

Novel trajectory clustering method based on distance dependent Chinese restaurant process

Reza Arfa^{1,2}, Rubiyah Yusof^{1,2} and Parvaneh Shabanzadeh^{1,2}

¹Centre for Artificial Intelligence and Robotics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

²Centre for Artificial Intelligence and Robotics, Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

ABSTRACT

Trajectory clustering and path modelling are two core tasks in intelligent transport systems with a wide range of applications, from modeling drivers' behavior to traffic monitoring of road intersections. Traditional trajectory analysis considers them as separate tasks, where the system first clusters the trajectories into a known number of clusters and then the path taken in each cluster is modelled. However, such a hierarchy does not allow the knowledge of the path model to be used to improve the performance of trajectory clustering. Based on the distance dependent Chinese restaurant process (DDCRP), a trajectory analysis system that simultaneously performs trajectory clustering and path modelling was proposed. Unlike most traditional approaches where the number of clusters should be known, the proposed method decides the number of clusters automatically. The proposed algorithm was tested on two publicly available trajectory datasets, and the experimental results recorded better performance and considerable improvement in both datasets for the task of trajectory clustering compared to traditional approaches. The study proved that the proposed method is an appropriate candidate to be used for trajectory clustering and path modelling.

Submitted 30 May 2018
Accepted 24 June 2019
Published 12 August 2019

Corresponding authors
Reza Arfa, rezaarfa@gmail.com
Rubiyah Yusof, rubiyah.kl@utm.my

Academic editor
Pablo Arbelaez

Additional Information and
Declarations can be found on
page 12

DOI 10.7717/peerj-cs.206

© Copyright
2019 Arfa et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Artificial Intelligence, Computer Vision, Visual Analytics

Keywords Path modelling, Trajectory clustering, Anomaly detection, Chinese restaurant process, Distance dependent CRP

INTRODUCTION

The trajectory of a moving object obtained by tracking the object's position from one frame to the next is a simple yet efficient descriptor of an object's motion. Trajectory analysis has long been a research focus in different fields of study (*Jonsen, Myers & Flemming, 2003; Pao et al., 2012; Reed et al., 1999; Fox, Sudderth & Willsky, 2007*). In the context of intelligent surveillance systems (ITS) (*Tian et al., 2017*), trajectory clustering is a critical core technology in many surveillance applications including activity analysis (*Morris & Trivedi, 2011*), path modelling (*Zhang, Lu & Li, 2009*), anomaly detection (*Dee & Velastin, 2008*), and road intersection traffic monitoring (*Aköz & Karsligil, 2014*).

Many trajectory analysis systems consist of two main steps. In the first step, trajectories are grouped into clusters based on their similarities. Most proposed methods assume the number of clusters to be known. After the trajectories are clustered, the path taken by agents

in each cluster will be modelled. There are at least two limitations with these approaches. First, in real-world problems, the number of clusters is usually unknown or is expensive to acquire. Furthermore, trajectory clusters and path models are closely related, whereby the knowledge of one helps in improving the performance of the other.

Most existing trajectory analysis methods can be categorized into similarity-based models and Probabilistic Topic Models (PTM). The main stages of similarity-based approaches are calculating a similarity matrix and clustering the trajectories based on the similarity matrix. At the first stage, pairwise similarities between trajectories are obtained via a similarity function and stored into a $N \times N$ matrix, where N is the total number of available trajectories. Defining a suitable similarity measure is a challenging task that directly affects the overall accuracy of the system (Zhang, Kaiqi & Tieniu, 2006). Well-known similarity measures used for trajectory analysis include Euclidean distance, dynamic time wrapping (DTW) (Keogh & Pazzani, 2000), Hausdorff distance (Atev, Miller & Papanikolopoulos, 2010), and Longest Common Sub-Sequences (LCSS) (Vlachos, Kollios & Gunopulos, 2002). After the similarity matrix is obtained, the second stage uses any standard clustering algorithm to cluster the trajectories into K clusters based on their similarities. Typical clustering algorithms include spectral clustering (Ng, Jordan & Weiss, 2002), agglomerative clustering (Xi, Weiming & Wei, 2006), and fuzzy c-means (Weiming et al., 2006). The main disadvantage of similarity-based approaches is that it requires the number of clusters, K , to be known in advance.

When trajectories are clustered, some studies perform path modelling in a further stage. Path models are useful in intelligent surveillance systems and used for compact representation of clusters, performing real-time anomaly detection (Morris & Trivedi, 2011), and high-level scene understanding (Lei et al., 2014), and route planning (Joseph et al., 2011). Makris & Ellis (2005) modelled the path as an envelope, which denotes the extent of a path by finding the two farthest samples in a cluster. Morris & Trivedi (2011) used the weighted average of trajectories of each cluster to form the path model for that cluster. Based on the dominant set clustering approach, Yiwen et al. (2014) proposed a system that obtains the scene structure from clustered trajectories.

All these approaches, however, model the path after the trajectories are clustered. Therefore, the performance of the modelled path is limited to how well trajectories are clustered. Also, the modelled path is not used to improve the trajectory clustering.

Another well-known class of approaches in trajectory analysis is based on probabilistic topic model (PTM) (Papadopoulos, 2008). In PTM approaches, trajectories are first converted into a set of symbols via a pre-defined codebook. This new representation of trajectories is then treated as documents while the symbols are treated as words. Compared to a similarity-based approach, trajectory analysis methods based on PTM do not usually require the number of clusters in advance.

Jeong, Chang & Choi (2011) used latent Dirichlet allocation (LDA) and the hidden Markov model (HMM) to discover the semantic regions and the temporal relationship between them. A two-level LDA topic model is proposed by Song et al. (Lei et al., 2014). The first level LDA models the motion of single-agent as distributions over patch-based

features. The second level LDA uses the output of the first-level to learn interactions over multi-agents. This model, however, does not perform trajectory clustering.

