USING GANS FOR GENERATION OF REALISTIC CITY-SCALE RIDE SHARING/HAILING DATA SETS

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper focuses on the synthetic generation of human mobility data in urban areas. We present a novel and scalable application of Generative Adversarial Networks (GANs) for modeling and generating human mobility data. We leverage actual ride requests from ride sharing/hailing services from four major cities in the US to train our GANs model. Our model captures the spatial and temporal variability of the ride-request patterns observed for all four cities on any typical day and over any typical week. Previous works have succinctly characterized the spatial and temporal properties of human mobility data sets using the *fractal dimensionality* and the *densification power law*, respectively, which we utilize to validate our GANs-generated synthetic data sets. Such synthetic data sets can avoid privacy concerns and be extremely useful for researchers and policy makers on urban mobility and intelligent transportation.

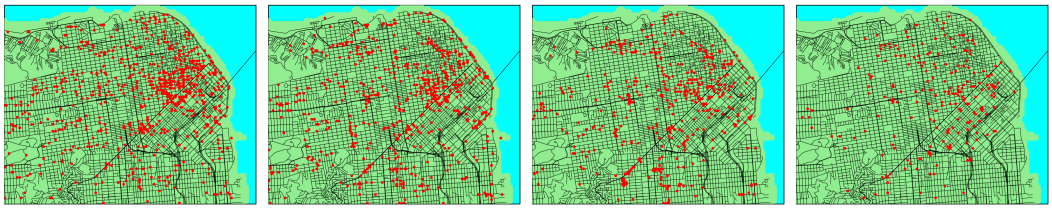
1 INTRODUCTION

Ride sharing or hailing services have disrupted urban transportation in hundreds of cities around the globe (Times, 2018; Cohen & Kietzmann, 2014). In United States, it has been estimated that between 24% to 43% of the population have used ride-sharing services in 2018 (Recode, 2018a). Uber alone operates in more than 600 cities around the globe (Recode, 2018b). Ride sharing services have turned urban transportation into a convenient utility (available any place at any time), and become an important part of the economy in large urban areas (Hahn & Metcalfe, 2017).

Ride request data from ride sharing services can potentially be of great value. Data gathered from ride sharing services could be used to provide insights about traffic and human mobility patterns which are essential for intelligent transportation systems. Ride requests in major cities with high penetration by such services exhibit spatial and temporal variability. Modeling of such variability is a challenging problem for researchers. Moreover, there are still unresolved challenges, such as: optimal algorithms for dynamic pooling of ride requests (Chen et al., 2017), real-time preplacement of vehicles (Jauhri et al., 2017b), and city scale traffic congestion prediction (Maystre & Grossglauser, 2016) and avoidance. Access to large amount of actual ride request data is essential to understanding and addressing these challenges.

Data from ride sharing services have been used for real-time sensing and analytics to yield insights on human mobility patterns (Song et al., 2010; Jauhri et al., 2017a). Each city exhibits a different pattern of urban mobility – there could be cultural or economical factors governing these patterns. If ride sharing services constitute a significant percentage of the riders in a city, can we build models from ride request data to model urban mobility for the whole city and provide societal benefit without compromising personal privacy? This question motivates us to explore the potential of using Generative Adversarial Networks (GANs) to generate synthetic ride request data sets that exhibit very similar attributes as the actual ride request data sets.

This work proposes a novel approach of generating synthetic ride request data sets using GANs. This approach involves viewing ride requests as a (temporal) sequence of (spatial) images of ride request locations. The approach uses GANs to match the properties of the synthetic data sets with that of real ride request data sets. Many recent works using neural networks have looked at demand prediction (Yao et al., 2018b; Zhou et al., 2018) and traffic prediction at intersections (Yao et al., 2018a). In our work, we are looking at generating actual ride requests for both spatially and temporal granular intervals. Also, we compare and validate the spatial and temporal variations of the synthetic



(a) 6pm (b) 9pm (c) 12am (d) 3am

Figure 1: Ride requests for a small region of downtown San Francisco for a typical week day. Each figure shows the aggregated ride requests (red dots) over a period of an hour. Each red dot may represent one or more ride requests. Ride request density varies spatially and temporally.

data sets with the real data sets. In dealing with large amount of data for many cities and long training times for GANs, we develop effective ways to parallelize and scale our GANs training runs using large CPU clusters on AWS. We present our GANs scaling approach and experimental results, and show that significant reduction in training times can be achieved.

