Probabilistic latent variable models for shape correspondence analysis: Model Approaches

Hernán F. García and Mauricio A. Álvarez

April 30, 2018

Abstract
Probabilistic approaches for the correspondence problem.

1 The model

Based on the Iwatta's paper (see [2, 3]), we set nonlinear approaches for the shape correspondence analysis using probabilistic latent variable models.

Suppose that we are given objects in D domains $\mathcal{X} = \{\mathbf{X}_d\}_{d=1}^{D}$ mapped to a Hilbert space $\mathcal{H}$, where $\mathbf{X}_d = \{\mathbf{x}_{dn}\}_{n=1}^{N_d}$ is a set of objects in the dth domain, and $\mathbf{x}_{dn} \in \mathbb{R}^{M_d}$ is the observed vector of the nth object in the dth domain. We can cluster groups of correspondences by using a non-linear function that represents the shape descriptors in the Hilbert space.

As in infinite Gaussian mixture models, our approach assumes that there are an infinite number of clusters related to each correspondence, and each cluster j has a latent vector $\mathbf{z}_j \in \mathbb{R}^K$ in a latent space of dimension K. Descriptors that have the same cluster assignments s_{dn} are related by the same latent vector and considered to match (establish a groupwise correspondence).

Each object in $\mathbf{x}_{dn} \in \mathcal{R}^{M_d}$ in the dth domain is generated depending on the domain-specific projection matrix $\mathbf{W}_d \in \mathbb{R}^{M_d \times K}$ and latent vector $\mathbf{z}_{s_{dn}}$ that is selected from a set of latent vectors $\mathbf{Z} = \{\mathbf{z}_j\}_{j=1}^{\infty}$. Here, $s_{dn} = \{1, \ldots, \infty\}$ is the latent cluster assignment of object $\mathbf{x}_{dn}$.

The proposed model is based on an infinite mixture model, where the probability of descriptor mapped in a Hilbert space $\mathbf{x}_{dn}$ is given by

$$p\left(\mathbf{x}_{dn}|\mathbf{Z}, \boldsymbol{W}, \boldsymbol{\theta}\right) = \sum_{j=1}^{\infty} \theta_j \mathcal{N}\left(\mathbf{x}_{dn}|\mathbf{W}_d \mathbf{z}_j, \alpha^{-1}\mathbf{I}\right), \tag{1}$$

where $\boldsymbol{W} = \{\mathbf{W}_d\}_{d=1}^{D}$ is a set of projections matrices, $\boldsymbol{\theta} = (\theta_j)_{j=1}^{\infty}$ are the mixture weights, θ_j represents the probability that the jth cluster is chosen, and $\mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ denotes a normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. One important contribution derived in [2], is that we can analyze multiples structures with different properties and dimensionalities, by employing projection matrices for each brain structure (domain-specific). Figure 1 shows the scheme of the proposed model, in which the relationship between shape descriptors, and latent vectors is described.

2 Model Approaches

The main ideas for the probabilistic correspondence problem are summarized

1. We can relax the assumption that the observations are linear with respect to their latent vectors by using nonlinear matrix factorization techniques ([4]). From these work we point out that by marginalizing out the mapping matrix $\mathbf{W}$(which goes from the latent space to the observed data space), derived in a Bayesian multi-output regression model.

   Our model can be formalized as

$$p\left(\mathbf{x}_{dn}|\mathbf{Z}, \mathbf{W}, \boldsymbol{\theta}\right) = \sum_{j=1}^{\infty} \theta_j \prod_{m=1}^{M_d} \mathcal{N}\left(\mathbf{x}_{dnm}|f_{dm}\left(\mathbf{z}_j\right), \alpha^{-1}\right) \tag{2}$$

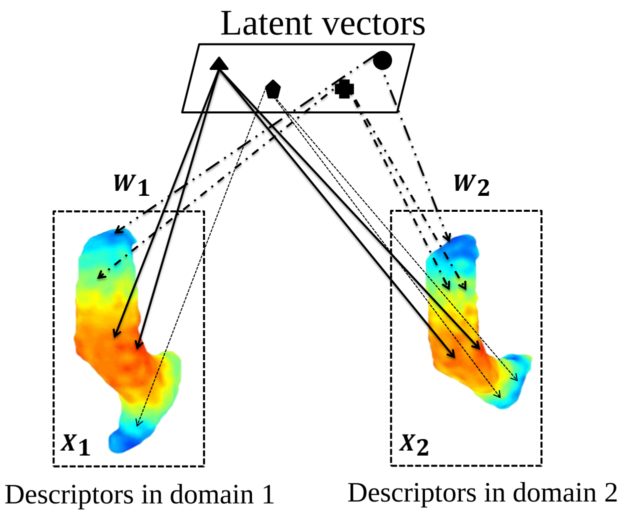

Figure 1: Scheme for the groupwise correspondence method. The figure shows an example of establishing clusters of correspondences in two domains (left ventrals).

