

AN EXEMPLAR-BASED HIDDEN MARKOV MODEL FRAMEWORK FOR NONLINEAR STATE-SPACE MODELS

Redouane Lguensat[#], Ronan Fablet[#], Pierre Ailliot^{#†}, Pierre Tandeo[#]

[#]Institut Mines-Telecom; Telecom Bretagne, CNRS UMR 6285 LabSTICC, pôle CID, Brest, France.

[†] University of Western Brittany, CNRS UMR 6205 LMBA, Brest, France.

ABSTRACT

In this work we present a data-driven method for the reconstruction of dynamical systems from noisy and incomplete observation sequences. The key idea is to benefit from the availability of representative datasets of trajectories of the system of interest. These datasets provide an implicit representation of the dynamics of this system, in contrast to the explicit knowledge of the dynamical model. This data-driven strategy is of particular interest in a large variety of situations, e.g., modeling uncertainties and inconsistencies, unknown explicit models, computationally demanding models, etc. We address this exemplar-based reconstruction issue using a Hidden Markov Model (HMM) and we illustrate the relevance of the method for missing data interpolation issues in multivariate time series. As such, our contribution opens new research avenues for a variety of application domains to exploit the wealth of archived observation and simulation data, aiming a better analysis and reconstruction of dynamical systems using past and future observation sequences.

Index Terms— Exemplar-based model, Missing data estimation, Hidden Markov Models, Analog method.

1. PROBLEM STATEMENT AND RELATED WORK

In many practical problems, one is interested in estimating an unobserved process $\{X_t\}_{t \in \llbracket 1, \dots, T \rrbracket}$ given a sequence of observations $\{Y_t\}_{t \in \llbracket 1, \dots, T \rrbracket}$. State-space methods provide a flexible framework to address this issue. They rely on the definition of two key components. Firstly, the dynamical model states the temporal dynamics of process $\{X_t\}_{t \in \llbracket 1, \dots, T \rrbracket}$, typically Markovian dynamics (as an illustration, we consider here a first-order Markov process). Secondly, the observation model relates the unknown state X_t at a given time t to the observed variable Y_t at the same time. Formally, it resorts to:

$$\begin{cases} X_t | X_{t-1} = x_{t-1} \sim f(\cdot, x_{t-1}) \\ Y_t | X_t = x_t \sim g(\cdot, x_t) \end{cases} \quad (1)$$

To be fully characterized, this state-space setting also involves the definition of the prior distribution of X_1 . From a Bayesian perspective, the reconstruction of the unknown

state sequence $\{X_t\}$ from a partial and/or noisy observation sequence $\{Y_t\}$ comes to evaluate filtering and smoothing posteriors, respectively $P(X_t | Y_{1:t})$ and $P(X_t | Y_{1:T})$. Among the classical techniques to infer these posteriors, one may refer to Kalman filtering and smoothing [1] for linear-Gaussian cases, to particle filtering and smoothing [2, 3] for nonlinear non-Gaussian cases, as well as further extensions for high-dimensional systems, such as the Ensemble Kalman filter and smoother [4]. All these models exploit the online evaluation of the dynamical model to forecast the state from a given time step to the next one. They then may be affected by modeling inconsistencies and uncertainties.

Meanwhile, over the last decades, observation, simulation and storage capabilities have been greatly enhanced, such that one may be provided with representative and realistic examples of the trajectories of a dynamical system of interest. We view these datasets as an implicit representation of system dynamics to overtake inconsistencies and uncertainties of explicit models and the absence of known explicit models as well. The key idea is to combine nearest neighbors search (referred to *Analog method* in the geoscience community) and stochastic filtering methods. Analog method introduced by Lorenz in 1969 [5] refers to a non-parametric forecasting method. It consists in searching for the state in historical data, referred to as the analog, that resembles the most to some current state and in forecasting the current state from the successor of the analog. This technique has recently gained much interest with proliferation of experimental and observational data, and have been applied in a variety of domains including, among others, data assimilation [6], climate and weather forecasting [7, 8], and the inpainting of dynamic textures [9].

