STATISTICAL ANOMALY DETECTION FOR MONITORING OF HUMAN DYNAMICS

K. Kamiya[a], T. Fuse[b]

[a] Dept. of Civil Engineering, The University of Tokyo, 7-3-1 Hongo Bunkyo Tokyo 113-8656, Japan
kamiya@trip.t.u-tokyo.ac.jp
[b] Dept. of Civil Engineering, The University of Tokyo, 7-3-1 Hongo Bunkyo Tokyo 113-8656, Japan
fuse@civil.t.u-tokyo.ac.jp

Commission V, WG V/4

ABSTRACT:

Understanding of human dynamics has drawn attention to various areas. Due to the wide spread of positioning technologies that use GPS or public Wi-Fi, location information can be obtained with high spatial-temporal resolution as well as at low cost. By collecting set of individual location information in real time, monitoring of human dynamics is recently considered possible and is expected to lead to dynamic traffic control in the future. Although this monitoring focuses on detecting anomalous states of human dynamics, anomaly detection methods are developed ad hoc and not fully systematized. This research aims to define an anomaly detection problem of the human dynamics monitoring with gridded population data and develop an anomaly detection method based on the definition. According to the result of a review we have comprehensively conducted, we discussed the characteristics of the anomaly detection of human dynamics monitoring and categorized our problem to a semi-supervised anomaly detection problem that detects contextual anomalies behind time-series data. We developed an anomaly detection method based on a sticky HDP-HMM, which is able to estimate the number of hidden states according to input data. Results of the experiment with synthetic data showed that our proposed method has good fundamental performance with respect to the detection rate. Through the experiment with real gridded population data, an anomaly was detected when and where an actual social event had occurred.

1 INTRODUCTION

Understanding of human dynamics such as people's mobility or distribution in a city has drawn attention to various areas; for example, urban planning and marking. Recently, due to the wide spread of positioning technologies that use GPS or public Wi-Fi, location information can be obtained with high spatial-temporal resolution as well as at low cost. Therefore, by collecting set of individual location information in real time, monitoring of human dynamics is considered possible and is expected to lead to dynamic traffic control in the future. This monitoring focuses on detecting anomalous states of human dynamics as well as normal states. Because there are uncountable possible factors that may trigger the anomalous states, it would not possible to monitor all of the factors. In contrast, if anomalous states are detected by analysis on observation data of human dynamics monitoring such as mesh population data or set of individual GPS trajectories, this will be very useful. The characteristics that human dynamics may drastically change depending on regions and times, and the massive observation data through 24-hour-monitoring would however make manual anomaly detection difficult. On the other hand, statistical anomaly detection method can learn normal states from regularly acquired data and automatically detect anomalies as a different state from normal states.

Several researches on statistical anomaly detection of human dynamics have been conducted. Candia (Candia et al., 2008) has used extensive cell phone records resolved in both time and space, focused on the occurrence of anomalous events and discussed how these spatial-temporal anomalies can be described using standard percolation theory tools. Horanont (Horanont, 2010) in terms of hot spots extraction and visualization, has interpolated aggregate cell phone Call Detail Records by using an implemented platform. As an example of researches focusing on GPS trajectory data, Pan (Pan et al., 2013) has identified anomalies according to drivers' routing behavior on an urban traffic network. Horiguchi (Horiguchi et al., 2013) has implemented a real time monitoring system that provides the visual comprehension of regional traffic situations in terms of the fluidity and the singularity indices. Moreover, there are other researches that try to detect an anomalous trajectory caused by external factors such as accidents or vicious drivers by collecting the vehicle trajectories over a long period of time.

Although these researches on statistical anomaly detection of human dynamics have been conducted so far, these methods are developed ad hoc due to the diversity of data and anomalies. Anomaly detection methods are not fully systematized yet. Additionally it is considered difficult even for a well-suited method for certain application domain to apply to another domain because of the discrepancy in the definition of anomalies between the domains.

For these reasons, this research aims to firstly define an anomaly detection problem with regard to the human dynamics monitoring. Because of the anonymity and the high accessibility, this research targets gridded population data that is aggregated individual GPS data into each grid to estimate population of each grid for each instance of time. Secondly, based on the definition, we develop a suitable anomaly detection method for the human dynamics monitoring, and finally evaluate the fundamental performance of the proposed method. Therefore, development of anomaly detection method and clarification of the fundamental performance of the proposed method will make it possible to detect anomalous states behind the human dynamics. In the future, the human dynamics monitoring is expected to lead an effective traffic control and a better understanding of social mechanisms.

