



Article

# Probabilistic Modelling for Unsupervised Analysis of Human Behaviour in Smart Cities

Yazan Qarout <sup>1,\*</sup>, Yordan P. Raykov <sup>1</sup> and Max A. Little <sup>2</sup><sup>1</sup> Department of Mathematics, Aston University, Birmingham B4 7ET, UK; yordan.raykov@gmail.com<sup>2</sup> Department of Computer Science, University of Birmingham, Birmingham B15 2TT, UK; maxl@mit.edu

\* Correspondence: qarouty@aston.ac.uk

Received: 10 December 2019; Accepted: 27 January 2020; Published: 31 January 2020



**Abstract:** The growth of urban areas in recent years has motivated a large amount of new sensor applications in smart cities. At the centre of many new applications stands the goal of gaining insights into human activity. Scalable monitoring of urban environments can facilitate better informed city planning, efficient security, regular transport and commerce. A large part of monitoring capabilities have already been deployed; however, most rely on expensive motion imagery and privacy invading video cameras. It is possible to use a low-cost sensor alternative, which enables deep understanding of population behaviour such as the Global Positioning System (GPS) data. However, the automated analysis of such low dimensional sensor data, requires new flexible and structured techniques that can describe the generative distribution and time dynamics of the observation data, while accounting for external contextual influences such as time of day or the difference between weekend/weekday trends. In this paper, we propose a novel time series analysis technique that allows for multiple different transition matrices depending on the data's contextual realisations all following shared adaptive observational models that govern the global distribution of the data given a latent sequence. The proposed approach, which we name Adaptive Input Hidden Markov model (AI-HMM) is tested on two datasets from different sensor types: GPS trajectories of taxis and derived vehicle counts in populated areas. We demonstrate that our model can group different categories of behavioural trends and identify time specific anomalies.

**Keywords:** time series; probabilistic modelling; trajectory analysis; smart city

## 1. Introduction

The rising population in modern cities introduces many challenges in urban city planning including problems associated with improving the inhabitants' quality of life and security. The utilisation of sensors and communication technology gives rise to the concept of Smart Cities and enables the adequate study of specific focus problems in urban development. Among the many examples are applications on structural health monitoring [1], where material conditions of civil infrastructure are monitored and automatically analysed to ensure population safety; smart waste management [2] to improve service provisioning by the utilisation of smart garbage bins and Internet of Things (IoT) technology; smart lighting solutions [3] for cost and greenhouse emission reduction by the installation of sensors and weather analytics on city streets; and, in the manufacturing industry [4], combining IOT with machine learning. With the growing interest and importance of sensor application in smart cities, there has also been a rise in the interest in human movement behavioural characteristics from sensors to identify movement trends and improve the understanding of traffic. Examples from recent literature on applications of human behavioural understanding include De Marsico et al. [5] user gait identification for automatic activation of controlled entry spaces, and Paul et al. [6], Torre-Bastida et al. [7] big data and social IOT was utilised for real time definition

of human behaviours. For traffic analysis, Bhatti et al. [8] introduced an IOT accident detection and notification system that relies on multiple, commonly available smart-phone sensors.

The analysis of car traffic can aid the understanding of population dynamics in smart cities. This, on the other hand, leads to better capabilities for congestion control and accident detection Ahmad et al. [9]. For example, Kanungo et al. [10] studied traffic light management using video surveillance analysis, whereas Ozkurt and Camci [11] use surveillance footage to estimate density patterns. This focus on video surveillance data has prompted research on camera placement for greater area coverage [12] to reduce the high cost accumulation of such sensors. However, the high costs and privacy concerns related to CCTV data are a key driver of the search of alternative sensor resources: such as density counts, noise levels or automated geolocation data (such as GPS).

Despite the prior work on behaviour understanding from GPS [13–15], a lot of the GPS data analysis remains challenging due to noise, irregular sampling, heterogeneity and frequent interruption of data collection as environmental factors cause missing data [16]. To address this Ellam et al. [17], a recently proposed parameter estimation framework for spatial interaction models was demonstrated to simulate accurately the flow of customers. In scenarios where large amounts of individual-level data is available, we can also try to approach the problem reversely and first search for recurring patterns from large databases of movement in cities. Frequently sampled monitoring data such as the GPS data can be highly multimodal. To account for this, Witayangkurn et al. [18] proposed the use of a Hidden Markov Model (HMM) to detect anomalies from large scale GPS data. Nonetheless, this approach does not consider the time self-dependence of the observation sequence and requires that the model complexity be predefined which may bias the unsupervised approach to the analysis. The geospatial relationship between the journeys of individuals is often dependent on various contextual factors such as differences between weekend/weekday behaviour, peak/off-peak behaviour and individual idiosyncrasies.

To address some of these challenges, we design an exploratory nonparametric probabilistic modelling approach suitable for this type of data. We deploy a flexible HMM with augmented latent space to account for various types of contextual information. This contextual information can include difference between peak/off-peak time driving, weekday/weekend differences and person specific manoeuvrability decisions. To achieve this, we propose an intuitive Adaptive Input Hidden Markov Model (AI-HMM). Building on the theory of the Vector Autoregressive Hidden Markov Model (VAR-HMM) [19] to allow for autoregressive self-dependence between observations, we introduce a new discrete and independent semi-Markov variable which acts as a parent to the state indicator variable in the graphical model to represent environmental factors that influence movement behaviour. For example, in our case study in Section 5 on GPS data, we assume that the growing complexity of human trips, as more trajectories are observed, can be well captured with Bayesian nonparametric structure and that environmental factors such as time of day (e.g., peak or off-peak) may also influence the modality of the behaviour and therefore should be considered in the design of the generative model structure. We show that we are able to learn common patterns occurring across different contextual realisations, in addition to their specific generative distributions. To summarise, the contributions of this paper are as follows.

- We develop an extension of the ubiquitous Hidden Markov model, which can robustly segment different behaviours in monitoring applications while accounting for present contextual factors.
- We have derived a principled framework for inference in the developed AI-HMM approach.
- We have shown the particularly suitability of the nonparametric switching autoregressive model for analysis of GPS and traffic count data.

This paper will continue as follows. Section 2 discusses related studies, Section 3 summarises the background of the used techniques, Section 4 explains the details of the proposed framework, Sections 5 and 6 summarise the algorithm experiments and its results and Section 7 concludes the report.

## 2. Related Work

Human behaviour profiling and automated analysis has great potential for both security and retail commercial applications in smart city designs. However, labelled sensor data of such nature is usually scarce and unsupervised clustering or direct time series methods must be used [20–22]. One commonly used pipeline is the online segmentation of crowd data from video feeds. For example, Mehran et al. [14] proposed a novel Social Force model for identifying abnormal behaviours in crowds. The approach assumes that located moving particles on sequences of frames are individuals and uses them to estimate their interaction forces throughout time and estimate the “normal” behaviour of the crowd. Then, a bag-of-words classifier is trained to separate normal and abnormal crowd behaviours. Mahadevan et al. [23] also proposed a supervised anomaly detection method based on crowd videos. However, [23] evaluated their methods on a significantly smaller database of videos. Rodriguez et al. [24] developed a partially Bayesian approach to the same problem which allows for robust analysis of crowd videos completely unseen during training. While civic value can be demonstrated through the analysis of videos at crowded locations, such data provides only localised snapshots of the movement within the city. It is available only for a small set of chosen locations and requires a large amount of labelled location specific training data. Video feeds have also been used extensively for activity recognition applications [20,25,26]; however, such techniques are likely far from being applicable on larger spatial scale across a city.

Continuous location data can be used to study a more diverse geographic perspective on personal mobility and daily activity patterns. Jiang et al. [21] investigated how to cluster daily patterns of activity in the city using both GPS data from individuals as well as additional survey data concerning their true activities. The study used the Chicago Regional Household Travel Inventory (CRHTI) dataset [27], which is one of the few data sources with fine labelled information in addition to GPS data. GPS trajectory data was not explicitly modelled but only used to identify different location areas for the participants. Jiang et al. [21] first constructed a calendar type array indicating which of nine activities (work, home, school, etc.) occurred. Principal Component Analysis (PCA) was then applied followed by K-means clustering to infer patterns of daily activity across different regions of the city. The inferred patterns were used to reveal socio-demographic information about the region. The results of the study are informative on human behaviour; however, if GPS information is included in the analysis, the complexity of the data may affect the effectiveness of the proposed framework.

