NAIST-IS-DD1361010

Doctoral Dissertation

Behavioral Analysis using Non-Stationary Time Series Modeling Method with Bayesian Nonparametrics

Ryunosuke Hamada

August 1, 2016

Department of Information Science
Graduate School of Information Science
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of ENGINEERING

Ryunosuke Hamada

Thesis Committee:

Professor Kazushi Ikeda	(Supervisor)
Professor Yuji Matsumoto	(Co-supervisor)
Associate Professor Junichiro Yoshimoto	(Co-supervisor)
Research Associate Professor Takatomi Kubo	(Co-supervisor)
Assistant Professor Toshihiro Hiraoka	(Kyoto University)

Behavioral Analysis using Non-Stationary Time Series Modeling Method with Bayesian Nonparametrics[*]

Ryunosuke Hamada

Abstract

Time series data are broadly collected and analyzed in diverse fields of science and engineering. To cope with such data, the non-stationary multidimensional time series modeling method plays a dominant role and is widely applied to the data for predicting future values and unveiling interesting structure of them.

The remarkable progress of Bayesian nonparametric methodology has expanded the scope of the non-stationary time series modeling methods, not only enabling us to automatically determine the number of parameters in the model according to the complexity of the dataset, but also tolerating a certain class of heterogeneity among a set of time series data. However, there is relatively little literature on utilizing the non-stationary time series modeling method with Bayesian nonparametrics for applications of time series data analysis, despite the recent extensive use of methods that are not based on any temporal or structural assumption of the time series data such as deep neural networks.

In this thesis, we present two application studies of the non-stationary time series modeling method with Bayesian nonparametrics. First, we apply the Bayesian nonparametric hidden Markov model to a birdsong dataset, and reveal that distinct syntactic rules are adopted by different groups of birds that have different tutors. Second, we analyze a set of multiple time series data of driver behaviors, and show that driver behaviors in future can be predicted by using the

---

[*] Doctoral Dissertation, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD1361010, August 1, 2016.

Bayesian nonparametric Markov-switching vector autoregressive processes without any model selection procedure. These results support the effectiveness of the non-stationary time series modeling method with Bayesian nonparametrics for behavior analysis.

Keywords:

Non-stationary time series, hidden Markov model, Bayesian nonparametrics, driving behavior modeling, bird song modeling.

ノンパラメトリックベイズ非定常時系列モデリング法による行動解析[*]

濱田 龍之介

内容梗概

時系列データは科学・工学の諸分野において盛んに収集・解析が行われている．時系列データの未来の値を予測したり，その中に隠された興味深い構造を明らかにすることを目的として，多次元非定常時系列モデリング法が広く利用されている．

近年ノンパラメトリックベイズ法のめざましい発展により，非定常時系列モデリング法においてモデルのもつパラメータ数をデータから自動的に決定することが可能になったのみならず，従来法では適切に扱えなかったある種の不均質さをもつデータセットをも扱える手法が開発されている．しかしながら，深層ニューラルネットワークに代表されるような時系列の時間的・構造的な仮定に基づかない手法が頻繁に用いられている一方で，このようなノンパラメトリックベイズ法を用いた非定常時系列モデリング法による解析を行った研究はあまり見られない．

本論文では，我々の行った非定常時系列モデリングの応用研究の2例を紹介する．一つ目は，ノンパラメトリックベイズ隠れマルコフモデルを鳥の歌声データセットに適用し，異なる教師をもつ被験者の2つのグループがそれぞれ異なる統語ルールをもっていることを明らかにする．二つ目は，運転行動時系列データを解析した研究で，ノンパラメトリックベイズ法を用いることでモデル選択を行うことなくマルコフ転換ベクトル自己回帰過程により運転行動をモデル化し，未来の運転行動を予測可能であることを示す．これら2つの結果から，ノンパラメトリックベイズ非定常時系列モデリング法の行動解析における有効性を示す．

キーワード

---

[*] 奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 博士論文, NAIST-IS-DD1361010, 2016年8月1日.

非定常時系列, 隠れマルコフモデル, ノンパラメトリックベイズ法, 運転行動モデリング, 鳥歌モデリング.

Contents

List of Figures

Notational Conventions

Symbol	Definition
$\boldsymbol{y}_t^{(i)}$	Observation vector at time t in time series i
$\boldsymbol{y}_{1:T_i}^{(i)}$	Set of observation vectors for $t = 1, 2, ..., T_i$ in time series i
$\boldsymbol{\varepsilon}_t^{(i)}$	Gaussian noise at time t in time series i
$z_t^{(i)}$	Hidden variable at time for $t = 1, 2, ..., T_i$ in time series i
$z_{1:T_i}^{(i)}$	Set of hidden variables at time t in time series i
$\boldsymbol{\pi}^{(i)}$	State transition probabilities of time series i
T_i	Length of time series i
$\boldsymbol{\mu}_k$	Mean vector of state (Gaussian emission) k
$\mathbf{A}_k$	Coefficient matrix of state (VAR process) k
$\boldsymbol{\Sigma}_k$	Covariance matrix of Gaussian noise of state k
$B \sim \mathrm{BP}(c, B_0)$	Draw of beta process $\mathrm{BP}(c, B_0)$
B_0	Base measure of beta process
c	Concentration parameter of beta process
$X \sim \mathrm{BeP}(B)$	Draw of Bernoulli process $\mathrm{BeP}(B)$
$\boldsymbol{f}_i$	Binary vector whose element f_{ik} represents emergence of state k in time series i

Chapter 1

Introduction

1.1 General Background

Time series data appear in diverse fields of science and engineering. The methodology of time series modeling has been developed and applied to finance [1], behavior analysis such as acoustic signal processing [2] etc. Some intrinsic properties of time series make the analysis of time series difficult compared to those of data that show no temporal dependency. Such difficulty arises especially when we treat non-stationary time series [3]. One of the modeling methods of non-stationary time series is to use an autoregressive integrated moving average (ARIMA) model [4] which models the temporal difference of the original time series as an autoregressive moving average (ARMA) process. The ARIMA models are commonly used to analyze the difference-stationary processes that are often appear in financial time series, but lack the ability to capture sudden changes in a mean or variance of time series, referred to as a structural break [5].

A promising approach of non-stationary time series modeling is to use the models that have their internal states or regimes [3]. A hidden Markov model (HMM) is widely applied to analysis of non-stationary time series, including the outstanding success in speech signal processing [6]. The vector autoregressive processes with Markov-switching regimes, also referred to as autoregressive hidden Markov models (AR-HMM), have been shown their effectiveness in finance [7] and behavior analysis [8]. These models are based on symbolizing or segmenting the non-stationary time series, then considering each symbol or segment to be piecewise stationary.

The main problem of using such non-stationary time series modeling methods with internal states comes from the model selection, that is, how to determine the number of states or regimes. The number can be determined according to certain quantities like the information criterion or estimate of the likelihood. These quantities, however, are inappropriate for the HMM or AR-HMM to determine the number, and calculating the quantities tends to require the expensive computational or human cost.

Bayesian nonparametric methodology can provide a solution to the problem by incorporating stochastic processes with the models. The stochastic processes enable the models to be arbitrarily complex or have countably infinite number of model parameters [9]. Bayesian nonparametric extensions of the HMM and AR-HMM were proposed and can automatically determine the number of states according to dataset [10,11]. In addition, these models can also consider a certain class of heterogeneity in inter-state switching probabilities for the dataset that consists of multiple time series.

In this dissertation we present two studies of behavior analysis. First, we develop the automatic annotation and clustering method for birdsong data by using the Bayesian nonparametric HMM, which has no autoregressive processes in the model. Our method of the automatic annotation is shown to achieve consistent annotation performance with a human specialist, and the clustering method identify subjects of birdsong time series without any information on subject ID or relationship of the subjects. Second, we develop the prediction method of driving behavior data in real cars by using the Bayesian nonparametric AR-HMM. Our method of predicting future behaviors is shown to have the better prediction performance than those based on the methods without any autoregressive assumption on observable variables. From these two results we verify the effectiveness of non-stationary time series modeling methods with Bayesian nonparametrics for the behavior analysis. The measurements of bird song data and driving behavior data were performed and provided by collaborators in RIKEN Brain Science Institute and DENSO CORPORATION, respectively.

1.2 Organization of Dissertation

The rest of this dissertation is organized as follows. Chapter 2 provides brief introductions of the hidden Markov model (HMM), autoregressive hidden Markov model (AR-HMM), Bayesian nonparametric methodology, and Bayesian nonparametric extensions of the HMM and AR-HMM. Chapter 3 describes the analysis of bird songs and clustering of the birdsong dataset without any prior knowledge. Chapter 4 describes the analysis of driving behaviors and prediction of the future driving behaviors. Chapter 5 concludes this dissertation and give discussions of our works.

Chapter 2

Hidden Markov model and Bayesian nonparametrics

In this chapter, we shortly review a hidden Markov model and its extensions, Bayesian nonparametric methods, and Bayesian nonparametric extensions of the hidden Markov model.

2.1 Hidden Markov model

2.1.1 Formulation of hidden Markov model

The hidden Markov model (HMM) is a probabilistic model of time series [6], which is widely used to model time series data in various applications, e.g., speech processing [12, 13], genomics [14], and human motion [15]. The HMM mainly consists of two stochastic components: first is the emission of observable variables that is characterized by a set of distributions and parameters of them, and second is a discrete-time stochastic process referred to as a Markov chain [16] that has finite states associated with the sets of parameters. These stochastic components enable us to convert time series into a sequence of symbols, that is, the HMM transforms an observed value of the signal into a certain state at each time point, according to similarity to representative values of clusters and probabilities of transitions between the clusters. This symbolization of the signal simplifies time series that have inherent noise, and results in clustering or recognition of the signals.

At each time point t of time series, the HMM assigns a hidden variable z_t whose value is in $\{1, 2, \ldots, K\}$. Here the number of states K is a fixed constant. We say a *state* is k at time t when z_t equals to k. Let the length of the time series to be T. A sequence of hidden variables $[z_1, z_2, \ldots, z_T]$ is a Markov chain, then for $1 < t \leq T$

$$p(z_t = l | z_{t-1} = k) = \pi_{kl}, \qquad k, l = 1, 2, \ldots, K, \tag{2.1}$$

$$\sum_l \pi_{kl} = 1, \qquad k = 1, 2, \ldots, K, \tag{2.2}$$

where $\boldsymbol{\pi} = (\pi_{kl})$ is a $K \times K$ transition probability matrix of the states. An initial state, namely z_1, is determined according to the unconditional probability of the states,

$$p(z_1 = k) = \pi_k^{init}, \qquad k = 1, 2, \ldots, K, \tag{2.3}$$

$$\sum_k \pi_k^{init} = 1, \tag{2.4}$$

where $\boldsymbol{\pi}^{init}$ is a K dimensional vector. If the state is given at time t an observation vector $\boldsymbol{y}_t$ is generated by a certain class of distribution, typically and in our case a Gaussian distribution. This distribution is associated with the given state for time $1 \leq t \leq T$,

$$\boldsymbol{y}_t \mid z_t = k \ \sim \ \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \qquad k = 1, 2, \ldots, K, \tag{2.5}$$

where $\mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ denotes a Gaussian distribution with mean and variance of $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$, respectively. Graphical model of the HMM that omits initial probability $\boldsymbol{\pi}^{init}$ is shown in Fig. 2.1. This formulation of the HMM results in the likelihood of whole time series $\boldsymbol{y}_{1:T} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T]$ and $z_{1:T} = [z_1, \ldots, z_T]$,

$$p(z_{1:T}, \boldsymbol{y}_{1:T}) = p(z_1) p(\boldsymbol{y}_1 | z_1) \prod_{t=2}^{T} p(z_t | z_{t-1}) p(\boldsymbol{y}_t | z_t). \tag{2.6}$$

Direct maximization of the likelihood function is intractable, so instead EM algorithm is often used [17, 18].

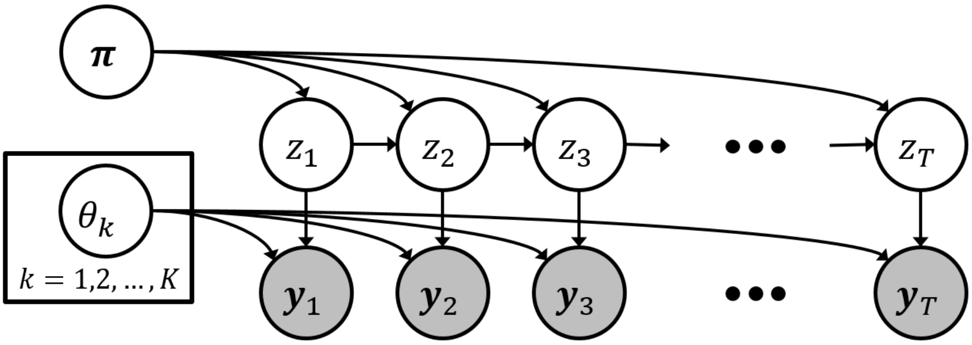


Figure 2.1. Graphical model of HMM. $\boldsymbol{y}_t$, observable variable; z_t, hidden state at time point t; $\theta_k = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$, parameters of k-th Gaussian distribution; $\boldsymbol{\pi}$, state transition probabilities.

