Analyzing quality of life survey using constrained co-clustering model for ordinal data and some dynamic implication

Margot Selosse, Julien Jacques, Christophe Biernacki, Florence Cousson-Gélie

Analyzing quality of life survey using constrained co-clustering model for ordinal data and some dynamic implication

Margot Selosse
Université de Lyon, Lyon 2, ERIC EA 3083, Lyon, France.
Julien Jacques
Université de Lyon, Lyon 2, ERIC EA 3083, Lyon, France.
Christophe Biernacki
Inria, Université de Lille, CNRS, Lille, France.
Florence Cousson-Gélie
Université Paul Valéry Montpellier 3, Université Montpellier, EPSLYON EA 4556, F34000, Montpellier, France.

Summary. The dataset which motivated this work is a psychological survey on women affected by a breast tumor. Patients replied at different moments of their treatment to questionnaires with answers on ordinal scale. The questions relate to aspects of their life called dimensions. To assist the psychologists in analyzing the results, it is useful to emphasize a structure in the dataset. The clustering method achieves that by creating groups of individuals that are depicted by a representative of the group. From a psychological position, it is also useful to observe how questions may be clustered. The simultaneous clustering of both patients and questions is called co-clustering. However, getting questions into a same group when they are not related to the same dimension does not make sense from a psychologist stance. Therefore, a constrained co-clustering has been performed to prevent questions from different dimensions from getting assembled in a same column-cluster. Then, evolution of co-clusters along time has been investigated. The method relies on a constrained Latent Block Model embedding a probability distribution for ordinal data. Parameter estimation relies on a Stochastic EM-algorithm associated to a Gibbs sampler, and the ICL-BIC criterion is used for selecting the numbers of co-clusters.

1. Introduction

Persons with cancer usually go through traumatizing hardships as chemotherapy and intense stress. The disease and its treatment has an impact on different domains of their environment as social life, or emotional state. In psychology, these domains are divided into dimensions. For example, in Table 1 the domain *quality of life* is divided into six dimensions (physical functioning, role functioning, social functioning, emotional functioning, cognitive functioning, global health evaluation). Differently, the domain *emotional state* is defined with the two dimensions anxiety and depression (Zigmond and Snaith (1983)). Other psychological dimensions have been identified as a quality of life predictor like perceived control of the illness, which corresponds to the general belief whereby evolution of the disease depends either on internal factors (action, effort, personal abilities) or on external factors (hazard, destiny) (Cousson-Gélie (2014)), or

social support which assesses perceived availability (number of people on whom the individual thinks they can count if necessary) and the degree of satisfaction relating to this support (Sarason et al. (1983)).

When psychology experts set up surveys using questionnaires, they often collect a lot of data, both in terms of number of individuals and in terms of number of variables (questions). This is generally interesting because the more data they collect, the more complete they conclusions will be. Nevertheless, just after having collected the data, a first phase of data mining is necessary. This phase of apprehension makes it possible to summarize the data, to distinguish structures internal to the data but also to detect anomalies if they exist. It also allows to better visualize the data and to have a better overall knowledge of these data. When the dataset is not tagged, it is therefore interesting to use unsupervised algorithms. The dataset which has initiated this work is a survey realized on women affected by breast cancer (Cousson-Gélie (2014)). The patients were asked to reply to various questionnaires related to distinct dimensions, the answers being of the ordinal kind with different numbers of levels (Agresti (2010)). They repeated this work at six different moments of their treatment. Therefore the resulting dataset is a set of six tables, the lines representing the patients, and the columns representing the questions. First of all, the psychologists are interested in identifying psychological profiles. Particularly, they are willing to analyze the mutual influence of the different dimensions for each profile. To help them on this task, a constrained co-clustering was performed. Co-clustering is a technique which operates simultaneous clustering of the rows and columns of a matrix (Govaert and Nadif (2014)). As a result, a co-clustering emphasizes an internal structure in the dataset, which in this case allows to detect typical psychological profiles and the groups of questions that differentiate them. The term "constrained" is used because the co-clustering was forced to keep separated the questions (columns) that did not relate to a common dimension. In a second time, the experts want to investigate the evolution of their patients answers. Indeed, they also focus on the changes in their psychological state, which is called the trajectory (Annema et al. (2017)). Realizing a co-clustering at each time the patients had to answer the questionnaires gives a better idea of the evolution of the patients on different perspectives. On a global scale, it shows how groups of persons evolve, and how replies changed along the study period. On a more precise scale, the co-clustering makes possible to analyze the behavior of a single patient, by noticing her row-clusters change over time.

The dataset exclusively contains values of the ordinal type. Unlike categorical data, ordinal data have received less attention from a clustering aspect. Therefore, confronted to such data, the practitioners often transform them into continuous data, by associating an arbitrary number to each level (Kaufman and Rousseeuw (2008); Lewis et al. (2005)) or into nominal data (Vermunt and Magidson (2005)). These choices allow to use well-known distributions but either loose the information given by the existing order among levels (when considering them as nominal) or introduce an arbitrary notion of distance between levels (when transforming them as continuous). In the CUB model Piccolo (2003), an answer is interpreted as the result of a cognitive process where the decision is intrinsically continuous but is expressed in a discrete scale of m levels. This approach interprets the choice of the respondent as a weighted combination of two components.

The first one reflects a personal feeling and is expressed by a shifted binomial random variable. The second component reflects an intrinsic uncertainty and is expressed by a uniform random variable (Iannario (2010)). However, the CUB model can not be easily used in a clustering context since a mixture of CUB is not identifiable. Several recent contributions have defined clustering algorithms specific to ordinal data (Jollois and Nadif (2009); Giordan and Diana (2011); McParland and Gormley (2011); Deldossi and Zappa (2014); Ranalli and Rocci (2016); Biernacki and Jacques (2016)). In a co-clustering context, Jacques and Biernacki (2017) defines a model-based algorithm relying on the Latent block Model (Govaert and Nadif (2014)). It embeds a recent distribution for ordinal data (BOS for Binary Ordinal Search model, Biernacki and Jacques (2016)) on an SEM-Gibbs algorithm. This model presents strong advantages through its parsimony and the significance of its parameter. Nevertheless, the weakness of this model is its inability to treat variables with different numbers of levels, which is an actual issue for this dataset.

In this work, the Latent Block Model is adapted as a constrained version, so that certain questions cannot be part of the same column-cluster. This extension solves two issues. First, it allows to force the column-clusters to be formed with questions of a same psychological dimension. Furthermore, it entitles to separate the questions that do not have the same number of levels, so that the BOS distribution can be used. This distribution is defined with two meaningful parameters (μ,π), mu indicating a position (mode), π indicating a scaling (precision), and it makes sense to compare the parameters of two samples only when the number of levels of these samples are equal. Indeed, Figure 1 shows two samples with different numbers of levels m, with the same number of observations N and with the same (μ, π). We notice that the probability distributions can not be compared because they do not have the same support even if parameters are identical. Note also that the overall shape of both distrbutions is quite different.

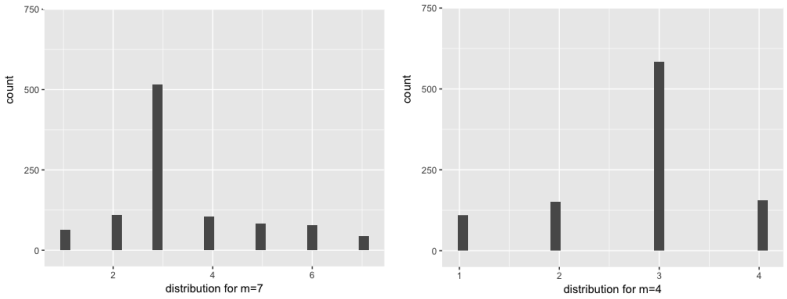

Fig. 1. Two ordinal data samples following a BOS distribution, with $N = 1000$, $\mu = 3$ and $\pi = 0.5$. On the left, $m = 7$, on the right, $m = 4$. It is easily noticed that the two probability distributions are different.

The paper is organized as follows: Section 2 presents the dataset and the notations, while Section 3 explains the statistical models that were used. At last, Section 4 describes the obtained results on the psychological dataset.

Table 1. Table of domains and dimensions that were brought up in the questionnaires.

Domains				
Quality of life (Aaronson et al. (1993))	Social Support (Sarason et al. (1983))	Specific Social Support (Pierce et al. (1997))	Emotional State (Zigmond and Snaith (1983))	Control perception (Cousson-Gélie (2014))
Dimensions				
Physical functioning, Role functioning, Emotional functioning, Cognitive functioning, Social functioning, Global health evaluation.	Satisfaction, Quantity.	Intensity, Perception of availability Conflicts.	Anxiety, Depression.	Causal attribution, Control perception, Religion control.