Shih et al. [22] modelled different mobility patterns based on GPS data from the Geolife dataset [28] to classify trips of people diagnosed with Alzheimer’s disease. The basis for this is the common assumption that a symptom of Alzheimer’s patients is spatio-temporal confusion. The abnormal trajectories were predicted based on thresholding the similarity to the training set. Although the results showed up to 97% accuracy rate in detecting abnormal mobility patterns, there was no ground truth data, and explicitly, no real information on the Alzheimer state of each subject. Instead, external sequences were added to the data of some individuals to represent Alzheimer’s patients. This can add clear biases to the analysis and may indicate that the accuracy of the technique on more representative data will not mimic that of the presented study. More recently, Yao et al. [15] also proposed a novel behaviour recognition framework for GPS data, demonstrated on trajectories of ships. They used a Long-Short Term Memory (LSTM) autoencoder to infer the latent structure in features which were extracted from sliding windows of the trajectory time series. By passing the dataset through a trained network, the inferred embeddings were extracted and stored for each input time series. These embeddings were then clustered with K-means to identify different behavioural patterns in ship data. However, trajectories of ship movement contain significantly smaller variations across patterns and a constrained space of behaviours which can be explored. The LSTM autoencoder is likely to require multiple different behaviour sequences in order to group the trajectories in the latent space. Also, when focused on segmenting anomalous events, K-means is known to lead to misleading clustering [29].

### 3. Dynamic Time Series Modelling

#### 3.1. Vector Autoregressive Model (VAR)

The Vector Autoregressive (VAR) model [30] is a widely used linear time series model for dynamic multidimensional data. An order  $p$  VAR assumes each observation in a sequence can be modelled as a linear combination of  $p$  previous points in the sequence and a stochastic term. The VAR model is also a parametric model of the spectral density. Unlike Fourier-based methods which require a windowing strategy, VARs are defined in the time domain and we can use  $p$  to control the number of spikes we assume in the power spectrum of the input. Let  $Y = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T)$  be a stationary time series where  $\mathbf{y}_t \in \mathbb{R}^D = (y_t^{(1)}, \dots, y_t^{(D)})^T$ , a VAR( $p$ ) model will have the form of

$$\mathbf{y}_t = \Theta_1 \mathbf{y}_{t-1} + \dots + \Theta_p \mathbf{y}_{t-p} + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \quad (1)$$

where  $\Theta_i$  is a  $(D \times D)$  matrix of weights,  $\boldsymbol{\mu}$  is the mean of the sequence  $Y$  and  $\boldsymbol{\epsilon}_t$  is a  $(D \times 1)$  vector of white noise with zero mean and covariance matrix  $\Sigma$ . Using the matrix notation  $Y_{p+1:T} = (\mathbf{y}_{p+1}, \mathbf{y}_{p+2}, \dots, \mathbf{y}_T)$  we can then write the VAR in the following form,

$$Y_{p+1:T} = \Theta \bar{Y} + E, \quad (2)$$

with  $\Theta = (\boldsymbol{\mu}, \Theta_1, \dots, \Theta_p)$ ,  $\bar{Y} = (\mathbf{1}, Y_{p:T-1}^T, \dots, Y_{1:T-p}^T)^T$ ,  $\mathbf{1}$  being a vector of ones with the same length as  $Y_{p+1:T}$  and  $E = (\boldsymbol{\epsilon}_{p+1}, \dots, \boldsymbol{\epsilon}_T)$ . Using this formulation, the parameters of a VAR model may be estimated in few main ways depending on the adopted formalism: via maximum likelihood estimation, using the Yule–Walker equations and via Bayesian inference. Inference in Bayesian VAR models is practically identical to statistical inference in Bayesian linear regression. Assuming, for practical convenience, a conjugate prior over the VAR parameters, the joint prior over the parameters  $\Theta$  and the covariance matrix  $\Sigma$  is a coupled Matrix-Normal Inverse-Wishart ( $\mathcal{MN}\mathcal{IW}$ ) distribution, implied by the Gaussian likelihood of the outputs.

$$\begin{aligned} \Sigma &\sim \mathcal{IW}(n_0, S_0), \\ \Theta | \Sigma &\sim \mathcal{MN}(M, \Sigma, K), \\ Y_{p+1:T} | \bar{Y}, \Theta, \Sigma &\sim \mathcal{MN}(\Theta \bar{Y}, \Sigma, \mathbf{I}), \end{aligned} \quad (3)$$

where  $n_0$  is the number of degrees of freedom;  $S$  is the scale matrix;  $M, \Sigma$  and  $S_0$  are the mean, covariance of the rows and covariance of the columns of parameter  $\Theta$ , respectively; and  $\mathbf{I}$  is the identity matrix. The posterior of the VAR coefficients is then

$$p(\Theta | Y, \Sigma) \propto p(Y | \bar{Y}, \Theta, \Sigma) p(\Theta | \Sigma). \quad (4)$$

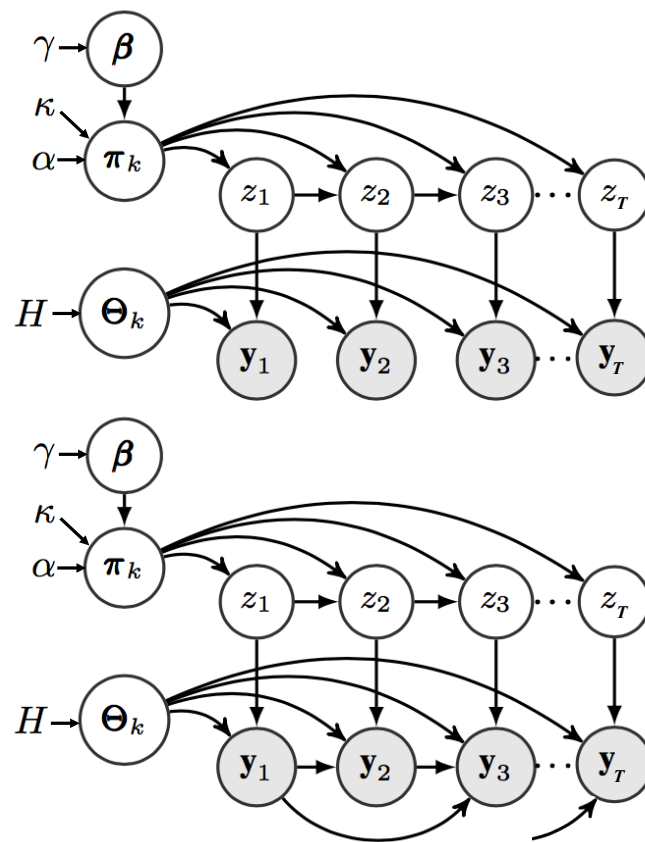
To compute posterior of the VAR noise covariance, it is more efficient to use the marginal likelihood, instead giving

$$\begin{aligned} p(\Sigma | Y) &\propto p(Y | \Sigma) p(\Sigma), \\ p(Y | \Sigma) &\propto \int_{\Theta} p(Y | \bar{Y}, \Theta, \Sigma) p(\Theta | \Sigma) d\Theta. \end{aligned} \quad (5)$$

#### 3.2. Autoregressive Infinite Hidden Markov Model (AR-iHMM)

The infinite Hidden Markov Model (iHMM) also known as the Hierarchical Dirichlet Process HMM (HDP-HMM) is a Bayesian nonparametric approach to the HMM first proposed by Beal et al. [31] then formalised by Teh et al. [32]. The nonparametric formalisation tackled the limitations of the traditional HMM of needing to pre-specify the structure of the model allowing it to automatically infer in adaptive manner the complexity depending on the amount of data observed. Building on the iHMM, Fox [33] included linear AR dependence between the observation sequence to

form the AR-iHMM (or also known as HDP-AR-HMM). Figure 1 depicts the Probabilistic Graphical Model (PGM) representation of the iHMM and the AR-iHMM.



**Figure 1.** iHMM (top) and AR-iHMM (bottom) probabilistic graphical models. Mixing Parameters  $\beta$  are used to sample the transition distribution  $\pi_k$ , which the state indicators  $z$  are sampled from. The observations  $y_t$  are generated from functions with parameters  $\Theta_{z_t}$  which are in turn sampled from base distribution  $H$ . The AR-iHMM differs such that there are autoregressive dependence between observations.

