MODEL-BASED SIMULTANEOUS CLASSIFICATION AND REDUCTION FOR THREE-WAY ORDINAL DATA

Monia Ranalli [1], Roberto Rocci[1]

[1] Department of Statistics, Sapienza University, (e-mail: monia.ranalli@uniroma1.it, roberto.rocci@uniroma1.it)

ABSTRACT: A finite mixture model for the unsupervised classification of three-way ordinal data is proposed. Technically, it is a finite mixture of Gaussians observed only through a discretization of its variates. Group specific means and covariances are reparameterized according to parsimonious models. Estimation is carried out through a composite approach to reduce the computational burden.

KEYWORDS: three-way ordinal data, mixture models, composite likelihood, EM algorithm.

1 Introduction

In a cluster analysis context, finite mixtures of Gaussians are frequently used to classify a sample of observations (see for example Hennig *et al.*, 2015), even with complex data structure. This may happens when there are different types of variables or different occasions, i.e. same observations and variables measured at different time points or under different experimental settings. The Gaussian mixture model has been generalized to mixtures of matrix Normal distributions under a frequentist (Viroli, 2011a) and a Bayesian (Viroli, 2011b) framework. The main disadvantage of this model is given by the large number of parameters involved. In the literature, there is a broad consensus in identifying as a possible solution an approach based on performing clustering and dimensionality reduction simultaneously. Indeed, several authors have already proposed such methods (see for example: Rocci & Vichi, 2005, Vichi *et al.*, 2007, Tortora *et al.*, 2016) but only using an optimization approach. In this paper, we focus on three-way ordinal data following a model based approach. We assume that the ordinal variables are variates of a mixture only partially observed through a discretization (Ranalli & Rocci, 2016). This allows us to capture the cluster structure underlying the data, since each component of the mixture corresponds to an underlying group. To reduce the dimensionality, group-specific mean vectors and group-specific covariance matrices are reparametrized according to parsimonious models that are able to highlight the discrimination power of both variables and occasions while taking into account the three-way structure of the data. The presence of ordinal variables makes the maximum likelihood estimation unfeasible (see for details Ranalli & Rocci, 2016). To overcome the computational issues due to the presence of high dimensional integrals, a composite likelihood (Lindsay, 1988) approach is proposed. The computation of parameter estimates is carried out through an EM-like algorithm.

2 The model

Let $\mathbf{x} = [x_{11}, x_{21}, \ldots, x_{P1}, \ldots, x_{1R}, x_{2R}, \ldots, x_{PR}]'$ be a random vector of P ordinal variables observed at R different occasions. For each ordinal variable we observe $c_p = 1, \ldots, C_p$ categories with $p = 1, \ldots, P$ in each occasion. Following the underlying response variable approach, the observed ordinal variables $\mathbf{x}$ are considered as a discretization of some continuous latent variables $\mathbf{y} = [y_{11}, y_{21}, \ldots, y_{P1}, \ldots, y_{1R}, y_{2R}, \ldots, y_{PR}]'$. The relationship between $\mathbf{x}$ and $\mathbf{y}$ is

$$\gamma_{c_p-1}^{(p)} \leq y_{pr} < \gamma_{c_p}^{(p)} \Leftrightarrow x_{pr} = c_p,$$

where $-\infty = \gamma_0^{(p)} < \gamma_1^{(p)} < \ldots < \gamma_{C_p-1}^{(p)} < \gamma_{C_p}^{(p)} = +\infty$ are non observable thresholds defining the C_p categories and constant over the occasions. We assume that $\mathbf{y}$ follows a heteroscedastic Gaussian mixture model, which is only partially observed,

$$f(\mathbf{y}) = \sum_{g=1}^{G} p_g \phi(\mathbf{y}; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g), \tag{1}$$

where ϕ is the multivariate normal density with mean $\boldsymbol{\mu}_g$ and covariance matrix $\boldsymbol{\Sigma}_g$, while p_g is the group-specific weight, with $p_g > 0 \ \forall g = 1, \ldots, G$ and $\sum_{g=1}^{G} p_g = 1$. To reduce the number of parameters, the group-specific covariance matrix is modelled as follows (Browne, 1984)

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}_{O,g} \otimes \boldsymbol{\Sigma}_{V,g}, \tag{2}$$

where $\otimes$ is the Kronecker product of matrices; while $\boldsymbol{\Sigma}_{O,g}$ and $\boldsymbol{\Sigma}_{V,g}$ represent the group-specific covariance matrices of occasions and variables, respectively. The dimensionality reduction is performed on the group-specific mean vectors following a Tucker2 model (Tucker, 1966). The $G \times (PR)$ matrix collecting the group-specific means is given by

$$\mathbf{M} = (\mathbf{C} \otimes \mathbf{B})\mathbf{N}, \tag{3}$$

where $\mathbf{N}$ collects the scores of the G groups on the Q latent variables under S latent occasions, $\mathbf{B}$ is the loadings matrix that connects the P variables with Q latent variables, $\mathbf{C}$ is the loadings matrix that connects R occasions with the S latent occasions. This trilinear model allows us to project the within-group means, lying into a PR dimensional space, onto a reduced subspace of dimension QS. The number of parameters can be further reduced by observing that $\mathbf{B}$ can be decomposed as follows,