Wang et al. (2011) proposed a dual hierarchical Dirichlet process (Dual-HDP). Unlike previous PTM models, Dual-HDP is capable of clustering the trajectories and modelling the semantic scene at the same time. Each semantic region is modelled as a distribution over grids, and the scene is modelled as a distribution over the semantic regions. The number of clusters and the semantic scene is decided automatically. Since the model relies only on bag-of-grids representation, it cannot capture the long-term dependency between observations. This results in having a partial path model for each cluster. Having a full path model is an important step for interpreting agents' movement in scenarios such as highways and junctions.

Furthermore, since only quantised trajectories are used, the overall performance of Dual-HDP is highly sensitive to grid size. Choosing a large grid size rapidly decreases the performance due to quantisation error. On the other hand, choosing a small grid size requires considerably more amount of data to learn the trajectory patterns.

This study proposed a trajectory clustering and path modelling system that clusters the trajectories and models the path taken by each cluster at the same time. Our approach is based on distant dependent Chinese restaurant process (DDCRP) (*Blei & Frazier, 2011*), which is a generalisation of the normal Chinese restaurant process (CRP) (*Pitman, 2002*).

METHODS

Distance dependence chinese restaurant process

The Chinese restaurant process (CRP) is a distribution on partitions of integers proposed by *Pitman (2002)*. CRP can be explained by the following analogy: Imagine a Chinese restaurant with an infinite number of tables. The first customer enters the restaurant and sits at the first table with probability 1. Next, customers enter the restaurant and sit at occupied tables with probability proportional to the number of customers sitting on that table or sit at an empty table with the probability relative to a parameter α . After this process, which is known as a customer-table assignment, customers sitting on the same table will share a similar dish. This process can be described as follows:

$$P(z_i = k | z_{-i}, \alpha) \propto \begin{cases} n_k, & k \leq K \\ \alpha, & k = K + 1 \end{cases} \quad (1)$$

where z_i denotes table assignment for the i th customer, K is the total number of occupied tables, and z_{-i} is table assignment the ent of all other customers except i th customer, and n_k is the total number of customers sitting on the i th table. More details of CRP and its connection to Dirichlet process can be found in *Gershman & Blei (2012)*.

The distance dependence Chinese restaurant process (DDCRP) generalises the CRP and allows for a non-exchangeable distribution over partitions (*Blei & Frazier, 2011*). Unlike CRP, where each customer is assigned to a table, in DDCRP each customer is assigned to another customer with a probability relative to their distance/similarity. Therefore, the more similar two customers, the more probable they will get a direct link. It is important to

note that it is still possible for two customers with small similarities to be indirectly linked to each other via intermediate customers. After this procedure, which is also known as a customer to customer assignment, customers who are directly or indirectly linked will sit down at a table and share a similar dish.

More formally, let d_{ij} represent the distance between i th and j th customers. Probability of customer i have a direct link with customer j is calculated as:

$$P(c_i = j | \mathbf{D}, f, \tau) \propto \begin{cases} \tau, & \text{if } i = j \\ f(d_{ij}), & \text{otherwise} \end{cases} \quad (2)$$

where $f(d)$ denotes a monolithically decreasing decaying function that satisfies $f(\infty) = 0$, \mathbf{D} is the matrix of pairwise distance between customers, and τ is a constant that indicates the probability of self-link.

The DDCRP was proposed originally for modelling non-exchangeable text documents where the distance between the dates of documents determines their similarity. The documents are converted into their bag-of-words (BoW) representation before the posterior probability of DDCRP is calculated. Such a conversion to BoW representation is a crucial step that makes the inference of DDCRP computationally tractable.

Recently researchers have adopted DDCRP for problems beyond language processing. [Ghosh et al. \(2011\)](#) proposed a hierarchical extension of DDCRP for producing coarser image segmentations in the form of human-like segmentations. In a more recent study, [Baldassano, Beck & Li \(2015\)](#) used DDCRP to model a complex web of connections with a small number of interacting units. The proposed method is used to model the connectivity between sub-regions of the human brain and analysing human migration behaviour. Also, [Ren et al. \(2016\)](#) used DDCRP for key frame selection from unordered image sets, where the selected frames are used for dense 3D reconstruction.

Trajectory analysis with distance dependent CRP

Unlike text data where observations in documents are words sampled from a corpus with a limited number of words, observations in trajectories are not discrete. Trajectories are vectors with varying length where each observation gets a real value bounded by the scene's size. One can divide the scene into blocks of equal sizes and convert a trajectory into its discrete form. After such a conversion, the resulting quantized trajectories are equal length vectors and each observation gets a discrete value. The size of grids in this case, however, will have a direct impact on the system performance. While theoretically smaller grids can improve the performance, they require substantially more data for training.

Another disadvantage of treating trajectories as documents is the bag-of-words representation. Such representation discards the order between observations. Discarding the orders between samples in trajectory data is problematic since it is possible for agents from opposite directions to share the same observations over grids. One solution to avoid this problem is to quantise the direction of observations ([Wang et al., 2011](#)). Estimating the direction of observation requires further processing and sometimes includes an inaccurate estimation. Such a quantisation increases the size of the corpus and, therefore, requires more data for training. In addition, with bag-of-word representation alone long-term

dependencies between observation cannot be captured which results in having partial path models in existing PTM approaches.

We addressed these problems by using similarity between trajectories as the prior probability in DDCRP. Using such a prior probability limits the assignment of trajectories and promotes trajectories to get linked based on how similar two trajectories are. In addition to the similarity measure, whether the trajectories are linked together or not, also will depend on their discrete observation over the grids. Since most similarity measures can be applied prior to converting the trajectory into discrete form, such a formulation is less sensitive to the choice of grid size. In addition, since some similarity measures, including Modified Hausdorff and LCSS, also take the order of the observations into account, it is not required to quantise the direction anymore.

Any raw trajectory T_i , is usually represented by a sequence of its n_i observation $T_i = [o_{i,1}, \dots, o_{i,l}, \dots, o_{i,n_i}]$. In this representation, $o_{i,l}$ indicates l th observed position of i th object. Let d_{ij} to indicate pairwise distance between i th and j th trajectories. This distance can be of any general distance used to measure similarity between trajectories. The result of pairwise distance between N trajectories can be stored in a distance matrix and denoted as $D \in \mathfrak{R}^{N \times N}$.

