Bayesian Multiview Manifold Learning Applied to Hippocampus Shape and Clinical Score Data

Giorgos Sfikas[1(✉)] and Christophoros Nikou[2]

[1] Computational Intelligence Laboratory,
Institute of Informatics and Telecommunications, NCSR "Demokritos",
Athens, Greece
sfikas@iit.demokritos.gr
[2] Department of Computer Science and Engineering,
University of Ioannina, Ioannina, Greece
cnikou@cs.uoi.gr

Abstract. In this paper, we present a novel Bayesian model for manifold learning, suitable for data that are comprised of multiple modes of observations. Our data are assumed to be lying on a non-linear, low-dimensional manifold, modelled as a locally linear structure. The manifold local structure and the manifold coordinates are latent stochastic variables that are estimated from a training set. Through the use of appropriate prior distributions, neighbouring points are constrained to have similar manifold coordinates as well as similar manifold geometry. A single set of latent coordinates is learned, common for all views. We show how to solve the model with variational inference. We also exploit the multiview aspect of the proposed model, by showing how to estimate missing views of unseen data. We have tested the proposed model and methods on medical imaging data of the OASIS brain MRI dataset [6]. The data are comprised of four views: two views that correspond to clinical scores and two views that correspond to hippocampus shape extracted from the OASIS MR images. Our model is successfully used to map the multimodal data to probabilistic embedding coordinates, as well as estimate missing clinical scores and shape information of test data.

1 Introduction

Using low-dimensional structures to model data is a widely used and studied practice in the context of a vast range of problems. Methods that deal with low-dimensional modeling may assume either a linear or a non-linear structure of data. Linear models like principal component analysis (PCA), are naturally simpler and more straightforward in their application. Non-linear models on the other hand, allow a more accurate and flexible representation of the data structure. A wealth of models exists for non-linear dimensionality reduction, or otherwise known as (non-linear) manifold learning [4].

Manifold modeling techniques typically treat data and model parameters as deterministic (in the sense of being non-probabilistic). The linear PCA algorithm, as well as the closely related canonical correlation analysis (CCA), have

H. Müller et al. (Eds.): MCV/BAMBI 2016, LNCS 10081, pp. 160–171, 2017.
DOI: 10.1007/978-3-319-61188-4_15

been shown to be expressible as equivalent probabilistic models [1,2]. In terms of probabilistic PCA/CCA, the latent variable acts as an embedding coordinate vector. In [1], a graphical model was introduced that was proved to be equivalent to CCA. Both models can be solved with Expectation-Maximization (EM) [2]. Interestingly, in both probabilistic models a single set of normally distributed latent variables is defined, while they differ in that probabilistic CCA defines two, instead of a single one, sets of observed variables (*views* in CCA parlance) and two sets of projections from the common latent space to the view spaces.

Non-linear manifold learning schemes are typically deterministic in the way they treat data and parameters, with few extensions to probabilistic models. One exception to this rule is the recently proposed locally linear latent variable model (LL-LVM) [7]. LL-LVM employs a probabilistic graphical model to describe observations, manifold coordinates and tangents [7]. The manifold is defined in terms of a patchwork of locally linear subspaces, that are represented using the tangent space to each point. The model is solved with standard variational inference (VI) [2]. LL-LVM is closely related to the Gaussian Process Latent Variable Model (GP-LVM) [5].

Manifold modeling has been extensively used in medical imaging in the recent years [4,10]. In [4], manifolds are learned on sets of brain structural MR images. New brain images are projected onto the manifold and a regression model is proposed, linking the MRI structure with subject clinical scores. In [10], an embedding is learned over brain MRIs that is used for atlas propagation. Registering one image to another is broken down to a set of subsequent registrations, following the shortest path over the learned manifold.

In this paper, we present a novel Bayesian model for manifold learning that can handle multiple observed views. Views here are to be understood as different sets of observations or different modes of measurements per observed datum, with each view typically having different dimensionality and statistics. This setup is in contrast to standard manifold learning techniques that typically assume a single source of observations and a non-probabilistic setup. In the same way that probabilistic CCA can be viewed as probabilistic PCA with multiple outputs [1], hence generalizing linear manifold learning to multiple views, the current model extends the LL-LVM model of [7]. Under this consideration we name the proposed model multiview locally linear latent variable model (MLL-LVM), underpinning its relation with LL-LVM. The proposed model is solved using variational inference. We show that a set of useful operations like out-of-sample extensions, predicting missing views, and generating new observations given the embedding coordinates, are all naturally defined in terms of the Bayesian model.