Let us assume that we are provided with a representative dataset of state sequence, that consists of historical realizations of that state. We are interested in state space methods where instead of using the model equation, the forecasts are sampled from the nearest-neighbors (analog) of the current state, this estimate is then updated using the observation at that time. This idea was first introduced by Tandeo et al. [6], where authors have proposed non-parametric Ensemble Kalman Filters and Smoothers (EnKF and EnKS) and have shown the relevance of their algorithm to reconstruct chaotic systems from noisy observations. Here, we show that this

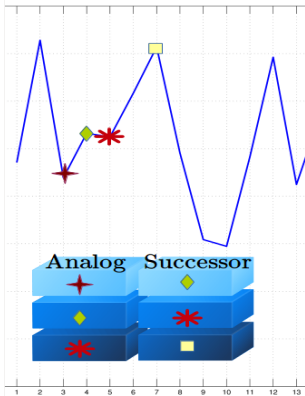


Fig. 1: The signal shown represents the historical trajectory of the system of interest. The catalog is built by matching every analog with its successor.

non-parametric setting can be restated using a discrete hidden Markov model, where the discrete state values refer to the different elements of the dataset. The key features of this HMM setting are two-fold compared to Ensemble Kalman filtering methods: the use of forward-backward procedures to derive exact filtering and smoothing posteriors, and the ability to address non-Gaussian multi-modal distributions.

This paper is organized as follows. Section 2 introduces the Analog HMM and the Analog Forward-Backward algorithms. In Section 3, we report numerical experiments for two dynamical systems, namely chaotic and nonlinear Lorenz-63 dynamics and human motion data. Section 4 summarizes our main contributions and states starting points for future work.

2. ANALOG HIDDEN MARKOV MODELS

2.1. Hidden Markov Models

With a view to formally introducing the considered HMM-based setting, we consider the following definitions. Let us denote by X_t the state at time t , Y_t the corresponding observation at time t . State X_t will be here a discrete random variable with values within a predefined finite set of possible states, as detailed in the next section. The HMM describes the joint probability of the hidden and observed discrete random variables. Let us suppose that the HMM involves S discrete states. Following [10], we denote by $\Lambda = (A, B, \pi_1)$ the parameters of the HMM. Without loss of generality, we here consider a first-order HMM, where:

- The **transition matrix** is given by: $A = \{a_{ij}\} = P(X_t = s_j | X_{t-1} = s_i)$
- The **observation matrix** (also called the emission matrix) refers to the likelihoods of each observation Y_t for each state s_j : $B = \{b_j(Y_t)\}$ where $b_j(Y_t) = P(Y_t | X_t = s_j)$.
- The **initial state distribution** (i.e. when $t = 1$) is given by: $\pi_1 = \{\pi_i\}$ where $\pi_i = P(X_1 = s_i)$.

Knowing the above HMM parameters, for any given observation sequence, we can compute the filtering and smoothing posteriors respectively $P(X_t | Y_{1:t})$ and $P(X_t | Y_{1:T})$ using

the classical *forward-backward algorithm* [10]. As applied in the next section, these posteriors are the basic component of different criteria for the inference of the state sequence best matching the observation time series, such as the Maximum A Posteriori, the Maximum Posterior Mode or the Posterior mean.

2.2. Analog HMM

Let us assume we are provided with a database \mathcal{D} storing historical trajectories of the system of interest. In a pre-processing step, we extract from \mathcal{D} every historical state with its associated successor in time. We note by $\mathcal{D}_a = (s_i)_{i \in [1, \dots, P]}$ the set of retrieved states that we refer to as *analog*s, while corresponding successors will be noted as $\mathcal{D}_s = (s_i)_{i \in \mathcal{F}}$, where \mathcal{F} contains indices of states at future times i.e. $s_{\mathcal{F}(i)}$ is the successor of s_i . Keep in mind that analogs and successors are just elements of \mathcal{D} organized in such a way to simplify algorithmic implementation. $\mathcal{D}_a \cap \mathcal{D}_s$ is not empty, as an analog can be the successor of another analog. The setting aforementioned helps us to build a "catalog" \mathcal{C} by coupling every analog with its successor, this can be written as:

$$\mathcal{C} = \{(s_i, s_{\mathcal{F}(i)}), i \in [1, \dots, P]\} \quad (2)$$

An illustration of the construction of the catalog is given in Fig.1. For practical applications, the construction of the catalog may benefit from realistic numerical simulations of the system of interest, as illustrated with Lorenz-63 case-study, and/or available observation datasets, as illustrated by the considered human motion case-study.