Let  $\mathbf{z} = (z_1, \dots, z_T)$  denote the state indicator variables and  $\pi_k$  be the state specific transition distribution for state  $k$  where  $z_t \sim \text{Categorical}(\pi_{z_{t-1}})$ . The observations  $y_t$  are conditionally independent given  $z_t$  and  $\tilde{Y}_t = (y_{t-1}^T, \dots, y_{t-p}^T)^T$ ; each observation is modelled with emission distributions  $y_t|z_t, \tilde{Y}_t \sim f(\Theta_{z_t}, \tilde{Y}_t)$  where  $\Theta_{z_t}$  are the emission parameters for state indicated by  $z_t$ . The complete data likelihood can then be written as

$$\begin{aligned}
 p(\mathbf{z}, Y) &= \prod_{t=p+1}^T p(z_t|z_{t-1})p(y_t|z_t, \tilde{Y}_t), \\
 p(\mathbf{z}, Y) &= \prod_{t=p+1}^T \sum_{k=1}^{K^+} \pi_{z_{t-1},k} f(y_t; \Theta_k, \tilde{Y}_t).
 \end{aligned}
 \tag{6}$$

using  $K^+$  to denote the inferred number of represented states.

The iHMM defines a set of random probability measures  $G_k$  for each state representing the dynamic observations  $Y$  sampled from a Dirichlet Process (DP) with a global probability measure  $G_0$  and concentration parameter  $\alpha$ . The base probability measure itself is DP distributed with concentration parameter  $\gamma$  and a base probability measure  $H$

$$\begin{aligned} G_0 | \gamma, H &\sim DP(\gamma, H), \\ G_k | \alpha, G_0 &\sim DP(\alpha, G_0), \end{aligned} \quad (7)$$

where measures  $G_k$  are conditionally independent given  $G_0$ . Intuitively, Equation (7) means that random distributions  $G_k$  vary around  $G_0$  with variability governed by  $\gamma$ , whereas  $G_0$  varies around the base distribution  $H$  with variability  $\alpha$ . Including explicit hyperparameter controlling the self-transitions probability in the AR-iHMM [31,33] we can summarize the full AR-iHMM as follows,

$$\begin{aligned} \beta | \gamma, H &\sim GEM(\gamma), \\ \pi_k | \alpha, \beta &\sim DP\left(\alpha + \kappa, \frac{\alpha\beta + \kappa\delta_k}{\alpha + \kappa}\right), \\ \Theta_k | H &\sim H, \\ z_t | z_{t-1}, (\pi_k)_{k=1}^{\infty} &\sim \pi_{z_{t-1}}, \\ \mathbf{y}_t | z_t, \tilde{Y}_t, (\Theta_k)_{k=1}^{\infty} &\sim \mathbf{f}(\Theta_{z_t}, \tilde{Y}_t), \end{aligned} \quad (8)$$

where  $\beta = (\beta_k)_{k=1}^{\infty}$  is both a mixing prior for the top level DP and a base measure for the lower level DP, GEM (which stands for Griffiths, Engen and McCloskey [34]) represents the stick breaking construction and  $(\cdot)_{k=1}^{\infty}$  represents an infinite set.

#### 4. Adaptive Input Infinite Hidden Markov Model (AI-iHMM)

##### 4.1. Model Specification

AR-iHMMs are powerful models that are capable of modelling complex dynamic data effectively. However, as a generative model, they assume that the state indicator  $z_t$  is only dependent on the state indicator at the previous time step  $z_{t-1}$ . The model also assumes that the full time series share the same upper layer mixing prior  $\beta$  and therefore the same set of transition distributions  $\Pi = (\pi_1^T, \pi_2^T, \dots, \pi_K^T)^T$ . In some real-world problems, there are independent and discrete features  $\tau = (\tau_1, \tau_2, \dots, \tau_T)$  that affect the model's state indicators and transition distributions. The semi-Markov variable  $\tau$  can represent certain events or periods for the dataset where data generated during any fixed realisation of  $\tau_t = v$  can share the same mixing prior  $\beta$  and same set of transition distributions  $\Pi$ , but have different distributions for different realisations  $v$ . For example, in the case of human movement analysis, different periods of the day  $\tau_t \in \{1, 2\}$ , such as off-peak and peak time, respectively, can have different behavioural traits with separate parameters  $\{\Pi_1, \Pi_2\}$  and  $\{\beta_1, \beta_2\}$ . Although the desired separation can be achieved by modelling the data generated for every realisation separately with different models, it will likely result in assigning exclusive states and state emission parameters  $\Theta_k$  for each model. This will raise the problem of identifying a suitable method to measure state overlap across different models, specifically in cases like human movement behaviour where any small change in the states parameters may be important.

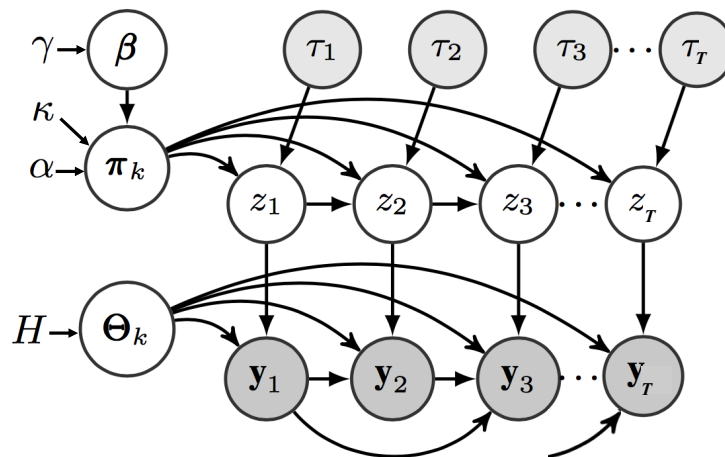
To avoid this problem, we propose a novel Adaptive Input infinite Hidden Markov Model (AI-iHMM) where an additional DP layer is added to the models hierarchy allowing for the generation of a different base measure  $G_0^{(v)}$  for each unique value of  $\tau$  while still sharing the same upper level base distribution  $H$  for the sampling of the model parameters  $\Theta_k$ . Figure 2 depicts the PGM of the proposed model with the respective joint probability distribution and formalisation.

$$P(\boldsymbol{\tau}, \mathbf{z}, Y) = \prod_{t=p+1}^T P(\tau_t) P(z_t | z_{t-1}, \tau_t) P(\mathbf{y}_t | z_t, \tilde{Y}_t), \quad (9)$$



$$\begin{aligned}
 Q|\psi, H &\sim DP(\psi, H), \\
 G_0^{(v)}|\gamma, H &\sim DP(\gamma, Q), \\
 G_k^{(v)}|\alpha, G_0^{(v)} &\sim DP(\alpha, G_0^{(v)}),
 \end{aligned}
 \tag{10}$$

where  $H$  is the base probability measure for the upper layer DP with concentration parameter  $\psi$ ,  $Q$  is the master probability measure for the middle layer DP associated with the possible values of  $\tau$  with concentration parameter  $\gamma$ ,  $G_0^{(v)}$  is the global probability measure for the lower layer DP associated with  $\tau_t = v$  with shared concentration parameter  $\alpha$  across all values of  $v$ , and  $G_j^{(v)}$  is the random probability measure for state  $j$  when  $\tau_t = v$ .



**Figure 2.** PGM of the AI-HMM structure. Observation  $y_t$  being dependent on the indicator variable  $z_t$  and  $p$  previous observations, whereas  $z_t$  is dependent on  $z_{t-1}$  in addition to the semi-Markov feature  $\tau_t$ . Parameters  $\beta$  are used to sample the transition distribution  $\pi_k$  which the state indicators  $z$  are sampled from. The observations  $y_t$  are generated from functions with parameters  $\Theta_{z_t}$  which are in turn sampled from base distribution  $H$ .

