https://doi.org/10.20965/jaciii.2020.p0040

Paper:

A Heuristic-Based Model for MMMs-Induced Fuzzy Co-Clustering with Dual Exclusive Partition

Katsuhiro Honda, Yoshiki Hakui, Seiki Ubukata, and Akira Notsu

Osaka Prefecture University
1-1 Gakuen-cho, Nakaku, Sakai, Osaka 599-8531, Japan
E-mail: honda@cs.osakafu-u.ac.jp
[Received March 15, 2019; accepted August 29, 2019]

MMMs-induced fuzzy co-clustering achieves dual partition of objects and items by estimating two different types of fuzzy memberships. Because memberships of objects and items are usually estimated under different constraints, the conventional models mainly targeted object clusters only, but item memberships were designed for representing intra-cluster typicalities of items, which are independently estimated in each cluster. In order to improve the interpretability of co-clusters, meaningful items should not belong to multiple clusters such that each co-cluster is characterized by different representative items. In previous studies, the item sharing penalty approach has been applied to the MMMs-induced model but the dual exclusive constraints approach has not yet. In this paper, a heuristic-based approach in FCM-type co-clustering is modified for adopting in MMMs-induced fuzzy co-clustering and its characteristics are demonstrated through several comparative experiments.

Keywords: fuzzy co-clustering, dual exclusive partition, multinomial mixture models

1. Introduction

Fuzzy co-clustering is a basic technique for summarizing co-cluster structures varied in cooccurrence information matrices among objects and items and has been utilized in such cooccurrence analysis as bag-of-word document analysis with document-keyword cooccurrence or market analysis with customer-product purchase history. Fuzzy c-means (FCM) [1, 2] has been extended to fuzzy co-clustering models [3–5] in such a way that the within-cluster-error measure is replaced with the aggregation measures among objects and items. Dual partition of objects and items are achieved by estimating two different types of fuzzy memberships. Although object memberships follow the same constraint with FCM for achieving exclusive assignment to clusters under the sum-to-one condition with respect to the cluster index, item memberships represent the relative typicality in each cluster independently under the sum-to-one condition with respect to the item index. Because the FCM-type co-clusterings have no clear connection with any statistical co-clustering models, they often suffer from the lack of statistical guidance in fuzziness tuning. Fuzzy co-clustering induced by MMMs (FCCMM) [6] is a fuzzy counterpart of statistical multinomial mixture models (MMMs) [7], whose partition quality was shown to be improved through careful tuning of fuzziness degrees.

A possible improvement of above fuzzy co-clustering models is introduction of the exclusive constraint to item partition for interpretability. For example, if some items are very popular and have large typicalities in multiple clusters, several clusters are characterized by common items and their differences may be concealed. Then, it will be useful to emphasize the differences among co-clusters by avoiding the sharing of common representative items in multiple clusters.

In order to avoid item sharing in multiple clusters, we can adopt two approaches. The first approach is to introduce the penalty on item sharing without changing the original constraints. Honda et al. added an item sharing penalty to the objective function of FCM-type co-clustering [8], in which the penalty term is composed of the sum of item memberships such that it becomes large if some items have large memberships in multiple clusters. Then, each item can have large memberships in at most one cluster or has small memberships in all clusters.

The second approach is to force the FCM-like exclusive constraints not only to object memberships but also to item memberships. Tjhi and Chen [9, 10] proposed heuristic techniques for avoiding *whole-data clusters* by modifying the objective function of FCM-type co-clustering, one of which is easier to implement without additional hyper-parameters [10]. In contrast to the first approach, the dual exclusive constraints play a role for clarifying the cluster assignment of each item and all items have relative memberships among clusters.

In this paper, a heuristic dual exclusive partition model is adopted in the MMMs-induced fuzzy co-clustering context and its characteristics are compared with the item sharing penalty approach. The remaining parts of this paper are as follows: Section 2 reviews the conventional fuzzy co-clustering models and its modification with dual exclusive partition. Section 3 introduces FCCMM with item sharing penalty and proposes a novel approach of extended MMMs-induced fuzzy co-clustering with dual

© Fuji Technology Press Ltd. Creative Commons CC BY-ND: This is an Open Access article distributed under the terms of the Creative Commons Attribution-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nd/4.0/).

exclusive constraints. The characteristics of the proposed method is demonstrated through several numerical experiments in Section 4 and the summary conclusion is presented in Section 5.

2. Fuzzy Co-Clustering and its Variants

2.1. FCM-Type Co-Clustering

Let $R = \{r_{ij}\}$ be an $n \times m$ cooccurrence matrix among n objects and m items, whose element r_{ij} represents the relation between object i and item j. The goal of co-clustering with soft partitions is to estimate the membership degrees of both object u_{ci} and item w_{cj} to cluster c in all C clusters.

