2014 International Symposium on Nonlinear Theory and its Applications
NOLTA2014, Luzern, Switzerland, September 14-18, 2014

Latent Variable Models Combined with Clustering

Daming Li†, Quan Wang‡, Fei Wang‡, and Xuan Wang‡

† Beijing Institute of Spacecraft System Engineering, Beijing, China
No. 104 Friendship Rd, Beijing Areoscpace City, China
‡ Xi'an Jiaotong University, Xi'an, Shaanxi Province, China
No. 28 West Xianning Rd, Xi'an, Shaanxi Province, China
Email: damli@163.com, bhxspring@stu.xjtu.edu.cn, wfx@mail.xjtu.edu.cn, xwang.cv@gmail.com

Abstract– Given a known dataset to learn a latent variable model, previous methods fail to focus on how to get labeled samples, some just choosing them randomly. As a result, it is likely that those methods gain models with a very limited representation capability. In this article, we propose a novel method in which we select representative samples to be labeled ones for latent variable models. To this end, the G-means clustering algorithm is adopted to automatically cluster latent variables and obtain those corresponding representative samples. We learn the Gaussian Process Latent Variable Model (GPLVM) and the Constrained Latent Variable Model (CLVM) respectively combined with the G-means, and compare them to those without clustering, in the context of non-rigid 3D reconstruction from monocular images. Our experimental results show that our methods present a more powerful representation capability.

1. Introduction

Latent variable techniques provide the substantial insight into the inherent structure of complex data. Latent Variable Models (LVMs) are commonly applied to learn the appropriate models from training data, since they can project the high-dimensional data into a low-dimensional space and represent unknown mappings by learning. The authors in [1] propose a generative LVM for dependency parsing. [2] presents a scalable parallel framework for efficient inference in LVMs over streaming web-scale data. [3] introduces a method for Gaussian Process Classification using LVMs trained with discriminative priors over the latent space. [4] combines sparse approximations with GPLVM [5,6] and achieves better quality on a benchmark visualization data set. GPLVM is used to deal with the trouble of high dimensionality in the space of possible deformations in [7]. In [8], the GPLVM is used to learn local deformation models from data. [9] has proposed a new model CLVM, and made the GPLVM as a comparison. CLVM [9] really gains improvements over GPLVM on the problem of non-rigid 3D reconstruction from monocular images.

Although these LVMs perform well, we think that the generating LVMs will be more effective if we are able to select representative samples to be labeled samples. Given a dataset with vast samples, there is a need to select representative samples with an appropriate method as labeled ones since there must be many whose intrinsic attributes are similar. If we just choose labeled samples randomly, the probability will increase that some whose intrinsic attributes are similar are chosen as labeled samples. If all the labeled samples have similar attributes, which of course is the worst situation, the model will only adapt to prediction in a very small range. Clustering can be used to get labeled samples. G-means [10] assumes that subsets of data follow the Gaussian distribution, which we think is applicative to the data we will use in the latent space. In fact, due to the advantage of low dimensionality in latent space, the G-means expenses little additional cost of time. Therefore, it's a good choice to make use of the G-means to obtain labeled samples.

More specifically, we make use of the G-means to cluster latent variables obtained by Principal Component Analysis (PCA) or other dimensionality reduction techniques. Those latent variables closest to clustering centers are taken to be labeled samples. We then learn the LVMs using these labeled samples. This method can ensure that labeled samples are as far different from each other as possible in the latent space. Consequently, the resulting model can represent a wider range of shapes as to the problem of monocular non-rigid 3D shape recovery.

By introducing the G-means to LVMs, we can improve the performance of LVMs. We have demonstrated the effectiveness of the LVMs combined with clustering on synthetic data in the context of non-rigid 3D shape recovery from monocular images. The LVMs combined with clustering do outperform the original LVMs.

