Neural Processing Letters manuscript No.
(will be inserted by the editor)

Hierarchical tensor SOM network for multilevel-multigroup analysis

Hideaki Ishibashi · Tetsuo Furukawa

Received: date / Accepted: date

Abstract The aim of this work is to develop a visualization method for multilevel-multigroup analysis based on a multiway nonlinear dimensionality reduction. The task of the method is to visualize what kinds of members each group is composed and to visualize the similarity between the groups in terms of probability distribution of constituent members. To achieve the task, the proposed method consists of hierarchically coupled tensor self-organizing maps (TSOMs), corresponding to the member/group level. This architecture enables more flexible analysis than ordinary parametric multilevel analysis, as it retains a high level of interpretability supported by strong visualization. We applied the proposed method to one benchmark dataset and two practical datasets: one is the survey data on the football players belonging to different teams and the other is the employee survey data belonging to different departments in a company. Our method successfully visualizes the types of the members that constitute each group as well as visualizes the differences or similarities between the groups.

Keywords multilevel analysis · multigroup analysis · tensor decomposition · self-organizing map · SOM

This work is supported by Okawa Foundation for Information and Telecommunications.

Hideaki Ishibashi
Department of Human Intelligence Systems, Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu-ku, Kitakyushu 808-0196, Japan
Tel.: +81-93-695-6003
Fax: +81-93-695-6008
E-mail: ishibashi-hideaki@edu.brain.kyutech.ac.jp

Tetsuo Furukawa
Department of Human Intelligence Systems, Kyushu Institute of Technology, Japan
2-4 Hibikino, Wakamatsu-ku, Kitakyushu 808-0196, Japan
Tel.: +81-93-695-6003
Fax: +81-93-695-6008
E-mail: furukawa@brain.kyutech.ac.jp

1 Introduction

To analyze survey data obtained from multiple groups, it is not easy to grasp the whole image of the data, because we need to analyze it at both individual level and group level, and to find out the relation between them. The aim of this study is to develop a visualization method for multigroup dataset based on the self-organizing map (SOM), which enables us to understand the whole image of the dataset intuitively.

A typical example is a psychological investigation of sports players belonging to different teams. To discover knowledge from the dataset, we need to know the psychological state of each player by finding out the factors representing the states. We also need to know how each player is similar or different to others by comparing their psychological states, and to compare the teams in terms of their constituent players. Another example is a morale survey of departments in a company, in which the task is to analyze the departments in terms of the morale states of constituent employees. We often encounter such kind of datasets at schools, at local governments, and so on.

The simplest method for group analysis is to use the mean value of the constituent members. In this case, the individual data are averaged over the members for each group, which is regarded as the feature vector of the group [1]. This approach is convenient if the mean value represents the members' property well. However, if the data distribution is not unimodal, this approach produces false results. For example, if the members of a group form some distinct clusters, then the average only represents the intermediate point between clusters, and there may not be any members around the mean point. This approach can be extended to other parametric models such as linear regression or principal component analysis (PCA), by which the obtained model parameters are regarded as the feature vector of each group [1]. Even in these cases, the obtained parameters also need to represent the member distribution well.

The alternative is to use a nonparametric representation such as a histogram. Like the cases of Bag-of-Features (BoF), the obtained histograms are regarded as the feature vectors of the groups. This provides a more flexible representation than the parametric approach, but is only available for low dimensional cases. To overcome this limitation, a combination of nonparametric representation and dimensionality reduction is required. Thus, high dimensional data are mapped to a low dimensional latent space in advance, and the data distributions in the latent space are estimated by a nonparametric model [2]. This approach seems to be promising, but it sacrifices the interpretability of the results, especially in nonlinear cases. Thus, one can see the member distribution in the latent space, but it is not easy to grasp the member property from the low dimensional expression.

In this paper, we introduce a multiway nonlinear dimensionality reduction based on the tensor SOM (TSOM) [3]. TSOM is a nonlinear tensor decomposition method to analyze relational dataset, and it visualizes a relational dataset by producing two (or more) maps. Though designed to analyze relational datasets, the TSOM can also visualize ordinary high dimensional datasets by regarding them as relational ones. In this case, the TSOM produces a map of the target objects and a map of the data components, with the latter visualizing the intrinsic factors underlying the data. Thus, the TSOM enhances the interpretability of high dimensional datasets. By introducing the TSOM to multilevel-multigroup analysis, we achieve both a tractable representation of the member distributions and a interpretable result.

Furthermore, the TSOM can be employed to visualize the groups, as the member–group affiliations are represented by relational data. As a result, the entire architecture becomes a hierarchical network of two TSOMs, one for the member level analysis and the other for the group level analysis. Using the TSOM network, multilevel-multigroup analysis can be executed in a consistent and seamless manner.

The remainder of this paper is organized as follows. Sec. 2 formulates the task, before the algorithm is described in Sec. 3. In Sec. 4, some visualization methods are introduced. Related works are summarized in Sec. 5. In Sec. 6 reports the experiment results of this method. The last section is the conclusion of this study. Some preliminary results were reported as a conference proceeding [4].

2 Problem formulation

As a typical example, let us consider a case where a questionnaire survey is conducted on I members with J rating queries. Thus the dataset becomes a matrix $\mathbf{X} = (x_{ij}) \in \mathbb{R}^{I\times J}$. Suppose that there are K groups, and each member belongs to at least one of the groups. (Duplication is allowed.) This member–group affiliation is represented by a matrix $\mathbf{Y} = (y_{ik}) \in \{0, 1\}^{I\times K}$, where $y_{ik} = 1$ if the ith member belongs to the kth group, and otherwise $y_{ik} = 0$.

The aim of the proposed method is to visualize the relationships between the members, the groups, and the queries by mapping them to low-dimensional latent spaces i.e., *member-map*, *query-map*, and *group-map*, respectively. Thus, our task is to estimate the three latent variable sets $\left\{\mathbf{z}_i^{(\text{member})}\right\}$, $\left\{\mathbf{z}_j^{(\text{query})}\right\}$, $\left\{\mathbf{z}_k^{(\text{group})}\right\}$. This task consists of two subtasks, corresponding to the member-level analysis and the group-level analysis.

