Fuzzy Co-Clustering and Application to Collaborative Filtering

Katsuhiro Honda(✉)

Osaka Prefecture University, Osaka, Sakai 599-8531, Japan
honda@cs.osakafu-u.ac.jp

Abstract. Cooccurrence information analysis became more popular in many web-based system analyses such as document analysis or purchase history analysis. Rather than the conventional multivariate observations, each object is characterized by its cooccurrence degrees with various items, and the goal is often to extract co-cluster structures among objects and items, such that mutually familiar object-item pairs form a co-cluster. A typical application of co-cluster structure analysis can be seen in collaborative filtering (CF). CF is a basic technique for achieving personalized recommendation in various web services by considering the similarity of preferences among users. This paper introduces a fuzzy co-clustering model, which is motivated from a statistical co-clustering model, and demonstrates its applicability to CF tasks following a brief review of the CF framework.

Keywords: Co-clustering · Fuzzy clustering · Collaborative filtering

1 Introduction

Cooccurrence information analysis becomes much more popular especially in recent web-based systems such as document-keyword frequencies in web document analysis and customer-product purchase transactions in web market analysis. Cooccurrence information among objects and items are often summarized into some co-cluster structures, where mutually familiar object-item pairs are grouped into co-clusters. Then, the goal of co-clustering is to extract pairwise clusters of dual partitions among both objects and items, rather than the conventional multivariate observations on objects, where cluster structures are extracted by focusing on mutual similarities among objects only.

In simultaneously extracting dual partitions of objects and items, we have roughly two different partition concepts. First, if we assume that each co-cluster should be formed by familiar objects and items without considering the priority between objects and items in analysis, both objects and items are handled under common constrains. Non-negative Matrix Factorization (NMF) [1] and its extension to co-clustering [2] is typical examples of this approach, where each of both objects and items is often forced to be assigned to a solo co-clusters such that both of them form non-overlapping clusters. This category also includes such

V.-N. Huynh et al. (Eds.): IUKM 2016, LNAI 9978, pp. 16–23, 2016.
DOI: 10.1007/978-3-319-49046-5_2

other model as sequential co-cluster extraction [3] and coVAT [4], where objects and items are mixed and an object $\times$ item cooccurrence matrix is enlarged into an (object + item) $\times$ (object + item) matrix.

Second, if our purpose is mainly to find object clusters in conjunction with associating typical items to each cluster, we can utilize the conventional clustering concepts under the cluster-wise Bag-of-Words concept. Typical examples of this category are multinomial mixture models (MMMs) [5] or its extensions [6], which utilize multinomial distribution or Dirichlet distribution of items as component densities and try to find object clusters under the mixture density concept. In these mixture models, each component density is estimated in each cluster under the constraint of sum-to-one probabilities of items, and objects are exclusively assigned to nearest clusters under the constrains of sum-to-one probability of objects. The traditional k-means clustering model [7] and fuzzy c-means (FCM) [8] have also been extended to co-clustering models. Fuzzy clustering for categorical multivariate data (FCCM) [9] and fuzzy clustering for document and keywords (Fuzzy CoDoK) [10] modified the within-cluster-error measure of FCM by introducing two types of fuzzy memberships for objects and items, and adopted an aggregation measure to be maximized in each cluster under the similar partition constrains to MMMs. Considering the conceptual similarity among MMMs and FCCM, Honda *et al.* proposed fuzzy co-clustering induced by MMMs concept (FCCMM) [11], which introduced a mechanism for tuning intrinsic fuzziness of the MMMs pseudo-log-likelihood function and demonstrated the advantage of tuning intrinsic fuzziness in MMMs.

Some hybrid models of the above two concepts have also been proposed by extending FCM-based frameworks. By introducing the exclusive nature of item memberships to FCCMM, we can produce dual exclusive partition [12,13]. It was also tried to apply the FCM-like exclusive condition to both objects and items [14,15].

In this paper, the characteristics of fuzzy co-clustering models are discussed considering to the applicability to collaborative filtering (CF) [16,17], which is a promising application of co-clustering models. CF is a recommendation system for reducing information overloads in recent network societies, where promising items are recommended to users by analyzing user $\times$ item evaluation matrices. A traditional approach of neighborhood-based model is often implemented through a two-stage procedure: user neighborhood searching and evaluation averaging in neighborhood. Then, the process is expected to be imitated through user clustering in conjunction with estimating item typicalities in each cluster. In this sense, the second co-clustering concept of 'object clustering with item typicality estimation' seems to be plausible in this task.