Fuzzy clustering for categorical multivariate data (FCCM) [3] is an FCM-type co-clustering model, in which the FCM objective function was modified with a within-cluster-aggregation measure followed by the entropy-based fuzzification penalty [2, 11] as:

$$L_{fccm} = \sum_{c=1}^{C}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ci} w_{cj} r_{ij} - \lambda_u \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}\log u_{ci} - \lambda_w \sum_{c=1}^{C}\sum_{j=1}^{m} w_{cj}\log w_{cj}, \quad \ldots\ldots (1)$$

where u_{ci} and w_{cj} obey the sum-to-one conditions as $\sum_{c=1}^{C} u_{ci} = 1$ and $\sum_{j=1}^{m} w_{cj} = 1$, respectively. Then, u_{ci} mainly represents cluster assignment of object i while w_{cj} implies the typicality of item j in cluster c. The two types of memberships are estimated under the alternating optimization scheme like FCM.

It is often difficult to select plausible values for fuzzification weights λ_u and λ_w because there is no guideline due to the lack of connection with statistical co-clustering models.

2.2. Fuzzy Co-Clustering Induced by Multinomial Mixture Models (FCCMM)

Honda et al. [6] proposed a fuzzy co-clustering method induced by the concept of multinomial mixture models (MMMs) [7]. Since the pseudo-log-likelihood function in MMMs can be interpreted as realization of soft partition by adding K-L information to hard aggregation degree criterion, the following objective function was proposed by considering regularization by K-L information [12, 13]:

$$L_{fccmm} = \sum_{c=1}^{C}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ci} r_{ij}\log w_{cj} + \lambda_u \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}\log\frac{\alpha_c}{u_{ci}}. \quad \ldots\ldots (2)$$

Supported by the MMMs-induced concept, each w_{cj} is forced to be $\sum_{j=1}^{m} w_{cj} = 1$ while u_{ci} are estimated under $\sum_{c=1}^{C} u_{ci} = 1$.

Since α_c represents the capacity of cluster c and u_{ci} becomes close to α_c by the penalty term of K-L information, fuzzification of u_{ci} is realized by taking into account the capacity of each cluster. Fuzzification weight λ_u tunes the degree of fuzziness of memberships u_{ci}, and when $\lambda_u = 1$, Eq. (2) is exactly equivalent to the pseudo-log-likelihood function to be maximized in MMMs. If $\lambda_u > 1$, FCCMM gives a much fuzzier object partition than MMMs while it becomes more crisp with $\lambda_u < 1$.

2.3. Two Approaches for Dual Exclusive Partition in FCM-Type Co-Clustering

In order to improve the interpretability of fuzzy co-cluster structures, each item should belong to a single cluster and two approaches for achieving exclusive partition were proposed in the FCM-type co-clustering (FCCM) context.

In the exclusive partition penalty approach [8], a dual exclusive partition model was achieved by introducing a sharing penalty on items, which is based on the concept used in sequential fuzzy cluster extraction [14]. Item sharing can be avoided by degrading w_{cj} when $\sum_{t \neq c} w_{tj}$ is large. Then, the trade-off between $\sum_{i=1}^{n} u_{ci} r_{ij}$ and $\sum_{t \neq c} w_{tj}$ was implemented in membership updating.

In the second approach, along with an exclusive constraint on objects $\sum_{c=1}^{C} u_{ci} = 1$, a heuristic method of imposing an exclusive constraint, which makes the sum of item memberships among clusters constant as $\sum_{c=1}^{C} w_{cj} = 1$, was proposed [10]. Generally, in FCM-type fuzzy co-clustering [3, 4], direct introduction of the same exclusive constraint for items causes a *whole data cluster*, into which all objects and items are crowded. Therefore, the objective function was slightly modified at each step of membership updating in order to achieve dual exclusive partition of both object and item memberships avoiding whole data clusters. Supported by the exclusive condition $\sum_{c=1}^{C} w_{cj} = 1$ on all items, w_{cj} play a role for clarifying the cluster assignment of each item such that all items belong to at least one cluster.

In the next section, the above two approaches are adopted in MMMs-induced fuzzy co-clustering (FCCMM) [6], which is more useful than FCCM through easy tuning of fuzziness penalty.

3. Dual Exclusive Partition in MMMs-Induced Fuzzy Co-Clustering

3.1. Exclusive Partition Penalty on Item Memberships in FCCMM

Honda et al. [15] introduced an item sharing penalty to FCCMM and showed that the interpretability of co-cluster structures is improved. In order to degrade w_{cj} with large $\sum_{t \neq c} w_{tj}$, a penalty was defined as:

$$s_{cj} = \exp\left(-\beta \sum_{t \neq c} w^*_{tj}\right), \quad \ldots\ldots\ldots\ldots (3)$$

where w^*_{tj} are the values of w_{tj} estimated in the previous iteration and are temporally fixed in updating step of w_{cj}. If item j is not shared and is assigned only in cluster c,

$\sum_{t \neq c} w_{tj}^* \to 0$ and $s_{cj} \to 1$. On the other hand, if item j is severely shared by multiple clusters, the object-item pair has little influences in the next iteration by penalizing with small s_{cj}. Weight β tunes the degree of exclusiveness of item partition such that each item tends to belong to at most one cluster when β is large.

Introducing the penalty s_{cj}, the FCCMM objective function (2) was modified as:

$$L_{fccmm'} = \sum_{c=1}^{C} \sum_{j=1}^{m} \left(\sum_{i=1}^{n} u_{ci} r_{ij} s_{cj} \right) \log w_{cj} + \lambda_u \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log \frac{\alpha_c}{u_{ci}}. \quad (4)$$

The aggregation criterion for w_{cj} calculation is degraded as $\sum_{i=1}^{n} u_{ci} r_{ij} s_{cj}$ by reflecting the sharing degrees of items.

