Neurocomputing 173 (2016) 442–449

Contents lists available at ScienceDirect

Neurocomputing

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

Generalized topographic block model

Rodolphe Priam [a,*], Mohamed Nadif [b], Gérard Govaert [c]

[a] S3RI, Southampton University, University Road, Southampton SO17 1BJ, United Kingdom
[b] LIPADE, Université Paris Descartes, UFR Mathématiques-Informatique, 45, rue des Saints Pères, 75270 Paris, France
[c] UTC, UMR CNRS 6599 Heudiasyc, Centre de Recherches de Royallieu, 60205 Compiègne, France

ARTICLE INFO

Article history:
Received 6 July 2014
Received in revised form
24 February 2015
Accepted 8 April 2015
Available online 8 September 2015

Keywords:
Latent block mixture model
Exponential family
Generative topographic mapping
Block expectation–maximization
Visualisation

ABSTRACT

Co-clustering leads to parsimony in data visualisation with a number of parameters dramatically reduced in comparison to the dimensions of the data sample. Herein, we propose a new generalized approach for nonlinear mapping by a re-parameterization of the latent block mixture model. The densities modeling the blocks are in an exponential family such that the Gaussian, Bernoulli and Poisson laws are particular cases. The inference of the parameters is derived from the block expectation–maximization algorithm with a Newton–Raphson procedure at the maximization step. Empirical experiments with textual data validate the interest of our generalized model.

1. Introduction

For the visualisation [1,2] of a data matrix, the main proximities or the higher correlations are summarized by a comprehensible and low dimensional graphical view. When the number of variables is large, the visualisation may combine a preprocessing stage by selection or linear transformation [3–5]. In a co-clustering method, both sides of the matrix are partitioned [6], hence the reduction of the variables space and the row clustering occur simultaneously. A earliest co-clustering formulation called direct clustering was introduced by Hartigan [7] who proposed a greedy algorithm for hierarchical co-clustering. We can also mention the following works [8–12] and the reviews in [13–15]. These methods are dedicated to a simultaneous clustering but not to visualisation.

Co-clustering is combined to self-organizing maps (SOM) for visualisation or clustering purposes in many ways in the literature [16–22], with illustrations to biological or textual data. Such combination can improve the quality of the clustering [23,24], with two contributing modeling factors or four sub-ones. Roughly speaking, the co-clustering leads to (a) the parsimony of the parameters and (b) the groups of variables. And, the auto-organization leads to (c) the partition of each class into several clusters and (d) the connections between neighbour clusters. Note that the subfactors (c) can enhance the classification [25], while (a) and (b) the regression [26].

The family of methods SOM counts the variants and the extensions of the Kohonen's map [27] which is a sequential clustering algorithm with decreasing connections of vicinity between the clusters for mapping continuous data. Modified versions are adapted to the analysis of discrete, sequential or block matrices for instance. Moreover, generative models for self-organizing maps has been justified [28–30]. The Generative Topographic Mapping (GTM) [31] is a probabilistic model of SOM for data visualisation [32–35]. In GTM the auto-organization of the clusters is directly induced by the parameterization. The algorithm of Kohonen's map is re-formulated by embedding the auto-organization process at the level of the means of a Gaussian mixture model (GMM) [36].

Herein we are interested on a probabilistic co-clustering model, the latent block mixture model (LBM) [12,37,38], in order to visualize the natural classes in a block matrix with a parameterization similar to GTM. First, we define a general model of LBM with the help of an univariate exponential family [39] which is well suited for most kinds of numerical variables. Then we introduce a parameterization of the central parameters in order to simultaneously perform the clustering and the reduction of the obtained clusters in a low dimensional space. The model is general enough to be related not only to self-organizing maps but also to recent approaches in factorization [40–42]. This offers a broad perspective for data analysis as illustrated through a generalized method for block generative topographic mapping or block GTM (BGTM) [43,44]. If the previous models of block generative topographic mapping have been proposed for only one particular

* Corresponding author.
E-mail addresses: rpriam@gmail.com (R. Priam), mohamed.nadif@parisdescartes.fr (M. Nadif), gerard.govaert@utc.fr (G. Govaert).

http://dx.doi.org/10.1016/j.neucom.2015.04.115
0925-2312/

distribution for the blocks, the new generalization is able to provide a unified framework for the visualisation of data block matrices and can help for implementing and comparing alternative distributions in future.