2. Material

2.1. Dataset

2.1.1. Inquiry population description

Several questionnaires were given to $N = 161$ women who had their first surgery for suspicious breast tumor. These patients were from 31 to 77 years old with an average age of 56.25 years (standard deviation=9.99). Most were married or lived maritally (77.0%). Near half of the patients were active professionally (49.7%) and 38.5% were retired at the moment they started the study. These 161 patients were asked to answer several questionnaires, at different moments of their treatment: one at their first surgery, and 1, 4, 7, 10, 13 months after this assessment. As a result, the patients replied 6 times to 134 questions and each answer was given on an ordinal scale (with a number of levels varying from 4 to 7). Therefore, the dataset is a set of 6 matrices of ordinal data such that the observations (rows) correspond to the patients, and the variables (columns) correspond to the questions. The dataset also contains missing values, for which we distinguish two types. The first one concerns patients that decided not to answer to the questions at a moment of their treatment. The second one concerns a few questions to which some patients exceptionally did not replied. In the first case, for each moment, the co-clustering was performed without taking into account the patients that did not want to answer at this moment. In the second case, the missing values (18 values in total) were included in the method, which handles missing data. The way of dealing with missing data is described afterward.

2.1.2. Psychological dimensions

The questionnaires that were given to the patients were precise. Actually, the conception of questionnaires is a highly-specialized work in psychology. Each questionnaire relates to domains of life, and each domain is itself divided into dimensions (e.g: MaloneBeach and Zarit (1995)). Table 1 lists the domains and the corresponding dimensions that were present in the study. In the questionnaires, most of the questions are associated to a dimension. The few questions that are not related to one of these psychological dimensions concern the treatment and disease symptoms (nausea, tiredness...).

2.2. Data representation and conventions

First of all, the dataset has been recoded so that for all the questions, the most positive answer is given the level "1". For example, for the question: *"Have you had trouble sleeping?"* with possible responses: *"Not at all." "A little." "Quite a bit." "Very much."*, the following levels number are assigned to the replies: 1 *"Not at all."*, 2 *"A little."*, 3 *"Quite a bit."*, 4 *"Very much."*, because it is perceived as more positive not to have had trouble sleeping.

Secondly, a graphical way of representing the data has been defined. Figure 2 exposes it: the women are projected on lines and the questions are projected on columns. Therefore, the cell (i, j) is the reply of patient i to question j. The shades of gray indicates how positively the person replied. For example, for the question *"Have you had trouble sleeping?"*, if the patient answers *"Not at all."*, the corresponding cell will be white, whereas a response as *"Very much."* will correspond to black cell.

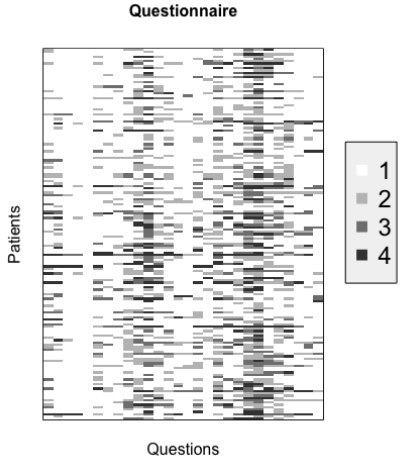

Fig. 2. Graphical representations of the patients replies. The women are in lines and the questions are in columns. A cell is the answer of a person to a question. The darker the cell is, the more pessimistic the patient responded.

2.3. Notations

First of all, an ordinal variable x with m levels $\{l_1, ..., l_m\}$ is a categorical variable whose levels order is significant. The order between the levels is quoted by the sequel "$<$": $l_1 < ... < l_m$. Furthermore, for simplicity the levels are numbered $\{1, ..., m\}$ according to their order. Following this notation, an ordinal variable x is an element of $\{1, ..., m\}$.

The representation of the questionnaires responses at a given time is now detailed. The questions are separated according to two criteria: the number of levels m and the dimension it is related to. Indeed, the variables that do not have the same number of levels and the variables that are not related to the same dimension are pulled apart. This results in a matrix split up in D tables, such that the d^{th} table is a $N \times J_d$ matrix written $\boldsymbol{x}^d$, where N is the number of observations (patients here) and J_d the number

of questions in the d-th table. The matrix $\boldsymbol{x}^d$ is made of ordinal data with number of levels m_d. Figure 3 illustrates these notations.

$$\boldsymbol{x} = \left[\left[\begin{array}{c} \\ \boldsymbol{x}^1 \\ \\ \end{array}\right] \ ... \ \left[\begin{array}{c} \\ \boldsymbol{x}^D \\ \\ \end{array}\right]\right], \text{ with } \boldsymbol{x}^d = (x_{ij}^d)_{i=1,...,N;\ j=1,...,J_d}$$

Fig. 3. Representation of the patients and questions at a given time. Questions related to different dimensions or with different number of levels m are separated.

The goal of the co-clustering is to partition the rows of $\boldsymbol{x}$ into G row-clusters, and the column of each submatrix $\boldsymbol{x}^d$ into H_d column-clusters.

The dataset contains missing data. The whole dataset will be written $\boldsymbol{x} = (\check{\boldsymbol{x}}, \hat{\boldsymbol{x}})$, $\check{\boldsymbol{x}}$ being the observed data, and $\hat{\boldsymbol{x}}$ being the missing data. Consequently a cell of $\boldsymbol{x}$ will be annotated as follows: $\check{x}_{ij}$, whether x_{ij} is observed, $\hat{x}_{ij}$ otherwise.

Finally, the bounds for the indices $i, j, g, h : 1 \leq i \leq N$, $1 \leq j \leq J$, $1 \leq g \leq G$, $1 \leq h \leq H$ (or $1 \leq h \leq H_d$ from Section 3.2.2) will not be written explicitly. For example, the matrix $\boldsymbol{x} = (x_{ij})_{1 \leq i \leq N, 1 \leq j \leq J}$ will be written $(x_{ij})_{i,j}$. Furthermore, the sums and the products relating to rows, columns, row-clusters and column-clusters will be subscripted respectively by the letters i, j, g, and h. So the sums and products will be written $\sum_i$, $\sum_j$, $\sum_g$ and $\sum_h$ and $\prod_i$, $\prod_j$, $\prod_g$ and $\prod_h$.

3. Methods

3.1. The BOS distribution for ordinal data

The Binary Ordinal Search (BOS) model (Biernacki and Jacques (2016)) is a probability distribution for ordinal data parametrized by a position parameter $\mu \in \{1, ..., m\}$ and a precision parameter $\pi \in [0, 1]$. This distribution rises from the uniform distribution when $\pi = 0$ to a more peaked distribution around the mode μ when π grows, and reaches a Dirac distribution at the mode μ when $\pi = 1$. Figure 4 illustrates the shape of the BOS distribution with different values of μ and π. It is shown in Biernacki and Jacques (2016) that the BOS distribution is a polynomial function of π with degree $m-1$, whose coefficients depend on the position parameter μ. For a univariate ordinal variable, the path in the stochastic binary search can be seen as a latent variable. Therefore, maximum likelihood estimation of model parameters can be simply performed using an EM algorithm (Dempster et al. (1977)).

3.2. Latent Block Model extension

In this section, the Constrained Latent Block Model is described but first the Latent Block Model concepts are recalled (Govaert and Nadif (2014)).

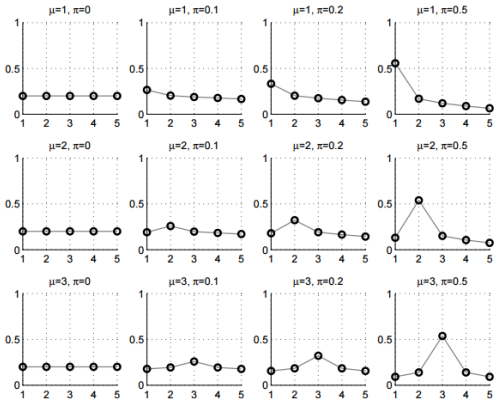

Fig. 4. BOS distribution $p(x;\mu,\pi)$: shape for $m = 5$ and for different values of μ and π.