In the member-level analysis, the subtask is to estimate the latent variables $\left\{\mathbf{z}_i^{(\text{member})}\right\}$, $\left\{\mathbf{z}_j^{(\text{query})}\right\}$, and a smooth nonlinear map f, so that the observed data is approximated as

$$x_{ij} \simeq f(\mathbf{z}_i^{(\text{member})}, \mathbf{z}_j^{(\text{query})}). \tag{1}$$

After these latent variables are estimated, the member distribution of each group is characterized by a probability density in the latent space. Here let $p_k(\mathbf{z}^{(\text{member})})$ denote the member distribution of the kth group.

In the group-level analysis, the subtask is to estimate the latent variable $\left\{\mathbf{z}_k^{(\text{group})}\right\}$ and a nonlinear smooth map q, so that the member distribution of the kth group is represented as

$$p_k(\mathbf{z}^{(\text{member})}) \simeq q(\mathbf{z}^{(\text{member})} \mid \mathbf{z}_k^{(\text{group})}). \tag{2}$$

3 Architecture and Algorithm

3.1 Architecture

In an ordinary SOM, the latent space (map space) $\mathcal{Z}$ is discretized to grid nodes, and a reference vector is assigned to every node. If the latent space is discretized to $\mathcal{Z} = \{\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_L\}$, then the SOM has L reference vectors $\{\mathbf{f}_1, \ldots, \mathbf{f}_L\}$. In the case of the TSOM of order 2, the TSOM has 2 latent spaces $\mathcal{Z}^{(1)}$, $\mathcal{Z}^{(2)}$ which are discretized to L and M nodes respectively, as in the ordinary SOM. However, unlike the conventional SOM, the reference vectors of the TSOM are assigned to all combinations of 2 nodes. Thus, there are $L \times M$ reference vectors $\{\mathbf{f}_{11}, \ldots, \mathbf{f}_{LM}\}$ [3].

The proposed architecture for multilevel-multigroup analysis consists of two TSOMs of order 2, one for the member-level and the other for the group-level (Fig. 1). Suppose that three latent spaces $\mathcal{Z}^{(\text{member})}, \mathcal{Z}^{(\text{query})}, \mathcal{Z}^{(\text{group})}$ are discretized to L, M, N nodes respectively, and $\{\boldsymbol{\zeta}_l^{(\text{member})}\}$, $\{\boldsymbol{\zeta}_m^{(\text{query})}\}$, $\{\boldsymbol{\zeta}_n^{(\text{group})}\}$ are the positional vectors of the nodes. In the

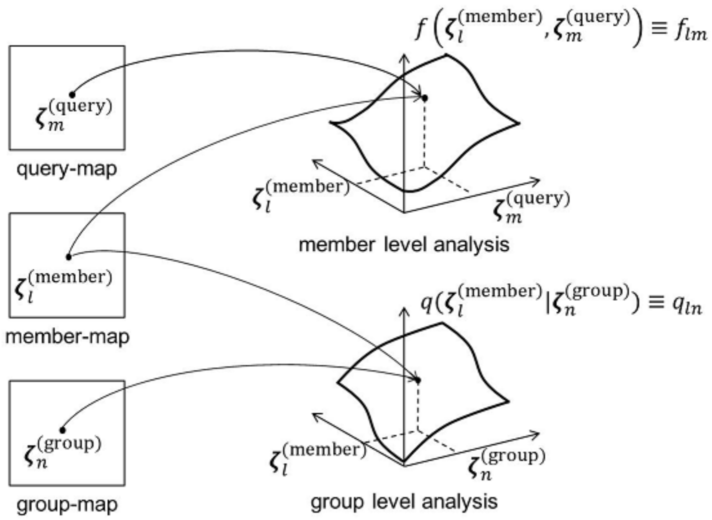


Fig. 1 The architecture of the proposed method. Two TSOMs are coupled by sharing the member-map.

TSOM for the member-level analysis, reference vectors are assigned to all combinations of $\left\{\left(\zeta_l^{(\text{member})}, \zeta_m^{(\text{query})}\right)\right\}$ so that $f_{lm} \equiv f(\zeta_l^{(\text{member})}, \zeta_m^{(\text{query})})$ [1]. Then, the entire set of reference vectors becomes a matrix $\mathbf{F} \equiv (f_{lm}) \in \mathbb{R}^{L\times M}$. A row vector $\mathbf{f}_l^{(\text{member})} \equiv (f_{l1}, \ldots, f_{lM})$ and a column vector $\mathbf{f}_m^{(\text{query})} \equiv (f_{1m}, \ldots, f_{Lm})$ act as conventional reference vectors for members and for queries, respectively.

Similarly, reference vectors in the group-level TSOM are assigned to all combinations of $\left\{\left(\zeta_l^{(\text{member})}, \zeta_n^{(\text{group})}\right)\right\}$ so that $q_{ln} \equiv q(\zeta_l^{(\text{member})} \,|\, \zeta_n^{(\text{group})})$. Thus, the entire set becomes a matrix $\mathbf{Q} = (q_{ln}) \in \mathbb{R}^{L\times N}$. In this case, a column vector $\mathbf{q}_n \equiv (q_{1n}, \ldots, q_{Ln})$ acts like a reference vector for groups.

Besides these reference vectors, the algorithm includes the vectors $\mathbf{g}_i^{(\text{member})} \in \mathbb{R}^L$, $\mathbf{g}_j^{(\text{query})} \in \mathbb{R}^M$, and $\mathbf{p}_k \in \mathbb{R}^N$. These play the role of data vectors for members, queries, and groups, and are calculated in the algorithm.

3.2 TSOM for member-level analysis

The proposed algorithm consists of two stages: the member-level TSOM learns first, followed by the group-level TSOM. The learning algorithm for the TSOM is the expectation-maximization (EM) algorithm in a broad sense, with the E and M steps iterating alternately.

The learning algorithm of the member-level TSOM can be described as follows.

Initialization First, all latent variables $\left\{\mathbf{z}_i^{(\text{member})}\right\}, \left\{\mathbf{z}_j^{(\text{query})}\right\}$ are randomly initialized, and the calculation time is set to $t = 0$. After that the EM algorithm starts from M step.

[1] Actually f_{lm} becomes a scalar here, but we call it a reference vector for consistent explanation.

