

Partially Exclusive Condition for Sequential Fuzzy Co-cluster Extraction

Katsuhiro Honda, Akira Notsu, Hidetomo Ichihashi
Department of Computer Science and Intelligent Systems
Osaka Prefecture University
1-1 Gakuen-cho, Nakaku, Sakai, Osaka 599-8531, Japan
Email: {honda, notsu, ichi}@cs.osakafu-u.ac.jp

Abstract—Sequential fuzzy co-cluster extraction has been proven to be useful for collaborative filtering tasks by extracting user-item co-clusters, in which promising items are connected to the corresponding users in each co-cluster. Since some popular items can be shared by multiple clusters in collaborative filtering problems, exclusive conditions, which force objects to belong to only one cluster, were used only for users. In this paper, it is demonstrated that such user-only exclusive conditions may cause poor clustering results, and partially exclusive conditions, in which part of items are also forced to be exclusive, are introduced. In addition, connections with neural approaches for co-clustering are discussed, in which co-clustering problems are identified using a modified model of principal component analysis.

Keywords—co-clustering, collaborative filtering, principal component analysis.

I. INTRODUCTION

Collaborative filtering is a promising approach for achieving personalized recommendation in network communities [1]. Many collaborative filtering algorithms are based on the concept of connecting promising items to users by considering mutual similarities among users (and/or items) [2], [3]. Co-clustering is a direct approach for extracting user-item clusters, in which not only similar users but also their associated items are gathered into a cluster. Once we estimate co-cluster structures buried in user-item relational data matrices, items that are supposed to be closely related are recommended to their corresponding users [4], [5].

Honda *et al.* [6], [7] proposed a sequential approach for the fuzzy co-clustering model, which is a combination of the structural balancing model in social science [9] and sequential fuzzy cluster extraction [10]. In collaborative filtering tasks, the co-clustering model plays a role for simultaneously partitioning users and items into several clusters by calculating the principal eigenvector of a modified adjacency matrix. In order to disjoint clusters, penalty weights are added to the diagonal elements of the adjacency matrix. Besides the fully exclusive condition where all users and items can belong to only a single cluster, the user-only exclusive condition can also be applied in such a case where some popular items are shared by multiple clusters. In a collaborative filtering problem, it was demonstrated that

This work was funded in part by the Ministry of Internal Affairs and Communications, Japan, under the Strategic Information and Communications R&D Promotion Programme (SCOPE) and in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, through a Grant-in-Aid for Scientific Research (20700214).

the sequential co-clustering model obtained using the user-only exclusive condition has a high recommendation ability with a low memory requirement. In addition, a modified calculation algorithm for reducing computational costs in cluster structure estimation was proposed, which makes it possible to easily apply the co-cluster structural information to new users [8].

In collaborative filtering tasks, it is also possible that some typical items can be used to distinguish a small community having a special preference. Such typical items should be assigned to only a single cluster, while popular items can be shared by multiple clusters. In this paper, a partially exclusive condition is introduced, in which some items are forced to be exclusive, while other popular items can be shared.

Recently, Oja [11] discussed co-clustering problems (bi-clustering problems) in the context of a modified principal component analysis (PCA) approach. This paper also discusses some connections with such neural approaches.

The remainder of this paper is organized as follows: Section II briefly reviews the sequential approach for fuzzy co-cluster extraction. Section III introduces the partially exclusive condition for items and presents a discussion of the relationship with other co-clustering models. Section IV presents several experimental results in order to demonstrate the characteristic features of the proposed approach. Section V summarizes the conclusions of this paper.

II. BRIEF REVIEW OF SEQUENTIAL APPROACH FOR FUZZY CO-CLUSTER EXTRACTION

A. Structural Balance and Sequential Fuzzy Cluster Extraction

Assume that we have a graph G defined by an adjacency matrix $A(G) = [a_{ij}]$, the elements of which are

$$a_{ij} = \begin{cases} 1 & : \text{if } i \text{ and } j \text{ have positive relation,} \\ -1 & : \text{if } i \text{ and } j \text{ have negative relation,} \\ 0 & : \text{if the relation of } i \text{ and } j \text{ does not exist.} \end{cases} \quad (1)$$

The structure theorem proposed by Cartwright and Harary [9] revealed that the set of vertices (notions) can be partitioned into two subsets (one of which may be empty), the elements of which have positive relations with members of the same subset and negative relations with members of different subsets in a balanced system.