The possible values of  $\tau$  can be inferred adaptively as will be described in Section 4.2 by placing an appropriate prior on them (for example the DP would be a conjugate choice) and estimate them using modified MCMC inference. This is useful when it is known that there is an underlying variable affecting  $z$  and the values for each time instance are unknown. For more intuitive analysis and better understanding of the effect the contextual category has on the transition dynamics of the time series data,  $\tau$  can also be set as a fixed categorical input to the model. This allows for more precise inference of the transition dynamics associated with each context  $v$  and for their custom selection based on the problem of interest. Depending on the complexity of the data, the estimation of input  $\tau$  in addition to the estimation of the remaining model parameters may be result in numerous local minima if no prior knowledge is known about the model. Therefore, it would be advantageous to train with partial knowledge of  $\tau$  or pre-train with a fixed input when inferring it adaptively. In the general case of a completely adaptive model, using the stick breaking construction, the model may be written as

$$\begin{aligned}
\boldsymbol{\lambda}|\psi, H &\sim GEM(\psi), \\
\boldsymbol{\beta}_v|\gamma, \boldsymbol{\lambda} &\sim DP(\gamma, \boldsymbol{\lambda}), \\
\boldsymbol{\pi}_k^{(v)}|\alpha, \boldsymbol{\beta}_v &\sim DP\left(\alpha + \kappa, \frac{\alpha\boldsymbol{\beta}_v + \kappa\delta_k}{\alpha + \kappa}\right), \\
\epsilon_{j,i}^{(v)} &\sim \boldsymbol{\pi}_{j,i}^{(v)} Dir\left(\frac{\eta}{V}\right) \lambda_{z_{t-1}}, \\
\Theta_k|H &\sim H, \\
\tau_t|z_t, z_{t-1}, \boldsymbol{\epsilon} &\sim \boldsymbol{\epsilon}_{z_{t-1}, z_t}, \\
z_t|z_{t-1}, \tau_t = v, \left(\boldsymbol{\pi}_k^{(v)}\right)_{k=1}^\infty &\sim \boldsymbol{\pi}_{z_{t-1}}^{(v)}, \\
\mathbf{y}_t|z_t, \bar{Y}_t, (\Theta_k)_{k=1}^\infty &\sim \mathbf{f}(\Theta_{z_t} \bar{Y}_t),
\end{aligned} \tag{11}$$

where  $\boldsymbol{\lambda}$  is the upper level mixing parameter and the middle level base measure,  $\eta$  is the prior parameter of the distribution of  $\tau$  and  $\epsilon_{j,i}^{(v)}$  is the posterior probability of  $\tau_t = v$  when  $z_t = i$  and  $z_{t-1} = j$ .

#### 4.2. Inference

Similarly to the AR-iHMM structure, the parameters of the AI-HMM can be inferred by blocked Gibbs sampling. A truncation level  $L$  is set to denote the maximum number of states expected for the time series modelling so that  $k \in \{1, 2, \dots, L\}$ . Given previously set transition distributions  $\{\Pi_v^{(n-1)}\}$ , mixing parameters  $\{\boldsymbol{\beta}_v^{(n-1)}\}$  and emission distribution parameters  $\{\Theta_k^{(n-1)}\}$  and assuming a normal distribution likelihood on the autoregressive observation sequence; the first step is to block sample the input contextual vector  $\boldsymbol{\tau}$ . Given its Markov blanket, the posterior of  $\tau_t$  is

$$p(\tau_t|z_t, z_{t-1}) \propto p(z_t|\tau_t, z_{t-1})p(\tau_t)p(z_{t-1}). \tag{12}$$

Here,  $p(z_t|\tau_t, z_{t-1})$  is simply the transition matrix elements  $\{\pi_{z_{t-1}, z_t}^{(v)}\}$  across all  $v$ ,  $p(\tau_t)$  is the prior  $Dir\left(\frac{\eta}{V}\right)$ , as  $\tau_t$  is assumed to be categorically distributed with a Dirichlet prior and  $p(z_{t-1})$  is the global prior of  $z_{t-1}$ . By design, changes in the value of  $\tau_t$  should be sparse and infrequent since they represent broad categories of transition dynamics and cover long contextual periods relative to the time series. For the case of  $z_t = z_{t-1} = k$ , it can be possible that  $\{\epsilon_{k,k}^{(v)}\}$  is similar for multiple values of  $v$  due to the closeness in value between probabilities  $\{\pi_{k,k}^{(v)}\}$ . A change in the value of  $\tau_t$  directly implies that the state for  $z_t$  has changed as well, due to the hierarchy of the model. This means that effectively  $\tau_t$  is only sampled on when  $z_t \neq z_{t-1}$ , which naturally enforces state persistence. To sample

$$\begin{aligned}
\tau_t &\sim \sum_{v=1}^V \epsilon_{z_{t-1}, z_t}^{(v)} \propto \sum_{v=1}^V \pi_{z_{t-1}, z_t}^{(v)} Dir\left(\frac{\eta}{V}\right) \lambda_{z_{t-1}} \delta(\tau_t, v) && \text{if } z_t \neq z_{t-1}; \\
\tau_t &= \tau_{t-1} && \text{if } z_t = z_{t-1}.
\end{aligned} \tag{13}$$

The next set of parameters to sample are  $\mathbf{z}$ . This is done using a variation of the forward-backward algorithm where the  $(1 \times L)$  backward message  $\mathbf{m}_{t,t-1}$  is calculated by

$$\mathbf{m}_{t,t-1} = \mathbf{m}_{t+1,t} \Pi_{\tau_t} \sum_{k=1}^L \mathcal{N}(\mathbf{y}_t; \Theta_k \bar{Y}_t, \Sigma_k). \tag{14}$$

The forward message  $\mathbf{f}(\mathbf{y}_t) = (f_1(\mathbf{y}_t), \dots, f_L(\mathbf{y}_t))$  for observation  $\mathbf{y}_t$  is

$$\mathbf{f}(\mathbf{y}_t) = \mathcal{N}\left(\mathbf{y}_t; \{\Theta_k \bar{Y}_t, \Sigma_k\}_{k=1}^L\right) \times \mathbf{m}_{t+1,t}, \tag{15}$$

where the first element of  $\mathbf{m}_{t+1,t}$  is equal to 1 and  $\mathcal{N}(\mathbf{y}_t; \{\Theta_k \bar{Y}_t, \Sigma_k\}_{k=1}^L) = (\mathcal{N}(\mathbf{y}_t; \Theta_1 \bar{Y}_t, \Sigma_1), \dots, \mathcal{N}(\mathbf{y}_t; \Theta_L \bar{Y}_t, \Sigma_L))$ . The indicator variable  $z_t$  can then be sampled by



$$z_t \sim \sum_{k=1}^L f_k(\mathbf{y}_t) \pi_{z_{t-1},k}^{(\tau_t)} \delta(z_t, k). \tag{16}$$

To sample the mixing parameters  $\beta_v$ , the following auxiliary variables must be sampled all of shape  $(L \times L)$ :  $M_v$  representing the count of transitions occurring due to sampling directly from the base probability;  $\bar{M}_v$  representing the counts of new transitions unobserved previously before in the sequence  $\mathbf{z}$  and  $N^{(v)}$  representing the counts of transitioning from state  $j$  to state  $k$ . To estimate  $M_v$ , for each  $(j, k) \in \{1, \dots, L\}^v$  set  $m_{j,k}^{(v)} = 0$  and for  $n = 1, \dots, n_{j,k}^{(v)}$  sample

$$x \sim \text{Ber} \left( \frac{\alpha \beta_k^{(v)} + \kappa \delta(j, k)}{n + \alpha \beta_k^{(v)} + \kappa \delta(j, k)} \right), \tag{17}$$

then increment  $n$  and if  $x = 1$  increment  $m_{j,k}^{(v)}$ . Then, for  $j \in \{1, \dots, L\}$ , estimate  $W_v$  by

$$\omega_j^{(v)} \sim \text{Binomial} \left( m_{j,j}^{(v)}, \frac{\kappa}{\kappa + \alpha \beta_j^{(v)}} \right), \tag{18}$$

where  $\omega_j^{(v)}$  denotes of the override counts when the base probability distribution was used to draw self transitions for state  $j$ . This can in turn be used to estimate  $\bar{m}_{j,k}^{(v)}$

$$\bar{m}_{j,k}^{(v)} = \begin{cases} m_{j,k}^{(v)} & j \neq k; \\ m_{j,j}^{(v)} - \omega_j^{(v)} & j = k. \end{cases} \tag{19}$$

