Fuzzy clusterwise generalized structured component analysis with non-negative constraint

Yoji Yamashita*, Kensuke Tanioka † and Hiroshi Yadohisa ‡

Abstract — **Hwang et al., (2006) proposed a Fuzzy clusterwise generalized structured component analysis (FCGSCA), which is a simultaneous analysis of generalized structured component analysis (GSCA) and fuzzy clustering (Bezdek, 1984). The method detects path structures between variables corresponding to clusters when the data consists of several path structures. The advantage of the FCGSCA is that researchers can reflect their hypotheses or knowledge on the analysis results. However, if the number of variables or clusters is larger, interpreting the results becomes difficult because the path weights have a sign. To overcome this problem, we propose an FCGSCA with non-negative constraint and its majorizing algorithm.**

Keyword: *Majorizing algorithm; entropy regularization; joint analysis, Non-negative matrix factorization*

1 Introduction

Researchers may find it is useful to analyze paths to validate their hypothesis based on given data. Therefore, Hwang and Takane (2004) proposed a generalized structured component analysis (GSCA), which has two advantages: First, the objective function is formulated according to the unified framework, and parameter estimation is easily conducted using the alternative least squares (ALS) framework. However, GSCA assumes that data has a homogeneity group. Therefore, if data includes heterogeneity groups, the estimation weights corresponding to paths are not meaningful.

Thus, researchers conduct a two-step approach by using existing methods: clustering and GSCA. In the first step, clustering, such as k-means, is applied to data. Next, GSCA is applied to each subgroup corresponding to the estimated cluster to obtain the results. However, such results cannot be evaluated because there is no objective function.

Hwang et al., (2006) proposed Fuzzy clusterwise GSCA (FCGSCA), which is a simultaneous analysis of GSCA and fuzzy clustering (Bezdek, 1984). The method can achieve its purpose when the data include heterogeneity groups and are useful. However, the application of the method has two problems in some situations. First, it is difficult to understand the results when the number of clusters or variables are large because researchers attempt to interpret the meaning of weights for considering their sign. Second, it is difficult to tune fuzzy parameters because it is difficult to imagine the meaning of fuzzy based on Bezdek et al.(1984).

To overcome this problem, we propose a new method for non-negative multivariate data: FCGSCA with non-negative constraint (FCGSCANC), which has three advantages. First, it is easy to interpret the estimated results because all the estimated weights are non-negative. Therefore, when we attempt to analyze the meaning of the estimated weights, we do not need to consider its sign. The good property is known to domain of non-negative matrix factorization (e.g., Lee and Seung, 1999). Second, we derived the majorization algorithm based on Jansen's inequation (e.g., Lee and Seung, 1999) and the inequation by Groenen et al. (2006), subject to non-negativity constraints. Third, we adopted the fuzzy clustering based on entropy regularization (Miyamoto and Mukaidono, 1999) and not Bezdek because understanding the tuning parameters of fuzziness is easy.

*Faculty of Nursing and Nutrition, Kagoshima Immaculate Heart University, 895-0011, Japan, E-mail: amashita@jundai.k-junshin.ac.jp, Tel: +81-996-23-5311

†Clinical Study Support Center, Wakayama Medical University, 641-8509, Japan

‡Department of Culture and Information Science, Doshisha University, 610-0394, Japan

2 FCGSCANC

2.1 Objective function of FCGSCANC

In this subsection, we provide the objective function of FCGSCANC. Eq. (2.1) is defined as the objective function of FCGSCANC and its parameters are estimated through minimization of Eq. (2.1).

$$L(\{\boldsymbol{A}_\ell\},\{\boldsymbol{B}_\ell\},\{\boldsymbol{C}_\ell\},\boldsymbol{U}|\ \boldsymbol{X},P_A,P_B,P_C,\ \delta)=\sum_{i=1}^{n}\sum_{\ell=1}^{k}u_{i\ell}\|\boldsymbol{x}_{(i)}^T\boldsymbol{V}_\ell-\boldsymbol{x}_{(i)}^T\boldsymbol{A}_\ell[\boldsymbol{B}_\ell,\boldsymbol{C}_\ell]\|^2+\delta\sum_{i=1}^{n}\sum_{\ell=1}^{k}u_{i\ell}\log u_{i\ell} \tag{2.1}$$