The paper is organized as follows: In section 2, we explain how the anomaly detection of human dynamics monitoring can be in-

terpreted with respect to the factors that characterize anomaly detection problem. Additionally, we give an explanation that the state-space model, especially the Hidden Markov Model based on Hierarchical Dirichlet Process, is considered possible to apply. In section 3, after introducing the properties of nonparametric Bayesian model such as Dirichlet process, detail of the proposed anomaly detection method is presented. We demonstrate our proposed method with synthetic and real data and discuss the fundamental performance in section 4. Finally, section 5 concludes with a summary and some directions for future works.

2 PROBLEM SETUP FOR ANOMALY DETECTION OF HUMAN DYNAMIC MONITORING

2.1 Interpretation of anomaly detection factors

Statistical anomaly detection aims to identify data that are not consistent with a pattern that most other data follow (Chandola et al., 2009). It is also referred to as novelty detection (Pimentel et al., 2014) or outlier detection (Hodge and Austin, 2004) in different application domains although the main principle is common. The importance of anomaly detection attributes to value of detected anomalies in data; anomalies are often important trends behind huge data or critical and significant data. Combined with the progress of data acquisition techniques and computer performance, this great importance makes anomaly detection gain much research attention in variety of application domains; intrusion detection, image processing, structural health monitoring, et cetera.

Firstly, we have comprehensively reviewed the previous studies of anomaly detection in variety of areas and a result of the review shows that there are four factors that characterize anomaly detection problem; the nature of the input data, the output of anomaly detection, availability or unavailability of the data labels, and the type of anomalies (Chandola et al., 2009). This section provides a clear explanation of these factors as well as the interpretation of human dynamics monitoring targeting gridded population data.

2.1.1 Nature of input data: The first factor that characterizes anomaly detection is the nature of the input data. This can be roughly divided into binary, categorical and continuous. Additionally, each input data might consist of univariate or multivariate. Input data can also be categorized in terms of relationship between each input data; for example, temporal data, spatial data, graph data, et cetera. In the case of this study with gridded population data, input data can be considered as temporal data of continuous values because each gridded population can be regarded as input value. Moreover, since it is commonly considered that urban population distribution has a greater influence with closer region, this temporal data may also have some spatial correlation. Therefore, development to handle temporal anomaly detection is necessary and spatial extension may be preferred.

2.1.2 Output of anomaly detection: The second important factor for any anomaly detection method is the manner in which the anomalies are reported. There are typically two ways: One is to calculate anomaly score for each instance, and the another is to assign "normal" label or "anomalous" label for each instance. If the anomaly score of every grid is computed, it can be utilized as an indicator of the traffic control priority. Also, if normal or anomalous label that represents what kind of state the grid is, it can be beneficial. In this manner, the human dynamics monitoring is considered to accept both anomaly score and labels.

2.1.3 Data labels: The normal or anomalous labels associated with a input data instance that is used to learn anomaly detection models is the third factor to characterize anomaly detection. Since it is difficult to obtain all of possible states, especially anomalous states, anomaly detection techniques can be divided into three modes based on the extent to which the labels are available; supervised anomaly detection with the need of both normal and anomalous labels in training dataset, semi-supervised anomaly detection with only normal labels, and unsupervised anomaly detection without any prior information with regard to input dataset. In the case of human dynamics monitoring, normal labels can be obtained in a relatively simple way by using observational data of the day when and the place where big social events such as serious accidents and natural disasters are not reported. On the other hand, the availability of anomalous labels is a big issue for not only human dynamics monitoring but any anomaly detection problem. For these reasons, semi-supervised or unsupervised anomaly detection is thought able to be employed.