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_U \\ \mathbf{B}_L \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{B}_L \mathbf{B}_U^{-1} \end{bmatrix} \mathbf{B}_U = \widetilde{\mathbf{B}} \mathbf{B}_U$$

where $\mathbf{B}_U$ is assumed to be invertible. The same can be done with $\mathbf{C}$, leading to a more parimonious model for the group-specific mean, that is

$$\begin{aligned} \mathbf{M} = (\mathbf{C} \otimes \mathbf{B})\mathbf{N} &= \left[(\widetilde{\mathbf{C}} \mathbf{C}_U) \otimes (\widetilde{\mathbf{B}} \mathbf{B}_U) \right] \mathbf{N} \\ &= (\widetilde{\mathbf{C}} \otimes \widetilde{\mathbf{B}})(\mathbf{C}_U \otimes \mathbf{B}_U)\mathbf{N} \\ &= (\widetilde{\mathbf{C}} \otimes \widetilde{\mathbf{B}})\widetilde{\mathbf{N}}. \end{aligned}$$

For a i.i.d. random sample of size N, the log-likelihood is given by

$$\ell(\boldsymbol{\theta}) = \sum_{l=1}^{L} n_l \log \left[\sum_{g=1}^{G} p_g \pi(\mathbf{x}_l; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \boldsymbol{\gamma}) \right]$$

where $\mathbf{x}_l = (c_{11}^{(1)}, \ldots, c_{P1}^{(P)}, \ldots, c_{1R}^{(1)}, \ldots, c_{PR}^{(P)})$ is a particular response pattern with the frequence n_l ($\sum_{l=1}^{L} n_n = N$) and

$$\pi(\mathbf{x}_l; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \boldsymbol{\gamma}) = \int_{\gamma_{c_1-1}^{(1)}}^{\gamma_{c_1}^{(1)}} \cdots \int_{\gamma_{c_(PR)-1}^{(P)}}^{\gamma_{c_{PR}}^{(P)}} \phi(\mathbf{y}; \boldsymbol{\mu}_g, \Sigma_g) d\mathbf{y}$$

is its probability in the g-th component of the mixture. This likelihood causes non trivial computational problems due to the presence of multidimensional integrals. To overcome computational issues, we adopt a composite likelihood, based on low-dimensional margins.
Further details will be given in the extended version of the paper along with simulation and real data results to show the effectiveness of the proposal.

References

BROWNE, MICHAEL W. 1984. The decomposition of multitrait-multimethod matrices. *British journal of mathematical and statistical psychology*, **37**(1), 1–21.

CATTELL, RAYMOND B. 1944. "Parallel proportional profiles" and other principles for determining the choice of factors by rotation. *Psychometrika*, **9**(4), 267–283.

GIORDANI, PAOLO, ROCCI, ROBERTO, & BOVE, GIUSEPPE. 2020. Factor Uniqueness of the Structural Parafac Model. *Psychometrika*, **85**(3), 555–574.

HENNIG, CHRISTIAN, MEILA, MARINA, MURTAGH, FIONN, & ROCCI, ROBERTO. 2015. *Handbook of cluster analysis*. CRC Press.

JÖRESKOG, KARL G. 1990. New developments in LISREL: analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity*, **24**(4), 387–404.

LINDSAY, BRUCE. 1988. Composite likelihood methods. *Contemporary Mathematics*, **80**, 221–239.

MCNICHOLAS, PAUL D., & MURPHY, THOMAS BRENDAN. 2010. Model-based clustering of microarray expression data via latent Gaussian mixture models. *Bioinformatics*, **26**(21), 2705–2712.

RANALLI, MONIA, & ROCCI, ROBERTO. 2016. Mixture models for ordinal data: a pairwise likelihood approach. *Statistics and Computing*, **26**, 529–547.

RANALLI, MONIA, & ROCCI, ROBERTO. 2017. Mixture models for mixed-type data through a composite likelihood approach. *Computational Statistics & Data Analysis*, **110**(C), 87–102.

ROCCI, ROBERTO, & VICHI, MAURIZIO. 2005. Three-mode component analysis with crisp or fuzzy partition of units. *Psychometrika*, **70**(4), 715.

TORTORA, CRISTINA, SUMMA, MIREILLE GETTLER, MARINO, MARINA, & PALUMBO, FRANCESCO. 2016. Factor probabilistic distance clustering (FPDC): a new clustering method. *Advances in Data Analysis and Classification*, **10**, 441–464.

TUCKER, LEDYARD R. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika*, **31**(3), 279–311.

VICHI, MAURIZIO, ROCCI, ROBERTO, & KIERS, HENK AL. 2007. Simultaneous component and clustering models for three-way data: within and between approaches. *Journal of Classification*, **24**(1), 71–98.

VIROLI, CINZIA. 2011a. Finite mixtures of matrix normal distributions for classifying three-way data. *Statistics and Computing*, **21**, 511–522.

VIROLI, CINZIA. 2011b. Model based clustering for three-way data structures. *Bayesian Analysis*, **6**(4), 573–602.