Apart from the calculation of distance matrix discussed above, raw trajectories are converted into bag-of-grids representation. For this, the scene is divided into M grid cells of equal size. Based on the cell in which it falls into each observation of a trajectory $o_{i,l}$, is individually quantised. Then a raw trajectory, T_i , is approximated by bag-of-grid representation $X_i \in \mathfrak{R}^M$. Each element of $X_i(s)$ indicates the number of times i th trajectory had an observation in the s th grid cell.

Using DDCRP's metaphor, we use the bag-of-grid representation of trajectories as customers, clusters as the tables and path models as dishes. Based on the definition of DDCRP, it is not possible to draw the table directly. Instead, the outgoing link for each customer needs to be drawn. Trajectories that directly or indirectly link together are considered to be in the same cluster. All trajectories in the similar cluster share the same path model which is a multinomial distribution over the grid cells. Each path model is independently drawn from a base distribution G_0 . In our case, G_0 is a Dirichlet distribution. The full generative process for the news program is as follows:

1. For each trajectory, sample customer assignment $C_i \sim ddCRP(\mathbf{D}, f, \tau)$ as explained in Eq. (2).
2. Drive table assignment from customer assignment. For each table, k , sample its parameter from the base distribution $\varphi_k \sim G_0$
3. For each trajectory, independently draw $X_i \sim Mul(\cdot | \varphi_{z_i})$

The decaying function, $f(\cdot)$, in Eq. (2) was defined as:

$$f(d; \gamma; \gamma_0) = \exp\left(-\frac{d}{\gamma}\right). \quad (3)$$

With this function, the probability of linking two trajectories becomes smaller as their distance increases. The parameter γ controls how fast this probability decays with increasing distance. The inference of DDCRP requires drawing samples for all samples which have the possibility of being linked.

Inference

The key problem that needs to be addressed is computing the posterior distribution of latent customer assignment conditioned on the bag-of-grid cell representation of trajectories $\mathbf{X}_{1:N}$. In our problem, the based distribution G_0 , is conjugate to the data generating distribution $P(X_i|Z_{c_i}, G_0)$. Therefore, the cluster parameters φ_k can be analytically marginalised. After such a calculation, the posterior distribution is expressed by:

$$P(c_{1:N}|\mathbf{X}_{1:N}, \mathbf{D}, f, \tau, \gamma, \gamma_0) \propto \prod_{i=1}^N P(c_i|\mathbf{D}, f, \tau, \gamma, \gamma_0) P(\mathbf{X}_{1:N}|Z(c_{1:N})) \quad (4)$$

where $Z(c_{1:N})$ denotes the table assignment and $P(\mathbf{X}_{1:N}|Z(c_{1:N}))$ is the likelihood function which can be expressed by [Blei & Frazier \(2011\)](#)

$$P(\mathbf{X}_{1:N}|Z(c_{1:N})) = \prod_{k=1}^{|Z(c_{1:N})|} P(\mathbf{X}_{z^k(C_{1:N})}|Z(c_{1:N})) \quad (5)$$

with $|Z(c_{1:N})|$ being the number of unique tables and $z^k(C_{1:N})$ denoting all customers assigned to table k .

Due to the combinatorial sum in the denominator, the analytical solution of the posterior given by [Eq. \(4\)](#) is intractable. Instead of exact inference, collapsed Gibbs strategy ([Blei & Frazier, 2011](#)) is used to derive the posterior inference where the customer assignment is iteratively sampled from the following equation:

$$P(c_i|c_{-i}, \mathbf{X}_{1:N}, \mathbf{D}, f, \tau, \gamma, \gamma_0) \propto P(c_i|\mathbf{D}, f, \tau) \times P(\mathbf{X}_{1:N}|z(c_i \cup c_{-i})) \quad (6)$$

where c_{-i} denotes all customer assignments except for c_i . The first term on the right side of the equation is DDCRP's customer assignment discussed in [Eq. \(2\)](#), and the second term is the likelihood term given by [Eq. \(5\)](#). More details can be found in the [Supplemental Material](#).

RESULTS AND DISCUSSION

The performance of the proposed approach was evaluated on the CROSS ([Morris & Trivedi, 2011](#)) and the Lankershim datasets ([NGSIM: Next Generation Simulation, 2008](#)).

The CROSS dataset provides objects trajectories and their ground truth activities. The data are organized into train and test sets. There are 1,900 and 9,700 trajectories in the train and test sets respectively. Two hundred samples in the test set are labeled as abnormal activities. These samples were discarded in this study and we evaluated the proposed model on 9,500 trajectories in the test set with legal activities ([Fig. 1](#)).

The Lankershim dataset is part of the Next Generation Simulation (NGSIM) program provided by the US Federal Highway Administration (FHWA). The dataset contains videos taken with overhead intersection cameras. The dataset also provided the trajectories of moving vehicles. Based on the time the videos are collected, the data are placed into 8:30 am to 8:45 am and 8:45 am to 9:00 am subsets. The trajectories took place near an intersection, and trajectories outside of this area were removed (see [Fig. 2](#)). The corresponding X and Y coordinate for this region were $-80 < X < 80$ and $300 < Y < 500$ respectively. After