In a nutshell, from a theoretical point of view the novel characteristics of the proposed model compared to LL-LVM [7] are: (a) An extension of the model to handle more than a single view/mode of observation, (b) derivation of VI updates for the extended model and (c) derivation of the required formulae to estimate missing views given observed views. Note that the latter point is only compatible with the present model and not with LL-LVM or other single-view models, since it applies only to a scenario where we have more than one view.

The proposed model is successfully applied in a medical imaging context, where various shape data and clinical ratings from a set of Alzheimer's Disease (AD) and controls are used to learn common latent manifold coordinates. Brain MR images are used to extract shape information about subject left and right hippocampi, which alongside clinical scores make up the set of observed views. All views, despite being heterogenous and following different statistics, are hence treated in a unified manner with our model. Also importantly, all estimates (out-of-sample coordinates, missing views) are computed in the form of posterior probability density functions, since the model is fully Bayesian.

The remainder of this paper is structured as follows. In Sect. 2, we present the proposed multi-view Bayesian model, we show how to solve it using variational inference, and show how to estimate missing views. In Sect. 3, we train our model on the OASIS data set and estimate unknown clinical scores and hippocampus shapes given the observed subject views. In Sect. 4, we discuss final conclusions and thoughts about the perspective of the proposed work.

2 Methods

The basis of the proposed method is a novel Bayesian model, trained on a set multimodal data of N observations and V views. After training, the model can be used on new data in order to estimate one or more of their views that may be missing. In this section, we present the proposed observational model, we show how to solve it with VI, and derive the formulae required to predict missing views.

2.1 Generative Model

Observed data: The input to our model is a set of observations y and a graph $\mathcal{G}$. Each observation y^n of the observation set y is itself a set of V observed views $y^n = \{y_1^n, y_2^n, \cdots, y_V^n\}$, with each view being a set of elements with corresponding per-view dimensionality $d_{y1}, d_{y2}, \cdots, d_{yV}$. N observed elements correspond to each of the V views, and for view v we have $\{y_v^1, y_v^2, \cdots, y_v^N\}$.

The graph $\mathcal{G}$ contains one node for each observation, and an edge exists between nodes (n,m) if and only if y^n and y^m are neighbours. A symmetric $N \times N$ adjacency matrix G corresponds to the graph structure of $\mathcal{G}$, with $G=[\eta^{nm}]$. Element $[\eta^{nm}]_{n=1..N,m=1..N}$ is equal to one if observations n and m are neighbours, otherwise it is zero.

In the assumed application context, each patient appears as a single observation y^n for the model, and each view corresponds to a different type of measurement for the patient. For example, for the n^{th} patient, y_1^n may contain brain MRI T1 data, y_2^n a scalar clinical rating and y_3^n a brain connectogram. Patients that are similar enough with respect to the available measurements are recorded as neighbours in $\mathcal{G}$.

The graphical representation for the proposed generative model can be examined in Fig. 1. Note that a single set of embedding coordinates x are defined,

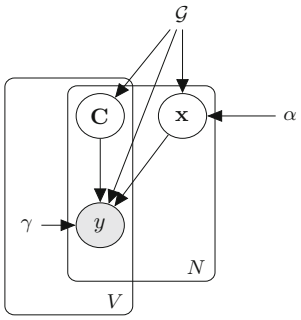


Fig. 1. The graphical model for the proposed MLL-LVM. V views are assumed for N observed data points. The latent variables x are embedding coordinates, common for all views. Latent variables C model the relation of the embedding coordinates x with each separate observed view. $\mathcal{G}$ is the fixed neighbourhood structure. γ and α are deterministic parameters that control the form of the likelihood function and the form of the prior on latent embedding coordinates respectively.

common for all views, while manifold geometry C and observations y are view-specific. In terms of the graphical model, this is the basic difference between the proposed model and LL-LVM [7]. The latter can be seen as a special case of our model, for $V = 1$.