2.1.2 Estimation of parameters of HMM

Parameters of the HMM are often estimated by using the EM algorithm, which consists of expectation step (E step) and maximization step (M step) [6, 19]. In E step given the parameters and observations $\boldsymbol{y}_{1:T}$, posterior probabilities $\gamma(z_t) \triangleq p(z_t|\boldsymbol{y}_{1:T})$ and joint posterior probabilities $\xi(z_{t-1}, z_t) \triangleq p(z_{t-1}, z_t|\boldsymbol{y}_{1:T})$ are efficiently calculated by the forward-backward algorithm [17, 18, 20]. From the Bayes theorem and the conditional independence of one order Markov chain, γ is factorized into

$$\begin{aligned}\gamma(z_t) = p(z_t|\boldsymbol{y}_{1:T}) &= \frac{p(\boldsymbol{y}_{1:T}|z_t)p(z_t)}{p(\boldsymbol{y}_{1:T})}\\ &= \frac{p(\boldsymbol{y}_{1:t}, z_t)p(\boldsymbol{y}_{t+1:T}|z_t)}{p(\boldsymbol{y}_{1:T})}\\ &\triangleq \frac{\alpha(z_t)\beta(z_t)}{p(\boldsymbol{y}_{1:T})}. \end{aligned} \tag{2.7}$$

For forward recursion, the $\alpha(z_t)$ $(t = 2, 3, \dots, T,\ z_t = 1, 2, ..., K)$ can be recursively calculated from a following relationship

$$\alpha(z_t) = p(\boldsymbol{y}_t|z_t) \sum_{z_{t-1}\in\{1,2,\dots,K\}} \alpha(z_{t-1})p(z_t|z_{t-1}), \tag{2.8}$$

where the initial value is obtained by multiplying an initial probability of states by an emission probability given a state

$$\alpha(z_1) = p(\boldsymbol{y}_1, z_t) = p(z_t)p(\boldsymbol{y}_1|z_t). \tag{2.9}$$

For backward recursion, the $\beta(z_t)$ $(t = T-1, T-2, \ldots, 1,\ z_{t+1} = 1, 2, ..., K)$ can be recursively calculated as follow

$$\beta(z_t) = \sum_{z_{t+1} \in \{1,2,\ldots,K\}} \beta(z_{t+1}) p(\boldsymbol{y}_{t+1}|z_{t+1}) p(z_{t+1}|z_t), \tag{2.10}$$

and the initial value is $\beta(z_T) = 1$. The $\xi(z_{t-1}, z_t)$ can be calculated by using $\alpha(z_{t-1}), \beta(z_t)$

$$\xi(z_{t-1}, z_t) = \frac{\alpha(z_{t-1}) p(\boldsymbol{y}_t|z_t) p(z_t|z_{t-1}) \beta(z_t)}{p(\boldsymbol{y}_{1:T})}. \tag{2.11}$$

In M step the estimation of means and covariance matrices $\{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$ are calculated by using the posteriors calculated in E step as follow

$$\boldsymbol{\mu}_k = \frac{\sum_{z_t=k} \alpha(z_t)\beta(z_t)\boldsymbol{y}_t}{\sum_{z_t=k} \alpha(z_t)\beta(z_t)}, \tag{2.12}$$

$$\boldsymbol{\Sigma}_k = \frac{\sum_{z_t=k} \alpha(z_t)\beta(z_t)(\boldsymbol{y}_t - \boldsymbol{\mu}_k)(\boldsymbol{y}_t - \boldsymbol{\mu}_k)^{\mathrm{T}}}{\sum_{z_t=k} \alpha(z_t)\beta(z_t)}. \tag{2.13}$$

Iterative execution of E step and M step continues until the likelihood satisfies a certain condition, typically such that the value of likelihood ceases to increase.

The number of states K is fixed for the HMM, and determining the number might be a crucial problem. To determine the number, an information criterion such as BIC is conventionally used [21]. Another approach is to estimate the generalization error under a certain K by cross validation [22]. Although these approaches are commonly adopted, there are some difficulties to use them. First, it takes relatively long time to train a lot of the HMMs that have different Ks. Second, we do not certain how much K possibly ranges in advance. Third, the information criterion is not appropriate to use for singular models such as the HMM [23].

2.1.3 Estimation of most probable sequence

We are sometimes interested in the state sequence itself. Given the observations $\boldsymbol{y}_{1:T}$ and parameters of the HMM, most probable sequence of states can be estimated for the time series. The most probable sequence can be obtained by maximizing the likelihood of hidden variables $z_{1:T}$. Its direct calculation needs the evaluation of the likelihood whose computational cost is proportional to all possible combinations of states K^T.

There exists an efficient algorithm based on dynamic programming to obtain the most probable sequence, called the Viterbi algorithm [6, 17, 18, 24]. The computational cost of the Viterbi algorithm is proportional to K^2T, which takes drastically smaller time than the direct computation. The Viterbi algorithm recursively calculates the joint probability of hidden variable sequence at time t,

$$\begin{aligned}\delta_t(k) &\triangleq \max_{z_{1:t-1}} \; p(z_{1:t-1}, z_t = k | \boldsymbol{y}_{1:t}) \\ &= \max_{j} \; \delta_{t-1}(j) p(z_t = k | z_{t-1} = j) p(\boldsymbol{y}_t | z_t = k).\end{aligned} \tag{2.14}$$

At the same time it stores history of the most probable previous state when a state is k at time t,

$$a_t(k) \triangleq \operatorname*{argmax}_{j} \; \delta_{t-1}(j) p(z_t = k | z_{t-1} = j) p(\boldsymbol{y}_t | z_t = k). \tag{2.15}$$

At first the algorithm forward calculates from

$$\delta_1(k) = p(z_1 = k) p(\boldsymbol{y}_1 | z_1 = k), \tag{2.16}$$

to $\delta_T(k)$. Next it gets the most probable state at time T by

$$z_T^* = \operatorname*{argmax}_{k} \; \delta_T(k). \tag{2.17}$$

Finally it traces back from time T to 1 by using the following relationship,

$$z_{t-1}^* = a_t(z_t^*), \tag{2.18}$$

then finally we get the most probable sequence of hidden states.

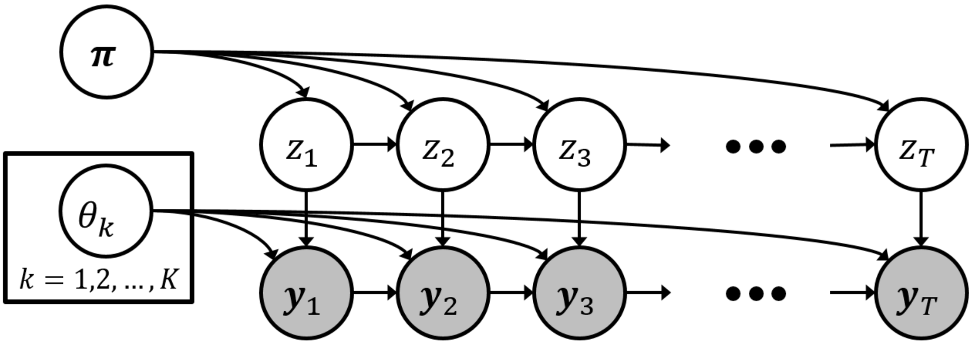


Figure 2.2. Graphical model of AR-HMM. $\theta_k = \{\mathbf{A}_k, \mathbf{\Sigma}_k\}$, VAR parameters of state k; the rest is the same as in Fig. 2.1.

2.2 Autoregressive hidden Markov model

2.2.1 Formulation of autoregressive hidden Markov model

The autoregressive hidden Markov model (AR-HMM), also known as Markov-switching vector autoregressions [25], is an extension of the HMM in which observation vectors $\boldsymbol{y}_{1:T}$ have temporal dependency and are subject to several vector autoregressive (VAR) processes [6, 26, 27]. As is for the HMM, the AR-HMM assigns a hidden variable z_t at each time t and it is subject to a Markov chain with a finite state set $\{1, 2, \ldots, K\}$. If we assume the order of VAR processes is 1, given the state is k the generative process of observation vectors is written as

$$\boldsymbol{y}_t \mid z_t = k \;\sim\; \mathcal{N}(\mathbf{A}_k \boldsymbol{y}_{t-1}, \mathbf{\Sigma}_k), \qquad k = 1, 2, \ldots, K, \tag{2.19}$$

where $\mathbf{A}_k, \mathbf{\Sigma}_k$ are a coefficient matrix and covariance matrix of a Gaussian distribution, respectively (here, we use the same notation for a covariance matrix as in the HMM). In the AR-HMM the observation vectors are subject to the identical VAR process as long as the hidden variables take the identical values. Graphical model of the AR-HMM that omits initial probability $\boldsymbol{\pi}^{init}$ is shown in Fig. 2.2.

2.2.2 Estimation of parameters and most probable state sequence of AR-HMM

Parameters of the AR-HMM can be estimated by using EM algorithm as well as the HMM. E step, that is, the calculation of the posteriors $p(z_t|\boldsymbol{y}_{1:T})$ and $p(z_{t-1}, z_t|\boldsymbol{y}_{1:T})$ is the same as the HMM. For M step, the calculation of transition matrix and initial probability is also the same as the HMM, but some modulations are necessary for VAR parameters $[\mathbf{A}_1, \ldots, \mathbf{A}_K, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K]$ [18, 26]. The covariance matrices $[\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K]$ are estimated as

$$\boldsymbol{\Sigma}_k = \frac{\sum_{z_t=k} \alpha(z_t)\beta(z_t)(\boldsymbol{y}_t - \mathbf{A}_k\boldsymbol{y}_{t-1})(\boldsymbol{y}_t - \mathbf{A}_k\boldsymbol{y}_{t-1})^{\mathrm{T}}}{\sum_{z_t=k} \alpha(z_t)\beta(z_t)}, \tag{2.20}$$

then the coefficient matrices $[\mathbf{A}_1, \ldots, \mathbf{A}_K]$ are estimated by solving K weighted linear regression problems

$$J(\mathbf{A}_k) = \sum_{t:z_t=k} \alpha(z_t)\beta(z_t)(\boldsymbol{y}_t - \mathbf{A}_k\boldsymbol{y}_{t-1})^{\mathrm{T}}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{y}_t - \mathbf{A}_k\boldsymbol{y}_{t-1}), \tag{2.21}$$

which can be efficiently solved by the Levinson-Durbin method [18]. The number of states K should be determined also for the AR-HMM, typically by using the information criterion or cross validation. The most probable state sequence can be efficiently estimated by using the Viterbi algorithm as well as the HMM.

2.3 Bayesian HMM

The HMM can be extended to the Bayesian modeling by introducing prior distributions of parameters. Posterior distributions of the parameters are calculated by multiplying the priors to a likelihood function in (2.6). Practically maximum a posteriori (MAP) parameters that maximize the posterior distributions are obtained as an estimate of parameters of the Bayesian HMM. Direct maximization is intractable as well as the conventional HMM, so the variational Bayes approximation [28, 29] or MCMC sampling inference [30, 31] are used. The Bayesian methodology can also be applied to the AR-HMM. In both the Bayesian HMM and AR-HMM, determining the number of states K is still a dominant problem.

2.4 Bayesian nonparametric extension of HMM

2.4.1 Bayesian nonparametric methods

Bayesian nonparametric methods have attracted interests of machine learning researchers in recent years, and give flexibility to the parameter space of statistical models. The core idea of Bayesian nonparametrics is to tolerate the number of model parameters being countably infinite by incorporating certain classes of stochastic processes as priors, in other words, to determine the complexity of models according to a dataset.

A natural question to the Bayesian nonparametrics would be, why should we use the stochastic process as a prior, not only the probability distributions? This necessity comes from the motivation that we set a prior on the probability distributions or functions themselves. For clustering problem, a Dirichlet process [32] is often used as a prior on the mixing weights of the mixture model like a Gaussian mixture model, and the Dirichlet process allow the number of mixture components to be an arbitrary positive integer [33]. For regression problem, a Gaussian process provides a prior on a nonlinear function that approximates the relationship between input and output variables, and the function is estimated being regularized by the Gaussian process [34]. In both cases the stochastic processes let the model be infinitely complex (or countably infinite number of the model parameters) in theory.

In this section we briefly explain a beta process, which is used as a building block of a Bayesian nonparametric extension of the HMM. For general information of Bayesian nonparametrics, see [9, 35–37].

2.4.2 Beta process and Bernoulli process

A beta process was originally proposed by Hjort as a prior of a cumulative hazard rate for survival analysis [38]. Thibaux and Jordan [39] defined a Bernoulli process as a conjugate process of the beta process, and found that their combinatorial process, a beta-Bernoulli process, is a de Finetti's mixing distribution of Indian buffet process proposed by Griffiths and Ghahramani [40].

The beta process is one of the classes of stochastic processes, called as Lévy processes [41], and is completely characterized by its Lévy measure. Let Θ denotes a parameter space of the model, and $\theta \in \Theta$ a point in the parameter space. Then a Lévy measure of the beta process on $[0,1] \times \Theta$ is defined as,

$$\nu_{\mathrm{BP}}(d\omega, d\theta) = c\,\omega(1-\omega)^{c-1} d\omega B_0(d\theta), \tag{2.22}$$

where $c > 0, B_0$ are a concentration parameter and a base measure that corresponds to a prior distribution of the model parameters, respectively [11,39]. Here $\omega \in [0,1]$ is an abstract random variable, but plays a relevant role in a Bayesian nonparametric HMM described in the next subsection. We say $B \sim \mathrm{BP}(c, B_0)$ is a beta process that has c and B_0 as its parameters.

If B_0 is continuous on Θ, the beta process B has a form

$$B = \sum_{k=1}^{\infty} \omega_k \delta_{\theta_k}, \tag{2.23}$$

where δ_{θ_k} is a Dirac measure at an atom θ_k on Θ [42]. Intuitive illustration of B is an countably infinite set of pillars whose heights are ω_k at the positions of θ_k on the parameter space Θ.

Let us think an independent series of random variables $X_1, X_2, \ldots, X_N$, each of them is actually a Bernoulli process. A Lévy measure of the Bernoulli process $X_i \sim \mathrm{BeP}(B)$ $(i = 1, 2, \ldots, N)$ on $\{0,1\}^{\infty} \times \Theta$ is

$$\nu_{\mathrm{BeP}}(df, d\theta) = \delta_1(df) B(d\theta), \tag{2.24}$$

where δ_1 is a Dirac measure at 1. Then X_i has a form

$$X_i = \sum_{k=1}^{\infty} f_{ik} \delta_{\theta_k}, \qquad i = 1, 2, \ldots, N, \tag{2.25}$$

where $f_{ik} \sim \mathrm{Bern}(\omega_k)$ is a binary valued variable that takes 1 with probability $0 \le \omega_k \le 1$ or 0 otherwise.

The important consequence of this formulation is a posterior of beta process B given $X_1, X_2, \ldots, X_N$,

$$B \mid X_1, X_2, \ldots, X_N, B_0, c \;\sim\; \mathrm{BP}\left(c+N,\; \frac{c}{c+N} B_0 + \sum_{k=1}^{\infty} \frac{m_k}{c+N} \delta_{\theta_k} \right), \tag{2.26}$$

where $m_k = \sum_{i=1}^{N} f_{ik}$. The base measure of the posterior is a weighted sum of the base measure of the prior and an empirical measure.

2.4.3 Beta process hidden Markov model

Formulation of beta process hidden Markov model

The problem of the HMM as mentioned above is that one should fix the number of the states K before estimating the model parameters, and then determine the number by comparing the fitness of different Ks according to the information criterion or performing the cross validation. Incorporating stochastic processes enable us to avoid this problem by introducing an unfixed dimensional parameter space in the model.

Bayesian nonparametric approaches imply prior stochastic processes to the emergence of parameters in a dataset, not only the prior distributions to the parameters. Fox et al. proposed a Bayesian nonparametric extension of the HMM, a sticky hierarchical Dirichlet process HMM (HDP-HMM) [43], which use a hierarchical Dirichlet process prior to automatically determine the number of the states K.

Although the HDP-HMM can avoid the model selection, there remains another problem for modeling a set of time series. The problem arises from a kind of heterogeneity among multiple time series, that is, when multiple time series are diverse in terms of transition probabilities between states that appear in the HMM, it might be appropriate to have a tailored state transition probability matrix for each single time series, rather have just one matrix for all time series.

The problem of heterogeneity can be solved by considering distinct emergences of states in different time series data. Such formulation of a Bayesian nonparametric HMM, called a beta process HMM (BP-HMM), was proposed by Hughes et al. [10]. The BP-HMM not only automatically determines the number of states, but also consider an emergence of states and a tailored state transition probability matrix for each time series, in contrast to the HDP-HMM that has only one transition probability matrix in the model.