2. LVMs combined with clustering

2.1. Latent Variable Models

Given a dataset full of varieties of available data $\boldsymbol{Y}$, the first thing is to get latent variables $\boldsymbol{X}$. Here, $\boldsymbol{Y}$ is a matrix which contains all the obtained samples in a high dimensional space and $\boldsymbol{X}$ is a matrix which contains all the corresponding latent variables in a low dimensional space. We use PCA to obtain the latent variables matrix $\boldsymbol{X}$.

2.1.1. GPLVM

First, n training pairs of samples and corresponding latent variables $[(\boldsymbol{x}_1, \boldsymbol{y}_1), \ldots, (\boldsymbol{x}_n, \boldsymbol{y}_n)]$ are randomly chosen from $\boldsymbol{Y}$ and $\boldsymbol{X}$. And then, the goal is to predict the output $\boldsymbol{y}' = M(\boldsymbol{x}')$ from a novel input $\boldsymbol{x}'$. Let $\boldsymbol{Y}_L = [\boldsymbol{y}_1, \dots, \boldsymbol{y}_n]^T$, $\boldsymbol{X}_L = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n]^T$. The likelihood

$$p(\boldsymbol{Y}_L|\boldsymbol{X}_L, \Theta) = \frac{1}{\sqrt{(2\pi)^{ND}|\boldsymbol{K}|^D}} \exp\left(-\frac{1}{2} tr\left(\boldsymbol{K}^{-1}\boldsymbol{Y}_L\boldsymbol{Y}_L{}^T\right)\right) \quad (1)$$

where $\boldsymbol{K}$ is the kernel matrix whose elements are defined by the covariance function, $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Here, we take this function to be the sum of a RBF and a noise term:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \theta_1 \exp\left(-\frac{\theta_2}{2}(\boldsymbol{x}_i - \boldsymbol{x}_j)^T(\boldsymbol{x}_i - \boldsymbol{x}_j)\right) + \theta_3 \quad (2)$$

GPLVM is learnt by maximizing $p(\boldsymbol{Y}_L|\boldsymbol{X}_L, \Theta)p(\Theta)$ with respect to Θ.At reference, given a latent variable $\boldsymbol{x}'$, the mean prediction can be expressed as

$$\mu(\boldsymbol{x}') = \boldsymbol{Y}_L{}^T\boldsymbol{K}^{-1}\boldsymbol{k}(\boldsymbol{x}') \quad (3)$$

where $\boldsymbol{k}(\boldsymbol{x}') = [k(\boldsymbol{x}', \boldsymbol{x}_1), \dots, k(\boldsymbol{x}', \boldsymbol{x}_n)]^T$. $\boldsymbol{y}'$ is taken to be this mean prediction.

2.1.2. CLVM

The CLVM explicitly imposes equality or inequality constraints on the model's output during learning. Actually, m unlabeled latent variables are randomly chosen from $\boldsymbol{X}$ to be used to impose constraints.

Let $\boldsymbol{X}_{\mathrm{U}} = [\boldsymbol{x}_1^{\mathrm{U}}, \dots, \boldsymbol{x}_{\mathrm{m}}^{\mathrm{U}}]^{\mathrm{T}}$, denoting m unlabeled latent variables. The latent variable model can be written as

$$\boldsymbol{y}' = \boldsymbol{W}\boldsymbol{\phi}(\boldsymbol{x}'). \quad (2)$$

The learning can be formulated as the constrained optimization problem

$$\min_{\boldsymbol{W}} \frac{1}{2}||\boldsymbol{W}\boldsymbol{\phi}(\boldsymbol{X}_L) - \boldsymbol{Y}_L||_F^2 + \frac{\gamma}{2}||\boldsymbol{W}||_F^2 \quad (3)$$

$$\text{s.t.} \quad \mathrm{C}(\boldsymbol{W}\boldsymbol{\phi}(\boldsymbol{X}_U)) = 0, \text{or}, \mathrm{D}(\boldsymbol{W}\boldsymbol{\phi}(\boldsymbol{X}_U)) \leq 0,$$

where $\boldsymbol{\phi}(\boldsymbol{X}_L) = [\boldsymbol{\phi}(\boldsymbol{x}_1), \dots, \boldsymbol{\phi}(\boldsymbol{x}_n)]$.