M Step **F** is calculated using Nadaraya-Watson kernel smoother:

$$f_{lm} = \sum_{i=1}^{I} \sum_{j=1}^{J} R_{li}^{(\text{member})} R_{mj}^{(\text{query})} x_{ij}, \tag{3}$$

where $R_{li}^{(\text{member})}$, $R_{mj}^{(\text{query})}$ are given by

$$R_{li}^{(\text{member})} = \frac{h(\boldsymbol{\zeta}_l^{(\text{member})}, \mathbf{z}_i^{(\text{member})})}{\sum_{i'} h(\boldsymbol{\zeta}_l^{(\text{member})}, \mathbf{z}_{i'}^{(\text{member})})} \tag{4}$$

$$R_{mj}^{(\text{query})} = \frac{h(\boldsymbol{\zeta}_m^{(\text{query})}, \mathbf{z}_j^{(\text{query})})}{\sum_{j'} h(\boldsymbol{\zeta}_m^{(\text{query})}, \mathbf{z}_{j'}^{(\text{query})})}. \tag{5}$$

Here $h(\boldsymbol{\zeta}, \mathbf{z}) = \exp\left(\|\boldsymbol{\zeta} - \mathbf{z}\|^2 / 2\sigma^2(t)\right)$ is the smoothing kernel, also referred as neighborhood function in SOM literatures. $\sigma(t)$ determines the kernel size, which is reduced gradually with the calculation time t.

After calculating **F**, $\{\mathbf{g}_i^{(\text{member})}\}$ and $\{\mathbf{g}_j^{(\text{query})}\}$ are also calculated by

$$g_{im}^{(\text{member})} = \sum_{j=1}^{J} R_{mj}^{(\text{query})} x_{ij} \tag{6}$$

$$g_{jl}^{(\text{query})} = \sum_{i=1}^{I} R_{li}^{(\text{member})} x_{ij}. \tag{7}$$

E Step The best matching nodes of members and queries are determined. For the ith member and the jth query, they are determined as follows:

$$l_i^\star = \arg\min_l \|\mathbf{g}_i^{(\text{member})} - \mathbf{f}_l^{(\text{member})}\|^2 \tag{8}$$

$$m_j^\star = \arg\min_m \|\mathbf{g}_j^{(\text{query})} - \mathbf{f}_m^{(\text{query})}\|^2. \tag{9}$$

Then the latent variables are estimated as $\mathbf{z}_i^{(\text{member})} = \boldsymbol{\zeta}_{l_i^\star}^{(\text{member})}$, $\mathbf{z}_j^{(\text{query})} = \boldsymbol{\zeta}_{m_j^\star}^{(\text{query})}$.

These M and E steps are iterated until the TSOM reaches the steady state with a reduced neighborhood size.

3.3 TSOM for group-level analysis

Once the first stage is finished, the group-level TSOM is trained.

Initialization Before starting the learning process, the member distributions of groups $\{\mathbf{p}_k\}$ are calculated by kernel density estimation as:

$$p_{lk} = \frac{1}{N_k} \sum_{i=1}^{I} \tilde{R}_{li}^{(\mathrm{member})} y_{ik}, \tag{10}$$

where

$$\tilde{R}_{li}^{(\mathrm{member})} = \frac{h(\zeta_l^{(\mathrm{member})}, \mathbf{z}_i^{(\mathrm{member})})}{\sum_{l'} h(\mathbf{z}_{l'}^{(\mathrm{member})}, \mathbf{z}_i^{(\mathrm{member})})}, \tag{11}$$

and N_k is the number of the members belonging to the kth group. Note that $\mathbf{p}_k$ represents the probability density $p_k(\mathbf{z}^{(\mathrm{member})})$, i.e., the member distribution of the kth group. After that the latent variables $\{\mathbf{z}_k^{(\mathrm{group})}\}$ are initialized randomly.

M Step In the M step, the reference vector $\mathbf{Q} = (\mathbf{q}_n)$ is updated by

$$\mathbf{q}_n = \sum_{k=1}^{K} R_{nk}^{(\mathrm{group})} \mathbf{p}_k, \tag{12}$$

where

$$R_{nk}^{(\mathrm{group})} = \frac{h(\zeta_n^{(\mathrm{group})}, \mathbf{z}_k^{(\mathrm{group})})}{\sum_{k'} h(\zeta_n^{(\mathrm{group})}, \mathbf{z}_{k'}^{(\mathrm{group})})}. \tag{13}$$

Note that $\mathbf{q}_n$ represents the conditional probability $q_{ln} \equiv q(\zeta_l^{(\mathrm{member})} \,|\, \zeta_n^{(\mathrm{group})})$. It is assured that $\mathbf{q}_n$ satisfies the axiom of the probability that it is nonnegative and the sum is 1, since Nadaraya-Watson kernel smoother generates mixture distributions of $\{\mathbf{p}_k\}$.

E Step The E step determines the best matching nodes. As $\mathbf{p}_k$ and $\mathbf{q}_n$ represent probabilities, the errors are evaluated by the cross entropy instead of Euclidean distance,

$$n_k^\star = \arg\max_n \sum_{l=1}^{L} p_{lk} \ln q_{ln}. \tag{14}$$

The latent variables are then estimated as $\mathbf{z}_k^{(\mathrm{group})} = \zeta_{n_k^\star}^{(\mathrm{group})}$.

These M and E steps are iterated alternately until the group map converges.

4 Coloring and exploring maps

4.1 Coloring

To discover knowledge from the generated maps, the maps can be colored by various ways. The most common coloring method of SOM is *U-matrix*, which represents Laplacian of $f(\mathbf{z})$ [5]. By using U-matrix, regions with large differentials are emphasized in the map, suggesting the cluster boundaries. In this paper, we used full-color representation for U-matrix, by which the cluster boundaries are indicated by red (see Fig. 2 (a)–(c) and Fig. 4).

Another common coloring method is the *component plane*, which enhances the region where the mapping $f(\mathbf{z})$ takes large value [6]. In TSOM, the method is extended to *Conditional Component Plane (CCP)*, by which one of the maps is colored according to f, while other latent variables are fixed as the condition [3]. For example, if one wants to see the score distribution of members for a query, the member map is colored by regarding $f(\mathbf{z}^{(\mathrm{member})}, \mathbf{z}^{(\mathrm{query})}) \equiv f(\mathbf{z}^{(\mathrm{member})} \,|\, \mathbf{z}^{(\mathrm{query})})$ as a function of $\mathbf{z}^{(\mathrm{member})}$, while $\mathbf{z}^{(\mathrm{query})}$ gives the condition. Thus the member map is colored according to the members' score when a query is specified by the position in the query map.