Unlike the classic state space formulation where X_t is a continuous variable (Eq.1), the Analog Hidden Markov Model setting relies on the discrete state space formed by the analogs \mathcal{D}_a and successors \mathcal{D}_s . Thereby the possible values of X_t are restricted to $\mathcal{S} = \mathcal{D}_a \cup \mathcal{D}_s$. The considered exemplar-based state-space model is stated as a discrete HMM with a large number of discrete states $P - \#(\mathcal{D}_a \cap \mathcal{D}_s)$. We resort to the Analog HMM characterized by its states \mathcal{S} and by parameters $\Lambda_a = (A, B, \pi_1)$:

$$\begin{cases} X_t = s_j | X_{t-1} = s_i \sim A = \{a_{ij}\} \\ Y_t = y_t | X_t = s_j \sim B = \{b_j(y_t)\} \end{cases} \quad (3)$$

The parameterization of the transition matrix relies on the determination of transitions between the states. We consider a sparse parameterization of the transition matrix, where each state $s_i \in \mathcal{S}$ involves K possible transitions as follows:

- We search for the K -nearest neighbors of s_i in set \mathcal{D}_a according to a predefined kernel in the state space.
- Let $\{s_n\}_{n \in \mathcal{I}(i)}$ denote the K nearest neighbors (analog) of s_i , where $\mathcal{I}(i) = \{i_1, i_2, \dots, i_K\}$ contains the K indices of these analogs. From catalog \mathcal{C} , we retrieve their successors $\{s_n\}_{n \in \mathcal{F}(\mathcal{I}(i))}$.

- the transition probabilities $a_{ij} = P(X_t = s_j | X_{t-1} = s_i)$ from state $s_i \in \mathcal{S}$ to state $s_j \in \mathcal{S}$ are non-null for successors $\{s_n\}_{n \in \mathcal{F}(i)}$

$$a_{ij} \propto \begin{cases} \exp\left(-\frac{\|s_i - s_k\|^2}{2\sigma^2}\right) & \text{if } j = \mathcal{F}(i_k) \\ 0 & \text{else} \end{cases} \quad (4)$$

where σ can be thought as a scale parameter.

The parameterization of the transition matrix involves the choice of the kernel, the number of nearest neighbors and the scale parameter σ . One may consider cross-validation schemes to set these parameters. In the reported experiments, we adopt a Radial Basis Function kernel. We use an adaptive setting for scale parameter σ : for every s_i we adapt σ to be the standard deviation of the K distances between s_i . Regarding parameter K , we typically consider between 5 and 15 states in the reported experiments. Overall, the transition matrix is a $M \times M$ matrix with only $M \times K$ non-null values.

The observation matrix of the HMM directly follows from the observation model $P(Y_t | X_t)$. The global observation matrix is a $M \times T$ matrix $b_j(y_t) = P(Y_t = y_t | X_t = s_j)$. In the reported numerical experiments, Gaussian observation models are considered:

$$b_j(y_t) \propto \exp\left(-\frac{(y_t - s_j)^T R^{-1} (y_t - s_j)}{2}\right) \quad (5)$$

where R is the observation covariance error matrix.

We may consider higher-order Markovian properties with a view to accounting for longer time dependencies. For a given time lag δ , it comes to consider the augmented state:

$$\hat{X} = (X_t, X_{t-1}, X_{t-2}, \dots, X_{t-(\delta-1)}) \quad (6)$$

Creating the catalog in this case and setting the parameters of the Analog HMM follow the same steps as aforementioned.

2.3. Analog Forward-Backward (AnFB) Algorithms

Given the above parameterization of the Analog HMM, we can apply the classical Forward-Backward procedure [10] to numerically evaluate the exact values of smoothing posteriors $P(X_t | Y_{1:T})$ and filtering posteriors $P(X_t | Y_{1:t})$ for any given observation sequence $Y_{1:T}$.