The remaining parts of this paper are organized as follows: Sect. 2 gives a brief survey of co-clustering based on soft partition, which includes comparison of two partition concepts. Section 3 introduces the basic concept of CF and discusses the applicability of fuzzy co-clustering to the task. A summary conclusion is presented in Sect. 4.

2 Co-clustering Based on Soft Partition

Assume that we have an $n \times m$ rectangular relational information matrix $R = \{r_{ij}\}$ among n objects and m items, where r_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$ is the degree of mutual connectivity among object i and item j such that a large value means a high connectivity. A typical example is the cooccurrence frequency among documents and keywords in document analysis, e.g., a large r_{ij} means a high frequency of appearance of keyword j in document i, where documents and keywords correspond to objects and items, respectively. This kind of datasets is also popular in network societies such as market analysis based on purchase transaction history.

A promising approach for summarizing the intrinsic information of such rectangular relational data is to extract pairwise clusters composed of mutually familiar object-item pairs. Such clustering tasks as extracting C disjoint pairwise clusters is called co-clustering (or bi-clustering). In this section, a brief survey on some co-clustering based on soft partition concepts is presented.

2.1 Co-clustering Models Supported by Dual Exclusive Partition Nature

When we have multivariate numerical observations on n objects such as an $n \times m$ numerical matrix X, a popular approach for low-dimensional summarization is principal component analysis (PCA), where the mathematical goal is to estimate an singular value decomposition $X \approx FA^\top$ such that the reconstruction error is minimized. The low-dimensional factors F can be often utilized for visually summarizing mutual relation among objects on the feature space spanned by basis A.

NMF [1] extended the matrix decomposition concept to non-negative cooccurrence information such that R is decomposed into $R \approx UW^\top$, where U and W are non-negative feature matrices of objects and items, respectively. In order to represent the cluster belongingness of objects and items, all elements of U and W are forced not to be negative, and the columns of U and W are mutually almost orthogonal such that each object and item has high feature values in at most one cluster. Oja *et al.* [2] has further extended NMF for emphasizing the co-cluster information of component matrices U and W. Because these models equally force the columns of U and W to have zero inner products, both objects and items are exclusively assigned to clusters without considering the priority under partition purposes.

Besides the singular value decomposition approach, objects and items can be combined into one set in such a way that an enlarged square relational information of $(n + m) \times (n + m)$ matrix.

$$S = \begin{pmatrix} G & R \\ R^\top & H \end{pmatrix} \tag{1}$$

Honda *et al.* [3] adopted the sequential fuzzy cluster extraction (SFCE) [18], which was designed for square relational data, to a modified relational data matrix. In this model, missing parts of the full-square matrix were given by $G = H = O$. Bezdek *et al.* [4] adopted visual assessment technique (VAT) to rectangular relational data, which was originally designed for square relational data, estimating G and H by calculating mutual similarity among rows or columns of R.

These co-clustering models are generally useful for summarizing the intrinsic connectivity features. However, if we have a primal target in clustering tasks, these models may not suit the goal of the task.

2.2 Co-clustering Models Considering Primal Target in Objects vs. Items

In such tasks as document clustering, the goal is often mainly to find document clusters in conjunction with selecting their cluster-wise typical keywords. That is, keywords are expected to contribute to characterizing each document cluster and are not intended to be exclusively assigned to clusters. Indeed, some common keywords may frequently occurred in various documents and should have larger typicalities in multiple clusters. So, the typicality of each keyword should be independently evaluated in each cluster.

MMMs [5] is a basic statistical model for co-cluster analysis, in which each component density is defined by multinomial distribution. Multinomial distribution is a multi-category extension of binomial distribution, where the probability of occurrence of each item is estimated following the relative frequencies of item appearances. Mixtures of C component multinomial distributions are estimated through the expectation-maximization (EM) algorithm [19], where the expectation step estimates the probability u_{ci} of each object i drawn from component density c and the maximization step estimates multinomial component densities of probability w_{cj} of item j in component c reflecting object assignments. Although the both two steps are based on a maximum likelihood estimation scheme with a common likelihood function, the partition concepts for objects and items are not common supported by different constraints. The item probability is estimated under the intra-cluster constraint of $\sum_{j=1}^{m} w_{cj} = 1$ so that it represents the relative typicality of item j in component c. On the other hand, the object probability is estimated under the inter-cluster constraint of $\sum_{c=1}^{C} u_{ci} = 1$ so that it represents the exclusive assignment of object i to C clusters. In this sense, Objects are regarded as the primal target in the clustering task and item probability plays a role just for characterizing each object cluster.