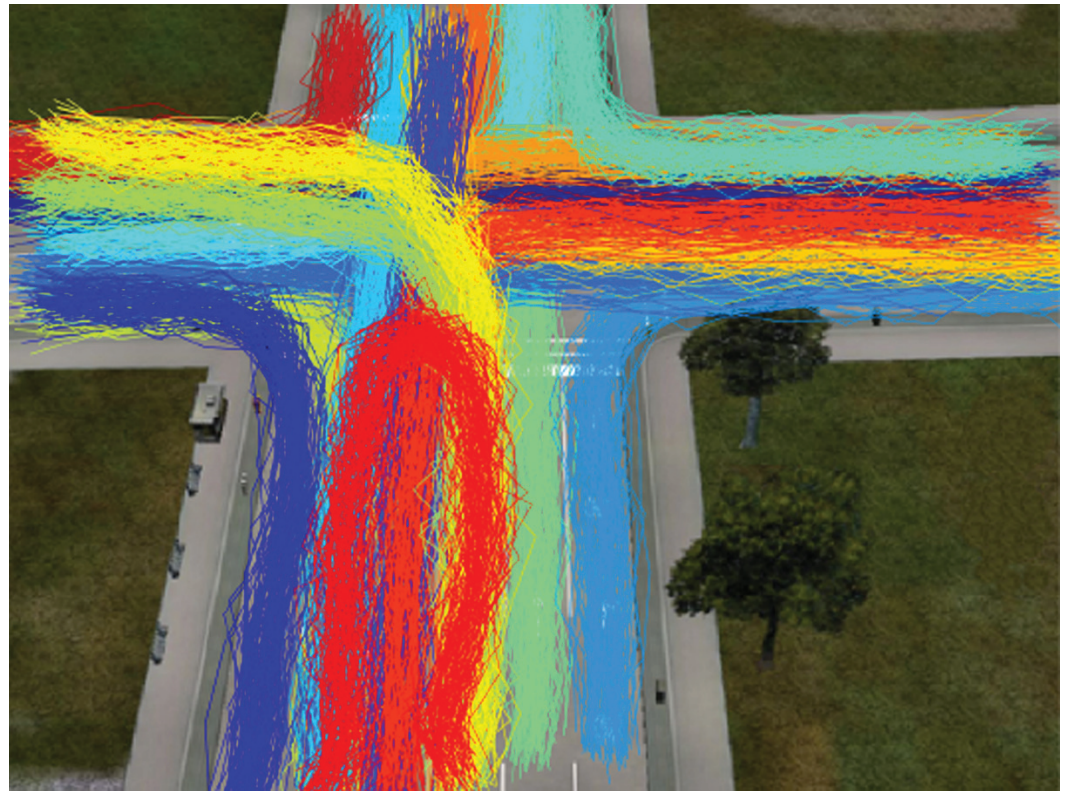


Figure 1 Vehicle trajectories in CROSS dataset. The colors of trajectories indicate the ground truth activity label.

Full-size  DOI: 10.7717/peerjcs.206/fig-1

filtering the trajectories having less than ten observations, a total of 2212 trajectories were obtained. Since this dataset does not provide activity labels for trajectories, the trajectories were manually labelled into 21 activities (19 legal activities, and two activities where agents took illegal maneuvers).

The main parameter that needs to be set prior to experiments is the size of the grid cells. Theoretically, smaller grid cells produce a better result with the cost of requiring more data. Based on the performed experiments, the cell size was set for the CROSS to 40×25 and for the Lankershim into 10×10 pixels. These choices of cell size divide the CROSS and Lankershim into 9 by 19 and 16 by 20 equal sized grid cells respectively. Each raw trajectory was converted into bag-of-grid representation mentioned in the section of Trajectory Analysis with Distance Dependent CRP. The dimensions of bag-of-grids representations are $X_i \in \mathfrak{R}^{1 \times 171}$ and $X_i \in \mathfrak{R}^{1 \times 320}$ for CROSS and Lankershim datasets respectively.

The correct clustering rate (CCR) is used to evaluate the clustering performance. The CCR has been used as evaluation criteria to verify trajectory clustering algorithms in several studies (*Morris & Trivedi, 2009; Weiming et al., 2013; Zhang, Kaiqi & Tieniu, 2006*). Given the ground truth set G and resulting clusters set C , corresponding cluster that maximizes

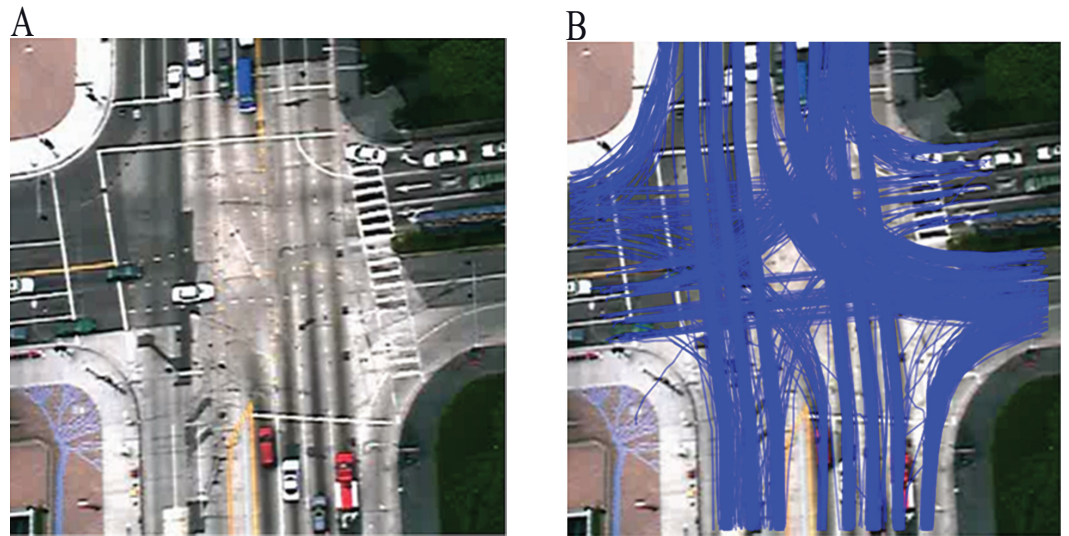


Figure 2 The Lankershim dataset: (A) area of interest, (B) vehicle trajectories in the interest area collected from 8:30 Am to 8:45 Am.

