The analysis based on principal matrix decomposition for 3-mode binary data

Haruka Yamashita* and Masayuki Goto

Department of Science and Engineering,
Waseda University,
3-4-1 Okubo, Shinjuku-ku, Tokyo, Japan
Email: h.yamashita@aoni.waseda.jp
Email: masagoto@waseda.jp
*Corresponding author

Abstract: Recently, principal points for a multivariate binary distribution (Yamashita and Suzuki, 2014, 2015) have been proposed as the binary vectors that optimally represent a distribution, in terms of the average Euclidian squared distance between a multivariate binary distribution and the vectors. In this paper, we propose a new analysis procedure for 3-mode binary data, based on principal points for a multivariate binary distribution (Yamashita and Suzuki, 2014, 2015). Moreover, we propose a method that decomposes principal matrixes for 3-mode binary data into a small number of vectors based on vector products. In order to investigate our method’s applicability to real-world data, we use the method to analyse 3-mode structured data from annual all-star games for Japanese professional baseball.

Keywords: principal points; 3-mode data; binary data; clustering; data analysis.


Biographical notes: Haruka Yamashita is a research associate of School of Creative Science and Engineering, Waseda University, Japan. She received her PhD from Keio University in 2015. Her research interests include applied statistics, multivariate analysis, and machine learning, statistical quality control, and big data analysis. She is a member of Japan Industrial Management Association, Japanese Society for Quality Control, Japanese Society of Applied Statistics, and Japan Association for Management Systems.

Masayuki Goto is a Professor of School of Creative Science and Engineering, Waseda University, Japan. He received his Dr.E. degree from Waseda University in 2000. His research interests include information theory, information statistics, management information, business analytics, and marketing analysis. He is a member of the Institute of Electronics, Information and Communication Engineers, Japan Industrial Management Association, the Japan Society for Management Information, the Japanese Society for Artificial Intelligence, IEEE, Business Model Association, and the Operations Research Society of Japan.
The analysis based on principal matrix decomposition

This paper is a revised and expanded version of a paper entitled ‘An analysis based on principal matrix decomposition for 3-mode binary data’ presented at Asian Conference for Management Science and Applications, Dalian, China, 12–16 September 2015.

1 Introduction

Summarising data with its small subset is often a useful approach in data analysis. Principal points (Flury, 1990) which have been studied for a few decades mainly in statistics allow us to carry out such an analysis in a variety of applications. In fact, the principal points analysis has been applied to many problems, such as the optimal sizing of protection masks, selective assembly for two mating parts in manufacturing, and discriminant analysis for drug and placebo responses. Although it has been widely demonstrated as shown by the above studies, that principal points are useful for analysing data from continuous distributions, it is not straightforward to extend a similar idea to the analysis of data from a multivariate binary distribution. In general, a set of $k$-principal points of a distribution $f$ is defined as a set of $k$ points that optimally represents $f$ in terms of the mean squared distance. However, such points do not generally have binary elements, although the data emitted from such a binary distribution has discrete elements on the binary region.

Recently, an approach for principal points for a multivariate binary distribution is proposed (Yamashita and Suzuki, 2015), where principal points are restricted to the binary region. This approach provides several representative patterns of multivariate binary data, and the points, which are not included in the data samples but important for representing the data, can be selected as representative patterns. Moreover, the definition of principal points for a multivariate binary distribution to principal matrixes for 3-way data, such as paired comparison data for each person is defined (Yamashita, 2015).

In this study, we propose an analysis method for 3-mode binary data, such as time series evaluation data of each item for each person. Our method is composed of two steps:

1. finding principal matrixes
2. decomposing each matrix into product of two factor vectors.

The proposed decomposition method based on vector product enables the principal matrixes to be compressed into a small number of factor vectors. 3-mode binary data is characterised by a combination of the two factor vectors. New findings about these two factor vectors can be obtained using this expression. In order to investigate the applicability of our method to real-world data, we analyse 3-mode structured data from annual all-star game for Japanese professional baseball teams, which includes teams and player position data from 2000 to 2015.