By solving this problem with Lagrange multiplier method and introducing kernel function, we can finally get the prediction for an input $\boldsymbol{x}^*$ in closed-form as

$$\boldsymbol{y}^* = \boldsymbol{A}\boldsymbol{K}_{:,*} \quad (4)$$

with

$$\boldsymbol{A} = \left[\boldsymbol{M} - \textstyle\sum_u \boldsymbol{G}_u^T \boldsymbol{\lambda}_u \boldsymbol{K}_{u,:}\right]\boldsymbol{B}^{-1}, \boldsymbol{Z} = [\boldsymbol{\phi}(\boldsymbol{X}_L), \boldsymbol{\phi}(\boldsymbol{X}_U)], \quad (5)$$

$$\boldsymbol{B} = \boldsymbol{K}_{:,L}\boldsymbol{K}_{L,:} + \gamma\boldsymbol{K}_{:,:}, \qquad \mathbf{M} = \mathbf{Y}_L\boldsymbol{K}_{L,:},$$

where $\mathbf{K}_{:,*} = [\boldsymbol{K}_{*,L}, \boldsymbol{K}_{*,U}]^T$, $\boldsymbol{K}_{:,:} = \boldsymbol{Z}\boldsymbol{Z}^T = \begin{bmatrix}\boldsymbol{K}_{L,L} & \boldsymbol{K}_{L,U} \\ \boldsymbol{K}_{U,L} & \boldsymbol{K}_{U,U}\end{bmatrix}$.

2.2. The G-means Algorithm

The G-means algorithm [10] circularly runs K-means with initial cluster centroid positions, whose data have undergone a statistical test, till all the subsets of data follow Gaussian distribution. Hence, at the heart of the G-means algorithm is the statistical test for the hypothesis that a subset of data follows Gaussian distribution.

The statistical test is a process in which decisions are made whether the cluster should be split into two sub-clusters or not. Anderson-Darling statistic test is used. Assuming that $\boldsymbol{c}$ is a center whose cluster is about to undergo the statistical test, the algorithm of the statistical test is as follows.

1. Initialize two centers with $\boldsymbol{c} \pm \boldsymbol{m}$, where $\boldsymbol{m} = \boldsymbol{s}\sqrt{2\lambda/\pi}$, which is obtained by doing PCA on X, data of the center $\boldsymbol{c}$, $\boldsymbol{s}$ being the main principal component corresponding to eigenvalue λ.
2. Run k-means on the above two centers in X and get two new centers, $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$.
3. Let vector $\boldsymbol{v} = \boldsymbol{c}_1 - \boldsymbol{c}_2$ and then project X onto $\boldsymbol{v}$: $x' = < \boldsymbol{x}, \boldsymbol{v} > /||\boldsymbol{v}||^2$. Transform X' into Y so that it has mean 0 and variance 1.
4. Let $z = \mathrm{F}(y)$, where F is the $N(0,1)$ cumulative distribution function. The Anderson-Darling statistic is

$$A^2(Z) = -\frac{1}{n}\sum_{i=1}^{n}(2i-1)[\log z_i + \log(1 - z_{n+1-i})] - n \quad (8)$$

and the corrected one is

$$A_*^2(Z) = A^2(Z)(1 + 4/n - 25/n^2) \quad (9)$$

5. If $A_*^2(Z)$ is lower than the critical value, keep the original center $\boldsymbol{c}$; if not, keep $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ instead of $\boldsymbol{c}$.