Katai and Iwai [12] discussed how to derive an approximately optimum two-cluster structure and proposed a tech-

nique for calculating a cluster indicator using the eigenvector of $A(G)$ corresponding to the maximum eigenvalue.

The calculation procedure in the eigenvalue problem has some connections with sequential fuzzy clustering models although they use different types of similarity matrices. Fuzzy clustering, most of the algorithms of which are motivated by fuzzy c -Means (FCM) [13], is an unsupervised classification method that can deal with ambiguous boundaries among clusters or overlapping clusters using fuzzy memberships. In addition to FCM-like batch algorithms, sequential clustering models extract clusters having high degrees of aggregation one by one.

Sequential fuzzy cluster extraction (SFCE) [10] estimates cluster memberships by calculating the eigenvector of a modified similarity matrix. Let e_{ij} be the similarity between samples i and j ($e_{ij} = e_{ji}$, $e_{ij} \geq 0$). Here, the goal is to extract a fuzzy cluster having the maximum degree of aggregation. The criterion function to be maximized is formulated as:

$$L_{sfce} = \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n e_{ij} w_{1i} w_{1j}, \quad (2)$$

where $\mathbf{w}_1 = (w_{11}, \dots, w_{1n})^\top$ is an n -dimensional membership indicator vector satisfying the following constraints:

$$0 \leq w_{1i} \leq 1, \quad i = 1, \dots, n, \quad (3)$$

$$\|\mathbf{w}_1\|^2 = \sum_{i=1}^n w_{1i}^2 = 1. \quad (4)$$

The optimal cluster indicator for the first cluster $\mathbf{w}_1$ is derived as the eigenvector corresponding to the largest eigenvalue of the modified similarity matrix $S = \{s_{ij}\}$ [10]:

$$s_{ij} = \begin{cases} e_{ij} & ; i \neq j \\ 0 & ; i = j \end{cases}$$

In order to obtain K disjoint clusters one by one, the optimization problem must be solved repeatedly under the additional constraint that $\mathbf{w}_k^\top \mathbf{w}_{k-1} = \dots = \mathbf{w}_k^\top \mathbf{w}_1 = 0$.

Tsuda et al. [10] introduced a penalty term to the aggregation criterion in order to avoid duplication, as follows:

$$L_{sfce} = \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n e_{ij} w_{ki} w_{kj} - \frac{1}{k-1} \sum_{t=1}^{k-1} \beta_t \text{dup}(\mathbf{w}_k, \mathbf{w}_t), \quad (5)$$

where the penalty term $\text{dup}(\mathbf{w}_k, \mathbf{w}_t)$ is the duplication measure to be minimized, and β_t is the weight for the penalty to be set by users. In [10], β_t was determined by the degree of cluster aggregation, as follows:

$$\beta_t = \eta \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n e_{ij} w_{ti} w_{tj}. \quad (6)$$

The weight η is tuned so that the additional penalty term is sufficiently larger than the aggregation criterion.

The optimal solution for $\mathbf{w}_k$ is then given as the eigenvector corresponding to the largest eigenvalue of the modified similarity matrix $S_k = \{s_{kij}\}$:

$$s_{kij} = \begin{cases} e_{ij} & ; i \neq j \\ -\frac{1}{k-1} \sum_{t=1}^{k-1} \beta_t w_{ti}^2 & ; i = j \end{cases}$$

B. Sequential Fuzzy Co-cluster Extraction

In order to apply the SFCE clustering model to rectangular relational matrices given in collaborative filtering problems, Honda *et al.* [6], [7] combined the concepts of the structural balancing approach. Davis [14] implied that a signed graph having two clusters is further clusterable if each positive line joins two nodes in the same subset but is not clusterable for different subsets. Taking Davis's suggestion into account, Honda et al. introduced a modified relational data matrix, in which only positive relations in the rectangular relational matrix were considered in order to extract subsets.