2.1.4 Type of anomalies: Anomalies can be classified into following three categories; point anomaly, contextual anomaly and collective anomaly. An individual data is regarded as a point anomaly if the data is considered anomalous with respect to the rest of the data, a data instance is regarded as a contextual anomaly if anomalous in the specific context (but not separately), and a set of data instance is regarded as a collective anomaly if anomalous with respect to the entire data set. As examples of anomalies that are possibly detected in gridded population data, there is an extreme increase or decrease caused by traffic congestion and a change of the population pattern affected by a change of the traffic demands. Figure 1 illustrates these possible anomalies in a gridded population temporal data. Aimed at the detection of these temporal anomalies as well as spatial anomalies due to the spatial correlations of gridded population data, contextual anomaly detection techniques are considered suitable to this study.
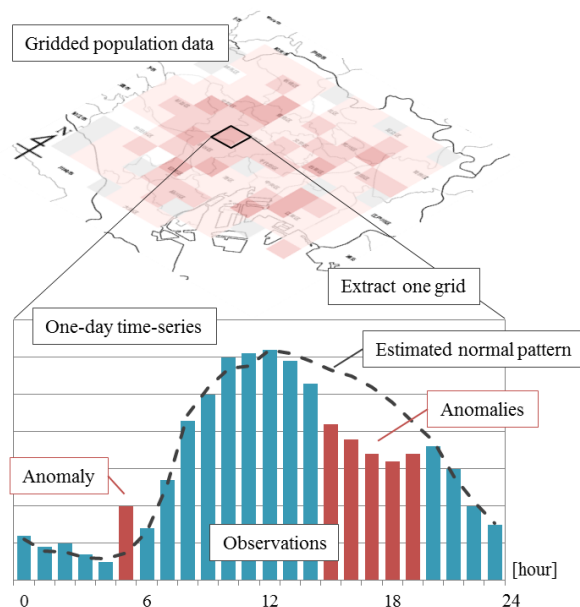

Figure 1: Anomalies in gridded population data

Based on these analyses showed in Table 2, we define the anomaly detection of human dynamics monitoring as a problem that detects contextual anomalies in time-series data of gridded population data by learning normal states with training dataset.

Table 2: Characteristics of the anomaly detection of human dynamics monitoring

Factors	Interpretation of human dynamics monitoring
Input data	Temporal data of continuous value of an individual gridded population
	Temporal data with spatial correlation
Output	Both anomaly scores and labels
Data labels	Semi or unsupervised anomaly detection
	Normal labels obtained from daily data
Type of anomaly	Contextual anomaly
	An extreme increase or decrease caused by traffic congestion
	A change of the population pattern by a change of the traffic demands

2.2 State-space model for anomaly detection

According to the problem setup defined above, anomaly detection techniques based on state-space models are considered possible to be applied. State-space models are models that consist of hidden (or latent) state variables and observed variables, and assume that hidden variables emit observed variables while evolving through time. This representation makes it possible to model complex temporal data, and this is the reason why state-space models are often used for anomaly detection in time-series data (Pimentel et al., 2014). Additionally, the high capability of addition of new variables such as neighbor population can make spatial extension of models possible.

There is said to be two most common techniques that use state-space models for anomaly detection. One assumes that anomalies are detected by calculating likelihood or emission probability of each observation, while the another assumes that anomalies are assigned to state variables which indicate anomalous states.

When considering the application for the anomaly detection of human dynamics monitoring, the construction of data generation models can be seriously problematic with regard to the former technique. Since there are few studies with respect to prediction or simulation of gridded population data, it is considered difficult to develop system models and observation models of which accuracy is enough for anomaly detection.

On the other hand, the later technique often suffers from the decision of the number of the hidden states because it generally needs to be fixed beforehand. Unfortunately, the accurate number of the hidden states that gridded population data possibly come into is unknown. This problem, however, can be conquered when the number is estimated from the collection of observation data.

This research regards that Hidden Markov Models based on Hierarchical Dirichlet Process, which are ones of nonparametric Bayesian models, can estimate the number of states of temporal data, and can be applied to the anomaly detection of human dynamics monitoring. Especially, we use a sticky Hierarchical Dirichlet Process - Hidden Markov Model (sHDP-HMM) which allows more robust learning of smoothly varying dynamics by limiting rapid state transitions. In the next section, an anomaly detection method based on a sHDP-HMM is proposed.

3 THE PROPOSED ANOMALY DETECTION METHOD

3.1 Nonparametric Bayesian models

In this section, following a brief explanation about the Dirichlet Process that is important aspect for understanding of nonparametric Bayesian models, a sticky HDP-HMM is referred which is applied Dirichlet Process to Hidden Markov Model.