Let us consider a set of N time series data $[\boldsymbol{y}^{(1)}_{1:T_1}, \boldsymbol{y}^{(2)}_{1:T_2}, \ldots, \boldsymbol{y}^{(N)}_{1:T_N}]$ each of which is of length T_i $(i = 1, 2, \ldots, N)$. Each time series i is associated with a Bernoulli process $X_i = \sum_k f_{ik}\delta_{\theta_k}$ that represents an emergence of states in the time series. If $\boldsymbol{f}_i \triangleq [f_{i1}, f_{i2}, \ldots, f_{ik}, \ldots]$ has K_i non-zero entries, then time series i is modeled by the HMM with K_i states. Following a notation in [11], the j-th row of a tailored transition matrix, referred to as a feature-constrained transition matrix of time series i, can be written as

$$\pi_k^{(i)} \mid \boldsymbol{f}_i, \gamma, \kappa \;\sim\; \mathrm{Dir}\left([\gamma, \ldots, \gamma, \gamma+\kappa, \gamma, \ldots,] \odot \boldsymbol{f}_i\right), \tag{2.27}$$

where $\mathrm{Dir}(\cdot)$ denotes a Dirichlet distribution, $\odot$ element-wise product, and γ, κ are hyperparameters. Here the k-th element of $\pi_k^{(i)}$ has a heavier weight than the other elements, which makes each state sticky to itself.

Now we see the generative process of the BP-HMM

$$z_t^{(i)} \mid z_{t-1}^{(i)} \;\sim\; \pi_{z_{t-1}^{(i)}}^{(i)}, \tag{2.28}$$

$$\boldsymbol{y}_t^{(i)} \mid z_t^{(i)} \;\sim\; \mathcal{N}(\boldsymbol{\mu}_{z_t^{(i)}}, \boldsymbol{\Sigma}_{z_t^{(i)}}). \tag{2.29}$$

Graphical model of the BP-HMM is shown in Fig. 2.3.

Prior settings and estimation of parameters and most probable state sequence of BP-HMM

Prior distributions on Gaussian emission parameters, means and covariance matrices $\{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$, are set to normal distributions given $\boldsymbol{\Sigma}_k$, and inverse-Wishart distributions, respectively. Posterior distributions of the emission parameters are again the normal distributions and inverse-Wishart distributions with different hyper parameters.

Parameter estimation of the BP-HMM is conducted by using Markov chain Monte Carlo (MCMC) sampling [10]. To briefly describe, the sampling from the beta process and Bernoulli process is based on the Metropolis-Hastings proposal, birth and death proposal, and split-merge moves. Parameters of the Gaussian distributions and state transition probabilities are sampled from the posterior, and the most probable state sequences are block sampled by using dynamic programming. The hyperparameters for the beta process, α and c, and the Dirichlet distribution prior, γ and κ, can be also estimated. α is sampled from the posterior. The posteriors of c, γ and κ are not conjugate, so these hyperparameters are updated by Metropolis-Hastings steps [44].

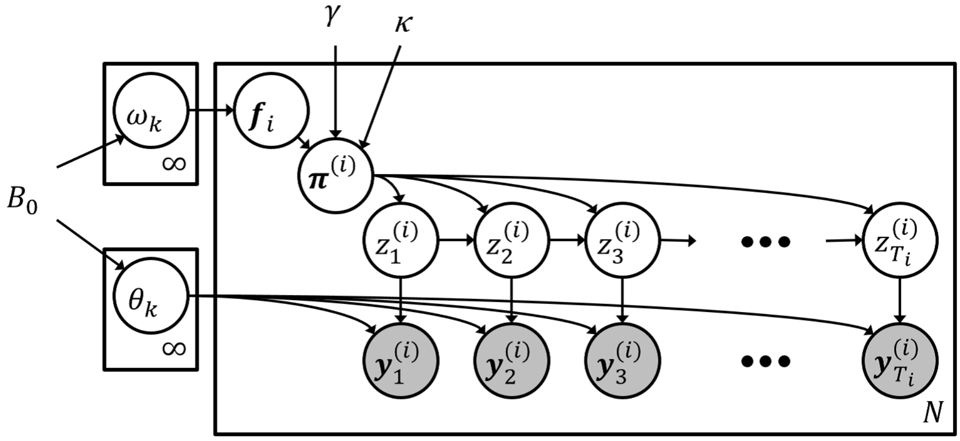


Figure 2.3. Graphical model of BP-HMM. $\boldsymbol{y}_t^{(i)}$, observable variable; $z_t^{(i)}$, hidden state at time point t in time series i; $\theta_k = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$, parameters of k-th Gaussian distribution; $\boldsymbol{\pi}^{(i)}$, a set of state transition probabilities $\pi_k^{(i)}$; $\boldsymbol{f}_i$, emergence of the states in time series i; κ, γ, hyperparameters; B_0, base measure.

2.4.4 Beta process autoregressive hidden Markov model

Formulation of beta process autoregressive hidden Markov model

The AR-HMM can be incorporated with the beta process and Bernoulli processes in the similar way for the HMM. Fox et al. proposed a beta process AR-HMM (BP-AR-HMM) [11], as a Bayesian nonparametric extension of the AR-HMM. The difference from the BP-HMM is that the generative process is the switching VAR processes

$$z_t^{(i)} \mid z_{t-1}^{(i)} \sim \pi_{z_{t-1}^{(i)}}^{(i)}, \tag{2.30}$$

$$\boldsymbol{y}_t^{(i)} \mid z_t^{(i)} \sim \mathcal{N}(\mathbf{A}_{z_t^{(i)}} \boldsymbol{y}_{t-1}^{(i)}, \boldsymbol{\Sigma}_{z_t^{(i)}}). \tag{2.31}$$

Graphical model of the BP-AR-HMM is shown in Fig. 2.4.

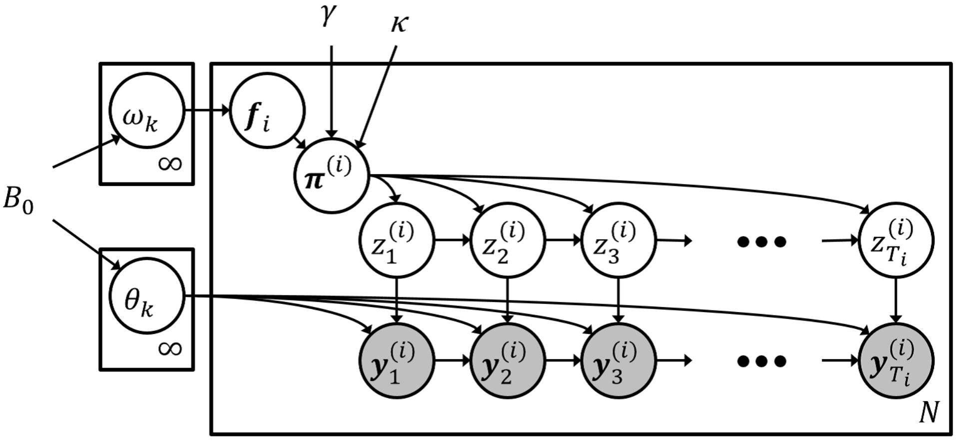


Figure 2.4. Graphical model of BP-AR-HMM. $\theta_k = \{\mathbf{A}_k, \mathbf{\Sigma}_k\}$, VAR parameters of state k; the rest is the same as in Fig. 2.3.

Prior settings and estimation of parameters and most probable state sequence of BP-AR-HMM

Prior distributions on dynamic parameters, VAR coefficient matrices and covariance matrices $\{\mathbf{A}_k, \mathbf{\Sigma}_k\}_{k=1}^K$, are set to matrix normal distributions given $\mathbf{\Sigma}_k$, and inverse-Wishart distributions, respectively. As is the BP-HMM, posterior distributions of the emission parameters are again the matrix normal distributions and inverse-Wishart distributions with different hyper parameters.

The parameters are estimated by using the MCMC sampling with birth and death proposals, without split-merge moves [11]. The VAR parameters and state transition probabilities are sampled from the posterior, and the most probable state sequences are block sampled by using dynamic programming. The hyperparameter of the beta process c is set to 1. The other hyperparameters, α, γ and κ are estimated as is for the BP-HMM.

Chapter 3

Syntactic rule analysis of birdsong data

3.1 Introduction

Human beings can speak languages, which are syntactically and hierarchically organized behaviors. In spoken languages, syllables constitute words, and words further sentences. The order of words provides meaning of sentences with complex sequencing rule, *syntax*. The process how neural circuits in the human brain regulate this syntactic behavior is still unknown and difficult to examine, because of its high complexity.

A certain kind of oscine, called Bengalese finch (*Lonchura striata* var. *domestica*), is known to sing with apparently complex syntactic rules. The Bengalese finch is one of the most ideal animal model for exploring neural circuits of speech production [45]. Transition between syllables of the Bengalese finch song is regarded as non-deterministic, which is different from deterministic songs of the zebra finch, White-rumped Munia (an ancestral species of Bengalese finch), etc.

For a study of the complex syntactic rule of Bengalese finches, large scale analysis is necessary including comparison between syllable sequencings of different birds. Such large scale analysis, however, needs large amount of manual annotation of the songs, which is impractical or requires high labor costs. Therefore developing automatic annotation and analysis for the songbird is imperative.

For the annotation of the Bengalese finchs' songs the HMM was used [46].

Also, annotation performance of the first-order HMM was comparable to that of the second order HMM. They individually modeled the song of each subject by using its individual HMM. The HMMs were trained for the different numbers of states.

The conventional HMM is not appropriate for comparison of syntactic rules of different birds. Such comparison of syntactic rules is necessary for a study that investigate how songs propagate from a father to its children. In addition to this, the comparison of the songs of an identical subject is also necessary. This is due to variability in sequencing even within the songs of the identical bird [47, 48]. The conventional HMM does not allow the transition probability matrix to have variability, so evaluation of the variations in sequencing is difficult.

In order to model multiple time series of diverse songs simultaneously, it is necessary to utilize a novel method that can share the set of common features corresponding to sharing of the syllables, and consider variations of syntactic rules among multiple different condition, e.g. different individuals, development stages, experimental intervention, etc. In this study, we use the BP-HMM as the method that can model multiple time series data considering common or different features across them and automatically determine the number of the features that corresponds to that of hidden states.

There are two purposes in this study. One is to evaluate the performance of automatic annotation with the BP-HMM and the other is to assess whether the BP-HMM can discriminate songs of different birds, and different syntactic rules of an identical bird in an unsupervised way.

3.2 Birdsong modeling

In this section we first explain the recording condition of Bengalese finches' songs and labeling by human experts. Next we introduce the formulation of birdsong with beta process hidden Markov model (BP-HMM), which can model multiple bird son time series considering common and different features among them. In the last of this section, we introduce similarity analysis of the estimated syntactic rules, or transitional probability matrices, of the bird songs.

3.2.1 Birdsong time series data

In this study, a part of data from those used in [46] was used[1]. The data consist of undirected songs (songs in the absence of a female) of 4 male adult Bengalese finches (BF1–4). All experimental procedures and housing conditions were approved by the Animal Care and Use Committee at RIKEN (approval ID: H22-2-217). BF1 and BF2 learned songs from their father, tutor A, as well as BF3 and BF4 from tutor B. Their songs were recorded during 24 hours using a microphone placed in sound attenuation chambers. The recorded songs were separated into lumps of twittering, referred to as *bouts*, which are sequences of song elements, referred to as *syllables*.

Three acoustic features were extracted for each syllable: syllable duration, mean pitch, and mean Wiener (spectral) entropy [49]. These features were normalized to have means of zero and variances of one. Thirty bouts for each subject were randomly picked and used for analysis below. Distributions of acoustic features fitted by Gaussian mixture model are visualized (Fig. 3.1). Some distributions of syllables are overlapped in the acoustic feature space, and others have multiple modes.

Three human experts annotated syllables in all bouts of subjects based on visual inspections on sonogram. These annotations were confirmed by calculating Fleiss's κ coefficient [50] from the annotated syllable sequences among different annotators, and the values fell within the range of "almost perfect agreement" $(0.81 - 1.00)$. Then we use the annotation result by a single expert.

3.2.2 Modeling birdsong with BP-HMM

We modeled the syntactic rules of Bengalese finches' songs from multiple time series data of acoustic feature by using one of the Bayesian nonparametric approaches, a BP-HMM [10, 11]. The emission distributions are Gaussian distributions in this study. Each time series intuitively corresponds to a bout of a subject, and its state sequence to a series of syllables. We used the code of the BP-HMM developed by Hughes et al. (available from [51]). With this code, the parame-

---

[1] The data were measured and provided by a group of Dr. Kazuo Okanoya in RIKEN Brain Science Institute, Hirosawa, Wako-shi, Saitama 351-0198, Japan.

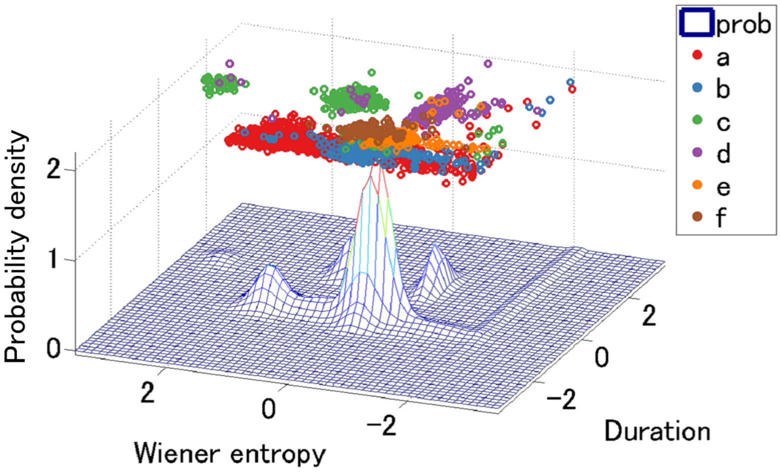


Figure 3.1. Distributions of acoustic features of BF3. The distributions is fitted by Gaussian mixture model, plotted with a scatter plot of acoustic features. Axes of a horizontal plane are duration and mean Wiener entropy of a syllable, which are normalized to have means of zero and variances of one respectively. Each color of the scatter plot represents a manually annotated syllable a–f.

ters were estimated with Markov chain Monte Carlo (MCMC) sampling method, utilizing split-merge moves [52] and data-driven reversible jump MCMC [53].

The BP-HMM is applied to 120 time series data (30 bouts for each subject), and 120 transition matrices of hidden states, and 120 state sequences are obtained. The sets of Gaussian distribution parameters corresponding to the states are also estimated, and the number of the states is automatically determined according to the dataset.

