A Variational Autoencoder Based Generative Model of Urban Human Mobility

Dou Huang*, Xuan Song*†, Zipei Fan*, Renhe Jiang*†, Ryosuke Shibasaki*, Yu Zhang‡, Haizhong Wang§, Yugo Kato¶
*Center for Spatial Information Science, The University of Tokyo
†Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology
‡Department of Computer Science and Engineering, Hong Kong University of Science and Technology
§Civil and Construction Engineering, Oregon State University
¶Transport Consulting Division, NAVITIME JAPAN Co., Ltd
{huangd, songxuan, jiangrh, shiba}@csis.u-tokyo.ac.jp, fanzipei@iis.u-tokyo.ac.jp
yuzhangcse@cse.ust.hk; Haizhong.Wang@oregonstate.edu

***Abstract*—Recently, big and heterogeneous human mobility data inspires many revolutionary ideas of implementing machine learning algorithms for solving some traditional social issues, such as zone regulation, air pollution, and disaster evacuation el at.. However, incomplete datasets were provided owing to both the concerns of violation of privacy and some technique issues in many practical applications, which leads to some limitations of the utility of collected data. Variational Autoencoder (VAE), which uses a well-constructed latent space to capture salient features of the training data, shows a significant excellent performance in not only image processing, but also Natural Language Processing domain. By combining VAE and sequence-to-sequence (seq2seq) model, a Sequential Variational Autoencoder (SVAE) is built for the task of human mobility reconstruction. It is the first time that this kind of SVAE model is implemented for solving the issues about human mobility reconstruction. We use navigation GPS data of selected greater Tokyo area to evaluate the performance of the SVAE model. Experimental results demonstrate that the SVAE model can efficiently capture the salient features of human mobility data and generate more reasonable trajectories.**

***Keywords*-human mobility; generative model; meachine learning**

I. Introduction

Many big cities have grown thanks for the rapid urbanization progress, which have modernized many people's lives but also engendered big challenges[1]. Years ago, solving this kind of challenges seems to be impossible because of the complex and dynamic settings of cities. Nowadays, some impressive methods of locational datasets collection have shown an opportunity for the human mobility applications. Although the usage of those kinds of datasets, which owned by enterprises or government, can give us opportunities to some potential applications, they have some limitations in two-fold: 1) it has the risk of a violation of privacy in some cases if used directly; 2) it will contain some bias or the sampling rate is low.

In general, a generative model is a model of the conditional

[1]NAVITIME JAPAN is a private company developing navigation technologies and providing various kinds of web application services such as route navigation, travel guidance, and other useful information services for moving people.

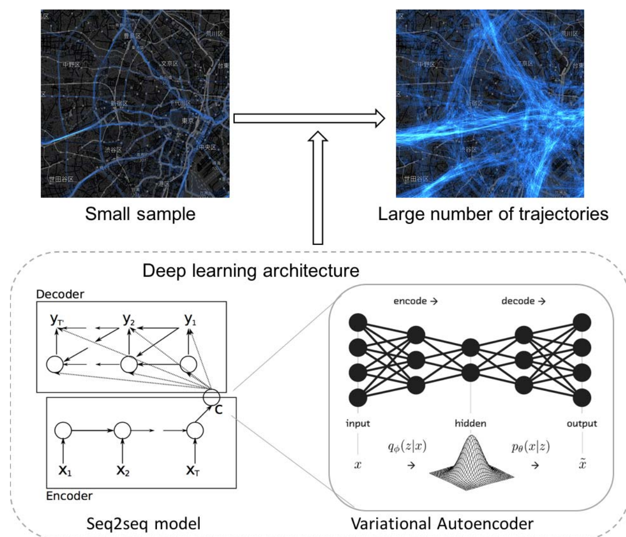

Figure 1. Limited datasets were usually provided owing to the risk of violation of privacy. How can we scale up the limited dataset to meet the requirements of better visualization or other further analysis such as traffic demand prediction? A Variational autoencoder based deep learning architecture provide us an opportunity to generate large number of trajectories which following the same probabilistic features learned from exited small samples.

probability of the observable X, given a target y, symbolically, $P(X|Y=y)$[2]. It can be used to generate random outcomes, either of an observation and target (x, y), or of an observation x given a target value y.

