Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Variational algorithms for biclustering models

Q1 Duy Vu, Murray Aitkin

Department of Mathematics and Statistics, University Of Melbourne, VIC 3010, Australia

ARTICLE INFO

Article history:
Received 27 February 2014
Received in revised form 12 February 2015
Accepted 24 February 2015
Available online xxxx

Keywords:
Biclustering
Stochastic block models
EM algorithms
Generalized EM algorithms
Variational EM algorithms
MM algorithms

ABSTRACT

Biclustering is an important tool in exploratory statistical analysis which can be used to detect latent row and column groups of different response patterns. However, few studies include covariate data directly into their biclustering models to explain these variations. A novel biclustering framework that considers both stochastic block structures and covariate effects is proposed to address this modeling problem. Fast approximation estimation algorithms are also developed to deal with a large number of latent variables and covariate coefficients. These algorithms are derived from the variational generalized expectation–maximization (EM) framework where the goal is to increase, rather than maximize, the likelihood lower bound in both E and M steps. The utility of the proposed biclustering framework is demonstrated through two block modeling applications in model-based collaborative filtering and microarray analysis.

1. Introduction

We consider a random matrix $\boldsymbol{Y} = [Y_{ij}] \in \Re^{n\times m}$ where observed values y_{ij} can be binary, count, or real depending on the applications involved. Our biclustering task is to simultaneously arrange rows and columns into groups of similar response patterns (Mirkin, 1996). In collaborative filtering, for example, Y_{ij} is a binary variable equal to 1 if user i liked movie j; and the modeling task is to cluster users into groups of similar reviewing patterns, while concurrently detecting movie groups with similar attractiveness levels (Su and Khoshgoftaar, 2009). Another example is from bipartite network modeling where Y_{ij} is the number of posts from user i to topic j in a Web forum; and the goal is to cluster users and topics into blocks of similar activity patterns (Borgatti and Everett, 1997; Freeman, 2003; Doreian et al., 2004).

The biclustering task that we are considering is also called co-clustering (Hanisch et al., 2002), and latent or stochastic block modeling (Arabie et al., 1978). The common assumption made by these models is that each row or column belongs exclusively to only one row or column group in contrast with other overlapping biclustering models where each row or column can also belong to more than one group or none (Cheng and Church, 2000; Lazzeroni and Owen, 2002). These overlapping biclustering models and their applications in microarray analysis are discussed in detail by Madeira and Oliveira (2004). In this paper, we consider only the non-overlapping group assumption and make three significant contributions to stochastic block modeling of bipartite datasets.

First, we describe a biclustering framework that can jointly model latent block structures and covariate effects. Although many biclustering models have been proposed in the literature, only few take into account the covariate effects on response patterns (Madeira and Oliveira, 2004). Moreover, if considered, these covariates are only controlled in an indirect manner. Q2 Their effects are either removed before biclustering algorithms are applied to the resulting residuals (Flynn and Perry, 2012)

E-mail addresses: duy.vu@unimelb.edu.au (D. Vu), murray.aitkin@unimelb.edu.au (M. Aitkin).

http://dx.doi.org/10.1016/j.csda.2015.02.015
0167-9473/

Please cite this article in press as: Vu, D., Aitkin, M., Variational algorithms for biclustering models. Computational Statistics and Data Analysis (2015), http://dx.doi.org/10.1016/j.csda.2015.02.015

Table 1
The percentages of movies in each genre in the *MovieLens 100K* dataset.

Action	Adventure	Animation	Children	Comedy	Crime
15%	8%	2%	7%	30%	6%
Documentary	Drama	Fantasy	Film-Noir	Horror	Musical
3%	43%	1%	1%	5%	3%
Mystery	Romance	Sci-Fi	Thriller	War	Western
4%	15%	6%	15%	4%	2%

or only estimated after the group structures are obtained (Wang et al., 2013b). In addition, our direct consideration of covariate structures can also help to reduce the latent class search space. The best biclustering model tends to have a smaller number of row and column groups since a large amount of variation is already explained by these observed covariates.