where f_{dm} can handled from two perspectives

(a) Set the same covariance matrix for all domains

$$f_{dm} \sim \mathcal{GP}\left(\mu_{dm}\left(\mathbf{z}_j\right), k\left(\mathbf{z}_j, \mathbf{z}_j'\right)\right) \tag{3}$$

(b) One covariance matrix for each domain

$$f_{dm} \sim \mathcal{GP}\left(\mu_{dm}\left(\mathbf{z}_j\right), k_d\left(\mathbf{z}_j, \mathbf{z}_j'\right)\right) \tag{4}$$

2. As in the in the linear model of coregionalization (LMC), the outputs are expressed as linear combinations of independent random functions [1]. Consider a set of D outputs $\{\mathbf{f}_d(\mathbf{z_j})\}_{d=1}^{D}$ with $\mathbf{f}_d(\mathbf{z_j}) \in \mathbb{R}^{M_d}$. By adopting this framework, our model can be formulated as

$$p\left(\mathbf{x}_{dn}|\mathbf{Z}, \mathbf{W}, \boldsymbol{\theta}\right) = \sum_{j=1}^{\infty} \theta_j \mathcal{N}\left(\mathbf{x}_{dn}|\mathbf{f}_d(\mathbf{z_j}), \alpha^{-1}\mathbf{I}\right), \tag{5}$$

where each component can be expressed as:

(a) By using the linear model of coregionalization

$$\mathbf{f}_d(\mathbf{z_j}) \sim \mathcal{GP}\left(\mathbf{0}, \sum_{q=1}^{Q} \mathbf{B}_q k_d\left(\mathbf{z}_j, \mathbf{z}_j'\right)\right), \tag{6}$$

where $\mathbf{B}_q = \mathbf{L}_q^\top \mathbf{L}_q \in \mathbb{R}^{M_d \times M_d}$ is the coregionalization matrix (computed from the Cholesky decomposition)

(b) By using simplified version of the LMC, known as the intrinsic coregionalization model (ICM) (see [1]), assumes that the elements of the coregionalization matrix $\mathbf{B}_q$ can be written as a scaled version of the elements b_q, which do not depend on the particular output functions $f_d(\mathbf{z}_j)$.

$$\mathbf{f}_d(\mathbf{z_j}) \sim \mathcal{GP}\left(\mathbf{0}, \mathbf{B}k\left(\mathbf{z}_j, \mathbf{z}_j'\right)\right), \tag{7}$$

3 Multiview Warped Mixture Models

We can use the single view problem of the warped mixture model from [5] in which they warp a latent mixture of Gaussians into nonparametric cluster shapes. The low-dimensional latent mixture model summarizes the properties of the high-dimensional density manifolds describing the data.

Our idea is to introduce a model which warps a multiview latent mixture of Gaussians (possibly MRD) to produce nonparametric cluster shapes. [1]