3.2.3 Consistency of estimated syllable sequences with manual annotation

For the automatic annotation, it is necessary to translate the estimated state sequences into syllable sequences. Mapping from the set of the BP-HMM's states to the set of syllables labeled manually is defined, according to the highest conditional probability of a syllable given a state. These mappings are allowed to be many-to-one mappings (from a state to a syllable). After translating the state

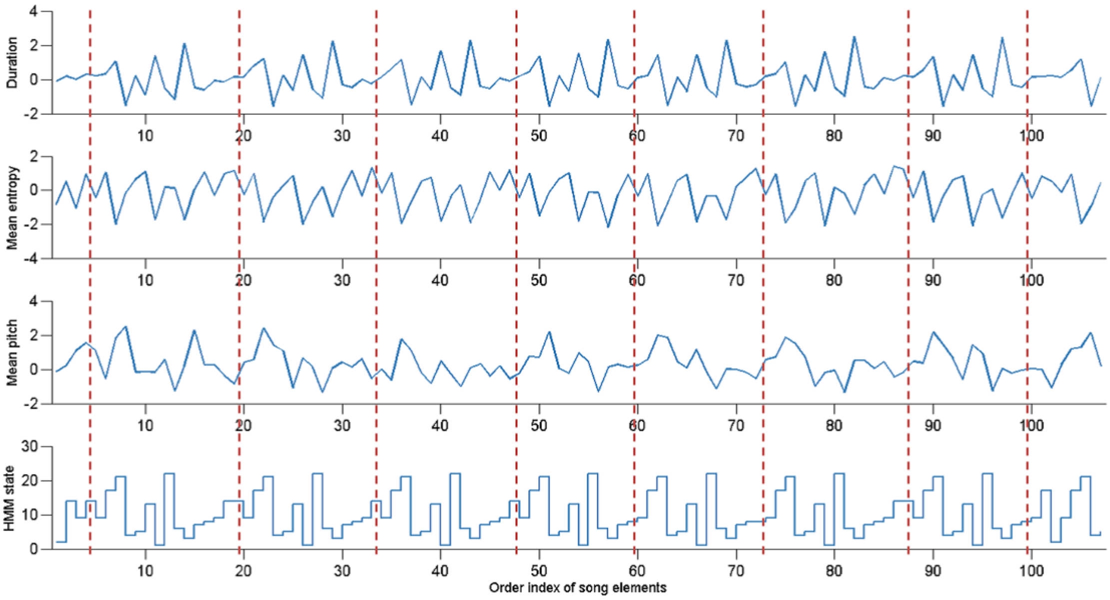


Figure 3.2. Examples of relationship between state sequence and time-courses of the features (a bout of BF1). Horizontal axis represents the index of song elements in the bout arranged in time order. Vertical dashed red lines are partitions of similar state transition patterns that occurred periodically. The partitions are drawn by visual inspection.

sequences to the estimated syllable sequences, the consistency of the sequences and syllable sequences manually annotated is evaluated by calculating Cohen's κ coefficient between them [50].

3.2.4 Calculating similarity between transition matrices and hierarchical clustering

To evaluate variations of syntactic rules, the similarity measure between two syllable transition rules are needed. For such evaluation of the syllable transition probabilities of Bengalese finches, Okanoya and Yamaguchi proposed that Pearson's correlation coefficient of syllable transition matrices be calculated after vectorization [45]. In this study, we calculate Pearson's correlation coefficients between all possible combinations of state transition matrices of the bouts, in order to examine how state transition probabilities differ within and between individual subjects. In addition, we conduct the hierarchical clustering analysis according to these Pearson's correlation coefficients [54].

3.3 Result

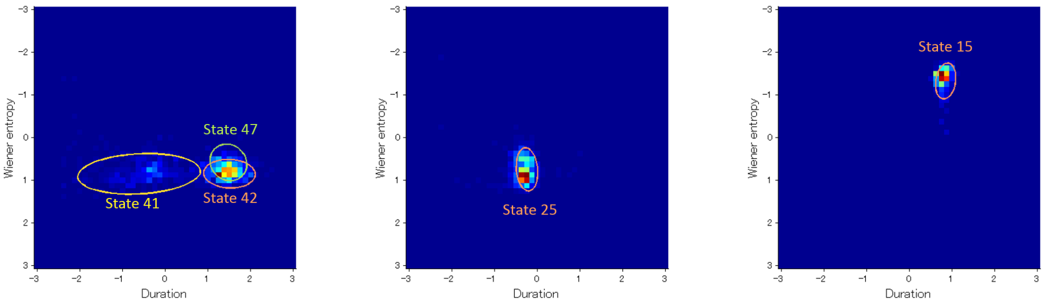
Automatic annotation of syllables

We obtained a state sequence for each bout and 72 sets of Gaussian distribution parameters corresponding to the states, 120 transition matrices, and 120 state sequences. In bouts of each subject, 52, 52, 47, and 53 states appeared for BF1, BF2, BF3, and BF4, respectively. BF1 and BF2 shared 43 states, on the other hand BF3 and BF4 34 states. Even within the individuals that have different tutors, some states were shared (39, 43, 38 and 41 for BF1-BF3, BF1-BF4, BF2-BF3 and BF2-BF4, respectively). Fig. 3.2 shows an example of relationship between state sequences and feature's time-course. Periodical change in the feature spaces are reflected by the state sequence with high reproducibility.

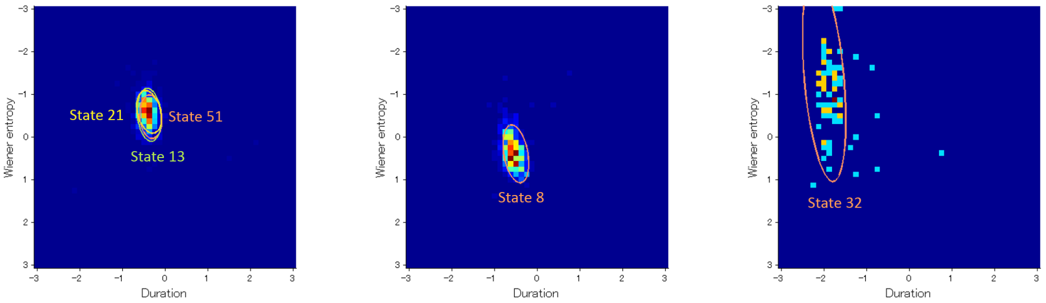
In order to evaluate the annotation performance, we calculated the consistency of estimated syllable sequences and human annotation. Cohen's κ coefficients are 0.8084, 0.7373, 0.8153, and 0.9332 for BF1, BF2, BF3 and BF4, respectively. For BF1, BF3 and BF4 the values are within the range of "almost perfect agreement (0.81–1.00)", and for BF2 "substantial agreement (0.61–0.80)". This result suggests that the automatic annotations based on BP-HMM are consistent with the labels annotated by human.

The estimate of Gaussian emission parameters were also consistent with the actual distributions of acoustic features of syllables (for BF4 in Fig. 3.3). The Gaussian distributions covered the actual distributions of syllables. Several Gaussian distributions overlapped in the feature space but were distinguished as different states (state 42 and 47 for syllable a; state 13, 21 and 51 for syllable d), due to distinct transition probabilities.

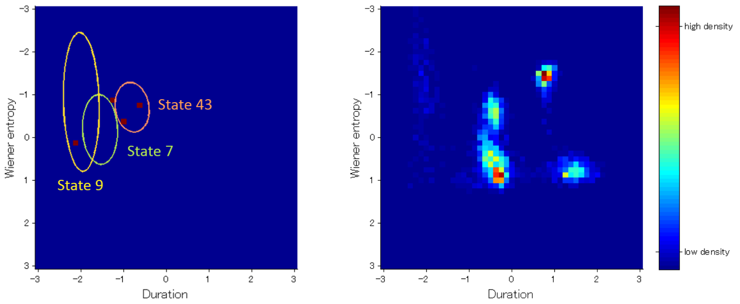
Some states were shared across different subjects (Fig. 3.4). State 47 (corresponds to syllable a of BF3 and BF4) only appeared in bouts of BF3 and BF4, for 278 and 127 times, respectively. Distributions of acoustic features of individual



(a) Syllable a (b) Syllable b (c) Syllable c

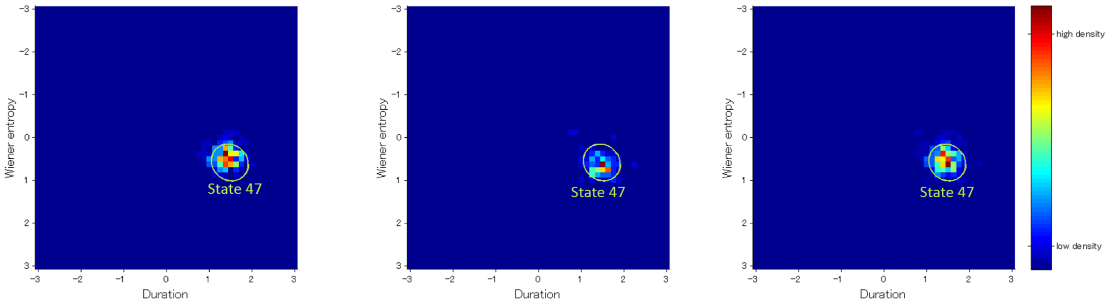


(d) Syllable d (e) Syllable e (f) Syllable f



(g) Syllable g (h) All syllables

Figure 3.3. 2D histogram of syllables of BF4 in feature space. Each color of the histogram represents the degree of density. Ellipsoids are contour lines of Gaussian distributions corresponding to each syllable (standard deviation multiplied by 2). The value of the features were normalized.



(a) BF3 (b) BF4 (c) Both BF3 and BF4

Figure 3.4. 2D histogram of state 47 of BF3 and BF4 in feature space. Color and ellipsoids are the same as in Fig. 3.3.

subjects were apparently similar. BP-HMM unwrapped the sharing of syllables among the different birds.

Clustering analysis of bouts

Correlations within individual subjects are higher than those between subjects (Fig. 3.5). Histograms of the correlation of transition matrices within subjects are shown in Fig. 3.6. Individual correlation coefficients of BF1 and BF2 were significantly low in comparison with those of BF3 and BF4 (one-sided Wilcoxon rank sum test; $p = 6.5983\text{e-}22$). The difference of the correlation suggests that the magnitudes of variation in sequencing is inherited from a tutor to some extent. In addition, the histogram of BF3 and BF4 have multiple modes.

Subsequently, we then conducted hierarchical clustering of all state transition matrices by using correlation coefficients of them (Fig. 3.7). As a result, the bouts from BF1 and BF2 can be perfectly separated from those from BF3 and BF4. In other words, correlations calculated between BF1 and BF2 in Fig. 3.8 are significantly higher than those calculated between a learner of tutor A and a learner of tutor B (one-sided Wilcoxon rank sum test; p = 1.7401e-24). In addition, the bouts from different individual subjects can be almost clearly discriminated from each other. These results insist that different sequencing rule of states is employed by each subject, although subjects share some states.

Some similar transition matrices seem to form subclusters within individuals,

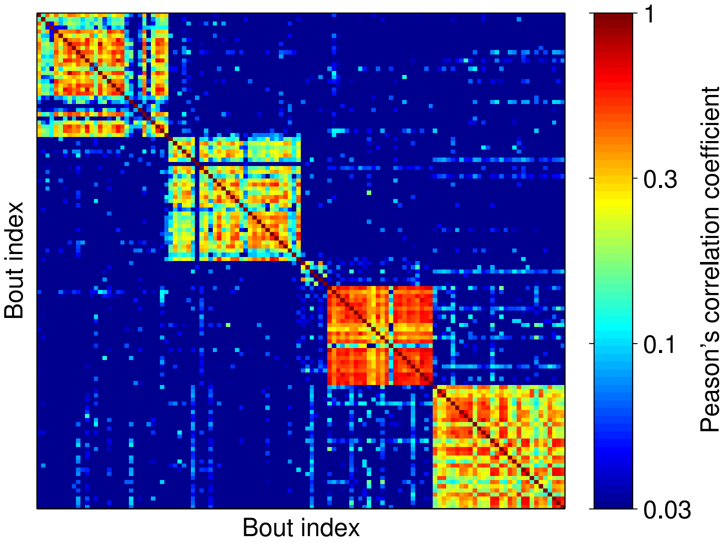


Figure 3.5. Correlation matrix calculated between bouts. Bout indices are arranged to line up from bouts of BF1 to those of BF4, downward or left to right. Each color represents the magnitude of correlation coefficient. Negative components of smaller magnitude than 0.1 are taken absolute values for visualization.

that is, the bouts in SC1 and SC2 within BF4 (Fig. 3.9). The entropy of transition probabilities from syllable a, which were calculated from human annotations, were significantly larger for SC2 than for SC1 (one-sided Wilcoxon rank sum test; $p = 0.0473 < 0.05$, effect size (Pearson's correlation) $r = 0.3284$). This indicates that the hierarchical clustering analysis based on the BP-HMM disclosed distinct syntactic rules of songs within an identical birds.

3.4 Discussion

Estimated syllable sequences were almost perfect agreement (0.81–1.00) with manually annotated labels for three subjects out of four. This insist that the BP-HMM can estimate sequencings of multiple birds simultaneously, even if the sequencings are not homogeneous.

Songs of different subjects could be discriminated and the bouts of a subject had subclusters in terms of transition probabilities. Such variability in sequencing of an identical bird can also arise from social context [47], LMAN lesions [55], and differential reinforcement [48]. The clustering based on the BP-HMM can promote the study of such variability in sequencing. The quantitative analysis scheme for birdsong also enables us to evaluate syntactic rules of bouts in heterogeneous

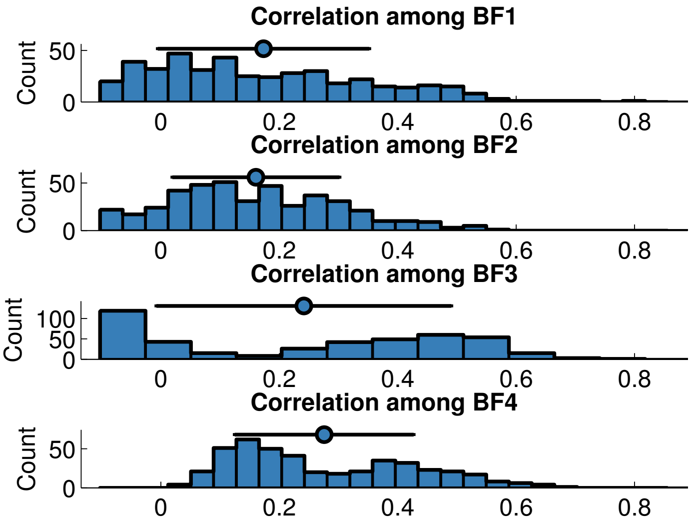


Figure 3.6. Comparison of distributions of correlation coefficients calculated from transition matrices within individual subjects. Each blue circle and error bar represent the mean and ±1 standard deviations of correlations.

conditions, e.g., different development stages of an identical bird.

Magnitudes of the similarity among the bouts of identical subjects varied according to subjects' tutors. This suggests that the difference of tutors modulates the magnitudes of variations in syllable sequencing.

3.5 Conclusion

We modeled songs of four Bengalese finches by using a beta process hidden Markov model that allows each time series data to have its own transition matrix. Our result of automatic annotation was within the range of almost perfect agreement (0.81–1.00) for three subjects out of four. As a result of analysis on transition matrices, it was shown that the BP-HMM can capture the larger similarity of transition matrices within individuals than those among the subjects. In addition, the BP-HMM discovered subclusters within the songs of an individual subject. These results are promising, especially considering to deal with the diversity of bird songs.