This paper is organised as follows. Section 2 provides a discussion on the original principal points, principal points for a multivariate binary distribution, and the principal matrix for 3-way binary data. In Section 3, the principal matrix for 3-way binary data is applied to the principal matrix for 3-mode binary data, and we propose a method that decomposes the principal matrix for 3-mode binary data into a small set of vectors. In
Section 4, an example for the application of principal matrixes for 3-mode binary data is described in detail, and in Section 5 we conclude the paper.

2 Principal points and principal matrixes

2.1 Conventional principal points

Recently, numerous research efforts have focused on ‘principal points’ (Flury, 1990). A set of $k$-principal points of a distribution $f$ is defined by a set of $k$ points that optimally represent the distribution $f$ in terms of mean squared minimum distance. To state it more precisely, let $X = (X_1, ..., X_p)$ be a $p$-variate random vector following a distribution $f$. A set of $k$ points $\xi_1, ..., \xi_k$ in $\mathbb{R}^p$ is called a set of $k$-principal points of the random vector $X$, if the $k$ points satisfy the following condition,

$$\{\xi_1, ..., \xi_k\} = \arg \min_{y_j \in \mathbb{R}^p, 1 \leq j \leq k} E_f[d^2(X|y_1, ..., y_k)],$$

(1)

where $d^2(x | y_1, ..., y_k)$ is the minimum squared distance between a point $x \in \mathbb{R}^p$ and $k$ points $y_1, ..., y_k \in \mathbb{R}^p$ defined as follows:

$$d^2(x | y_1, ..., y_k) = \min_{1 \leq j \leq k} (x - y_j)(x - y_j)^T.$$  

(2)

Properties of principal points have been widely discussed in many papers. For univariate distributions, the conditions for the uniqueness of principal points have been studied (Zoppe, 1995; Yamamoto and Shinozaki, 2000). For multivariate distributions, Tarpey et al. (1995) established the principal subspace theorem, which specifies a linear subspace in which a set of principal points of an elliptically symmetric distribution exists. Recently, the principal subspace theorem has been extended to handle distributions consisting of several subgroups such as location mixtures of spherically symmetric distributions (Kurata, 2008; Matsuura and Kurata, 2011) and a mixture of elliptically symmetric distributions that form an allometric extension model (Matsuura and Kurata, 2014). Principal points have also been applied to many data analysis problems, such as optimal sizing of protection masks (Flury, 1993), selective assembly for two mating parts in manufacturing (Mease et al., 2004; Mease and Nair, 2006; Matsuura, 2011), functional data analysis (Tarpey and Kinateder, 2003; Shimizu and Mizuta, 2007, 2008), and discriminant analysis for drug and placebo responses (Tarpey et al., 2010).

2.2 Principal points for a multivariate binary distribution

Although the above studies have demonstrated that principal points are very useful for analysing continuous distributions, difficulties remain in analysing categorical distributions. When $X$ is a multivariate binary random variable, that is, follows a multivariate binary distribution $f$ that has a total mass on the $p$-dimensional space $S = \{0, 1\}^p$ and $|S| = 2^p$, the original principal
points defined by equation (1) are not binary values. For example, consider a 5-variate binary random vector $X$: $P[X = (0, 0, 0, 0, 0)] = P[X = (0, 0, 0, 1, 0)] = P[X = (0, 0, 1, 0)] = P[X = (0, 1, 0, 0)] = P[X = (1, 0, 0, 0)] = \frac{1}{4}$. Then the 2-principal points are $(0, 0, 0, \frac{1}{4}, \frac{1}{4})^T$ and $(1, 1, 1, \frac{1}{2}, 0)$, which are not in $S$; in other words, they are not realisations from the random vector $X$. Hence, it is difficult to give an interpretation of the principal points, especially when the variables are categorical (e.g., male or female).