The updating rules for the modified FCCMM are derived as follows:

$$\alpha_c = \frac{1}{n} \sum_{i=1}^{n} u_{ci}, \quad (5)$$

$$u_{ci} = \frac{\alpha_c \prod_{j=1}^{m} (w_{cj})^{\frac{r_{ij} s_{cj}}{\lambda_u}}}{\sum_{l=1}^{C} \alpha_l \prod_{j=1}^{m} (w_{lj})^{\frac{r_{ij} s_{lj}}{\lambda_u}}}, \quad (6)$$

$$w_{cj} = \frac{\sum_{i=1}^{n} r_{ij} s_{cj} u_{ci}}{\sum_{l=1}^{m} \left(\sum_{i=1}^{n} r_{il} s_{cl} u_{ci} \right)}. \quad (7)$$

When $\beta \to 0$, all penalty $s_{cj} \to 1$ and the model is reduced to the conventional FCCMM, i.e., non-exclusive item partition. In order to avoid the illegal effect of the sharing penalty, β should be first initialized as $\beta = 0$ and be gradually increased so that soft transition from non-exclusive to exclusive partition is achieved [8].

3.2. A Novel Heuristic Model with Dual Exclusive Constraint in FCCMM

Next, a novel heuristic model for realizing dual exclusive partition in FCCMM is proposed, which is a hybrid of the heuristic FCCM model [10] and MMMs-induced fuzzy co-clustering [6].

If we impose the FCM-like exclusive constraints on both object and item memberships with the conventional FCCMM objective function (2), all objects and items are crowded into a single cluster. For example, the aggregation term of Eq. (2) can be maximized such that $\sum_{c=1}^{C} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ci} r_{ij} \log w_{cj} \to 0$ if $u_{1i} = 1$, $w_{1j} = 1$ and $u_{ci} = 0$, $w_{cj} = 0$, $\forall c > 1, i, j$. In the conventional FCCMM, the cluster-wise typicality constraint of $\sum_{j=1}^{m} w_{cj} = 1$ is expected to play the role for avoiding the trivial whole-data cluster.

In FCM-type co-clustering, Tjhi and Chen [10] introduced the *intra-cluster typicality* concept in calculation of the *exclusive partition memberships* for both objects and items. In this paper, the same heuristic concept is introduced into the MMMs-induced fuzzy co-clustering.

Assume that we try to maximize Eq. (2) under the dual exclusive constraints of $\sum_{c=1}^{C} u_{ci} = 1$ and $\sum_{c=1}^{C} w_{cj} = 1$. In order to introduce the *intra-cluster typicality* concept in object membership updating, Eq. (2) is remodeled as:

$$L_{fccmm^{o}} = \sum_{c=1}^{C} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ci} r_{ij} \log \left(\frac{w_{cj}}{\sum_{k=1}^{m} w_{ck}} \right) + \lambda_u \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log \frac{\alpha_c}{u_{ci}}, \quad (8)$$

where $w_{cj} / \sum_{k=1}^{m} w_{ck}$ is identified with the intra-cluster typicality like the conventional FCCMM so that object memberships do not form a whole-data cluster. Then, the updating rule is modified as:

$$u_{ci} = \frac{\alpha_c \prod_{j=1}^{m} \left(\frac{w_{cj}}{\sum_{k=1}^{m} w_{ck}} \right)^{\frac{r_{ij}}{\lambda_u}}}{\sum_{l=1}^{C} \alpha_l \prod_{j=1}^{m} \left(\frac{w_{lj}}{\sum_{k=1}^{m} w_{lk}} \right)^{\frac{r_{ij}}{\lambda_u}}}. \quad (9)$$

Next, item memberships are updated by considering the following remodel:

$$L_{fccmm^{i}} = \sum_{c=1}^{C} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{u_{ci}}{\sum_{k=1}^{n} u_{ck}} r_{ij} \log(w_{cj}) + \lambda_u \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log \frac{\alpha_c}{u_{ci}}, \quad (10)$$

where $u_{ci} / \sum_{k=1}^{n} u_{ck}$ work like the intra-cluster typicality so that item memberships do not form a whole-data cluster. Then, the updating rule is modified as:

$$w_{cj} = \frac{\sum_{i=1}^{n} r_{ij} \frac{u_{ci}}{\sum_{k=1}^{n} u_{ck}}}{\sum_{l=1}^{C} \sum_{i=1}^{n} r_{ij} \frac{u_{li}}{\sum_{k=1}^{n} u_{lk}}}. \quad (11)$$

A sample procedure for the proposed iterative algorithm is summarized as follows:

[Heuristic-based FCCMM with Dual Exclusive Constraints]

1. Randomly initialize u_{ci} and w_{cj} such that $\sum_{c=1}^{C} u_{ci} = 1$ and $\sum_{c=1}^{C} w_{cj} = 1$, respectively.
2. Update cluster volumes α_c by Eq. (5).
3. Update object memberships u_{ci} by Eq. (9).
4. Update item memberships w_{cj} by Eq. (11).