$\beta_v$  may now be sampled by

$$\beta_v \sim \text{Dir} \left( \bar{\mathbf{m}}_{\cdot,1}^{(v)}, \dots, \bar{\mathbf{m}}_{\cdot,L}^{(v)} \right), \tag{20}$$

where  $\bar{\mathbf{m}}_{\cdot,j}^{(v)}$  denotes the  $j^{\text{th}}$  column of  $\bar{M}_v$ . The mixing parameter may then be used to sample  $\pi_k^{(v)}$

$$\pi_k^{(v)} \sim \text{Dir} \left( \alpha \beta_1^{(v)} + n_{k,1}, \dots, \alpha \beta_k^{(v)} + \kappa + n_{k,k}, \dots, \alpha \beta_L^{(v)} + n_{k,L} \right), \tag{21}$$

where the hyperparameter  $\kappa$  is added to the parameters for state  $k$  to enforce self transitions. The final step in the iterative optimisation procedure is to sample the state specific vector autoregressive parameters  $\Theta_k$  and  $\Sigma_k$ . When using a  $\mathcal{MN}\mathcal{IW}$  prior as explained in Section 3.1, the parameters may be updated given the observations  $Y$  and state indicator variables  $\mathbf{z}$ . For more details on the derivation of the update, please see Fox [33].

$$\begin{aligned} S_{\bar{Y}\bar{Y}}^{(k)} &= \bar{Y}^{(k)} \bar{Y}^{(k)T} + K, \\ S_{Y\bar{Y}}^{(k)} &= Y^{(k)} \bar{Y}^{(k)T} + MK, \\ S_{Y Y}^{(k)} &= Y^{(k)} Y^{(k)T} + MKM^T, \\ S_{Y|\bar{Y}}^{(k)} &= S_{Y Y}^{(k)} + S_{Y\bar{Y}}^{(k)} S_{\bar{Y}\bar{Y}}^{-(k)} S_{Y\bar{Y}}^{(k)T}, \end{aligned} \tag{22}$$

where  $Y^{(k)}$  and  $\bar{Y}^{(k)}$  are the observations and lag matrix of lagged observations, respectively, that are classified to state  $k$ .  $S_{\bar{Y}\bar{Y}}^{(k)}$ ,  $S_{Y\bar{Y}}^{(k)}$ ,  $S_{Y Y}^{(k)}$  and  $S_{Y|\bar{Y}}^{(k)}$  are the posterior parameters such that

$$\begin{aligned} \Sigma_k &\sim \mathcal{IW} \left( n_0 + \sum_v c_k^{(v)}, S_{Y|\bar{Y}}^{(k)} + S_0 \right), \\ \Theta_k | \Sigma_k &\sim \mathcal{MN} \left( S_{Y\bar{Y}}^{(k)} S_{\bar{Y}\bar{Y}}^{-(k)}, \Sigma_k, S_{Y\bar{Y}}^{(k)} \right), \end{aligned} \tag{23}$$

where  $c_k^{(v)}$  is the count of the number of observations in state  $k$  through when  $\tau_t = v$ .

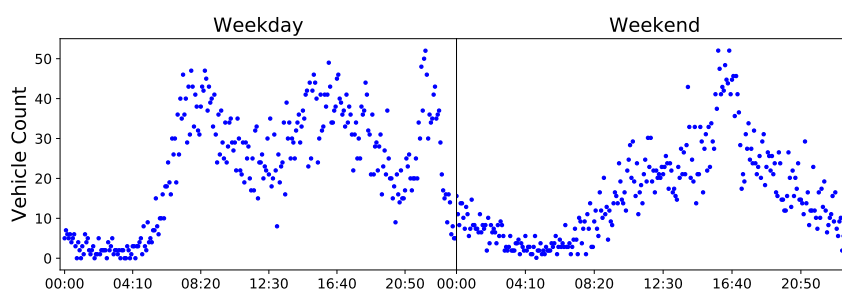
The AI-HMM is a generative model that can represent the observation data statistically with the optimised model parameters. Therefore, given a pre-trained model and initial starting points, it may be possible to sample new observations that resemble the dynamics of the training data. The parameters can even be altered manually to represent different types of behaviours and trends. This allows for generation of new custom datasets that can be used for research across different fields.

## 5. Modelling With Fixed Inputs

To demonstrate the effectiveness of the AI-HMM, we tested the model on two different datasets that represent human and traffic behaviour in smart city environments. The first was the Dodgers loop dataset [35] containing loop sensor data counting the number of vehicles that pass through the Glendale on-ramp for the 101 North Freeway in Los Angeles in 5 min (288 reading per day) intervals over a period of 25 weeks. In addition to the presence of weekday morning and afternoon traffic-peak trends in the dataset, a baseball stadium in the vicinity of the ramp allows for the observation of the post-game traffic rise. This allows us to study the performance of the model at identifying commonly repeating daily trends that may change depending on external contextual information (weekend/weekday trends) as well as studying the outlier detection capabilities with the spike in density counts due to sparsely occurring games at the nearby stadium. The dataset is simple and uniformly sampled with easily identifiable trends that can be detected by simple data visualisation. Results can be seen in Section 5.1. The second dataset was the T-drive dataset [13,36] which contains GPS trajectory data from more than 10,000 taxis collected over a period of 1 week within the city of Beijing, China. Analysing this dataset is more challenging as GPS data can be very noisy, sparse, irregularly sampled, heterogeneous and its data collection can be frequently interrupted due to environmental factors causing missing data (rarely missing at random). This type of data is complex and allows us to explore the effect of multiple contextual realisations on the observation data such as time of day and user identity. It also relates well traffic analysis, an important topic in smart city research. The AI-HMM identifies interesting patterns and results as seen in Section 5.2.

### 5.1. Dodgers Loop Sensor Dataset

The AI-HMM was used to model the complete dodgers loop sensor dataset. Figure 3 shows the count scatter for an average game day on a weekday and on a weekend. As is can be seen from the sharp peaks on the plots, games on weekdays often occur in the evening, whereas games on the weekends can occur at a similar time in the evening or, as depicted in the figure, they can occur in the afternoon.

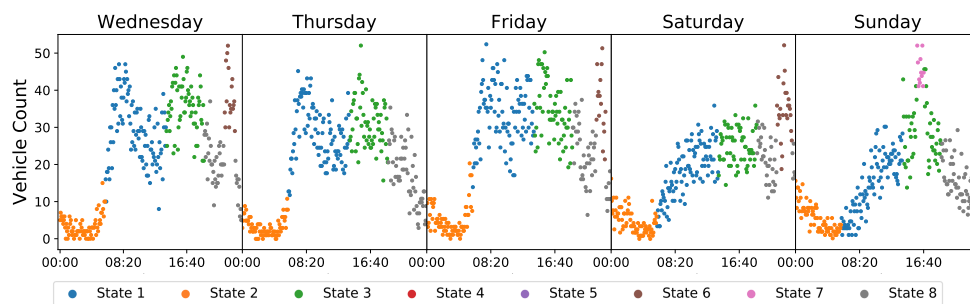


**Figure 3.** Scatter plots for vehicular density counts on an average game day on a weekday (left) and weekend (right). The first two distinct peaks on the weekday represent the morning and afternoon traffic peaks respectively, whereas the sharp peaks around 23:00 on the weekday and on 16:30 on the weekend represent traffic caused by fans leaving the stadium after a baseball game.

For this experiment, the observation instances  $\mathbf{y}_t = (c, h)$  where  $c$  is the vehicle count for the 5 min interval at time  $t$  and  $h$  is the hour of day when the observation took place. As weekday and weekend

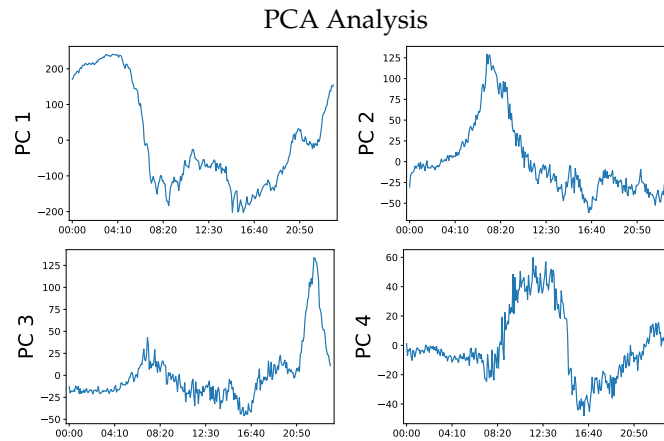
traffic dynamics commonly differ in density flow, the input feature  $\tau_t \in \{1, 2\}$  where  $\tau_t = 1$  denotes weekdays and  $\tau_t = 2$  is assigned for weekends. This choice is supported by the assumption that traffic dynamics generally repeat on a daily basis for weekends and weekdays separately, enforcing the belief of having different transition behaviours for each in the generative model. The VAR order was set to  $p = 288$  being the number of observations in a day.

