Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

CrossRef DOI of original article:

Automatic Identification of Driving Maneuver Patterns using a Robust Hidden Semi-Markov Models

Yang Chen

Received: 1 January 1970 Accepted: 1 January 1970 Published: 1 January 1970

Abstract

There is an increase in interest to model driving maneuver patterns via the automatic unsupervised clustering of naturalistic sequential kinematic driving data. The patterns learned are often used in transportation research areas such as eco-driving, road safety, and intelligent vehicles. One such model capable of modeling these patterns is the Hierarchical Dirichlet Process Hidden Semi-Markov Model (HDP-HSMM), as it is often used to estimate data segmentation, state duration, and transition probabilities. While this model is a powerful tool for automatically clustering observed sequential data, the existing HDP-HSMM estimation suffers from an inherent tendency to overestimate the number of states. This can result in poor estimation, which can potentially impact impact transportation research through incorrect inference of driving patterns. In this paper, a new robust HDP-HSMM (rHDP-HSMM) method is proposed to reduce the number of redundant states and improve the consistency of the model?s estimation. Both a simulation study and a case study using naturalistic driving data are presented to demonstrate the effectiveness of the proposed rHDP-HSMM in identifying and inference of driving maneuver patterns.

Index terms— hidden markov model, driving maneuver, dirichlet process, naturalistic driving data

1 I. Introduction

The analysis of vehicle driving styles is prominent to the field of intelligent transportation and vehicle calibration [1,2]. The term driving style can be referred as a set of dynamic activities or steps that a driver uses when driving. Hence, this type of research impacts eco-driving, road safety, and intelligent vehicles [3,4,5]. To model these driving styles, one popular approach is the use of a Hierarchical Dirichlet Process Hidden Semi-Markov Model (HDP-HSMM) [6]. This model is powerful in that it considers the sequential nature of driving kinematic signals, and estimates data segmentation, behavior state duration, and state transition probabilities. The HDP-HSMM provides semantical way for analyzing driver behaviors, and is thus popularly used for describing driving styles. Figure 1b shows an While the HDP-HSMM is powerful, literature outside of the field of transportation details how the model's use of an HDP prior can lead to redundant and inconsistent state estimations. This detail is important as it needs to be considered by researchers attempting to utilize the HDP-HSMM to describe driving styles. For example, Figure 1 clearly has redundant states as seen by the green shaded states. The redundant states can make analysis of HDP-HSMM outputs across multiple datasets difficult for researchers hoping to utilize the HDP-HSMM to model driving styles. This paper addresses this issue by presenting an algorithm that reduces redundant states to improve consistency while still aligning to the structure of a basic HDP-HSMM. The presented algorithm results a more robust HDP-HSMM (rHDP-HSMM) that is expected to output a more consistent data segmentation, behavior state duration, and state transition probabilities than a basic HDP-HSMM. This will impact the transportation field in that driving maneuver patterns can be better grouped together for classification or behavioral studies.

The remainder of this paper is as follows. Section 2 will provide the background about HDP-HSMM's from a statistical perspective, and highlight the current set of approaches towards addressing the issues derived from the HDP prior. Section 3 will provide the data description and the model formulation of a basic HDP-HSMM.

Section 4 discusses the details of inference for a HDP-HSMM, and how this paper's algorithm can be included within the inference to produce a more robust HDP-HSMM. Section 5 presents a simulation study, in which the rHDP-HSMM is compared to the basic HDP-HSMM based on simulated data. Section 6 presents a case study that uses realistic, naturalistic driving data to compare the rHDP-HSMM with the original HDP-HSMM method on the basis of describing driving patterns. Finally, Section 7 summarizes new contributions and major conclusions of the paper.

The HDP-HSMM was designed to improve upon the structure of a discrete state-space Hidden Markov Model (HMM). HMM's are also popularly used for describing sequential data [7,8,9,10,11,12]. In particular, the HMM [13,14] utilizes a two-layer structure (Figure 2a) to represent sequential data observed at equally spaced time points. In this model, data is assumed to be generated from a set of probability distribution functions dependent on corresponding hidden states. The hidden states determine the data segmentation. Transitions among hidden states are modeled as a Markov Chain. This allows for the consideration of time sequence information during inference and further aids in the prediction of future states. One condition of using the Markov Chain is that the state duration of each hidden state is assumed to be Geometrically distributed. While the HMM is able to define data segmentation and state transitions, its definition of state duration is severely limited by the model's structure. This limitation lead to the development of the Hierarchical Dirichlet Process Hidden Semi-Markov Model (HDP-HSMM) [15] which provided two key improvements to the HMM. The first improvement was the removal of the HMM's assumption of geometrically distributed state duration. As the HDP-HSMM uses a Semi-Markovian approach to model the state transitions ?zs , this removes self-transitions from the transition matrix. As a consequence, this frees the geometric distribution restriction on the duration D s , which leads to a three-layer structure model as shown in (Figure 2b). In other words, users can choose different models for representing state duration, while allowing the segmentation of hidden states to be directly represented by z s .

The second improvement was the introduction of Dirichlet Processes to the model. The Dirichlet processes is an extension to the Dirichlet distribution, as atoms can be sampled from it based on an input distribution. However, one key difference is that the Dirichlet Process assigns a probability of drawing a new atom from the input distribution and a separate probability of drawing an atom based on the atoms seen in previous samples. The resulting distribution is discrete and similar to the input distribution, but also has the possibility of having infinite discrete atoms if infinite samples were drawn. This phenomenon is interesting in the context of HMMs and HSMMs, as the Dirichlet process can be used as a prior to the state transition probability vector [16,17,15]. Doing this allows the probability vector length (i.e. models' number of states) to grow without limit during inference, which implies the Dirichlet process also acts like a prior on the number of clusters. In the HDP-HSMM, a Hierarchical Dirichlet Process (HDP) is used as a prior on the state transitions, which allows all the state transition probabilities to share a similar base distribution. This is beneficial, as all the states represented in the base distribution are shared between all the different state transition probabilities, while allowing each transition probability be dependent on the exit state. Hence, for the context of modeling of driving maneuvers, the HDP-HSMM is preferred as it allows greater flexibility in defining the relationship between the data and segmentation, state duration, and state transitions.