Assumed distributions and relations with latent variables: Embedding coordinates can be concatenated to a single vector $x = [x^{1T} x^{2T} \cdots x^{NT}]^T$, where $x \in \mathcal{R}^{d_x N}$. The prior on latent variables x constraints elements that are neighbours to have embedding coordinates that lie close to each other:

$$\log p(x|G, \alpha) = -\frac{1}{2} \sum_{n=1}^{N} (\alpha \|x^n\|^2 + \sum_{m=1}^{N} \eta^{nm} \|x^n - x^m\|^2) + const. \quad (1)$$

The set of linear projections that correspond to the v^{th} view can be concatenated to a single matrix $C_v = [C_v^1 C_v^2 \cdots C_v^N]$, where $C_v \in \mathcal{R}^{d_{y_v} \times d_x N}$. For all sets of linear maps C_v, a prior is defined that constrains neighbouring maps to be close to each other in the sense of the Frobenius norm:

$$\log p(C_v|G) = -\frac{\epsilon}{2} \|\sum_{n=1}^{N} C_v^n\|_F^2 - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \eta^{nm} \|C_v^n - C_v^m\|_F^2 + const. \quad (2)$$

where ϵ is set to a constant, small value. Local manifold tangents of neighbouring points are equivalently constrained to be similar, favouring smooth solutions with low-curvature for all views.

Observed views are assumed to be conditionally independent given x. Hence the model likelihood is defined as the sum of V terms, each corresponding to a different view:

$$\log p(y|C,x,\gamma,G) = \sum_{v=1}^{V} \log p(y_v|C_v,x,\gamma_v,G) \tag{3}$$

where $\gamma = \gamma_1, \gamma_2, \cdots, \gamma_V$ is a set of scale parameters. The log-likelihood component specific to each view is given by:

$$\log p(y_v|C_v,x,\gamma_v,G) = -\frac{\epsilon}{2}\|\sum_{n=1}^{N} y_v^n\|^2 - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \eta^{nm}\gamma_v \|\Delta_{y_v}^{m,n} - C_v^n \Delta_x^{m,n}\|^2 + const. \tag{4}$$

where $\Delta_x^{m,n} = x^m - x^n$ and $\Delta_{y_v}^{m,n} = y_v^m - y_v^n$. The double-summation term in the above equation encodes the assertion that $C_v^n \Delta_{x_v}^{m,n} \approx \Delta_{y_v}^{m,n}$, or that the assumed manifolds are locally linear.

Following [7], it is straightforward to show that x and y_v $\forall v \in [1..V]$ are normally distributed, and C_v $\forall v \in [1..V]$ follow the matrix-normal distribution. More specifically,

$$x|G,\alpha \sim \mathcal{N}(0, \Sigma_x^0), \tag{5}$$

$$C_v|G \sim \mathcal{MN}(0, I_{d_{y_v}}, \Sigma_{C_v}^0), \forall v \in [1..V], \tag{6}$$

$$y_v|C_v,x,\gamma_v,G \sim \mathcal{N}(\mu_{y_v}^0, \Sigma_{y_v}^0), \forall v \in [1..V], \tag{7}$$

where for $\Sigma_x^{-1} = \alpha I_{d_x N} + 2L \otimes I_{d_x}$ and $L = diag(G1_N) - G$ is the graph Laplacian matrix of G. The prior covariance ${\Sigma_C^0}^{-1} = \epsilon JJ^T + 2L \otimes I_{d_x}$ is the same for all view distributions. The likelihood parameters are ${\Sigma_{y_v}^0}^{-1} = (\epsilon 1_N 1_N^T + 2\gamma_v L) \otimes I_{d_{y_v}}$, $\mu_{y_v}^0 = \Sigma_{y_v}^0 e_v$, where $e_v = [e_v^{1T}, e_v^{1T}, \cdots, e_v^{NT}]^T \in \mathcal{R}^{d_{y_v}N}$, $e_v^n = -\sum_{m=1}^{N} \eta^{mn}\gamma_v(C_v^m + C_v^n)\Delta_x^{m,n}$.