If one wants to see the property of a group, then the query-map is colored by the expectation of CCP (ECCP), as

$$\overline{f}(\mathbf{z}^{(\mathrm{query})}|\mathbf{z}^{(\mathrm{group})}) \equiv \mathbb{E}_{q(\mathbf{z}^{(\mathrm{member})}|\mathbf{z}^{(\mathrm{group})})}\left[f(\mathbf{z}^{(\mathrm{query})}|\mathbf{z}^{(\mathrm{member})})\right]. \tag{15}$$

Here $\overline{f}(\mathbf{z}^{(\mathrm{query})}|\mathbf{z}^{(\mathrm{group})})$ represents the average score of the members belonging to the group $\mathbf{z}^{(\mathrm{group})}$.

In this paper, CCP and ECCP are represented by blue–red colors, where red/blue correspond to high/low regions. For example, Fig. 2 (e), (f), (h), (i) are colored by CCP. Fig. 6 (c)–(e) and Fig. 7 (c)–(e) are also colored by CCP, while Fig. 6 (b) and Fig. 7 (b) are ECCP.

The last coloring visualizes a member distribution. Since $q(\mathbf{z}^{(\mathrm{member})}|\mathbf{z}^{(\mathrm{group})})$ represents a probability distribution at $\mathbf{z}^{(\mathrm{group})}$, the member distribution of $\mathbf{z}^{(\mathrm{group})}$ can be visualized by coloring the member-map according to q. In Fig. 2 (d) (g), the member distribution is represented by monochromatic scale.

4.2 Interactive analysis by graphical user interface

To make the above analysis easier, we developed an interactive graphical user interface (GUI) for TSOM network[2]. When a user wants to analyze the result with respect to a group, all one need is to point the group on the map in the display. Similarly, when one wants to analyze a member cluster, the score is indicated in the query map by pointing the cluster position on the member map. With the help of the interactive GUI system, it is easy to grasp the entire image of the multigroup dataset intuitively. This is an advantage of using TSOM in such multilevel-multigroup analysis.

[2] `http://www.brain.kyutech.ac.jp/~furukawa/tsom/multilevel`

5 Related works

5.1 Multigroup analysis and multilevel analysis

In statistics, methods for simultaneous analysis of multiple datasets are called multigroup analysis. In general, multigroup analysis estimates the common structure among groups, and analyzes individual groups under the common structure. A representative is simultaneous component analysis (SCA) which is the extension of PCA for multigroup analysis [7]. In SCA, the loading matrix is shared among groups as the common structure, while the component scores are evaluated for each group. Various variations of SCA have been developed [8–12]. Factor analysis and partial least squares (PLS) have been also extended to multigroup analysis [13, 14]. From the viewpoint of multigroup analysis, the proposed method projects all data points to the same nonlinear subspace as the common structure, whereas SCA projects them to the common linear subspace.

While multigroup analysis mainly focuses to intra-group analysis under the common structure, the methods focusing to inter-group analysis are called multilevel analysis [15]. In the multilevel SCA (MLSCA), PCA is applied to analyze the inter-group distribution as well as to analyze the intra-group distributions [1]. Many other hierarchical modeling methods for multilevel analysis have been proposed [16–18]. From the viewpoint of multilevel analysis, the proposed method achieves it by the hierarchically combined TSOMs.

In the machine learning field, there is a class of methods called multitask learning [19]. The purpose of multitask learning is to improve learning accuracy by sharing knowledge among different learning tasks [20–22]. Though the aim of multitask learning is different from multigroup analysis, both categories deal with similar datasets.

From the viewpoint of visualization, this work is most relevant to document analysis using Bag-of-Words (BoW) [23, 24]. Especially visualization method combining BoW and dimensionality reduction is strongly related to this work [25–28]. These methods visualize a set of documents or topics in terms of the word occurrence probability, while our method visualizes a set of groups in terms of the distribution of members.

5.2 Multiway Dimensionality reduction

Roughly speaking dimensionality reduction methods are classified to two categories [29]. The first category is the spectral method, which aims to map data points to a low dimensional space with keeping their similarities or distances. This group contains multidimensional scaling (MDS), kernel PCA (kPCA), manifold learnings such as ISOMAP, and so on. An advantage of this category is that the task becomes a convex optimization problem, and a disadvantage is that it is not easy to recover the original data from the low dimensional representation. The latter is called pre-image problem [30].

On the other hand, the methods in the second category aim to estimate a mapping from the low dimensional latent space to the high-dimensional data space. SOM [31], Generative Topographic Mapping (GTM) [32], and Gaussian Process Latent Variable Model (GPLVM) [33] are the representatives of this category. An advantage of this category is that it is easy to recover the original data from the low dimensional representation as well as their interpolations. Thus they do not suffer the pre-image problem. The visualization by the component plane is based on this property, and it enhances the interpretability of the analysis result. This is the reason why we chose SOM in this study.

As an alternative to SOM, coloring by component plane is also possible by GTM and GPLVM. The essential difference between these methods and SOM is that GTM and GPLVM use Gaussian process (GP) to produce smooth manifold, whereas SOM uses Nadaraya-Watson kernel smoother. To deal with a set of probability distributions, Nadaraya-Watson smoother is much easier than GP, because it produces nonnegative mixtures of given distributions. Thus the axiom of probability is always assured for the smoothing result. This is the reason why we chose SOM. The disadvantages of SOM are the discrete expression of the map, and the difficulty of Bayesian treatment. These are our future issues.

To make a multiway analysis, singular value decomposition (SVD) or tensor decomposition are usually used [34]. Iwasaki and Furukawa extended SOM and GTM for multiway analysis, namely, TSOM and Tensor GTM (TGTM). These are extensions of Tucker type tensor decomposition [3]. TSOM and TGTM provide various coloring methods such as CCP. These are the reason why we chose TSOM for multilevel-multigroup analysis.

6 Experiments

The proposed method was applied to three datasets. The first one is the sushi preference dataset, used as a benchmark test. In this experiment, groups of users are organized by their gender and age. The second dataset is a psychological survey of football players in a university league. The aim of this experiment is to analyze the football teams in terms of their constituent player types. The third dataset is a morale survey in a company. The aim is to analyze the departments of the company in terms of the morale condition of their constituent employees.

6.1 Sushi preference data

The sushi dataset[3] consists of the ranking scores of 10 sushi toppings evaluated by 5,000 respondents [35]. Thus, 10 toppings are ranked from 1 to 10 by each respondent. In this work, they are regarded as scalar scores. The respondents were grouped according to age and gender.