First of all, a generative model can learn a low dimensional feature space which can infer the travelers pattern from the complex redundant collected locational datasets. Then, we can utilize the learned feature space to transportation planning and applications. And if necessary, we can resample from the learned low dimensional feature space to generate a fake dataset that has the similar pattern with the real dataset for further use. There are two reasons for generating fake datasets: 1) using generated fake datasets can avoid the risk of invasion of customers privacy; 2) obtain enough data samples if the dataset is too small to be used.

978-1-7281-1198-8/19/$31.00 ©2019 IEEE
DOI 10.1109/MIPR.2019.00086

In this paper, we combine Variational autoencoder and sequence-to-sequence model to build a Sequential Variational Autoencoders (SVAE) model for tackling trajectories of human mobility. In the model architecture we focus on learning the hidden space which using multivariate Gaussian distribution to approximate the real posterior distribution of the real human mobility. Then we can resample from the learned distribution, and using a learned decoder to reconstruct the trajectories of human mobility. Our contributions are as follows:

1) In our limited knowledge, it is the first time to implement the SVAE for trajectories of human mobility.

2) Diversity. We can obtain some reasonable trajectories, which are not contained in the original data set, to achieve diversity.

3) We use some metrics to quantitatively evaluate the performance of SVAE model for trajectories of human mobility.

We use real navigation GPS data to conduct the experiment. The data we used is GPS navigation data contains trajectories of human mobility of entire Japan.

II. Related works

A definition of Urban Computing is given by Y. Zheng[1]: it is a process of tackling the major issues which cities face using big and heterogeneous data collected by a diversity of sources in urban areas.

Techniques of data collection have been improved rapidly, which lead to some revolutionary ideas of implementing machine learning algorithms for solving some traditional social issues, such as zone regulation[3], air pollution[4], disaster evacuation[5] el at., since collected big and heterogeneous data makes tasks which are nearly impossible years ago become possible.

Recently, there are many researches conducted on human mobility data, such as mobile phone GPS log data, taxi GPS data and navigation GPS data. These kinds of researches often related with building intelligent city system[6], [7], [8], [9], [10], [11], [12], [13].

Meanwhile in recent years, Variational Autoencoders (VAEs) have been widely used for approximate some complicated distributions[14]. The ability of VAEs has been proved to be promise in the works of generating many kinds of complicated data in image processing domain. However, some researches[15], [16], [17] also using this framework in other domains such as Natural Language Processing (NLP), which inspired its implementation for tackling issues based on sequential data.

Input-Output Hidden Markov Model (IO-HMM)[18] was proposed to enable activity based travel demand models which can protect the privacy of mobile phone users while using this cellular data to simulate synthetic agent travel patterns. Their models achieve a reasonable accuracy when conducting an agent-based microscopic traffic simulation. To improve the HMMs, Baratchi et al.[19] proposed Hidden Semi-Markov Model(HSMM), which including the duration of the state into the hidden variables. In general, their works are all based on Hidden Markov Model, and focus on reconstruct the trajectories of human mobility following a specific probability distribution.

Very recently, a non-Parametric generative model for human trajectories has been proposed[20]. They use Generative Adversarial Network (GAN) to produce data points after a simple and intuitive yet effective embedding for locations traces designed. It is the first time that deep learning methods implemented in building a generative model for human mobility in our knowledge.

The Sequential Variational Autoencoder we build in this research has significant differences comparing with their model. Their work is a GAN based model which aims to generate fake data that can be recognized as true data by the trained discriminator. While the model described in this research aims to learn the approximated latent distribution of training data first, then resample the fake data from this learned latent space. Besides, there is no need of trajectory transformation for trajectories when using our model.

