Learning Non-linear Dynamical Systems by Alignment of Local Linear Models

Masao Joko
Dept. of Aerospace Engineering
The University of Tokyo
Japan
mjoko@space.rcast.u-tokyo.ac.jp

Yoshinobu Kawahara
The Inst. of Scientific and Industrial Res. (ISIR)
Osaka University
Japan
kawahara@ar.sanken.osaka-u.ac.jp

Takehisa Yairi
Dept. of Aerospace Engineering
The University of Tokyo
Japan
yairi@space.rcast.u-tokyo.ac.jp

Abstract—Learning dynamical systems is one of the important problems in many fields. In this paper, we present an algorithm for learning non-linear dynamical systems which works by aligning local linear models, based on a probabilistic formulation of subspace identification. Because the procedure for constructing a state sequence in subspace identification can be interpreted as the CCA between past and future observation sequences, we can derive a latent variable representation for this problem. Therefore, as in a similar manner to the recent works on learning a mixture of probabilistic models, we obtain a framework for constructing a state space by aligning local linear coordinates. This leads to a prominent algorithm for learning non-linear dynamical systems. Finally, we apply our method to motion capture data and show how our algorithm works well.

Keywords—dynamical system; non-linear system; manifold learning; subspace identification;

I. INTRODUCTION

Learning dynamical systems is an important problem in several fields. Since we may not know the exact model parameter values, or even the gross model structure of the dynamical system itself in many cases, the dynamics of the system have to be learned from the sequences of observations only. One major approach to learn such dynamical systems is the subspace identification method, which involves geometric operations on subspaces spanned by the column or row vectors of certain block Hankel matrices formed by observed data [1]. After the proposition in the late 1980s, the subspace methods have been actively researched in the field of system control. However when the system is complex such as human motion, the model with linear expression is not suitable for modeling the system. Recently, nonlinear extensions have been actively researched [2]. On the other hand, in the fields of pattern recognition and machine learning, a wide range of techniques have been proposed for unsupervised learning of nonlinear manifolds [3], which assume that the high dimensional observations, such as images, are confined to low dimensional subspaces of the original high dimensional spaces. In this paper, we integrate these researches respectively developed in system identification and pattern recognition, and obtain a framework for constructing a state space by aligning local linear coordinates. This approach leads to a prominent algorithm for learning non-linear dynamical systems in the case with no-input signals, based on a probabilistic interpretation of the subspace identification in an analogous fashion with probabilistic CCA [4]. Our framework is based on the same intuitions as in earlier works that learn a mixture of latent variable density models on the original training data so that each mixture component acts as a local feature extractor, and thus can be seen a novel extension of this idea to the domain of learning dynamical systems. In Section II, we derive a latent variable model in an analogous form to CCA by interpreting subspace identification probabilistically, and propose a method for learning non-linear dynamical systems by global alignment of local linear models. Then we show a filtering procedure in Section III. Finally, some illustrative examples are presented in Section IV, and we give a conclusion in Section V.

Notation:: Let $y(\tau)$, $\tau=0, \pm 1, \cdots$ be a discrete-time vector process, then the past and the future subsequences of $k$-steps, corresponding to the present time $t$, will be referred to as

$$
\bar{y}_k(t) \equiv [y(t-1)^T, y(t-2)^T, \cdots, y(t-k)^T]^T
$$

$$
y_k(t) \equiv [y(t)^T, y(t+1)^T, \cdots, y(t+k-1)^T]^T.
$$

II. LEARNING NON-LINEAR DYNAMICAL SYSTEMS BY ALIGNMENT APPROACH

Consider a discrete-time wide-sense stationary vector process $y(t), t=0, 1, \cdots$ with dimensionality $p$, where the component $y$ models the output of the unknown stochastic system which we want to identify. We often assume that the stochastic system can be modeled as a linear state-space system:

$$
x(t+1) = Ax(t) + v(t), \ y(t) = Cx(t) + w(t), \quad (1)
$$

where $x$ is the state vector, and $v$, $w$ are the system and observation noises. In this paper, in order to consider the complex non-linear system such as human motion, we model the system as a mixture of state-space systems probabilistically as described later (cf.(4)). In Section II, $A$, we introduce a probabilistic interpretation of the subspace identification, and consider a method to construct a state space by integrating several local coordinates into a
single global representation. This is achieved in the similar manner with the recent works on coordination methods for probabilistic models [3]. Then in Section II. B, we obtain the system parameters of the mixture model by solving regression problems and then derive an overall learning algorithm.