The paper is organized as follows. In Section 2, we introduce the latent block model for an exponential family and add the constraints. In Section 3, we present the related objective function to optimize for the estimation of parameters. We deduce the learning algorithm in a general setting, from the block expectation–maximization (BEM) [45]. In Section 4 we present the connection of our approach with GTM and discuss the resulting nonlinear visualisation. In Section 5 we present the numerical experiments for testing the proposed approach. Finally, in Section 6 we summarize our contribution.

2. Generalized LBM

Let us have $\mathbf{x}:=\{x_{ij}; i=1,\dots,n; j=1,\dots,d\}$ stands for a data matrix of size $n\times d$. When $\mathbf{x}$ is a two-way contingency table it is associated to two categorical variables that take values in sets $I=\{1,\dots,n\}$ and $J=\{1,\dots,d\}$. In this case, the entries x_{ij} are co-occurrences of row and column categories, each of them counts the number of entities that fall simultaneously in the corresponding row and column categories. Let $\mathbf{z}$ and $\mathbf{w}$ be partitions in g row clusters and m column clusters of I and J of $\mathbf{x}$. The partition $\mathbf{z}$ will be represented by the vector of labels $(z_1,\dots,z_i,\dots,z_n)$ where $z_i\in\{1,\dots,g\}$ or, by the classification matrix $\{z_{ik}; i=1,\dots,n; k=1,\dots,g\}$ where $z_{ik}=1$ if i belongs to the kth cluster and 0 otherwise. A similar notation will be used for the partition $\mathbf{w}$ which will be represented by the vector $(w_1,\dots,w_j,\dots,w_d)$ where $w_j\in\{1,\dots,m\}$ or the classification matrix $\{w_{j\ell}; j=1,\dots,d; \ell=1,\dots,m\}$. Note that $z_{ik}w_{j\ell}=1$ if x_{ij} belongs to the $(k\ell)$th block and 0 otherwise. For a latent block model, the $n\times d$ random variables that generate the observed cells x_{ij} are assumed to be independent, once $\mathbf{z}$ and $\mathbf{w}$ are fixed, they make it possible to define a co-clustering model. Hereafter, to simplify the notation, the sums and the products relating to rows, columns or clusters will be subscripted respectively by the letters i, j, k, or ℓ without indicating the limits of variation, which are implicit.

2.1. Latent block model (LBM)

The probability density function (pdf) of a latent block model is denoted $f_{LBM}(\mathbf{x};\boldsymbol{\theta})$ and defined as the following decomposition. It is obtained by independence of $\mathbf{z}$ and $\mathbf{w}$, by summing over all the assignments [12] and takes the following form:

$$\sum_{(\mathbf{z},\mathbf{w})\in\mathcal{Z}\times\mathcal{W}}\prod_i p_{z_i}\prod_j q_{w_j}\prod_{i,j}\varphi(x_{ij};\alpha^{ij}_{z_iw_j}),$$

where the set of all the possible assignments is denoted $\mathcal{Z}$ for I and $\mathcal{W}$ for J, while $\varphi(.;\alpha^{ij}_{k\ell})$ is a probability density function defined for cell (ij) on the set of reals $\mathbb{R}$ while $\alpha^{ij}_{k\ell}$ depends on the parameter $\alpha_{k\ell}$ as given in (1). The vectors of the probabilities p_k and q_ℓ that a row (resp. a column) belongs to the kth component (resp. ℓth component) are denoted $\mathbf{p}=(p_1,\dots,p_g)$ (resp. $\mathbf{q}=(q_1,\dots,q_m)$). The set of parameters is denoted $\boldsymbol{\theta}$ and is a compound of $\mathbf{p}$ and $\mathbf{q}$ plus $\boldsymbol{\alpha}$ which aggregates all the parameters from the pdf of the cells, $\boldsymbol{\theta}=\{\mathbf{p},\mathbf{q},\boldsymbol{\alpha}\}$. The set of parameters $\boldsymbol{\theta}$ of the model can be estimated by maximizing the log-likelihood:

$$L(\mathbf{x};\boldsymbol{\theta})=\log f_{LBM}(\mathbf{x};\boldsymbol{\theta}).$$

The block model is dramatically more parsimonious than the usual mixture model where each dimension of the data table is modeled separately. Next, we describe the latent block model where φ is in an exponential family.