Yamashita and Suzuki proposed the principal points for a multivariate binary distribution based on Flury’s definition by restricting the possible values of principal points to be binary (Yamashita and Suzuki, 2014, 2015). The definition of $k$-principal points for a multivariate binary distribution $f$ is given as follows: a set $\{\xi_1, \ldots, \xi_k\}$ in $S$ is called a set of $k$-principal points for a $p$-variate binary distribution $f$, if the following condition is satisfied,

$$\{\xi_1, \ldots, \xi_k\} = \arg \min_{\mathbf{y} \in S, \ 1 \leq j \leq k} E_f[d^2(X \mid y_1, \ldots, y_k)] = \arg \min_{\mathbf{y} \in S, \ 1 \leq j \leq k} \sum_{\mathbf{x} \in S} P[X = x]d^2(x \mid y_1, \ldots, y_k),$$

(3)

where $d^2(x \mid y_1, \ldots, y_k)$ is the minimum squared distance between a binary point $x \in S$ and $k$ binary points $y_1, \ldots, y_k$ in $S$ defined as follows:

$$d^2(x \mid y_1, \ldots, y_k) = \min_{1 \leq j \leq k} (x - y_j)(x - y_j)^T.$$

(4)

This formulation allows us to obtain $k$-principal points for a multivariate binary distribution which are realisations from the distribution $f$.

Note that other metrics can be considered for the definition of principal points for a multivariate binary distribution; however, the Manhattan distance and Hamming distance are equivalent to the squared Euclidean distance in $S$, and Euclidean distance is one of the most basic distances. Because this distance is widely used in many cases, we apply Euclidean distance for defining principal points for a multivariate binary distribution.

In case in which we estimate principal points for a multivariate binary distribution estimated from given data $x_1, \ldots, x_n$, we cannot know the true distribution $f$; therefore, non-parametric and parametric estimation method (Yamashita et al., 2015) of principal points for a multivariate binary distribution have been proposed. In a non-parametric estimation method, principal points for a multivariate binary distribution are obtained from the empirical distribution by using the number of each observation as follows:

$$\hat{P}[X = x] = \frac{1}{n} \sum_{i=1}^{n} \delta(x = x_i),$$

(5)

where $\delta(x = x_i)$ is an indicator function that returns 1 if $x = x_i$, otherwise it returns 0. It should be noted that the parametric estimation method finds principal points for a multivariate binary distribution from the estimated distribution by assuming a multinomial distribution of the given data and applying log-linear model (Yamashita et al., 2015). In case that the amount of the data is small, the algorithm for the parameters may not converge, and we may not estimate principal points for multivariate binary data. On the other hand, because the non-parametric approach assumes an
empirical distribution of the given data, it is able to estimate a distribution regardless of whether the amount of the data is small or large. Hence, the non-parametric method is useful for the real-world data analysis.

Here, we describe an example of an analysis with principal points for a multivariate binary distribution. Suppose that we have questionnaire data collected from customers, in which they were asked to select quality factor(s) important to them when selecting clothes. The probability of each pattern $x_i$ is shown in Table 1, and illustrated in Figure 1. As defined in equation (3), 2-principal points are (110) (i.e., price and shape) and (001) (i.e., utility). Note that the most frequently occurring pattern (100) (i.e., price) with the highest probability is not selected as a principal point. The search of principal points for a multivariate binary distribution is an NP-hard problem; therefore, several approximation methods have been proposed (Yamashita and Suzuki, 2014, 2015). These approximation methods enable an analysis of multivariate binary data using the principal points even if the number of variables is large. Moreover, Yamashita et al. (2015) have discussed the method for estimating principal points for a multivariate binary distribution.

Table 1  Example of a multivariate binary distribution

<table>
<thead>
<tr>
<th>Price</th>
<th>Shape</th>
<th>Utility</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.08</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.13</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.21</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.21</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.13</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Notes: The right-most column represents the true probability of each pattern.
For example, the second row indicates 25% of customers chose only Price.

Figure 1  Illustration of principal points for multivariate binary data (see online version for colours)

Note: The 2-principal points for multivariate binary data are (110) and (001).
2.3 Principal matrix for 3-way binary data

A principal matrix for 3-way binary data (Yamashita, 2015) was proposed by expanding the model of principal points for multivariate binary data; this method enables us to analyse complicated structured data, such as paired comparison data for each person.