[Full-size](#) DOI: [10.7717/peerjcs.206/fig-2](https://doi.org/10.7717/peerjcs.206/fig-2)

the number of matched labels is found. The CCR is defined as

$$CCR = \frac{1}{N} \sum_{i=1}^K p_i \quad (7)$$

where N is the number of trajectories, K is the number of clusters in the ground truth. Given the assignment between ground truth and estimated cluster labels, p_i is computed as (Zhang, Kaiqi & Tieniu, 2006):

$$p_i = \begin{cases} |c_i \cap g_m|; & \text{given } c_i \in C \text{ assigned to } g_m \in G \\ 0; & \text{otherwise} \end{cases} \quad (8)$$

The proposed method was compared with dual-HDP and three well-known distance measure methods, LCSS, DTW, and modified Hausdorff (MH). For each distance, four unsupervised clustering algorithms were used: K-mean clustering, spectral clustering, agglomerative clustering, and graph-based clustering. The average CCR of clustering algorithms for each distance method is reported in this study. One limitation of distance-based clustering techniques is that they require the number of clusters to be given to them.

To show the effect of choosing the number of clustering on the performance the experiments were run with the different number of clusters, including the true value. The other parameters of competitor methods were set during the course of experiments to achieve their maximum accuracy. For the proposed methods, collapsed Gibbs was performed for 100 samples. After each sampling, CCR was evaluated based on the customer assignment result. Figure 3 shows CCR per sample for the Lankershim and CROSS datasets.

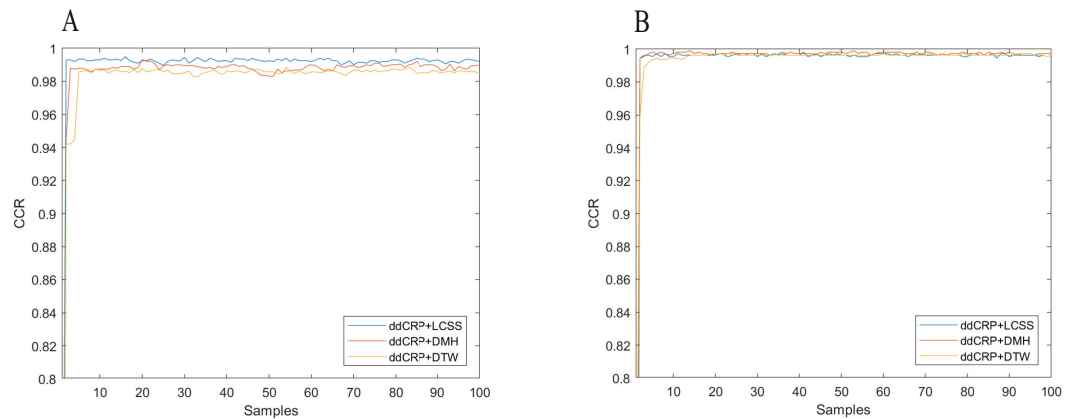


Figure 3 Clustering accuracy of (A) the CROSS dataset, (B) the Lankershim dataset.

Full-size [DOI: 10.7717/peerjcs.206/fig-3](https://doi.org/10.7717/peerjcs.206/fig-3)

Table 1 The CCR Performance of different methods for the CROSS Dataset.

Number of Clusters	5	10	15	19	20	21	25	30
DTW	0.292	0.559	0.806	0.971	0.984	0.968	0.916	0.857
LCSS	0.291	0.555	0.805	0.986	0.971	0.952	0.864	0.792
MH	0.556	0.559	0.807	0.986	0.986	0.973	0.934	0.879
Dual HDP	–	–	–	–	–	0.801	–	–
DDCRP (DTW)	–	–	–	0.986	–	–	–	–
DDCRP (LCSS)	–	–	–	0.993	–	–	–	–
DDCRP (MH)	–	–	–	0.989	–	–	–	–

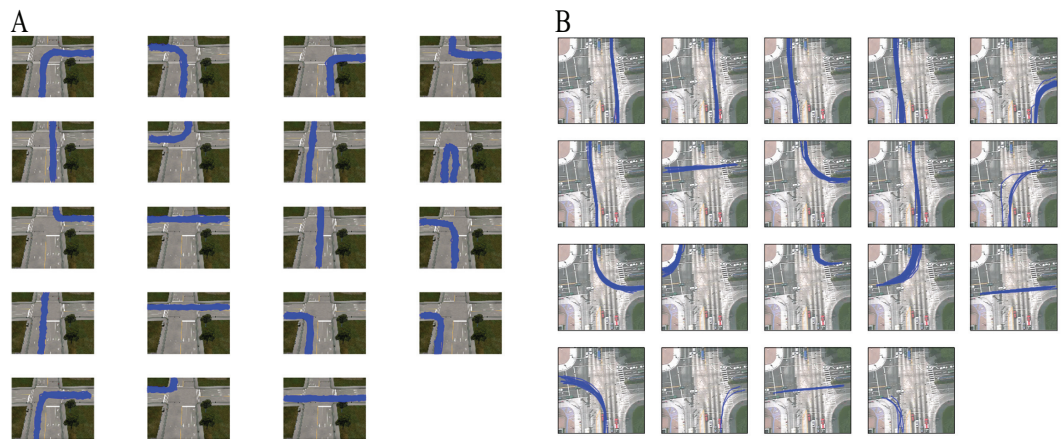
In all methods, CCR achieves greater than 0.9 after the 3rd sample. The average CCR is obtained by averaging the CCR values after neglecting the first ten samples.

The results of trajectory clustering accuracy for the CROSS dataset are summarized in [Table 1](#). The best correct clustering rate is obtained by DDCRP when using LCSS as a distance measure which produces 0.993. The average correct clustering rate of LCSS with traditional clustering algorithm is 0.986. While this value is slightly less than the performance produced by LCSS and DDCRP, it needs to be highlighted that traditional clustering techniques achieved 0.986 correct clustering rate with the assumption of knowing the true total number of clusters. Also, the proposed method improves the correct clustering rate regardless of which similarity method is used. In other words, using DTW and MH as similarity measure along with DDCRP achieve better average CCR compared to traditional clustering algorithms.