Let $\boldsymbol{X}=(\boldsymbol{x}_{(1)},\boldsymbol{x}_{(2)},\cdots,\boldsymbol{x}_{(n)})^T$, then $\boldsymbol{x}_{(i)}=(x_{ij})\in\mathbb{R}^+$ $(i=1,2,\cdots,n;\ j=1,2,\cdots,p)$ is the $n\times p$ multivariate data, where n is the number of objects and p is the number of variables, Further, let P_A and P_B be the path structures of cluster ℓ $(\ell=1,2,\cdots,k)$, where k is the number of clusters, from observed to latent variables, and from latent to observed variables, respectively. $\delta>0$ is the fuzziness tuning parameter, $\boldsymbol{A}_\ell=(a_{jo\ell})$ $(j=1,2,\cdots,p;\ o=1,2,\cdots,d;\ \ell=1,2,\cdots,k)$ be weights matrices of cluster ℓ from the observed to latent variables, where d is the number of components, such that $a_{jo\ell}\geq 0$ if $(j,o,\ell)\in P_A$, else $a_{jo\ell}=0$. Let $\boldsymbol{B}_\ell=(b_{os\ell})$ $(o=1,2,\cdots,d;\ s=1,2,\cdots,p;\ \ell=1,2,\cdots,k)$ be the weight matrices of cluster ℓ from latent to observed variables, such that $b_{os\ell}\geq 0$ if $(o,s,\ell)\in P_B$, else $b_{os\ell}=0$. Let $\boldsymbol{C}_\ell=(c_{oq\ell})$ $(o=1,2,\cdots,d;\ q=1,2,\cdots,d;\ \ell=1,2,\cdots,k)$ be weight matrices of cluster ℓ from latent to latent variables, such that $c_{oq\ell}\geq 0$ if $(o,q,\ell)\in P_C$, else $c_{oq\ell}=0$, and $\boldsymbol{U}=(u_{i\ell})$ $u_{i\ell}\in[0,1]$ $(i=1,2,\cdots,n;\ \ell=1,2,\cdots,k)$ is the membership matrix. Here, $\boldsymbol{V}$ is defined as $\boldsymbol{V}=(\boldsymbol{I}_p,f(\boldsymbol{A}))$, where $\boldsymbol{I}_p$ is the $p\times p$ identity matrix and $f(\boldsymbol{A})$ is the subset of the column space of $\boldsymbol{A}$.

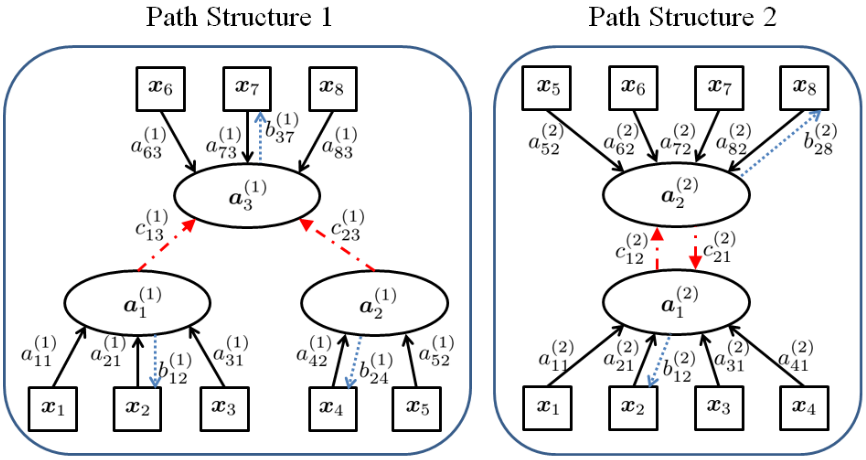

Figure 1: Example of path structures

Figure 1 illustrates a path structure, in which the models corresponding to clusters are as follows:

$$\boldsymbol{X}[\boldsymbol{I}_8,\boldsymbol{a}_3^{(1)}]=\boldsymbol{X}\boldsymbol{A}_1[\boldsymbol{B}_1,\boldsymbol{C}_1]+\boldsymbol{E}_1\quad\text{and}\quad\boldsymbol{X}[\boldsymbol{I}_8,\boldsymbol{a}_1^{(2)},\boldsymbol{a}_2^{(2)}]=\boldsymbol{X}\boldsymbol{A}_2[\boldsymbol{B}_2,\boldsymbol{C}_2]+\boldsymbol{E}_2$$

where

$$\boldsymbol{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_8),\ \boldsymbol{x}_j\in\mathbb{R}^n,$$

$$\boldsymbol{A}_1=\begin{bmatrix} a_{11}^{(1)} & a_{21}^{(1)} & a_{31}^{(1)} & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & a_{42}^{(1)} & a_{52}^{(1)} & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & a_{63}^{(1)} & a_{73}^{(1)} & a_{83}^{(1)} \end{bmatrix}^T,\quad \boldsymbol{A}_2=\begin{bmatrix} a_{11}^{(2)} & a_{21}^{(2)} & a_{31}^{(2)} & a_{41}^{(2)} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & a_{52}^{(2)} & a_{62}^{(2)} & a_{72}^{(2)} & a_{82}^{(2)} \end{bmatrix}^T,$$

$$[\boldsymbol{B}_1, \boldsymbol{C}_1] = \left[\left[\begin{array}{cccccccc} 0 & b_{12}^{(1)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_{24}^{(1)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & b_{37}^{(1)} & 0 \end{array}\right], \left[\begin{array}{c} c_{13}^{(1)} \\ c_{23}^{(2)} \\ 0 \end{array}\right]\right] \text{ and}$$

$$[\boldsymbol{B}_2, \boldsymbol{C}_2] = \left[\left[\begin{array}{cccccccc} 0 & b_{12}^{(2)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{28}^{(2)} \end{array}\right], \left[\begin{array}{cc} 0 & c_{12}^{(2)} \\ c_{21}^{(2)} & 0 \end{array}\right]\right].$$

Then, objects are assigned to path structure 1 or 2 by this method subject to non-negativity constraints.

2.2 Algorithm of FCGSCANC

This subsection shows the FCGSCANC algorithm. The parameters of FCGSCANC are estimated through ALS.

Algorithm of FCGSCANC

Step 0: Set initial values of $\boldsymbol{B}_\ell,\ \boldsymbol{C}_\ell,\ (\ell = 1,2,\cdots,k)$, supplemental variables $\lambda_{joi\ell}$ $(j = 1,2,\cdots,p; o = 1,2,\cdots,d; i = 1,2,\cdots,n; \ell = 1,2,\cdots,k)$, $\eta_{jois\ell}$ $(j = 1,2,\cdots,p; o = 1,2,\cdots,d; i = 1,2,\cdots,n; s = 1,2,\cdots,p+m;\ \ell = 1,2,\cdots,k)$ and $\boldsymbol{U}$.

Step 1: Update $\boldsymbol{A}_\ell$, given $\boldsymbol{B}_\ell,\ \boldsymbol{C}_\ell,\ \lambda_{joi\ell},\ \eta_{jois\ell}$ and $\boldsymbol{U}$

Step 2: Update $\boldsymbol{B}$ and $\boldsymbol{C}_\ell$, given $\boldsymbol{A}_\ell$, suppremental variables $\lambda_{joi\ell},\ \eta_{jois\ell}$, and $\boldsymbol{U}$

Step 3: Update $\lambda_{joi\ell}$ and $\eta_{jois\ell}$, given $\boldsymbol{A}_\ell, \boldsymbol{B}_\ell,\ \boldsymbol{C}_\ell$ and $\boldsymbol{U}$

Step 4: Update $\boldsymbol{U}$, given $\boldsymbol{A}_\ell,\ \boldsymbol{B}_\ell,\ \boldsymbol{C}_\ell,\ \lambda_{joi\ell}$ and $\eta_{jois\ell}$,

Step 5: If stop rule is satisfied, this algorithm is stopped, else back to Step 1.