3.1.1 Dirichlet process: The dirichlet process (DP) is called a distribution over a distribution or a countably infinite probability measure, uniquely defined by a base measure H on a parameter space Θ and a concentration parameter α. We denote it by $\mathrm{DP}(\alpha, H)$. DP is also regarded as a stochastic process in which a probability distribution called as a base measure is approximated by a random discrete distribution. A random draw $G_0 \sim \mathrm{DP}(H, \alpha)$ can be expressed as

$$G_0 = \sum_{k=1}^{\infty} \pi_k \delta(\theta_k), \quad \theta_k \sim H \qquad (1)$$

where the $\delta(\theta_k)$ indicates a Dirac delta, and the weights π_k are sampled via a stick-breaking construction, denoted below:

$$\begin{aligned} \nu_k &= \mathrm{Beta}(1, \alpha) \\ \pi_k &= \nu_k \prod_{l=1}^{k-1} (1 - \nu_l) \end{aligned} \qquad (2)$$

Here, let them be denoted by $\pi = (\pi_1, \pi_2, ..., \pi_\infty) \sim \mathrm{GEM}(\alpha)$. The stick-breaking process can be interpreted as dividing a stick of 1 in length, which is a total sum of probability, by weights π_k drawn from a Beta distribution. These infinitely divided sticks make an infinite multinomial distribution π. Additionally, $\delta(\theta_k)$ is called an atom and assigned to parameter drawn from the parameter space Θ. This representation gives us an insight into how the parameter α controls the model complexity of G_0 in terms of the expected numbers of components.

The DP is commonly used as a prior on the parameters of a mixture model with a random number of components. Such a model is called a dirichlet process mixture model and its graphical model is depicted in Figure 3(a). Here, let N be the number of the observations y_i's and z_i be an indicator for the assignment of observations to the parameters. An indicator z_i is drawn from the countably infinite multinomial distribution π constructed by the stick-breaking process $\mathrm{GEM}(\alpha)$, a parameter θ_k is assigned by the indicator z_i, and the observation y_i is emitted from the likelihood distribution F with the parameter θ_k.

3.1.2 Hierarchical dirichlet process: The hierarchical dirichlet process (HDP) defines first a global base measure G_0, drawn by a DP prior $\mathrm{DP}(\gamma, H)$ as an average distributions of the group specific distributions G_j, which is sampled from a $\mathrm{DP}(\alpha, G_0)$. Since G_0 is discrete due to the characteristics of the DP, the group specific distributions G_j will share the same atoms $\delta(\theta_k)$. This ensures that the mixture models in the different groups share mixture components.

With respect to the representation of the stick-braeking process, the HDP can be interpreted that the global parameter β is sampled from $\mathrm{GEM}(\gamma)$, then the group specific distribution π_j is sampled from the $\mathrm{DP}(\alpha, \beta)$, and finally the group specific indicator z_{ji} and the ith observation are generated in the same way as the DP, as depicted in Figure 3(b).

3.1.3 Hierarchical dirichlet process - hidden markov model: The model that hierarchical dirichlet process applies to a hidden
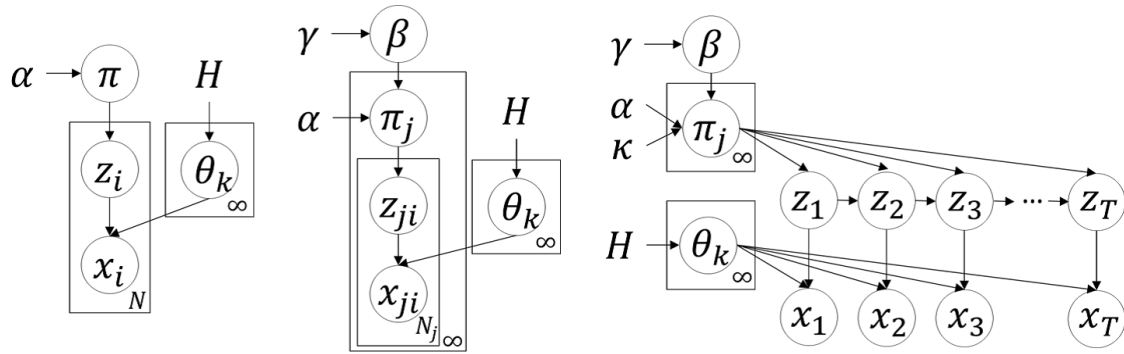
Figure 3: Dirichlet process mixture model(a), Hierarchical Dirichlet Process(b) and sticky HDP-HMM(c)