3.2.1. Latent Block Model

Let $\boldsymbol{x} = (x_{ij})_{i,j}$ be a data matrix. It is assumed that there exists a partition $\boldsymbol{v} = (v_{ig})_{i,g}$ and a partition $\boldsymbol{w} = (w_{jh})_{j,h}$ such that each element x_{ij} is generated under a parameterized probability density function $f(x_{ij};\alpha_{gh})$ where g denotes the cluster of row i while h denotes the cluster of column j. The univariate random variables x_{ij} are assumed to be conditionally independent given the row and column partitions $\boldsymbol{v}$ and $\boldsymbol{w}$. Therefore, the conditional probability density function of $\boldsymbol{x}$ given $\boldsymbol{v}$ and $\boldsymbol{w}$ can be expressed in the following form:

$$p(\boldsymbol{x}|\boldsymbol{v},\boldsymbol{w};\boldsymbol{\alpha}) = \prod_{i,j,g,h} f(x_{ij};\alpha_{gh})^{v_{ig}w_{jh}},$$

considering that $v_{ig} = 1$ if i belongs to cluster g, whereas $v_{ig} = 0$ otherwise, and that $w_{jh} = 1$ when j belongs to cluster h, but $w_{jh} = 0$ otherwise.

Different univariate distributions can be used regarding the type of data (e.g: Gaussian, Bernoulli, Poisson...). In the present case, the BOS distribution is chosen. The label of row i is called v_i and belongs to $\{1,...,G\}$. Similarly, the label for column j is called w_j and belongs to $\{1,...,H\}$. They are latent variables, and as usual in the latent variables theory, they are assumed to be independent (Everitt (1984)). So we have $p(\boldsymbol{v},\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\rho}) = p(\boldsymbol{v};\boldsymbol{\gamma})p(\boldsymbol{w};\boldsymbol{\rho})$ with:

$$p(\boldsymbol{v};\boldsymbol{\gamma}) = \prod_i p(v_i;\boldsymbol{\gamma}) = \prod_{i,g} \gamma_g^{v_{ig}} \text{ and } p(\boldsymbol{w};\boldsymbol{\rho}) = \prod_j p(w_j;\boldsymbol{\rho}) = \prod_{j,h} \rho_h^{w_{jh}},$$

knowing that $\gamma_g = p(v_{ig} = 1)$ with $g \in \{1,...,G\}$ and $\rho_h = p(w_{jh} = 1)$ with $h \in \{1,...,H\}$. This implies that, for all i, the distribution of v_i is the multinomial distribution $\mathcal{M}(\gamma_1,...,\gamma_G)$ and does not depend on i. In a similar way, for all j, the distribution of w_j is the multinomial distribution $\mathcal{M}(\rho_1,...,\rho_H)$ and does not depend on j. From these considerations, the parameter of the latent block model is defined as $\boldsymbol{\theta} = (\boldsymbol{\gamma},\boldsymbol{\rho},\boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\alpha_{gh})_{g,h}$, with $\alpha_{gh} = (\mu_{gh},\pi_{gh})$ being the position and precision BOS parameters of the distribution of block (g,h). Additionally, $\boldsymbol{\gamma} = (\gamma_1,...,\gamma_G)$ and $\boldsymbol{\rho} = (\rho_1,...,\rho_H)$ are the mixing proportions. Therefore, if V and W are the sets of all possible labels $\boldsymbol{v}$ and $\boldsymbol{w}$, the probability density function $p(\boldsymbol{x};\boldsymbol{\theta})$ of $\boldsymbol{x}$ can be written:

$$p(\boldsymbol{x};\boldsymbol{\theta}) \;=\; \sum_{(v,w)\in V\times W} \prod_{ig}\gamma_g^{v_{ig}} \prod_{jh}\rho_h^{w_{jh}} \prod_{i,j,g,h} f(x_{ij};\alpha_{gh})^{v_{ig}w_{jh}}.$$

3.2.2. Constrained Latent Block Model

In this section, the Latent Block Model is extended as in Robert (2017) so that the questions from different dimensions are kept separated, as well as the questions with different number of levels. In this co-clustering framework, $\boldsymbol{x}$ is a $N \times (J_1 + ... + J_D)$ matrix, and it is seen as matrices $\boldsymbol{x}^1, ..., \boldsymbol{x}^D$ stored side by side as explained in Section 2.3. It is supposed that there is a row-partition $\boldsymbol{v}$, and that for all $d \in \{1, ..., D\}$, there exists a column-partition $\boldsymbol{w}^d$ such that each element x_{ij}^d is generated under a parameterized probability density function $f(x_{ij}^d; \alpha_{gh})$. Here, h denotes the cluster of column j, with $j \in \{1, ..., J_d\}$ and $h \in \{1, ..., H_d\}$. The univariate random variables x_{ij}^d are assumed to be conditionally independent given the row and column partitions $\boldsymbol{v}$ and $\boldsymbol{w^d}$. Therefore, the conditional probability density function of $\boldsymbol{x}$ given $\boldsymbol{v}$ and $\boldsymbol{w} = (\boldsymbol{w}^d)_{d\in\{1,...,D\}}$ can be written in the following form:

$$p(\boldsymbol{x}|\boldsymbol{v},\boldsymbol{w};\boldsymbol{\alpha}) \;=\; \prod_{i,j,d} f(x_{ij}^d;\alpha_{v_i w_j^d}) \;=\; \prod_{d,i,j,g,h} f(x_{ij}^d;\alpha_{gh})^{v_{ig}w_{jh}^d},$$

knowing that: $\forall d \in \{1, ..., D\},\ w_{jh}^d = 1$ when j belongs to cluster h, but $w_{jh}^d = 0$ otherwise.

The labels $v_1, ..., v_N, (w_1^d, ..., w_{J_d}^d)_{d\in\{1,...,D\}}$ are latent variables assumed to be independent: $p(\boldsymbol{v},\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\rho}) \;=\; p(\boldsymbol{v};\boldsymbol{\gamma})\prod_d p(\boldsymbol{w}^d;\boldsymbol{\rho}^d)$ with:

$$p(\boldsymbol{v};\boldsymbol{\gamma}) \;=\; \prod_i p(v_i;\boldsymbol{\gamma}) \;=\; \prod_{i,g}\gamma_g^{v_{ig}} \text{ and } p(\boldsymbol{w}^d;\boldsymbol{\rho}^d) \;=\; \prod_j p(w_j^d;\boldsymbol{\rho}^d) \;=\; \prod_{j,h}\rho_h^{d\,w_{jh}^d},$$

knowing that $\rho_h^d \;=\; p(w_{jh} = 1)$ with $h \in \{1, ..., H\}$. Again, for all i, the distribution of v_i is the multinomial distribution $\mathcal{M}(\gamma_1, ..., \gamma_G)$ and does not depend on i. Evenly, for all j and for all d, the distribution of w_j^d is the multinomial distribution $\mathcal{M}(\rho_1^d, ..., \rho_{H_d}^d)$ and does not depend on j. By analogy with the classic Latent Block Model, the probability density function $p(\boldsymbol{x};\boldsymbol{\theta})$ is written:

$$p(\boldsymbol{x};\boldsymbol{\theta}) \;=\; \sum_{(v,w^1,..,w^D)\in V\times W_1\times...\times W_D} \prod_{i,g}\gamma_g^{v_{ig}} \prod_{d,j,h}\rho_h^{d\,w_{jh}^d} \prod_{i,j,g,d,h} f(x_{ij}^d;\alpha_{gh})^{v_{ig}w_{jh}^d}.$$

3.3. Model inference with an SEM-Gibbs algorithm

This section details the model inference in the case of the constrained latent block model. The aim is to estimate $\boldsymbol{\theta}$ by maximizing the observed log-likelihood $l(\boldsymbol{\theta};\check{\boldsymbol{x}}) = \sum_{\hat{x}} \log p(\boldsymbol{x};\boldsymbol{\theta})$. In a co-clustering context, the EM algorithm is not computationally feasible (see Govaert and Nadif (2014)). Indeed, the E step requires the calculation of the joint conditional probability of the missing labels $p(v_{ig} = 1, w_{jh} = 1|\boldsymbol{x};\boldsymbol{\theta}^{(q)})$ for $1 \le i \le N,\ 1 \le g \le G,\ 1 \le d \le D,\ 1 \le j \le J_d,\ 1 \le h \le H_d$, with $\boldsymbol{\theta}^{(q)}$ the current value of the parameter. Therefore, this step implies to compute $N\times G\times(J_1\times H_1+...+J_D\times H_D)$ terms that cannot be factorized as for a standard mixture, due to the dependence of the row and

column labels conditionally to the observations. There exists several alternatives to the EM algorithm like variational EM algorithm, the SEM-Gibbs algorithm, or Bayesian inference (Govaert and Nadif (2014)). The SEM-Gibbs is known to avoid spurious solutions (Keribin et al. (2010)), this is why it is used in this paper.

3.3.1. SEM-Gibbs algorithm

Starting from an initial value for the parameter $\boldsymbol{\theta}^{(0)}$, the q^{th} iteration of the algorithm is composed of two steps.