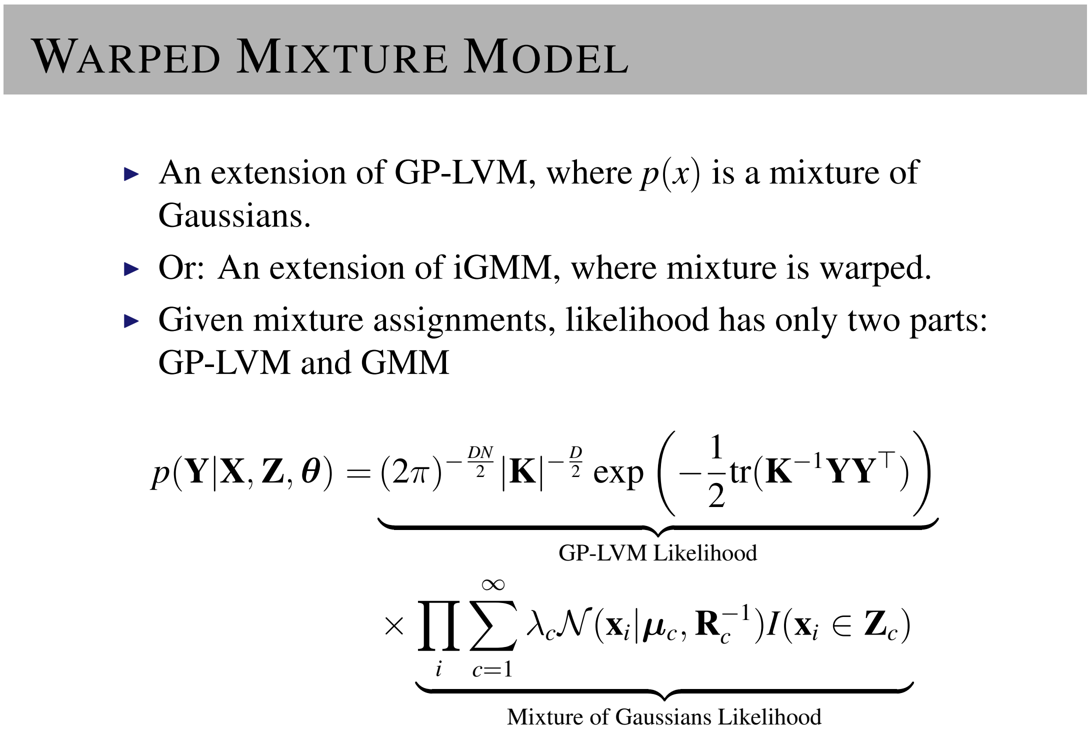

Figure 2: WMM from [5]

3.1 The model

Let as define a multi-view data set as $\mathcal{Y} = \{\mathbf{Y}^v\}_{v=1}^V$, where each view is defined as $\mathbf{Y}^v \in \mathbb{R}^{N_v \times D_v}$. This leads to the likelihood

$$p\left(\mathbf{Y}^{\mathcal{V}}|\mathbf{X},\mathbf{Z},\boldsymbol{\theta}\right) = \sum_{v=1}^{V} p(\mathbf{Y}^v|\mathbf{X},\boldsymbol{\theta}) \times \prod_i \sum_{c=1}^{\infty} \lambda_c \mathcal{N}\left(\mathbf{x}_i|\boldsymbol{\mu}_c,\mathbf{R}_c^{-1}\right), \quad \mathbf{x}_i \in \mathbf{Z}_c \tag{8}$$

Based on the iWMM (see [5]) our generative model generates multiple observations $\mathbf{Y}^{\mathcal{V}}$ according to the following generative process:

1. Draw mixture weights $\boldsymbol{\lambda} \sim \mathrm{GEM}(\eta)$
2. For each cluster $c = 1, \cdots, \infty$
   (a) Draw precision $\mathbf{R}_c \sim \mathcal{W}(\mathbf{S}^{-1}, v)$
   (b) Draw mean $\boldsymbol{\mu}_c \sim \mathcal{N}(\mathbf{u}, (r\mathbf{R}_c)^{-1})$

[1]The possibly low-dimensional latent mixture model allows us to summarize the properties of the high-dimensional clusters (or density manifolds) describing the data. The number of manifolds, as well as the shape and dimension of each manifold is automatically inferred.

3. For each view $v = 1, \cdots, \mathcal{V}$
   (a) For each observation $n = 1, \cdots, N_v$
      i. Draw latent assignment $z_{nv} \sim \text{Mult}(\boldsymbol{\lambda})$
      ii. Draw latent coordinates $\mathbf{x}_{nv} \sim \mathcal{N}(\boldsymbol{\mu}_{z_{nv}}, \mathbf{R}_{z_{nv}}^{-1})$
4. For each view $v = 1, \cdots, \mathcal{V}$
   (a) For each observed dimension $d = 1, \cdots, D_v$
      i. Draw function $\mathbf{f}_d^v \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}^v)$
5. For each view $v = 1, \cdots, \mathcal{V}$
   (a) For each observed dimension $d = 1, \cdots, D_v$
      i. For each observed dimension $d = 1, \cdots, D_v$
         A. Draw feature $y_{nd}^v \sim \mathcal{N}\left(\mathbf{W}^v f_d^v(\mathbf{x}_{nv}), \beta^{-1}\right)$

We define the dimensionalities of our variables as:

- K: real number of clusters
- Q: dimensionality of the Latent Space
- D_v: dimensionality of the input data in the v-th view
- $\boldsymbol{\lambda} \in \mathbb{R}^{K \times 1}$
- $\mathbf{R}_c \in \mathbb{R}^{Q \times Q}$
- $\boldsymbol{\mu}_c \in \mathbb{R}^{Q \times 1}$

References

[1] Mauricio A. Álvarez, Lorenzo Rosasco, and Neil D. Lawrence. Kernels for vector-valued functions: A review. *Found. Trends Mach. Learn.*, 4(3):195–266, March 2012.

[2] Tomoharu Iwata, Tsutomu Hirao, and Naonori Ueda. Unsupervised cluster matching via probabilistic latent variable models, 2013.

[3] Tomoharu Iwata, Tsutomu Hirao, and Naonori Ueda. Probabilistic latent variable models for unsupervised many-to-many object matching. *Information Processing and Management*, 52(4):682 – 697, 2016.

[4] Neil D. Lawrence and Raquel Urtasun. Non-linear matrix factorization with gaussian processes. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 601–608, New York, NY, USA, 2009. ACM.

[5] Zoubin Ghahramani Tomoharu Iwata, David Duvenaud. Warped mixtures for nonparametric cluster shapes. In *29th Conference on Uncertainty in Artificial Intelligence*, pages 311–319, 2013.