The result is shown in Fig. 2. Fig. 2 (a)–(c) are the maps of respondents, sushi toppings, and groups. In these maps, more similar/different items are arranged closer/farther. For example, respondents with similar preferences are located closer in the map (a), and sushi toppings preferred by similar respondents are also arranged nearer in the map (b). These results are consistent with the previous work [3]. In the group map, men (labelled by M) and women (labelled by F) are located separately. In addition, the groups are ordered according to the generation (labelled by 10 to 50, indicating 10s to 50s)[4]. Note that the group map represents the similarities of groups in terms of their member distributions, and it does not represent the similarities of preference directly.

Fig. 2 (d)–(f) show the member distribution of 20s women and their preferences, while (g)–(i) are of 40s women. In the case of 20s women, there are two distinct clusters (labelled by *a* and *b* in (d)), one of which prefers sea eel and egg, and the other does not. Similarly, 40s women constitute two clusters, one prefers salmon roe and sea urchin, and the other do not. When a group consists of such opposite types of members, the average of the group has no meaning.

[3] http://www.kamishima.net/sushi/

[4] 50F and 50M include 60s people.

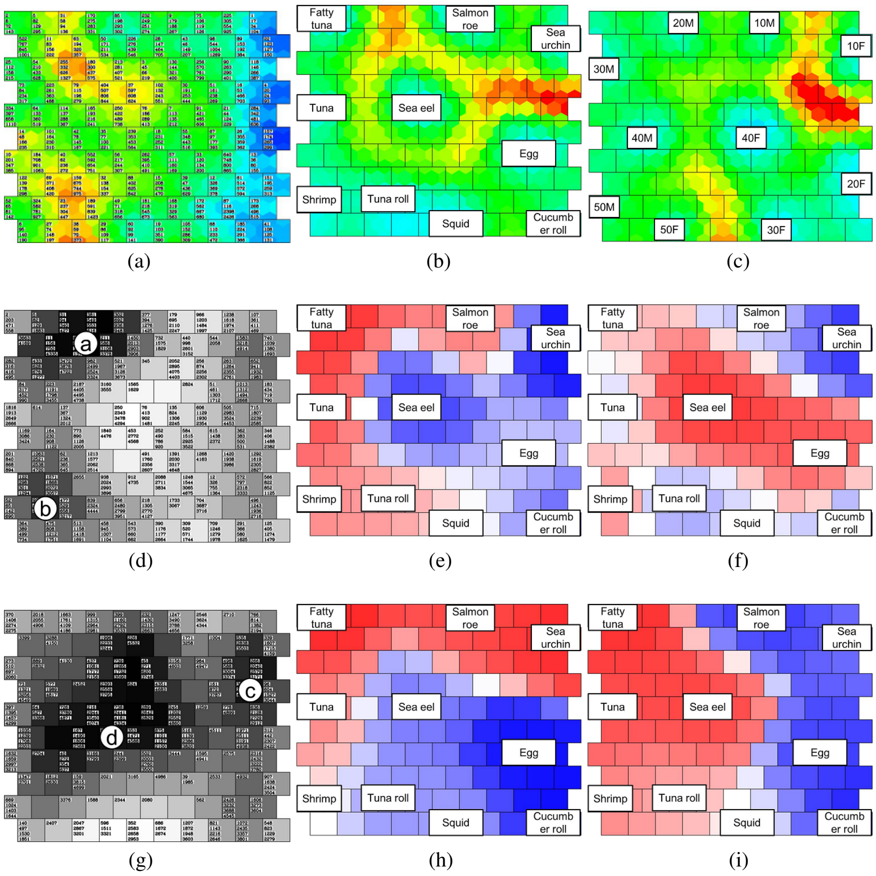


Fig. 2 The result of the sushi preference dataset. (a) Respondent map (b) Sushi map (c) Group map. 10F means 10s women, and 40M means 40s men. (a)–(c) are colored by U-matrix. (d) Member distribution of 20s women in the member map. Two clusters a and b are labelled. (e) (f) Sushi preference of 20s women at cluster centers a and b. The red/blue regions represent favorite/dislike items. (g) Member distribution of 40s women in the member map. Two clusters c and d are labelled. (h) (i) Sushi preference of 40s women of cluster centers c and d.

To verify the result, we analyzed each group by k-means clustering separately. Fig. 3 shows the result of women in 20s and 40s. The reference vectors of k-means represent the preferences of clusters in the group (upper low). The corresponding preferences obtained from TSOM at the cluster center show consistent results (lower row). Note that k-means represent the clusters in the high dimensional data space, whereas clusters of TSOM are in the low dimensional latent space. Nevertheless both results are consistent.

6.2 Football team data

The second dataset is a psychological questionnaire survey of football players in a university league in Japan. The survey was taken for the 439 players of 16 universities. The questionnaire consisted of 112 queries of Likert scale asking how players assess themselves and their