2.3. LVMs combined with clustering for Non-rigid 3D Reconstruction

The essence of LVMs combined with clustering is to take advantage of the G-means to automatically obtain representative samples as labeled samples for LVMs. The procedure of the LVMs combined with clustering for non-rigid 3D reconstruction is as follows:

1. Given a dataset $\boldsymbol{Y}$ representing 3D shapes, randomly choose M test samples $\boldsymbol{Y}_T$. Perform PCA on the rest samples $\boldsymbol{Y}_R$ to get latent variables $\boldsymbol{X}$. $\boldsymbol{X}_R$ corresponds to $\boldsymbol{Y}_R$. $\boldsymbol{X}_T$ contains test latent variables.
2. Cluster $\boldsymbol{X}_R$ with G-means according to subsection 2.2. Choose the latent variables closest to the clustering centers in $\boldsymbol{X}_R$ to be N labeled latent variables $\boldsymbol{X}_L$, the corresponding $\boldsymbol{Y}_L$ to be labeled samples. Choose V unlabeled latent variables $\boldsymbol{X}_U$ randomly from the remaining latent variables specially for the CLVM.
3. Learn GPLVM and CLVM according to Eq. (2) and Eq. (7) respectively.
4. Predict test samples according to Eq. (3) and Eq. (6) respectively. As a test, $\boldsymbol{X}_T$ can be known latent variables to predict shapes $\boldsymbol{Y}'_T$. During simulating reconstruction, optimized latent variables can be used to predict the shapes $\boldsymbol{Y}'_T$. The mean of reconstruction errors between $\boldsymbol{Y}'_T$ and $\boldsymbol{Y}_T$ can judge the performance of models.

The optimization is necessarily used to reconstruct. Now let's see the process of the optimization in brief. First a latent variable is initialized with some knowledge or by intuition. Then we iteratively obtained the prediction of the 3D shape with LVMs and minimized the reprojection error. The optimization can be written as

$$\min_x \sum_{\mathrm{i}} ||(u_{\mathrm{i}}', v_{\mathrm{i}}') - (u_{\mathrm{i}}, v_{\mathrm{i}})||_2^2 \qquad (10)$$

where (u_i', v_i') and (u_i, v_i) are the reprojection and the real location of the input image's i^{th} feature point. The details of solution to reprojection can be found in [11].

3. Experimental Results

We verified the effectiveness of LVMs combined with clustering when applying it to non-rigid 3D reconstruction from a single camera. For all the experiments, we made use of the cardboard dataset and the cloth dataset which had been employed in [9].

3.1. Variations of Errors on LVMs combined with clustering

For both datasets, we first chose *M*=300 test samples, then clustered the latent variables of the rest samples with G-means to attain labeled samples, and finally chose *V*=50 unlabeled samples for CLVM. We learnt CLVM as well as GPLVM and predicted 300 shapes of the test samples with known latent variables obtained by PCA. Equality constraints were used here. Varying the critical value and maintaining the unlabeled samples unchanged, we observed the variations of the mean reconstruction error over the 300 test samples and the number of labeled samples with the critical value. Fig.1 gave the variations.

It can be observed from Fig.1 that the mean reconstruction errors roughly showed a trend of increase with the critical value, which meant that a lower critical value resulted in a better model. However, there was a truth that when the critical value increased, the number of labeled samples decreased, which was beneficial to lowering the computational burden. So a balance lay between the quality of models and the computation cost.

3.2. Comparison to LVMs without Clustering

We picked an appropriate critical value 60 to do the following experiments.

3.2.1. Prediction with Known Latent Variables

A comparison was made between two original LVMs (GPLVM and CLVM) and LVMs combined with clustering (GPLVM combined with clustering and CLVM combined with clustering) using known latent variables.

We still used M=300 test samples and V=50 unlabeled samples. For the LVMs combined with clustering, indices of labeled samples derived from the G-means, and the

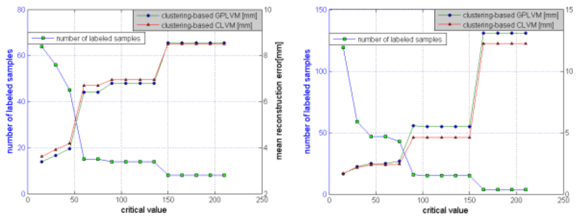

Fig.1. Variations of the mean reconstruction error and the number of labeled samples with the critical value for the cardboard (left) and the cloth (right).