While the Dirichlet Process's clustering properties have been seen as a tool to address the model selection for Bayesian nonparametric approaches [18,19], the Dirichlet Process is known to have inconsistency issues regarding estimation of the true number of states.

posterior does not concentrate at the true number of components, and instead introduces extra clusters even if they are not needed. Under the context of HMMs, [21] showed how the Dirichlet Process also leads to the creation of redundant states, which presents an unrealistic rapid switching between states in the inferred transition matrices. Under the context of HSMM's, Figure 1 shows how this side effect occurs even in the HDP-HSMM. However, for the HDP-HSMM, the redundancy issue also affects the inference of transition probabilities and duration estimation.

A few works exist that focus on solving this issue for HMM's. [22] discussed HMM's utilizing a Dirichlet prior, and the assumptions on the prior required for the consistency. [23] developed the sticky HDP-HMM (sHDP-HMM) to consider the issue of redundant states. This model adds a bias to the prior on the rows of the transition matrix which emphasizes self-transitions. This results in an increased state duration for each learnt state, which allows the sHDP-HMM to avoid redundant states with short state duration. However, this strategy cannot be applied to HDP-HSMM as the modeling structure of HMM's is inherently different from HSMM's. Outside of HMM and HSMM modeling, [24] focused on the Dirichlet Process Mixture model, and presented the Merge-Truncate-Merge algorithm, which guaranteed a consistent estimate to the number of mixture components. This post-processing procedure takes advantage of the fact that the posterior sample tends to produce a large number of atoms with small weights, and probabilistically merges atoms together.

Given these approaches, this paper attempts to address the HDP's inconsistency problem by taking inspiration from both the sticky HDP-HMM and the Merge-Truncate-Merge algorithm. The idea is to apply a merging procedure during inference which promotes longer durations and the avoidance of redundant states. In doing so, this paper's contribution will include demonstrating how the HDP-HSMM becomes robust to the inconsistencies brought by the HDP prior and how this paper's method can reduce the number of redundant states to better define driving maneuvers existing in Figure 1a. A brief summary, which describes where our model fits in relation to the other models described in HMM literature, is given in Table 1.

2 State Duration

Distribution Model Extension (not sensitive to prior) Geometric HDP-HMM [14] sticky HDP-HMM [7] Any Discrete Distribution HDP-HSMM [15] robust HDP-HSMM (This paper) In this paper, a sequential dataset consists of a series of observations collected at T chronologically ordered time points. At each time point t, y t ? R p represents the p-dimensional signal responses. The sequential data is assumed to follow multiple phases; there exists a partition [20] provided an example for Dirichlet Process Mixture Models which demonstrates how the segment, and (3) identify the probability of transitioning from one distribution to another. The challenge lies in little information being available relating to the number of states, the states' durations, and the transition probability matrix.1 = t 1 1 ? t 1 2 ? ... ? t 1 S = T -D S ,

3 b) Basis of HDP-HSMMs and Notations

The HDP-HSMM accomplishes this objective with the following structure. The multivariate sequential data is represented by the sequence (y t) t=1:T := {y t ? R p : t = 1, ..., T } and is assumed to transit among K different hidden states. The hidden states at each time point t are represented by the sequence (x t) t=1:T := {x t ? {1, 2, . . . , K} : t = 1, ..., T }, and can be further divided into S segments. Within each data segment s ? {1, 2, . . . , S}, all hidden states share the same index (labeled by the super-state z s ? {1, 2, . . . , K}), and the state duration of the segment is denoted by D s . As such, the start and end times of each segment s are indexed by time stamps t 1 s and t 2 s , respectively. They can be calculated as t 1 s = s<s D s and t 2 s = t 1 s + D s -1 where s represents all the segments before segment s. The state of segment s is assumed to be Markovian with a transition probability? i,j = Pr(z s = j | z s-1 = i),

where the rows of the transition matrix are denoted as ? i = [? i,1 ? i,2 . . . ? i,K]. However, as each state has a random state duration D s ? g(? zs), the HSMM does not permit selftransitions to occur. To consider this, the transition rows of ? i are adjusted to ?i with each element being ?i,j =? i,j 1-? i,i (1 -? i,j) (where ? i,j = 1 if i = j; ? ij = 0 otherwise).

The relationship between the observation sequence and the segmentation described above can be seen by the emission distribution functions f (? zs) and the state duration probability mass functions g(? zs) with parameters ? zs and ? zs being dependent on segment s. The priors on ? zs and ? zs are denoted by H and G respectively.

A Hierarchical Dirichlet Process (HDP) is used to define a prior on the rows of the transition matrix (? i) to learn the number of unknown states. The HDP creates a countably infinite state-space and utilizes a stick-breaking process ? ? Beta(?) [25] to determine the number of unknown states (K). A smaller ? (?? 0) yields more concentrated distributions, which plays a part in shaping the transition pattern. Each row of the Markovian transition probability matrix is sampled from a Dirichlet process (? i iid ? DP(?, ?)) and its similarity to the stick-breaking process depends on the concentration parameter ? ? (0, ?).

The HDP-HSMM is shown in Figure 2b and can be formulated as follows:? ? Beta(?), ? i iid ? DP(?, ?) (? i , ? i) iid ? H × G i = 1, 2, . . . , z s ? ?z s-1 D s ? g(? zs) s = 1, 2, . . . , x t 1 s :t 2 s = z s , y t 1 s :t 2 s iid ? f (? zs) t 1 s = s<s D s t 2 s = t 1 s + D s -1.**(1)**

Typically, Gibbs sampling approaches are used for statistical inference of the model parameters of the HDP-HSMM, which requires the full conditional distributions of the model parameters [26]. The details of the general Gibbs sampling procedure and how this paper applies a merging algorithm within it to create a robust HDP-HSMM is presented in the next section.

IV. Proposed Robust HDP-HSMM a) Inference

The details of the block sampling procedure presented in [15] to infer the parameters for the HDP-HSMM are discussed here. Additional insight regarding this paper's proposed changes will also be included in this section. Assume initial values have been set for the state sequence, the emission parameters, the duration parameters, and the transition probabilities:(x t) (0) , {? i } (0) , {? i } (0) , {? i } (0) .

Step 1: The block sampling procedure begins iteration m = 1 with the sampling of the emission, duration, and transition distribution parameters. The distributional parameters can be sampled independently of one another, conditional on data assigned to each state i under the current state sequence (x t) (m-1) . Assuming distributions with conjugate priors are utilized within the HDP-HSMM, this step can be simplified significantly into the following statement:{? i } (m) ? h ? i (? i |(x t) (m-1) , (y t), H, G, ?) {? i } (m) ? h ? i (? i |(x t) (m-1) , (y t), H, G, ?) {? i } (m) ? h ? i (? i |(x t) (m-1) , (y t), H, G, ?),

where h ? refers to the updated posterior corresponding to the conditional distribution with parameter ?.