Assume that there are n users and m items and that we have an $n \times m$ partial adjacency matrix $Z = [z_{ij}]$, where z_{ij} is given by:

$$z_{ij} = \begin{cases} 1 & \text{if } x_{ij} \geq \epsilon \text{ (positive relation),} \\ -1 & \text{if } x_{ij} < \epsilon \text{ (negative relation),} \\ 0 & \text{if } i \text{ did not evaluate } j, \end{cases} \quad (7)$$

where x_{ij} is the evaluation value for item j by user i and ϵ is a pre-defined threshold.

This type of relational matrix is called "a rectangular relational" matrix and most relational clustering models can be applied only after reconstruction of the full-square relational matrix. For example, Bezdek *et al.* [15] performed visual assessment of the cluster tendency by imputing missing relations (user-user relations and item-item relations) based on a concept similar to the Expectation-Maximization (EM) algorithm [16]. The goal of collaborative filtering, however, is the estimation of missing relations, and such preprocessing as imputation is not applicable.

In [6], the influences of missing relations (user-user relations and item-item relations) were ignored based on the concept of the structural balancing approach, and the diagonal block matrices corresponding to user-user relations and item-item relations were given by zero-matrices. In addition, all elements of the relational data matrix must be non-negative elements if we want to apply SFCE to the square relational matrices. Taking Davis's suggestion into account, in [6], only positive relations in the rectangular relational matrix are considered to extract subsets, such that each positive line joins users and items in the same subset, and the partial adjacency matrix $Z = [z_{ij}]$ is generated by ignoring not only negative relations but also unknown relations, as follows:

$$z_{ij} = \begin{cases} 1 & \text{if } x_{ij} \geq \epsilon \text{ (positive relation),} \\ 0 & \text{if } x_{ij} < \epsilon \text{ (negative relation),} \\ 0 & \text{if } i \text{ did not evaluate } j. \end{cases} \quad (8)$$

For example, this type of relational data is common in purchase history data, in which the zero-relation of 'not yet bought'

includes both ‘negative feelings toward an item’ and ‘not yet but might buy’.

Then, the modified similarity matrix $S = \{s_{ij}\}$ is given as the following $(n + m) \times (n + m)$ block matrix:

$$S = \begin{pmatrix} O & Z \\ Z^\top & O \end{pmatrix}. \quad (9)$$

If all users and items are disjoint, then the constraint for indicator vectors $w_1, \dots, w_k$ is given by $w_k^\top w_{k-1} = \dots = w_k^\top w_1 = 0$. On the other hand, if all users are disjoint, but some items can be shared, then the exclusive condition is forced only on users. Let $w_t = (w_t^u, w_t^i)^\top$, where $w_t^u = (w_{t1}, \dots, w_{tn})^\top$ and $w_t^i = (w_{t(n+1)}, \dots, w_{t(n+m)})^\top$. The user-only exclusive constraint is given as $w_k^{u\top} w_{k-1}^u = \dots = w_k^{u\top} w_1^u = 0$, and the diagonal elements of the modified similarity matrix s_{ii} for k -th cluster are re-formulated as

$$s_{ii} = \begin{cases} -\frac{1}{k-1} \sum_{t=1}^{k-1} \beta_t w_{ti}^2 & ; i \leq n \\ 0 & ; i > n \end{cases} \quad (10)$$

In order to reduce the computational cost of eigenvalue problems, an improved procedure was proposed [8], in which two membership vectors of w_1^u for users and w_1^i for items are calculated separately. In the improved procedure, either of w_1^u or w_1^i is first calculated by solving an eigenvalue problem of a smaller relational matrix (the size of the relational matrix is reduced by 75% or more) and the remaining membership indicator is calculated by a simple calculation process. Divided calculation processes are also useful for handling new users or items without re-calculating the eigenvalues.