Similarly, [Table 2](#) summarizes the clustering accuracy for the Lankershim dataset. Using DDCRP along with MH distance produces the best correct clustering rate of 0.998. Same as CROSS dataset, the proposed method improves correct clustering rate regardless of which similarity measure is used. The most notable improvement is when DTW is used as

Table 2 The CCR Performance of different methods for the Lankershim dataset.

Number of Clusters	5	10	15	18	19	20	25	30
DTW	0.453	0.705	0.901	0.864	0.868	0.864	0.828	0.789
LCSS	0.529	0.846	0.901	0.924	0.925	0.931	0.912	0.899
MH	0.488	0.840	0.973	0.985	0.977	0.974	0.937	0.902
Dual HDP	–	–	–	0.974	–	–	–	–
DDCRP (DTW)	–	–	–	–	0.996	–	–	–
DDCRP (LCSS)	–	–	–	–	0.996	–	–	–
DDCRP (MH)	–	–	–	–	0.998	–	–	–

**Figure 4** Found clusters: (A) the CROSS dataset by using the DDCRP and the LCSS distance methods, (B) the Lankershim dataset by using the DDCRP and the MH distance methods.

Full-size DOI: [10.7717/peerjcs.206/fig-4](https://doi.org/10.7717/peerjcs.206/fig-4)

a similarity measure. In this case, the average CCR for similarity-based clustering is 0.868 while the combination of DTW and DDCRP results in the CCR of 0.996.

After removing clusters with single trajectory and ignoring the initial samples, methods based on DDCRP discovered 19 clusters for both the CROSS and the Lankershim datasets. Figure 4 shows the discovered clusters in the 100th sample for the CROSS and Lankershim datasets. The results shown in this figure are obtained by DDCRP using LCSS and MH distances for the CROSS and Lankershim respectively. The discovered clusters are typical activities in an intersection and include crossing the intersection, turning left, turning right, and u-turn.

As discussed in the Trajectory Analysis with Distance Dependent CRP section, the size of the grid impacts the accuracy of any PTM-based trajectory analysis system. Another advantage of the proposed method compared to the Dual-HDP method is that it is less sensitive to the choice of grid size. This is due to the fact that most PTM models, including dual HDP, are based only on bag-of-grids representation of the trajectories. The proposed method, however, uses both bag-of-grids and pairwise distance between raw trajectories. Therefore, it can be expected that the proposed method is less sensitive to the choice of grid sizes.

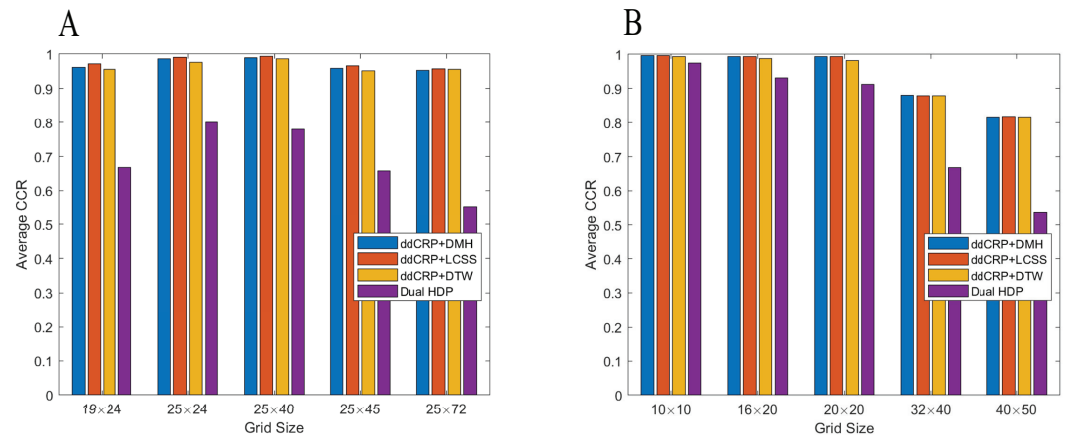


Figure 5 The impact of grid size on the clustering accuracy: (A) the CROSS dataset, (B) the Lankershim dataset.

Full-size DOI: [10.7717/peerjcs.206/fig-5](https://doi.org/10.7717/peerjcs.206/fig-5)

Figure 5 shows the average the CCR of the DDCRP and dual-HDP systems for different sizes of the grid. The grid size of 25×24 and 10×10 pixels produces 0.801 and 0.974 correct clustering rate for the dual-HDP method in the CROSS and Lankershim datasets respectively. However, the accuracy substantially decreases by increasing or decreasing the grid size. The proposed method, however, is more robust to the choice of grid size since the pairwise distance between trajectories is independent of the choice of grid size.

The aim of trajectory path modelling is to discover the paths commonly taken by objects in each cluster. One benefit of our method is its ability to model the path simultaneous to trajectory clustering. In our study, each path is characterized by the distribution over grid cells in a scene. Each cell for a path can be associated to any number in the range of 0 to 1, where 0 are the cells that have no chance of being observed in that path. As the values of a cell are closer to 1, this cell become more essential for the path, and the probability of it being passed by trajectories belonging to that path increases.

The path modelling experiments were conducted with the same parameter setup discussed earlier in this section. Figure 6 shows the cluster models for the CROSS and Lankershim datasets. The blue cells are less likely to be observed by trajectories in that cluster. Conversely, the red cells are more probably observed by trajectories. Then most paths have their probable grid cells in the middle of their route, while when moving further away to the edges of the routes, the probability of grid cells decreases.