Step 2: Once a new set of parameters have been sampled, it is practical to apply some identifiability constraints to the parameters to help ensure state switching does not occur during the sampling procedure. State switching is a problem mentioned in literature [27,28], in which the permutation of defined states is not considered during the sampling procedure. Identifiability constraints ensure the order of states does not change between iterations of the sampling procedure, and helps ensure the posterior chain is not multimodal at the end of the sampling procedure. While many types of constraints can be applied, such as rearranging the states such that ? 1 < ? 2 < ? 3 < . . . , the constraints used in this paper are be mentioned in each section directly.

Step 3: After identifiability constraints have been applied, the new state sequence can be sampled. [15]'s procedure makes use of the following backwards messages:B t (i) :=p(y t+1:T |x t = i, F t = 1) = j B * t (j)p(x

t+1 = j|x t = i) B * t (i) :=p(y t+1:T |x t+1 = i, F t = 1) = T -t d=1 B t+d (i)p(D t+1 = d|x t = i)p(y t+1:t+d |x t+1 = i, D t+1 = d) + p(D t+1 > T -t|x t+1 = i)p(y t+1:T |x t+1 = i, D t+1 > T -t) B T (i) :=1,

where F t = 1 denotes a new segment begins at t + 1, and D t+1 denotes the duration of the segment that begins at time t + 1 [29]. The procedure for obtaining the posterior state sequence begins by drawing a sample for the first state using the following formula:p(x 1 = k|y 1:T) ? p(x 1 = k)B * 0 (k).

Next, a sample is drawn from the posterior duration distribution by conditioning on sampled initial state x1 :p(D 1 = d|y 1:T , x 1 = x1 , F 0 = 1) = p(D 1 = d)p(y 1 : d|D 1 = d, x 1 = x1 , F 0 = 1)B d (x 1) B * 0 (x 1) .

The rest of the state sequence can be sampled assuming the new initial state has distribution p(x D 1 +1 = i|x 1 = x1) and repeating the process, until a state is assigned for all indices t = 1, . . . , T .

Step 4: Once the new state sequence is sampled, the Gibbs sampling procedure normally returns back to Step 1, increments m by 1, and repeats Steps 1 to 3 until posterior convergence. However, before doing that, this paper propose adding an additional sampling Step 4 that removes redundant states from the posterior state sequence(x t) (m) ? h (xt) ((x t)|{? i } (m) , {? i } (m) , {? i } (m) (x t) (m) , (y t), H, G, ?),**(2)**

where h (xt) (?) represents a sampling step proposed by this paper to promote robustness.

The proposed Step 4 is the main contribution of this paper. This section will provide the details on how to implement Equation 2 described in Step 4 above. The procedure is described by first defining redundancy between two states: Definition 4.1. In the state sequence (x t) t=1:T , the states i and j are identified as redundant states if D f (? i), f (? j) ? ? , where ? is the decision threshold and D f (? i), f (? j) is a measure of divergence that gets larger when the distributions f (? i) and f (? j) are more different from one another.

Although D f (? i), f (? j) can be any measure of divergence satisfying Definition 4.1, the remainder of the paper will assume D f (? i), f (? j) = ||(? i -? j)|| 2 is the â??" 2 norm of the difference in parameters. Now that redundancy has been defined, the details of Equation 2 can be represented by Algorithm 1. In short, the procedure samples a new state sequence that contains no redundant states. [15] describes a weak-limit approximation to the Dirichlet Process prior,?|? ? Dir(?, . . . , ?) ? j |? ? Dir(?? 1 , . . . , ?? K), j = 1, . . . , K, b) Implementation of Step 4

as well as an augmentation that introduces auxiliary variables which are added to the ? vector to preserve conjugacy. This approximation eases the use of sampling procedures when dealing Dirichlet Processes [30]. Taking this approach, the ? vector takes no consideration of redundant states, which may negatively impact the posterior of ? j . The presence of redundant states means the posterior transition probabilities contain extra transitions to and from redundant states, which dilute the underlying transition process. To counter this, h (xt) (?) aims to adjust the ? vector in this step as to discourage transitions to redundant states in future steps, and preserve the true underlying transition process.

Algorithm 1 describes h (xt) (?) entirely. The procedure begins by initializing a new vector ?, a new state sequence (x t) (m) , and taking the input of a similarity threshold ? . Taking inspiration from [24], the states order is firstly randomized in which redundancy is checked. This is to ensure the start of the merging procedure begins at a point close to the "central mass" of the emission distribution clusters with a high probability. Going through the order, if the state exists within the new state sequence (x t) (m) , the algorithm proceeds to find similar states based on our similarity metric and similarity threshold. Weights are then defined which will determine the probability of retaining a state from the set of redundant states. These weights are determined by the probability of other non-similar states transitioning to the state of interest and then normalized. The state by which to retain is selected randomly in accordance to the probabilistic weights, and the rest of the similar states are erased from the state sequence. Vector ? is further updated by weakening the unselected similar states values in the vector.

After implementing Algorithm 1, the sampling procedure is allowed to return to Step 1. Noticeably, every time this step is implemented, the algorithm begins with the originally sampled ? and (x t) (m) , but ends with a ? and (x t) (m) that encourages the transition matrix in Step 1 to promote transitions to non-redundant states and allow larger sample sizes for

4 Algorithm 1 Sample a State Sequence Containing No Redundant States

Initialize ? = ?, (x t) (m) = (x t) (m) , and define similarity threshold ? Reorder {? i : i ? (x t) (m) } into new order {? I i : i ? (x t) (m) } using random sampling without replacement where

? i corresponds to index the unique states existing in (x t) (m)

? I i corresponds to the new index of state i in the new order I = {1, 2, 3, . . . } while I is not an empty set do Let i correspond to the first I i appearing in the new order I Calculate D f (? i), f (? j) for all j ? = i where j ? (x t) (m) ? Similarity metric. Define set J = {j : D f (? i), f (? j) ? ? } and set J ? = {j : D f (? i), f (? j) > ? } for j ? J do ? j = i?J ? ? i,j ? Weights depend on transition probabilities from non-similar states. end for Sample j * from P(j *) where P(j * = j) = ? j /(j ? j) ? j * is the redundant state to keep. Update ?j = 0.1 * ? j for all j ? J where j ? = j * ? Influence transition prior. Update xt = j * for all {t : xt ? J} ? Influence data used for inference. Remove I j from I for all j ? (J ? i).

? Prevent merging these states in future iterations. end while Output final ? and (x t) (m) ? These will be used in next iteration of Gibbs sampling.

