# Semiotic Prediction of Driving Behavior using Unsupervised Double Articulation Analyzer

Tadahiro Taniguchi, Shogo Nagasaka, Kentarou Hitomi, Naiwala P. Chandrasiri, and Takashi Bando

Abstract—In this paper, we propose a novel semiotic prediction method for driving behavior based on double articulation structure. It has been reported that predicting driving behavior from its multivariate time series behavior data by using machine learning methods, e.g., hybrid dynamical system, hidden Markov model and Gaussian mixture model, is difficult because a driver's behavior is affected by various contextual information. To overcome this problem, we assume that contextual information has a double articulation structure and develop a novel semiotic prediction method by extending nonparametric Bayesian unsupervised morphological analyzer. Effectiveness of our prediction method was evaluated using synthetic data and real driving data. In these experiments, the proposed method achieved long-term prediction 2-6 times longer than some conventional methods.

### I. INTRODUCTION

#### A. Prediction of driving data

To assist driving and to provide additional benefit to drivers, an intelligent vehicle's capability for recognizing and predicting driver's intention is important. It enables advanced driving assistance systems e.g. information service to assist intended driving maneuver, risk evaluation of current driving scene and maneuvers. Many kinds of approaches have been taken for modeling driving behaviors. Especially, statistical time series modeling techniques, such as hidden Markov model (HMM), hybrid dynamical system (HDS), and Gaussian mixture model (GMM), have been frequently employed. Previous studies, however, indicate it is difficult to predict concrete values of time series driving data including acceleration throttle, brake, and steering angle. Takano et al. [1] used HMM for modeling driving behavior and reported that it is difficult to use their prediction method with the degree of accuracy for automated driving system. HDS including Piece Wise Auto Regressive eXogenous model (PWARX) [2], Stochastic Switched Auto Regressive eXogenous [3] and Auto Regressive Hidden Markov Model [4]

S.Nagasaka is with College of Information Science and Engineering, Ritsumeikan University, 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan s.nagasaka@em.ci.ritsumei.ac.jp

K.Hitomi is with Research & Development Division, Toyota InfoTechnology Center Co.,Ltd. ke-hitomi@jp.toyota-itc.com

N.P.Chandrasiri is with Research & Development Division, Toyota InfoTechnology Center Co.,Ltd. np-chandrasiri@jp.toyota-itc.com

T.Bando is with Corporate R&D Div.3, DENSO CORPORATION takashi\_bandou@denso.co.jp

were also used to predict driving behaviors. HDS has several discrete states corresponding to elemental linear dynamical systems. Although HDS seems to be more complicated than HMM and GMM, HDS does not always outperform them<sup>1</sup>. Angkitirakul et al. [5] showed that GMM and PWARX have almost same prediction capability.

These previous studies imply that predicting concrete value of driving behavior is difficult because a driver is affected by contextual information and his/her intention. Modeling and predicting such various hidden information are important to achieve long-term prediction of driving behavior. In HMM or GMM, the context or the intention is usually modeled as their hidden states. Prediction of hidden states itself is important in many situations. Even if an intelligent vehicle cannot predict concrete value of driving behavior, it can support its driver by predicting contextual information, i.e. hidden states of these probabilistic models. Therefore, we discuss about prediction of a sequence of hidden states in this paper. However, HMM and GMM have Markovian property in their hidden state transition. These cannot model longterm context. Oliver and Pentland[6] enumerated drivers' typical behaviors and prepared left-to-right HMM model for recognition and prediction of drivers' behavior. However, this kind of approach requires preparing drivers' model and it is difficult to enumerate elemental human drivers' behavior completely. Therefore, an unsupervised learning approach is required to model and predict longer-term contextual information. To achieve this, we focus on double articulation structure in driving behaviors. From that point of view, We extends an unsupervised morphological analysis method proposed in natural language processing field and derives a novel semiotic prediction algorithm of driving behavior.