Similar co-clustering models have been also proposed supported by k-means-like clustering concepts. In order to estimate the compactness of clusters by considering the intra-cluster typicalities of items instead of centroids of k-means families, FCCM [9] and Fuzzy CoDoK [10] defined the objective function to be maximized with the following aggregation measure:

$$cluster\ aggregation = \sum_{c=1}^{C} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ci} w_{cj} r_{ij}, \tag{2}$$

under the same constraints with MMMs. Co-clusters are extracted in such a way that familiar object i and item j having large cooccurrence r_{ij} tend to have large memberships u_{ci} and w_{cj} in a same cluster c. Besides the crisp k-means-like linear objective function of Eq. (2), FCCM and Fuzzy CoDoK achieved fuzzy partition by introducing such fuzzification schemes as the entropy-based regularization [20, 21] and the quadratic term-based regularization [21, 22]. Although the regularization-based fuzzy partition models have some advantages against the traditional statistical models such as MMMs in arbitrarily tuning the fuzziness degrees of object and item partitions, it is often difficult to find the plausible fuzziness degrees for dual partitions without guidelines or comparative statistical models.

Considering the conceptual similarity among MMMs and FCCM, Honda *et al.* proposed FCCMM [11], which introduced a mechanism for tuning intrinsic fuzziness of the MMMs pseudo-log-likelihood function. The objective function to be maximized was defined as follows:

$$L_{fccmm_1} = \sum_{c=1}^{C}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ci} r_{ij} \log w_{cj} + \lambda_u \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci} \log \frac{\alpha_c}{u_{ci}}, \tag{3}$$

where α_c represents the volume of cluster c such that $\sum_{c=1}^{C} \alpha_c = 1$. Eq. (3) is equivalent to the pseudo-log-likelihood function for MMMs when $\lambda_u = 1$, and the intrinsic fuzziness degree of object partition can be arbitrarily tuned by adjusting the penalty weight λ_u. A larger λ_u ($\lambda_u > 1$) brings a fuzzier partition rather than MMMs while a smaller λ_u ($\lambda_u < 1$) causes a crisper one. Indeed, $\lambda_u \to 0$ implies a crisp partition.

Additionally, the fuzziness degree of item partition can be also tuned by adjusting the non-linearity with respect to w_{cj} such as:

$$L_{fccmm2} = \sum_{c=1}^{C}\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{1}{\lambda_w} u_{ci} r_{ij} \left((w_{cj})^{\lambda_w} - 1 \right) + \lambda_u \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci} \log \frac{\alpha_c}{u_{ci}}, \tag{4}$$

where the additional weight λ_w is responsible for tuning the item partition fuzziness. $\lambda_w \to 0$ implies the same fuzziness degree with MMMs and a smaller λ_w brings a fuzzier partition.

Because FCCMM was designed supported by the MMMs concept, the plausible fuzziness degrees of dual partitions are expected to be achieved under the guideline of MMMs. That is, the plausible partition can be often searched for by slightly tuning MMMs partitions.

2.3 Hybrid Approaches of Object Targeting Partition and Exclusive Item Partition

Even if the primal target is object partitioning like FCCM and FCCMM, exclusive partition of items is also expected to be useful in order to improve the interpretability of dual partitions. Then, some researches tried to introduce exclusive natures into the object targeting co-clustering models.

In order for each item can have large membership w_{cj} at most in one cluster, FCCM and FCCMM were modified by introducing an additional sharing penalty on items such as $\sum_{t \neq c} w_{cj} w_{tj}$ into the conventional objective function [12,13]. This model was demonstrated to be useful for improving the interpretability by selecting cluster-wise unique items in each object cluster. It is also possible to force the exclusive nature only to some selected items not to be shared by multiple clusters. By carefully selecting the specific items, partition quality can be improved [23].