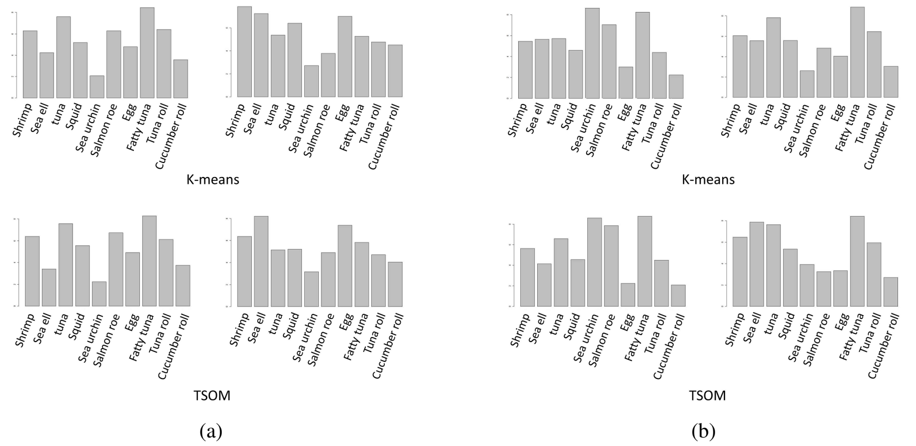


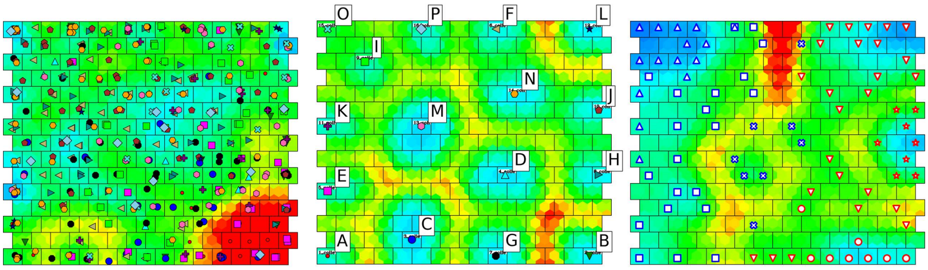
Fig. 3 The reference vectors of k-means obtained from subsets of the sushi dataset. First 2 primary components are indicated (upper row). The corresponding results of TSOM are also indicated (lower row). (a) Result from 20s women. These correspond to the clusters *a, b* in Fig. 2d. (b) Result from 40s women. These correspond to the clusters *c, d* in Fig. 2g.

coaches. The queries were categorized as six types: (1) motivation, (2) self-management skill, (3) football skill, (4) coach acceptance, (5) performance and maintenance (PM) of coaching, and (6) perceived coaching effectiveness (PCE). Questions (1)–(3) concern self-assessment, and (4)–(6) concern the assessments of the coach by the players. The data were standardized in advance so that the mean and variance were zero and one, respectively.

By applying the proposed method, three maps, i.e., member-map, group-map, and query-map were produced as shown in Fig. 4. In the query map Fig. 4 (c), the six categories are separated. In particular, a clear cluster boundary appeared between self-assessment and coach-assessment queries, suggesting the result is plausible.

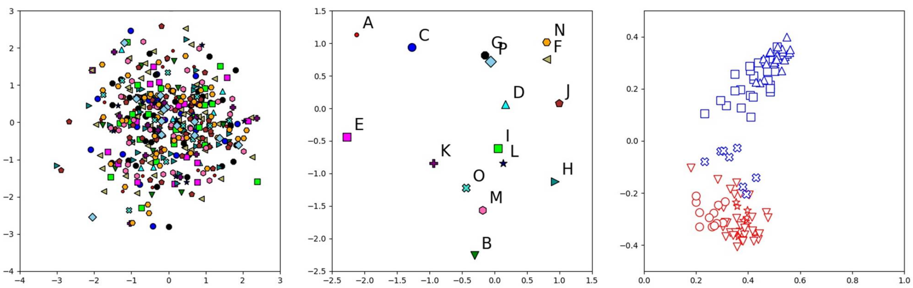
To compare with existing parametric approach, we analyzed this dataset by Multi-level Simultaneous Component Analysis (MLSCA) [36]. In MLSCA, a set of groups are analyzed according to the average values of the constituent members. The results are shown in Fig. 5. Compared with Fig. 4, the arrangement of the teams (b) and queries (c) are almost similar to each other, indicating the consistent results. (In MLSCA, the member map (a) is arranged according to the relative scores within each group, whereas it is arranged by the absolute scores in our method. Thus the member maps of two methods cannot be compared.) However, some teams are located differently. For example, team B and O are located at opposite corners in our method, whereas they are almost located similar positions in MLSCA result.

Fig. 6 and Fig. 7 show the constitution of teams B and O, indicating that both consist of three clusters of players (Fig. 6 (a), Fig. 7 (a)), and the score distributions of those clusters are quite different (Fig. 6 (c)–(e), Fig. 7 (c)–(e)). The proposed method clearly discriminates between such differences. However, the averaged score distributions are almost equal in the query map (Fig. 6 (b), Fig. 7 (b)). In MLSCA, the team map represent the average scores of the teams, and the differences between members within each group are ignored. This is the reason why MLSCA arranged team B and O at similar positions in the team map.



(a) (b) (c)

Fig. 4 The maps of football players and teams generated by our algorithm. (a) The member (player) map. The symbols represent the belonging teams. (b) The group (team) map. (c) The query map. (△: Percieved Coaching Effectiveness, □: Performance and Maintenance, ×: Coaching Acceptance, ▽: Self-Management, ◯: Football Skill, ★: Motivation). These maps are colored by U-matrix method.



(a) (b) (c)

Fig. 5 Result of the team analysis by MLSCA. (a) The member map. The symbols represent the belonging teams. (b) The group (team) map. (c) The query map. Symbols represent the same as in Fig. 4.

6.3 Morale survey data

In the third experiment, we applied the method to a morale survey dataset in a company. The task of the analysis is to compare the departments in the company in terms of the morale state of the employees. The survey was taken for the 482 employees belonging to 15 departments. The survey consists of 16 queries of 5 grade Likert scale, consisting of 5 categories. (1) Relationship with supervisors, (2) human resource development, (3) work content and operation efficiency, (4) management policy of the company, (5) safety management, (6) paid leave.

The result of the proposed method is shown in Fig. 8, and the MLSCA result is shown in Fig. 9. Departments B, F, I have the similar average scores of their members, and are plotted in the same place in the MLSCA result. However, as shown in Fig. 10, the departments have different member distributions. Our method successfully represents the difference.

7 Conclusion

In this paper, we proposed a visualization method for multilevel-multigroup analysis using a hierarchical TSOM network. The proposed method provides flexible modeling by multiway

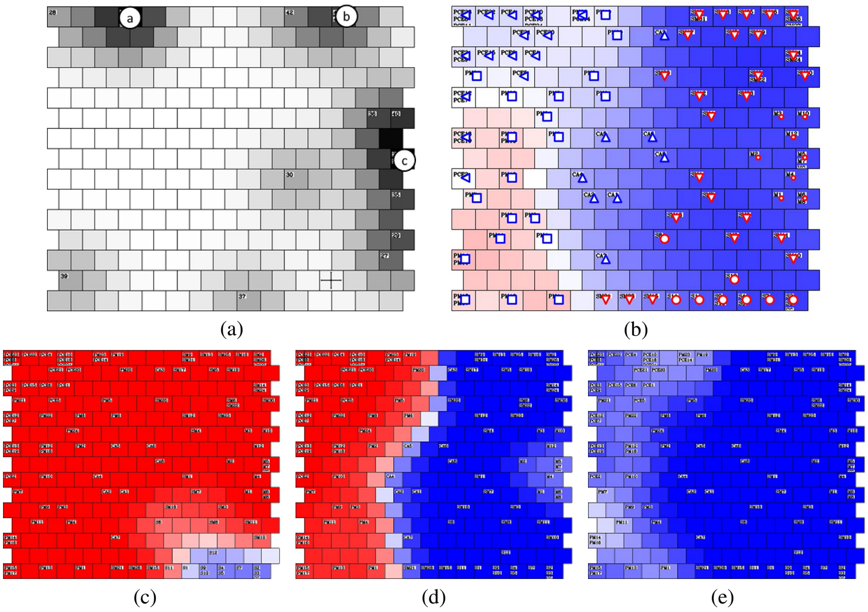


Fig. 6 The player distribution of team B in Fig. 4b. (a) The distribution of players from team B in the member map. (b) The averaged score of the members in the query map. (c)–(e) The scores of clusters a, b, c in the query map.