III. PARTIALLY EXCLUSIVE CONDITION AND CONNECTION WITH OTHER MODELS

A. Partially Exclusive Condition for Sequential Fuzzy Co-cluster Extraction

In collaborative filtering tasks, it is often the case that some popular items may be preferred by many people having various preference characteristics. Therefore, such popular items should be shared by multiple clusters and the user-only exclusive condition appears to be reasonable for finding items to be recommended to users classified into disjoint clusters. It is also possible that some typical items can be used to distinguish a small community having special preferences. Such typical items should be assigned to only a single cluster in order to be assigned to users having such special preferences for the single cluster while popular items can be shared by multiple clusters. In this paper, a partially exclusive condition is introduced, in which certain items are forced to be exclusive while other popular items can be shared.

Without loss of generality, assume that the first p among m items are to be assigned only to a single cluster and the others can be shared by multiple clusters. The exclusive condition, which forces an item to have a large membership value only in a single cluster, is used only for certain items.

$$w_{k(n+j)} \cdot w_{(k-1)(n+j)} = \dots = w_{k(n+j)} \cdot w_{1(n+j)} = 0, \quad j = 1, \dots, p. \quad (11)$$

Here, all users are forced to be exclusive as $w_k^{u\top} w_{k-1}^u = \dots = w_k^{u\top} w_1^u = 0$.

Then, in order to achieve the partially exclusive condition, the diagonal elements of the modified similarity matrix s_{ii} for the k -th cluster are re-defined as

$$s_{ii} = \begin{cases} -\frac{1}{k-1} \sum_{t=1}^{k-1} \beta_t w_{ti}^2 & ; i \leq n + p \\ 0 & ; i > n + p \end{cases} \quad (12)$$

The clustering algorithm is equivalent to the conventional clustering algorithm without the penalty weights.

B. Connection With Other Models

Fuzzy clustering for categorical multivariate (FCCM) data [17] and fuzzy co-clustering of documents and keywords (Fuzzy CoDoK) [18] are FCM-type co-clustering models, in which the relational fuzzy c -Means (RFCM) [19] for square relational matrices is extended for rectangular matrices using two different types of memberships. The first type of membership is the conventional fuzzy membership for objects, which are forced to be exclusive, and the second type of membership is the additional membership for items, which represents the relative typicalities of items in each cluster. In this clustering model, the item memberships are used not for partitioning items, but rather for finding representative items in each cluster. Therefore, certain items can be shared by multiple clusters, and the clustering model has similar features to sequential fuzzy co-cluster extraction under the user-only exclusive condition although FCM-type models are easily influenced by model parameters and missing elements [7]. In addition, FCM-type models cannot consider the partially exclusive condition for items.

Recently, Oja [11] discussed co-clustering problems (bi-clustering problems) in the context of a modified principal component analysis (PCA) approach. When we have an $n \times m$ data matrix X and the m columns are considered to be data vectors, the approximation $X \approx WW^\top X$ performs sparse feature extraction. On the other hand, if n rows are considered to be data vectors, an approximation $X \approx XWW^\top$ performs clustering of the rows. If all of the elements of the data matrix X are non-negative, as in the case of the modified adjacency matrix given by Eq.(9) for sequential fuzzy co-cluster extraction, the matrix factorization process can be reduced to the estimation of an almost orthogonal non-negative matrix W [20]. Then, a biclustering model is constructed based on the approximation $X \approx LL^\top XRR^\top$, in which L and R are the cluster indicators for row and column objects, respectively. The multiplicative rules for neural learning have been reported in [11]. In addition, as suggested in [21], non-negative learning tends to yield high orthogonality. Therefore, the biclustering model essentially forces the exclusive condition to both row and column objects and has features similar to sequential fuzzy co-cluster extraction under the fully exclusive condition. In order to achieve the partially exclusive condition in the neural approach, some modifications are necessary in the approximation part.