Figure 4 shows an example of the results for five days in one week. Looking at the weekday plot, it can be seen that the morning and afternoon peak periods were clustered into states 1 (blue) and 4 (green), respectively. The sharp peaks in the evening ( $\sim 23:00$ ) on some weekdays corresponding to traffic rises due to the baseball games were clustered into state 6 (brown). On the weekend, the evening game on Saturday was also clustered into state 6; however, the afternoon ( $\sim 16:30$ ) game on Sunday was clustered into state 7 (pink); a state which is not encountered on weekdays when  $\tau_t = 1$ . This demonstrates the dynamics of the AI-HMM. States are shared across the structures relating to the different values of  $\tau_t$ , however, there may be differences when a set of observations following a certain behaviour trends occur exclusively for realisation  $i$ .



**Figure 4.** Scatter plots for vehicular density counts. Each colour represents a specific state that the respective observations was clustered into. Morning and evening traffic peaks on weekdays are clustered into state 1 (blue) and state 3 (green), respectively. Traffic caused by evening baseball games near 23:00 (Wednesday, Friday and Saturday) have been clustered into state 6 (brown), whereas traffic peaks caused by afternoon games at about 16:30 (Sunday) were clustered into state 7 (pink).

To compare the results, we use Principal Component Analysis (PCA) on the density counts of vehicles per day. Figure 5 shows the first 4 PCs plotted against time. It can be seen that PC 2 seems to capture the dynamics of the morning peak hours, whereas PC 4 captures the dynamics of the lighter weekend midday traffic. PC 3 captures the trend of the evening game which commonly occurs on weekdays and Saturdays. These results inform us which orthogonal direction in the multidimensional data distribution we can expect to see these respective dynamics. However, this is specific to this data only, any additional day vectors added may alter the expected results. Moreover, using this technique alone, it is not possible to identify which days contain the identified trends and which days do not. Within the first four PCs, there was no PC that appears to identify the evening peaks nor the afternoon baseball games. All these shortcomings are solved with the AI-HMM approach demonstrating the effectiveness of the technique.



**Figure 5.** Figure depicting the first four (PCs with the largest eigen values in the PCA analysis of the Dodgers loop sensor dataset). PC 2 capture the dynamics of the morning peak; PC 3 represents the sharp traffic rise due to the evening baseball game, which commonly occurs on weekdays and Saturdays; and PC 4 captures the dynamics of the weekend light midday traffic.

## 5.2. T-Drive Dataset

### 5.2.1. Data Preprocessing

GPS data are often in the form of a collection of geolocation data points in the form of  $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)$  where  $\mathbf{x}_t = (time, \lambda, l) \in \mathbb{R}^3$ ,  $\lambda$  is the longitude coordinate and  $l$  is the latitude coordinate. The altitude is also included in some data sources; however, it is neglected for the purpose of this paper since it is not always available. Given that the altitude is recorded in the dataset, the feature be added for analysis in the proposed technique with no changes to the structure. Extreme outliers relating to errors in the sensors measurement and very short trajectories (a trajectory consisting of only five data points for example) may be neglected since they do not generally hold much information about the journey.

Using only the trajectory time series  $X$ , the feature sequence  $Y$  is calculated to describe the trajectory path where  $Y = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T)$ ,  $\mathbf{y}_t = (\lambda, l, h, \Delta v) \in \mathbb{R}^4$ ,  $h$  is the hour of day and  $\Delta v$  is the average velocity given the distance travelled between two points  $\Delta d$ . The distance is calculated with the Haversine Equation [37]

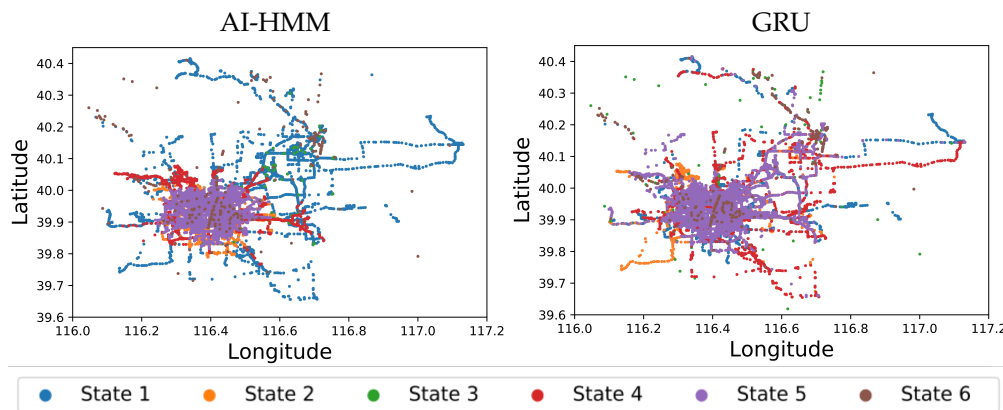
$$h\left(\frac{\Delta d_t}{r}\right) = h(\Delta \lambda) + \cos(\lambda_{t-1}) \cos(\lambda_t) h(\Delta), \quad (24)$$

where  $h(\phi) = \sin^2\left(\frac{\phi}{2}\right)$  and  $r = 6371km$  is the radius of the Earth assumed to be constant on all locations on its surface.  $\Delta v$  is then calculated by  $\frac{\Delta d_t}{\Delta time_t}$ . From the T-drive dataset, the recorder journeys of 10 taxis were selected randomly and modelled with the proposed AI-HMM. GPS data is often recorded as a sequence of geolocation observations  $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)$ , where  $\mathbf{x}_t \in \mathbb{R}^3 = (time, \lambda, l)$ ,  $\lambda$  is the longitude coordinate and  $l$  is the latitude coordinate. The feature sequence  $Y$  is calculated to describe the trajectory path where  $Y = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T)$ ,  $\mathbf{y}_t \in \mathbb{R}^4 = (\lambda, l, h, \Delta v)$ ,  $h$  is the hour of day and  $\Delta v$  is the average velocity given the distance travelled between two points as calculated by the Haversine Equation [37].

### 5.2.2. AI-HMM Modelling with Predefined Input $\tau$

Beijing is a busy city that is known for congested streets throughout most hours of the day [38]. However, the trajectories of trips in peak and off-peak times are likely to have very different latent dynamics. For this part of the case study, we assume that  $\tau_t \in \{1, 2\}$  can represent peak and off-peak traffic times respectively, and they are preset prior to modelling based on the time of day the trajectory

point was recorded in. Due to the smoothness of the sampled GPS trajectories, we opted for a low VAR order of  $p = 2$ . The hyperparameters  $\gamma$ ,  $\alpha$  and  $\kappa$  can be set experimentally by Bayesian model selection or by placing an updatable prior over them as seen in Fox et al. [33]. After running the experiment until convergence, the model identified six states corresponding to different movement behaviour as presented in Figure 6.



**Figure 6.** Results obtained from analysing 10 taxi trajectories using the novel AI-HMM framework (left) and the GRU technique (right). Each data point corresponds to a GPS observation where its colour represents the state to which it was assigned. The plots depict the longitude  $\lambda$  against latitude  $l$  reconstructions of the movements of the taxi locations on a map.

For comparison, a t-distributed Stochastic Neighbour Embedding (tSNE)-based technique [39] was applied following the methodology of Singleton [40]. However, due to the complexity of the data, the technique failed at identifying meaningful clusters; experimentally converging to the sample data average. We also compare our framework with a deep learning Gated Recurrent Unit (GRU) autoencoder embedding model with K-means clustering, similar to that proposed by Yao et al. [15]. To enable direct comparison, the number of clusters for the K-means algorithm was set to 6, following the number of states that were automatically identified with the nonparametric AI-HMM framework.

