Computational Statistics and Data Analysis 93 (2016) 46–75

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Mixture-based clustering for the ordered stereotype model

D. Fernández [*], R. Arnold, S. Pledger

School of Mathematics, Statistics and Operations Research, Victoria University of Wellington, Wellington, New Zealand

HIGHLIGHTS

- New methodology for clustering rows and columns from a matrix of ordinal data.
- Establishes likelihood-based methods via finite mixtures with the stereotype model.
- Tests the reliability of this methodology through a simulation study.
- Illustrates this new approach with two examples.
- Reviews and compares the performance several model choice measures.

ARTICLE INFO

Article history:
Received 20 March 2014
Received in revised form 6 November 2014
Accepted 8 November 2014
Available online 18 November 2014

Keywords:
Biclustering
Cluster analysis
Dimension reduction
EM-algorithm
Finite mixture model
Fuzzy clustering
Likert scale
Ordinal data
Stereotype model

ABSTRACT

Many of the methods which deal with the reduction of dimensionality in matrices of data are based on mathematical techniques such as distance-based algorithms or matrix decomposition and eigenvalues. Recently a group of likelihood-based finite mixture models for a data matrix with binary or count data, using basic Bernoulli or Poisson building blocks has been developed. This is extended and establishes likelihood-based multivariate methods for a data matrix with ordinal data which applies fuzzy clustering via finite mixtures to the ordered stereotype model. Model-fitting is performed using the expectation–maximization (EM) algorithm, and a fuzzy allocation of rows, columns, and rows and columns simultaneously to corresponding clusters is obtained. A simulation study is presented which includes a variety of scenarios in order to test the reliability of the proposed model. Finally, the results of the application of the model in two real data sets are shown.

1. Introduction

An ordinal variable is one with a categorical data scale which describes order, and where the distinct levels of such a variable differ in degree of dissimilarity more than in quality (Agresti, 2010). This is different from nominal variables which vary in quality, not in quantity, and thus the order of listing the categories is irrelevant. For example, Likert scale responses in a questionnaire might be "disagree", "neither agree nor disagree" or "agree". In his seminal paper, Stevens (1946) called a scale ordinal if "any order-preserving transformation will leave the scale form invariant". Although the collection and use of ordinal variables is common, most of the current methods for analyzing them treat the data as if they were nominal (Hoffman and Franke, 1986) or continuous data (Agresti, 2010). On the one hand, treating an ordered categorical variable as ordinal rather than nominal provides advantages in the analysis such as simplifying the data description and allowing the use of more parsimonious models. The nominal approach ignores the intrinsic ordering of the data and thus the statistical

[*] Correspondence to: MSOR, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand. Tel.: +64 4 463 5341.
E-mail addresses: daniel.fernandez@msor.vuw.ac.nz (D. Fernández), richard.arnold@msor.vuw.ac.nz (R. Arnold), shirley.pledger@vuw.ac.nz (S. Pledger).

http://dx.doi.org/10.1016/j.csda.2014.11.004
0167-9473/© 2014 Elsevier B.V. All rights reserved.

results are less powerful than they could be. On the other hand, models for continuous variables have similarities to those for ordinal variables although the use of them with ordinal variables has disadvantages such as the treatment of the output categories as equally spaced, which they may not be (see Agresti, 2010, Sections 1.2–1.3 for a list of advantages from treating an ordinal variable as ordinal rather than nominal or continuous).

Categorical data analysis methods were first developed in the 1960s and 1970s (Bock and Jones, 1968; Snell, 1964), including loglinear models and logistic regression (see the review by Liu and Agresti, 2005). An increasing interest in ordinal data has since produced the articles by Goodman (1979) and McCullagh (1980) on loglinear modeling relating to ordinal odds ratios, and logit modeling of cumulative probabilities respectively. Recently, new ordinal data analysis methods have been introduced such as the proportional odds model version of the cumulative logit model, and the stereotype model with ordinal scores (Agresti, 2010, Chap. 3 and 4) from which new lines of research have developed. Two recent examples of these are the application of a stereotype model in a case-control study by Ahn et al. (2009), and a new methodology to fit a stratified proportional odds model by Mukherjee et al. (2008). In particular, the stereotype model is a paired-category logit model which is an alternative when the fit of cumulative logits and adjacent-categories logit models in their proportional odds version is poor. Anderson (1984) proposed this model as nested between the adjacent-categories logit model and the standard baseline-category logits model (see the review by Agresti, 2002, Chapter 6).