2.2 Solution with Variational Inference

Solving the model amounts to calculating the posterior distributions for the shared coordinates x and the sets of linear projections $C_v, \forall v \in [1..V]$, as well as the non-stochastic parameters $\{\gamma_v\}_{v=1}^V$ and α. As an exact calculation of the posterior is intractable, we employ variational inference [2] to approximate it. In VI, the model is solved by iterating between optimizing the Kullback-Leibler divergence $KL(q||p)$ of the posterior estimate q and the actual posterior p, and optimizing a lower bound $\mathcal{L}$ of the model likelihood. In our model, the variational lower bound $\mathcal{L}$ is defined as

$$\mathcal{L}(q,C,x,\gamma,\alpha) = \int_{C,x} q(C,x) \log \frac{p(y,C,x|G,\gamma,\alpha)}{q(C,x)} dCdx. \tag{8}$$

According to VI theory, the posteriors of the latent variables are estimated by taking expectations of the model joint distribution, in our case $p(C,x|\mathcal{G},\gamma,\alpha)$, over all latent variables except the one that is being computed. Formally, for the approximate posteriors $q^*(x), q^*(C_1), q^*(C_2), \cdots, q^*(C_V)$ we have

$$\log q^*(x) = <\log p(y,C,x|G,\gamma,\alpha)>_C + const. \tag{9}$$

$$\log q^*(C_v) = <\log p(y, C, x|G, \gamma, \alpha)>_{x, C_1, \cdots, C_{v-1}, C_{v+1}, \cdots, C_V} + const., \forall v \in [1..V] \quad (10)$$

Key to model tractability with VI is the fact that the log-likelihood term (Eq. 4) can be written as a quadratic function in both x and C. More specifically,

$$\log p(y|C, x, \gamma, G) = -\frac{1}{2}[x^T\{\sum_{v=1}^{V} A_v\}x - 2x^T\{\sum_{v=1}^{V} b_v\}] + Z_x, \quad (11)$$

$$= -\frac{1}{2}\sum_{v=1}^{V} Tr[\Gamma_v C_v^T C_v - 2\gamma_v C_v^T H] + Z_C \quad (12)$$

where we followed [7] in a related calculation, and Z_x,Z_C contain terms not depending on x or C respectively. Matrix A_v is of size $Nd_x \times Nd_x$ and b_v is of size $Nd_x \times 1$. Matrix Γ_v is of size $Nd_x \times Nd_x$. Hence all priors $x, C_1, C_2, \cdots, C_V$ are conjugate to the likelihood and VI is tractable.

Variational E step update of $q(x)$: Equation (9) can be further decomposed to

$$\log q^*(x) = <\log p(y|C, x, \gamma, G)>_C + \log p(x|G, \alpha) + const.$$

$$= -\frac{1}{2}\sum_{v=1}^{V}[x^T A_v x - 2x^T b_v] - \frac{1}{2}[x^T \Sigma_x^{-1(0)}] + const.$$

where we have used Eqs. (5) and (11). As the non-constant terms are quadratic in x, the approximate posterior of x is Gaussian. Thus we have $q^*(x) = \mathcal{N}(x|\mu_x, \Sigma_x)$ with

$$\Sigma_x^{-1} = \Sigma_x^{-1(0)} + \sum_{v=1}^{V} <A_v>_{C_v}, \quad (13)$$

$$\mu_x = <x> = \Sigma_x \sum_{v=1}^{V} <b_v>_{C_v} . \quad (14)$$

We also calculate the expectation $<xx^T>$, useful for some later updates,

$$<x^n x^{mT}> = \Sigma_x^{nm} + <x^n><x^m>^T, \quad (15)$$

where Σ_x^{nm} is the $(n, m)^{th}$ chunk of size $d_x \times d_x$ of this matrix. Expectations for A_v and b_v can be derived following a related calculation in [7]. We show updates for all $d_x \times d_x$-sized chunks of the $Nd_x \times Nd_x$-sized matrix A_v, and updates for all d_x-sized chunks of the Nd_x-sized matrix b_v:

$$<A_v^{nm}>_{C_v} = \gamma_v^2 \sum_{p=1}^{N}\sum_{q=1}^{N}\{[\hat{L}_v^{pq} - \hat{L}_v^{pm} - \hat{L}_v^{nq} + \hat{L}_v^{nm}]\eta^{pn}\eta^{qm}$$

$$\times <C_v^{pT} C_v^q + C_v^{pT} C_v^m + C_v^{nT} C_v^q + C_v^{nT} C_v^m>_{C_v}\}, \quad (16)$$

$$<b_v^n>_{C_v} = \gamma_v \sum_{m=1}^{N} \eta^{nm}\{<C_v^m>^T (y_v^n - y_v^m) - <C_v^n>^T (y_v^m - y_v^n)\}, \quad (17)$$

where the quantity $\hat{L}_v$ for each v is equal to $(\epsilon 11^T + 2\gamma_v L)^{-1}$.

