.479</b>	<b>0.472</b>	<b>0.484</b>
4 (GBD+GBA)	<b>0.462</b>	<b>0.463</b>	<b>0.465</b>	<b>0.467</b>
5 (GBD)	0.440	0.429	<b>0.458</b>	<b>0.460</b>
5 (GBA)	0.441	0.438	0.411	0.409

the classification ability of semi-supervised fuzzy co-clustering when we do not have enough semi-supervision in datasets.

However, it seems that too many virtual objects degrade the partition quality. In this experiment, the quality was not improved when 5 or more virtual objects per supervised objects were generated. This indicates that a small number of virtual objects can contribute to improving the quality of semi-supervision while too many virtual objects bring a negative impact on the effect of semi-supervision. It may be because too many virtual objects can bring the overfitting to the virtual objects and cause poor generalization capability. Then, it will be important to generate a proper volume of virtual objects.

**4.3. Validation of Class-Wise Partition Quality.** Finally, the validity of class-wise partition quality is studied, where the best results of the virtual sample approach are compared with the original one given without virtual objects. The best results were given with  $\{1(\text{GBD}), t = 0.05\}$  for the test set of *CiteSeer* dataset and  $\{1(\text{GBD}), t = 0.1\}$  for the test set of *Cora* dataset, respectively.

TABLE 8: Example of classification result of a class  $c$ .

Actual class label	Output class label	
	Positive	Negative
True	$\alpha$	$\gamma$
False	$\beta$	$\delta$

TABLE 9: Comparison of classification ability of each class, *CiteSeer*.

Class label	No virtual objects	$\{1(\text{GBD}), t = 0.05\}$
Agents	0.655	0.651
AI	0.103	<b>0.107</b>
DB	0.346	<b>0.374</b>
IR	0.484	<b>0.512</b>
ML	0.464	<b>0.465</b>
HCI	0.629	0.627

TABLE 10: Comparison of classification ability of each class, *Cora*.

Class label	No virtual objects	$\{1(\text{GBD}), t = 0.1\}$
Case_based	0.533	<b>0.584</b>
Genetic_algorithms	0.400	<b>0.406</b>
Neural_networks	0.632	<b>0.662</b>
Probabilistic_methods	—	<b>0.416</b>
<i>Reinforcement_learning</i>	—	—
Rule_learning	0.170	<b>0.196</b>
Theory	0.248	<b>0.278</b>

The quality of result of each class was validated by  $F$ -measure. Assume that the classification result in class  $c$  is shown as in Table 8.  $F$ -measure is defined as follows:

$$F = \frac{2PR}{P + R}, \quad (5)$$

where  $P$  and  $R$  are precision and recall, respectively, as follows:

$$P = \frac{\alpha}{\alpha + \beta}, \quad (6)$$

$$R = \frac{\alpha}{\alpha + \gamma}.$$

Precision  $P$  means the proportion of objects with actual class label  $c$  out of objects classified to class  $c$ . Recall  $R$  means the proportion of objects classified to class  $c$  out of objects with actual class label  $c$ .  $F$ -measure is the harmonic mean of precision  $P$  and recall  $R$ , and larger  $F$ -measure implies better classification ability. Supported by the 5-fold cross-validation scheme, the average of 5 trials is compared.

Tables 9 and 10 show  $F$ -measure of each class for test objects. “—” denotes the failure of  $F$ -measure calculation caused by missing of  $P$  or  $R$ , where the clustering model failed to find the corresponding class objects in a trial after maximum memberships classification with very fuzzy boundaries. These tables indicate that the results with virtual objects are superior to those without virtual objects for most

classes except for *Agents* and *HCI* in *CiteSeer dataset* and *Reinforcement.Learning* in *Cora dataset*. Therefore, virtual objects can mostly contribute to improving partition quality from the viewpoint of class-wise recognition.

## 5. Conclusion

In this paper, the effect of virtual objects in semi-supervised fuzzy co-clustering was demonstrated. Following the previous study [13], two novel procedures for generating virtual objects for cooccurrence data analysis were proposed and their utility was investigated through numerical experiments. In the numerical experiments with two benchmark datasets, the effects of virtual objects in conjunction with the effects of the number of virtual objects per original supervised object were compared and it was indicated that classification quality of semi-supervised fuzzy co-clustering can be improved without additional cost for generating supervised objects by adding several virtual objects while the classification quality can be degraded with too many virtual objects.

Future work includes improving the quality of virtual samples. For example, it may be possible to improve the quality of the virtual samples by evaluating the plausibility of the additional samples with some cluster validity measures for fuzzy co-clustering [22]. Furthermore, besides the simple virtual copy of objects, virtual copy of items may be another possible direction.

## Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

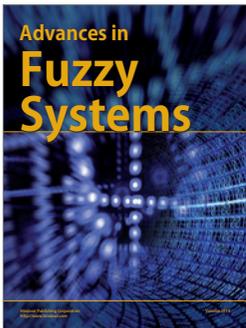
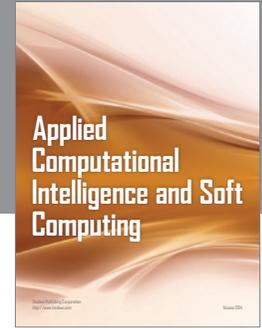
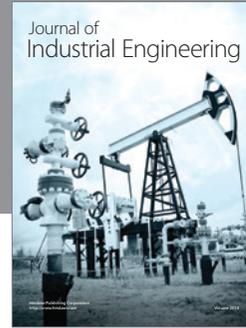
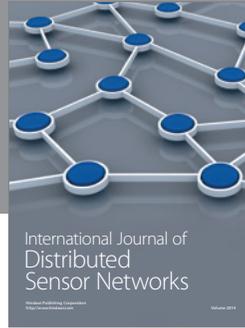
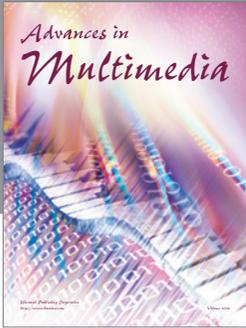
## Acknowledgments

This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under Grant-in-Aid for Scientific Research (no. 26330281).