V.

5 Simulation Study

In this section, simulations are used to demonstrate the advantages of the proposed rHDP-HSMM method. The robustness and modeling accuracy is compared with the existing HDP-HSMM method. The simulation is designed as follows.

For each simulation, a sequence of observed data is generated with 30 total change points based on the distributions and parameters in Table 2. The emission parameters were specifically selected as they feature some small overlap between their distributions.

The generated sequence begins with a state being randomly selected from the three listed in Table 2. A length of duration is sampled from the selected state's duration distribution, which determines how many samples to draw from that state's emission distribution. Once the emission samples are collected, they are stored in the sequence, and the next state is sampled according the to that state's transition probability. The process is repeated 30 times to create a simulated sequence of "observed" data. An example of a simulated dataset can be observed in the Figure **??**. In each simulation, both the HDP-HSMM and the rHDP-HSMM are trained on the observed data with the same initial distributions and priors. The prior distributional forms were selected as to allow models to make use of conjugate relationships. Their parameters were selected as to ensure the true distributional parameters could be inferred with high probability. Each simulation's initial parameter values for the HDP-HSMM and rHDP-HSMM were drawn according to the selected prior. The maximum number of states for both models was set to 20. Each state's initial emission distribution was assumed Normal(µ, ? 2). The mean's prior distribution was set to µ ? Normal(µ 0 = 0, ? 2 0 = 4). The variance's prior distribution was set to A models had identifiability constraints implemented such as to order their states in increasing order of the posterior mean of their emission distribution. Furthermore, both models performed their respective Gibbs procedure over a maximum of 10000 iterations, or until their Gelman-Rubin statistic [31] reached less than 1.1. The burn-in period for both models was set to 100 iterations. Every 5th iteration of the sampled parameter chains was collected as to remove autocorrelation (resulting in a chain of 2000 length if convergence was not met). The rHDP-HSMM threshold for removing redundant states was set to 1.5. The posterior parameter values for each state was calculated as the mean of the most recent 20% of samples collected from the posterior parameter chains. The posterior sequence was selected to be the mode of the most recent 20% of samples collected from the posterior state sequence.

The results of a single simulation are shown in Figures 4,5, and 6. Figure 4 compares the The simulation is repeated 100 times, and the results are shared in Figure **??** and Table 3. Looking at the number of estimated states between the HDP-HSMM and the rHDP-HSMM, it is clear that the rHDP-HSMM's inference procedure removes states that would be otherwise present in a standard HDP-HSMM (Figure **??**). In fact, 80 of the 100 simulations resulted in the rHDP-HSMM correctly inferring the true number of states. Furthermore, Table 3 shows that the rHDP-HSMM converged on average with fewer iterations than the HDP-HSMM. This table also shows that while both models are able to correctly capture all the true change points, the standard HDP-HSMM tends to estimate many more change points than the rHDP-HSMM. This is due to the redundancy issue, which the rHDP-HSMM eliminates through its modified inference procedure. The benefit of the proposed rHDP-HSMM is demonstrated via the real-world application of modeling vehicle driving maneuver patterns. This type of modeling is useful for the development intelligent driving assistant systems and autonomous driving vehicles. The dataset analyzed in this study was collected by University of Michigan's Transportation Research Institute [32]. Several kinematic driving signals were collected from human-driven vehicles during their everyday activities. This naturalistic dataset is rich with information related to discover common driving maneuvers and behaviors. [1]. Signals are recorded on trip by trip basis, which begins when the vehicle is turned-on and ends when the vehicle is turned-off. An example of a trip can be seen in Figure 8.

The kinematic signals of interest are acceleration, lane offset, and yaw rate. Acceleration and lane offset reflect a driver's intention of moving in the longitudinal and lateral directions respectively. Yaw rate captures a driver's intention of of changing the forward direction of the car. Together, they form a multivariate time-series sampled at 10 Hz which should be highly correlated with human-driving behaviors. An example of the collected signals is The colors in Figure 8 represent the labeling results after training the 0.5 threshold rHDP-HSMM. Noticeably, the rHDP-HSMM segments the road into 9 states. Looking deeper at Figure 8b, it is clear that each state is primarily dictated by changes in yaw rate. Hence this model is able to capture portions of the road where various turning maneuvers are intended by the driver (Figure 8a). Comparing Figure 8a with the HDP-HSMM segmentation shown in Figure 1a, it is clear how the rHDP-HSMM merged the HDP-HSMM's 17 states into a more clear representation of maneuvers used on the road.

The rHDP-HSMM and HDP-HSMM are further compared in Figure 9 by using states obtained from the curved portion of the road marked in Figure 8a. Six other trips existed where the same driver drove on that part of the road. Hence, both the HDP-HSMM and the rHDP-HSMM are trained again on each of the other trips under the same initial parameters. The learned states from each model which occurred on the marked portion are analyzed in Figure 9. Figures 9a and9b shows the emission means and durations learned by the HDP-HSMM and

the rHDP-HSMM respectively. Interestingly, Figure 9b shows how the rHDP-HSMM concentrates the emission means in various quadrants of the graph. These quadrants relay a positive yaw rate, a negative lane offset, and a positive acceleration in all the learnt means. The concentration of these means in each quadrant indicate a consistency in maneuvers among the various trips, which translates to a left turning action intended by the driver. This same conclusion is not easily recognizable in Figure 9a, as the HDP-HSMM loses this consistency in the learnt means. The difference in learning procedure between the HDP-HSMM and the rHDP-HSMM suggests that the HDP-HSMM's lack of concentrated means derives from the HDP-HSMM overestimating the number of states. As the rHDP-HSMM inference procedure merges similar states together, the emission means of each state can be inferred with a greater amount of data, providing both more consistent estimates and more consistent conclusions.

The HDP-HSMM is a powerful model for discovering driving maneuver patterns from kinematic driving data. This paper details an extension to the HDP-HSMM in which this paper refers to as a robust HDP-HSMM (rHDP-HSMM). This model provides a solution to the inconsistency problem caused by the HDP prior. Looking through the lens of a weak-limit 9a shows the means from the original HDP-HSMM, while Figure 9b shows the means from the proposed rHDP-HSMM

6 VII. Discussion and Conclusion

approximation of the HDP prior, the problem typically occurs as the Dirichlet distribution takes no consideration for redundant states, which dilutes the underlying transition process. The rHDP-HSMM solves this issue by adjusting the sample from Dirichlet distribution by checking which states can be merged together. The model then scales down the weights which encourage transitions to redundant states. As a result, the rHDP-HSMM learns fewer redundant states and estimates longer state durations when compared to the original HDP-HSMM. This change leads to improved segmentation and more accurate transition probability representation, which is useful for the application of learning driving maneuvers.