SE-step The SE-step consists in simulating the latent variables according to their joint conditional probability by a Gibbs sampling. Therefore, it repeats, for a given number of iterations, the generation of the row partitions conditionally on the column partitions and the generation of the column partitions conditionally on the row partitions. The generation of the row partitions $v_{ig}^{(q+1)} \mid \boldsymbol{x}, \boldsymbol{w}^{(q)}$ is done according to:

$$p(v_{ig}^{(q+1)} = 1 \mid \boldsymbol{x}^d, \boldsymbol{w}^{(q)}; \boldsymbol{\theta}^{(q)}) \propto \gamma_g^{(q)} \times \prod_d t_g^d(\boldsymbol{x}_{i.}^d | \boldsymbol{w}^{d^{(q)}}; \boldsymbol{\alpha}^{(q)}),$$

where $t_g^d(\boldsymbol{x}_{i.}^d | \boldsymbol{w}^{d(q)}; \boldsymbol{\alpha}^{(q)}) = \prod_{j,h} f(x_{ij}^d; \mu_{gh}^d, \pi_{gh}^d)^{w_{jh}^{d\ (q)}}$ with $\boldsymbol{x}_{i.}^d = (x_{ij}^d)_j$. The generation of the column partitions $w_{jh}^d \mid \boldsymbol{x}, \boldsymbol{v}^{(q+1)}$ for the d^{th} table $\boldsymbol{x}^d$ ($d \in \{1, ..., D\}$) is done according to:

$$p(w_{jh}^d = 1 \mid \boldsymbol{x}^d, \boldsymbol{v}^{(q+1)}; \boldsymbol{\theta}^{(q)}) \propto \rho_h^{d\,(q)} \times s_h^d(\boldsymbol{x}_{.j}^d | \boldsymbol{v}^{(q+1)}; \boldsymbol{\alpha}^{(q)})$$

where $s_h^d(\boldsymbol{x}_{.j}^d | \boldsymbol{v}^{(q+1)}; \boldsymbol{\alpha}^{(q)}) = \prod_{i,g} f(x_{ij}^d; \mu_{gh}^d, \pi_{gh}^d)^{v_{ig}^{(q+1)}}$ with $\boldsymbol{x}_{.j}^d = (x_{ij}^d)_i$.

M-step The M-step consists in maximizing the completed log-likelihood by updating the co-clusters parameters according to the results of the last SE step. It relies on the EM algorithm used in Biernacki and Jacques (2016) for the estimation of the BOS distribution on each block.

3.3.2. Imputation of missing values

The SEM-algorithm is able to take into account the missing data and to estimate them. It is assumed that the whole missing process is Missing At Random (see Little and Rubin (1986)). First, the notation of $\boldsymbol{x}$ becomes $\boldsymbol{x}^{(q)}$ since the missing variables are going to be imputed. Then, a third step is added to the SE-step. For all $d \in \{1, .., D\}$, it generates the missing data $\hat{x}_{ij}^{d^{(q+1)}} | \check{\boldsymbol{x}}^d, \boldsymbol{v}^{(q+1)}, \boldsymbol{w}^{d^{(q+1)}}$ as follows:

$$p(\hat{x}_{ij}^{d^{(q+1)}} | \check{\boldsymbol{x}}^d, \boldsymbol{v}^{(q+1)}, \boldsymbol{w}^{d^{(q+1)}}; \boldsymbol{\theta}^{(q)}) = \prod_{g,h} p(\hat{x}_{ij}^{d^{(q+1)}}; \mu_{gh}^{d\ (q)}, \pi_{gh}^{d\ (q)})^{v_{ig}^{(q+1)} w_{gh}{}^{d^{(q+1)}}}.$$

3.3.3. Estimation of partitions and model parameter

The SEM-algorithm repeats several times the aforementioned steps. The first iterations are called the burn-in period, which means the parameters are not stable yet. Consequently, the iterations that occurred after this burn-in period are taken into account, they are called the sample distribution. The final estimation of the position parameter μ_{gh} is the mode of the sampling distribution. The final estimation of the continuous parameters $(\pi_{gh}^d, \boldsymbol{\gamma}_g, \boldsymbol{\rho}_h^d)_d$ is the median of the sample distribution. It corresponds to a final estimation of $\boldsymbol{\theta}$ that is called $\hat{\boldsymbol{\theta}}$. Then, a sample of $(\hat{\boldsymbol{x}}, \boldsymbol{v}, (\boldsymbol{w}^d)_d)$ is generated by a SE-step with $\boldsymbol{\theta}$ fixed to $\hat{\boldsymbol{\theta}}$. The final partitions $(\hat{\boldsymbol{v}}, \hat{\boldsymbol{w}})$ and the missing observation $\hat{\boldsymbol{x}}$ are estimated by the mode of their sample distribution.

3.4. Model Selection

To select the number of clusters, G in rows and $H_1, ..., H_D$ in columns, a model selection criterion must be used. The most classical ones, like BIC (Schwarz (1978)) rely on penalizing the maximum log-likelihood value $l(\hat{\boldsymbol{\theta}}; \boldsymbol{x})$. However, due to the dependency of the observed data, this value is not available in a co-clustering context.

Alternatively, an approximation of the ICL information criterion (Biernacki et al. (2000)), called here ICL-BIC, can be invoked since allowing to overcome the previous problem due to the dependency structure in $\check{\mathbf{x}}$. The key point is that this latter vanishes since ICL relies on the complete latent block information $(\mathbf{v}, \mathbf{w})$, instead of integrating on it as it is the case in BIC. In particular, Keribin et al. (2015) detailed how to express ICL-BIC for the general case of categorical data and Jacques and Biernacki (2017) for the specific case of ordinal data using the BOS model. In the present work, the ICL-BIC is therefore adapted for the constrained latent block model:

$$\begin{aligned}\text{ICL-BIC}(G, H_1, ..., H_D) &= \log p(\check{\boldsymbol{x}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{w}}^1, \ldots, \hat{\boldsymbol{w}}^D; \hat{\boldsymbol{\theta}}) \\ &\quad - \frac{G-1}{2} \log N - \sum_d \frac{H_d - 1}{2} \log J_d - \sum_d \frac{G \times H_d}{2} \log(N \times J_d),\end{aligned}$$

where $\hat{\boldsymbol{v}}, \hat{\boldsymbol{w}}^1 ..., \hat{\boldsymbol{w}}^D$ are the row and column partitions discovered by the SEM-algorithm, and $\hat{\boldsymbol{\theta}}$ is the corresponding estimated model parameter.

Let's note that the co-clustering has to be performed for each possible values of G and H_d, $d \in \{1, ..., D\}$, then the result with the highest ICL-BIC is retained. Let n_G be the number of candidate values for G, while n_{H_d} is the number of candidate values for H_d, $d \in \{1, ..., D\}$. Thus, the number of co-clustering processes to execute is $n_G \times n_{H_1} \times ... \times n_{H_D}$. As an example, if $D = 3$ and the user wants to try 3 values for G and for each H_d, then it would require to execute $3^4 = 81$ co-clusterings. Depending on the dataset, it might take too much time to find the best solution.

We propose the following heuristic search. Let start by computing the ICL-BIC with minimum values $(G_{min}, H_{1_{min}}, ..., H_{D_{min}})$. Then, add 1 to each number of clusters, step by step, and compute the ICL-BIC. Retain the best solution (highest ICL-BIC), and continue the same process until the ICL-BIC stops increasing.

Table 2. Co-clustering result on dimensions anxiety, depression, symptoms: estimated BOS parameters (μ_{gh}, π_{gh}) for each cluster (g, h).

	Anxiety		Depression		Symptoms	
	col. cluster 1	col. cluster 2	col. cluster 1	col. cluster 2	col. cluster 1	col. cluster 2
row cluster 1	(2,0.77)	(2,0.77)	(1,0.70)	(2,0.83)	(2,0.46)	(1,0.74)
row cluster 2	(2,0.68)	(3,0.72)	(2,0.47)	(2,0.79)	(3,0.39)	(1,0.42)
row cluster 3	(1,0.64)	(2,0.44)	(1,0.77)	(2,0.70)	(2,0.58)	(1,0.71)
row cluster 4	(1,0.67)	(2,0.47)	(1,0.79)	(2,0.71)	(1,0.80)	(1,0.93)
row cluster 5	(2,0.72)	(3,0.55)	(2,0.64)	(2,0.75)	(2,0.66)	(1,0.77)

4. Application on the survey dataset

4.1. Constrained co-clustering on different dimensions

Several constrained co-clusterings were performed on the dataset, with different dimensions and at different times. This section presents some significant results that were obtained. In the following experiments, the heuristic search described in Section 3.4 was executed with $G_{min} = 3$ and $H_{d_{min}} = 1$ to choose the numbers of row-clusters and column-clusters (G and $(H_d), d \in \{1, ..., D\}$). All the ICL-BIC values are available in the appendix. The choice for a sufficient number of iterations for the SEM-algorithm and for the burn-in period was made empirically. It was noticed that the parameters would stabilize after 150 iterations (or less). Therefore, the burn-in period was set to 400 iterations and the total number of iterations was fixed to 500. In the appendix, Figure A1 shows how some example parameters converge along iterations.