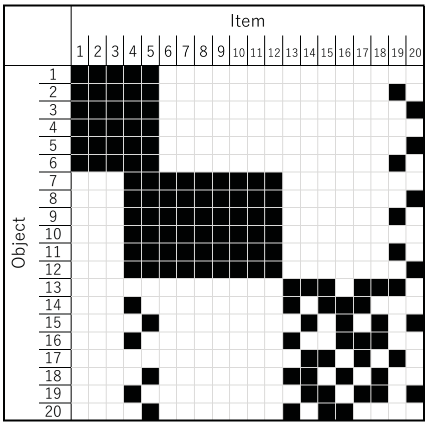
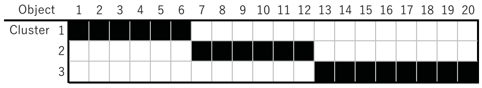
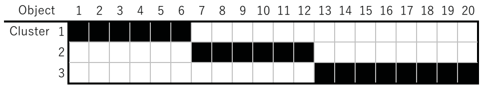


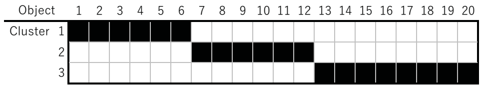
Fig. 1. Artificial data set (Section 4.1).



(a) Conventional FCCMM

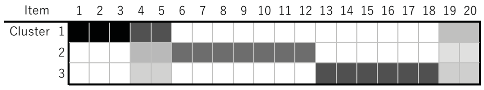


(b) Exclusive partition penalty model

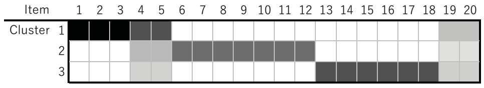


(c) Proposed method

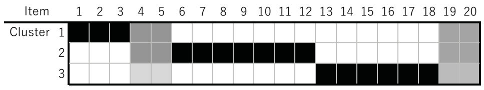
Fig. 2. Comparison of object memberships (Section 4.1).



(a) Conventional FCCMM



(b) Exclusive partition penalty model



(c) Proposed method

Fig. 3. Comparison of item memberships (Section 4.1).

5. If convergent, stop. Otherwise, return to Step 2.

The proposed algorithm is also a heuristic-based one like [10] and the iterative process is not exactly supported by the iterative optimization principle with a single objective function. However, the two sub-objective functions of Eqs. (8) and (10) have similar features and can have similar optimal solutions. Indeed, in the experiments of Section 4, it always converged to a (local) optimal solution like the conventional FCCMM even when items are heavily shared by multiple clusters. So, the algorithm is expected to often work well by jointly utilizing the two sub-objective functions.

4. Numerical Experiments

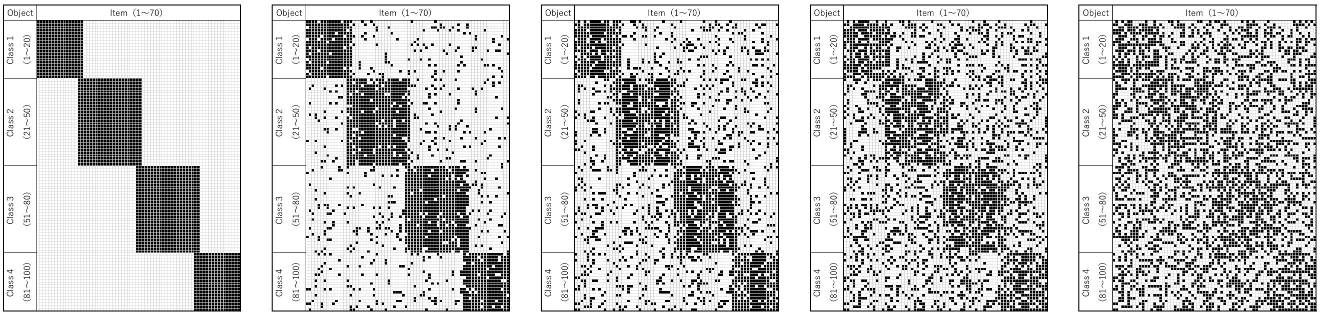
In this section, the characteristic features of the proposed method is demonstrated through comparative experiments with the conventional FCCMM algorithm [6] and the exclusive partition penalty approach [15] using several artificial data sets. Each experiment was performed in 100 trials with different random initializations, and the averages of 100 trials are compared in partition quality evaluations.

4.1. Exclusive Characteristics of Item Memberships

First, the features of fuzzy memberships are studied with an artificial data set shown in **Fig. 1**, which depicts the cooccurrence information matrix $R = \{r_{ij}\}$ among 20 objects ($n = 20$) and 20 items ($m = 20$) such that black and white cells represent $r_{ij} = 1$ and $r_{ij} = 0$, respectively. 20 objects are designed to be drawn from 3 classes ($C = 3$), which are characterized by different typical items. Here, some items are shared by multiple classes, i.e., they are typical in multiple co-clusters. In this experiment, fuzzification weight was set as $\lambda_u = 1.0$ and $\beta_{max} = 4.0$, respectively.