Two case studies are presented to further demonstrate the ability of the proposed rHDP-HSMM over the HDP-HSMM. The first study is a simulation which utilizes 1-dimensional normal distributions for the emission function. The rHDP-HSMM demonstrates a clear improvement with regards to the posterior chains. The emission parameters converge much faster, the duration posteriors have far less variance than the HDP-HSMM's duration posterior, and finally the posterior state sequence presents far less change points than the HDP-HSMM's. Over the course of 100 simulations, the rHDP-HSMM out performs the HDP-HSMM in terms of convergence and having less extra change points relative to the truth.

The second study demonstrates of the effectiveness of the model in identifying and inferring driving maneuver patterns from a naturalistic dataset of kinematic signals. It is shown how the rHDP-HSMM's merging procedure reduces the number of states to describe a trip from 17 to 9 states when compared to a regular HDP-HSMM. The states are highly interpretable and now specifically capture portions of the road where various turning maneuvers are intended by the driver. In addition to this, the study also compares the results from multiple trips occurring on a curved portion of the road. The results show how the rHDP-HSMM consistently estimates similar emission distributions from multiple trips when compared to the original HDP-HSMM estimates.

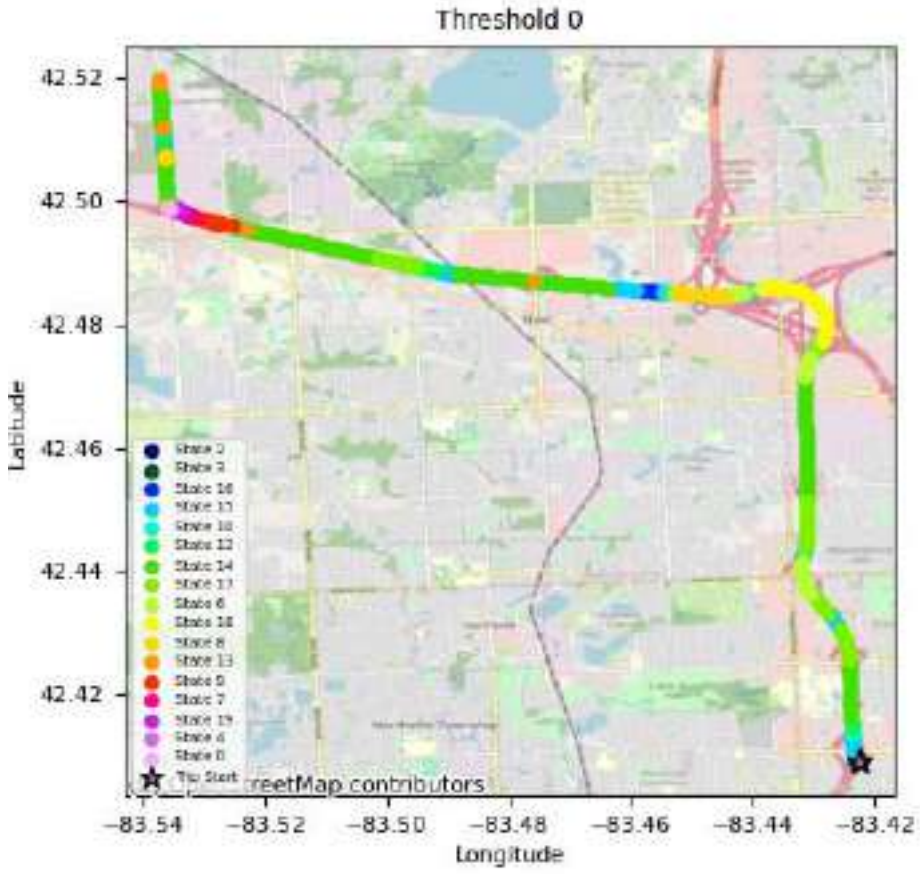
In both studies, the rHDP-HSMM outperforms the HDP-HSMM in terms of estimation and consistency. This paper concludes that the rHDP-HSMM is worth applying to datasets where an HDP prior may be generating redundant states. Further inspection as to how to select the threshold may be required, however it is clear that the merging procedure within the model is still able to learn consistent and highly interpretable states for the study of driving maneuvers.

7 Highlights

? A robust HDP-HSMM is proposed which produces more consistent results than the HDP-HSMM ? An algorithm is described as to combat the inconsistency issues that arise from using an HDP prior ? A simulation study is performed to show the impact of the proposed robust HDP-HSMM versus the basic HDP-HSMM in terms of parameter convergence and data segmentation ? Real kinematic data is used to further compare robust HDP-HSMM and the basic HDP-HSMM in terms of learned maneuver patterns. [1] [2]