4.1.1. Anxiety, depression and symptoms.

As a first experiment, it was decided to investigate the responses that were given at time $T5$, at the end of the treatment. The questions regarding the treatment's symptoms are interesting at this moment because the patients had been going through chemotherapy for one year at this moment. A constrained co-clustering was realized by fetching the questions related to the dimensions anxiety, depression and symptoms. In this case, all the questions have a number of levels m equals to 4. Therefore, the only constraint is the separation of the questions that are from different dimensions. The execution time of this set up is about 12 seconds with an Intel Xeon E5-2620 CPU 2.00 GHz and 8Go RAM. The result of the constrained co-clustering is illustrated by Figure 5. For all the figures, clusters are read from left to right and from top to bottom. Furthermore, Table 2 details the estimated BOS parameters (μ_{gh} and π_{gh}) for $g \in \{1, ..., G\}$ and $h \in \{1, ..., H_d\}, \forall d \in \{1, ..., D\}$.

Five row-clusters are highlighted by the co-clustering results. Table 2 shows that the positions parameters of the second row-cluster $(\mu_{2h})_{d,h}$ are globally greater than (or equal to) those of the other row-clusters. It means that the second group feels more anxiety and depression, and senses more intensively the disease symptoms than the other ones. It is also noticed that the fourth row-cluster is less inclined to anxiety and depression and suffers less from the symptoms than the others groups: indeed, parameters $(\mu_{4h})_{d,h}$ are globally the lowest. Furthermore, the precision parameters $(\pi_{4h})_{d,h}$ are quite high for this row-cluster, which means that the answers do not disperse much around the position $(\mu_{4h})_{d,h}$. By observing these two groups results, one could tell that the sense

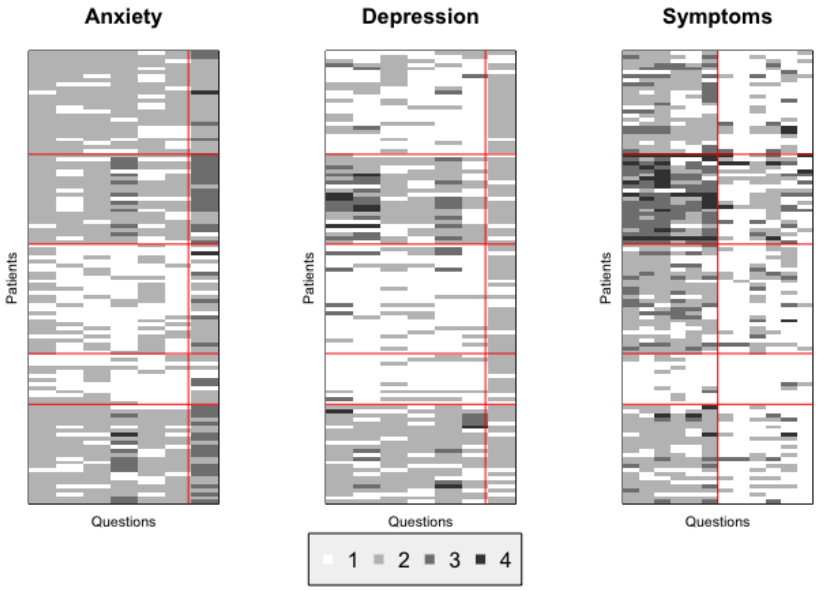

Fig. 5. Results from constrained co-clustering on dimensions anxiety, depression and symptoms.

of symptoms is closely associated to signs of anxiety and depression, which is a pretty logical and intuitive result. Yet, the first, the third and the fifth groups bring more information. They are effectively very similar about how much they suffer the disease symptoms dimensions. However, they differ a lot in the first column cluster of anxiety, and in the first column cluster of depression. It means than even if a link between symptoms, anxiety and depression can be deduced from the first observations, it is not totally confirmed when people do not sense the symptoms at the extremes (very much or not much). Moreover the column-clusters offer interesting result as well: there is a clear separation among the symptoms. By examining the questions in each cluster, it turns out that questions in the first cluster exclusively deal with pain and fatigue, while the second cluster deals with other symptoms such as nausea or loss of appetite. The co-clustering therefore detected two sub-dimensions for the symptoms dimension. What's more there is a big difference on how the patients sense these two clusters: it is easily noticed that the position parameters $(\mu_{g1})_{(symptoms)}$ are globally higher than $(\mu_{g2})_{(symptoms)}$. Therefore all the patients in general suffer more from pain and fatigue than the other symptoms.

4.1.2. Social support: satisfaction, availability, intensity and conflicts.

As a second experiment, questions related to the social support were used. The responses come from the fourth moment of the experiment: it is in the middle of the treatment for the patients, so they have gone through a lot, but know they have to keep on going for a few months. Their perception of their social support is therefore interesting at

Table 3. Co-clustering result on social support dimensions: estimated BOS parameters (μ_{gh}, π_{gh}) for each cluster (g, h).

	Satisfaction	Availability		Intensity		Conflicts		
	col. cluster 1	col. cluster 1	col. cluster 2	col. cluster 1	col. cluster 2	col. cluster 1	col. cluster 2	col. cluster 3
row cluster 1	(2,0.90)	(1,0.72)	(1,0.96)	(3,0.48)	(1,0.59)	(4,0.80)	(3,0.24)	(1,0.62)
row cluster 2	(3,0.87)	(1,0.64)	(1,0.50)	(3,0.46)	(2,0.48)	(4,0.47)	(3,0.42)	(1,0.49)
row cluster 3	(1,0.73)	(2,0.72)	(1,0.86)	(3,0.52)	(2,0.63)	(4,0.59)	(3,0.51)	(1,0.44)
row cluster 4	(2,0.79)	(1,0.61)	(2,0.64)	(3,0.31)	(2,0.50)	(3,0.27)	(3,0.32)	(1,0.44)
row cluster 5	(1,0.93)	(1,0.78)	(1,0.91)	(3,0.27)	(1,0.68)	(4,0.71)	(3,0.18)	(1,0.63)

this moment. This aspect includes questions of four dimensions: the satisfaction (with number of levels $m = 6$), the perception of availability, the intensity and the conflicts (with a number of levels $m = 4$). The questions which relate to the same dimension have the same number of levels. Again, the only constraint is the separation of the questions that are not from the same dimensions. The result of the constrained co-clustering is illustrated by Figure 6. Furthermore, Table 3 details the estimated BOS parameters (μ and π) for each co-cluster.

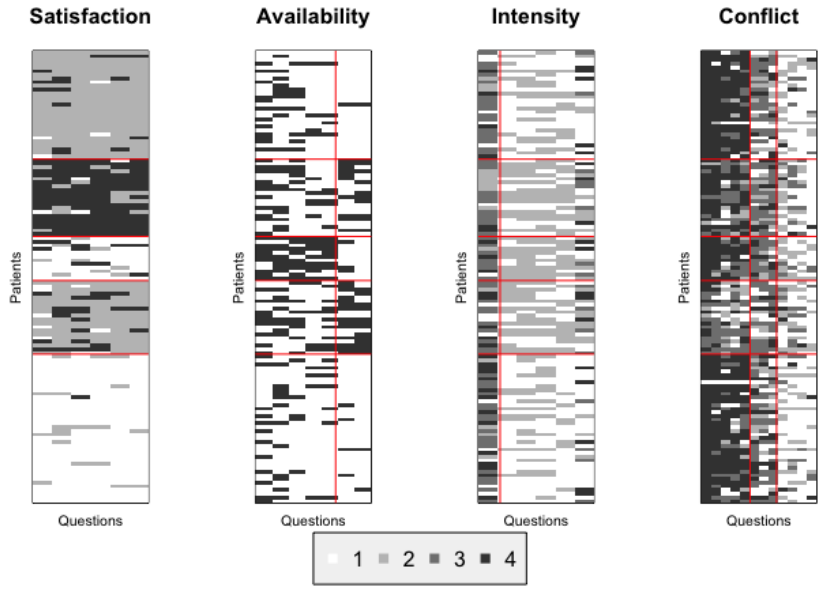

Fig. 6. Result from constrained co-clustering on dimensions related to social support.