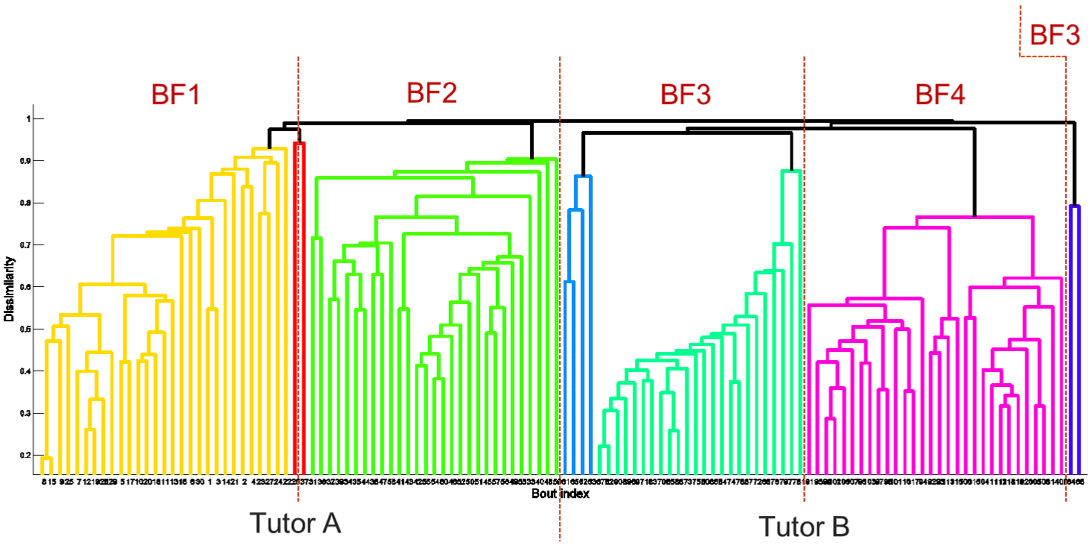


Figure 3.7. Result of hierarchical clustering based on the correlation coefficient.

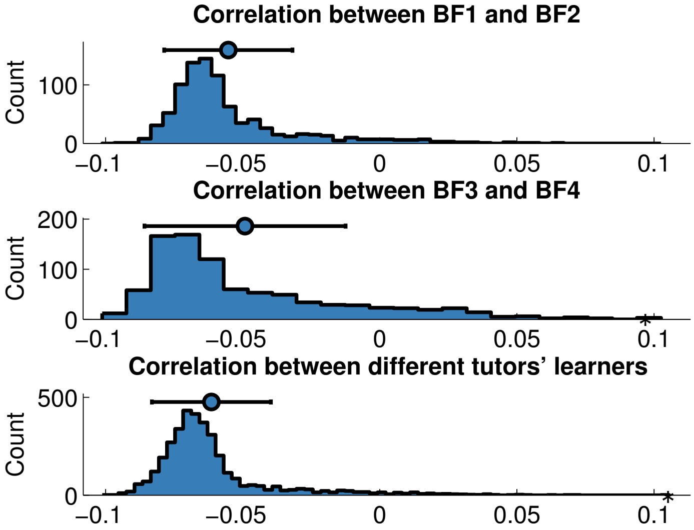


Figure 3.8. Comparison of distributions of correlation coefficients calculated from transition matrices between different subjects. Upper figure is a histogram of correlation coefficients calculated between BF1 and BF2, middle between BF3 and BF4, and lower between subjects that have different tutors. Each blue circle and error bar represent the mean and ±1 standard deviations of correlations.

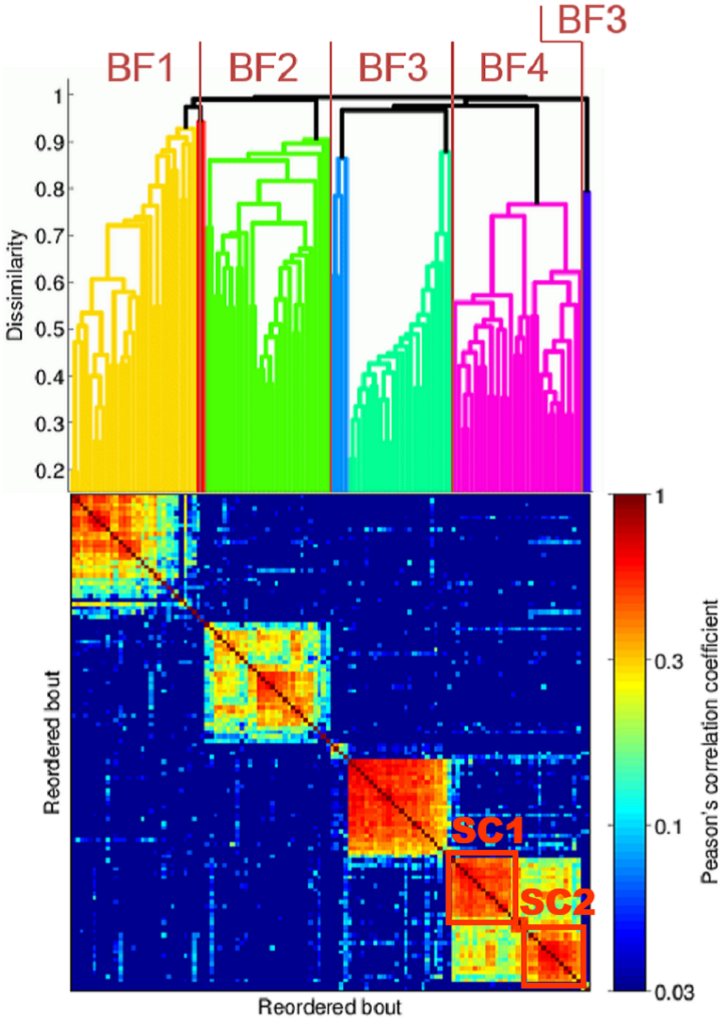


Figure 3.9. Reordered correlation matrix. The indices of correlation matrix in Fig. 3.5 was reordered according to the result of hierarchical clustering.

Chapter 4

Analysis and prediction of driving behavior

4.1 Introduction

Recent advanced driver assistance systems (ADASs) such as automatic braking systems [56, 57], adaptive cruise control or lane-keeping systems [58–60] and pedestrian protection [61–63] have reduced the number of traffic accidents [64]. These systems detect a dangerous condition and warn the driver of the condition. This means that the driver falls into the dangerous condition once, which should be avoided in advance.

To prevent a car from a dangerous condition, the future movement of the car must be estimated, which results in the prediction of driving behaviors since the car is operated by a driver (Fig. 4.1). Some systems have successfully predicted

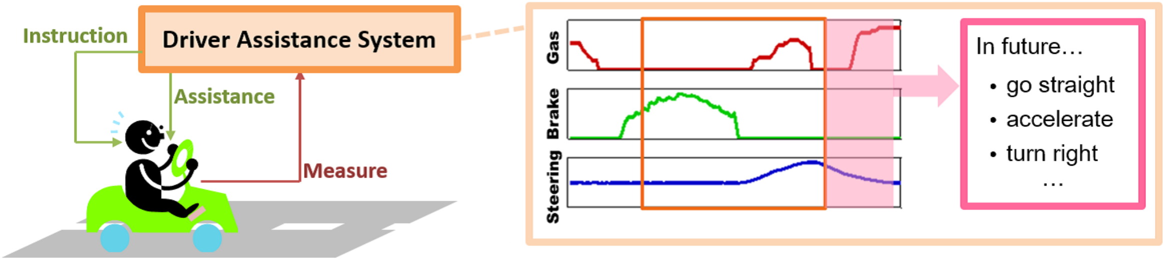


Figure 4.1. Typical application of driver assistance system.

specific behaviors, for example, braking behaviors [65], “approaching a traffic light” behaviors [66], lane departure behaviors [67], and behaviors at intersections [68,69]. Although they successfully predict specific behaviors in a short time scale, few systems have achieved to predict general behaviors in a longer time scale.

To treat general behaviors, a prediction system must divide a sequence of behaviors into segments and extract sequences of segments. Taniguchi et al. [70] formulated this problem as a two-level structure and solved it based on an analogy with language (letters/words) by a nonparametric Bayesian method called the nested Pitman-Yor language model (NPYLM) [71]. They modeled driving behaviors by the NPYLM and predicted sequences of segments that corresponded to more than eight seconds. However, they predicted only symbolized behaviors, not driving behaviors themselves, that are necessary to predict the future movement of the car.

Inspired by the success of the NPYLM in predicting driving behaviors, in this chapter, we proposed to apply a nonparametric Bayesian method with dynamical systems to predicting driving behaviors, not the sequence of symbols. Driving behaviors are well modeled by a set of linear dynamical systems called a hybrid dynamical system (HDS), wherein the dynamics switches from one to another [72,73]. The HDSs have some variants depending on the switching method and the dynamical systems therein such as the Markov dynamic model [73], the switching linear dynamical systems [74] and the autoregressive hidden Markov model (AR-HMM) [8]. In this study, the AR-HMM was employed because it is a simplest model to express dynamics although more complicated models were employed in the literature [75, 76]. The AR-HMM must determine the number of AR models (the number of kinds of behaviors in driving) in advance. To avoid this difficulty, we incorporated the nonparametric Bayesian technique into the AR-HMM, which was proposed as the beta process autoregressive hidden Markov model (BP-AR-HMM) [11].

The BP-AR-HMM divides a sequence of behaviors into segments (called driving letters in [70]) in an unsupervised way and assigns an AR dynamical system to each segment. When given a sequence, the BP-AR-HMM is trained by alternatively carrying out the segmentation of the sequence according to the estimated dynamical systems and the identification of the dynamical system in each segment. After trained, the BP-AR-HMM predicts a sequence of segments (driving letters) using the state transition probability of the HMM and then predicts the behaviors using the AR model in each segment.

In the following, we show the soundness of the BP-AR-HMM for prediction of driving behaviors using the real driving data.

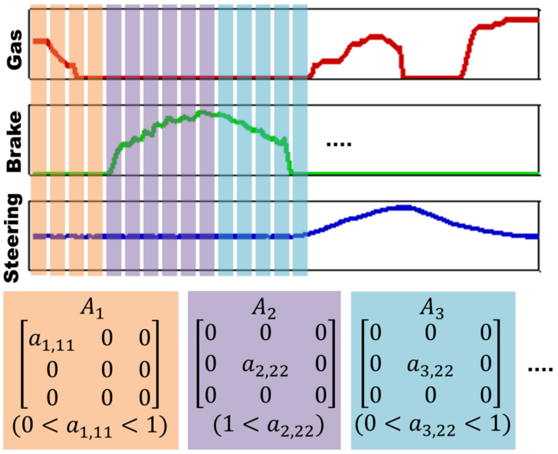


Figure 4.2. A typical example of BP-AR-HMM applied to driving behaviors.

4.2 Materials and Methods

4.2.1 Generative model

We used the BP-AR-HMM to model driving behaviors [11]. The behavior vector $\boldsymbol{y}_t^{(i)}$ at time t that consists of the accelerator opening rate, the brake pressure and the steering angle is expressed as

$$\boldsymbol{y}_t^{(i)} = \mathbf{A}_{z_t}\boldsymbol{y}_{t-1}^{(i)} + \boldsymbol{\varepsilon}_t, \tag{4.1}$$

where z_t and $\boldsymbol{\varepsilon}_t$ are the hidden state and the Gaussian noise at time t, respectively.

The BP-AR-HMM assigns a hidden state (VAR model) in the same way as the AR-HMM. Differently from the AR-HMM, the BP-AR-HMM makes a new state not assigned so far in a certain probability according to a beta process. Thus, the beta process generates a prior probability of emergence ω_k of state k [11]. The AR coefficients of the AR process are also chosen according to the base measure B_0. This model can produce an infinite number of states in principle and determines the total number of states according to the intrinsic complexity of given data (Fig. 4.2).

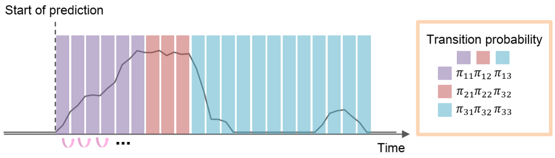
4.2.2 Training and prediction

When sequences of behaviors were given, the unknown variables are the state (AR model) at each time step and the AR coefficients of the AR model. To estimate them from the given sequence, the BP-AR-HMM assigns a state to each time step using the current criterion for assignment and updates the estimates of the AR coefficients in the states and the state transition probability using the assigned data, and iterates this procedure until the estimates satisfy a certain condition [11].

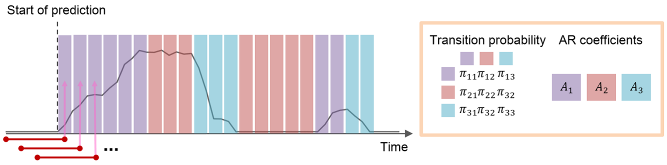
The BP-AR-HMM code in [77] was used in our experiments, where the parameters were estimated using a Markov chain Monte Carlo (MCMC) sampling with a sum-product algorithm [78] and reversible jump MCMC [79]. Their hyper-parameters, γ and κ, were assumed to have gamma-distribution priors and the other parameters were set as follows: The order of the VAR processes, one; the prior distribution of observation noises, Gaussian with mean zero where the variance was the covariance of the observed data multiplied by 0.75. In the experiments for Dataset 2 (See the next subsection), the parameters were estimated using the Viterbi algorithm [6] according to the past driving behaviors to reduce the computational complexity.

The BP-AR-HMM can predict how the states and the behaviors change in the future. First, the trained BP-AR-HMM made a sequence of states according to the estimated state transition probability. Since each state expressed a VAR process, behaviors in the future were predicted using the VAR process of the state at each time point.

For the prediction of the state sequences, we took two different methods. The first method chose the most probable state as the predicted state (Fig. 4.3(a)). This method is easy to implement with less computational complexity. The second method predicted the states successively using the Viterbi algorithm with the state transition probability, the AR coefficients, and a batch of past behaviors (Fig. 4.3(b)). This prediction method corresponds to the Bayesian inference but requires high computational complexity.



(a) Naive prediction



(b) Prediction with Viterbi

Figure 4.3. Overview of state sequence prediction.

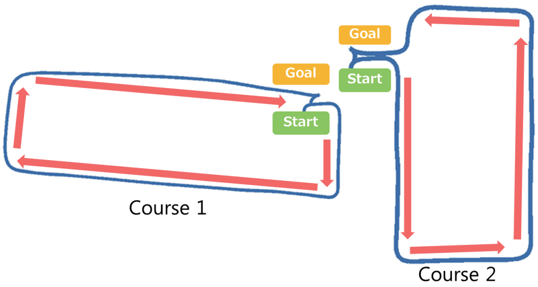


Figure 4.4. Courses 1 and 2 of a short track experiment. Subject was instructed to drive car clockwise along course 1, and counterclockwise along course 2.