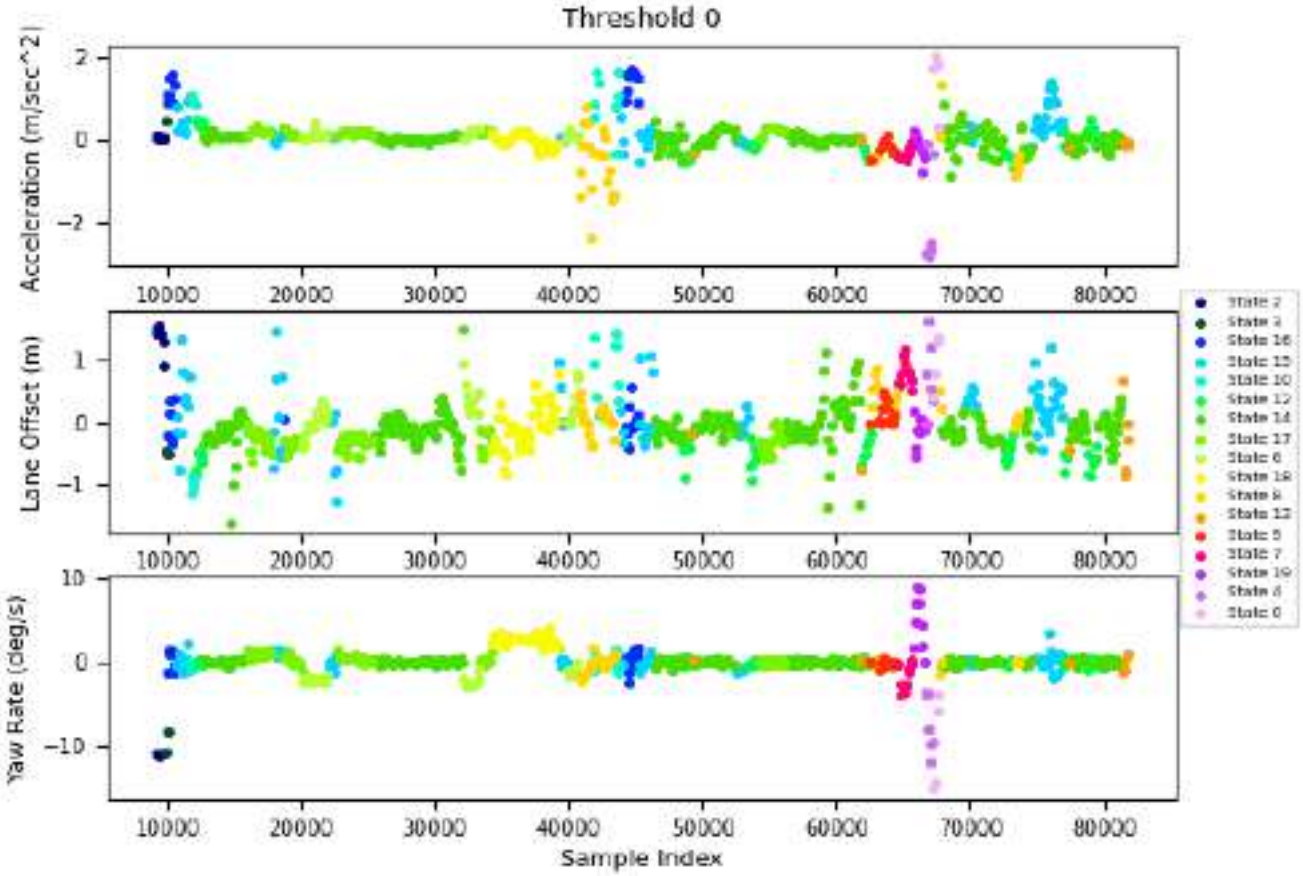
[1] © 2023 Global Journ als als

[2] © 2023 Global Journals



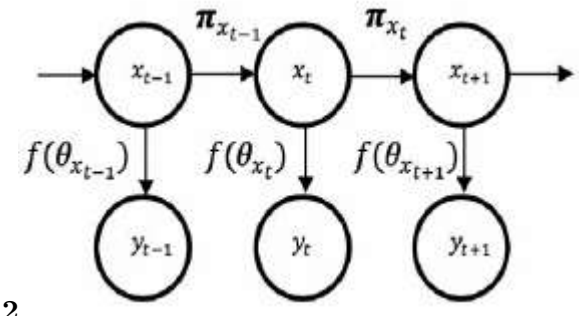
1

Figure 1: Figure 1 :



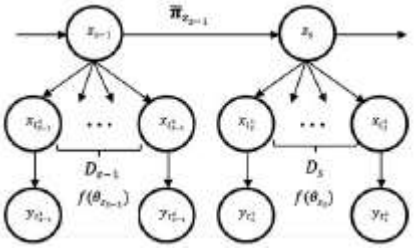
2

Figure 2: Figure 2 :



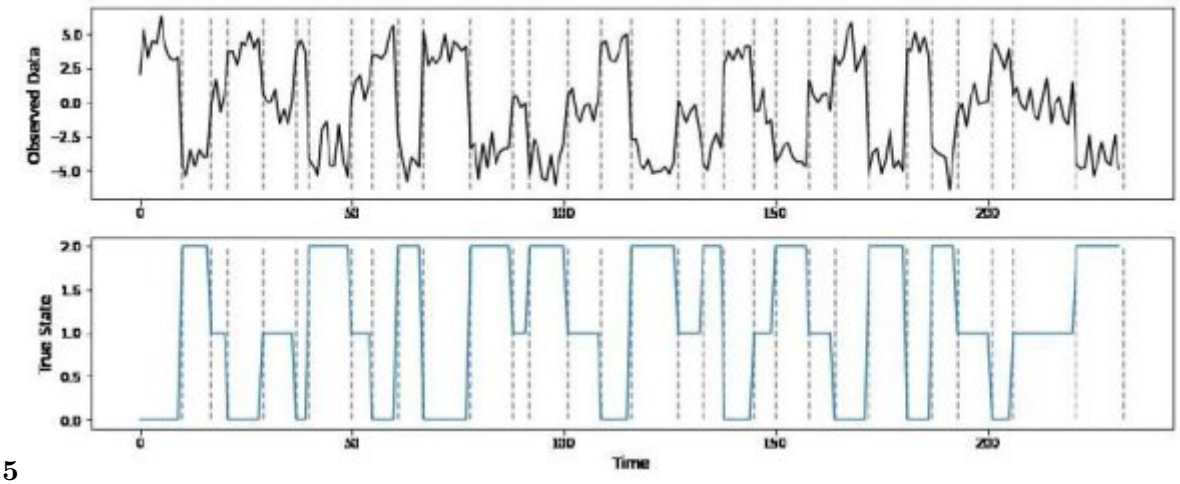
2

Figure 3: ? 2 ?



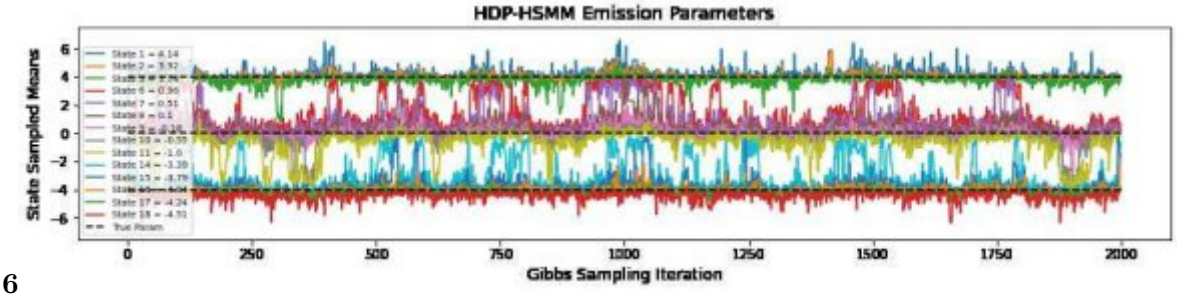
4

Figure 4: Figure 4 :