The co-clustering detected five row-clusters. The third and the fifth ones are clearly satisfied with the social support they have. Indeed, their position parameters $\mu_{31(satisfaction)}$ and $\mu_{51(satisfaction)}$ are equal to 1. Furthermore, the precisions $\pi_{31(satisfaction)}$ and $\pi_{51(satisfaction)}$ are really high, which means that most of the patients effectively replied in the most positive way to the questions regarding their satisfaction. In contrast, the women in the first group are quite dissatisfied by their social support compared to the other ones. An other result is that the third group, which is one of the most satisfied, has one of the worst perception of availability from their close family and friends

$(\mu_{31(availability)} \geq \mu_{g1(availability)})$. Furthermore, it is also interesting to observe the column-clusters that were detected by the co-clustering for the conflicts dimension. The first group of questions is about the efforts the patient has to make not to enter in conflict with their close ones. The second group gathered questions about changes in the relationship, whereas the last cluster concerns the sentiments of anger towards their close family and friends.

4.1.3. Symptoms at different times.
In this experiment, the questions related to symptoms were selected for different moments (at time $T0$, $T2$ and $T5$). The constraint is therefore not to separate the questions from different dimensions, but to separate the questions that are not from the same time. The point of performing a co-clustering on such a dataset is that the row-clusters gather people that had a similar evolution regarding this dimension. Furthermore, the column-clusters give information about how the patients symptoms globally worsens (or get better) throughout the treatment. BOS parameters for this experiment are available in Table 4, and Figure 7 illustrates the results.

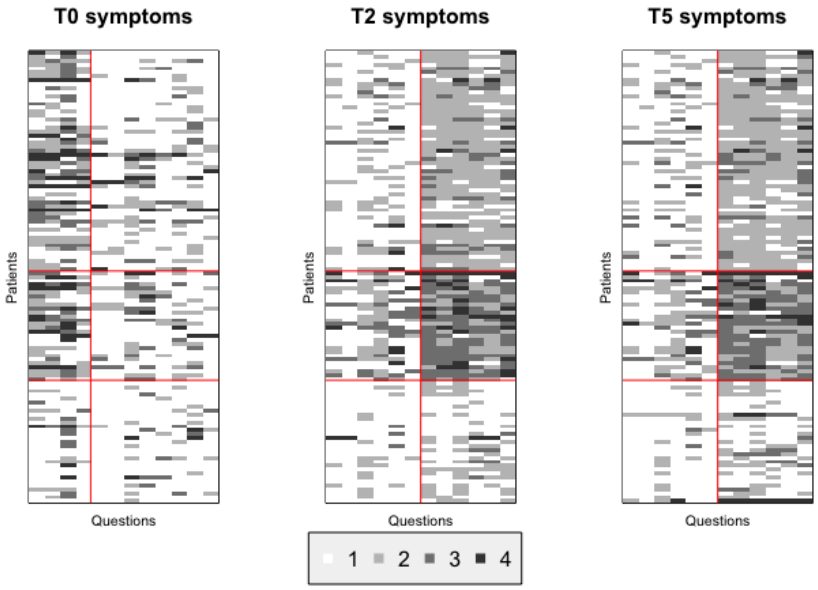

Fig. 7. Co-clustering results with questions related to symptoms, at three different times.

The co-clustering emphasizes three row-clusters. The third one gathers people that felt less the disease symptoms than the others: the position parameters $(\mu_{3h})_{d,h}$ are all equal to 1. What's more, the precision parameters $(\pi_{3h})_{d,h}$ are pretty high, which implies that the responses do not spread a lot around the value 1. It is also interesting to investigate how the column-clusters evolve. To begin, for each time, the symptoms are separated into two column-clusters: systematically, the first one is globally worse than the second one, because $(\mu_{g1})_{(T0,T2,T5)} \geq (\mu_{g2})_{(T0,T2,T5)}, \forall g \in \{1, ..., G\}$. It is observed

Table 4. Co-clustering result on symptoms dimension, at three different times: estimated BOS parameters (μ_{gh}, π_{gh}) for each co-cluster (g, h).

	T0 Symptoms		T2 Symptoms		T25 Symptoms	
	col. cluster 1	col. cluster 2	col. cluster 1	col. cluster 2	col. cluster 1	col. cluster 2
row cluster 1	(2,0.20)	(1,0.67)	(2,0.62)	(1,0.72)	(2,0.64)	(1,0.74)
row cluster 2	(2,0.09)	(1,0.62)	(3,0.43)	(1,0.40)	(3,0.42)	(1,0.46)
row cluster 3	(1,0.66)	(1,0.84)	(1,0.58)	(1,0.85)	(1,0.54)	(1,0.84)

that at time $T0$ there is less symptoms in the column-cluster 1 than in column-cluster 2, whereas they are equally shared at times $T2$ and $T5$.

4.2. Handling of the dynamical aspect of the data

The patients answered to the same questionnaires at 6 different moments of their treatment and there is a clear interest in the evolution of the responses. Defining a model to study this evolution is essential, but it is not the purpose of this paper, and it will be the matter of another work. However, here we focus on providing a tool that allows the psychologists to visualize the row-clusters evolution, without defining a mathematical reasoning. To handle this perspective, visualization were created so that the psychologists could have a first impression of the evolution, with the `Javascript` library `D3js`. First of all, the dimensions of questionnaire EORTC QLQ-C30 were selected. Then, a co-clustering was performed, similarly to Section 4.1, for each time $T0$, $T1$, $T2$, $T3$, $T4$, $T5$. The visualization represents the row-clusters on the ordinate axis, and the time line on the abscissa axis: Figure 8 illustrates the home page of a visualization that was created with dimensions dealing with quality of life and emotional state. If the expert wishes to observe the evolution of a single patient, they can click on the list of patients on the right to see the row-clusters it belongs to through time, like in Figure 9. Moreover, if the expert wants to know the co-cluster BOS parameters, they click on the row-clusters, and is able to read the (μ, π) of the corresponding co-clusters, as in Figure 10.

These visualizations showed that the patients globally got stabler with time. Indeed, it is noticed that whereas a lot of patients changed of row-clusters at the three first moments $T0, T1, T2$, these transitions get rarer after time $T2$.

5. Conclusion

In this paper, a constrained co-clustering algorithm is proposed to analyze psychological questionnaires given to women affected by breast cancer. This dataset has a lot of specificities, which makes the use of classical techniques difficult without changing the information. First, it is made of questionnaires with answers on an ordinal scale. Furthermore it included a temporal aspect because the patients answered 6 times to these questionnaires. Then, the questions are assimilated to psychological dimensions, which can not be ignored. Finally, just like a lot of real dataset, this one contains some missing values.

To adapt to the particularities of the survey, an extension of the latent block model is defined, and the parsimonious BOS distribution for ordinal data is employed. What's more, model inference is performed with an SEM-Gibbs algorithm, which allows to take missing values into account. An R package called *ordinalClust* with a full implementation

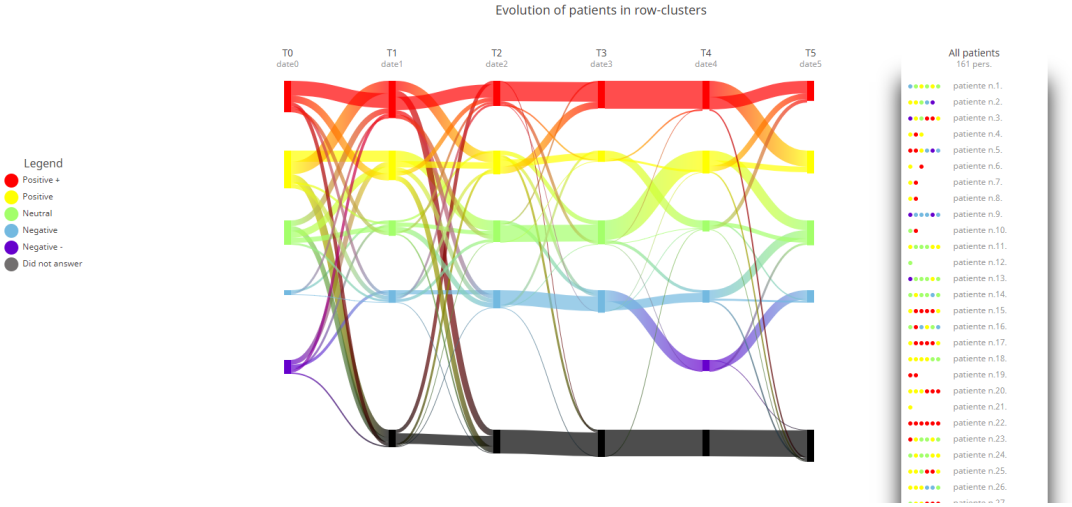

Fig. 8. Home-page: the row-clusters are represented on the ordinate axis, and the time line is on the abscissa axis.