markov model is called hierarchical dirichlet process - hidden markov model (HDP-HMM) (Teh et al., 2006). In HDP-HMMs, the inditator z_t acts as the hidden state at the time t, π_j means the state specific transition distribution for state j, and β is called as the global transition distribution. Due to the fact that each group specific ditribution G_j shares the common components, each state specific distribution π_j share the common countably infinite states. This also means the state transition matrix π can be regarded to have raw and column of countably infinite.

The ability of HDP-HMMs to have countably infinite states theoretically make it possible to estimate the number of states according to input data, and has leads to some application such as clustering and anomaly detection, especially where the number of clusters or states is unknown (Lello et al., 2012).

An extension of the HDP-HMM, nemed a sticky HDP-HMM or sHDP-HMM, has been deveploed by Fox (Fox et al., 2009) to solve the issue which the standard HDP-HMM learns a model with unrealistically fast dynamics that causes the reduction of the model precision. The sHDP-HMM introduces the hyper-parameter $\kappa > 0$ that controls the expected probability of the self-transition to the stage where the state specific ditribution π_j is sampled. The graphical representation of the sHDP-HMM is depicted in Figure 3(c). The resulting generative model is given by:

$$
\begin{aligned}
\beta|\gamma &\sim \text{GEM}(\gamma) \\
\theta_k|H &\sim H \\
\pi_j|\alpha, \beta, \kappa &\sim \text{DP}(\alpha + \kappa, \tfrac{\alpha\beta + \kappa\delta_j}{\alpha + \kappa}) \\
z_t|z_{t-1}, \{\pi_j\}_{j=1}^{\infty} &\sim \pi_{z_{t-1}} \\
x_t|z_t, \{\theta_j\}_{j=1}^{\infty} &\sim F(x_t|\theta_{z_t})
\end{aligned}
\tag{3}
$$

In terms of the human dynamics monitoring, since it is considered that hidden states behind human dynamics gradually transit through time, setting the accurate expected probability of tself-transition can lead to the model with high predictive performance. Therefore, we consider that the adequate number of states behind human dynamics data can probably be estimated by using a sHDP-HMM for the an anomaly detection model.

3.2 THE PROPOSED ANOMALY DETECTION METHOD

The proposed method consists of the following three steps: It firstly learns hidden state that represent normal states sequence and the hyper-parameters from training data (step 1). Secondly, the sHDP-HMM with the learned hyper-parameters infers hidden states of test data (step 2). By comparing the two estimated states for each instant of time, an anomaly will finally be detected if they are unequal (step 3).

In the step 1, a "normal" training temporal data $\{x_t\}_{t=1}^{T}$ is set as

the observation variables of a sHDP-HMM, where T indicates the length of d dimensional temporal data. We use d dimensional Gaussian distribution $\text{N}(\mu_k, \Sigma_k)$ for the likelihood distribution of the kth hidden state, uniquely defined by parameter $\theta_k = (\mu_k, \Sigma_k)$. Also, a Gaussian distribution $\text{N}(\mu_0, \Sigma_0)$ and a Inverse-Wishart distribution $\text{IW}(\nu, \Delta)$ are used for the prior distribution of the mean μ_k and the variance-covariance matrix Σ_k each. Since hyper-parameters are unknown, we give vague prior distributions and try infer them by sampling algorithm. For the sHDP-HMM explained above, the Blocked Gibbs Sampler is employed to infer all of the hyper-parameters and the parameters including the state sequence $\{z_t\}_{t=1}^{T}$.

The step 2 applies the inferred sHDP-HMM to training temporal data $\{x_t'\}_{t=1}^{T}$ and estimates its hidden state sequence $\{z_t'\}_{t=1}^{T}$ which may contain anomalies. With respect to the parameter setting, the step 2 uses the expected values which are calculated from posterior distributions of the hyper-parameters sampled the in the step 1. Moreover, the initialization of the parameters such as β, π, θ is given not by new sampled values, but by using the last ones in step 1.