III. Methodology

A. Variational Bayesian

Variational Bayes is a particular variational method which aims to find some approximate joint distribution $Q(x,\theta)$ over hidden variables x to approximate the true joint $P(x)$, and defines the distance as the Kullback-Leibler divergence $KL(Q(x,\theta)||P(x))$.[21] Kingma, Diederik P., and Max Welling[22] introduce a stochastic variational inference and learning algorithm that scales to large datasets and, under some mild differentiability conditions, even works in the intractable case, and propose a Variational Autoencoder framework, which is widely used in recent years.

In variational inference, a Gaussian distribution is used for approximating the real posterior distribution $p(z|x)$. We use $q_\lambda(z|x)$ denote the approximate distribution. As it is a Gaussian distribution, the latent variables can be given by the mean and variance $\lambda_{x_i} = (\mu_{x_i}, \sigma^2_{x_i})$. Then the problem is that how can we measure the difference between the real posterior distribution $p(z|x)$ and the approximate distribution $q_\lambda(z|x)$. In information theory, the Kullback-Leibler divergence is often used for solving such problem:

$$KL(q_\lambda(z|x)||p(z|x)) =$$
$$E_q[logq_\lambda(z|x)] - E_q[logp(x,z)] + logp(x)$$

By minimizing this divergence respect to the parameters λ, we can find the optimal approximate distribution. This process can be written as follows:

$$q^*_\lambda(z|x) = argmin_\lambda KL(q_\lambda(z|x)||p(z|x))$$

However, aforementioned Kullback-Leibler divergence cannot calculated directly since the evidence $p(x)$ appears.

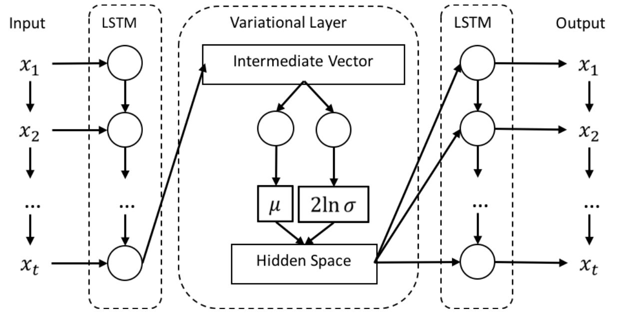

Figure 2. A graphical model of proposed Sequential Variational Autoencoder.

To tackle this problem, the solution comes to the Evidence Lower Bound (ELBO):

$$ELBO(\lambda) = E_q[logp(x, z)] - E_q[logq_\lambda(z|x)]$$

Combining the above Kullback-Leibler divergence and this ELBO function, we can get the formula for the evidence:

$$logp(x) = ELBO(\lambda) + KL(q_\lambda(z|x)||p(z|x))$$

According to the Jensen's inequality, the result of Kullback-Leibler divergence between two probability distribution always greater than or equal to zero, and zero can be achieved only when these two probability distribution are the same. Since the information of the evidence is a constant value, we can know that minimizing the Kullback-Leibler divergence is equivalent to maximizing the ELBO.

B. Recurrent Neural Networks

Deep neural networks that are mainly feedforward fully connected neural network are powerful but not really appropriate for sequential data such as time series data or language. They are very good to map input data to discrete output or continuous variable but not sequence to sequence mapping. Sequence-to-sequence (Seq2seq) model learns from variable sequence input fixed length sequence output. It uses two Long Short Term Memory (LSTM) model, one learns vector representation from input sequence of fixed dimensionality, and another LSTM learns to decode from this input vector to target sequence. LSTM is a variant of recurrent neural network that solves problem of handling long sequences using different gates. Seq2seq model was recently proposed, and demonstrated excellent result for Natural Language Processing (NLP)[23], [24], [25]. This model proved to be more effective than previous methods at NMT, and is apparently now used by Google Translate.

Long-Short-Term Memory (LSTM) was designed to combat vanishing gradients through a gating mechanism[26]. The ability of modeling long-term dependencies is improved in LSTMs thanks to the gating mechanism.