Figures 2 and **3** compare the object and item memberships given by the conventional FCCMM, the exclusive partition penalty model and the proposed method, where grayscale presentation implies the real values of [white, black] $= [0, max]$. **Fig. 2** shows that all three algorithms successfully captured the 3 object clusters.

On the other hand, **Fig. 3** indicates that item memberships have different features, where the main role of item memberships are not item classification but characterization of clusters. Because the conventional FCCMM (**Fig. 3(a)**) do not consider the exclusive feature of item memberships, such shared items as 4 and 5 have same membership degrees with other non-shared items in each cluster. Then, they are not suitable for considering item assignment to co-clusters reflecting item sharing. Next, the exclusive partition penalty model (**Fig. 3(b)**) tried to achieve exclusive assignment of items to co-clusters by rejecting overlapping of items among clusters. Although some item memberships were degraded from the conventional FCCMM, it still cannot be used for considering item partition to clusters, e.g., items 4 and 5 have larger memberships in cluster 1 than cluster 2 while their typicality seems to be almost equivalent in **Fig. 1**. Finally, the proposed method fairly depicted the item assignment to clusters reflecting both cluster-wise typicality and item clusters such that **Fig. 3(c)** indicates that cluster-specific items have almost full memberships ($w_{cj} \simeq 1$) while the memberships of shared items were equally distributed as



(a) Noiseless R_0 (b) 10% noise (c) 20% noise (d) 30% noise (e) 40% noise

Fig. 4. Noisy data set with weak item sharing (Section 4.2).

Table 1. Comparison of partition quality with weak item sharing (Section 4.2).

Noise ratio	10%	20%	30%	40%
FCCMM	94.50	94.54	94.07	42.07
Exclusive penalty	93.42	93.83	82.67	37.28
Proposed method	**98.00**	**99.95**	**96.75**	**42.39**

$w_{1j} \simeq w_{2j}$.

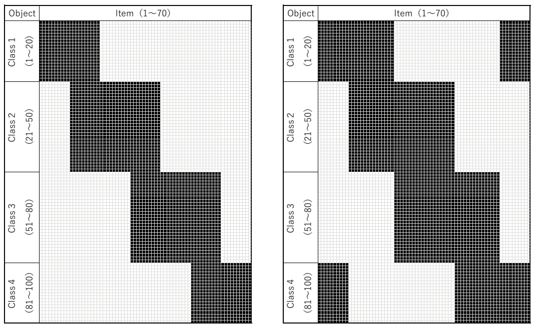
In this way, the proposed method is useful for depicting the dual exclusive partition in dual clustering.

4.2. Noise Robustness

Second, the partition quality of three methods is compared with noisy data sets. The data sets to be analyzed were constructed from the 100×70 matrix R_0 shown in **Fig. 4(a)**, which is composed of 4 co-clusters having fully crisp object partition and weak item sharing. In this experiment, the influences of element-wise noise are considered. For example, in document clustering tasks with document $\times$ keyword frequencies, we often have element-wise noise such that each document can include some occurrences of noise keywords, which are not necessarily relevant to the topic. Noisy data sets R shown in **Figs. 4(b)–(e)** were generated by randomly exchanging r_{ij} values as $0 \leftrightarrow 1$ under noise ratios $\{10\%, 20\%, 30\%, 40\%\}$. The object partition quality is compared with average Rand Index through class-cluster matching after maximum object membership assignment to clusters. In the following experiments, fuzzification weight was set as $\lambda_u = 1.0$ and $\beta_{max} = 1.0$, respectively.

Table 1 compares the partition quality of the three models, where the best performance for each noise ratio is emphasized in bold. In all noise ratios, the proposed method achieved the best partition quality while the exclusive partition penalty model derived the worst result. That is, the two exclusive partition approaches of Sections 3.1 and 3.2 have quite different features and should be used in different purposes.

The similar results are also given with other two situations: moderate item sharing situation (**Fig. 5(a)**) and



(a) Moderate sharing (b) Heavy sharing

Fig. 5. Moderate and heavy item sharing (Section 4.2).

Table 2. Comparison of partition quality with moderate and heavy item sharing (Section 4.2).

(a) Moderate item sharing

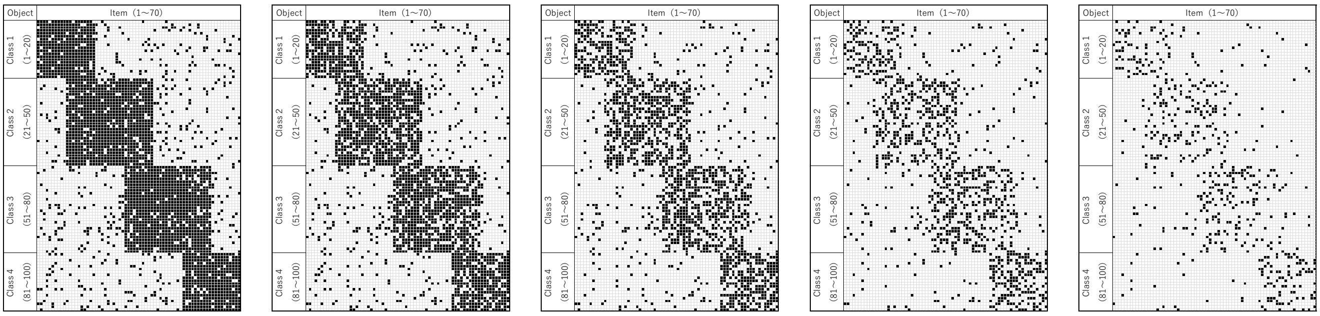
Noise ratio	10%	20%	30%	40%
FCCMM	92.37	93.67	89.16	44.72
Exclusive penalty	93.83	91.22	79.51	38.31
Proposed method	**97.77**	**97.71**	**92.21**	**44.99**