Variational E step update of $q(C_v)$, $v \in [1..V]$*:* We decompose Eq. (10) as:

$$\log q^*(C_v) = -\frac{1}{2}\sum_{v=1}^{V} Tr[\Gamma_v C_v^T C_v - 2\gamma_v C_v^T H] + \mathcal{MN}(0, I_{d_{y_v}}, \Sigma_C^0) + const., \quad (18)$$

where we wrote the likelihood function in terms of C_v using Eq. (12). The approximate posterior distribution for the v^{th} view projection matrix C_v can thus be written as a matrix normal distribution $q^*(C_v) = \mathcal{MN}(\mu_{C_v}, I_{d_{y_v}}, \Sigma_{C_v})$ with

$$\Sigma_{C_v}^{-1} = \Sigma_{C_v}^{-1(0)} + <\Gamma_v>_x, \quad (19)$$

$$<C_v^{nT} C_v^m>_x = <C_v^n>_x^T <C_v^m>_x + d_y \Sigma_{C_v}^{nm}, \quad (20)$$

where $\Sigma_{C_v}^{nm}$ is the $(n,m)^{th}$ chunk of size $d_x \times d_x$, and C_v^n is the n^{th} chunk of the respective matrices. Also,

$$\mu_{C_v} = <C_v>_x = \gamma <H_v>_x \Sigma_{C_v}. \quad (21)$$

Finally, expectations for quantities Γ_v and H_v are given as:

$$<\Gamma_v^{nm}>_x = \gamma_v^2 \sum_{p=1}^{N}\sum_{q=1}^{N}\{[\hat{L}_v^{pq} - \hat{L}_v^{pm} - \hat{L}_v^{nq} + \hat{L}_v^{nm}]\eta^{pn}\eta^{qm}$$

$$\times <x^p x^{qT} - x^p x^{mT} - x^n x^{qT} + x^n x^{mT}>_x\}, \quad (22)$$

$$<H_v^n>_x = \sum_{m=1}^{N} \eta^{nm}(y_v^m <x^m>_x^T - y_v^m <x^n>_x^T - y_v^m <x^m>_x^T + y_v^m <x^n>_x^T). \quad (23)$$

Variational M step update of α, γ_v, $\forall v \in [1..V]$*:* In the maximization step we optimize the variation lower bound with respect to non-stochastic parameters α and γ_v, $\forall v \in [1..V]$. The update of α is identical to the one for the single-view case [7]. The update for γ_v is similar to the update for γ of [7], save that for each view it is now calculated over y_v and the statistics of C_v instead of y and C respectively.

We alternate the aforementioned E-step updates for the approximate posterior of x (Eqs. 13–17), the approximate posterior of C (Eqs. 19–23) and M-step updates until convergence.

2.3 Estimation of Missing Views for New Data

Given a previously unseen datum y^{new} for which only part of all the V views are observed, we can use the trained model to estimate the missing views. In order to do this, first we add the new datum to the training set and re-compute the E-step for the new datum only, keeping posteriors for the original trained data and deterministic parameters fixed. The new observation is added to the previous

graph structure by computing its nearest neighbours. Using the E-step equations gives us an estimate of the posterior distributions $q(x^{new})$ and $q(C^{new})$ for the new datum. These steps let us effectively project the new observation onto the manifold, a process known in the literature as out-of-sample projection [7].

The set of missing views $\hat{v}$ of y^{new} are treated also as latent variables, for which we require their approximate posterior distribution $q(y)$. Hence the joint posterior now also includes $q(\{y_v^{new}\}_{v\in\hat{v}})$, decomposed using the mean field approximation [2] into $\{q(y_v^{new})\}_{v\in\hat{v}}$. In order to estimate the posterior for missing view v, we compute the expectation of the model evidence. This is formally written as

$$logq^*(y_v^{new}) = <logp(y_v, C_v, x|G, \gamma_v, \alpha)>_{x,C_v} + const. \tag{24}$$

The above equation, combined with the likelihood formula (Eq. 4), where we have kept all observations fixed except y^{new}, gives a posterior Normal distribution $\mathcal{N}(y_v^{new}|m_v^{new}, S_v^{new})$ with statistics given by

$$S_v^{new} = (2\gamma_v \sum_{m=1}^{N} \eta^{m,new} + \epsilon)^{-1} I_{d_{y_v}} \tag{25}$$

$$m_v^{new} = S_v^{new}(2\gamma_v \sum_{m=1}^{N} \{\eta^{m,new}[y^n + 1/2(<C_v^{new}> + <C_v^m>) <\Delta_x^{new,m}>]\} - \frac{\epsilon}{2} \sum_{m=1}^{N} y_v^m) \tag{26}$$