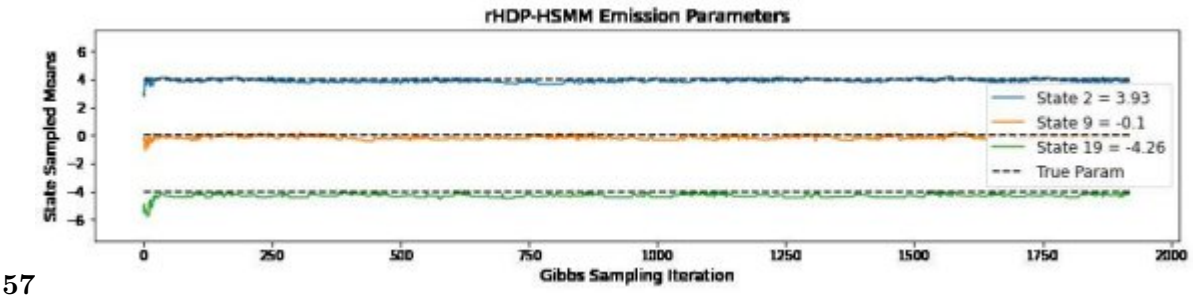
5

Figure 5: Figure 5 :



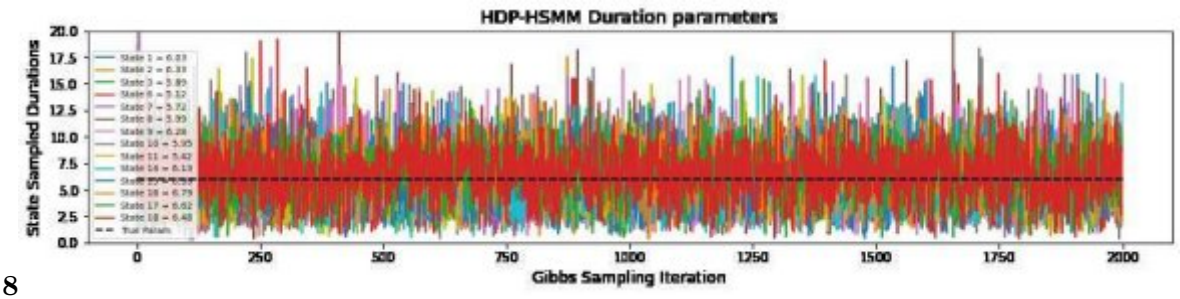
6

Figure 6: Figure 6 :



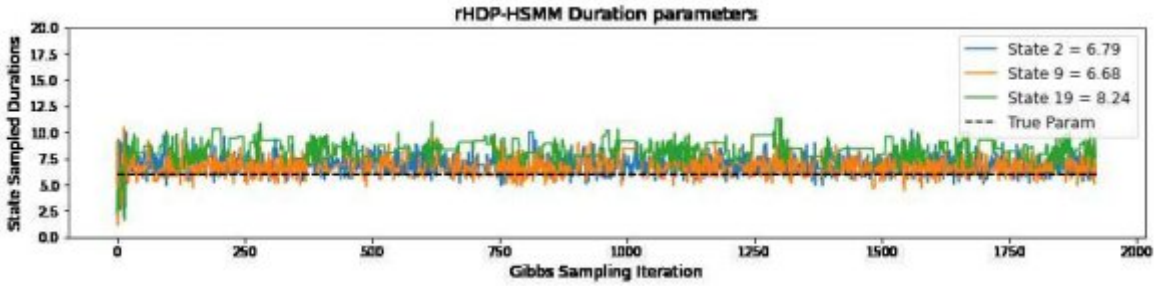
57

Figure 7: 5 Figure 7 :



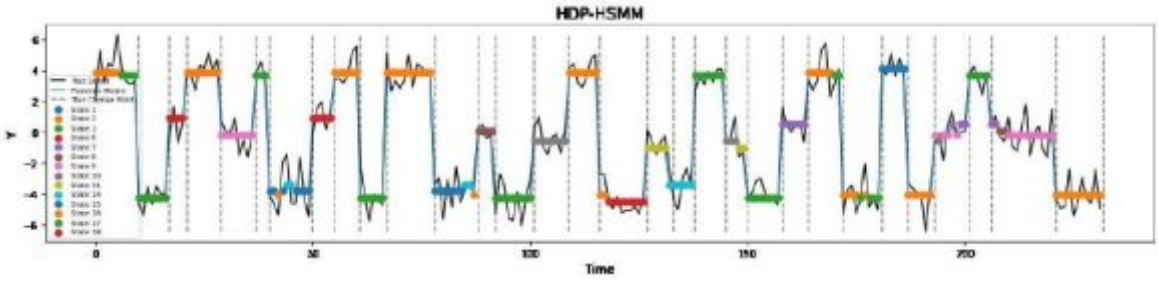
8

Figure 8: 8 .



8

Figure 9: Figure 8 :



9

Figure 10: Figure 9 :

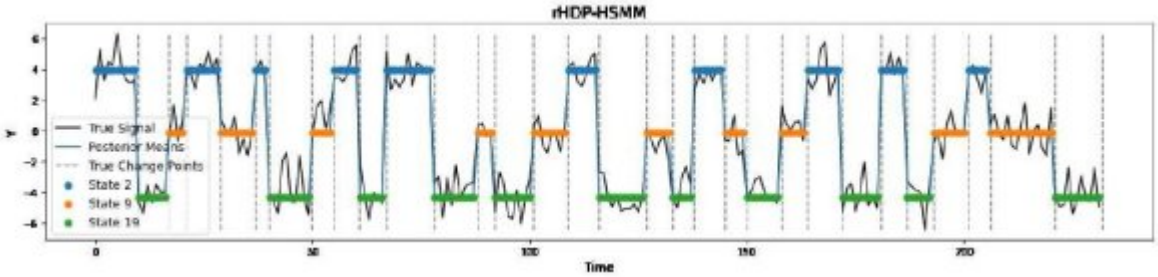


Figure 11: Figure

1

of Various HMM-based Models Versus Our Proposed Robust HDP-HSMM (rHDP-HSMM) III. Problem Formulation a) Data Description

Figure 12: Table 1 :

( ) Volume Xx XIII Issue I V ersion I B
Journal of Researches in Engineering
Global

such that the elements within the s-th segment, denoted by y t 1 s :t 2 s , are independent and identically distributed (i.i.d.) for a state duration of D s ? 1, 2, . . . , S.