4.2.3 Driving behavior datasets

We used two datasets[1]. Dataset 1 was the same dataset as Driving Data in Factory Course in [70], where one participant drove our experimental car along two courses in a factory five laps for each (Fig. 4.4). Dataset 2 was the same dataset as Driving Data on Public Road also in [70], where one participant drove our experimental car along a course in a public road in Japan for nine roundtrips (Fig. 15 in [70]). The eighteen runs took 42.9 minutes on average with standard deviation 6.6 minutes. During the experiments, the accelerator opening rate,

---

[1] The two datasets were measured and provided by DENSO CORPORATION in 1-1 Showa-cho, Kariya-shi, Aichi 448-8661, Japan.

the brake pressure, the steering angle, and the speed of the car were measured through control area network (CAN) at a sampling rate of 10 Hz.

4.2.4 Evaluation

We compared the prediction performance of our model, the BP-AR-HMM, with the simple HMM, the sticky hierarchical Dirichlet process HMMs (HDP-HMMs), the AR-HMM and the HDP-AR-HMM to see the effectiveness of introducing AR models and beta processes. The evaluation was done in two ways using the five-fold (Dataset 1) or nine-fold (Dataset 2) cross validation method. The one is the sequence of states as in [70] and the other is the accuracy of the behaviors, that is, the accelerator opening rate, the brake pressure and the steering angle. The former calculated the ratio of the correspondence between the prediction in the cross validation and the true state determined by using all the data. The latter calculated the mean absolute error (MAE) between the measured and predicted driving behaviors.

4.3 Results

4.3.1 State sequence prediction

4.3.1.1 Dataset 1

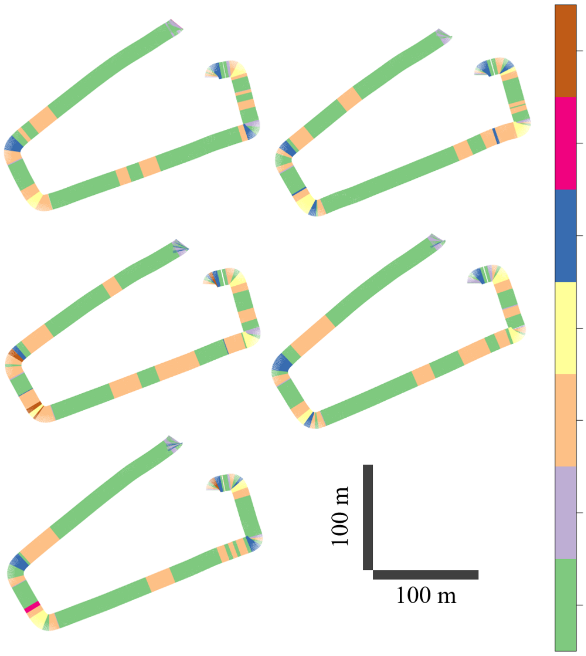
Given the whole data with Courses 1 and 2 in Dataset 1, the BP-AR-HMM produced seven states (AR models) and assigned one of them to each time point (Fig. 4.5). Although several disturbances occurred such as a pedestrian crossing a road and another vehicle in front of the experimental car, the assignment of the states was consistent and 76.9% of the states at the same position were coincident in three or more laps among five (Fig. 4.6). Moreover, the same state sequence frequently appeared in the same situations. For example, when the driver turned left, the sequence of states, 3-7-1, appeared in Course 2, which was analyzed hereafter.

To evaluate the prediction ability of our model, the BP-AR-HMM predicted the state sequence of a lap in the cross validation procedure. The number of states were determined to be eight on median across folds of the cross validation. The model trained with the rest correctly predicted the state sequence of a lap for 23.4 time points (2.34 seconds) on average (Fig. 4.7).

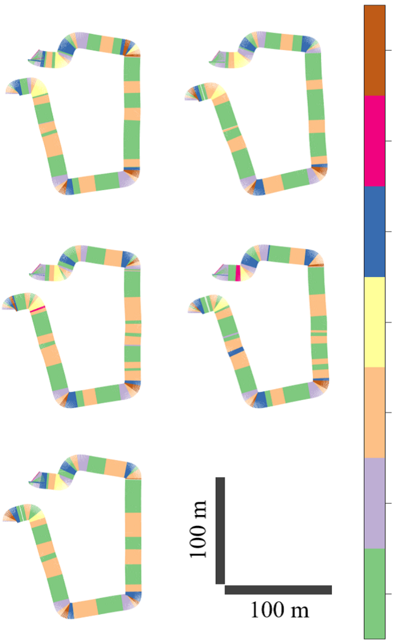
4.3.1.2 Dataset 2

Given the whole data (nine go-runs and nine return-runs) in Dataset 2, the BP-AR-HMM produced eight states (AR models) and assigned one of them to each time point (Fig. 4.8). Here, we concentrated the driving behaviors on the left-turns at an intersection because the driving behaviors in other situations widely diverged. For example, the driver changed the lane in some runs and did not in others.

For Dataset 2, the state sequences are not consistent as the case for Dataset 1. However, they seem to be classified into four classes (Fig. 4.8, leftmost) and the class depends on the vehicle's speed of the car. Note that the state sequences were estimated using the Viterbi algorithm because the estimation of driving states is not so easy task compared with a short track case.



(a) Course 1



(b) Course 2

Figure 4.5. Determined state sequences with our model for Dataset 1. The five laps are shown.

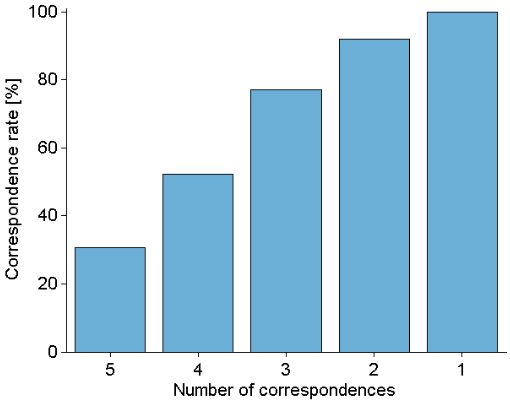


Figure 4.6. Correspondence rate of the assigned states in five laps.

4.3.2 Behavior prediction

4.3.2.1 Dataset 1

Using the predicted states and the corresponding AR models, the BP-AR-HMM predicted the behaviors of the driver (the brake pressure and the steering angle) during left-turn corners in Course 2 in Dataset (Fig. 4.5), where the initial state was set to State 3 since the sequence of states, 3-7-1, frequently appeared in left-turn corners.

The BP-AR-HMM had a smaller mean absolute error (MAE) in the brake pressure than the other models but did not have a significantly smaller MAE in the steering angle (Fig. 4.9), where the accelerator opening rates were omitted because they almost always took the value of 0% during the corners. Here, the numbers of states of HMM and AR-HMM were selected so that their MAEs took minimum among the different numbers of states. This means that nonparametric Bayesian methods were comparable in performance without model selection. This is because the BP-AR-HMM predicted the sudden decrease in brake pressures in four laps in the five although it did not predict the gradual increase in steering angles except for one lap (Fig. 4.10). Note that the HMM without AR models could not predict any of the above.

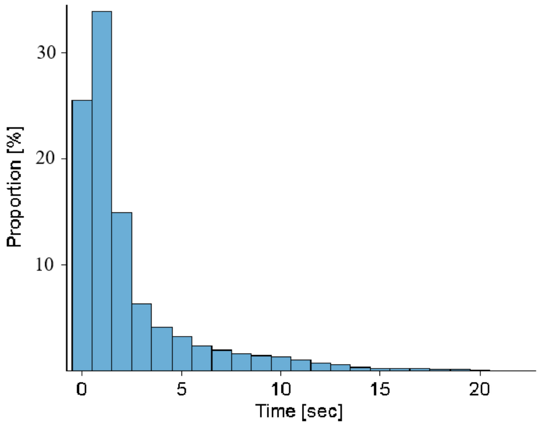


Figure 4.7. Histogram of lengths the sequence prediction methods could predict.

4.3.2.2 Dataset 2

Since the driving behaviors were strongly affected by the vehicle's speed, we included the vehicle's speed to the state variables, that is, each state consisted of the accelerator opening rate, the brake pressure, the steering angle and the vehicle's speed. The number of states were determined by BP-AR-HMM to be ten on median across folds of the cross validation, which was not so larger compared to Dataset 1. The HMM with AR models, i.e., AR-HMM, HDP-AR-HMM and BP- AR-HMM, had smaller MAE in predicting the vehicle's speed but they didn't have significant difference in the driving behaviors (Fig. 4.11). The numbers of states of HMM and AR-HMM were selected as were in Dataset 1.

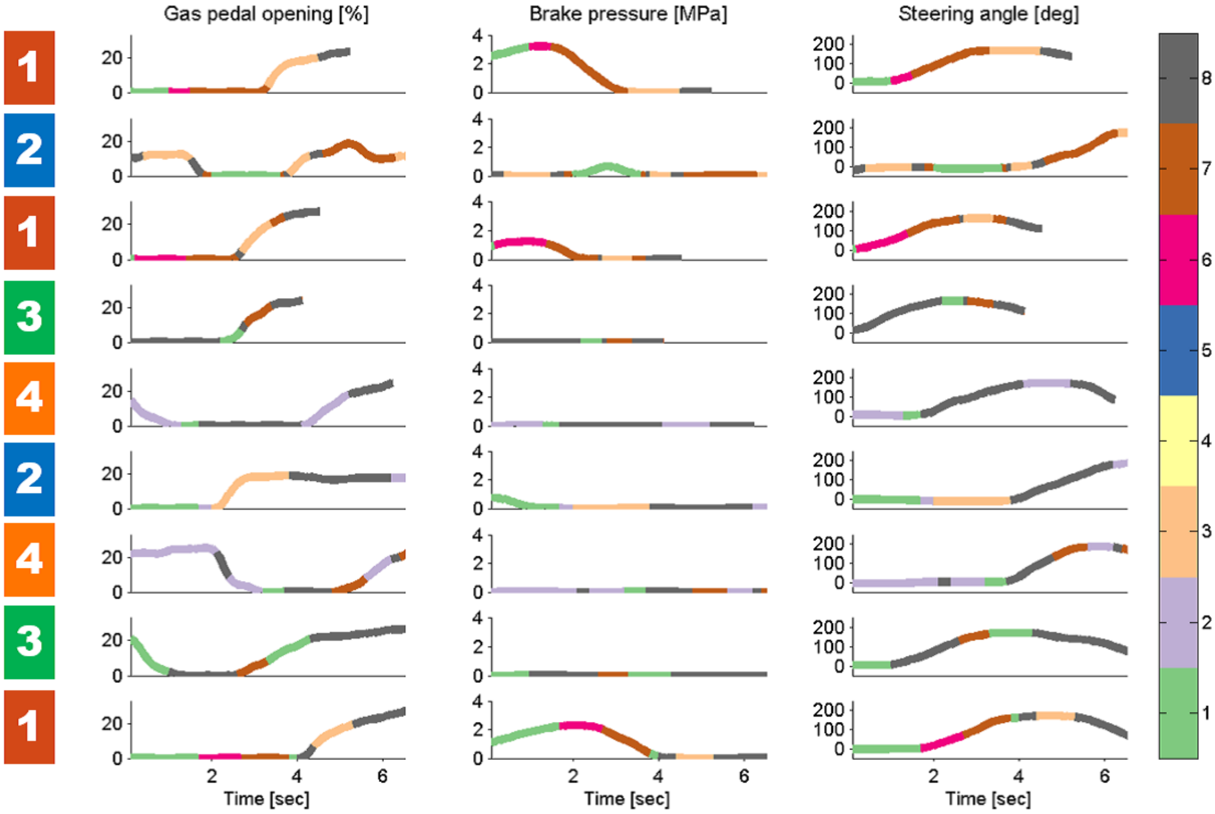


Figure 4.8. Determined state sequences with our model. The nine return-runs are shown.

4.4 Discussion

In our experiments using real driving data, the BP-AR-HMM successfully predicted not only states but also driving behaviors themselves for Dataset 1.

The duration time successfully predicted was shorter than the double articulation analyzer [70] (Fig. 4.7). This is because the BP-AR-HMM did not treat the states as sequences explicitly as the language model does [70, 71]. Introduction of such hierarchical dependency in state sequences will enables us to predict more complex driving behaviors. Nonetheless, it could predict the sequences of states by virtue of the dynamics (AR model) in each state and this implies the soundness of introducing AR models to HMM models.

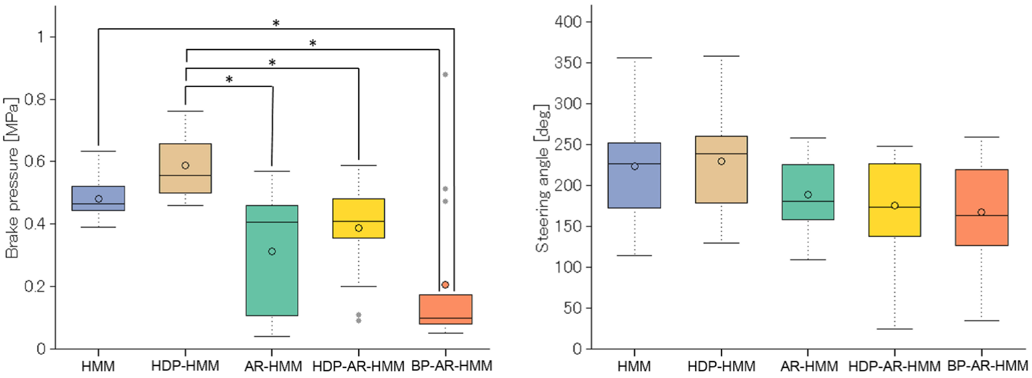
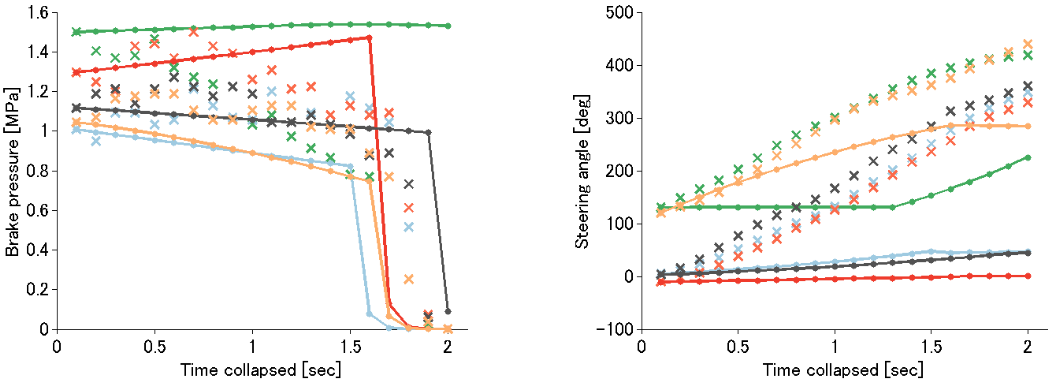


Figure 4.9. MAEs of the models for brake pressures (left) and steering angles (right). Wilcoxon rank-sum test, Bonferroni corrected, $p \leq 0.05$.