In the research literature, multiple algorithms and techniques have been developed which deal with the clustering of data such as hierarchical clustering (Johnson, 1967; Kaufman and Rousseeuw, 1990), association analysis (Manly, 2005) and partition optimization methods such as the k-means clustering algorithm (Jobson, 1992; Lewis et al., 2003; McCune and Grace, 2002). There has been research on cluster analysis for ordinal data based on latent class models (see Agresti and Lang, 1993; Moustaki, 2000; Vermunt, 2001; DeSantis et al., 2008; Breen and Luijkx, 2010; McPartland and Gormley, 2013 and the review by Agresti, 2010, Section 10.1). There are a number of clustering methods based on mathematical techniques such as distance metrics (Everitt et al., 2001), association indices (Wu et al., 2008; Chen et al., 2011), matrix decomposition and eigenvalues (Quinn and Keough, 2002; Manly, 2005; Wu et al., 2007). However, these do not have a likelihood based formulation, and do not provide a reliable method of model selection or assessment. A particularly powerful model-based approach to one-mode clustering based on finite mixtures, with the variables in the columns being utilized to cluster the subjects in the rows, is provided by McLachlan and Basford (1988), McLachlan and Peel (2000), Everitt et al. (2001), Böhning et al. (2007), Wu et al. (2008) and Melnykov and Maitra (2010).

The simultaneous clustering of rows and columns into row clusters and column clusters is called biclustering (or block clustering or two-mode clustering). Biclustering models based on double k-means have been developed in Vichi (2001) and Rocci and Vichi (2008). A hierarchical Bayesian procedure for biclustering is given in DeSarbo et al. (2004). Biclustering using mixtures has been proposed for binary data in Pledger (2000), Arnold et al. (2010) and Labiod and Nadif (2011), and for count data in Govaert and Nadif (2010). An approach via finite mixtures for binary and count data using basic Bernoulli or Poisson building blocks has been developed in Govaert and Nadif (2010) and Pledger and Arnold (2014). This work expanded previous research for one-mode fuzzy cluster analysis based on finite mixtures (McLachlan and Basford, 1988; McLachlan and Peel, 2000; Everitt et al., 2001) to a suite of models including biclustering. Finally, Matechou et al. (2011) have recently developed biclustering models for ordinal data using the assumption of proportional odds and having a likelihood-based foundation. The main difference with our work is that we use the assumption of ordinal stereotype model which has the advantage of allowing us to determine a new spacing of the ordinal categories, dictated by the data.

In this article, we present an extension of the likelihood-based models proposed in Pledger and Arnold (2014) by applying them to matrices with ordinal data by using finite mixtures to define a fuzzy clustering. We use the ordered stereotype model introduced by Anderson (1984) in order to formulate the ordinal approach, which has rarely been used so far. Two possible reasons for this lack of use might be the absence of standard software for model fitting and its unusual structure including the product of parameters in the linear predictor (Kuss, 2006). The plan of the article is as follows. Section 2 has definitions of the models and its formulation including fuzzy clustering via finite mixtures. Model fitting by using the iterative EM algorithm is described in Section 3. Section 4 presents a review of several model comparison measures and a comparison of eleven information criteria performance. Two real-life examples and simulation studies are given in Section 5, and we conclude with a discussion in Section 6.

2. Model formulation

In this section, we give the standard definition of the ordered stereotype model (Section 2.1) followed by a modification to include clustering (Section 2.2). The likelihood for the suite of basic models is provided next (Section 2.3).

2.1. Data and ordered stereotype model definition

For a set of m ordinal response variables each with q categories measured on a set of n units, the data can be represented by a $n \times m$ matrix Y where, for instance, the n rows represent the subjects of the study and the m columns are the different questions in a particular questionnaire. Although the number of categories might be different, we assume the same q for all such questions. If each answer is a selection from q ordered categories (e.g. strongly agree, agree, neutral, disagree, strongly disagree), then

$$y_{ij} \in \{1, \dots, q\}, \quad i = 1, \dots, n, \; j = 1, \dots, m.$$