Another challenge is to adopt the FCM-like exclusive condition to both object and item partitions such that $\sum_{c=1}^{C} u_{ci} = 1$ and $\sum_{c=1}^{C} w_{cj} = 1$. Generally, the objective functions of the object targeting co-clustering models have a trivial solution, where all objects and items are gathered into one cluster. Then, Tjhi and Chen [14,15] introduced a heuristic-based algorithm and demonstrated some advantages in revealing plausible item features.

3 Applicability of Co-clustering to Collaborative Filtering Tasks

This section considers the applicability of co-clustering to CF tasks. CF is a promising technique for reducing information overloads through personalized recommendation based on collaborative data analysis among users [16].

Assume that $R = \{r_{ij}\}$ is an $n \times m$ evaluation matrix, where r_{ij} is an implicit or explicit rating for item j given by user i such that a larger r_{ij} implies a better availability of item j for user i. Each element can be not only an explicit evaluation such as an active rating given by a user but also an implicit evaluation such as purchase transaction of an item by a user. The goal of personalized recommendation in CF can be identified with the task of predicting the availability of missing items [17], e.g., items the active user have not purchased or evaluated. Then, the recommendation system recommends the items, whose availability are expected to be maximum. Here, it should be noted that a missing element in an implicit rating matrix may be not only an unknown situation but also a negative feeling. For example, in purchase history, missing elements can be either of 'negative evaluation' and 'positive but not yet bought'.

The representative memory-based algorithm is GroupLens [16], which virtually implements *word-of-mouth* in network societies and is composed of main two phases: neighborhood users search and preferences averaging. Using the mutual similarity among users, the applicability of item j for an active user a is predicted as follows:

$$y_{aj} = \frac{\sum_{i=1}^{n} (r_{ij} - \bar{r}_i) \times s_{ai}}{\sum_{i=1}^{n} s_{ai}} + \bar{r}_a, \tag{5}$$

where $\bar{r}_i$ is the mean of evaluations by user i and s_{ai} is the mutual similarity measure among user i and the active user a such as Pearson correlation coefficient of two users' evaluations. In GroupLens, the deviations from mean evaluation of each user are considered for eliminating the influences of users' evaluation tendencies.

Because the process of co-clustering can be regarded as summarization of preference tendencies on items in each homogeneous user (object) clusters, a potential application of co-clustering is found in CF tasks [3,12,24]. Here, from the viewpoint of realization of *word-of-mouth*, co-clustering-based CF should consider users (objects) as the clustering target, where item preferences are independently summarized in each cluster. In this sense, MMMs-induced co-clustering models are expected to be available in various CF applications.

4 Conclusions

This paper presented a brief survey on co-clustering models, in which many algorithms were summarized into three categories considering their constraints: Dual Exclusive Partition Models, Object Targeting Partition Models and their Hybrid Models. Besides their algorithmic frameworks, they have different applicability to many real world tasks under different constraints on dual partitions.

In this paper, document-keyword analysis and CF are introduced as typical applications of Object Targeting Partition Models, where the goal is mainly to extract object clusters in conjunction with characterization emphasizing their representative items. However, we have also other application areas, where objects and items should be equally handled. For example, in a factory management problem, where many products are manufactured in several machining centers utilizing a wide range of machine tools, we should assign products to machining centers such that each machining center manufactures their products using only similar (common) tools. In this case, product-tool relational data should be summarized into product-tool co-clusters without targeting neither of products and tools.

Rectangular relational data becomes much more popular in various network societies but they should be processed under plausible constraints reflecting the characteristics of tasks. It is expected that many co-clustering researches can contribute to tackling various real applications.

Acknowledgment. This work was supported by JSPS KAKENHI Grant Number JP26330281.

References

1. Lee, D.-D., Seung, H.-S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**(6755), 788–791 (1999)
2. Oja, E., Ilin, A., Luttinen, J., Yang, Z.: Linear expansions with nonlinear cost functions: modeling, representation, and partitioning. In: 2010 IEEE World Congress on Computational Intelligence, Plenary and Invited Lectures, pp. 105–123 (2010)
3. Honda, K., Notsu, A., Ichihashi, H.: Collaborative filtering by sequential user-item co-cluster extraction from rectangular relational data. Int. J. Knowl. Eng. Soft Data Paradigms **2**(4), 312–327 (2010)
4. Bezdek, J.C., Hathaway, R.J., Huband, J.M.: Visual assessment of clustering tendency for rectangular dissimilarity matrices. IEEE Trans. Fuzzy Syst. **15**(5), 890–903 (2007)