## References

- [1] P. H. Thong and L. H. Son, "An overview of semi-supervised fuzzy clustering algorithms," *International Journal of Engineering and Technology*, vol. 8, no. 4, pp. 301–306, 2016.
- [2] W. Na, L. Xia, and L. Xuehui, "Semi-supervised Kernel-based fuzzy  $c$ -means with pairwise constraints," in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '08)*, pp. 1098–1102, Hong Kong, June 2008.
- [3] S. Miyamoto and A. Terami, "Semi-supervised agglomerative hierarchical clustering algorithms with pairwise constraints," in *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ '10)*, pp. 1–6, IEEE, Barcelona, Spain, July 2010.
- [4] Y. Hamasuna, Y. Endo, and S. Miyamoto, "Semi-supervised fuzzy  $c$ -means clustering using clusterwise tolerance based pairwise constraints," in *Proceedings of the IEEE International Conference on Granular Computing (GrC '10)*, pp. 188–193, San Jose, Calif, USA, August 2010.
- [5] N. Grira, M. Crucianu, and N. Boujemaa, "Unsupervised and semi-supervised clustering: a brief survey," Report of the MUSCLE European Network of Excellence, 2005.
- [6] S. Basu, A. Banerjee, and R. J. Mooney, "Semi-supervised clustering by seeding," in *Proceedings of the 19th International Conference on Machine Learning*, pp. 27–34, Sydney, Australia, 2002.
- [7] C.-L. Liu, T.-H. Chang, and H.-H. Li, "Clustering documents with labeled and unlabeled documents using fuzzy semi-Kmeans," *Fuzzy Sets and Systems*, vol. 221, pp. 48–64, 2013.
- [8] Y. Endo, Y. Hamasuna, M. Yamashiro, and S. Miyamoto, "On semi-supervised fuzzy  $c$ -means clustering," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, pp. 1119–1124, Jeju Island, South Korea, August 2009.
- [9] X. Yin, T. Shu, and Q. Huang, "Semi-supervised fuzzy clustering with metric learning and entropy regularization," *Knowledge-Based Systems*, vol. 35, pp. 304–311, 2012.
- [10] E. Bair, "Semi-supervised clustering methods," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 5, no. 5, pp. 349–361, 2013.
- [11] B. Schölkopf, C. Burges, and V. Vapnik, "Incorporating invariances in support vector learning machines," in *Proceedings of the International Conference on Artificial Neural Networks, Bochum, Germany, July 1996*, C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff, Eds., vol. 1112 of *Lecture Notes in Computer Science*, pp. 47–52, Springer, 1996.
- [12] P. Niyogi, F. Girosi, and T. Poggio, "Incorporating prior information in machine learning by creating virtual examples," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2196–2208, 1998.
- [13] M. Sassano, "Virtual examples for text classification with support vector machines," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 208–215, Sapporo, Japan, July 2003.
- [14] K. Kummamuru, A. Dhawale, and R. Krishnapuram, "Fuzzy co-clustering of documents and keywords," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, vol. 2, pp. 772–777, May 2003.
- [15] K. Honda, A. Notsu, and H. Ichihashi, "Collaborative filtering by sequential user-item co-cluster extraction from rectangular relational data," *International Journal of Knowledge Engineering and Soft Data Paradigms*, vol. 2, no. 4, p. 312, 2010.
- [16] K. Honda, S. Oshio, and A. Notsu, "Fuzzy co-clustering induced by multinomial mixture models," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 19, no. 6, pp. 717–726, 2015.
- [17] L. Rigouste, O. Cappé, and F. Yvon, "Inference and evaluation of the multinomial mixture model for text clustering," *Information Processing and Management*, vol. 43, no. 5, pp. 1260–1280, 2007.
- [18] K. Honda, S. Ubukata, A. Notsu, N. Takahashi, and Y. Ishikawa, "A semi-supervised fuzzy co-clustering framework and application to twitter data analysis," in *Proceedings of the International Conference on Informatics, Electronics & Vision (ICIEV '15)*, pp. 1–4, IEEE, Fukuoka, Japan, June 2015.
- [19] R. J. Hathaway, "Another interpretation of the EM algorithm for mixture distributions," *Statistics and Probability Letters*, vol. 4, no. 2, pp. 53–56, 1986.
- [20] H. Ichihashi, K. Miyagishi, and K. Honda, "Fuzzy  $c$ -means clustering with regularization by K-L information," in *Proceedings of the 10th IEEE International Conference on Fuzzy Systems*, vol. 2, pp. 924–927, IEEE, Melbourne, Australia, December 2001.

- [21] K. Honda and H. Ichihashi, "Regularized linear fuzzy clustering and probabilistic PCA mixture models," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 508–516, 2005.
- [22] K. Honda, M. Muranishi, A. Notsu, and H. Ichihashi, "FCM-type cluster validation in fuzzy co-clustering and collaborative filtering applicability," *International Journal of Computer Science and Network Security*, vol. 13, no. 1, pp. 24–29, 2013.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