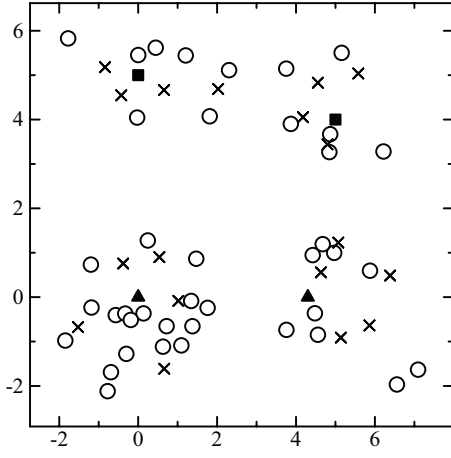


Fig. 1. Artificial data set (o: user, x: item)

IV. NUMERICAL EXPERIMENTS

In this section, experimental results are presented in order to demonstrate the characteristic features of the proposed approach.

A. Artificial Data Set

Figure 1 shows the artificially generated data set used in this experiment, in which $\circ$ and $\times$ denote users and items, respectively. The data set, which was composed of 40 users and 20 items, formed four loose clusters in a 2-D space, and the lower two masses are larger than the upper two masses. Black triangles and squares are shared items that should belong to two clusters, e.g., the item shown in the black triangle is shared by the lower two clusters. Then, a 40×20 relational data matrix $Z = [z_{ij}]$, the elements of which are given by an alternative process, was constructed by calculating mutual Euclidean distances d_{ij} among users and items. Here, d_{ij} for the shared items (black triangle and square) is given by the distance to the nearest mark. The element z_{ij} was given by

$$z_{ij} = \begin{cases} 1 & \text{if } d_{ij}^2 \leq 10, \\ 0 & \text{if } d_{ij}^2 > 10, \end{cases} \quad (13)$$

which is calculated by Eq.(8) with $\epsilon = 10$. The modified similarity matrix $S = \{s_{ij}\}$ was calculated using Eq.(9). The cluster number k and the penalty weight η were set as $k = 3$ and $\eta = 10$, respectively.

First, the data set was partitioned by sequential fuzzy co-cluster extraction under the fully exclusive condition. The maximum membership partition is shown in Fig. 2. The item indicated by the black square had very small membership degrees in all clusters. Because of the strong influence of the most popular item (blue triangle), the lower two masses (larger data masses) were merged into a dominant single cluster. The exclusive condition forcing all items to belong to only one cluster tends to generate a far more dominant cluster and may fail to find useful structural information.

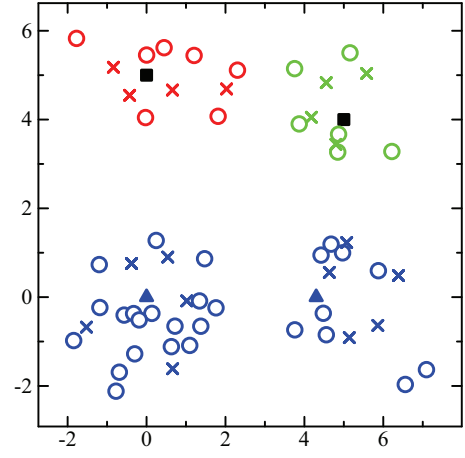


Fig. 2. Cluster partition obtained under fully exclusive condition

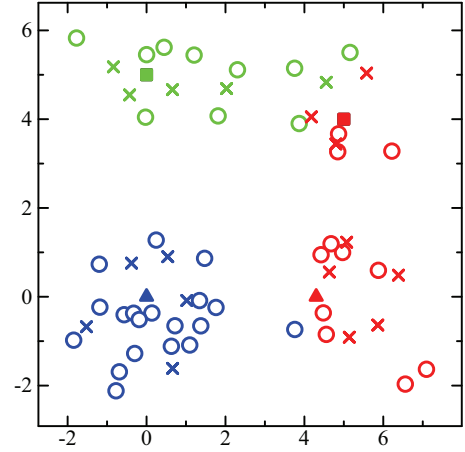


Fig. 3. Cluster partition obtained under user-only exclusive condition

Second, the data set was partitioned under the user-only exclusive condition. Figure 3 shows the derived cluster partition. Since the most popular item (filled triangle) is shared by two clusters (blue and red clusters), the lower two data masses were successfully partitioned. However, the items denoted by the filled squares were also illegally shared by the green and red clusters because these items were weakly shared by many users. Then, the upper right cluster was divided into two clusters. The results imply that the user-only exclusive condition may destroy the cluster structure of small data masses.