(b) Heavy item sharing

Noise ratio	10%	20%	30%	40%
FCCMM	92.29	95.59	94.47	46.73
Exclusive penalty	91.50	91.00	86.44	38.99
Proposed method	**96.89**	**99.57**	**97.04**	**48.63**

heavy item sharing situation (**Fig. 5(b)**). Comparative results are presented in **Table 2**. The exclusive partition penalty model often derives inferior object partition to the conventional FCCMM, i.e., it is not designed for improving partition quality but for emphasizing the cluster-wise typical items for cluster interpretability. The exclusive penalty can cause a local view in item selection by rejecting many items with very small weight s_{cj}, which can easily bring an inappropriate local solution.

On the other hand, the proposed method can contribute not only to clarifying both object and item cluster assignments but also to improvement of the object partition



(a) No sparsity (b) 20% sparsity (c) 40% sparsity (d) 60% sparsity (e) 80% sparsity

Fig. 6. Sparse data set with moderate item sharing and 10% noise (Section 4.3).

Table 3. Comparison of partition quality with sparse data set with moderate item sharing and 10% noise (Section 4.3).

Sparsity ratio	0%	20%	40%	60%	80%
FCCMM	92.37	93.17	93.87	86.37	**62.26**
Exclusive penalty	93.83	92.69	92.65	83.94	58.77
Proposed method	**97.77**	**98.59**	**97.38**	**87.63**	58.96

quality by rejecting inappropriate item sharing in noisy data sets, i.e., exclusive item partition can contribute to emphasizing the differences among co-clusters, which are useful in correct object assignment. Additionally, the proposed method is considered to outperform the exclusive penalty-based FCCMM because the proposed method has a global view in selection of items without item rejection by weights.

4.3. Influences of Sparseness

Third, the partition quality of three methods is compared with sparse data sets, which were constructed by randomly replacing $r_{ij} = 1 \rightarrow r_{ij} = 0$ in the datasets of Section 4.2 under some sparsity rations. The 10% noisy data set of **Fig. 4(b)** were arranged to sparse ones with sparsity ratios $\{20\%, 40\%, 60\%, 80\%\}$ as shown in **Fig. 6**.

Table 3 compares the partition quality of the three models, where the best performance for each sparsity ratio is emphasized in bold. The proposed method still works better than others when sparsity ratio is 60% or higher but is inferior to the conventional FCCMM when sparsity ratio is 80%.

The influences of sparseness is further studied with other noise ratios: 20% and 30%, whose comparative results are presented in **Table 4**. The tables imply that the proposed method works better than the conventional FCCMM regardless noise ratio when the sparsity is low. However, it can be influenced by sparsity when data sets contaminated by heavy noise while still better than the exclusive partition penalty model from the object partition viewpoint.

Table 4. Comparison of partition quality with more noisy data sets (Section 4.3).

(a) 20% noise

Sparsity ratio	0%	20%	40%	60%	80%
FCCMM	93.67	93.79	85.36	**67.79**	**45.50**
Exclusive penalty	91.22	91.11	80.59	62.54	43.15
Proposed method	**97.71**	**97.14**	**88.19**	64.29	43.47

(b) 30% noise

Sparsity ratio	0%	20%	40%	60%	80%
FCCMM	89.16	75.71	**56.74**	**41.83**	**37.26**
Exclusive penalty	79.51	69.97	52.50	40.92	35.47
Proposed method	**92.21**	**80.59**	53.92	41.52	36.28

4.4. Fuzzification Parameter

Third, the effects of fuzzification parameter λ_u is considered using three types of noisy and sparse data sets: 10% noise (**Fig. 4(b)**), 20% noise (**Fig. 4(c)**) and 20% noise + 20% sparse. The co-clustering algorithms were implemented with three different fuzziness degrees of $\lambda_u \in \{0.5, 1.0, 2.0\}$, i.e., crispy, MMMs, fuzzy. **Table 5** compares the partition qualities of the three models for each data set.

In the comparison of three clustering algorithms, even when the fuzziness degrees are tuned, we can still find the similar features to the previous experiments that the exclusive penalty model is comparative with or inferior to the conventional FCCMM while the proposed model always outperforms others.

In all situations, fuzzier models ($\lambda_u = 2.0$) could achieve the best performances while crisper models ($\lambda_u = 0.5$) brought the worst. Especially, the performances of crisp models are much more inferior to fuzzy models when data sets are more noisy and more sparse. The fuzzier models seem to be more tolerant than MMMs and crisper models when data sets are contaminated by noise and are sparse, and so, fuzziness tuning can contribute to improving the performance of MMMs in the proposed method.