Finally in the step 3, the normal state sequence $\{z_t\}_{t=1}^{T}$ and the test state sequence $\{z_t'\}_{t=1}^{T}$ are compared. To handle the problem that the indicators assigned to each state can be exchanged due to re-estimation of β, π, θ, the proposed method employs the Hungarian Algorithm to solve assignment problem of which cost function is given by Hamming distance between the normal and test state sequences. After the state assignment, an anomaly at time t is detected if $z_t \neq z_t'$.

Based on the characteristics of anomaly detection problem defined in section 2, this proposed method is categorized to a semi-supervised anomaly detection technique that uses normal time-series data and detects contextual anomalies with output of anomaly labels. This proposed method especially aims to detect contextual anomalies that are anomalous at specific time, but normal otherwise.

4 EXPERIMENTS AND RESULTS

In this section, we apply our proposed anomaly detection method to both synthetic and real data, and discuss the fundamental performance. In implementing the inference algorithm, we use a MATLAB toolbox supporting several inference algorithms based on Gibbs sampling for the sticky HDP-HMM, which has been made available by Fox (Fox et al., 2009).

4.1 Experiment with synthetic data

For an experiment with synthetic data, we generate 1 dimensional training and test time-series data with length of $T = 400$, depicted in Figure 5(a). We manually set twelve states represented by the numbers in Figure 5 and obtain i.i.d. samples from Gaussian distribution defined for each state. This synthetic temporal data imitates an average one-day pattern of a real gridded population data discussed later, by adjusting means and variances of Gaussian distributions. The test data, which is also sampled from the same distributions, is manually added anomalous values on 10% of the test data as follows; (i)5σ anomaly at $t = 46$, (ii)3σ anomalies at $t = 98 \sim 108$, (iii)3σ anomalies at $t = 196 \sim 210$, (iv)3σ anomalies at $t = 311 \sim 328$ and (v)10σ anomaly at $t = 381$.

We used 1 dimensional Gaussian emission distribution, and placed priors $N(\mu_0, \Sigma_0)$ and $IW(\nu, \Delta)$ on the space of mean and variance parameters. The μ_0 and Σ_0 are given by empirical mean and variance, degree of freedom ν is set as 4 and $\Delta = I$ where I indicates the identity matrix. To realize a steady inference, we normalize the input data to make its mean be 0 and variance be 1.

The upper side of Figure 5(b) depicts the learned state sequence $\{z_t\}_{t=1}^{T}$ where green solid and dotted lines represent the estimated mean and standard variation of each state. The estimated state sequence is also visualized. Although we have initially set twelve states, the final 10,000th sampling gave six states because a couple of states were unified into one common state. Yet, significant differences between the real Gaussian parameters and the estimated parameters were not found.

After the inference of the test data in the step 2, hidden states and their Gaussian parameters were estimated as illustrated in the lower side of Figure 5(b). The number of the hidden states, which was six at the time of training, was estimated seven. Besides, our anomaly detection method regarded 64 points as anomaly pointed by red circles in the figure. Table 4 shows a confusion matrix in which a result of this anomaly detection experiment is listed. The fact that results scored 80.0% of the detection rate shows that our proposed method has good fundamental performance of anomaly detection for pseudo gridded population data with respect to the detection rate. On the other hand, the precision was 50.0% because a number of anomalies including false positives were often detected near borders of states.

Table 4: Confusion matrix of the result with the synthetic data

		Result of anomaly detection	
		Anomaly 64	Normal 336
True	Anomaly 40	True Positive 32	False Negative 8
	Normal 360	False Positive 32	True Negative 328