Figure 13:

2

		Emission	Duration		Transition	
Distribution		Normal	Poisson		N/A	
Parameter(s)	Mean	Variance	Rate	State 1	State 2	State 3
State 1	4	1	6	0	0.3	0.7
State 2	0	1	6	0.8	0	0.2
State 3	-4	1	6	0.4	0.6	0

Figure 14: Table 2 :

3

Figure 15: Table 3 :

[Teh ()] , Y W Teh . 2010. (Dirichlet process.)

[Sethuraman ()] 'A constructive definition of dirichlet priors'. J Sethuraman . *Statistica sinica* 1994. p. .

[Jordan and Teh ()] 'A gentle introduction to the dirichlet process, the beta process, and bayesian nonparametrics'. M I Jordan , Y W Teh . *Dept. Statistics* 2014. UC Berkeley.

[Gelman and Rubin ()] 'A single series from the gibbs sampler provides a false sense of security'. A Gelman , D B Rubin . *Bayesian statistics* 1992. 4 p. .

[Fox et al. ()] 'A sticky hdp-hmm with application to speaker diarization'. E B Fox , E B Sudderth , M I Jordan , A S Willsky . *The Annals of Applied Statistics* 2011. p. .

[Gassiat and Rousseau ()] 'About the posterior distribution in hidden markov models with unknown number of states'. E Gassiat , J Rousseau . *Bernoulli* 2014. 20 p. .

[Fox et al. ()] 'An hdp-hmm for systems with state persistence'. E B Fox , E B Sudderth , M I Jordan , A S Willsky . *Proceedings of the 25th international conference on Machine learning*, (the 25th international conference on Machine learning) 2008. ACM. p. .

[Rabiner and Juang ()] 'An introduction to hidden markov models'. L Rabiner , B Juang . *ieee assp magazine* 1986. 3 p. .

[Johnson and Willsky ()] 'Bayesian nonparametric hidden semi-markov models'. M J Johnson , A S Willsky . *Journal of Machine Learning Research* 2013. 14 p. .

[Jordan ()] 'Bayesian nonparametric learning: Expressive priors for intelligent systems, Heuristics, probability and causality: A tribute to'. M I Jordan . *Judea Pearl* 2010. 11 p. .

[Beal et al. ()] M J Beal , Z Ghahramani , C E Rasmussen . *Advances in neural information processing systems*, 2002. p. . (The infinite hidden markov model)

[Wang et al. ()] 'Driving style analysis using primitive driving patterns with bayesian nonparametric approaches'. W Wang , J Xi , D Zhao . *IEEE Transactions on Intelligent Transportation Systems* 2018. 20 p. .

[Martinez et al. ()] 'Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey'. C M Martinez , M Heucke , F.-Y Wang , B Gao , D Cao . *IEEE Transactions on Intelligent Transportation Systems* 2017. 19 p. .

[Rahman et al. ()] 'Evaluation of driver car following behavior models for cooperative adaptive cruise control systems'. M Rahman , M Chowdhury , K Dey , M R Islam , T Khan . *Transportation Research Record* 2017. 2622 p. .

[Gelman et al. ()] A Gelman , J B Carlin , H S Stern , D B Dunson , A Vehtari , D B Rubin . *Bayesian data analysis*, 2013. CRC press.

[Guha et al. ()] A Guha , N Ho , X Nguyen . arXiv:1901.05078. *On posterior contraction of parameters and interpretability in bayesian mixture modeling*, 2019. (arXiv preprint)

[Teh et al. ()] 'Hierarchical dirichlet processes'. Y W Teh , M I Jordan , M J Beal , D M Blei . 10.1198/016214506000000302. doi:10.1198/ 016214506000000302. https://doi.org/10.1198/016214506000000302 *Journal of the American Statistical Association* 2006. 101 p. .

[Nodine et al. ()] *Integrated vehicle-based safety systems (IVBSS) light vehicle field operational test independent evaluation*, E Nodine , A Lam , S Stevens , M Razo , W Najm . 2011. United States. National Highway Traffic Safety Administration. (Technical Report)

[Johnson and Willsky ()] M J Johnson , A S Willsky . arXiv:1208.6537. *Dirichlet posterior sampling with truncated multinomial likelihoods*, 2012. (arXiv preprint)

[Symul and Holmes ()] 'Labeling self-tracked menstrual health records with semi-markov models'. L Symul , S P Holmes . *medRxiv* 2021.

[Jasra et al. ()] 'Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling'. A Jasra , C C Holmes , D A Stephens . *Statistical Science* 2005. 20 p. .

[Miller and Harrison ()] J W Miller , M T Harrison . arXiv:1301.2708. *A simple example of dirichlet process mixture inconsistency for the number of components*, 2013. (arXiv preprint)

[Kamson et al. ()] 'Multi-centroid diastolic duration distribution based hsmm for heart sound segmentation'. A P Kamson , L Sharma , S Dandapat . *Biomedical signal processing and control* 2019. 48 p. .

[Wooters and Huijbregts ()] *Multimodal Technologies for Perception of Humans*, C Wooters , M Huijbregts . 2007. Springer. p. . (The icsi rt07s speaker diarization system)

[Murphy ()] K P Murphy . *Hidden semi-markov models (hsmms)*, 2002.

[Sperrin et al. ()] 'Probabilistic relabelling strategies for the label switching problem in bayesian mixture models'. M Sperrin , T Jaki , E Wit . *Statistics and Computing* 2010. 20 p. .

7 HIGHLIGHTS

[Kolter and Johnson ()] 'Redd: A public data set for energy disaggregation research'. J Z Kolter , M J Johnson . *Workshop on data mining applications in sustainability (SIGKDD)*, (San Diego, CA) 2011. 25 p. .

[Sagberg et al. ()] F Sagberg , G F Selpi , J Bianchi Piccinini , Engstr Öm . *A review of research on driving styles and road safety*, 2015. 57 p. .

[Di Cairano et al. ()] 'Stochastic mpc with learning for driver-predictive vehicle control and its application to hev energy management'. S Di Cairano , D Bernardini , A Bemporad , I V Kolmanovsky . *IEEE Transactions on Control Systems Technology* 2013. 22 p. .

[Shumway and Stoffer ()] *Time series analysis and its applications: with R examples*, R H Shumway , D S Stoffer . 2017. Springer.

[Zhao et al. ()] 'Trafficnet: An open naturalistic driving scenario library'. D Zhao , Y Guo , Y J Jia . *IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017. 2017. IEEE. p. .

[Kim et al. ()] 'Unsupervised disaggregation of low frequency power measurements'. H Kim , M Marwah , M Arlitt , G Lyon , J Han . *Proceedings of the 2011 SIAM international conference on data mining*, (the 2011 SIAM international conference on data mining) 2011. p. .