5. Rigouste, L., Cappé, O., Yvon, F.: Inference and evaluation of the multinomial mixture model for text clustering. Inf. Process. Manag. **43**(5), 1260–1280 (2007)
6. Holmes, I., Harris, K., Quince, C.: Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS ONE **7**(2), e30126 (2012)
7. MacQueen, J.B.: Some methods of classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297 (1967)
8. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
9. Oh, C.-H., Honda, K., Ichihashi, H.: Fuzzy clustering for categorical multivariate data. In: Proceedings of Joint 9th IFSA World Congress and 20th NAFIPS International Conference, pp. 2154–2159 (2001)
10. Kummamuru, K., Dhawale, A., Krishnapuram, R.: Fuzzy co-clustering of documents and keywords. In: Proceedings of 2003 IEEE International Conference Fuzzy Systems, vol. 2, pp. 772–777 (2003)
11. Honda, K., Oshio, S., Notsu, A.: Fuzzy co-clustering induced by multinomial mixture models. J. Adv. Comput. Intell. Intell. Inf. **19**(6), 717–726 (2015)
12. Honda, K., Oh, C.-H., Matsumoto, Y., Notsu, A., Ichihashi, H.: Exclusive partition in FCM-type co-clustering and its application to collaborative filtering. Int. J. Comput. Sci. Netw. Secur. **12**(12), 52–58 (2012)
13. Honda, K., Nakano, T., Oh, C.-H., Ubukata, S., Notsu, A.: Partially exclusive item partition in MMMs-induced fuzzy co-clustering and its effects in collaborative filtering. J. Adv. Comput. Intell. Intell. Inf. **19**(6), 810–817 (2015)
14. Tjhi, W.-C., Chen, L.: A partitioning based algorithm to fuzzy co-cluster documents and words. Pattern Recogn. Lett. **27**, 151–159 (2006)
15. Tjhi, W.-C., Chen, L.: A heuristic-based fuzzy co-clustering algorithm for categorization of high-dimensional data. Fuzzy Sets Syst. **159**, 371–389 (2008)
16. Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gardon, L.R., Riedl, J.: Grouplens: applying collaborative filtering to usenet news. Commun. ACM **40**(3), 77–87 (1997)
17. Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J.: An algorithmic framework for performing collaborative filtering. In: Proceedings of 22nd Conference on Research and Development in Information Retrieval, pp. 230–237 (1999)
18. Tsuda, K., Minoh, M., Ikeda, K.: Extracting straight lines by sequential fuzzy clustering. Pattern Recogn. Lett. **17**, 643–649 (1996)
19. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B **39**, 1–38 (1977)
20. Miyamoto, S., Mukaidono, M.: Fuzzy c-means as a regularization and maximum entropy approach. In: Proceedings of the 7th International Fuzzy Systems Association World Congress, vol. 2, pp. 86–92 (1997)
21. Miyamoto, S., Ichihashi, H., Honda, K.: Algorithms for Fuzzy Clustering. Springer, Heidelberg (2008)
22. Miyamoto, S., Umayahara, K.: Fuzzy clustering by quadratic regularization. In: Proceedings 1998 IEEE International Conference Fuzzy Systems and IEEE World Congress Computational Intelligence, vol. 2, pp. 394–1399 (1998)
23. Nakano, T., Honda, K., Ubukata, S., Notsu, A.: MMMs-induced fuzzy co-clustering with exclusive partition penalty on selected items. Integrated Uncertainty in Knowledge Modelling and Decision Making **9376**, 226–235 (2015)
24. Honda, K., Muranishi, M., Notsu, A., Ichihashi, H.: FCM-type cluster validation in fuzzy co-clustering and collaborative filtering applicability. Int. J. Comput. Sci. Netw. Secur. **13**(1), 24–29 (2013)

Springer
http://www.springer.com/978-3-319-49045-8

Integrated Uncertainty in Knowledge Modelling and Decision Making
5th International Symposium, IUKM 2016, Da Nang, Vietnam, November 30- December 2, 2016, Proceedings
Huynh, V.-N.; Inuiguchi, M.; Le, B.; Le, N.B.; Denœux, T. (Eds.)
2016, XXIV, 728 p. 191 illus., Softcover
ISBN: 978-3-319-49045-8