Let \( n \times p \times p \) 3-way binary data \( X \in \{0, 1\}^{n \times p \times p} \) be the collection of elements

\[
\{x_{irs}|i = 1, ..., n, r = 1, ..., p, s = 1, ..., p\}. \tag{6}
\]

The elements are placed in a 3-dimensional block, such that (1) index \( i \) runs along the depth axis and shows the element of subset \( V \), where \( |V| = n \), (2) index \( r \) runs along the vertical axis and shows the element of subset \( W \), where \( |W| = p \), and (3) index \( s \) runs along the horizontal axis and shows the element of subset \( W \), where \( |W| = p \), that is, variable indexes \( r \) and \( s \) show the elements of a common subset \( W \).

For instance, for paired comparison data associated with a person, assume that index \( i \) identifies the person, and that \( r \) and \( s \) identify the objects being compared.

Following the definition of principal points (Flury, 1990) and principal points for a multivariate binary distribution (Yamashita and Suzuki, 2014, 2015), and applying the non-parametric estimation of principal points for a multivariate binary distribution (Yamashita et al., 2015) based on the given data, the definition of \( k \)-principal matrixes for binary 3-way data \( Z_j \in \{0, 1\}^{p \times p} (j = 1, ..., k) \) is given as follows:

\[
\{Z_1, \ldots, Z_k\} = \arg\min_{Y_j \in \{0, 1\}^{p \times p}, 1 \leq j \leq k} \sum_{i=1}^{n} d^2(X_i | Y_1, \ldots, Y_k), \tag{7}
\]

where \( d^2(X_i | Y_1, \ldots, Y_k) \) is the minimum squared Frobenius norm between \( X_i \) and \( \{Y_1, \ldots, Y_k\} \):

\[
d^2(X_i | Y_1, \ldots, Y_k) = \min_{1 \leq j \leq k} \|X_i - Y_j\|_F^2. \tag{8}
\]

Because equation (7) is a simple expansion of equation (3), the searching algorithms for principal points for a multivariate binary distribution (Yamashita and Suzuki, 2014, 2015) can be utilised for finding principal matrixes for 3-way binary data. Furthermore, Yamashita investigated the adequacy of the method by applying it to questionnaire data for the means of transportation between two Japanese cities (Yamashita, 2015).

3 An analysis based on principal matrix decomposition for 3-mode binary data

3.1 Principal matrix for 3-mode binary data

Yamashita proposed the principal matrix for 3-way binary data, such as paired comparison data of each person (Yamashita, 2015). This definition is applicable to 3-mode binary data, such as evaluation data for each item of each person in the small arrangement in the definition. 3-mode binary data \( X \) is composed of

\[
\{x_{irs}|i = 1, ..., n, r = 1, ..., p, s = 1, ..., q\}. \tag{9}
\]
where (1) index \( i \) runs along the depth axis and shows the element of subset \( V \), where \( |V| = n \), (2) index \( r \) runs along the vertical axis and shows the element of subset \( W \), where \( |W| = p \), and (3) index \( s \) runs along the horizontal axis and shows the element of subset \( U \), where \( |U| = q \). In 3-mode matrix, the variable indexes \( r \) and \( s \) do not indicate the elements of a common subset. For instance, for evaluation data containing items associated with a person, assume the index \( i \) identifies the person, \( r \) identifies an item and \( s \) identifies a question.

Then the definition of principal matrixes for 3-mode binary data \( Z_j \in \{0, 1\}^{p \times q} \) \( (j = 1, ..., k) \) is defined as follows:

\[
\{Z_1, \ldots, Z_k\} = \arg \min_{Y_j \in \{0, 1\}^{p \times q}, 1 \leq j \leq k} \sum_{i=1}^{n} d^2(X_i \mid Y_1, \ldots, Y_k),
\]

(10)

where \( d^2(X_i \mid Y_1, \ldots, Y_k) \) is a squared Frobenius norm between \( X_i \) and \( \{Y_1, \ldots, Y_k\} \) as follows:

\[
d^2(X_i \mid Y_1, \ldots, Y_k) = \min_{1 \leq j \leq k} \|X_i - Y_j\|_F^2.
\]

(11)

An illustration of 3-way data and 3-principal matrixes for 3-mode binary data is shown in Figure 2.