CONCLUSION

This paper proposed an unsupervised approach for trajectory clustering and modelling. The generative process of trajectory analysis was modelled via a probabilistic model. The pairwise distances were used as prior in DDCRP to promoting similar trajectories to be clustered. The DDCRP were used to combine the advantages of similarity-based and PTM-based approaches. Compared to probabilistic topic approaches, our method is able to model the full path taken by agents in each cluster. Unlike most similarity-based methods,

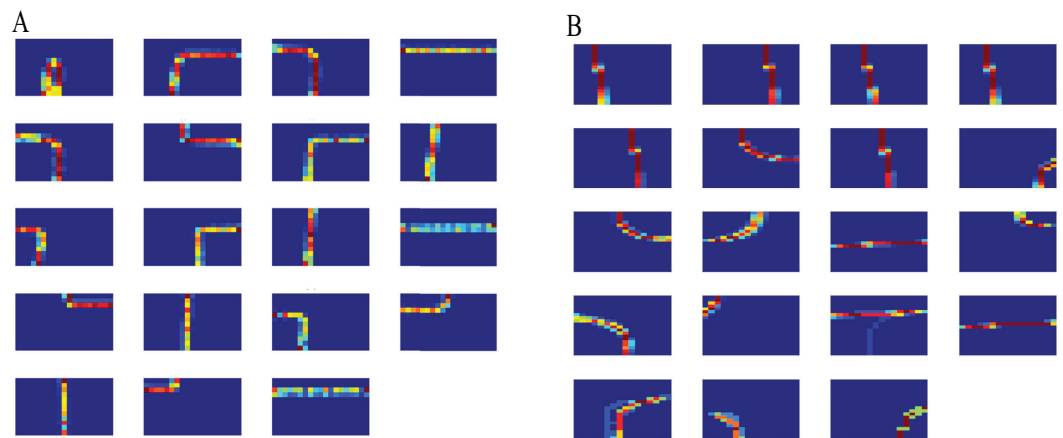


Figure 6 Cluster models: (A) the CROSS dataset, (B) the Lankershim dataset.

[Full-size](#) [DOI: 10.7717/peerjcs.206/fig-6](https://doi.org/10.7717/peerjcs.206/fig-6)

our method drives the number of clusters automatically. The proposed trajectory analysis system clusters the trajectories and models the clusters' paths at the same time. Specifically, raw trajectories were converted to bag-of-grid cells representation and considered each cluster with its distribution over the grids. Experimental results confirmed the effectiveness and usefulness of the proposed algorithm in trajectory clustering and modelling compared to other methods. The proposed approach is planned to have an online learning capability, where the cluster and path models keep updated as more data is observed.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Ministry of Education Malaysia by through a Research University Grant of University Technology Malaysia (UTM), project titled “Intelligent fault detection and diagnosing for process plant R.k430000.77434.4J010.” The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Ministry of Education Malaysia by through a Research University Grant of University Technology Malaysia (UTM).

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Reza Arfa, Rubiyah Yusof and Parvaneh Shabanzadeh conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, performed the

computation work, authored or reviewed drafts of the paper, approved the final draft, the authors contributed equally to the completion of the manuscript.

Data Availability

The following information was supplied regarding data availability:

The code is available at: <https://github.com/rezaarfa/motionlearning> and in the [Supplemental Information](#).

The CROSS dataset can be downloaded from http://cvrr.ucsd.edu/bmorris/datasets/dataset_trajectory_analysis.html. The Lankershim is part of the NGSIM dataset and can be downloaded from <https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm>.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.206#supplemental-information>.

REFERENCES

- Aköz Ö, Karşlıgil ME. 2014.** Traffic event classification at intersections based on the severity of abnormality. *Machine Vision and Applications* **25**:613–632 DOI [10.1007/s00138-011-0390-4](https://doi.org/10.1007/s00138-011-0390-4).
- Atev S, Miller G, Papanikolopoulos NP. 2010.** Clustering of vehicle trajectories. *IEEE Transactions on Intelligent Transportation Systems* **11**:647–657 DOI [10.1109/TITS.2010.2048101](https://doi.org/10.1109/TITS.2010.2048101).
- Baldassano C, Beck DM, Li F-F. 2015.** Parcellating connectivity in spatial maps. *PeerJ* **3**:e784 DOI [10.7717/peerj.784](https://doi.org/10.7717/peerj.784).
- Blei DM, Frazier PI. 2011.** Distance dependent Chinese restaurant processes. *Journal of Machine Learning Research* **12**:2461–2488.
- Dee HM, Velastin SA. 2008.** How close are we to solving the problem of automated visual surveillance? *Machine Vision and Applications* **19**:329–343 DOI [10.1007/s00138-007-0077-z](https://doi.org/10.1007/s00138-007-0077-z).
- Fox EB, Sudderth EB, Willsky AS. 2007.** Hierarchical Dirichlet processes for tracking maneuvering targets. In: *2007 10th international conference on information fusion*. IEEE, 1–8 DOI [10.1109/ICIF.2007.4408155](https://doi.org/10.1109/ICIF.2007.4408155).
- Gershman SJ, Blei DM. 2012.** A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology* **56**:1–12 DOI [10.1016/j.jmp.2011.08.004](https://doi.org/10.1016/j.jmp.2011.08.004).
- Ghosh S, Ungureanu AB, Sudderth EB, Blei DM. 2011.** Spatial distance dependent Chinese restaurant processes for image segmentation. In: *Proceedings of the 24th international conference on neural information processing systems*. Granada, Spain: Curran Associates Inc., 1476–1484.
- Jeong H, Chang HJ, Choi JY. 2011.** Modeling of moving object trajectory by spatio-temporal learning for abnormal behavior detection. In: *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE international conference on*. Piscataway: IEEE, 119–123 DOI [10.1109/AVSS.2011.6027305](https://doi.org/10.1109/AVSS.2011.6027305).