C. Sequential Variational Autoencoder

A Variational Autoencoder can build a hidden space which follow Gaussian distribution to approximate the real distribution of the observed trajectories. The reason for constructing a hidden space which follows a Gaussian distribution is that by learning the parameters of the Gaussian distribution representing the input observed trajectories, we can sample from the distribution and generate new samples of trajectory. The ability for constructing hidden space following a Gaussian distribution is exactly what we want in our Sequential Variational Autoencoder. However, the Variational Autoencoder lack the ability of tackling sequential data, which is the main limitation.

A seq2seq model framework usually use several Recurrent Neural Networks as encoder and decoder. Therefore, a seq2seq model can handle sequential data without difficulties, but the hidden space C is not well constructed.

We can regard the seq2seq model as a Sequential Autoencoder. By doing that, it is natural to consider that if we combine Variational Autoencoder and seq2seq model, we can combine their advantages. That means the Sequential Variational Autoencoder is well-designed generative model for sequential data.

Let $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_t})$ denote a high dimensional sequence, such as a trajectory of human mobility with t steps. We use a LSTM neural network as recurrent encoder to capture the information of the input trajectory $\mathbf{x}$. Then we will obtain a series of hidden state s_t, and a series of output o_t. In actual case, what we really care about is the final output o rather than a sequence of output value o_t. Since we only keep the final output, we can obtain an intermediate non-sequential vector o to represent the information captured from the input sequence using this recurrent encoder.

After intermediate vector o is obtained, we treat this vector as the input of the Variational Autoencoder part. Then we can write the joint probability of the model as $p(o, z) = p(o|z)p(z)$. In the SVAE model, we parametrize approximate posterior $q_\theta(z|o)$ using an inference network, approximate likelihood $p_\phi(o|z)$ using a generative network. Then the loss of the model will be:

$$loss = -E_{q_\theta(z|o)}[logp_\phi(o|z)] + KL(q_\theta(z|o)||p(z))$$

Finally, we use another LSTM neural network as recurrent decoder to reconstruct the trajectories of human mobility, $\mathbf{x}$ from parameters in learned latent distribution.

IV. Experiment

A. Data description

The data we used for this research is navigation GPS data, which is collected when vehicles were using navigation application. The coordinate system of this GPS data is WGS84, and the records of the locations cover all over Japan. However, owing to some reasons, such as privacy

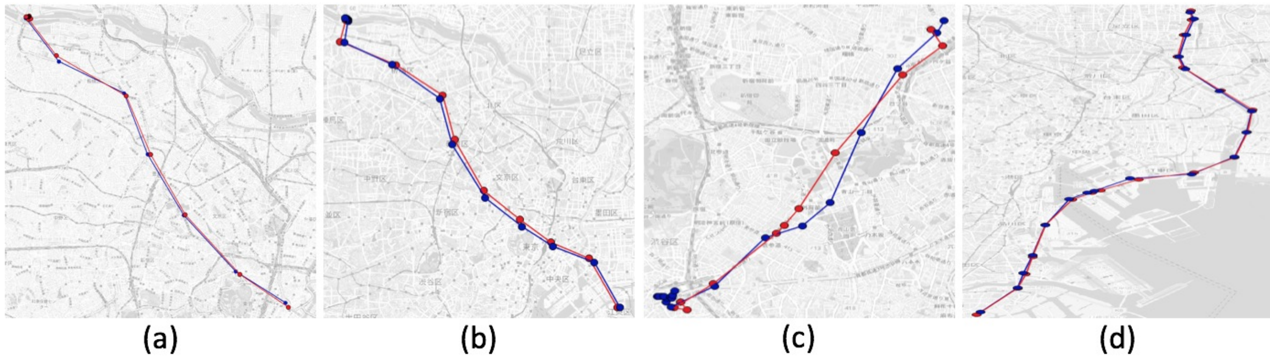

Figure 3. Visualization of four original trajectory samples and their corresponding reconstructed trajectories. Red lines are original trajectories, and blue lines are reconstructed trajectories.