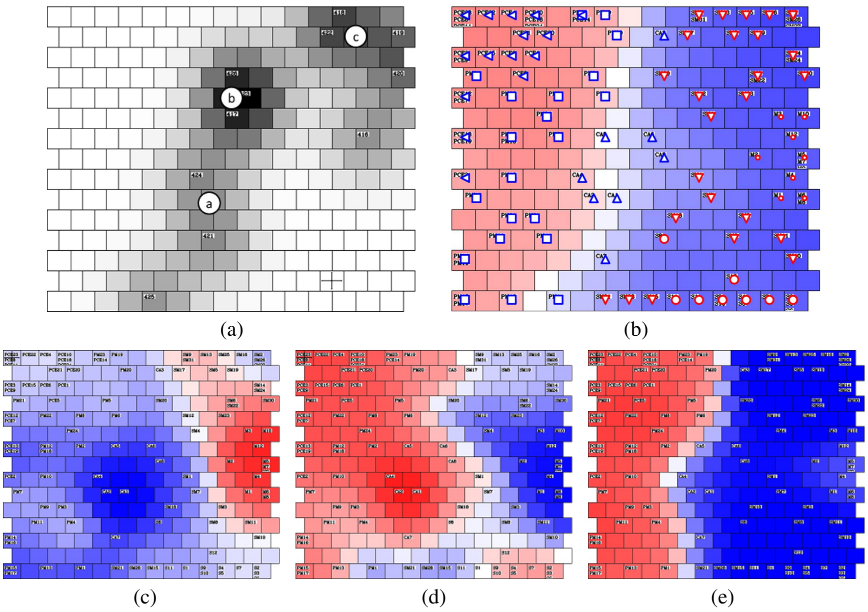
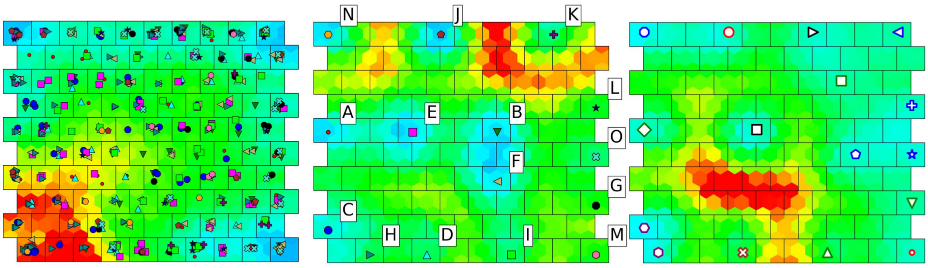


Fig. 7 The player distribution of team O in Fig. 4b. (b) The averaged score of the members in the query map. (c)–(e) The scores of clusters a, b, c in the query map.



(a) (b) (c)

Fig. 8 The maps of employees and departments in a company obtained from the morale survey dataset. (a) The member (employee) map. (b) The group (department) map. (c) The query map. (red: management policy, green: workcontent and operation efficiency, blue: relationship with superviser, purple: safety management, brown: paid leave.) These maps are colored by U-matrix method.

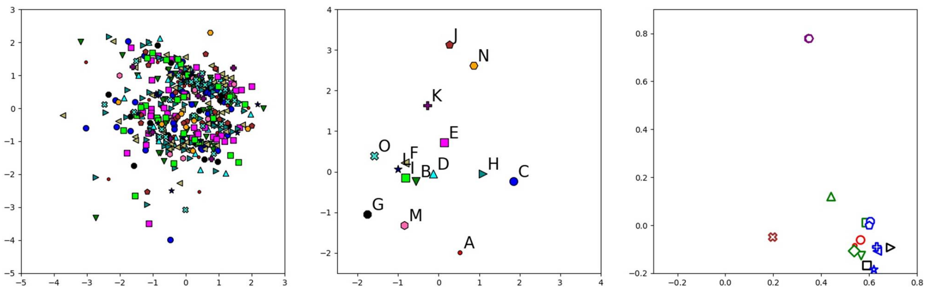
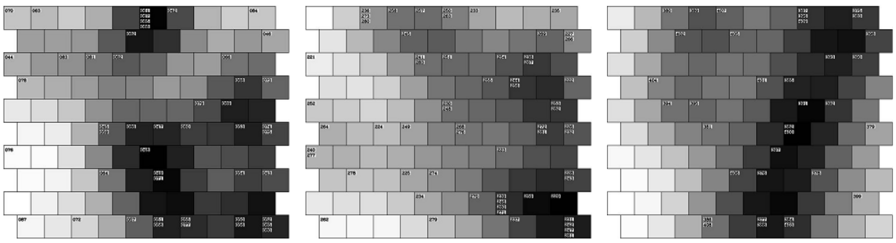


Fig. 9 Department analysis result by MLSCA. (a) The member (employee) map. (b) The group (department) map. (c) The query map.



(a) (b) (c)

Fig. 10 The employee distributions of three departments (a) Department B, (b) Department F, (c) Department I.

nonlinear dimensionality reduction. It also ensures ease of analysis and highly interpretability while avoiding information loss by averaging. With the interactive GUI, this method becomes more useful tool which enables us to grasp the entire image of multigroup dataset intuitively.