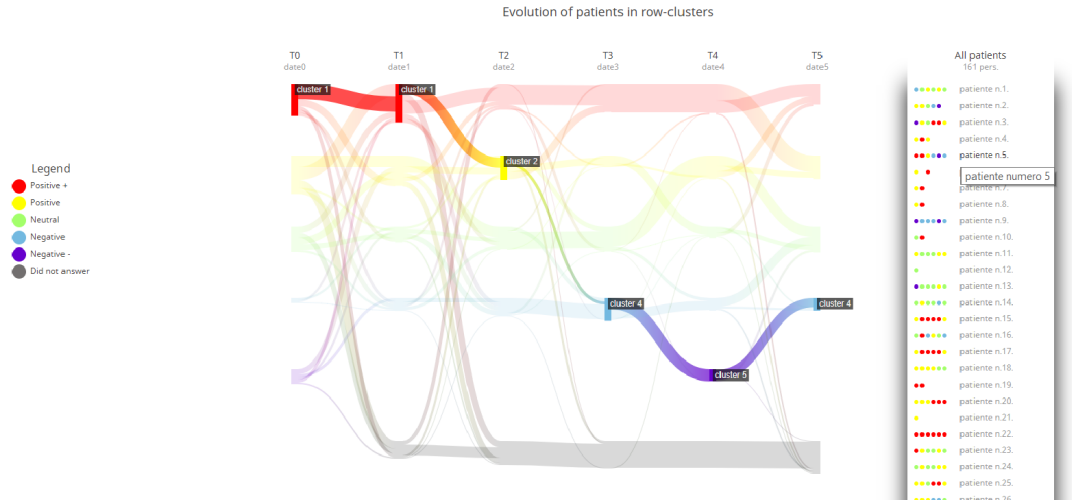

Fig. 9. When user clicks on a patient on the right list, they can observe the psychological trajectory of this patient.

of this work is available on CRAN, and will be soon available on the CRAN. Finally, visualizations have been created to help the psychologists to observe the evolution of their patients.

The results were particularly satisfying to the psychologist. The proposed technique provides a parsimonious way to cluster the patients by gathering the questions in a small number of groups, and the BOS parameters meaningfulness allows to easily interpret the resulting co-cluster. Furthermore, the constrained co-clustering overtakes two matters: the different numbers of levels for the questions, and the fact that the questions refer to

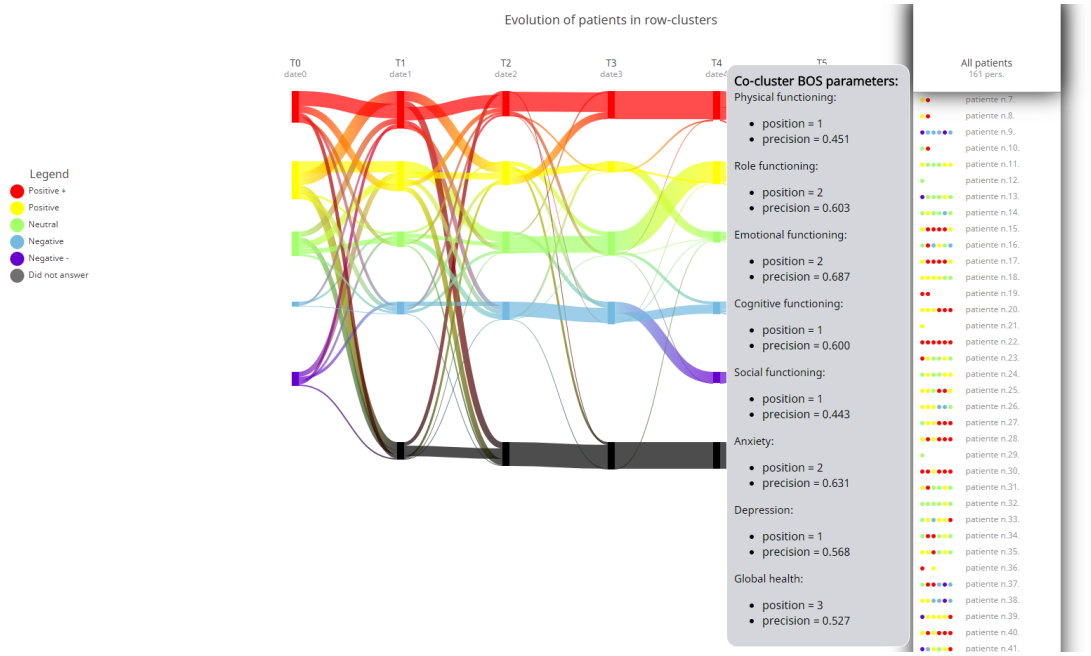

Fig. 10. When user clicks on a row-cluster they are able to see the BOS parameters of all corresponding co-clusters.

different psychological dimensions.

With the proposed approach, features with different number of levels can be treated in the co-clustering execution, but they are not allowed to be part of the same column-cluster. As a future work, it would be interesting to make that possible, and ideally, to perform co-clusterings with data of different kinds (continuous, functional...). At last, although the dynamical aspect of the data has been approached with visualizations it would be advantageous to define a mathematical model on this aspect.

6. Acknowledgement

We thank INCa (Institut National du Cancer), Institut Lilly, Institut Bergonié, Centre Régional de Lutte Contre le Cancer de Bordeaux (C. Tunon de Lara, J. Delefortrie, A. Rousovoal, A. Avril, E. Bussières) and Laboratoire de Psychologie de l'Université de Bordeaux (C. Quintrinc and S. de Castro-Lévèque).

References

Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., Haes, J. C. J. M. d., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, P. B., Schraub, S., Sneeuw, K., Sullivan, M. and Takeda, F. (1993) The european organization for research and treatment of cancer qlq-c30: A quality-of-life instrument for use in international clinical trials in oncology. *JNCI: Journal of the National Cancer Institute*, **85**, 365–376. URL: `http://dx.doi.org/10.1093/jnci/85.5.365`.

Agresti, A. (2010) *Analysis of Ordinal Categorical Data, 2nd Ed.* John Wiley & Sons, Inc.

Annema, C., Roodbol, P. F., Van den Heuvel, E. R., Metselaar, H. J., Van Hoek, B., Porte, R. J. and Ranchor, A. V. (2017) Trajectories of anxiety and depression in liver transplant candidates during the waiting-list period. *British Journal of Health Psychology*, **22**, 481–501.

Biernacki, C., Celeux, G. and Govaert, G. (2000) Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Mach. Intell.*, **22**, 719–725. URL: `http://dx.doi.org/10.1109/34.865189`.

Biernacki, C. and Jacques, J. (2016) Model-Based Clustering of Multivariate Ordinal Data Relying on a Stochastic Binary Search Algorithm. *Statistics and Computing*, **26**, 929–943.

Cousson-Gélie, F. (2014) Évolution du contrôle religieux la premire année suivant lannonce dun cancer du sein: quels liens avec les stratégies de coping, lanxiété, la dépression et la qualité de vie? *Psychologie Franaise*, **59**, 331 – 341. URL: `http://www.sciencedirect.com/science/article/pii/S0033298412000027`.

Deldossi, L. and Zappa, D. (2014) Evaluating r & r of ordinal classifications with cub model. *quaderni di statistica*.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, series B*, **39**, 1–38.

Everitt, B. S. (1984) *Introduction to Latent Variable Models*. Chapman and Hall.

Giordan, M. and Diana, G. (2011) A clustering method for categorical ordinal data. *Communications in Statistics - Theory and Methods*, **40**, 1315–1334.

Govaert, G. and Nadif, M. (2014) *Co-Clustering*. Computing Engineering series. ISTE-Wiley.

Iannario, M. (2010) On the identifiability of a mixture model for ordinal data. *METRON*, **68**, 87–94. URL: `https://doi.org/10.1007/BF03263526`.

Jacques, J. and Biernacki, C. (2017) Model-Based Co-clustering for Ordinal Data. Working paper or preprint.

Jollois, F.-X. and Nadif, M. (2009) Classification de données ordinales : modèles et algorithmes. In *41èmes Journées de Statistique, SFdS, Bordeaux*. Bordeaux, France, France.

Kaufman, L. and Rousseeuw, P. J. (2008) *Introduction*, 1–67. John Wiley and Sons, Inc.

Keribin, C., Brault, V., Celeux, G. and Govaert, G. (2015) Estimation and selection for the latent block model on categorical data. *Statistics and Computing*, **25**, 1201–1216.

Keribin, C., Govaert, G. and Celeux, G. (2010) Estimation d'un modèle à blocs latents par l'algorithme SEM. In *42èmes Journées de Statistique*. Marseille, France.

Lewis, S. J. G., Foltynie, T., Blackwell, A. D., Robbins, T. W., Owen, A. M. and Barker, R. A. (2005) Heterogeneity of parkinson's disease in the early clinical stages using a data driven approach. *Journal of Neurology, Neurosurgery & Psychiatry*, **76**, 343–348.