Table I
PERFORMANCE OF SVAE WITH DIFFERENT COEFFICIENT OF LOSS FUNCTION

Coefficient	0.00000	0.00001	0.00005	0.00010	0.00050	0.00100	0.00500	0.01000	0.05000	0.10000
KL divergence	134.05	69.15	42.69	33.84	16.46	13.61	8.92	7.62	4.76	3.56
MDE(/m)	294.99	363.79	336.69	429.37	471.75	538.53	852.36	1109.39	2073.37	2902.62

Figure 4. Reconstruction error(m) of each step between original trajectories and reconstructed trajectories.

protection, we can only use one month records which is from Oct 1, 2015 to Oct 31, 2015. Besides, the ID of the users were deleted, so the privacy is protected well. We can only get the information of the ID of each navigation route to distinguish different trajectories.
However, the raw data should be preprocessed before the training process. The data preprocessing of linear interpolation is done to simplify the input data, by forcing the trajectories have fixed timestamp. Therefore, the input only contains information about longitude and latitude but can still represent the dynamics of the trajectories.

B. Results and Evaluation

We use aforementioned navigation GPS data to conduct experiment. Mean Distance Error (MDE) between real trajectories and generated trajectories is used for evaluating the trajectory reconstruction error and the Kullback-Leibler divergence is used for evaluating the diversity of the generated trajectories. We test the performance of the SVAE model with different parameter settings:
(1) The dimensionality of hidden space is set to be 8, 12, 16 respectively to test the performance of the model for different dimensionality of hidden space;
(2) Three kinds of input (values of coordinate only, grid ID only and combination input of values of coordinate and grid ID) are tested.

$$MDE = \frac{\sum_i^N (dis(x_i, \hat{x}_i))}{N}$$

$$D_{KL} = \frac{\sum_i^N (KL_i(q_\theta(z|o)||p(z)))}{N}$$

where dis is haversine distance between two points, x_i is the i-th real trajectory, $\hat{x}_i$ is i-th reconstructed trajectory. $q_\theta(z|o)$ is the approximate Gaussian distribution learned from training data, $p(z)$ is the unit Gaussian distribution.
We firstly give four randomly selected pairs of original trajectories and reconstructed trajectories to show the SVAE's ability to reconstruct human mobility trajectories in figure 3. Meanwhile, we calculate the distance error of each step of trajectory pairs shown in figure 4.
These results show that the SVAE model can reconstruct the trajectories with an acceptable accuracy, namely under 500 meters in about 33,000 x 36,000 square meters area in Tokyo. In that case, we think the SVAE model is suitable for human mobility trajectory generation work.
We also give the results about the performance of the SVAE