The straightforward application of this procedure may involve matrix operations with very large matrices (depending on the size of the catalog). To reduce the computational complexity, we first benefit from the sparsity of the transition matrix. In both the forward and backward recursions, the sparsity of the transition matrix implies that some states may have a null posterior. Therefore, we do not compute the terms of the observation matrix within the recursions only for state values with non-null transition probabilities. It may be noted that this implementation does not involve any approximation. However, it does not guarantee that the computational complexity does not evolve as $O(M^2)$, as one may expect that all

possible discrete states may be associated with non-null posteriors whatever the initial priors for long-enough observation sequence.

As a consequence, when a $O(M^2)$ complexity becomes an issue, we introduce a truncated forward-backward procedure. During the forward recursion, at any time t , we only keep the N most likely states with respect to filtering posteriors $P(X_t | Y_{1:t})$. We proceed as follows. We apply the transition matrix to the states at time $t-1$ with non-null posteriors. Only, the resulting states at time t involve non-null posteriors $P(X_t | Y_{1:t-1})$. We then determine posteriors $P(X_t | Y_{1:t})$ as $P(X_t | Y_{1:t}) \propto P(Y_t | X_t) P(X_t | Y_{1:t-1})$ and only retain the N most likely states according to $P(X_t | Y_{1:t})$. Posteriors $P(X_t | Y_{1:t})$ are renormalized to account for this selection. This scheme can be viewed as a truncating of the posterior to force to zero $M - N$ terms to speed up the procedure and constrain the actual memory usage. A similar procedure is applied for the backward recursion.

Overall, given the posteriors computed from the forward-backward procedures, we may consider different criterions to infer the hidden state sequence associated with a given observation sequence $Y_{1:T}$:

- The MPM (Maximum Posterior Mode) sequence, given by $\arg \max_{s_i} P(X_t = s_i | Y_{1:T})$, maximizes the computed local posteriors.
- $\arg \max_{X_{1:T}} P(X_{1:T} | Y_{1:T})$ Gives the MAP sequence. To retrieve this sequence, Viterbi algorithm is the typical numerical solution.
- With a view to minimizing the mean square estimation error, one resorts to conditional mean $E(X_t | Y_{1:T}) = \sum_i s_i \cdot P(X_t = s_i | Y_{1:T})$. This is the version that we will consider in the experiments.

3. NUMERICAL EXPERIMENTS

To illustrate the relevance of the proposed exemplar-based HMM framework, we consider applications to missing data interpolation for two nonlinear dynamical systems, Lorenz-63 dynamics and human motion data.

3.1. Lorenz 63 chaotic model

Lorenz-63 model is a simplified model for atmospheric convection introduced by Edward Lorenz [11]. It consists of a three-dimensional state, whose non-linear dynamics are given by:

$$\begin{cases} \frac{dx}{dt} = \mu(y - x), \\ \frac{dy}{dt} = x(\rho - z) - y, \\ \frac{dz}{dt} = xy - \beta z. \end{cases} \quad (\mathcal{L}_{63})$$

where x, y, z form the state of the system, and μ, ρ, β are the parameters of the model. As in [11] we use the values $\rho = 28$, $\mu = 10$, and $\beta = 8/3$ which gives birth to the *Lorenz strange attractor*. This model is widely used in the literature

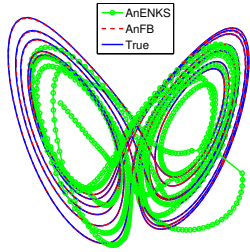


Fig. 2: Lorenz63 model: 1000 time steps of the true Lorenz attractor and reconstruction using the truncated AnFB and the AnEnKS.

	EnKS	AnEnKS		AnFB	
$dt_{obs} = 40$	\mathcal{L}_{63}	$M = 10^4$	$M = 10^5$	$M = 10^4$	$M = 10^5$
RMSE	3.7799	4.3279	3.2788	1.3183	0.5774
τ (s)	2 h	13 min	1.3 h	8 min	14 min

Table 1: RMSE comparison when increasing M between EnKS (state-of-the-art algorithm, pure dynamical model), and both the AnEnKS using 100 ensemble members and the truncated AnFB with $N = 1000$ when $dt_{obs} = 40$

to benchmark assimilation and filtering schemes with respect to nonlinear dynamics.