**Figure 2** Illustration of 3-principal matrixes for 3-mode binary data: the matrices on the right show 3-mode binary data, and the left 3 matrixes show the 3-principal matrixes for 3-mode binary data.
The analysis based on principal matrix decomposition

The obtained principal matrices $Z_j \in \{0, 1\}^{p \times q}$ ($j = 1, \ldots, k$) may help us to develop a rich understanding of the data. However, $Z_j$ may be considered in the direction of not only index $r$ but also $s$, and each $Z_j$ has $p \times q$ elements. Therefore, for interpreting the principal matrices for 3-mode binary data, it is a reasonable approach to decompose principal matrices into a small number of factor vectors, that is, the decomposing of $Z$ into $r$-directional vectors (i.e., vertical vectors) and $s$-directional vectors (i.e., horizontal vectors).

3.2 Principal matrix decomposition for 3-mode binary data

To describe principal matrices using a small number of vector products, one simple and effective method is to arrange $k$-principal matrices into a large rectangle shaped matrix and decompose the matrix into vertical and horizontal vectors.

In our method, the given $k$-principal matrices $Z_1, \ldots, Z_k$ found by (10) are rearranged into a large matrix $\Psi = [\psi_{ed}]_{e=1,\ldots,d=1,\ldots,\beta}$ as

$$\Psi = \begin{bmatrix}
\psi_{11} & \cdots & \psi_{1d} & \cdots & \psi_{1\beta} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\psi_{c1} & \cdots & \psi_{cd} & \cdots & \psi_{c\beta} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\psi_{\alpha 1} & \cdots & \psi_{\alpha d} & \cdots & \psi_{\alpha \beta}
\end{bmatrix},$$

where $\psi_{ed} \in \{Z_1, \ldots, Z_k\}$ and $k = \alpha \beta$. The number of rearrangements is represented by $k!$. For example, if $k = 4$, the given principal points are $Z_1, Z_2, Z_3$, and $Z_4$, and we set $\alpha = 2$ and $\beta = 2$, then there are $4!$ patterns for the arrangement of principal points, such as

$$\Psi = \begin{bmatrix} Z_1 & Z_2 \\ Z_3 & Z_4 \end{bmatrix},$$

and

$$\Psi = \begin{bmatrix} Z_2 & Z_3 \\ Z_1 & Z_4 \end{bmatrix}.$$  

The arranged large matrix is then decomposed into a vertical vector $a = (a_1, \ldots, a_{\alpha})^T$ and a horizontal vector $b = (b_1, \ldots, b_{\beta})$.

Hence, in our method, there are two optimisations: the optimisation for the rearrangement of principal matrices, and the optimisation for the decomposition of the large matrix. We formulate the procedure as follows:

$$\{\psi^*, a^*, b^*\} = \arg \min_{\psi = [\psi_{ed}] \in \{Z_1, \ldots, Z_k\}, a, b} \|\psi - ab\|_F^2. \quad (13)$$

Since the optimisation of $\psi$, $a$, and $b$ in equation (13) cannot be solved analytically, we decompose every pattern of $\psi$ whose number of cases is $k!$ and find the best solution.
Let $\Psi_t$ ($t = 1, ..., k$) be a $t$-th large matrix, and each large matrix is decomposed into $a_t = (a_{t1}, ..., a_{t\alpha})^T$ and the horizontal vector $b_t = (b_{t1}, ..., b_{t\beta})$ as

$$\{a_t, b_t\} = \arg \min_{a,b} ||\Psi_t - ab||_F^2.$$  (14)