#### B. Prediction based on Double Articulation

Taniguchi and Nagasaka [7] proposed a double articulation analyzer for extracting long-term human motion chunks by connecting several short-term natural segments of human motion. The method is based on analyzing hidden double articulation structure, which is well known in *Semiotics*. Fig. 1 is a conceptual figure of double articulation. Our spoken language and some various semiotic data have a double articulation structure. Most theories of speech recognition

This research was partially supported by a Grant-in-Aid Creative Scientific Research 2007-2011 (19GS0208) funded by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

T.Taniguchi is with College of Information Science and Engineering, Ritsumeikan University, 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan taniguchi@ci.ritsumei.ac.jp

<sup>&</sup>lt;sup>1</sup>HDS is fundamentally a type of HMM. The difference between HDS and HMM is only in a type of output distributions, i.e., HDS has linear dynamics and HMM has Gaussian distributions. However, if HDS assume Gaussian distribution as a prior distribution of input data, and if HMM adopt dynamic features, i.e., including first derivation of feature data, into their feature input data, the two models become equivalent.



Fig. 1. On-line estimation of hidden state using double articulation structure.

have the following assumptions. First, a spoken auditory signal is segmented into phonemes ("a", "b", "c" etc. in Fig. 1). Second, the phonemes are chunked into words ("abc", "e", "db" etc. in Fig. 1). In most cases, we assume that phonemes do not have any meanings, but words have certain meanings. A transition model between the words is represented as language model. In this paper, we presume that human driving behavior also has double articulation structure. We assume that unsegmented driving behavior is segmented into many short-term behaviors based on its linearity or its locality of distribution in its observed state space. The short-term behavior is called *segment* in this paper (see Fig. 1). The segments are chunked into a chunk which corresponds to a word in spoken language (see Fig. 1). The prediction method proposed by Okuda et al. [2] showed the effectiveness of incorporating the symbolized context longer than Markov transition, which was acquired through mostlyunsupervised procedure. Their prediction is a straightforward one depending on a given context with production rules. They also implied strongly the existence of a hierarchical structure in driving behavior.

Assuming that natural driving behavior has double articulation structure, a prediction method of the behavior can be derived. In most previous studies, direct prediction of time series behavior proved to be difficult. In this paper, the time series prediction problem is changed contrastively into semiotic prediction. We, human, recognize our environment abstractly or symbolically (cf. Fig. 1). Therefore, we developed a semiotic predictor which anticipates the next hidden state which is corresponding to a linear model in HDS. We extend double articulation analyzer [7] consisting of sticky Hierarchical Dirichlet Process HMM (HDP-HMM) and Nested Pitman-Yor Language Model (NPYLM), and then propose the novel semiotic predictor that can exploit contextual information in the next section.

#### **II. ALGORITHM**

#### A. Overview

An overview of our proposed semiotic predictor based on double articulation analyzer[7] is given in this subsection.

First, large amounts of high-dimensional time series data in a vehicle are observed. Sticky HDP-HMM[8] is used for segmentation and modeling the target driving behavior. By using sticky HDP-HMM, the analyzer can estimate segments and obtain sequences of hidden state labels (letters) without fixing the number of hidden states. A sequence of the letters that corresponds to observed time series is called a sentence. After obtaining the sentence, it is chunked into a sequence of words. A word corresponds to a sequence of letters. Then, after obtaining several sentences from observed behavior data, an unsupervised morphological analysis method based on NPYLM is applied to them. NPYLM, proposed by Mochihashi et al. [9], is a nonparametric Bayesian language model is comprised of two hierarchical Pitman-Yor (HPY) process. One is a language model and the other is a word model. The language model is an n-gram model of words and word model is an n-gram model of letters. The language model is named NPYLM because an HPY word model is nested by an HPY language model (HPYLM). The language model enables unsupervised morphological analysis, in other words, unsupervised chunking. By using sticky HDP-HMM and NPYLM collaboratively, double articulation analyzer can extract chunks from continuous time series data. By extending this method, semiotic predictor will be obtained.

#### B. Segmentation by Sticky HDP-HMM

The observed multivariate time series data is converted to a sentence using sticky HDP-HMM and we obtain a set of sentences. Sticky HDP-HMM is an extension of infinite HMM (iHMM). Infinite hidden Markov model, proposed by Beal et al. [10], is a first nonparametric Bayesian statistical model which can be substituted for a conventional HMM. A potentially infinite number of hidden states are assumed with iHMM. Through its inference process, iHMM can flexibly estimate the number of hidden states depending on given training data. However, an adequate generative model and an efficient inference algorithm were not shown in [10]. Teh et al. [11] extends hierarchical Dirichlet process mixture into hierarchical Dirichlet process-HMM (HDP-HMM), which is an adequate generative model for iHMM. Fox et al. [8] proposed sticky HDP-HMM with a self-transition bias. This model is an extension of HDP-HMM. By biasing the self-transition probability, sticky HDP-HMM can reduce the frequency of transition among hidden states. This model is more effectively used to model and segment a continuous real-world data stream, e.g., speaker diarization and speech recognition.