Third, the data set was partitioned with the proposed partially exclusive condition and the cluster partition shown in Fig. 4 was derived. Here, only the item indicated by the black square was forced to be exclusive. Using a partially exclusive condition, the intrinsic co-cluster structure was clarified.

In this way, the proposed approach, in which only some of the items are forced to be exclusive, is useful for decreasing the inappropriate influences of popular items and weakly shared items.

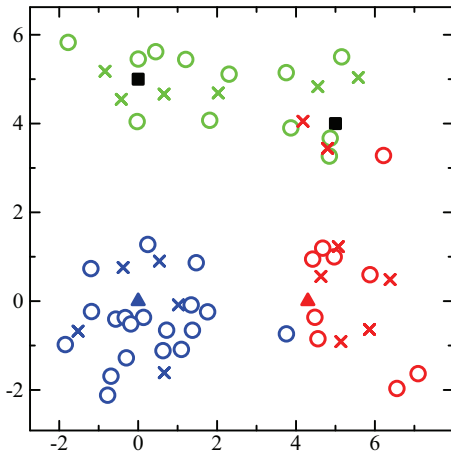


Fig. 4. Cluster partition obtained under partially exclusive condition

B. Purchase History Data

Next, a purchase history data set collected by Nikkei Inc. in 2000 is used in a collaborative filtering task. The data set used in [7] includes the purchase history of 996 users ($n = 996$) on 18 items ($m = 18$). The elements z_{ij} of 996×18 relational data matrix $Z = [z_{ij}]$ is 1 if user i has item j while otherwise 0. In this experiment, Piano (the 2nd item), VD (the 8th item) and Oven (the 14th item), which relatively fewer users have, are selected for items to be predicted, and the goal is to predict the applicability of the items for the second half of 996 users whose corresponding element z_{ij} were all withheld so as to be ' $z_{ij} = 0$ '.

First, the users and items were partitioned into 10 co-clusters under the constraint of $w_k^{u\top} w_{k-1}^u = \dots = w_k^{u\top} w_1^u = 0$, where all items can be shared by multiple clusters. Then, the applicability of the target items for each user was predicted by considering the fuzzy membership degrees of the items to the cluster, in which the user have the maximum membership. If the membership degree of the target item to the cluster is higher than a threshold, the item is recommended to the user.

Second, 10 co-clusters were extracted by forcing the 15th item (Coffee Maker) to be exclusive. The 15th item was selected because its average membership degree was the smallest in the first co-cluster extraction, i.e., it might be only weakly shared by all clusters without belonging to a certain cluster. So, in this second co-cluster extraction, the 15th item was assigned only to a single cluster under exclusive condition without belonging to all clusters.

Table I compares the recommendation abilities of the two co-cluster partitions. The recommendation ability is assessed by ROC sensitivity [22]. 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 are of the curve becomes large as the recommendation ability is higher. The table implies that the proposed model with the partially exclusive condition has higher recommendation ability than the conventional model.

TABLE I
COMPARISON OF RECOMMENDATION ABILITIES: AREA OF LOWER PART OF ROC CURVES

item	conventional	partial exclusive
Piano	0.644	0.651
VD	0.547	0.578
Oven	0.575	0.598

V. CONCLUSION

In this paper, a modified condition for fuzzy co-clustering was considered. It was demonstrated that the partially exclusive condition is useful for clarifying intrinsic co-cluster structures when some items are very popular with or are weakly shared by many users. In a comparative discussion, it was revealed that the sequential fuzzy co-clustering model can derive similar partitions to both FCM-type co-clustering models and neuro-learning-based biclustering models using different exclusive conditions while the partially exclusive condition yields another type of cluster partition.

In a collaborative filtering task, an item that was weakly shared by all clusters was forced to be exclusive so that it assigned to a single cluster without evenly belonging to all clusters. Then, the recommendation ability for items that relatively fewer users have could be improved by considering the partially exclusive condition.