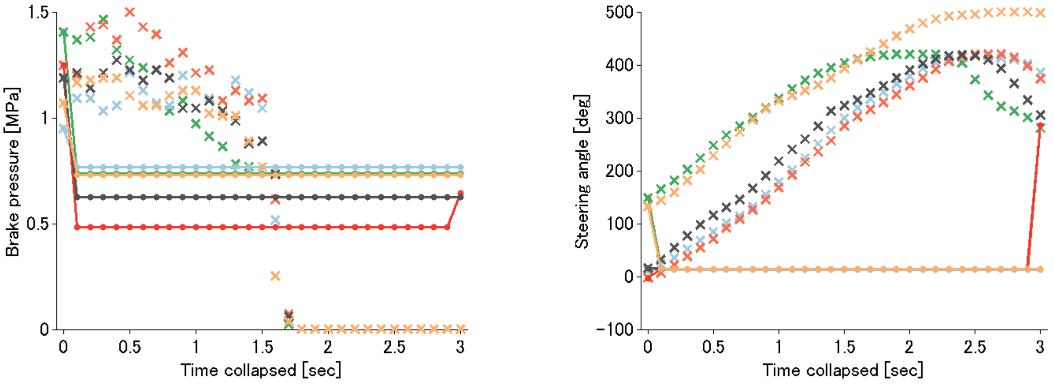
The prediction accuracies of the modeling methods for driving behaviors were compared in terms of MAE and the BP-AR-HMM outperformed the other models that do not have dynamics such as the HMM and the HDP-HMM (Fig. 4.9). In addition, the BP-AR-HMM showed a little smaller MAE than the AR-HMM and the HDP-AR-HMM that include AR models. The HDP-AR-HMM is a nonparametric Bayesian method of the AR-HMM as the BP-AR-HMM. One property of the HDP-AR-HMM is to share a state transition probability among sequences [27] although the BP-AR HMM assigns a different one to each state. This may be the reason why the BP-AR-HMM would work better since the driving behaviors were not homogeneous but heterogeneous due to the variety of road conditions.

For Dataset 2, however, the models had comparable MAE (Fig. 4.11). This may be because the driving behaviors at intersections are not the same (Fig. 4.8). Although we succeeded to predict the vehicle's speed by adding the speed to the measurement, we need to study more to improve the behavior prediction.



(a) Brake pressures, BP-AR-HMM.

(b) Steering angle, BP-AR-HMM.



(c) Brake pressures, HMM.

(d) Steering angle, HMM.

Figure 4.10. Predicted and actual behaviors by the BP-AR-HMM and the HMM. x's, measured; lines, predicted; the same color shows the same run.

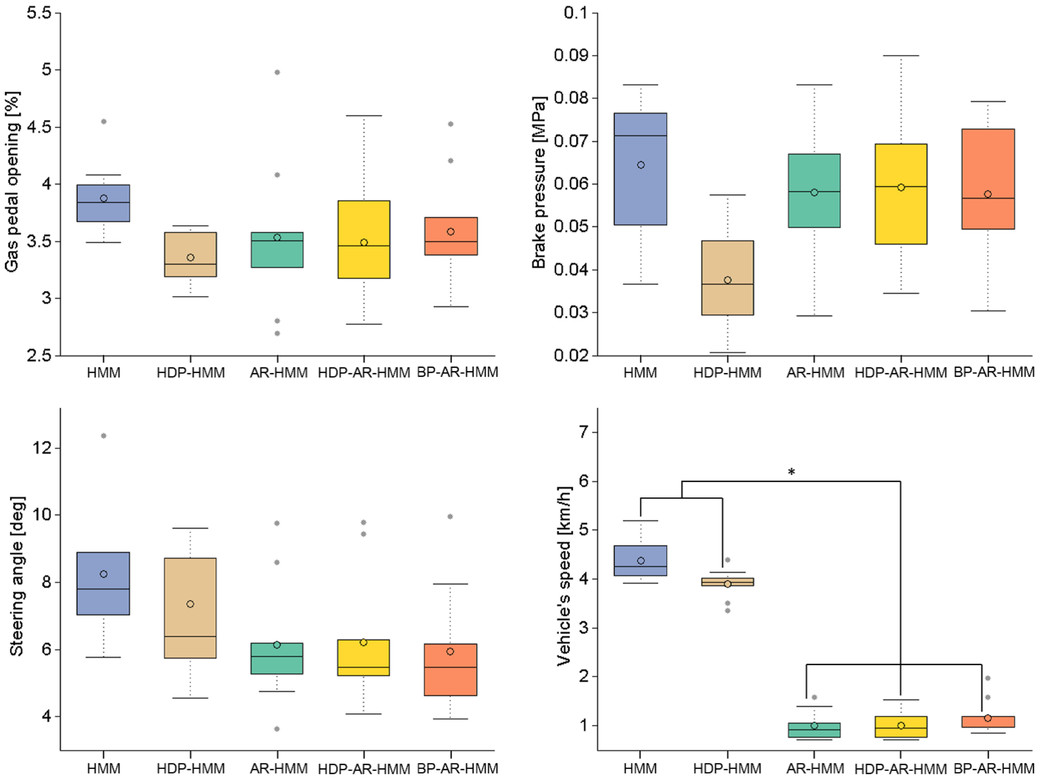


Figure 4.11. MAEs of the models for the accelerator opening rate (upper right), the brake pressure (upper right), the steering angles (lower left) and the vehicle's speed (lower right). Wilcoxon rank-sum test, Bonferroni corrected, $p \leq 0.05$.

4.5 Conclusion

We showed the BP-AR-HMM successfully predicted the driving behaviors. The BP-AR-HMM automatically segmented the past driving behaviors into discrete states each of which corresponded to an autoregressive dynamical model and predicted the state sequences as well as the driving behaviors better in our experiments. Although the BP-AR-HMM fails to predict in some cases, the AR models are found to be effective to predict the driving behaviors that might be useful for a new type of advanced driver assistance systems that predict dangerous conditions in advance.

Chapter 5

Conclusion

5.1 Summary

This dissertation verify the effectiveness of non-stationary time series modeling methods with Bayesian nonparametrics for the analysis of behavior data. First, we developed the automatic annotation and clustering method for birdsong data by using the Bayesian nonparametric HMM. Our method could archive consistent annotation with human annotators, and identify the subjects of bird song data without any information of the subject ID or tutor-learner relationships of the subjects. Second, we developed the prediction method of driving behaviors by using Bayesian nonparametric AR-HMM, and showed that our method predicted the driving behaviors better than the other methods without any model selection procedure. These results insist that the non-stationary time series modeling methods with Bayesian nonparametrics are promising to analyze the behavior time series dataset.

5.2 Discussions and Future work

The automatic annotation and clustering of birdsong data can encourage the research of evolution or developmental process of birdsong. The comparison of songs of the Bengalese finch and its ancestral species is important to reveal how the Bengalese finches evolutionary acquire their songs, but difficult to conduct with conventional HMMs because the single HMM has a single state transition probability matrix. Our modeling method of birdsong makes it possible to compare different transition matrices of different subjects. Such comparison study will give insight into the evolutionary acquisition of complex grammar. On the other hand the comparison of songs of an identical subject but in different developmental stages is important for the study of song learning. Our modeling method of birdsong enable us to investigate whether there are some differences between two bouts, and it is possible to see how the song and its syntactic rule change as the subject grows and learns his song. The developmental process of birdsong is reported to have similarity to that of human speech [80], so analyzing the developmental process of birdsong could be useful also for understanding language acquisition of human.

The prediction of driving behaviors has an impact on the development of automatic driving and driver assistance systems. The automatic driving systems have been developed and already realized for some driving environments such as in highways by many researchers and automobile manufacturers. Although the success in rather simple situations, the automatic driving in urban situations is still difficult because a lot of cars and vulnerable road users such as pedestrians and cyclists interact and their behaviors are not easily predicted with conventional methods. To develop more practical automatic driving systems in complex driving situations, more driving behavior data should be collected and analyzed. Our modeling method of driving behaviors enables us to segment time series data to obtain collections of complex driving situations, in which the state sequences typically have more transitions than in simple situations, then it results in collecting amount of data in complex driving situations that rarely occur such as accidents. The driver assistance systems will also benefit from our modeling method, because if the driving behaviors can be predicted the driver assistance systems can instruct a driver to avoid dangerous situations, or compensate for the control of a car to keep the car more stable. Modeling and analyzing professional drivers' behaviors will give us some useful knowledge to educate novice drivers.

Some model selection methodology can be used as alternatives to Bayesian nonparametrics. As pointed out above, information criteria like AIC and BIC cannot be applicable to singular models such as the HMM and AR-HMM. On the other hand, information criteria that are adopted to the singular models, e.g., WAIC, WBIC [81], FIC [82], can be applied. These methods need to compare the quantity calculated from different number of states but are still interesting approaches in theory and practice.

For our future work, it is necessary to modify and extend the non-stationary time series modeling method with Bayesian nonparametrics to harness prior knowledge of datasets. If we know some information of time series data, for example in our cases the course ID or subject ID, these information can be exploited to associate some time series with each other. Hierarchical beta processes [39] can be used to tie up multiple time series and introduce the grouping of them. Such semi-supervised extension of the modeling method make an explicit assumption of the hierarchical structure of time series dataset, and facilitate sharing or non-sharing of states among multiple time series. Modification of syllables in birdsong should be considered in the hierarchy for the birdsong dataset. Songbirds learn their songs and inherit syllables by their tutors, but modify their acoustic structures to some extent [83]. Explicitly modeling of the modification of syllables permits the errors by inheritance and can enhance the sharing of syllables between a tutor and his learners.

Acknowledgements

本研究の3章で用いたジュウシマツの歌声データは，理化学研究所 脳科学総合研究センターの岡ノ谷一夫教授のグループにより計測・ご提供いただきました．また4章で用いた自動車運転行動データは，株式会社デンソー様により計測およびご提供いただきました．これらのデータをご提供していただくことがなければ，本研究を行うことが不可能でした．データをご提供いただきましたこと，また私の解析や結果の解釈，研究の目的・意義について有益なご意見をいただきましたことを心より感謝いたします．

3年間の博士後期課程を通しご指導いただきました主指導教官の池田和司教授に深く感謝いたします．池田先生は私を在学中に幾度もの国際会議に出席させてくださり，また3年次の後半という時期にカリフォルニア大学デービス校への留学をお許しくださいました．博士前期課程の修了間際に私を研究室に受け入れてくださった時より，決して言い尽くせないほどの恩を賜りました．思うように結果が出せずに研究の方向性を見失いそうになったときもありましたが，池田先生のもとであればこそ博士後期課程の3年間を送ることができたことと存じます．重ねて心より感謝いたします．

研究全般にあたり多大なるご指導をいただきました久保孝富特任准教授に感謝いたします．久保先生は常日頃より研究内容に関する丁寧かつ適切なご助言のみならず，研究者としての生き方・心得について大きな示唆をいただきました．統計・機械学習・応用分野への深い知見から，私が思いつけないような視点を常にご教示くださったこと，また思うように結果を出せず苦しむ私に研究者として為すべきことを根気良く説いてくださったこと，これらが私の中の理想の研究者像を形成しました．心より感謝いたします．

お忙しいなか博士論文審査委員の任を引き受けてくださいました本学の松本裕治教授，吉本潤一郎准教授，京都大学の平岡敏洋助教に心より感謝いたします．

I would like to thank Dr. Alexander Aue and his students, especially Rex Cheung, for accepting me as a visiting student in University California, Davis. Studying and discussing our topic are so valuable and give me profound insight of statistics and a scent of rigor.

行動解析への広範な知識から自動車運転の研究について有益なご助言をいただきました九州工業大学の柴田智広教授に感謝を申し上げます．研究室内セミナーでの発表の際をはじめとして，常日頃より有益なコメントをくださりました吉本潤一郎准教授，爲井智也助教，丸野由希さんをはじめとする数理情報学研究室の

構成員のみなさま，および卒業されたOB・OGのみなさまに感謝いたします．研究室秘書の谷本史さんには，出張・留学をはじめ数々の手続きについてお世話になりました．数少ない同期の小出真子さん，Jimson Ngeoくんは研究・学生生活ともに良き話し相手となっていただきました．Matthew James Hollandくんは驚くほど深い数学・統計への知識，また刃のようにするどい直感力から数々の有益なコメントをくださいました．秋澤翔くん，下村環太朗くん，岩根史明くん，町田宗丈くん，田村真一くん，武田友希さんをはじめとする多くの後輩に深く感謝いたします．

最後に，厳しい研究生活で擦り切れた心身を支えてくれた妻恵里に感謝します．先の見えない博士後期課程のなか，見捨てずについてきてくれて本当にありがとう．平坦な道では決してないですが，一緒に頑張って歩いていきましょう．

本研究を通して，また本論文を執筆するにあたり，多くの方々のご協力をいただきました．決して私の力だけでは研究を為すことができませんでした．研究の遂行の際にお世話になった全ての方々に，ここに深く感謝の意を表します．本当にありがとうございました．

Bibliography

[1] R. S. Tsay, *Analysis of financial time series.* Hoboken, NJ, USA: John Wiley & Sons, 2005.

[2] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals.* Englewood Cliffs, NJ, USA: Prentice Hall, 1978.

[3] J. D. Hamilton, *Time series analysis.* Princeton, NJ, USA: Princeton university press, 1994.

[4] G. E. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal of the American statistical Association*, vol. 65, no. 332, pp. 1509–1526, 1970.

[5] A. Aue and L. Horváth, "Structural breaks in time series," *Journal of Time Series Analysis*, vol. 34, no. 1, pp. 1–16, Jan. 2013.

[6] L. R. Rabiner and B. H. Juang, "An introduction to hidden markov models," *IEEE ASSP Mag.*, vol. 3, no. 1, pp. 4–16, Jan. 1986.

[7] H.-M. Krolzig, *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis.* Berlin, Germany: Springer-Verlag, 1997.

[8] Y. Kishimoto and K. Oguri, "A modeling method for predicting driving behavior concerning with driver's past movements," in *Proc. IEEE Int. Conf. Veh. Electron. Safety (ICVES)*, Columbus, OH, USA, 2008, pp. 132–136.

[9] N. Hjort, C. Holmes, P. Müller, and S. Walker, *Bayesian nonparametrics.* Cambridge, UK: Cambridge University Press, 2010.

[10] M. C. Hughes, E. B. Fox, and E. B. Sudderth, "Effective split-merge Monte Carlo methods for nonparametric models of sequential data," in *Advances in Neural Information Processing Systems 25 (NIPS)*, 2012, pp. 1304–1312.

[11] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "Sharing features among dynamical systems with beta processes," in *Advances in Neural*

Information Processing Systems 22 (NIPS). Cambridge, MA, USA: MIT Press, 2009, pp. 549–557.