2 Data Sets from Ride Sharing Services

In this section, we introduce the actual (real) ride request data sets used for our GANs training and evaluation. We use the real data sets as the initial training data for the discriminator. We also use the real data sets to compare with and validate the GANs generated synthetic data sets.

Our real ride request data sets consist of all the ride requests for an entire week for the four cities. There is a strong repeating pattern from week to week. Hence the week-long data should be quite representative. For all four cities, the ride sharing services have significant penetration. Hence we believe the ride request data sets also reflect the overall urban mobility patterns for these cities.

2.1 Ride Requests

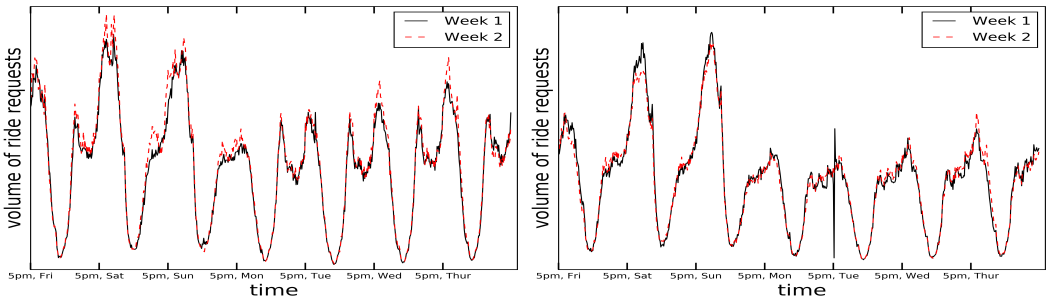
Our real data sets are actual ride requests for four cities over one week period from ride sharing services operating in the United States. Each ride request in a data set includes: request time and location (latitude & longitude), and drop-off time and location (latitude & longitude). For this work we focus only on ride request time and location.

We do some temporal and spatial quantization of the raw ride request data from ride sharing services. We partition the entire week into 2016 time intervals of 5 minutes each, and lump together all the ride requests within each interval. We partition spatially the area of the entire city into small $50m \times 50m$ squares, and lump together all the ride requests occurring within the same square area. Each square area is then represented by a single pixel in a 2-D image with the gray scale intensity of the pixel reflecting the number of ride requests in that square area.

Combining the temporal and spatial quantizations, the real ride request data set for each city becomes a time sequence of 2016 images with each image spatially capturing all the ride requests occurring in a particular 5-min interval. Each pixel within each image represents the presence (or absence) of ride request(s) in a particular $50m \times 50m$ square area of the city in a particular time interval.

2.2 Spatial and Temporal Patterns

The actual ride requests in every city exhibit distinct patterns of variability in both the spatial dimension (over geographical area of the city) and the temporal dimension (over each day and over each week). In Figure 1, this variability is illustrated. The request density is at its highest at 6pm, and continually decreases over time till 3am. Spatially there are dense patches of ride requests and these dense patches can shift with time, reflecting shifting concentrations of commuters in different areas at different times of day. We observe similar repeating patterns of temporal and spatial variability for all four cities.



(a) San Francisco

(b) Los Angeles

Figure 2: Weekly pattern of ride requests.

Previous works have been able to characterize these temporal and spatial variability patterns (Jauhri et al., 2017a). A graph can be used to model the ride requests within a 5-min interval, with nodes representing request and drop off locations and a directed edge connecting the request node and the drop off node of a ride request. It was shown in (Jauhri et al., 2017a) that the size and density of this Ride Request Graph (RRG) evolves in time in response to the fluctuation of ride requests during each day and through out each week; see Figure 2.

It was observed that these ride request graphs obey the *Densification Power Law* (DPL) property, similar to other graphs modeling human behaviors such as social networking graphs and publication citation graphs (Leskovec et al., 2005). It was further observed that the ride request graphs for each city exhibit a distinct degree or *exponent* of the DPL, and that this *DPL Exponent* (α) can be viewed as a very succinct quantitative characterization of the temporal variability of the ride request patterns for that city. For any time snapshot t:

$$e(t) \propto n(t)^{\alpha} \tag{1}$$

where $e(t)$ and $n(t)$ are the number of edges and number of nodes respectively, formed by all ride requests occurring in the time interval t. The number of edges grows according to a specific exponential power (α) of the number of nodes.