Equation (14) is solved by considering a partial differentiation of $||\Psi_t - ab||_F^2$ in equation (14) by $a$ and a partial differentiation of $||\Psi_t - ab||_F^2$ by $b$. The optimal $a$ and $b$ are the vectors that satisfy

$$\frac{\partial ||\Psi_t - ab||_F^2}{\partial a} = 0$$ and $$\frac{\partial ||\Psi_t - ab||_F^2}{\partial b} = 0$$ at the same time, that is,

$$a = \frac{\Psi_t 1^T bb^T}{bb^T}, \quad b = \frac{1}{a^T a}.$$  (15)

where $1$ denotes a vector which every element is 1. Therefore, the solutions $a_t$ and $b_t$ ($t = 1, ..., k$) for each $\Psi_t$ are found based on the alternating least squares method (ALS) described in Figure 3, and we can find the best set of solutions $\{\Psi_t, a_t, b_t\}$. In our method, the best $\{\Psi_t, a_t, b_t\}$ denotes the set of $\{\Psi^*, a^*, b^*\}$ of equation (13). Nevertheless the solution using ALS may not be a global solution, at least local optimum solution is calculated by ALS. This method enables to interpret the $k$-principal matrixes for 3-mode binary data by combinations of $a^*$ and $b^*$ that show a summary of 3-mode binary data.

**Figure 3** Decomposition of a large matrix $\Psi_t$ into $\alpha$ vectors $a_{tc}$ ($c = 1, ..., \alpha$) and $\beta$ vectors $b_{td}$ ($d = 1, ..., \beta$)

4 Application of principal matrix decomposition for 3-mode binary data

4.1 Data and analysis

In order to investigate the applicability of our method, we use the proposed method to analyse 3-mode structured data from annual all-star games of Japanese professional baseball teams. The Japanese professional baseball league holds annual all-star games between players from the Central League which includes six teams (Bay Stars, Giants, Dragons, Tigers, Carp, and Swallows), and Pacific League which includes six teams
The analysis based on principal matrix decomposition

The analysis based on principal matrix decomposition 33

(Fighters, Hawks, Lions, Blue Waves, Marines, and Eagles). The vertical axis represents the Central League team (Bay Stars, Giants, Dragons, Tigers, Carp, and Swallows) of each player, the horizontal axis represents position (Catcher, First, Second, Third, Short, and Outfield) of each player in the Central League, and the depth axis represents the year.

We found $k$-principal matrixes ($k = 1, 2, 3, 4, \text{and} 5$, respectively) and the result of the squared distance between the data and principal points (SD) was as Table 2.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Squared distance between the data and $k$-principal matrixes ($k = 1, 2, 3, 4, 5$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>1</td>
</tr>
<tr>
<td>SD</td>
<td>733</td>
</tr>
</tbody>
</table>

Judging from the value of squared distance, 4-principal matrixes should be preferable, and we see the result of 4-principal matrixes in detail (i.e., the value of SD and the number of the data for each class) as Table 3, we have decided the number of principal points $k = 4$.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Squared distance between the data and $k$-principal matrixes ($k = 1, 2, 3, 4, 5$) for each class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>1</td>
</tr>
<tr>
<td>SD</td>
<td>100</td>
</tr>
<tr>
<td>Data</td>
<td>11.5</td>
</tr>
</tbody>
</table>

This 3-mode binary data is analysed using 4-principal matrixes; the principal matrixes are decomposed into two vertical vectors and two horizontal vectors. For searching principal matrix, we applied the binary $k$-means approach (Yamashita and Suzuki, 2014). The principal matrixes divide the data into four clusters according to year and show representative binary matrixes of the 3-mode data. Furthermore, the principal matrixes are decomposed into two vertical vectors that show the score of each team and two horizontal vectors that show the score of each position.

4.2 Analysis results and interpretation

Subsection 4.2 contains the results of the principal matrixes (Table 4), the decomposed vertical and horizontal vectors (Tables 5 and 6), and the summarisation of the results (Table 6) are shown.

Table 4 shows that each cluster includes the principal matrixes of the closed years. This result suggests that trends showing the relationships between a team and its player positions gradually change year by year. Moreover, we see many suggestions from the given principal matrixes. For example, all elements of the catcher position of Swallows are 1, and the number of element 1’s for the Carp tends to be increasing.