Fox et al. [8] also describes a numerical computation algorithm using blocked Gibbs sampler. Straight-forward application of forward filtering-backward sampling algorithm for HMM[12] to iHMM is not feasible because it is impossible to accumulate forward messages for an infinite number of hidden states. Therefore, halting stick breaking process (SBP) [13] and truncating the number of hidden states are unavoidable. Blocked Gibbs sampler, proposed in [8] by adopting weak-limit approximation, accelerates the inference sampling process in HDP-HMM. Practically, the approximation is not so problematic for the purpose of analyzing real-world time series data. Therefore, blocked Gibbs sampler is employed in Taniguchi and Nagasaka [7].

# C. Chunking by Unsupervised Morphological Analyzer using Nested Pitman-Yor Language Model

By an unsupervised morphological analyzer, we can extract word set (dictionary) included in the set of sentences. To chunk sequential letters into several words, the double articulation analyzer[7] employed unsupervised morphological analysis methods using nonparametric Bayesian language models [9], [14]. In the analysis of real-world time series data based on double articulation, chunks are usually unknown in contrast with that a set of words is usually known in spoken language recognition. Therefore, the analyzer has to deal with unknown words. Where, the words are sequences of letters, and the letters correspond to hidden states in sticky HDP-HMM. Unsupervised morphological analysis does not assume preexisting dictionary. Mochihashi et al. [9] proposed an unsupervised morphological analysis method based on Nested Pitman-Yor language model (NPYLM). It consists of letter n-gram in addition to word n-gram model. The both of them use Pitman-Yor process to smooth generative probability of letters and words. NPYLM and probabilistic dynamic programming were employed for chunking sentences written in natural language.

HPYLM is an n-gram language model using HPY process. Pitman-Yor process is a stochastic process whose base measure is itself Pitman-Yor process, which is a generalization of Dirichlet process.

In HPYLM, probability of word w after a context  $h = w_{t-n+1} \dots w_{t-1}$  is calculated as

$$p(w|h) = \frac{c(w|h) - d \cdot t_{hw}}{\theta + c(h)} + \frac{\theta + d \cdot t_h}{\theta + c(h)} p(w|h'), \qquad (1)$$

where, h' is a context whose order is one less than h, i.e.,  $h' = w_{t-n+1} \dots w_{t-1}$ . Therefore, p(w|h') represents a prior probability of w after h, and their probability is calculated recursively. c(w|h) is a count of w after in a context h, and c(h) is summation of all words' counts in a context of h. A context is a sequence of words which are observed before a target word.  $t_{hw}$  is a count that w is estimated to be generated from the context of h', and  $t_h$  is summation of  $t_{hw}$  in a context of h. Discount parameter d and concentration parameter  $\theta$ are hyper parameters of HPYLM (see [14]). Any n-gram probability can be calculated using this equation recursively, except in the case of p(w|h) is unigram probability. In this case, p(w|h') does not exist, and most of previous studies employed some heuristics based on preexisting dictionary to calculate the word unigram. To overcome this problem, NPYLM calculate a base measure of word unigram using letter n-gram smoothed by letter HPYLM. Thus, NPYLM

enables calculation of word n-gram probability without preexisting dictionary, and the word model gives reasonable base measure without the dictionary. In addition, blocked Gibbs sampler and probabilistic dynamic programming enables NPYLM to chunk given letter sequences without heavy computational time[9].