A potential future work is to develop a technique for evaluating the degree of responsibility of items in order to automatically select items to be made exclusive.

REFERENCES

- [1] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gardon and J. Riedl, "GroupLens: applying collaborative filtering to usenet news," *Communications of the ACM*, vol.40, no.3, pp.77-87, 1997.
- [2] J. L. Herlocker, J. A. Konstan, A. Borchers and J. Riedl, "An algorithmic framework for performing collaborative filtering," *Proc. of Conference on Research and Development in Information Retrieval*, 1999.
- [3] G. Linden, B. Smith and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *IEEE Internet Computing*, Jan-Feb, pp.76-80, 2003.
- [4] L. Ungar and D. Foster, "Clustering methods for collaborative filtering," *Proc. of Workshop on Recommendation Systems*, 1998.
- [5] R. Jin, L. Si and C. X. Zhai, "A study of mixture models for collaborative filtering," *Information Retrieval*, vol.9, no.3, pp.357-382, 2006.
- [6] K. Honda, A. Notsu and H. Ichihashi, "Collaborative filtering by sequential extraction of user-item clusters based on structural balancing approach," *Proc. of 2009 IEEE International Conference on Fuzzy Systems*, pp.1540-1545, 2009.
- [7] 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, pp.312-327, 2010.
- [8] K. Honda, A. Notsu and H. Ichihashi, "An efficient algorithm for collaborative filtering based on sequential user-item co-cluster extraction," *Proc. of Joint 5th International Conference on Soft Computing and Intelligent Systems and 11th International Symposium on Advanced Intelligent Systems*, pp.4-8, 2010.
- [9] D. Cartwright and F. Harary, "Structural balance, a generalization of Heider's theory," *Psychological Review*, vol. 63, pp. 167-293, 1956.
- [10] K. Tsuda, M. Minoh and K. Ikeda, "Extracting straight lines by sequential fuzzy clustering," *Pattern Recognition Letters*, vol. 17, pp. 643-649, 1996.

- [11] E. Oja, A. Iljin, J. Luttinen and Z. Yang, "Linear expansions with nonlinear cost functions: modeling, representation, and partitioning," *2010 IEEE World Congress on Computational Intelligence, Plenary and Invited Lectures*, pp.105-123, 2010.
- [12] O. Katai and S. Iwai, "On the characterization of balancing processes of social systems and the derivation of the minimal balancing processes," *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-8, no. 5, pp. 337-348, 1978.
- [13] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981.
- [14] J. A. Davis: Clustering and structural balance in graph; *Human Relation*, Vol. 20, pp. 181-187 (1967)
- [15] J. C. Bezdek, R. J. Hathaway and J. M. Huband, "Visual assessment of clustering tendency for rectangular dissimilarity matrices," *IEEE Trans. Fuzzy Systems*, vol. 15, no. 5, pp. 890-903, 2007.
- [16] A. P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. of the Royal Statistical Society, Series B*, vol. 39, pp. 1-38, 1977.
- [17] C.-H. Oh, K. Honda and H. Ichihashi, "Fuzzy clustering for categorical multivariate data," *Proc. of Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, pp. 2154-2159, 2001.
- [18] K. Kummamuru, A. Dhawale and R. Krishnapuram, "Fuzzy co-clustering of documents and keywords," *Proc. of 12th IEEE international conference on fuzzy systems*, pp. 772-777, 2003.
- [19] R. J. Hathaway, J. W. Davenport and J. C. Bezdek, "Relational duals of the *c*-means clustering algorithms," *Pattern Recognition*, vol. 22, no. 2, pp. 205-212, 1989.
- [20] Z. Yang and E. Oja, "Linear and nonlinear projective nonnegative matrix factorization," *IEEE Transactions on Neural Networks*, vol.21, no.5, pp.734-749, 2010.
- [21] Z. Yang and J. Laaksonen, "Multiplicative updates for nonnegative projections," *Neurocomputing*, vol.71, no.13, pp.363-373, 2007.
- [22] J. A. Swets, "Measuring the accuracy of diagnostic systems," *Science*, vol.240, no.4857, pp.1285-1289, 1988.