Table 5 shows two estimated vertical vectors. We see that the vertical vector corresponding to matrixes (1) and (2) (i.e., the left vector) has high score elements or the Tigers and Swallows, and the vertical vector corresponding to matrixes (3) and (4) (i.e., the right vector) has high score elements of the Giants and Carp. Table 6 shows
two estimated horizontal vectors. These estimators indicate that the horizontal vector corresponding to (1) and (3) (i.e., the upper vector) has high score elements for Cather and Outfield, and the horizontal vector corresponding to (2) and (4) (i.e., the lower vector) has high score elements for the Cather and Outfield, and First. These results are summarised in Table 7, and show the combinations of the high scoring teams from the vertical vectors and the high scoring positions from the horizontal vectors.

**Table 4** Principal matrixes [(1) to (4)] for the 3-mode data

<table>
<thead>
<tr>
<th>Year</th>
<th>Catch</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>Short</th>
<th>Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ba</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Gi</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dr</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ti</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ca</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Sw</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ba</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Gi</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Dr</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ti</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ca</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sw</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ba</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Gi</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Dr</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ti</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ca</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Sw</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ba</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Gi</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Dr</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ti</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ca</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Sw</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
The analysis based on principal matrix decomposition

Table 5  Decomposed vertical vectors showing the scores of each team

<table>
<thead>
<tr>
<th>Matrix</th>
<th>(1) and (2)</th>
<th>(3) and (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ba</td>
<td>0.123</td>
<td>0.542</td>
</tr>
<tr>
<td>Gi</td>
<td>0.564</td>
<td>0.672</td>
</tr>
<tr>
<td>Dr</td>
<td>0.850</td>
<td>0.897</td>
</tr>
<tr>
<td>Ti</td>
<td>0.564</td>
<td>0.897</td>
</tr>
<tr>
<td>Ca</td>
<td>0.360</td>
<td>0.859</td>
</tr>
<tr>
<td>Sw</td>
<td>0.868</td>
<td>0.821</td>
</tr>
</tbody>
</table>

Table 6  Decomposed horizontal vectors showing the scores of each position

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Catch</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>Short</th>
<th>Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) and (3)</td>
<td>0.952</td>
<td>0.447</td>
<td>0.499</td>
<td>0.349</td>
<td>0.438</td>
<td>1.244</td>
</tr>
<tr>
<td>(2) and (4)</td>
<td>0.867</td>
<td>0.572</td>
<td>0.379</td>
<td>0.494</td>
<td>0.397</td>
<td>1.197</td>
</tr>
</tbody>
</table>

Table 7  Summarisation of the results; the high score teams of the decomposed vertical vectors and positions of the decomposed horizontal vectors for each matrix

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Team</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>Ti &amp; Sw</td>
<td>Catcher, Outfield</td>
</tr>
<tr>
<td>(2)</td>
<td>Ti &amp; Sw</td>
<td>Catcher, Outfield, and First</td>
</tr>
<tr>
<td>(3)</td>
<td>Gi &amp; Ca</td>
<td>Catcher, Outfield</td>
</tr>
<tr>
<td>(4)</td>
<td>Gi &amp; Ca</td>
<td>Catcher, Outfield, and First</td>
</tr>
</tbody>
</table>

These results show new findings regarding the statistical characteristics of the data, which cannot be found by applying only the conventional model (i.e., principal matrixes for 3-mode data). Therefore, we have shown the applicability of our method. It should be noted that when we use our method, we should decide the number of clusters considering not only the value of squared distance but also the balance of the number of data in each cluster, and the interpretation, comprehensively.

In this study, only one application has been demonstrated. This model is applicable to real data such as such as time series evaluation data of each item for each person. Thus, we should reinforce the effectiveness of our proposed model by testing a wider variety of data.

5  Conclusions

In this study, we have proposed a new analysis method for 3-mode binary data, such as time series evaluation data of each item for each person by applying the principal points for 3-way binary data (Yamashita, 2015) to 3-mode binary data and decomposing the principal matrices into small number of vectors based on vector product.

Moreover, in order to investigate the applicability of our method to real-world data, we have analysed 3-mode structured data from annual all-star game of Japanese professional baseball teams. Specifically, we analyses the data of teams and positions of players recorded in all-star games held between 2000 and 2015, using the proposed
method based on the principal matrixes for 3-mode binary data. Our results show the applicability of our method and the proposed procedure can be useful for several problems of data analysis.

References