Next, updated formula of these parameters are shown.

Updated formula of $\boldsymbol{A}_\ell$ $(\ell = 1,2,\cdots,k)$

Given $\boldsymbol{B}_\ell,\ \boldsymbol{C}_\ell$ $(\ell = 1,2,\cdots,k)$, and $\boldsymbol{U}$, the updated formula of $\boldsymbol{A}_\ell$ $(\ell = 1,2,\cdots,k)$ is derived according to the majorizing function based on both Jansen's and Groenen's inequations (2006), as follows:

$$a_{jo\ell} = \frac{\sum_{i=1}^{n} u_{i\ell}\left[x_{ij}(\sum_{j^\dagger=1}^{p} x_{ij^\dagger} w_{jq\ell} + \sum_{t=1}^{d}\sum_{j^*=1}^{p} x_{ij^*} c_{j^*o\ell} z_{to\ell}) + \sum_{s=1}^{p+m} x_{ij} z_{os\ell}(\sum_{j^\dagger=1}^{p} x_{ij^\dagger} w_{js\ell} + \sum_{t=1}^{d}\sum_{j^*=1}^{p} x_{ij^*} w_{j^*t\ell} b_{ts\ell})\right]}{\sum_{i=1}^{n} u_{i\ell}\left(\lambda_{joi\ell}^{-1} x_{ij}^2 + \sum_{s=1}^{p+m} \eta_{jois\ell}^{-1} x_{ij}^2 z_{os\ell}^2\right)} \tag{2.2}$$

$$a_{jo\ell} = \frac{\sum_{i=1}^{n} u_{i\ell} \sum_{s=1}^{p+m} \eta_{jois\ell}^{-1} x_{ij}^2 b_{ts\ell}^2}{\sum_{i=1}^{n} u_{i\ell} \sum_{j^\dagger=1}^{n} x_{ij^\dagger} v_{j^\dagger t\ell} \sum_{s=1}^{p+m} x_{ij} b_{ts\ell}} \tag{2.3}$$

where $(j,o,\ell) \in P_A$ and $\lambda_{joi\ell}$ and $\eta_{jois\ell}$, $\boldsymbol{Z}_\ell = [\boldsymbol{B}_\ell,\ \boldsymbol{C}_\ell] = (z_{os\ell})$ $(o = 1,2,\cdots,d; s = 1,2,\cdots,p+m; \ell = 1,2,\cdots,k)$ are supplemental variables and $\boldsymbol{W}_\ell = (w_{jo\ell})$ is pre step values of $\boldsymbol{A}_\ell$. Here, m is the number of subspaces of a column space of $\boldsymbol{A}$. If $(j,o,\ell) \in P_A$ and $\boldsymbol{a}_{o\ell}$ belongs to $f(\boldsymbol{A}_\ell)$, Eq. (2.2) is adopted, else Eq. (2.3) is adopted.

Updated formula of $\boldsymbol{B}_\ell,\ \boldsymbol{C}_\ell$ ($\ell = 1, 2, \cdots, k$)

Given $\boldsymbol{A}_\ell$ ($\ell = 1, 2, \cdots, k$), and $\boldsymbol{U}$, the updated formulae of $\boldsymbol{B}_\ell\ \boldsymbol{C}_\ell$ ($\ell = 1, 2, \cdots, k$) is derived according to the majorizing function based on Jansen's inequation as follows

$$z_{os\ell} = \frac{(\sum_{j=1}^{p+m} x_{ij} v_{js})(\sum_{j=1}^{p} x_{ij} a_{jo\ell})}{\sum_{j=1}^{p} \eta_{jois\ell}^{-1} x_{ij}^2 a_{jo\ell}^2} \quad ((o, s, \ell) \in P_B \cup P_C) \tag{2.4}$$

where $\boldsymbol{Z}_\ell = [\boldsymbol{B}_\ell,\ \boldsymbol{C}_\ell]$.