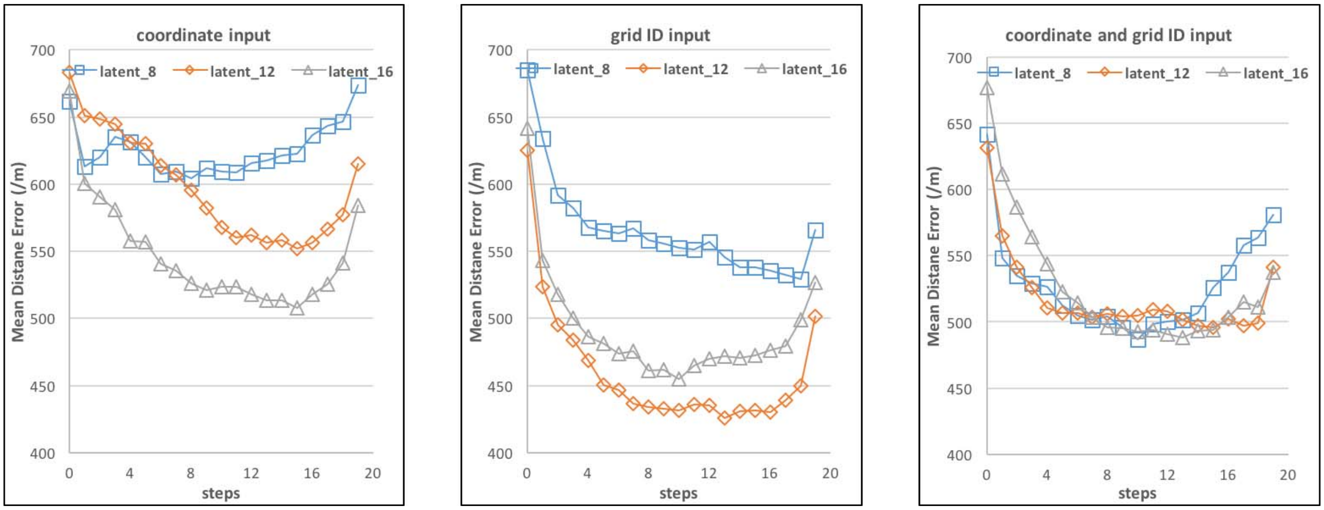

Figure 5. Mean Distance Error of results with different parameters. We made comparison with different latent dimensionality (8, 12, 16) and different input (grid ID, coordinate, both coordinate and grid ID)

model with different parameters. In figure 5, we aim to compare the performance of different input strategies. From the figure, we can say that just using coordinate values as inputs of the SVAE model is not a good choice in most cases. However, we cannot point out the advantage of using grid ID as inputs in this figure.

In general, the results of evaluation using MDE shows that the reconstruction error of SVAE model is smaller than 800 meters. Actually, there is a trade-off between the accuracy of the reconstruction trajectories and the robustness of the ability to generate resampled trajectories. As mentioned before, the loss function of the SVAE model is consist of reconstruction error and Kullback-Leibler divergence. In practical training process, minimizing the reconstruction error will increase the accuracy of reconstructing input trajectories, while minimizing the Kullback-Leibler divergence will reduce the complexity of learned latent space.

In table I, we give a quantitatively comparison of Mean Distance Error and Kullback-Leibler divergence using different coefficient w. The result shows that when the coefficient w is large, the KL-divergence is small while the Mean Distance Error is large, and vice versa. It means that we can achieve high diversity using a large coefficient w, while it will lose some reconstruction accuracy.

V. Conclusions

In this research, we make a brief introduction about Variational Autoencoder and Sequence-to-sequence model, then combine these two frameworks to build a Sequential Variational Autoencoder. It is believed that this Sequential Variational Autoencoder is first time implemented in modeling trajectories of human mobility. We use navigation GPS data of cars in Tokyo to evaluate the performance of SVAE model. The performance of SVAE with different parameter settings and its explanation have been discussed. In general, the SVAE model can capture the salient features of input trajectories using a latent space constructed by following Gaussian distribution, then reconstruct the input trajectories. As a generative model, the ability of generating fake resampled trajectories of SVAE is also proved. Using this SVAE model, we can generate more trajectories of human mobility which have similar pattern with training data to solve the low sampling rate problem since the generated trajectories show some diversity.

We also note the limitation of SVAE model when implemented in trajectories of human mobility, which is that many points of reconstructed trajectories is not located in road network. Implementing map matching to the generated trajectories may solve the problem, but we believe a better choice is that change the current coordinate and grid based model to a node based model. Another idea for this problem is changing the current resampling from Gaussian distribution strategy to resampling from historical trajectories.

VI. Acknowledgments

This work was partially supported by Leading Initiative for Excellent Young Researchers (LEADER) Program and Grant in-Aid for Scientific Research B (17H01784) of Japans Ministry of Education, Culture, Sports, Science, and Technology(MEXT); and JST, Strategic International Collaborative Research Program (SICORP).

References

[1] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 3, p. 38, 2014.

[2] S. PERMISSION, "Generative and discriminative classifiers: Naive bayes and logistic regression," 2005.

[3] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and pois," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 186–194, ACM, 2012.

[4] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1436–1444, ACM, 2013.