Figure 6 shows the results obtained from the AI-HMM and the GRU frameworks while Table 1 shows the median velocity  $v_{median}$  interquartile range (IQR) velocity  $v_{IQR}$ , and the most prominent time ranges  $h_{prominent}$  for each of the states using both techniques. The plots depict a reconstructed scatter map of the data where each point represents an observation with its corresponding value and time-step, whereas the colours represent which state the points belong to according to the legend. States 2 (orange) and 5 (purple) are mostly focused in the middle of the city with median velocities of 18 km/h and 19 km/h, respectively. Their most prominent hours of day indicate that they represent the morning and afternoon peak periods when the roads are highly congested. As night approaches and traffic subsides, they are replaced by state 4 (red), which is also most prominent in the center of the city with a median velocity of 23 km/h. State 1 (blue) is focussed on the highway network where velocities are expectedly higher with a median velocity of 26 km/h. The observations belonging to state 3 (green) with a median velocity of 0 km/h are mostly distributed near Beijing capital international airport, where taxis often spend prolonged periods of time in stationary motion awaiting customers. As previously mentioned, GPS data can be very noisy with sporadic measurement errors causing possible location jumps. This noise was not filtered prior to analysis; however, the model was capable of identifying the noisy observations and separating from the rest of the data into state 6 (purple).

**Table 1.** Cluster statistics (velocity in km/h).

State	AI-HMM			GRU		
	$v_{median}$	$v_{IQR}$	$h_{prominent}$	$v_{median}$	$v_{IQR}$	$h_{prominent}$
1	26	35	00:00–23:00	18	24	12:00–18:00
2	18	19	17:00–23:00	18	34	00:00–06:00
3	0	0	00:00–23:00	2147	7697	00:00–23:00
4	23	27	00:00–06:00	16	29	06:00–13:00
5	19	19	10:00–18:00	16	22	18:00–23:00
6	497	1336	00:00–23:00	186	176	00:00–23:00

The GRU analysis was also successful at identifying the noisy observations and filtering them into states 3 (green) and 6 (brown); however, the remaining states have very similar median velocities. Inspection of state velocity histogram distributions indicates that the clustering values have been affected by long periods of stationarity (influenced by taxis at the airport). The distribution of the data points in Figure 6 also shows that there is minimal location based clustering since most of the states overlap between all regions of the city (e.g., city centre and highways). Therefore, the GRU was effective at identifying the most separated clusters; however, it identified more specific and less diverged behaviour patterns. The AI-HMM framework was capable of identifying meaningful states that represent behaviour trends combining location information, velocity and hour of day patterns, yielding more informative and representative results.

## 6. Adaptive Input

In cases where the contextual input  $\tau$  is known or can be assigned to the data, the AI-HMM demonstrates highly informative unsupervised analysis results outperforming other commonly used time series sensor analysis techniques in the literature. However, it is often the case that  $\tau$  is partially missing, and must therefore be inferred from the data adaptively. To demonstrate the performance of adaptively inferring the input vector  $\tau$  given a pre-trained model, another experiment was conducted on the T-drive dataset where  $\tau$  was used to indicate the taxis identity. The full data of two different taxis was extracted and separated into 75% training set and 25% test set. For the training data points, the input was set to  $\tau_t \in \{1, 2\}$  representing taxi 1 and taxi 2, respectively, and they were fit with the AI-HMM model to learn all remaining parameters and variables  $\Theta_k, \mu_k, \Sigma_k, \Pi_v, \beta_v$  and  $\mathbf{z}_{train}$ . The aim for the test data was to infer the values of  $\mathbf{z}_{test}$  as well as  $\tau_{test}$  using Equation (13) assuming they are unknown.

As would be expected, the transition distributions of the two different taxis given the training data showed some overlap in behaviour time dynamics with the addition of some differences. This firstly proves the hypothesis that sensor data belonging to different contextual factors (i.e., person identity in this example) follow different time dynamics in an HMM structure. Correctly identifying these transition distributions allows for more accurate model fitting; the framework will refer to a more specific transitions matrix representing the state transitions under the specified contextual representation during optimisation, generating more informative clustering results. For example, if transitions into state 4 representing late night, off-peak travel is absent from taxi 2's behaviour dynamics, indicating that this taxi avoids late night fairs and prefers to work at different times of day, whereas taxi 1 may prefer to work in such less congested times.

The presence of differences in the transition dynamics allow for accurate estimation of  $\tau$  given the trained parameters on unseen data as reflected by the results obtained from the test data run. Figure 7 shows the AI-HMMs capability of estimating the correct identity of the taxi using adaptive  $\tau$  inference with an accuracy of 93%. The state transition behaviour of an entity can help discover its identity.



**Figure 7.** Prediction of the value of  $\tau_t$  corresponding to the taxi identity on the test. The top plot shows the sample labels, whereas the bottom plot shows the predicted  $\tau_t$  values using the AI-HMM with an accuracy of 93%.

## 7. Conclusions

This paper demonstrates a novel Bayesian nonparametric model for the identification of human behavioural trends from smart city sensors. We address the unsupervised problem of analysing complex data sequences representing human movement by designing a generative PGM that permits for the adaptive contextual switching of transition dynamics. Results obtained from applying the model to complex real-world data demonstrated effectiveness in identifying different dynamic behavioural trends and noise filtering capabilities. We have motivated the value of the presented approach via piratical problems in monitoring and compared the results to examples of alternative tools such as RNNs. The presented framework allows for easy augmentation of existing systems with an adaptive factor which can estimate the effect of discrete contextual information. By adaptively estimating the input value given a pre-trained model, it is possible to accurately identify the contextual realisation which the data belongs to, such as predicting the identity of the individual that generated the data.

As of now, the model is only limited to low dimensional time-series data. However, future work will include broadening the model to scenarios where the observation model can be high dimensional and where multiple dependent contextual variables can be included.

**Author Contributions:** Methodology, Y.Q.; Software, Y.Q.; Supervision, Y.P.R. and M.A.L.; Writing—original draft, Y.Q.; Writing—review & editing, Y.P.R. All authors have read and agreed to the published version of the manuscript.

**Acknowledgments:** We would like to thank the Engineering and Physical Sciences Research Council (EPSRC) for their support and contribution to this work, without whom, this research will not be possible.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AR	Autoregressive
AI-HMM	Adaptive Input Hidden Markov Model
AR-iHMM	Autoregressive infinite Hidden Markov Model
CRHTI	Chicago Regional Household Travel Inventory
DP	Dirichlet Process



FF-BS	Forward Filtering Backward Sampling
GEM	Griffiths, Engen and McCloskey
GPS	Global Positioning System
GRU	Gated Recurrent Unit
HDP-HMM	Hierarchical Dirichlet Process Hidden Markov Model
HMM	Hidden Markov Model
iHMM	infinite Hidden Markov Model
IQR	interquartile Range
LSTM	Long-Short Term Memory
MCMC	Markov Chain Monte Carlo
MLP	Multi-Layer Perceptron
MNIW	Matrix-Normal Inverse-Wishart
PC	Principal Component
PCA	Principal Component Analysis
PGM	Probabilistic Graphical Model
tSNE	t-distributed Stochastic Neighbour Embedding
VAR	Vector Autoregressive
VAR-HMM	Vector Autoregressive Hidden Markov Model