Updated formula of suppremental variables

Given $\boldsymbol{A}_\ell,\ \boldsymbol{B}_\ell\ \boldsymbol{C}_\ell$ ($\ell = 1, 2, \cdots, k$), and $\boldsymbol{U}$, the updated formulae of supplemental variables are derived as follows:

$$\lambda_{joi\ell} = \frac{u_{i\ell}^{1/2} x_{ij} a_{jo\ell}}{\sum_{j^*=1}^{p} u_{i\ell}^{1/2} x_{ij^*} a_{j^*t^*}} \ (j = 1, 2, \cdots, p.; o = 1, 2, \cdots, d; i = 1, 2, \cdots, n; \ell = 1, 2, \cdots, k) \tag{2.5}$$

$$\eta_{jois\ell} = \frac{u_{i\ell}^{1/2} x_{ij} a_{jo\ell} z_{os\ell}}{\sum_{j^*=1}^{p} \sum_{o^*=1}^{d} u_{i\ell}^{1/2} x_{ij^*} a_{j^*o^*} z_{o^*s\ell}} (j = 1, 2, \cdots, p\ ; o = 1, 2, \cdots, d;\ i = 1, 2, \cdots, n; s = 1, 2, \cdots, p{+}m;\ \ell = 1, 2, \cdots, k) \tag{2.6}$$

Updated formula of $\boldsymbol{U}$

Given $\boldsymbol{A}_\ell,\ \boldsymbol{B}_\ell\ \boldsymbol{C}_\ell$ ($\ell = 1, 2, \cdots, k$), and $\boldsymbol{U}$, the updated formula of membership $\boldsymbol{U}$ is derived as follows:

$$u_{i\ell} = \frac{\exp\left\{ -\delta^{-1} \|\boldsymbol{x}_{(i)}^T \boldsymbol{V}_\ell - \boldsymbol{x}_{(i)}^T \boldsymbol{A}_\ell \boldsymbol{B}_\ell\|^2 \right\}}{\sum_{\ell^\dagger=1}^{k} \exp\left\{ -\delta^{-1} \|\boldsymbol{x}_{(i)}^T \boldsymbol{V}_{\ell^\dagger} - \boldsymbol{x}_{(i)}^T \boldsymbol{A}_{\ell^\dagger} \boldsymbol{B}_{\ell^\dagger}\|^2 \right\}} \ (i = 1, 2, \cdots, n; \ell = 1, 2, \cdots, k) \tag{2.7}$$

3 Concluding and Remarks

The numerical simulations and application are shown on the conference.

References

[1] Bezdek, J.C., Ehrlich, R., and Full, W. (1984). *FCM. Fuzzy c-means algorithm*. Computers and Geoscience, 10(2-3), p191–203.

[2] Groenen, P.J.F., Winsberg, S., Rodriguez, P. and Diday, E. (2006) *I-Scal: Multidimensional scaling of interval dissimilarities*. Computational Statistics & Data Analysis, 51(1), p360–378.

[3] Hwang, H., Desarbo, W.S., and Takane, Y. (2006) *Fuzzy clusterwise generalized structured component analysis*. Psychometrika, 72, p181–198.

[4] Hwang, H. and Takane, Y. (2004) *Generalized structured component analysis* . Psychometrika, 69, p81–99.

[5] Lee, P.D.D., and Seung, H.S. (1999) *Learning the parts of objects with nonnegative matrix factorization*.Nature, 401, p788–791.

[6] Miyamoto, S., and Mukaidono, M. (1997) *Fuzzy c-means as regularization and miximum entropy approach*. In:Proc of the 7th Fuzzy System Association World Confgress, Prague Czech Republic, 2, p86-92. , p788–791.