Table 5. Comparison of partition quality with different fuzziness degrees (Section 4.4).

	10% noise			20% noise			20% noise + 20% sparse		
λ_u	0.5	1.0	2.0	0.5	1.0	2.0	0.5	1.0	2.0
FCCMM	90.88	92.37	93.23	91.24	93.67	93.47	88.79	93.79	94.85
Exclusive penalty	91.25	93.83	94.64	86.73	91.22	93.00	84.33	91.11	93.94
Proposed method	96.45	97.77	**99.01**	93.70	97.71	**100.00**	90.21	97.14	**99.97**

5. Conclusions

In this paper, a novel approach for achieving dual exclusive partition in MMMs-induced fuzzy co-clustering was proposed. The proposed method is a hybrid of the heuristic-based FCCM model and MMMs-induced fuzzy co-clustering, and is expected to have advantages against both the conventional models of FCCMM with exclusive partition penalty [15] and heuristic-based FCCM [10], from the viewpoints of easy fuzziness tuning, probabilistic model compatibility and global view item selection.

The characteristic features of the proposed method were demonstrated through several comparative experiments with the conventional FCCMM and the exclusive partition penalty model. The proposed method can work better with non-sparse data sets even when data sets are contaminated by heavy noise, but, for sparse data sets, its performances can be degraded. However, it is always better than the exclusive partition penalty model from the object partition viewpoint.

Additionally, the advantage of tuning the intrinsic fuzziness degree of MMMs was demonstrated such that fuzzier models can contribute to improving partition quality of noisy and/or sparse data. Comparison of fuzzier and crisper results was also available in evaluating the degree of noisiness and/or sparseness because a greater degradation of crisper models was caused by higher degrees of them. There features are expected to be confirmed and utilized in future work with larger scale real applications.

A possible future work is to combine with other mechanisms for improving co-cluster partitions such as noise rejection for robust estimation [16] and deterministic annealing process for initialization robustness [17]. Another direction of future work is to introduce the similar dual partition concept to other statistical co-clustering models such as probabilistic latent semantic analysis (pLSA) [18].

Acknowledgements

This work was supported in part by JSPS KAKENHI Grant Number JP18K11474.

References:

[1] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms," Plenum Press, 1981.

[2] S. Miyamoto, H. Ichihashi, and K. Honda, "Algorithms for Fuzzy Clustering," Springer, 2008.

[3] C.-H. Oh, K. Honda, and H. Ichihashi, "Fuzzy clustering for categorical multivariate data," Proc. of Joint 9th IFSA World Congress and 20th NAFIPS Int. Conf., Vol.4, pp. 2154-2159, 2001.

[4] K. Kummamuru, A. Dhawale, and R. Krishnapuram, "Fuzzy co-clustering of documents and keywords," Proc. of 12th IEEE Int. Conf. on Fuzzy Systems, Vol.2, pp. 772-777, 2003.

[5] T.-C. T. Chen and K. Honda, "Fuzzy Collaborative Forecasting and Clustering," Springer, 2019.

[6] K. Honda, S. Oshio, and A. Notsu, "Fuzzy co-clustering induced by multinomial mixture models," J. Adv. Comput. Intell. Intell. Inform., Vol.19, No.6, pp. 717-726, 2015.

[7] L. Rigouste, O. Cappé, and F. Yvon, "Inference and evaluation of the multinomial mixture model for text clustering," Information Processing & Management, Vol.43, No.5, pp. 1260-1280, 2007.

[8] K. Honda, C.-H. Oh, Y. Matsumoto, A. Notsu, and H. Ichihashi, "Exclusive partition in FCM-type co-clustering and its application to collaborative filtering," Int. J. of Computer Science and Network Security, Vol.12, No.12, pp. 52-58, 2012.

[9] W.-C. Tjhi and L. Chen, "A partitioning based algorithm to fuzzy co-cluster documents and words," Pattern Recognition Letters, Vol.27, No.3, pp. 151-159, 2006.

[10] W.-C. Tjhi and L. Chen, "A heuristic-based fuzzy co-clustering algorithm for categorization of high-dimensional data," Fuzzy Sets and Systems, Vol.159, No.4, pp. 371-389, 2008.

[11] S. Miyamoto and M. Mukaidono, "Fuzzy c-means as a regularization and maximum entropy approach," Proc. of the 7th Int. Fuzzy Systems Association World Congress, Vol.2, pp 86-92, 1997.

[12] H. Ichihashi, K. Miyagishi, and K. Honda, "Fuzzy c-means clustering with regularization by K-L information," Proc. of 10th IEEE Int. Conf. on Fuzzy Systems, Vol.2, pp. 924-927, 2001.

[13] K. Honda and H. Ichihashi, "Regularized linear fuzzy clustering and probabilistic PCA mixture models," IEEE Trans. on Fuzzy Systems, Vol.13, No.4, pp. 508-516, 2005.

[14] K. Tsuda, M. Minoh, and K. Ikeda, "Extracting straight lines by sequential fuzzy clustering," Pattern Recognition Letters, Vol.17, No.6, pp. 643-649, 1996.