In summary, in order to estimate the missing views of an unseen datum y^{new} we iterate through the E- step updates for the approximate posterior of x^{new} (Eqs. 13–17), the approximate posterior of C^{new} (Eqs. 19–23) and the approximate posterior for y^{new} (Eqs. 25 and 26), keeping fixed the deterministic model parameters and all other point posteriors[1].

3 Experiments

3.1 Dataset

We have experimented with data from the OASIS database [6]. In our evaluation we have included the 198 subjects aged 60 or more found in the cross-sectional set of OASIS. 100 of these subjects have been diagnosed with very mild to moderate AD. The rest of the subjects are used as controls. We have used in total 4 views/modes for each subject. The two first views are the clinical scores Minimental State Exam (MMSE) and Clinical Dementia Rating (CDR). The other two views correspond to shape information for the left and the right hippocampus of each subject respectively. The volumetric characteristics of the hippocampus are known to be correlated with the advance of AD [9].

In order to create the shape views, we have first segmented the OASIS T1-modulated MR images with Freesurfer [3]. We have then computed deformation

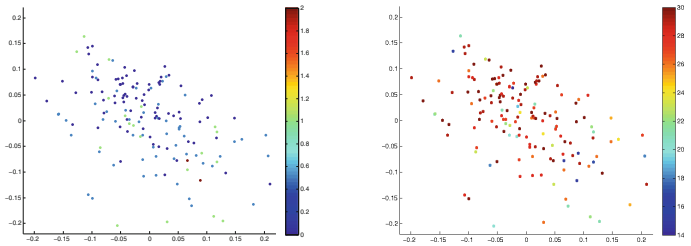
[1] MATLAB code that implements training and missing view estimation using the presented model is available at https://github.com/sfikas/mll-lvm/.

fields for each volume, given as the output of matching with an in-set template image. The template, one for each hippocampus, was chosen as the medoid image within the sets of left and right hippocampi. The medoid was taken with respect to a distance metric that is analogous to the total magnitude of the deformation field required to perform a matching non-rigid deformation between volumes [4]. Deformation fields are subsampled to 25% of the original length of each axis, resulting in $11 \times 15 \times 8$ and $12 \times 14 \times 9$-sized fields of $\mathcal{R}^3$ vectors. These volumes are further vectorized into descriptors of 3960 and 4536 dimensions respectively.

We partitioned our dataset into a training set and a test set. The training set was used to learn the parameters of our model, and the test set was used to evaluate the model. We assigned the first 80% of the data (first in the sense of lexicographical OASIS id order) to the training set, and the rest to the test set. Mean clinical scores for both training and test differ by less than 10^{-2} (CDR) and 0.5 (MMSE) to the respective statistics of the full set (CDR = 0.2 and MMSE = 27 respectively).

3.2 Experimental Setup

Before proceeding to any tests we computed the neighbourhood structure G. To this end, a distance ζ^{nm} for all pairs of subjects (n, m) in the training set was first calculated. This distance fusions view-specific distances are $\zeta^{nm} = \sum_{v=1}^{V} \zeta_v^{nm}$. The view-specific distances are Euclidean distances over normalized view data. We computed embeddings with $d_x = 2$, which can be examined in Fig. 2, along with an overlay of clinical score values.



(a) (b)

Fig. 2. Computed embedding given clinical score and shape information of the training data ($N = 158$ subjects). Approximate posterior mean values for x are shown, per subject. Point colours correspond to (a) CDR scores (b) MMSE scores. (Color figure online)

Estimating clinical scores given shape data: In the first experiment, we assumed that only shape information (views 3 and 4) was known for the test set subjects. For each test subject, we estimate the posterior distribution of its clinical scores (views

Table 1. Clinical score estimation given hippocampus shape data. We show the moments of the Gaussian posterior distribution of the clinical scores (mean ± st.deviation) for all the 40 test set subjects. Significant statistical correlation is reported between estimate means and ground truth for both CDR and MMSE (bottom row).