We call the method TSOM *network*, though it consists of only two TSOMs in this paper. Actually it is easy to extend this method for more complex cases, and it becomes a network of TSOMs. For example, in the case of user–item rating data of an on-line shop, both user groups and item groups may exist. In this case, it needs a network consisting of three TSOMs. Another example is the e-mail analysis in a company, in which e-mails are grouped by the senders and the receivers, and they are grouped according to their departments and the positions. Our method easily adapts to such highly complex cases. Thus, this approach presents a powerful and general tool for multilevel-multigroup analysis.

Acknowledgements We would like to thank Prof. Isogai and Dr. Shinriki, Kyushu Institute of Technology, for allowing us to use football player data. We also would like to thank Dr. Iwasaki for much assistance in developing analysis tools and useful discussion.

References

1. Timmerman M (2006) Multilevel component analysis. British Journal of Mathematical and Statistical Psychology 59:301–320
2. Friedman J (1987) Exploratory projection pursuit. Journal of the American Statistical Association 82:259–266
3. Iwasaki T, Furukawa T (2016) Tensor SOM and Tensor GTM: Nonlinear tensor analysis by topographic mappings. Neural Networks 77:107–125
4. Ishibashi H, Furukawa T (2016) Multilevel-multigroup analysis by hierarchical tensor SOM network. In: Proc. of ICONIP, pp 459–466
5. Ultsch A, Siemon HP (1990) Kohonen's self organizing feature maps for exploratory data analysis. In: Proc. INNC'90, Int. Neural Network Conf., pp 305–308
6. Stefanovic P, Kurasova O (2011) Visual analysis of self-organzing maps. Nonlinear Analysis: Modelling and Control 16(4):488–504
7. Ten Berge JMF, Kiers HAL, Van der Stel V (2013) Simultaneous components analysis. Statistica Applica 4:377–392
8. Ten Berge JMF, Kiers HAL, Van der Stel V (2011) A clusterwise simultaneous component method for capturing within-cluster differences in component variances and correlations. British Journal of Mathematical and Statistical Psychology 66:81–102, DOI 10.1111/j.2044-8317.2012.02040.x
9. De Roover K, Timmerman ME, Mesquita B, Ceulemans E (2013) Common and cluster-specific simultaneous component analysis. PLoS One 8:e62,280, DOI 10.1371/journal.pone.0062280
10. De Roover K, Ceulemans E, Giordani P (2016) Overlapping clusterwise simultaneous component analysis. Chemometrics and Intelligent Laboratory Systems 156:249–259
11. Schouteden M, Van Deun K, Wilderjans TF, Van Mechelen I (2013) Performing DISCO-SCA to search for distinctive and common information in linked data. Behavior Research Methods 46:576–587
12. Zhou G, Cichocki A, Zhang Y, Mandic D (2015) Group component analysis for multiblock data: Common and individual feature extraction. IEEE Transactions on Neural Networks and Learning Systems
13. French BF, Finch WH (2008) Multigroup confirmatory factor analysis: Locating the invariant referent sets. Structural Equation Modeling 15:96–113
14. Eslamia A, Qannarib EM, Kohlerd A, Bougeard S (2014) Algorithms for multi-group PLS. Journal of Chemometrics 28:192–201, DOI 10.1002/cem.2593
15. de Leeuw J, Meijer E (2008) Handbook of multilevel analysis. New York, NY: Springer
16. Wulsin D, Jensen S, Litt B (2012) A hierarchical dirichlet process model with multiple levels of clustering for human eeg seizure modeling. In: Proceedings of the 29th International Conference on Machine learning
17. Nguyen V, Phung DQ, Nguyen X, Venkatesh S, Bui HH (2014) Bayesian nonparametric multilevel clustering with group-level contexts. In: Proc. of International Conference on Machine Learning (ICML), pp 288–296
18. Hu L, Cao J, Xu G, Wang J, Gu Z, Cao L (2013) Cross-domain collaborative filtering via bilinear multilevel analysis. IJCAI
19. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering 22:1345–1359
20. Caruana R (1997) Multitask learning. Machine Learning 28:41–75
21. Argyriou A, Evgeniou T, Pontil M (2007) Multi-task feature learning. In: Proceedings of the 20th Annual Conference on Neural Information Processing Systems, pp 41–48
22. Evgeniou T, Pontil M (2004) Regularized multi-task learning. In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 109–117
23. Tsai CF (2012) Bag-of-words representation in image annotation: A review. In: Proceedings of the 9th European Conference on Computer Vision

24. Csurka G, Bray C, Dance C, Fan L (2004) Visual categorization with bags of keypoints. In: ECCV International Workshop on Statistical Learning in Computer Vision, pp 1–22
25. Cao N, Cui W (2016) Overview of Text Visualization Techniques, vol 1. Vol 1 the series Atlantis Briefs in Artificial Intelligence
26. Iwata T, Yamada T, Ueda N (2008) Probabilistic latent semantic visualization: topic model for visualizing documents. In: KDD, pp 363–371
27. Iwata T, Saito K, Ueda N, Stromsten S, Griffiths TL, Tenenbaum JB (2007) Parametric embedding for class visualization. Neural Computation 19:2536–2556
28. Le T, Lauw HW (2015) Semantic visualization for spherical representation. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 1007–1016
29. Lawrence ND (2005) Probabilistic non-linear principal component analysis with gaussian process. Journal of Machine Learning Research 6:1783–1816
30. Kwok JTY, Tsang IWH (2004) The pre-image problem in kernel methods. Neural Networks, IEEE Transactions on 15(6):1517–1525, DOI 10.1109/tnn.2004.837781, URL http://www.hpl.hp.com/conferences/icml2003/papers/345.pdf
31. Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biological Cybernetics 43(1):59–69
32. Bishop CM, Svensen M, Williams CKI (1998) GTM: The generative topographic mapping. Neural Computation 10:215–234
33. Lawrence ND (2003) Gaussian process latent variable models for visualisation of high dimensional data. In: Neural Information Processing Systems (NIPS), vol 16, pp 329–336
34. Kolda TG, Bader BW (2009) Tensor decompositions and applications. SIAM REVIEW 51(3):455–500
35. Kamishima T, Kazawa H, Akaho S (2010) A survey and empirical comparison of object ranking methods. In: Fürnkranz J, Hüllermeier E (eds) Preference Learning, Springer, pp 181–201
36. Ceulemans E, Wilderjans TF, Kiers HAL, Timmerman ME (2016) Multilevel simultaneous component analysis: A computational shortcut and software package. behavior research methods. Behavior Research Methods 48:1008–1020, DOI doi:10.3758/s13428-015-0626-8