# D. Unsupervised Morphological Analysis for Incomplete Sentence

As shown in Fig. 1, a sentence given to a semiotic predictor is an incomplete sentence. When an unsupervised morphological analyzer tries to parse a given incomplete sentence including probably incomplete word on its tail such as "nicetomeety", it requires an appropriate probability to the incomplete word. For example, in the case of "nicetomeety", its hidden original sentence is "nice to meet you". Thus the unsupervised morphological analyzer has to give appropriate probability to the possible final incomplete word "y", "ty", "ety" or the other suffix of the unsegmented sentence. Possibly, incomplete words have not been registered in an obtained language model, i.e., sampled word set. The probability to the incomplete word is obtained by marginalizing over infinite possible words that include the incomplete word as a prefix. Here, we define a symbol  $\Box$  representing prefix relation between two strings.  $w_1 \sqsubset w_2$  means  $w_1$  is a prefix of  $w_2$ , e.g., "adb"  $\sqsubset$  "abdef" and "24"  $\sqsubset$  "243". The n-gram probability of incomplete word *u* is

$$p(u|h) = \sum_{\forall w \sqsupset u} p(w|h)$$
(2)

by marginalizing the probability of infinite number of possible words based on NPYLM. The set of possible words includes infinite number of unobserved words.

The infinite summation of right side of equation (2) is computable owing to the nonparametric Bayesian characteristics of NPYLM. By using equation (1), we obtain

$$\sum_{\forall w \sqsupset u} p(w|h) = \frac{\sum_{\forall w \sqsupset u} \{c(w|h) - d \cdot t_{hw}\}}{\theta + c(h)} + \frac{\theta + d \cdot t_h}{\theta + c(h)} \sum_{\forall w \sqsupset u} p(w|h').$$
(3)

The first term of right side is summation over observed words who have u as a prefix of itself. The operation of the summation requires obviously finite computational time. The summation in the second term is the same form of left side. Therefore, p(u|h) can be calculated by applying the equation (3) recursively.

Also in the case of considering the probability incomplete words, base measure  $G_0$ , 0-gram probability, is required when the length of context *h* reaches 0. In NPYLM,  $G_0$  is given by its word model. Here, we have to be careful about the difference between incomplete word  $u = (l_1, l_2, ..., l_m)$ and complete word  $w = (l_1, l_2, ..., l_m)$ . An incomplete word *u* does not have *EOW* which is a special letter representing the end of the word.  $w = (l_1, l_2, ..., l_m)$  is obtained by adding *EOW* to the end of  $u = (l_1, l_2, ..., l_m)$ . Therefore, if *EOW* is described explicitly, the *w* becomes  $w = (l_1, l_2, ..., l_m, l_{m+1} =$  *EOW*). As a result, the probability for incomplete word can also be computable by using similar marginalization, i.e.,

$$G_0(u) = \sum_{\forall w \sqsupset u} p(w) \tag{4}$$

$$= p(u) \sum_{\forall w \supseteq u} p(w - u|h = u)$$
(5)

$$= p(u), \tag{6}$$

where w - u is a substring obtained by deleting prefix u from w. This means the incomplete word probability can be calculated easily by using word model of HPYLM. The calculation process is completely same as the one for complete words except for *EOW*. Now, unsupervised morphological analyzer can parse incomplete sentence and it can be applied to incoming time series data.

#### E. Semiotic Prediction for Time Series Data

Various multimodal sensor data of intelligent vehicle is observed while a driver drives the vehicle. However, the observed data is still a part of complete driving data, and then, the sequence of hidden states is considered as an incomplete sentence from the viewpoint of double articulation. Therefore, an unsupervised morphological analyzer for incomplete sentence is applied to the estimated sequence of hidden states. After obtaining chunked sentence and an incomplete last word, the proposed semiotic predictor gives MAP estimation for the last word and subsequent words. If the observed data has double articulation structure clearly, this method is expected to predict the transition of the hidden states more properly than conventional Markov model.

#### **III. EXPERIMENT 1: SYNTHETIC DATA**

The proposed algorithm assumes target time series data to have double articulation structure. To evaluate prediction performance on such a data, the proposed semiotic predictor was applied to a synthetic symbolic data.

#### A. Conditions

The semiotic predictor assumed that the data set consists of a number of sentences, and that a number of words and letters (states) are unknown. To generate a synthetic data with double articulation structure, we trained a generative language model in which a word model is trained with a document data obtained from a web site and a word transition probability is generated by stick breaking process. Finally, as a synthetic data set, 100 artificial sentences which include 50 words in total were generated by using the language model.