OASIS id	CDR		MMSE	
	Estimate	Actual	Estimate	Actual
Control subjects				
363	0.16 ± 0.32	0.00	28.3 ± 4.0	30.0
365	0.41 ± 0.32	0.00	24.4 ± 4.0	30.0
369	0.23 ± 0.05	0.00	27.3 ± 0.7	28.0
371	0.22 ± 0.07	0.00	27.6 ± 0.9	30.0
373	0.16 ± 0.32	0.00	28.4 ± 4.0	30.0
374	0.16 ± 0.32	0.00	28.3 ± 4.0	29.0
380	0.16 ± 0.32	0.00	28.3 ± 4.0	29.0
382	0.16 ± 0.32	0.00	27.8 ± 4.0	28.0
388	0.27 ± 0.09	0.00	26.8 ± 1.1	29.0
390	0.16 ± 0.32	0.00	27.3 ± 4.0	28.0
398	0.21 ± 0.10	0.00	27.9 ± 1.2	29.0
399	0.41 ± 0.32	0.00	26.9 ± 4.0	29.0
400	0.23 ± 0.07	0.00	27.0 ± 0.8	30.0
402	0.22 ± 0.11	0.00	27.4 ± 1.3	29.0
404	0.18 ± 0.10	0.00	28.0 ± 1.2	28.0
405	0.16 ± 0.32	0.00	28.4 ± 4.0	30.0
411	0.41 ± 0.32	0.00	27.4 ± 4.0	26.0
AD subjects				
418	0.66 ± 0.32	1.00	20.8 ± 4.0	20.0
422	0.16 ± 0.09	0.50	27.8 ± 1.1	29.0
423	0.16 ± 0.32	0.50	26.3 ± 4.0	18.0
424	0.28 ± 0.23	1.00	25.9 ± 2.8	15.0
425	0.41 ± 0.13	1.00	26.1 ± 1.6	22.0
426	0.24 ± 0.19	0.50	28.0 ± 2.3	24.0
428	0.66 ± 0.32	1.00	24.4 ± 4.0	29.0
430	0.41 ± 0.32	0.50	23.4 ± 4.0	25.0
432	0.67 ± 0.32	0.50	26.3 ± 4.0	30.0
438	0.16 ± 0.32	1.00	27.9 ± 4.0	23.0
440	0.32 ± 0.10	0.50	27.4 ± 1.2	29.0
441	0.18 ± 0.10	0.50	27.7 ± 1.2	28.0
445	0.29 ± 0.23	1.00	28.3 ± 2.8	20.0
446	0.41 ± 0.32	1.00	24.8 ± 4.0	23.0
447	0.41 ± 0.32	1.00	22.3 ± 4.0	17.0
449	0.23 ± 0.12	0.50	27.0 ± 1.5	26.0
451	0.32 ± 0.19	0.50	26.0 ± 2.3	27.0
452	0.41 ± 0.32	0.50	28.4 ± 4.0	29.0
453	0.41 ± 0.32	0.50	26.4 ± 4.0	24.0
454	0.16 ± 0.32	0.50	28.4 ± 4.0	27.0
455	0.25 ± 0.19	1.00	27.9 ± 2.3	22.0
456	0.16 ± 0.23	0.50	28.1 ± 2.8	29.0
457	0.16 ± 0.32	0.50	27.9 ± 4.0	23.0
	r	p-value	r	p-value
corr.coeff.	**0.43**	**0.006**	**0.44**	**0.004**

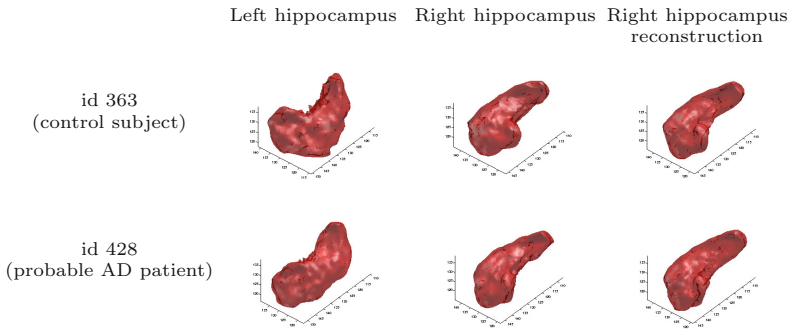


Fig. 3. Estimation of the right hippocampus given the left hippocampus shape data. We show reconstructions for a probable AD patient as well as for a control subject. Left column: Left hippocampus shapes on which the estimate is conditioned. Middle column: Right hippocampus ground truth data, shown here for comparison. Right column: Right hippocampus posterior mean, calculated with the proposed algorithm.