Our second contribution is a novel application of the variational generalized EM (VGEM) framework (Vu et al., 2013) to the estimation of the biclustering models. Based on the minorization–maximization principle (Hunter and Lange, 2004), our estimation algorithms seek only to increase, rather than maximize, the likelihood lower bound in both E and M steps of each variational EM iteration. More specifically, we first extend the variational generalized E step in Vu et al. (2013) to deal with double latent class structures and handle both discrete and continuous responses. A comparison experiment between this new MM update and the popular fixed-point update is then carried out to demonstrate its superior performance. Furthermore, in contrast with Vu et al. (2013) who only considered the MM principle in the variational E step, we derive new MM updates for the variational M step so that a large number of covariate effects can be estimated. These biclustering algorithms are complete examples of how the VGEM framework can be fully operationalized in practice.

Finally, the broad application of our biclustering framework and estimation algorithms is demonstrated through two different application datasets. In the *MovieLens 100K* dataset, for example, we allow the effects of 3 row and 18 column covariates to vary over columns and rows respectively, which results in the total number of 22,020 coefficients to estimate. Consequently, this application not only illustrates the important role of covariate data in model selection, but also shows how the MM principle can help to estimate such a large number of parameters. In these examples, we have also explored the application of the bootstrapping procedure in network analysis (Hunter et al., 2008) to biclustering model diagnostics. Besides model selection criteria, these goodness-of-fit plots can provide us with a powerful visual tool for model checking.

The rest of this paper is organized as follows. In Section 2, two example datasets from different areas are discussed to motivate the application of biclustering models. The general biclustering framework is then presented in Section 3 followed by the derivation of VGEM estimation algorithms in Section 4. Section 5 shows the advantage of this VGEM framework by comparing the MM update with the FP update on two empirical datasets of both discrete and continuous responses. In Section 6, the broad application of the biclustering framework and MM algorithms is then demonstrated with two analyses of example datasets in Section 2.

2. Data

To motivate the application of the biclustering framework, we consider two real datasets corresponding to binary and real response types, respectively. They also illustrate the application of the biclustering framework in two different areas: model-based collaborative filtering and microarray analysis. Appendix E discusses another application of our biclustering framework in network modeling with count-valued data.

MovieLens 100K. The dataset contains 100,000 ratings from 943 users of 1682 movies (Herlocker et al., 1999).[1] We limit our interest in this dataset to detecting viewing patterns across users and movies. Since most users rate movies only after watching them, we convert all non-zero ratings to 1 while assigning 0 to entries without ratings. This results in a binary matrix of 943 rows and 1682 columns. The dataset also comes with 2 row covariates for age and gender, and 18 binary column covariates for movie genres. The first, second, and third quantiles of age are 25, 31, and 43, respectively; while the percentage of female reviewers is 29%. Table 1 shows the distribution of movies across different movie genres. It is important to note that movie genres are not exclusive, i.e., a movie might have multiple genres. For example, the movie *Toy Story* was simultaneously classified into three genres *Animation*, *Children*, and *Comedy*. Our modeling goals are to both extract the latent class structures of users and movies, and estimate the effects of these covariates on viewing patterns. For those readers who are interested in the modeling of rating scores, the discussion on admixture models in (Zhou and Lange, 2009) provides a starting point.

AGEMAP. This dataset is used to demonstrate our extension of the VGEM framework to continuous responses. The matrix contains the expression levels of 39 mice across 17,864 genes of two tissue types: cerebellum and cerebrum (Zahn et al., 2007).[2] It also comes with 2 covariates on mice: age and gender. There are 4 different age groups of 1, 6, 16, and 24 months. Each combination of age and gender has 5 mice, except the group of male mice at 24 months only has 4 samples. In this application, our exploratory goal is to group samples and genes into blocks of similar expression patterns.

[1] http://movielens.umn.edu.
[2] http://cmgm.stanford.edu/~kimlab/aging_mouse/.

Please cite this article in press as: Vu, D., Aitkin, M., Variational algorithms for biclustering models. Computational Statistics and Data Analysis (2015), http://dx.doi.org/10.1016/j.csda.2015.02.015