Table 1. Comparison between the original LVMs and the LVMs combined with clustering when predicting shapes from known latent variables

		Equality Constraints		Inequality Constraints	
		R.E.	C.E.	R.E.	C.E.
cardboard	GPLVM	2.9791 ±0.2056	0.8852 ±0.0525	2.9791 ±0.2056	0.8646 ±0.0535
	with clustering	**2.4973**	**0.7566**	**2.4973**	**0.7366**
	CLVM	2.7459 ±0.1631	0.6309 ±0.0664	2.9026 ±0.1708	0.7457 ±0.0446
	with clustering	**2.3719**	**0.6221**	**2.4577**	**0.6824**
cloth	GPLVM	7.1611 ±0.2603	1.5163 ±0.1098	7.1611 ±0.2603	0.9041 ±0.0862
	with clustering	**6.4033**	**1.3906**	**6.4033**	**0.8648**
	CLVM	7.4518 ±0.2697	0.9268 ±0.0931	7.1061 ±0.1963	0.6520 ±0.0599
	with clustering	**6.8238**	**0.9045**	**6.3690**	**0.6345**

number N was 47 for the cardboard and 15 for the cloth. For the original LVMs, we performed 20 times learning with only labeled samples randomly chosen every time and with the test and unlabeled samples maintained the same as LVMs combined with clustering, and predicted with known test latent variables every time, taking the mean values and the standard deviations over 20 times results of mean reconstruction and constraint errors of 300 predictions. Equality and inequality constraints were imposed respectively. Table 1 presented the comparison about the reconstruction and the constraint errors.

The comparison clearly pointed out that reconstruction errors and constraint errors of the LVMs combined with clustering were never outside the range of those of the original LVMs. In fact, all the errors of the LVMs combined with clustering were lower than the mean values of the original LVMs. This revealed that the LVMs combined with clustering could improve the performance of original LVMs in nature.

3.2.2. Reconstruction with Optimized Latent Variables

We now generated synthetic data for both datasets. 1000 random points were spread on each mesh-represented shape to be 3D feature points. We specified their barycentric coordinates. 2D locations of the 3D feature points could be created by projection. We then added Gaussian noise with standard deviation 1 to these 2D locations. Having owned the synthetic data, we could use them to optimize latent variables. For this part, we just performed 5 times learning for the original LVMs since we took it into account that the process of optimization

Table 2. Comparison between the original LVMs and the LVMs combined with clustering when reconstructing the shapes with the optimized latent variables

		Equality Constraints		Inequality Constraints	
		R.E.	C.E.	R.E.	C.E.
cardboard	GPLVM	3.3470 ±0.1814	0.9690 ±0.0414	3.4804 ±0.1765	0.9301 ±0.0465
	with clustering	**2.9379**	**0.8380**	**2.9379**	**0.7913**
	CLVM	2.9032 ±0.0892	0.7664 ±0.0524	3.2640 ±0.1633	0.8469 ±0.0433
	with clustering	**2.4912**	**0.7520**	**2.8599**	**0.7776**
cloth	GPLVM	8.0354 ±0.2738	1.7189 ±0.0681	8.0354 ±0.2738	1.2257 ±0.1050
	with clustering	**6.9284**	1.7362	**6.9284**	**1.1490**
	CLVM	7.2406 ±0.2174	1.6788 ±0.0974	7.9362 ±0.2560	1.0760 ±0.1203
	with clustering	**6.4653**	**1.5678**	**6.8498**	**0.9488**

was rather time-consuming. Labeled, test and unlabeled samples were chosen in the same way with the part 3.2.1.

As observed from Table 2, nearly all the errors of the LVMs combined with clustering were smaller than the mean of errors of the original LVMs. That was to say, the LVMs combined with clustering gained improvements during the optimization which was a necessary process in the real situation. Therefore, we could draw a conclusion that the LVMs combined with clustering had advantages over their original ones. Fig. 2 showed shapes of a frame of deformation for the cardboard and the cloth. The shapes were reconstructed with equality constraints for the cardboard and inequality constraints for the cloth.