The reported experiments address the reconstruction of the three-dimensional state sequence at a unit sampling rate ($dt = 1$) from noisy observations of only the variable x provided at a lower sampling rate of $dt_{obs} = 40$. The construction of the catalog is issued from numerical simulations of Lorenz-63 dynamics. Regarding the test sequence, we run an independent simulation and sample the observation sequence for variable x every dt_{obs} time step and add white noise sample with variance $R = 2$. We report in Fig.2 the reconstruction of Lorenz-63 attractor from only the partial and noisy observation of x . It shows the relevance of the AnFB algorithm to address complex non-linear and chaotic dynamics. We further compare the reconstruction performance of the AnFB algorithm to those of the AnEnKS [6] and of the Ensemble Kalman filter and Smoother (EnKS) [4] which uses explicit dynamical model (\mathcal{L}_{63}). We report Root Mean Square Error (RMSE) performance calculated over the three variables when increasing M in Table.1 in the case of $dt_{obs} = 40$. The situation when $dt_{obs} = 40$ exhibits highly nonlinear patterns, which can hardly be captured by a linear dynamical model. Execution time τ is also shown in the same table.

The proposed HMM-based setting, AnFB, clearly outperforms the other exemplar-based filtering strategies (Fig.2). It proves more robust than the AnEnKS, which assumes a Gaussian distribution for the state sequence. Table.1 shows how performance improves when increasing M . With $N = 1000$ the truncated AnFB outperforms both the AnEnKS and the state-of-the-art EnKS which relies on pure equations of the model. With scarce observations both in time ($dt_{obs} = 40$) and space (observing only x), we still achieve a good and faster reconstruction of the attractor trajectories.

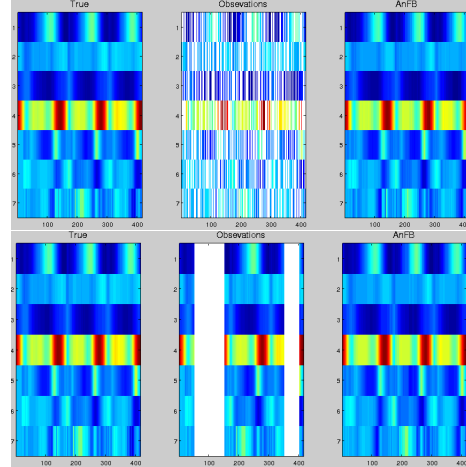


Fig. 3: Color map representation of the 7 DOFs. From left to right: the true dynamics of the test walking sequence, the observations with missing data (blanks), the reconstruction using the AnFB. Top is experiment (A), bottom is (B).

3.2. Human motion data

We also consider an application to motion capture data. Due to sensor failure or lighting and environmental effects [12], some markers may not be tracked correctly and one has to deal with missing data interpolation. Whereas this issue is often stated as an exemplar-driven synthesis [12], we apply here the proposed exemplar-based filtering, which allows to benefit from all the available observations.

In this section we report the results of using an Analog HMM with data retrieved from the CMU Graphics Lab Motion Capture Database. We consider series of 23 walk sequences performed by subject 35. The last walking sequence was taken out to be used as test data, and we design an Analog HMM using the others. We rescale data and evaluate the transition matrix which involves 9154 states. Test data consists of 415 frames describing the body dynamics by a 62-dimensional vector representing degrees of freedom (DOF) of joint angles with respect to a skeletal structure. In this work we consider only the dynamics of the left leg which is 7-dimensional, the seven DOFs are ordered as the following: femur (3), tibia (1), foot (2), and toes (1). We consider two experiments: (A) Erasing randomly 50% of test data and adding a white noise of covariance $0.5\mathbf{I}_7$, (B) a case where we erase all data from frame 50 to 150 and from 350 to 400 and add a white noise of covariance $0.1\mathbf{I}_7$. We run the AnFB with $N = 500$ and compare it with the AnEnKS [6]. Evaluation is done by calculating the RMSE between the reconstruction and ground truth data on the missing data values coordinates. A cross validation step is done to tune the parameter K , we test different values and get finally $K = 6$. We notice that the AnFB perform very well when the variable presents a repeated pattern like in DOFs $\{1,3,4,5,6\}$. Performance degrades slightly when the variable may present a scenario

Experiment	NNN	AnEnKS		AnFB	
	A	A	B	A	B
RMSE	15.5247	1.7696	2.8151	1.7499	2.2576
τ (s)	0.5 h	25 s	25 s	18 s	80 s