There is also a comparable quantitative characterization of the spatial variability of the ride request patterns for each city. *Correlation Fractal Dimension* (Schroeder, 2009; Belussi & Faloutsos, 1998) provides a succinct description of a k-dimensional point-set to provide statistics about the distribution of points; it provides a quantitative measure of self-similarity. The spatial distribution of ride requests in each time interval can be viewed as a point-set image. We can measure the Correlation Fractal Dimension (D_2) as described in (Jauhri et al., 2017b). Values for correlation fractal dimension fall within a range for each city indicating the degree of self-similarity, and the consistent weekly pattern. For our 2-dimenional space, we impose a 2D-grid with cells of side ϵ. For the i-th cell, let $C_{\epsilon,i}$ be the count of requests in each cell. The correlation fractal dimension is defined as:

$$D_2 \equiv \frac{\partial \log \sum_i C_{\epsilon,i}^2}{\partial \log \epsilon} = \text{constant} \qquad \epsilon \in (\epsilon_1, \epsilon_2) \tag{2}$$

For self-similar data sets, we expect the derivative to be constant for a certain range of ϵ (Traina Jr et al., 2010). We observe that this range varies for our four cities, and each city exhibits a distinct value range for its correlation fractal dimension (D_2).

We use the *Densification Power Law Exponent* (α) and the *Correlation Fractal Dimension* (D_2) to capture and characterize the temporal and spatial variability, respectively, for the ride request patterns for each city. We then use these two parameters to confirm the similarity between the real data sets and the GANs generated synthetic data sets. We can claim strong similarity if the values of these two parameters (α and D_2) of the synthetic data sets match closely the values of the same two parameters of the real data sets.

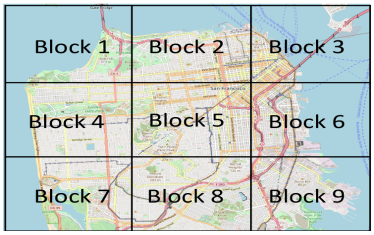


Figure 3: An illustration of how the geographical region of each city is divided into smaller blocks of equal size and trained independently.

3 GENERATING RIDE REQUESTS USING GANS

3.1 IMAGE GENERATING USING GANS

Generative Adversarial Networks learn to generate high quality samples (Goodfellow et al., 2016) i.e. sample from the data distribution $p(x)$. Previous works by (Denton et al., 2015; Ledig et al., 2017) synthesized images of a higher quality using GANs which were hard for humans to distinguish from real images. Conditional GANs are an extension of GANs to sample from a conditional distribution given each image has an associated label which is true for our case of ride requests.

In our framework, we would apply conditional GANs to ride request data in the form of images; similar to as shown in Figure 1 but without the base map shown in color, and a label to generate a series of synthetic ride requests.

3.2 USING GANS FOR RIDE REQUEST GENERATION

GANs learn a mapping from a random noise vector z to output image x. Conditional GANs learn a mapping from noise vector z and a label y to x Mirza & Osindero (2014); Gauthier. The additional variable in the model allows to generate and discriminate samples conditioned on y. The generator accepts noise data z along with y to produce an image. The discriminator accepts an image x and condition y to predict the probability under condition y that x came from the empirical data distribution rather than from the generative model. The objective function can be expressed as:

$$\mathop{\mathbb{E}}_{x,y\sim p_{data}(x,y)}[logD(x,y)] + \mathop{\mathbb{E}}_{z\sim p(z),y\sim p_y}[log(1-D(G(z,y),y))] \tag{3}$$

where G tries to minimize to this objective function against an adversarial D that tries to maximize it.