[15] K. Honda, T. Nakano, C.-H. Oh, S. Ubukata, and A. Notsu, "Partially exclusive item partition in MMMs-induced fuzzy co-clustering and its effects in collaborative filtering," J. Adv. Comput. Intell. Intell. Inform., Vol.19, No.6, pp. 810-817, 2015.

[16] K. Honda, N. Yamamoto, S. Ubukata, and A. Notsu, "Noise rejection in MMMs-induced fuzzy co-clustering," J. Adv. Comput. Intell. Intell. Inform., Vol.21, No.7, pp. 1144-1151, 2017.

[17] T. Goshima, K. Honda, S. Ubukata, and A. Notsu, "Deterministic annealing process for pLSA-induced fuzzy co-clustering and cluster splitting characteristics," Int. J. of Approximate Reasoning, Vol.95, pp. 185-193, 2018.

[18] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," Machine Learning, Vol.42, No.1-2, pp. 177-196, 2001.

Name:
Katsuhiro Honda

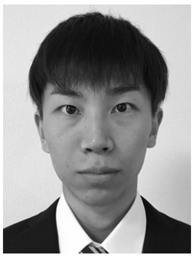
Affiliation:
Graduate School of Engineering, Osaka Prefecture University

Address:
1-1 Gakuen-cho, Nakaku, Sakai, Osaka 599-8531, Japan

Brief Biographical History:
1999- Research Associate, Graduate School of Engineering, Osaka Prefecture University
2007- Assistant Professor, Graduate School of Engineering, Osaka Prefecture University
2009- Associate Professor, Graduate School of Engineering, Osaka Prefecture University
2013- Professor, Graduate School of Engineering, Osaka Prefecture University

Main Works:
- "Linear Fuzzy Clustering With Selection of Variables Using Graded Possibilistic Approach," IEEE Trans. on Fuzzy Systems, Vol.15, No.5, pp. 878-889, 2007.
- "Fuzzy PCA-guided Robust *k*-Means Clustering," IEEE Trans. on Fuzzy Systems, Vol.18, No.1, pp. 67-79, 2010.

Membership in Academic Societies:
- The Institute of Electrical and Electronics Engineers (IEEE)
- Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
- The Institute of Systems, Control and Information Engineers (ISCIE)
- Japan Industrial Management Association (JIMA)



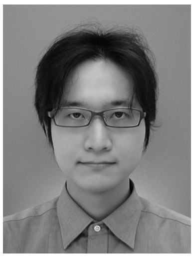
Name:
Yoshiki Hakui

Affiliation:
Graduate School of Engineering, Osaka Prefecture University

Address:
1-1 Gakuen-cho, Nakaku, Sakai, Osaka 599-8531, Japan

Brief Biographical History:
2018 Received the B.Eng. degree in Department of Computer Sciences and Intelligent Systems, Osaka Prefecture University

Main Works:
- "Dual Exclusive Partition Model for MMMs-induced Fuzzy Co-clustering," Proc. of 15th Int. Conf. on Modeling Decisions for Artificial Intelligence, 2018.



Name:
Seiki Ubukata

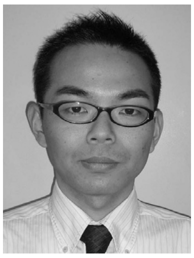
Affiliation:
Graduate School of Engineering, Osaka Prefecture University

Address:
1-1 Gakuen-cho, Nakaku, Sakai, Osaka 599-8531, Japan

Brief Biographical History:
2014- Assistant Professor, Graduate School of Engineering Science, Osaka University
2015- Assistant Professor, Graduate School of Engineering, Osaka Prefecture University

Main Works:
- "Characteristics of Rough Set C-Means Clustering," J. Adv. Comput. Intell. Intell. Inform., Vol.22, No.4, pp. 551-564, 2018.
- "Rough Set-Based Clustering Utilizing Probabilistic Memberships," J. Adv. Comput. Intell. Intell. Inform., Vol.22, No.6, pp. 956-964, 2018.

Membership in Academic Societies:
- The Institute of Electrical and Electronics Engineers (IEEE)
- Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
- The Institute of Systems, Control and Information Engineers (ISCIE)
- Japan Society of Kansei Engineering (JSKE)



Name:
Akira Notsu

Affiliation:
Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University

Address:
1-1 Gakuen-cho, Nakaku, Sakai, Osaka 599-8531, Japan

Brief Biographical History:
2005- Research Associate, Graduate School of Engineering, Osaka Prefecture University
2007- Assistant Professor, Graduate School of Engineering, Osaka Prefecture University
2012-2015 Associate Professor, Graduate School of Engineering, Osaka Prefecture University
2016- Associate Professor, Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University

Main Works:
- "Visualization of Learning Process in "State and Action" Space Using Self-Organizing Maps," J. Adv. Comput. Intell. Intell. Inform., Vol.20, No.6, pp. 983-991, 2016.
- "Visualization of Balancing Systems Based on Naïve Psychological Approaches," AI & Society, Vol.23, No.2, pp. 281-296, 2009.

Membership in Academic Societies:
- Society of Instrument and Control Engineers (SICE)
- Human Interface Society (HIS)