Leave-one-out procedure was employed to test prediction performance of the predictor. One sentence was removed from the data set as a test data, and the other sentences were used as training data for NPYLM. The test data was modified into an incomplete sentence by elimination of a last part of the sentence. The semiotic predictor predicts the subsequent erased letters from the incomplete sentence. We compared the proposed method, *NPYLM with Prediction*, with *Letter Markov Model* which predicts the next state based on the



Fig. 2. Averaged length the semiotic predictor could predict



Fig. 3. A histogram of length the semiotic predictor could predict

present states<sup>2</sup> and *Simple NPYLM* which does not care the incompleteness of target sentence.

#### B. Result

Fig. 2 shows the averaged length each model can predict the subsequent letters i.e., it shows how many subsequent letters each method could estimate correctly. This shows that the NPYLM with Prediction outperforms the other methods. Fig. 3 shows a histogram of correctly predicted length. In contrast to the predicted length of the Letter Markov Model which decays exponentially with the prediction length, NPYLM with Prediction achieves much longer prediction of letter sequence so many times. This result owes to the context information and obtained knowledge of words acquired from the data set in unsupervised way.

#### IV. EXPERIMENT 2: DRIVING DATA

We applied the proposed method to a real-world driving data and evaluated the predictability of driving data based on double articulation structure.

# A. Conditions

In this experiment, a driver drove a car through two types of courses. Fig. 4 shows the example view from a driver. The

<sup>&</sup>lt;sup>2</sup>Of course, sticky HDP-HMM assumes that hidden states transit from one to another based on this Markov model.



Fig. 4. An example of a view from driver's seat



Fig. 5. Abstract figures of the two driving courses in the experiment

abstract figures of the two courses are shown in Fig. 5. At first, the car stopped at a parking space. The driver started driving into one of the courses. After driving the course, the car returned to the original position and stopped in every trial. The driver drives each course for five times. Finally, we collected 10 time series data of driving behavior; five tracks for each course in Fig. 5. . In each track, some unpredictable disturbances were observed, e.g., pedestrians walked across a road, traffic light turned into red, and there was a leading vehicle or not.

Driving behavior data is consisted of velocity of the car, steering angle, brake pressure, and accelerator position. By adding two dynamic features, temporal difference of velocity and steering angle, to the four time series data, we obtained the six dimensional time series data as the driving behavior data. By applying sticky HDP-HMM to the obtained data, multivariate sensory time series data was encoded into a sequence of segments representing hidden state labels of sticky HDP-HMM. Then, three prediction methods are evaluated using the sentences.

# B. Result

First, the unsupervised double articulation analyzer segmented and chunked the sequence of driving through the course. The relationship between positions on the courses and estimated segments (represented by colors) and chunks (represented by gray bar) are shown in Fig.  $6^3$ . As a result, it is observed that similar behaviors on the course are often encoded into the same sequence of segments (for example, see lower left corner and lower right corner in track06). Next, the prediction performance of the proposed method

<sup>3</sup>The position was estimated from encoder's record. Therefore, the trajectories are deviated from the true position.



Fig. 6. Top view of estimated segments and chunks (top) in the 1st track and (bottom) in the 6th track



Fig. 7. Averaged length the semiotic predictor could predict

was evaluated. The experimental procedure is same as the experiment 1. Fig. 7 shows the average length of the predicted sequence of segments by each method. Fig. 8 shows that histogram of the predicted length. These figures show the NPYLM with Prediction outperform the Markov model and the Simple NPYLM which does not take care about incompleteness of a hidden sentence. The relationship between the number of erased segments from the tail of the original sentence and average predictable length is also shown in Fig. 9. It shows that our proposed semiotic predictor, the NPYLM with Prediction, can predict a longer sequence from any breaking points than the other two methods. However, the difference between the performance of the proposed method and the baseline methods was smaller than synthetic data.



Fig. 8. A histogram of length the semiotic predictor could predict



Fig. 9. Relation between prediction point of test data and averaged predictable length

In the real-world condition, the obtained letter sequences might contain recognition errors. Because the proposed semiotic predictor has the cascade structure of sticky HDP-HMM for segmentation and NPYLM for chunking, segmentation errors contaminate the subsequent process, unsupervised morphological analyzer. There are still several possibilities that could reduce performance of our proposed method, e.g., the sparseness of training data set, and ambiguousness of double articulation structure in driving data. To overcome these problems, an unsupervised learning architecture should be improved to segment and chunk simultaneously.