[12] B. H. Juang and L. R. Rabiner, "Hidden Markov models for speech recognition," *Technometrics*, vol. 33, no. 3, pp. 251–272, Mar. 1991.

[13] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, and T. Kitamura, "Speech parameter generation algorithms for HMM-based speech synthesis," in *Proc. 25th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, 2000, pp. 1315–1318.

[14] A. Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer, "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes," *Journal of molecular biology*, vol. 305, no. 3, pp. 567–580, Jan. 2001.

[15] M. Elmezain, A. Al-Hamadi, J. Appenrodt, and B. Michaelis, "A hidden Markov model-based continuous gesture recognition system for hand motion trajectory," in *Proc. 19th International Conference on Pattern Recognition (ICPR)*, Tampa, FL, USA, 2008, pp. 1–4.

[16] S. M. Ross, *Introduction to Probability Models.* Burlington, MA, USA: Academic press, 2014.

[17] C. M. Bishop, *Pattern recognition and machine learning.* New York, NY, USA: Springer, 2009.

[18] K. P. Murphy, *Machine learning: a probabilistic perspective.* Cambridge, MA, USA: MIT Press, 2012.

[19] A. P. Dempster, N. M. Laird, and D. B. Lubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[20] R. Douc, E. Moulines, and D. Stoffer, *Nonlinear time series: Theory, Methods and Applications with R Examples.* Boca Raton, FL, USA: CRC Press, 2014.

[21] G. Schwarz, “Estimating the dimension of a model,” *The annals of statistics*, vol. 6, no. 2, pp. 461–464, Mar. 1978.

[22] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proc. 14th International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 14, no. 2, 1995, pp. 1137–1145.

[23] S. Watanabe, “Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory,” *Journal of Machine Learning Research*, vol. 11, pp. 3571–3594, Dec. 2010.

[24] G. D. Forney, “The Viterbi algorithm,” *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.

[25] H. Krolzig, “Statistical analysis of cointegrated VAR processes with Markovian regime shifts,” Citeseer, Tech. Rep., 1996.

[26] Y. Ephraim, D. Malah, and B. Juang, “On the application of hidden Markov models for enhancing noisy speech,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1846–1856, Dec. 1989.

[27] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, “Bayesian nonparametric methods for learning markov switching processes,” *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 43–54, Nov. 2010.

[28] M. J. Beal, “Variational algorithms for approximate Bayesian inference,” *PhD thesis, Gatsby Unit, University College London*, 2004.

[29] K. Katahira, K. Watanabe, and M. Okada, “Deterministic annealing variant of variational Bayes method,” *Journal of Physics: Conference Series*, vol. 95, no. 1, 2008.

[30] S. Goldwater and T. L. Griffiths, “A fully Bayesian approach to unsupervised part-of-speech tagging,” *Annual meeting-association for computational linguistics*, vol. 45, no. 1, pp. 744–751, 2007.

[31] S. Guha, Y. Li, and D. Neuberg, “Bayesian hidden Markov modeling of array CGH data,” *Journal of the American Statistical Association*, vol. 103, pp. 485–497, 2008.

[32] T. S. Ferguson, "A bayesian analysis of some nonparametric problems," *The Annals of Statistics*, vol. 1, no. 2, pp. 209–230, Mar. 1973.

[33] C. E. Rasmussen, "The infinite Gaussian mixture model," in *Advances in Neural Information Processing Systems 12 (NIPS).* Cambridge, MA, USA: MIT Press, 2000, pp. 554–560.

[34] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning.* Cambridge, MA, USA: MIT Press, 2006.

[35] P. Orbanz and Y. W. Teh, *Bayesian nonparametric models in Encyclopedia of Machine Learning.* New York, NY, USA: Springer US, 2010.

[36] S. J. Gershman and D. M. Blei, "A tutorial on bayesian nonparametric models," *Journal of Mathematical Psychology*, vol. 56, no. 1, pp. 1–12, Feb. 2012.

[37] P. Orbanz, "Lecture notes on Bayesian nonparametrics," `http://stat.columbia.edu/~porbanz/talks/npb-tutorial.html`, 2014.

[38] N. L. Hjort, "Nonparametric bayes estimators based on beta processes in models for life history data," *The Annals of Statistics*, vol. 18, no. 3, pp. 1259–1294, Sep. 1990.

[39] R. Thibaux and M. I. Jordan, "Hierarchical beta processes and the Indian buffet process," in *Proc. 11th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2007, pp. 564–571.

[40] T. L. Griffiths and Z. Ghahramani, "Infinite latent feature models and the Indian buffet process," in *Advances in Neural Information Processing Systems 18 (NIPS).* Cambridge, MA, USA: MIT Press, 2005, pp. 475–482.

[41] K. Sato, *Lévy processes and infinitely divisible distributions.* Cambridge, UK: Cambridge university press, 1999.

[42] R. A. Johnson, "Atomic and nonatomic measures," *Proceedings of the American Mathematical Society*, vol. 254, no. 3, pp. 650–655, Jul. 1970.

[43] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, “A sticky HDP-HMM with application to speaker diarization,” *Ann. Appl. Stat.*, vol. 5-2A, pp. 1020–1056, 2011.

[44] E. B. Fox, M. C. Hughes, E. B. Sudderth, and M. I. Jordan, “Supplement to Joint modeling of multiple time series via the beta process with application to motion capture segmentation,” 2014.

[45] K. Okanoya and A. Yamaguchi, “Adult Bengalese finches (*lonchura striata* var. *domestica*) require real-time auditory feedback to produce normal song syntax,” *Journal of neurobiology*, vol. 33, no. 3, pp. 343–356, Dec. 1998.

[46] K. Katahira, K. Suzuki, K. Okanoya, and M. Okada, “Complex sequencing rules of birdsong can be explained by simple hidden Markov processes,” *PloS ONE*, vol. 6, no. 9, p. e24516, Sep. 2011.

[47] J. T. Sakata, C. M. Hampton, and M. S. Brainard, “Social modulation of sequence and syllable variability in adult birdsong,” *Journal of neurophysiology*, vol. 99, no. 4, pp. 1700–1711, Apr. 2008.

[48] T. L. Warren, J. D. Charlesworth, E. C. Tumer, and M. S. Brainard, “Variable sequencing is actively maintained in a well learned motor skill,” *The Journal of Neuroscience*, vol. 32, no. 44, pp. 15 414–15 425, Oct. 2012.

[49] O. Tchernichovski, F. Nottebohm, C. E. Ho, B. Pesaran, and P. P. Mitra, “A procedure for an automated measurement of song similarity,” *Animal Behaviour*, vol. 59, no. 6, pp. 1167–1176, Jun. 2000.

[50] J. L. Fleiss, “Measuring nominal scale agreement among many raters,” *Psychological bulletin*, vol. 76, no. 5, pp. 378–382, Nov. 1971.

[51] M. C. Hughes, “NPBayesHMM by michaelchughes,” `http://michaelchughes.github.io/NPBayesHMM/`, accessed on Oct. 21, 2013.

[52] D. B. Dahl, “Sequentially-allocated merge-split sampler for conjugate and nonconjugate Dirichlet process mixture models,” *Submitted to Journal of Computational and Graphical Statistics*, 2005.

[53] M. C. Hughes and E. B. Sudderth, “Nonparametric discovery of activity patterns from video collections,” in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Providence, RI, USA, 2012, pp. 25–32.

[54] S. C. Johnson, “Hierarchical clustering schemes,” *Psychometrika*, vol. 32, no. 3, pp. 241–254, Sep. 1967.

[55] C. M. Hampton, J. T. Sakata, and M. S. Brainard, “An avian basal ganglia-forebrain circuit contributes differentially to syllable versus sequence variability of adult Bengalese finch song,” *Journal of neurophysiology*, vol. 101, no. 6, pp. 3235–3245, Jun. 2009.

[56] A. Broggi, P. Cerri, S. Ghidoni, P. Grisleri, and H. Jung, “A new approach to urban pedestrian detection for automatic braking,” *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 4, pp. 594–605, Dec. 2009.

[57] C. Keller, T. Dang, H. Fritz, A. Joos, C. Rabe, and D. Gavrila, “Active pedestrian safety by automatic braking and evasive steering,” *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1292–1304, Dec. 2011.

[58] A. Vahidi and A. Eskandarian, “Research advances in intelligent collision avoidance and adaptive cruise control,” *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 3, pp. 143–153, Sep. 2003.

[59] P. A. Ioannou and C. C. Chien, “Autonomous intelligent cruise control,” *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 657–672, Nov. 1993.

[60] T. Pilutti and A. G. Ulsoy, “Identification of driver state for lane-keeping tasks,” *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 29, no. 5, pp. 486–502, Sep. 1999.

[61] D. M. Gavrila, “Pedestrian detection from a moving vehicle,” in *Proc. 6th European Conf. Computer Vision (ECCV)*, Dublin, Ireland, 2000, pp. 37–49.

[62] T. Gandhi and M. M. Trivedi, “Pedestrian protection systems: Issues, survey, and challenges,” *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 413–430, Jun. 2007.

[63] D. Gerónimo, A. M. López, A. D. Sappa, and T. Graf, “Survey of pedestrian detection for advanced driver assistance systems,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1239–1258, Jul. 2010.

[64] T. Vaa, M. Penttinen, and I. Spyropoulou, “Intelligent transport systems and effects on road traffic accidents: state of the art,” *IET Intell. Transp. Syst.*, vol. 1, no. 2, pp. 81–88, Jun. 2007.

[65] J. C. McCall and M. M. Trivedi, “Human behavior based predictive brake assistance,” in *Proc. IEEE Intell. Veh. Symp. (IV)*, Tokyo, Japan, 2006, pp. 8–12.

[66] M. G. Ortiz, J. Fritsch, F. Kummert, and A. Gepperth, “Behavior prediction at multiple time-scales in inner-city scenarios,” in *Proc. IEEE Intell. Veh. Symp. (IV)*, Baden-Baden, Germany, 2011, pp. 1068–1073.

[67] P. Angkititrakul, R. Terashima, and T. Wakita, “On the use of stochastic driver behavior model in lane departure warning,” *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 174–183, Mar. 2011.

[68] G. S. Aoude, V. R. Desaraju, L. H. Stephens, and J. P. How, “Driver behavior classification at intersections and validation on large naturalistic data set,” *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 724–736, Feb. 2012.

[69] M. Liebner, F. Klanner, M. Baumann, C. Ruhhammer, and C. Stiller, “Velocity-based driver intent inference at urban intersections in the presence of preceding vehicles,” *IEEE Intell. Transp. Syst. Mag.*, vol. 5, no. 2, pp. 10–21, Apr. 2013.

[70] T. Taniguchi, S. Nagasaka, K. Hitomi, N. P. Chandrasiri, T. Bando, and K. Takenaka, “Sequence prediction of driving behavior using double articulation analyzer,” *IEEE Trans. Syst., Man, Cybern., Syst.*, to appear.

[71] D. Mochihashi, T. Yamada, and N. Ueda, “Bayesian unsupervised word segmentation with nested Pitman-Yor language modeling,” in *Proc. Joint Conf. 47th Annu. Meeting ACL 4th Int. Joint Conf. Nat. Lang. (AFNLP)*, Singapore, 2009, pp. 100–108.

[72] J. A. Michon, “A critical view of driver behavior models: what do we know, what should we do?” in *Human Behavior and Traffic Safety.* New York, NY, USA: Springer US, pp. 485–524, 1986.

[73] A. Pentland and A. Liu, “Modeling and prediction of human behavior,” *Neural Comput.*, vol. 11, no. 1, pp. 229–242, Jan. 1999.

[74] T. Kumagai and M. Akamatsu, “Prediction of human driving behavior using dynamic bayesian networks,” *IEICE Trans. Inf. Syst.*, vol. 89, no. 2, pp. 857–860, Feb. 2006.

[75] S. Sekizawa, T. Suzuki, N. Tsuchida, T. Tsuda, and H. Fujinami, “Modeling and recognition of driving behavior based on stochastic switched ARX model,” *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 4, pp. 593–606, Dec. 2007.

[76] H. Okuda, N. Ikami, T. Suzuki, Y. Tazaki, and K. Takeda, “Modeling and analysis of driving behavior based on a probability-weighted ARX model,” *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 98–112, Dec. 2013.

[77] E. B. Fox, “Homepage of Emily B. Fox,” `http://stat.wharton.upenn.edu/~ebfox/software.html`, accessed on Aug. 30, 2012.

[78] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[79] P. J. Green, “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination,” *Biometrika*, vol. 85, no. 4, pp. 711–732, Dec. 1995.

[80] A. J. Doupe and P. K. Kuhl, “Birdsong and human speech: common themes and mechanisms,” *Annual review of neuroscience*, vol. 22, no. 1, pp. 209–230, Mar. 1999.

[81] S. Watanabe, “A widely applicable Bayesian information criterion,” *Journal of Machine Learning Research*, vol. 14, pp. 867–897, Mar. 2013.

[82] R. Fujimaki and K. Hayashi, “Factorized asymptotic Bayesian hidden markov models,” in *Proc. 29th International Conference on Machine Learning (ICML)*, Scotland, UK, 2012.

[83] O. Fehér, H. Wang, S. Saar, P. P. Mitra, and O. Tchernichovski, “*De novo* establishment of wild-type song culture in the zebra finch,” *Nature*, vol. 459, no. 7246, pp. 564–568, May 2009.

Publication List

Journal Papers

1. Ryunosuke Hamada, Takatomi Kubo, Kazushi Ikeda, Zujie Zhang, Tomohiro Shibata, Takashi Bando, Kentarou Hitomi, and Masumi Egawa. Modeling and Prediction of Driving Behaviors Using a Nonparametric Bayesian Method with AR Models. IEEE Transaction on Intelligent Vehicles. Accepted.

International Conferences and Workshops

1. Ryunosuke Hamada, Takatomi Kubo, Kazushi Ikeda, Zujie Zhang, Tomohiro Shibata, Takashi Bando, and Masumi Egawa. Towards prediction of driving behavior via basic pattern discovery with BP-AR-HMM. IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, Canada, May 2013.

2. Ryunosuke Hamada, Takatomi Kubo, Kazushi Ikeda, Zujie Zhang, Tomohiro Shibata, Takashi Bando, and Masumi Egawa. Applying Bayesian nonparametrics to non-homogeneous driving operation data towards prediction. 21st European Signal Processing Conference, Marrakech, Morocco, September 2013.

3. Ryunosuke Hamada, Takatomi Kubo, Kazushi Ikeda, Zujie Zhang, Takashi Bando, and Masumi Egawa. A comparative study of time series modeling for driving behavior towards prediction. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Kaohsiung, Taiwan, October 2013.

4. Ryunosuke Hamada, Takatomi Kubo, Kentaro Katahira, Kenta Suzuki, Kazuo Okanoya, and Kazushi Ikeda. Birdsong analysis using beta process hidden Markov model. IEEE International Workshop on Machine Learning for Signal Processing, Reims, France, September 2014.