A. Construct a State Space by Aligning Local Coordinates

Here, we present a latent variable model corresponding to the balanced stochastic realization of a linear state-space model, putting the relation with probabilistic CCA [4] into context. As stated in Section I, one solution to construct the state vector process in a linear state-space model (1) is the subspace identification method based on CCA [1][5], which calculate the balanced stochastic realization [6]. On the other hand, recently a probabilistic interpretation of CCA was reported [4], in which latent variable model for CCA was presented. Thus we can introduce a latent variable model corresponding to the balanced stochastic realization:

\[ x \sim N(0, \Psi), y_k|x \sim N(F_1x, \Psi_1), \hat{y}_k|x \sim N(F_2x, \Psi_2), \]

where \( \Psi, F_1, F_2, \Psi_1 \) and \( \Psi_2 \) are the parameters of this model. That is, the latent variables obtained by CCA between future and past observation sequences correspond to the state vector in the dynamical system. In this way, subspace identification can be interpreted probabilistically. Moreover, considering a mixture model and the alignment condition as in an analogous fashion to [3], we can extend the subspace identification method into nonlinear method; we can construct a global state space by integrating several local coordinates into a single global representation. Figure 1 shows the graphical models of CCA-based subspace identification and local alignment approach. In the current setting, in order to obtain the global coordinates, we add to the data log-likelihood a penalty term which measures how much the mixture \( p(x|\hat{y}_k(n), y_k(n)) \) resembles a Gaussian \( q_n(x) = \mathcal{N}(x(t), \Sigma(t)) \), in the sense of KL divergence \( D(q_n(x) \parallel p(x|\hat{y}_k(n), y_k(n)). \) Thus we try to maximize the following penalized log-likelihood objective function:

\[
L' = \sum_{n=1}^{N} [\log p(\hat{y}_k(n), y_k(n)) - D(q_n(x) \parallel p(x|\hat{y}_k(n), y_k(n))].
\]  \hspace{1cm} (2)

This objective function can be lower bounded by \( \Phi \)

\[
L' \geq \Phi = \sum_{n=1}^{N} H(q_n(s)) + H(q_n(x)) + \sum_{n=1}^{N} \sum_{s=1}^{C} q_n(s) \int q_n(x) \log p(\hat{y}_k(n), y_k(n), x, s) dx, \]

where \( H \) denotes the entropy of a distribution, and \( q_n(s) \) denotes the distribution over the mixture components. This objective function can be maximized through the EM-like procedure as in [3], by replacing two i.i.d. observed variables and an hidden variable by the subsequences \( \hat{y}_k \) and \( y_k \), and the state vector \( x \), respectively. The parameters are also initialized as in [3] using extended LLE [7]. Thus we can calculate \( p(x|\hat{y}_k(n), y_k(n)) \) as the aligned global coordinates.

B. Learning Algorithm

Once a state sequence is obtained, a dynamical system can be estimated by computing two regression problems in each linear subspace \( E\{p(x(t)|\hat{y}_k(t), y_k(t), s) \} \) obtained in the previous section. Then, both the transition and observation probability densities also become a mixture, expressed as

\[
p(x(t+1)|x(t)) = \sum_{s=1}^{C} p(s|x(t))p(x(t+1)|x(t), s)
\]

\[
p(y(t)|x(t)) = \sum_{s=1}^{C} p(s|x(t))p(y(t)|x(t), s), \hspace{1cm} (4)
\]

where \( s \) is an index over the \( C \) components of the mixture and \( p(s|x(t)) \) is the weight on the \( s \)-th mixture given the current position in the global coordinate computed via \( p(x(t), s)/p(x(t)) \). So the parameters that we need to estimate, in the regressions for the model:

\[
p(x(t+1)|x(t), s) = \mathcal{N}(\mu^s_x + A_s x(t), \Sigma^s_x)
\]

\[
p(y(t)|x(t), s) = \mathcal{N}(\mu^s_y + C_s x(t), \Sigma^s_y), \hspace{1cm} (5)
\]

are the regression matrices \( A_s \) and \( C_s \), the covariances \( \Sigma^s_x \) and \( \Sigma^s_y \), and the residuals \( \mu^s_x \) and \( \mu^s_y \). This regression problem can be calculated using the estimates of the state in each linear subspace \( E\{p(x(t)|\hat{y}_k(t), y_k(t), s) \} \) by applying the least-square method. As a result, we conclude the overall procedure for learning non-linear dynamical systems by alignment approach as follows.

1. Construct the set of subsequences from a given finite time series data with length \( N + 2k - 1 \).
2. Initialize the parameters by extended LLE [7] with the subsequences \( \hat{y}_k(t) \) and \( y_k(t) \).
3. Optimize the objective function (2) by the EM-like algorithm as in [3].
4. Solve the regression problems by the least-square methods and estimate the parameters in (5).
III. FILTERING IN GLOBAL STATE SPACE