## V. CONCLUSIONS AND FUTURE WORKS

We proposed a novel semiotic prediction method for driving behavior based on double articulation structure. To construct the predictor, the unsupervised morphological analyzer proposed by Mochihashi et al. [9] is extended to deal with incomplete sentence. The extension enables an analyzer to parse online incoming time series data which potentially has double articulation structure. The proposed method was applied to a synthetic data set, and our method outperformed the conventional Markov model and Simple NPYLM which did not care about incompleteness of the observed sentence. We also applied the proposed method to a real driving data set measured by commercially available vehicle. Our proposed method outperformed the baseline methods.

As we mentioned in the first section, many previous

studies about predicting driving behavior used HMM, HDS or the other switching model. Most of them assume Markov transition of the hidden states. However, our result shows that driving data have double articulation structure and that using this structure improves long-term prediction performance. This result means statistical time series modeling techniques for driving behavior should take hierarchical structure like double articulation into consideration. An HMM is not enough to model driving behaviors. To apply this method to more natural daily driving behaviors and to evaluate our method are future work. To develop integrated theoretical method for simultaneous optimization of segmentation and chunking which has a proper total generative model is also our future work.

#### REFERENCES

- W. Takano, A. Matsushita, K. Iwao, and Y. Nakamura, "Recognition of human driving behaviors based on stochastic symbolization of time series signal," in *Intelligent Robots and Systems*, 2008. *IROS 2008. IEEE/RSJ International Conference on*. IEEE, 2008, pp. 167–172.
- [2] H. Okuda, T. Suzuki, A. Nakano, S. Inagaki, and S. Hayakawa, "Multi-Hierarchical Modeling of Driving Behavior Using Dynamics-Based Mode Segmentation," *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 92, no. 11, pp. 2763–2771, 2009.
- [3] S. Sekizawa, S. Inagaki, T. Suzuki, S. Hayakawa, N. Tsuchida, T. Tsuda, and H. Fujinami, "Modeling and recognition of driving behavior based on stochastic switched arx model," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8, no. 4, pp. 593–606, 2007.
- [4] Y. Kishimoto and K. Oguri, "A modeling method for predicting driving behavior concerning with driverfs past movements," in *Vehicular Electronics and Safety, 2008. ICVES 2008. IEEE International Conference* on. IEEE, 2008, pp. 132–136.
- [5] P. Angkititrakul, T. Ryuta, T. Wakita, K. Takeda, C. Miyajima, and T. Suzuki, "Evaluation of driver-behavior models in real-world carfollowing task," in *Vehicular Electronics and Safety (ICVES), 2009 IEEE International Conference on*, 2010, pp. 113–118.
- [6] N. Oliver and A. Pentland, "Graphical models for driver behavior recognition in a smartcar," in *Intelligent Vehicles Symposium*, 2000. *IV 2000. Proceedings of the IEEE*. Ieee, 2000, pp. 7–12.
- [7] T. Taniguchi and S. Nagasaka, "Double articulation analyzer for unsegmented human motion using pitman-yor language model and infinite hidden markov model," in *Proceedings of 2011 IEEE/SICE International Symposium on System Integration*, 2011.
- [8] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "The sticky hdp-hmm: Bayesian nonparametric hidden markov models with persistent states," MIT Laboratory for Information and Decision Systems, Tech. Rep. 2777, 2007.
- [9] D. Mochihashi, T. Yamada, and N. Ueda, "Bayesian unsupervised word segmentation with nested Pitman-Yor language modeling," in Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1, 2009, pp. 100–108.
- [10] M. Beal, Z. Ghahramani, and C. Rasmussen, "The infinite hidden Markov model," *Advances in Neural Information Processing Systems*, vol. 1, pp. 577–584, 2002.
- [11] Y. Teh, M. Jordan, M. Beal, and D. Blei, "Hierarchical dirichlet processes," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1566–1581, 2006.
- [12] S. Scott, "Bayesian methods for hidden markov models: Recursive computing in the 21st century," *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 337–351, 2002.
- [13] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, vol. 4, no. 2, pp. 639–650, 1994.
- [14] Y. Teh, "A hierarchical bayesian language model based on pitman-yor processes," in *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2006, pp. 985–992.