1 and 2), using the method described in Sect. 2.3. In order to fit the new datum onto the neighbourhood structure $\mathcal{G}$, we assigned neighbours according to a distance threshold chosen so that the mean number of neighbours is closest to $k = 5$. Data without neighbours are assigned their nearest neighbour to their neighbourhood.

We can see an overview of the results in Table 1. Note that all estimates are computed as posterior probability density functions. The moments of the posterior Gaussians are reported for all test set subjects, alongside with the ground truth values. The correlation coefficient between estimate mean values and ground truth is also computed. The results clearly indicate that there is statistically significant correlation between estimates and actual values. This result agrees with the fact, known from the related literature, that hippocampus shape and the progression of neurogenerative diseases such as AD are correlated [4], hence validating the usefulness of the proposed MLL-LVM model. Furthermore, our results come all in the form of pdfs, measuring estimation uncertainty in a natural and principled manner, in line with the model assumptions.

Estimating shape data given shape data: We have experimented with using the proposed model to calculate an estimate of missing shape data given existing shape data. To this end, we have trained our model with the set of left and right hippocampus shape data. We have assumed that the test set now contains information only about the left hippocampus. In other words, for the 40 images of the test we now assume only the right hippocampus shape view as available, while the left hippocampus shape is missing. We have calculated posterior distribution approximations of the right hippocampus shape given the model and the observed test left hippocampus. We show visual results in Fig. 3. The results show that the estimate right hippocampus is reasonably similar to the ground truth right hippocampus. Again, estimates are computed in the form of pdfs. Here however we show only mean volumes, due to visualization constraints.

4 Conclusion

We have presented a novel Bayesian model for manifold learning, and tested it on a set of medical data. The model assumes that observed values are comprised of a number of heterogenous views. The solution has been shown to be feasible with approximate inference. The proposed model also allows new test data to have one or more of their views missing; we have shown how to compute estimates of these views, in a manner that is consistent with the definition of model. All estimates are computed in the form of posterior probability distributions.

In perspective, the model can be used with any number and combination of modes. Other imaging modalities could be used as modes, or other descriptors that characterize other parts of the brain. Extensions of the probabilistic model could also be considered. For example, replacing the binary neighbourhood graph with a more flexible alternative could be envisaged, in the spirit of the continuous line process model of [8].

References

1. Bach, F.R., Jordan, M.I.: A probabilistic interpretation of canonical correlation analysis. Technical Report 688, Department of Statistics, University of California, Berkeley (2005)
2. Bishop, C.M.: Pattern Recogn. Mach. Learn. Springer, New York (2006)
3. Fischl, B.: FreeSurfer. Neuroimage **62**(2), 774–781 (2012)
4. Gerber, S., Tasdizen, T., Fletcher, P.T., Joshi, S., Whitaker, R.: Alzheimers disease neuroimaging initiative: manifold modeling for brain population analysis. Med. Image Anal. **14**(5), 643–653 (2010)
5. Lawrence, N.D.: Gaussian process latent variable models for visualisation of high dimensional data. Adv. Neural Inf. Process. Syst. **16**(3), 329–336 (2004)
6. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. **19**(9), 1498–1507 (2007)
7. Park, M., Jitkrittum, W., Qamar, A., Szabó, Z., Buesing, L., Sahani, M.: Bayesian manifold learning: the locally linear latent variable model (LL-LVM). In: Advances in Neural Information Processing Systems, pp. 154–162 (2015)
8. Sfikas, G., Nikou, C., Galatsanos, N., Heinrich, C.: Spatially varying mixtures incorporating line processes for image segmentation. J. Math. Imaging Vis. **36**(2), 91–110 (2010)
9. Wang, L., Swank, J.S., Glick, I.E., Gado, M.H., Miller, M.I., Morris, J.C., Csernansky, J.G.: Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging. Neuroimage **20**(2), 667–682 (2003)
10. Wolz, R., Aljabar, P., Hajnal, J.V., Hammers, A., Rueckert, D.: Alzheimer's disease neuroimaging initiative: LEAP: learning embeddings for atlas propagation. NeuroImage **49**(2), 1316–1325 (2010)