[5] X. Song, Q. Zhang, Y. Sekimoto, T. Horanont, S. Ueyama, and R. Shibasaki, "Modeling and probabilistic reasoning of population evacuation during large-scale disaster," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1231–1239, ACM, 2013.

[6] R. Jiang, X. Song, Z. Fan, T. Xia, Q. Chen, Q. Chen, and R. Shibasaki, "Deep roi-based modeling for urban human mobility prediction," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, p. 14, 2018.

[7] Z. Fan, X. Song, R. Shibasaki, and R. Adachi, "Citymomentum: an online approach for crowd behavior prediction at a citywide level," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 559–569, ACM, 2015.

[8] Y. Zheng, Y. Liu, J. Yuan, and X. Xie, "Urban computing with taxicabs," in *Proceedings of the 13th international conference on Ubiquitous computing*, pp. 89–98, ACM, 2011.

[9] Z. Fan, X. Song, R. Shibasaki, T. Li, and H. Kaneda, "Citycoupling: bridging intercity human mobility," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 718–728, ACM, 2016.

[10] X. Song, Q. Zhang, Y. Sekimoto, and R. Shibasaki, "Intelligent system for urban emergency management during large-scale disaster.," in *AAAI*, pp. 458–464, 2014.

[11] X. Song, Q. Zhang, Y. Sekimoto, and R. Shibasaki, "Prediction of human emergency behavior and their mobility following large-scale disaster," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 5–14, ACM, 2014.

[12] X. Song, Q. Zhang, Y. Sekimoto, R. Shibasaki, N. J. Yuan, and X. Xie, "A simulator of human emergency mobility following disasters: Knowledge transfer from big disaster data.," in *AAAI*, pp. 730–736, 2015.

[13] Q. Chen, X. Song, H. Yamada, and R. Shibasaki, "Learning deep representation from big and heterogeneous data for traffic accident inference.," in *AAAI*, pp. 338–344, 2016.

[14] C. Doersch, "Tutorial on variational autoencoders," *arXiv preprint arXiv:1606.05908*, 2016.

[15] Y. Fan, G. Wen, D. Li, S. Qiu, and M. D. Levine, "Video anomaly detection and localization via gaussian mixture fully convolutional variational autoencoder," *arXiv preprint arXiv:1805.11223*, 2018.

[16] J. Mueller, D. Gifford, and T. Jaakkola, "Sequence to better sequence: continuous revision of combinatorial structures," in *International Conference on Machine Learning*, pp. 2536–2544, 2017.

[17] S. Asakawa, "Comparison between variational autoencoder and encoder-decoder models for short conversation," 2017.

[18] M. Yin, M. Sheehan, S. Feygin, J.-F. Paiement, and A. Pozdnoukhov, "A generative model of urban activities from cellular data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 6, pp. 1682–1696, 2018.

[19] M. Baratchi, N. Meratnia, P. J. Havinga, A. K. Skidmore, and B. A. Toxopeus, "A hierarchical hidden semi-markov model for modeling mobility data," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 401–412, ACM, 2014.

[20] K. Ouyang, R. Shokri, D. S. Rosenblum, and W. Yang, "A non-parametric generative model for human trajectories,"

[21] C. W. Fox and S. J. Roberts, "A tutorial on variational bayesian inference," *Artificial intelligence review*, vol. 38, no. 2, pp. 85–95, 2012.

[22] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[23] R. J. Weiss, J. Chorowski, N. Jaitly, Y. Wu, and Z. Chen, "Sequence-to-sequence models can directly translate foreign speech," *arXiv preprint arXiv:1703.08581*, 2017.

[24] T. Liu, K. Wang, L. Sha, B. Chang, and Z. Sui, "Table-to-text generation by structure-aware seq2seq learning," *arXiv preprint arXiv:1711.09724*, 2017.

[25] S. Gu and F. Lang, "A chinese text corrector based on seq2seq model," in *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2017 International Conference on*, pp. 322–325, IEEE, 2017.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.