Little, R. J. A. and Rubin, D. B. (1986) *Statistical Analysis with Missing Data*. New York, NY, USA: John Wiley & Sons, Inc.

MaloneBeach, E. E. and Zarit, S. H. (1995) Dimensions of social support and social conflict as predictors of caregiver depression. *International Psychogeriatrics*, **7**, 2538.

McParland, D. and Gormley, I. C. (2011) Clustering ordinal data via latent variable models. *Berthold Lausen, Dirk Van den Poel, Alfred Ultsch (eds.). Algorithms from and for Nature and Life : Classification and Data Analysis*.

Piccolo, D. (2003) On the moments of a mixture of uniform and shifted binomial random variables. **5**.

Pierce, G. R., Sarason, I. G., Sarason, B. R., Solky-Butzel, J. A. and Nagle, L. C. (1997) Assessing the quality of personal relationships. *Journal of Social and Personal Relationships*, **14**, 339–356.

Ranalli, M. and Rocci, R. (2016) Mixture models for ordinal data: A pairwise likelihood approach. *Statistics and Computing*, **26**, 529–547. URL: `http://dx.doi.org/10.1007/s11222-014-9543-4`.

Robert, V. (2017) *Classification croisée pour l'analyse de bases de données de grandes dimensions de pharmacovigilance*. Ph.D. thesis. Thèse de doctorat dirigée par Celeux, Gilles Mathématiques appliquées Paris Saclay 2017.

Sarason, I. G., Levine, H. M., Basham, R. B. and Sarason, B. R. (1983) Assessing social support: The social support questionnaire. *Journal of Personality and Social Psychology*, 139.

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.

Vermunt, J. and Magidson, J. (2005) Latent gold 4.0 user's guide. belmont, massachusetts:statistical innovations inc.

Zigmond, A. S. and Snaith, R. P. (1983) The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, **67**, 361–370.

Table 1. ICL-BIC values for experiment with dimensions: anxiety, depression and symptoms.

iteration number	tested set	ICL-BIC value
0	**3111**	**-3182.636**
1	3211	-3202.003
	4111	-3178.63
	3121	-3135.95
	3112	**-2981.384**
2	4112	-2955.255
	3212	-3012.662
	3122	**-2931.426**
	3113	-3003.892
3	4122	-2909.457
	3222	**-2907.208**
	3132	-2937.025
	3123	-2941.77

iteration number	tested set	ICL-BIC value
4	**4222**	**-2897.574**
	3322	-2899.502
	3232	-2915.243
	3223	-2914.43
5	**5222**	**-2887.428**
	4322	-2901.103
	4232	-2902.765
	4223	-2911.143
6	6222	-2890.224
	5322	-2890.043
	5232	-2898.423
	5223	-2900.492

Table 2. ICL-BIC values for experiment with dimensions of social support.

iteration number	tested set	ICL-BIC value
0	**31111**	**-4159.044**
1	41111	-4148.478
	32111	-4169.625
	31211	-4167.269
	31121	-4109.996
	31112	**-3890.939**
2	41112	-3861.417
	32112	-3901.316
	31212	-3966.906
	31122	-3847.206
	31113	**-3792.995**
3	**41113**	**-3759.687**
	32113	-3800.504
	31213	-3793.978
	31123	-3760.164
	51113	-3803.808

iteration number	tested set	ICL-BIC value
4	51113	-3745.543
	42113	-3792.785
	41213	-3807.943
	41123	**-3708.452**
	41114	-3782.372
5	**51123**	**-3684.643**
	42123	-3723.946
	41223	-3710.339
	41133	-3814.453
	41124	-3807.466
6	61123	-3672.869
	52123	-3689.939
	51223	**-3646.392**
	51133	-3670.815
	51124	-3674.937

iteration number	tested set	ICL-BIC value
7	61223	-3654.905
	52223	-3664.249
	51323	-3660.886
	51233	-3670.111
	51224	-3665.377

Appendix

The following tables present the ICL-BIC obtained by executing the heuristic search described in Section 3.4 on the applications described in Section 4. At each iteration, the values in bold represent the highest ICL-BIC values of the iteration. The underlined values are the final chosen values for $(G, H_1, ..., H_D)$.

Figure A1 presents the evolution of some parameters through the SEM-algorithm iterations.

Table 3. ICL-BIC values for experiment with symptoms dimensions at times ($T0$, $T1$, $T2$).

iteration number	tested set	ICL-BIC value
0	**3111**	**-4479.951**
1	4111	-4473.039
	3211	-4419.312
	3121	-4278.773
	3112	**-4259.51**
2	4112	-4206.192
	3212	-4200.733
	3122	**-4021.611**
	3113	-4262.249
3	4122	-4033.372
	3222	**-3967.913**
	3132	-4012.178
	3123	-4103.417
4	4222	-3981.241
	3322	-4082.588
	3232	-4080.828
	3223	-4046.097

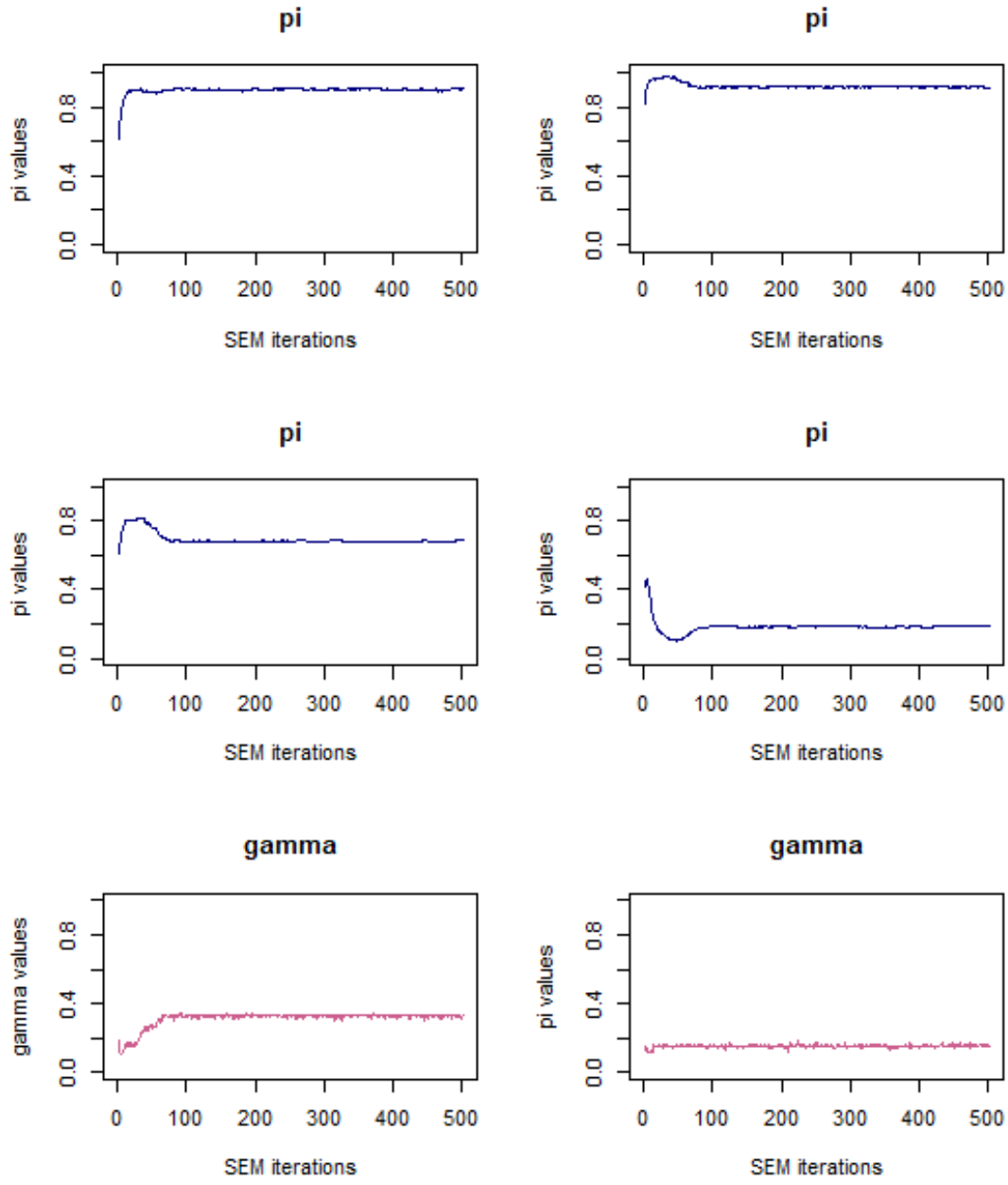

Fig. A1. Parameters evolution along time in the SEM-algorithm.