4.2 Experiment with real data

As a real gridded population data for this experiment, we use 500m-1hour resolved population of 30 October 2012 and 6 November 2012 in Tokyo, provided by ZENRIN DataCom Co., Ltd. Because a serious accident on the Tobu Isesaki Line has been recognized at about 6 am on 6 November to end up an hour-and-a-half service interruption between Asakusa Sta. and Kita-Senju Sta., we use the data of Octber 30 for training and the another for test. Especially, we target on a grid where Asakusa Sta. is located (Figure 6) because it was probably considerably impacted
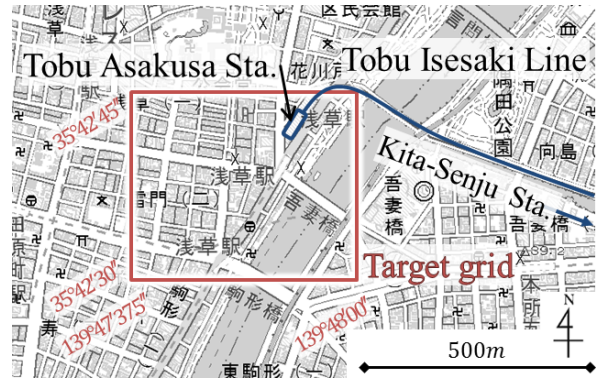

Figure 6: Location of the target grid

by the accident. We employ our proposed anomaly method with the normalized temporal data for observation variables and the same sHDP-HMM setting as before.

The upper side of Figure 7 illustrates a result of the inference of training data and shows that the number of normal states was estimated three, each of which stands for morning, afternoon and evening. After the test-data inference, new two states have been assigned for data in morning and evening, and then, two anomalies were detected as pointed by red circles in the bottom of Figure 7. This result shows that an anomaly was detected at the time when train services had been stopped. On the other hand, while the proposed method detected the another anomaly in evening, there is not recognized serious events such as traffic accidents or social events around the target area. An extensive analysis is required on the reason of this detection, whether it is result of latent anomaly or a false alarm.
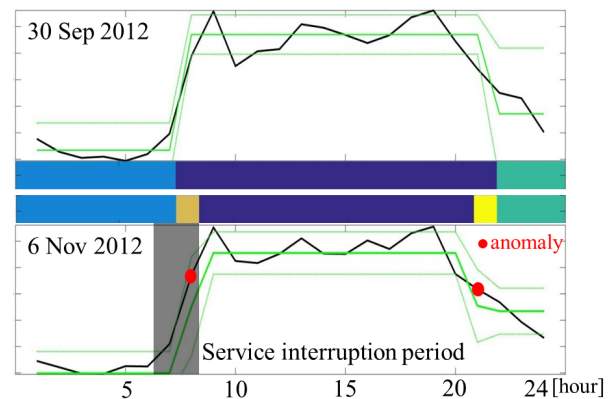

Figure 7: learning result of training data(upper), anomaly detection result(lower) and compared states(middle)

5 CONCLUSIONS

In this study, we firstly defined an anomaly detection problem of the human dynamics monitoring with respect to gridded population data. Based on the result of review we have comprehensively done, we discussed how the anomaly detection of human dynamics monitoring can be interpreted in terms of the four factors; nature of data, output of anomaly detection, data labels and type of anomalies. Additionally, we gave an explanation that state-space model, especially a HDP-HMM is considered possible to apply due to its property that a HDP-HMM can estimate the number of hidden states according to input data. Besides, we developed an anomaly detection method based on a sticky HDP-HMM, which
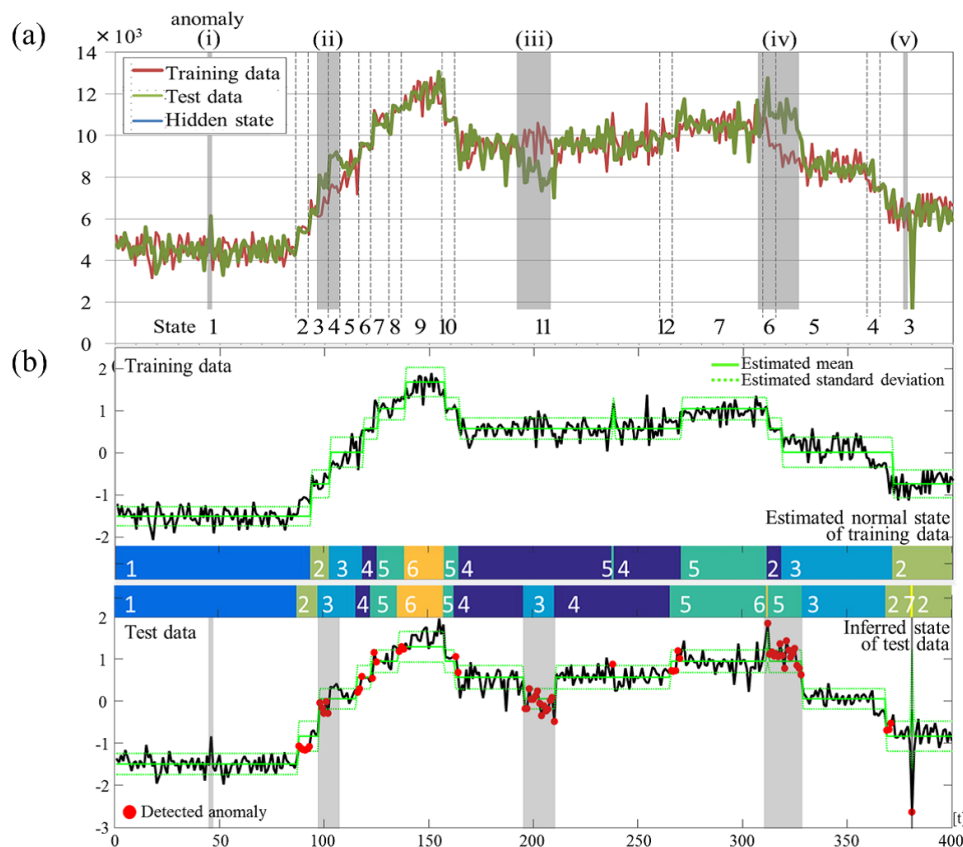
Figure 5: Synthetic training / test time-series data(a), inference and anomaly detection results(d)