- Jonsen ID, Myers RA, Flemming JM. 2003.** Meta-analysis of animal movement using state-space models. *Ecology* **84**:3055–3063 DOI [10.1890/02-0670](https://doi.org/10.1890/02-0670).
- Joseph J, Doshi-Velez F, Huang AS, Roy N. 2011.** A Bayesian nonparametric approach to modeling motion patterns. *Autonomous Robots* **31**:383–400 DOI [10.1007/s10514-011-9248-x](https://doi.org/10.1007/s10514-011-9248-x).
- Keogh EJ, Pazzani MJ. 2000.** Scaling up dynamic time warping for datamining applications. In: *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. New York: ACM, 285–289 DOI [10.1145/347090.347153](https://doi.org/10.1145/347090.347153).
- Lei S, Fan J, Zhongke S, Molina R, Katsaggelos AK. 2014.** Toward dynamic scene understanding by hierarchical motion pattern mining. *IEEE Transactions on Intelligent Transportation Systems* **15**:1273–1285 DOI [10.1109/TITS.2014.2299403](https://doi.org/10.1109/TITS.2014.2299403).
- Makris D, Ellis T. 2005.** Learning semantic scene models from observing activity in visual surveillance. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **35**:397–408 DOI [10.1109/TSMCB.2005.846652](https://doi.org/10.1109/TSMCB.2005.846652).
- Morris B, Trivedi M. 2009.** Learning trajectory patterns by clustering: experimental studies and comparative evaluation. In: *IEEE conference on computer vision and pattern recognition, 2009*. Piscataway: IEEE, 312–319 DOI [10.1109/CVPR.2009.5206559](https://doi.org/10.1109/CVPR.2009.5206559).
- Morris BT, Trivedi MM. 2011.** Trajectory learning for activity understanding: unsupervised, multilevel, and long-term adaptive approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**:2287–2301 DOI [10.1109/TPAMI.2011.64](https://doi.org/10.1109/TPAMI.2011.64).
- Ng AY, Jordan MI, Weiss Y. 2001.** On spectral clustering: analysis and an algorithm. In: *Proceedings of the 14th international conference on neural information processing systems: natural and synthetic*. Vancouver, British Columbia, Canada: MIT Press, 849–856.
- NGSIM: Next Generation Simulation. 2008.** In: 2008. Available at www.ngsim.fhwa.dot.gov: FHWA, U.S. Department of Transportation.
- Pao H-K, Fadlil J, Lin H-Y, Chen K-T. 2012.** Trajectory analysis for user verification and recognition. *Knowledge-Based Systems* **34**:81–90 DOI [10.1016/j.knosys.2012.03.008](https://doi.org/10.1016/j.knosys.2012.03.008).
- Papadopoulos AN. 2008.** Trajectory retrieval with latent semantic analysis. In: *Proceedings of the 2008 ACM symposium on applied computing*. New York: ACM, 1089–1094.
- Pitman J. 2002.** Combinatorial stochastic processes. Technical Report 621. Available at <https://www.stat.berkeley.edu/~pitman/621.pdf>.
- Reed M, Johansen Ø, Brandvik PJ, Daling P, Lewis A, Fiocco R, Mackay D, Prentki R. 1999.** Oil spill modeling towards the close of the 20th century: overview of the state of the art. *Spill Science & Technology Bulletin* **5**:3–16 DOI [10.1016/S1353-2561\(98\)00029-2](https://doi.org/10.1016/S1353-2561(98)00029-2).
- Ren Q, Wang Q, Zhang J, Chen S. 2016.** Unordered images selection for dense 3D reconstruction based on distance dependent Chinese restaurant process. In: *Intelligent Control and Automation (WCICA), 2016 12th world congress on*. IEEE, 2969–2973 DOI [10.1109/WCICA.2016.757842](https://doi.org/10.1109/WCICA.2016.757842).

- Tian B, Morris BT, Tang M, Liu Y, Yao Y, Gou C, Shen D, Tang S. 2017.** Hierarchical and networked vehicle surveillance in its: a survey. *IEEE Transactions on Intelligent Transportation Systems* **18**:25–48 DOI [10.1109/TITS.2016.2552778](https://doi.org/10.1109/TITS.2016.2552778).
- Vlachos M, Kollios G, Gunopulos D. 2002.** Discovering similar multidimensional trajectories. In: *18th international conference on data engineering, 2002*. Piscataway: IEEE, 673–684 DOI [10.1109/ICDE.2002.994784](https://doi.org/10.1109/ICDE.2002.994784).
- Wang X, Ma KT, Ng G-W, Grimson WEL. 2011.** Trajectory analysis and semantic region modeling using nonparametric hierarchical bayesian models. *International Journal of Computer Vision* **95**:287–312 DOI [10.1007/s11263-011-0459-6](https://doi.org/10.1007/s11263-011-0459-6).
- Weiming H, Xi L, Guodong T, Maybank S, Zhongfei Z. 2013.** An incremental DPMM-based method for trajectory clustering, modeling, and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**:1051–1065 DOI [10.1109/TPAMI.2012.188](https://doi.org/10.1109/TPAMI.2012.188).
- Weiming H, Xuejuan X, Zhouyu F, Xie D, Tieniu T, Maybank S. 2006.** A system for learning statistical motion patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**:1450–1464 DOI [10.1109/TPAMI.2006.176](https://doi.org/10.1109/TPAMI.2006.176).
- Xi L, Weiming H, Wei H. 2006.** A coarse-to-fine strategy for vehicle motion trajectory clustering. In: *18th international conference on pattern recognition, 2006*. 591–594 DOI [10.1109/ICPR.2006.45](https://doi.org/10.1109/ICPR.2006.45).
- Yiwen W, Tze IY, Keathly D, Buckles B. 2014.** Dynamic scene modelling and anomaly detection based on trajectory analysis. *IET Intelligent Transport Systems* **8**:526–533 DOI [10.1049/iet-its.2012.0119](https://doi.org/10.1049/iet-its.2012.0119).
- Zhang T, Lu H, Li SZ. 2009.** Learning semantic scene models by object classification and trajectory clustering. In: *2009 IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE, 1940–1947 DOI [10.1109/CVPR.2009.5206809](https://doi.org/10.1109/CVPR.2009.5206809).
- Zhang Z, Kaiqi H, Tieniu T. 2006.** Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes. In: *18th international conference on pattern recognition, 2006*. 1135–1138 DOI [10.1109/ICPR.2006.392](https://doi.org/10.1109/ICPR.2006.392).