Table 2: RMSE comparison and time execution for the two experiments **A** and **B**.

which is far from the scenarios of the catalog like for DOFs $\{2,7\}$. In Fig.3 we show a color map of the ground truth for the 7 DOFs, the positions of the missing data in the observations (blanks) and the reconstruction using the AnFB which gives an idea of the quality of missing data estimation. Videos of the observations with missing data and those of both the AnFB and a naive nearest neighbors (NNN) reconstruction can be seen at the website: perso.telecom-bretagne.eu/redouanelguensat/anhmm. Table 2 shows how the AnFB outperforms the AnEnKS in terms of RMSE and time execution except for experiment **B** where the AnFB takes more time in the frames where we do not have observations, this is due to the fact that the AnFB tends to explore more states when not directed by the observations.

4. CONCLUSION AND FUTURE WORK

In this study, we propose exemplar-based filtering techniques for state-space models. We exploit discrete HMM and derive sequential and exact computations of the posterior likelihoods associated to the inference of the hidden state sequence given a partial and noisy observation sequence. Our results show that we achieve remarkably good performance when the size of the database is sufficiently important with the advantage of being model-free. In this work we used a special structure for the transition matrix, where the user can tune two parameters K and σ , and choose a suitable kernel for the considered data. We introduced a truncating scheme which can make execution time reasonable in case of big data. Dealing with the curse of dimensionality is our bigger challenge given that we work with big HMMs and since k-NN algorithms performance degrades in high dimensions. Exploring kernel methods for a better selection of the analogs and integration of dimensionality reduction methods is an assuring step to our future work.

5. ACKNOWLEDGMENT

Motion capture data was obtained from mocap.cs.cmu.edu, created with funding from NSF EIA-0196217. Matlab playback of motion and generation of videos was done using Neil Lawrence’s Motion Capture toolbox. This work was supported by ANR (Agence Nationale de la Recherche, grant ANR-13-MONU-0014) and Labex Cominlabs (project SEACS).

6. REFERENCES

- [1] Rudolph Emil Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Fluids Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [2] Genshiro Kitagawa, “Monte Carlo filter and smoother for non-Gaussian nonlinear state space models,” *Journal of computational and graphical statistics*, vol. 5, no. 1, pp. 1–25, 1996.
- [3] Neil J Gordon, David J Salmond, and Adrian FM Smith, “Novel approach to nonlinear/non-Gaussian Bayesian state estimation,” in *IEE Proceedings F (Radar and Signal Processing)*. IET, 1993, vol. 140, pp. 107–113.
- [4] Geir Evensen, “The ensemble Kalman filter: Theoretical formulation and practical implementation,” *Ocean dynamics*, vol. 53, no. 4, pp. 343–367, 2003.
- [5] Edward N Lorenz, “Atmospheric predictability as revealed by naturally occurring analogues,” *Journal of the Atmospheric sciences*, vol. 26, no. 4, pp. 636–646, 1969.
- [6] Pierre Tandeo et al., “Combining Analog Method and Ensemble Data Assimilation: Application to the Lorenz-63 Chaotic System,” *Machine Learning and Data Mining Approaches to Climate Science*, pp. 3–12, 2015.
- [7] Luca Delle Monache, F Anthony Eckel, Daran L Rife, Badrinath Nagarajan, and Keith Searight, “Probabilistic weather prediction with an analog ensemble,” *Monthly Weather Review*, vol. 141, no. 10, pp. 3498–3516, 2013.
- [8] Zhizhen Zhao and Dimitrios Giannakis, “Analog forecasting with dynamics-adapted kernels,” *arXiv preprint arXiv:1412.3831*, 2014.
- [9] Redouane Lguensat, Pierre Tandeo, Ronan Fablet, and Pierre Ailliot, “Non-parametric Ensemble Kalman methods for the inpainting of noisy dynamic textures,” in *IEEE International Conference on Image Processing (ICIP), 2015*, Sept 2015, pp. 4288–4292.
- [10] Lawrence R Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [11] Edward N Lorenz, “Deterministic nonperiodic flow,” *Journal of the atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.
- [12] Graham W Taylor, Geoffrey E Hinton, and Sam T Roweis, “Modeling human motion using binary latent variables,” in *Advances in neural information processing systems*, 2006, pp. 1345–1352.