In this section, we give an algorithm of filtering non-linear dynamical system. Filtering is the recursive calculation of \( p(x(t)|y(1:t)) \), where \( y(1:t) \) means all observations up to the present time \( t \). This can be carried out by the iteration of two steps, called the prediction step and the update step. This can be calculated exactly as well as the Kalman filter except for including the marginalization on mixture components by the prediction step and the update step. Here since local coordinates are learned as less overlapped as possible, \( p(x(t)|y(1:t)) \) can be approximated by the component \( s^* \) which has the highest weight \( p(s^*|y(1:t)) \),

\[
p(x(t)|y(1:t)) = \sum_{s=1}^{C} p(s|y(1:t)) p(x(t)|y(1:t), s) \\
\approx p(x(t)|y(1:t), s^*).
\]

Thus our filter only needs to keep a Gaussian distribution on the global coordinate in each time \( t \), and we can escape the exponential increase of the number of mixtures on the state as in the case of the switching Kalman filter. In fact, the calculational complexities for a update in filtering is \( O(C \times d^3) \) in the proposed algorithm. This property of the proposed algorithm has an immense significance in the conduct of filtering.

IV. EXPERIMENTAL RESULTS

In Section IV. A, we compared our algorithm with some major linear approaches in the performance of prediction. Then in Section IV. B, we applied our method to the motion capture, where data is high dimensional and expected to be generated from a complicated system.

A. Comparative results

The compared algorithms are the Kalman filter-smoother-based algorithm for linear dynamical systems (we abbreviate as KFS) [8] and orthogonal-decomposition-based subspace identification method (ORT) [1]. We tested with the famous non-linear dynamical system example known as Lorenz attractor. The latent variables are the solutions of the following differential equations

\[
\begin{align*}
\dot{x}_1(t) &= -ax_1(t) + ax_2(t) \\
\dot{x}_2(t) &= -x_1(t)x_3(t) + rx_1(t) - x_2(t) \\
\dot{x}_3(t) &= x_1(t)x_2(t) - bx_3(t),
\end{align*}
\]

and the observations are the latent variables with some Gaussian noise. We set \( a=10 \), \( b=28 \), and \( r=8/3 \) and the initial values were \((-10.0, -10.0, 30.0)\). We gathered 2000 data points for training (every 0.1 second for 200 seconds), and following new 200 data points for test. The dimensionality of the state and the length of subsequence \( d \) are \((4, 5)\). As for our algorithm (MIX), we selected \( C=6 \), and the nearest neighborhood for the LLE as 20. We evaluated these algorithms with the predictive log-likelihood of filtering (i.e. 1 step prediction). The comparative results are shown in Table I, and you can see that the performance of prediction is improved drastically. As an illustrative result of our algorithm, the Figure 2 illustrates how each component distributes in the vector field corresponding to \( y_1(t) \) and the vertical axis \( y_1(t+1) \). This type of plot is useful to analyze the dynamics in time series. You can see that the samples of each component are distributed in successive areas in the vector field. And Figure 3 shows how filtering works well.

![Figure 2](image.png)

**Figure 2.** Samples that belong to each component in Lorenz simulation(each color corresponds to each mixture component)

B. Application to motion capture

The data used for this experiment is human motion capture data from the Carnegie Mellon University motion capture database.\(^2\) We used a simplified skeleton, where each pose

\(^1\)The parameters we need to tune beforehand are the number of components \( C \), the dimensionality of intrinsic manifold \( d \), the length of subsequence \( k \) and the number of nearest neighbors for the LLE. For \( k \), it is known that \( k \) should be large to a certain extent from the viewpoint of subspace identification, but the selection of \( C \) and \( d \) is our future tasks.

\(^2\)http://mocap.cs.cmu.edu/
is defined by 50 dimensions [9]. We trained the model on the first 150 frames of the walking sequence, which corresponded to 2 walking cycles. The dimensionality of the state and the length of subsequence (d, k) were (5, 5), and we selected C=4 and the nearest neighborhood for the LLE as 8. Figure 4 shows the latent space projected onto visible 3D vector fields in the training phase. You can see that the components are distributed separately. Furthermore, Figure 5 shows a time series of reconstructed human motion images which correspond to Figure 4. The human motion dynamical system is divided into each charasteric movement in each component, and we can understand the system easily.

V. CONCLUSION

A novel algorithm for learning non-linear dynamical systems which works by aligning local coordinates has been proposed, based on a probabilistic formulation of subspace identification. The procedure for constructing a state sequence in subspace identification can be interpreted as the CCA, and based on this, we presented a latent variable model for this problem. So as in an analogous fashion to the recent works on learning a mixture of probabilistic models, we presented a framework for learning non-linear dynamical systems. Our work is expected to give an insight on the connection of dynamical systems with manifold learning and a mixture of probabilistic models. We also give the motion capture example and the comparative results that show the usefulness of our method. In future work, we will investigate the relationship between dynamical systems and manifold learning, and also develop the model selection for this framework for choosing the number of components.

REFERENCES

[10] V. Pavlović, J. M. Rehg, T. J. Cham and K. P. Murphy, A Dynamic Bayesian Network Approach to Figure Tracking Using Learned Dynamic Models: Conference on Computer Vision (ICCV 99), September 1999.