## References

- Jang, S.; Jo, H.; Cho, S.; Mechitov, K.; Rice, J.A.; Sim, S.H.; Jung, H.J.; Yun, C.B.; Spencer, B.F., Jr.; Agha, G. Structural health monitoring of a cable-stayed bridge using smart sensor technology: Deployment and evaluation. *Smart Struct. Syst.* **2010**, *6*, 439–459. [[CrossRef](#)]
- Hong, I.; Park, S.; Lee, B.; Lee, J.; Jeong, D.; Park, S. IoT-based smart garbage system for efficient food waste management. *Sci. World J.* **2014**, *2014*, 646953. [[CrossRef](#)] [[PubMed](#)]
- Castro, M.; Jara, A.J.; Skarmeta, A.F. Smart lighting solutions for smart cities. In Proceedings of the 2013 27th International Conference on Advanced Information Networking and Applications Workshops, Barcelona, Spain, 25–28 March 2013; pp. 1374–1379.
- Masse, R.A.C.; Ochoa-Zezzatti, A.; García, V.; Mejía, J.; Gonzalez, S. Application of IoT with haptics interface in the smart manufacturing industry. *Int. J. Comb. Optim. Probl. Inform.* **2019**, *10*, 57–70.
- De Marsico, M.; Mecca, A.; Barra, S. Walking in a smart city: Investigating the gait stabilization effect for biometric recognition via wearable sensors. *Comput. Electr. Eng.* **2019**, *80*, 106501. [[CrossRef](#)]
- Paul, A.; Ahmad, A.; Rathore, M.M.; Jabbar, S. Smartbuddy: Defining human behaviors using big data analytics in social internet of things. *IEEE Wirel. Commun.* **2016**, *23*, 68–74. [[CrossRef](#)]
- Torre-Bastida, A.I.; Del Ser, J.; Laña, I.; Ildardia, M.; Bilbao, M.N.; Campos-Cordobés, S. Big Data for transportation and mobility: Recent advances, trends and challenges. *IET Intell. Transp. Syst.* **2018**, *12*, 742–755. [[CrossRef](#)]
- Bhatti, F.; Shah, M.A.; Maple, C.; Islam, S.U. A novel internet of things-enabled accident detection and reporting system for smart city environments. *Sensors* **2019**, *19*, 2071. [[CrossRef](#)] [[PubMed](#)]
- Ahmad, A.; Paul, A.; Rathore, M.M.; Chang, H. Smart cyber society: Integration of capillary devices with high usability based on Cyber-Physical System. *Future Gener. Comput. Syst.* **2016**, *56*, 493–503. [[CrossRef](#)]
- Kanungo, A.; Sharma, A.; Singla, C. Smart traffic lights switching and traffic density calculation using video processing. In Proceedings of the 2014 Recent Advances in Engineering and Computational Sciences (RAECS), Chandigarh, India, 6–8 March 2014; pp. 1–6.
- Ozkurt, C.; Camci, F. Automatic traffic density estimation and vehicle classification for traffic surveillance systems using neural networks. *Math. Comput. Appl.* **2009**, *14*, 187–196. [[CrossRef](#)]
- Ma, X.; He, Y.; Luo, X.; Li, J.; Zhao, M.; An, B.; Guan, X. Camera placement based on vehicle traffic for better city security surveillance. *IEEE Intell. Syst.* **2018**, *33*, 49–61. [[CrossRef](#)]
- Yuan, J.; Zheng, Y.; Xie, X.; Sun, G. Driving with knowledge from the physical world. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, San Diego, CA, USA, 21–24 August 2011; pp. 316–324.

14. Mehran, R.; Oyama, A.; Shah, M. Abnormal crowd behavior detection using social force model. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 935–942.
15. Yao, D.; Zhang, C.; Zhu, Z.; Huang, J.; Bi, J. Trajectory clustering via deep representation learning. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 3880–3887.
16. Wang, Y.; Zhu, Y.; He, Z.; Yue, Y.; Li, Q. *Challenges and Opportunities in Exploiting Large-Scale GPS Probe Data*; Technical Report HPL-2011-109; HP Laboratories: Palo Alto, CA, USA, 21 July 2011.
17. Ellam, L.; Girolami, M.; Pavliotis, G.; Wilson, A. Stochastic modelling of urban structure. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2018**, *474*, 20170700. [[CrossRef](#)] [[PubMed](#)]
18. Witayangkurn, A.; Horanont, T.; Sekimoto, Y.; Shibasaki, R. Anomalous event detection on large-scale gps data from mobile phones using hidden markov model and cloud platform. In Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication, Zurich, Switzerland, 8–12 September 2013; pp. 1219–1228.
19. Fox, E.; Sudderth, E.B.; Jordan, M.I.; Willsky, A.S. Nonparametric Bayesian learning of switching linear dynamical systems. In Proceedings of the Neural Information Processing Systems 2008, Vancouver, BC, Canada, 8–10 December 2008; pp. 457–464.
20. Suzuki, N.; Hirasawa, K.; Tanaka, K.; Kobayashi, Y.; Sato, Y.; Fujino, Y. Learning motion patterns and anomaly detection by human trajectory analysis. In Proceedings of the 2007 IEEE International Conference on Systems, Man and Cybernetics, Montreal, QC, Canada, 7–10 October 2007; pp. 498–503.
21. Jiang, S.; Ferreira, J.; González, M.C. Clustering daily patterns of human activities in the city. *Data Min. Knowl. Discov.* **2012**, *25*, 478–510. [[CrossRef](#)]
22. Shih, D.H.; Shih, M.H.; Yen, D.C.; Hsu, J.H. Personal mobility pattern mining and anomaly detection in the GPS era. *Cartogr. Geogr. Inf. Sci.* **2016**, *43*, 55–67. [[CrossRef](#)]
23. Mahadevan, V.; Li, W.; Bhalodia, V.; Vasconcelos, N. Anomaly detection in crowded scenes. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 1975–1981.
24. Rodriguez, M.; Sivic, J.; Laptev, I.; Audibert, J.Y. Data-driven crowd analysis in videos. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 1235–1242.
25. Fu, Z.; Hu, W.; Tan, T. Similarity based vehicle trajectory clustering and anomaly detection. In Proceedings of the IEEE International Conference on Image Processing 2005, Genova, Italy, 14 September 2005.
26. Piciarelli, C.; Micheloni, C.; Foresti, G.L. Trajectory-based anomalous event detection. *IEEE Trans. Circuits Syst. Video Technol.* **2008**, *18*, 1544–1554. [[CrossRef](#)]
27. Frank, P. Chicago Regional Household Travel Inventory. Available online: [https://www.cmap.illinois.gov/documents/10180/77659/TravelTracker\\_ModeShareReport20100604.pdf/a67bf419-c05a-45c2-a127-a18d7984cd7d](https://www.cmap.illinois.gov/documents/10180/77659/TravelTracker_ModeShareReport20100604.pdf/a67bf419-c05a-45c2-a127-a18d7984cd7d) (accessed on 28 January 2020).
28. Zheng, Y.; Xie, X.; Ma, W.Y. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.* **2010**, *33*, 32–39.
29. Raykov, Y.P.; Ozer, E.; Dasika, G.; Boukouvalas, A.; Little, M.A. Predicting room occupancy with a single passive infrared (PIR) sensor through behavior extraction. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; pp. 1016–1027.
30. Kunst, R.M. Vector Autoregressions. Available online: <https://homepage.univie.ac.at/robert.kunst/var.pdf> (accessed on 28 January 2020).
31. Beal, M.J.; Ghahramani, Z.; Rasmussen, C.E. The infinite hidden Markov model. In Proceedings of the Advances in Neural Information Processing Systems 2002, Vancouver, BC, Canada, 9–14 December 2002; pp. 577–584.
32. Teh, Y.W.; Jordan, M.I.; Beal, M.J.; Blei, D.M. Hierarchical Dirichlet Processes. *J. Am. Stat. Assoc.* **2006**, *101*, 1566–1581. doi:10.1198/016214506000000302. [[CrossRef](#)]
33. Fox, E.B. Bayesian Nonparametric Learning of Complex Dynamical Phenomena. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.
34. Pitman, J. Poisson–Dirichlet and GEM invariant distributions for split-and-merge transformations of an interval partition. *Comb. Probab. Comput.* **2002**, *11*, 501–514. [[CrossRef](#)]

35. Ihler, A.; Hutchins, J.; Smyth, P. Adaptive event detection with time-varying poisson processes. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, 20–23 August 2006; pp. 207–216.
36. Yuan, J.; Zheng, Y.; Zhang, C.; Xie, W.; Xie, X.; Sun, G.; Huang, Y. T-drive: Driving directions based on taxi trajectories. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; pp. 99–108.
37. Van Brummelen, G. *Heavenly Mathematics: The Forgotten Art of Spherical Trigonometry*; Princeton University Press: Princeton, NJ, USA, 2012; pp. 160–162.
38. Zhang, H.H. Beijing Still Struggling Deal Traffic Congestion. Available online: <https://www.scmp.com/news/china/article/1298559/beijing-still-struggling-deal-traffic-congestion> (accessed on 7 March 2019).
39. Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
40. Singleton, A.D.; Spielman, S.; Folch, D. *Urban Analytics*; Sage Publishing: Thousand Oaks, CA, USA, 2017.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).