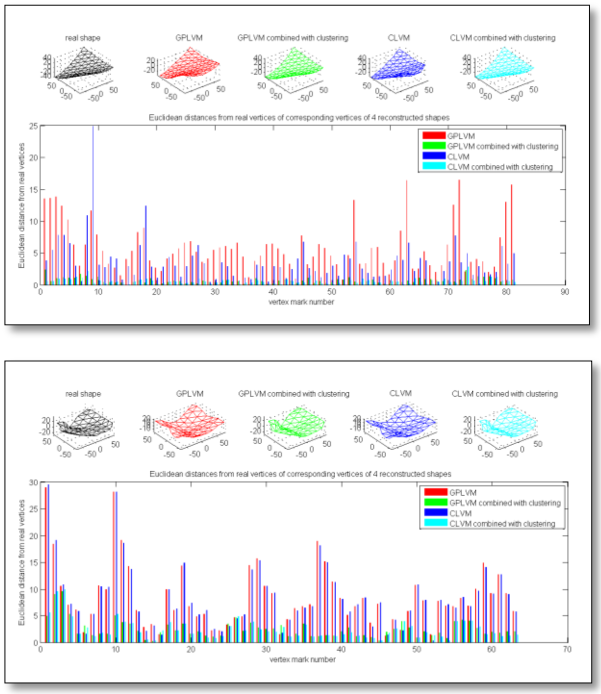

Fig.2. Shapes and distances from real vertices at each vertex about the 1821th and the 1729th deformation for the cardboard (top) and the cloth (bottom)

4. Conclusion

In this paper, we presented LVMs combined with clustering. The G-means clustering algorithm was introduced into LVMs. In theory, the LVMs combined with clustering can predict in a wider range, thus performing better. We conducted experiments on the cardboard dataset and the cloth dataset and compared the LVMs combined with clustering to the original ones, GPLVM and CLVM, in the context of monocular non-rigid 3D reconstruction. The results demonstrated the effectiveness of the LVMs combined with clustering. In future work, it is worth considering how to give a good initial latent variable during the optimization.

Acknowledgments

This work was supported by the Fund of National Natural Science Foundation of China (61231018), the Programme of Introducing Talents of Discipline to University (B13043) and the National Key Technology R&D Program (2013BAH62F03-3).

References

[1] I. Titov, and J. Henderson, “A Latent Variable Model for Generative Dependency Parsing,” IWPT 2007, pp. 144-155, 2007

[2] A. Ahmed, M. Aly, J. Gonzalez, S. Narayanamurthy, and A. Smola, “Scalable Inference in Latent Variable Models,” WSDM 2012, pp. 123-132, 2012

[3] R. Urtasun and T. Darrell, “Discriminative Gaussian Process Latent Variable Model for Classification,” ICML 2007, pp. 927-934, 2007

[4] N. D. Lawrence, “Learning for Larger Datasets with the Gaussian Process Latent Variable Model,” AISTATS 2007

[5] N. D. Lawrence, “Gaussian Process Latent Variable Models for Visualisation of High Dimensional Data,” NIPS 2004, Vol. 16, pp. 329-336, 2004

[6] N. D. Lawrence, “Probabilistic non-linear principal component analysis with Gaussian process latent variable models,” The Journal of Machine Learning Research, Vol. 6, pp. 1783-1816, 2005

[7] M. Salzmann, R. Urtasun, P. Fua, “Local deformation models for monocular 3D shape recovery,” CVPR 2008, June 23-28, pp. 1-8, 2008

[8] A. Varol, A. Shaji, M. Salzmann, P. Fua, “Monocular 3D Reconstruction of Locally Textured Surfaces,” PAMI-34(6), pp. 1118-1130, 2012

[9] A. Varol, M. Salzmann, P. Fua, R. Urtasun, “A constrained latent variable model,” CVPR 2012, June 16-21, pp. 2248-2255, 2012

[10] G. Hamerly, C. Elkan, “Learning the k in k-means,” NIPS 2004, Vol. 16, pp. 281-288, 2004

[11] M. Salzmann, P. Fua, “Deformable surface 3D reconstruction from monocular images,” Synthesis Lectures on Computer Vision 2(1), pp. 1-113, 2010