is categorized to a semi-supervised anomaly detection technique that uses normal time-series data and detects contextual anomalies with output of anomaly labels. Results of the experiment with synthetic data showed that our proposed method has good fundamental performance with respect to the detection rate. Through the experiment with real gridded population data, an anomaly was detected at the location where a station was located and the time when train services had been stopped.

Future works include the improvement of the precision of the proposed method, the application of the proposed method to a bigger dataset of different gridded population data, and the deepening the interpretation of the estimated states and the detected anomalies on the traffic networks. The final goal of this research is to construct a dynamic traffic control model by integrating the spatial extension of the proposed method, the on-line anomaly detection method and the control theory.

REFERENCES

Candia, J., González, M. C., Wang, P., Schoenharl, T., Madey, G. and Barabási, a. L., 2008. Uncovering individual and collective human dynamics from mobile phone records. Journal of Physics A: Mathematical and Theoretical 41(22), pp. 1–11.

Chandola, V., Banerjee, A. and Kumar, V., 2009. Anomaly detection: a survey. ACM Computing Surveys 41(3), pp. 15:1–15:58.

Fox, E. B., Sudderth, E. B., Jordan, M. I. and Willsky, A. S., 2009. Nonparametric bayesian learning of switching dynamical systems. Advances in Neural Information Processing Systems 21, pp. 457–464.

Hodge, V. J. and Austin, J., 2004. A survey of outlier detection methodologies. Artificial Intelligence Review 22(2), pp. 85–126.

Horanont, T., 2010. A study on urban mobility and dynamic population estimation by using aggregate mobile phone sources. CSIS Discussion Paper.

Horiguchi, R., Iijima, M., Kobayashi, M. and Hanabusa, H., 2013. Traffic anomaly detection for surface street networks with the mesh-wised traffic indices on macroscopic fundamental diagram. OPTIMUM 2013 International Symposium on Recent Advances in Transport Modelling pp. 1–7.

Lello, E. D., Laet, T. D. and Bruyninckx, H., 2012. Hierarchical dirichlet process hidden markov models for abnormality detection in robotic assembly. Workshop on Bayesian Nonparametric Models (BNPM) For Reliable Planning And Decision-Making Under Uncertainty pp. 1–8.

Pan, B., Zheng, Y., Wilkie, D. and Shahabi, C., 2013. Crowd sensing of traffic anomalies based on human mobility and social media. Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems pp. 344–353.

Pimentel, M. A., Clifton, D. A., Clifton, L. and Tarassenko, L., 2014. A review of novelty detection. Signal Processing 99, pp. 215–249.

Teh, Y. W., Jordan, M. I., Beal, M. J. and Blei, D. M., 2006. Hierarchical dirichlet processes. Journal of the american statistical association 101(476), pp. 1–30.