2.2. Univariate exponential family of distributions

When the cells are generated with an exponential family, the latent block model is denoted in the following ELBM and the density function for the $(k\ell)th$ block is written as

$$\varphi(x_{ij};\alpha^{ij}_{k\ell})=\exp\left(x_{ij}A(\alpha^{ij}_{k\ell})-B(\alpha^{ij}_{k\ell})+C(x_{ij})\right),$$

where $A(\alpha^{ij}_{k\ell})$ is the natural parameter, while $B(\alpha^{ij}_{k\ell})$ and $C(x_{ij})$ ensure that φ is a probability density function. The considered form of distributions is defined without nuisance parameter and without loss of generality. Note that a more general expression is possible for modeling more particular distributions. For instance a function of x_{ij} could be used instead of the identity one or A could be chosen multivariate. It is also supposed that the quantities $\alpha^{ij}_{k\ell}$ are written as a function of a fixed parameter depending on the data β_{ij} and an unknown parameter named $\alpha_{k\ell}$, such that

$$\alpha^{ij}_{k\ell}=\beta_{ij}\alpha_{k\ell}. \tag{1}$$

The model can be represented by a graphical model depicted in Fig. 1. Here two aggregating matrices are involved, $\boldsymbol{\alpha}=(\alpha_{k\ell})_{g\times m}$ for the parameters and $\boldsymbol{\beta}=(\beta_{ij})_{n\times d}$ for the multiplicative effects. Three cases of distributions which belong to this family for discrete and continuous matrices are considered. The different distributions are listed in Table 1, where the cells are drawn from one particular distribution. For a Bernoulli law with the parameters $\alpha_{k\ell}$, the model is denoted BLBM [12]. For a Poisson law with the parameters $\alpha_{k\ell}$, it is denoted PLBM [46], with $\boldsymbol{\mu}=(\mu_1,\dots,\mu_n)^T$ where $\mu_i=\sum_j x_{ij}$ and $\boldsymbol{\nu}=(\nu_1,\dots,\nu_d)^T$ where $\nu_j=\sum_i x_{ij}$. For a normal law, it is denoted GLBM [47] with the means $\alpha_{k\ell}$ and the variances $\sigma_{k\ell 0}$ assumed constant here. For the three cases, the support for the variables generating the observation x_{ij} and the parameter range of $\alpha_{k\ell}$ are defined in Table 1. Note that in the case of a Poisson law, the parameters $\alpha_{k\ell}$ can be chosen unconstrained as introduced in [46], and the quantities β_{ij} can be optimized too in certain cases. Next, each parameter $\alpha_{k\ell}$ is written with a link function as explained.

2.3. Re-parameterization of the model

The parameters of the exponential latent block model are parameterized with two sets of (unknown) vectors,

$$\{\xi_k\in\mathbb{R}^h, 1\le k\le g\},$$
$$\{w_\ell\in\mathbb{R}^h, 1\le\ell\le m\}. \tag{2}$$

where $h\in\mathbb{N}^*_+$ is the dimension of the latent space. These two sets of vectors are used for modeling the blocks $(k\ell)$ because each $\alpha_{k\ell}$ is dependent on two indices, k and ℓ. As an effect from the kth and

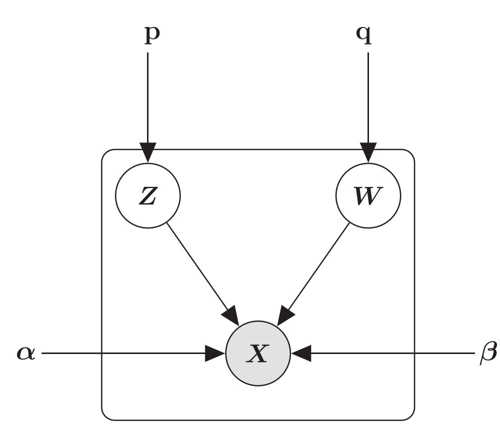

Fig. 1. Graphical notation for the Latent block model with random variables $\boldsymbol{X}$, $\boldsymbol{Z}$ and $\boldsymbol{W}$ generating the observations and latent labels.