3.3 TRAINING PROCESS & ARCHITECTURE

Every image is assigned a label from the set $\{0, 1, 2, ..., 23\}$ representing the hour of a day. All twelve instances of five minute snapshots within an hour are assigned the same hour label [1]. To accelerate our training using multiple machines, we exploit spatial parallelism by dividing the entire geographical region of a city into an array of blocks. Figure 3 illustrates the division of San Francisco downtown into nine blocks. We arbitrarily sized each block to represent an image of 24×24 pixels, with each pixel representing one $50m \times 50m$ square area. Hence, each block covers an area of $1200m \times 1200m$.

Each block, representing a grey scale image of 24×24 pixels depicting all the ride requests in that block. Each image of that block is labeled with a time interval (for our experiments, the hour in a day). The synthetically generated images from an array of blocks with the same time interval label will be combined by stitching together once all the blocks are processed.

[1] One could easily extend this approach to a label within the set $\{0, 1, ..., 267\}$ if looking at labels associated with any five minute slots of a day or the set $\{0, 1, ..., 2015\}$ if looking at labels associated with any five minutes slots of a week.

The generator network takes an input of a 100-dimensional Gaussian noise sample as well as a one-hot vector encoding of the time snapshot to be generated. It has a single, fully-connected hidden layer without any convolution (Goodfellow, 2018) consisting of 128 ReLU-activated neurons which then passes to a sigmoid activated output layer with the same number of output neurons as the total number of pixels in each block.

The discriminator network has a single hidden layer of 128 ReLU-activated neurons with a single sigmoid activated output neuron. We find that small networks are appropriate for the training data and allow for a quick and stable convergence to be achieved between the discriminator and the generator. Using relatively simple network architectures makes it possible to ensure that the discriminator and generator are *evenly matched* such that the loss for either network does not saturate early in the training process.

In addition to the standard GANs architecture of generator and discriminator, an additional network is introduced which is referred to as the classifier (Lee & Seok, 2017). This network is pre-trained on the training data with the time sequence of the data serving as the classification target. In this way the time information that is encoded into the synthetic data by the generator network is then decoded by the classifier network. The generator is then trained on a weighted sum of the loss from both the classifier and discriminator networks as shown in the following equation:

$$\beta \log D(G(z, y)) + (1 - \beta) \log C(G(z, y)) \tag{4}$$

where β is a tune-able hyper-parameter.

This allows for more explicit loss attribution such that the generator receives two different error signals; one indicating the realism of the synthetic data and the other indicating accuracy relative to the conditioning values. This is found to increase the efficiency of the training process and result in higher quality synthetic data while incurring considerably less training time than other conditional GANs architectures we have experimented.

4 Experimental Results

In this section, we present the cloud infrastructure used for running our experiments. We also present performance results on scaling our GANs workloads on the cloud infrastructure.

4.1 Running GANs on AWS

All experiments are conducted on Amazon Web Services (AWS) using c5.18x instances with each instance containing an Intel Xeon Scalable Processor with 72 virtual cores (vCores) running at 3.0GHz and 144 GB of RAM. In this work we set the block size for each of the four cities to be 1200×1200 meters; each block is trained separately. Enlarging the block size will increase the computational time for training; and the complexity of the model can potentially impact scalability. The total number of blocks for each city are shown in Table 1. The number of blocks are mainly determined by the size of the greater metropolitan area of each city.

To help enhancing the scalability of our GANs workload across multiple nodes we make use of Ray (Moritz et al., 2017) from Berkeley, a distributed framework for AI Applications, to efficiently parallelize our workload across cluster of CPU nodes on AWS. Ray provides a convenient API in Python to scale deep learning workloads for numerous libraries, and support for heterogeneous resources like CPUs and GPUs. We also make use of Intel's Math Kernel Library Intel (2018) (MKL) which provides machine learning libraries for supporting operations like activation (ReLU), inner product, and other useful functions.

4.2 Training Time

Using Ray we scale our training runs by using from 2 to 8 c5.18x instances (containing from 144 cores to 576 cores) on AWS. The scalability results are shown in Figure 4. As can be seen increasing the number of c5.18X Xeon CPU instances can significantly reduce the GANs training time up to 8 c5.18x instances. For the city of Los Angeles, the training time can be reduced from over one hour to less than 20 minutes. For New York City the training time can be reduced to just minutes.

City	*Number of Blocks*
San Francisco	1402
Los Angeles	1978
New York	765
Chicago	1155

Table 1: Training Workload for Different Cities

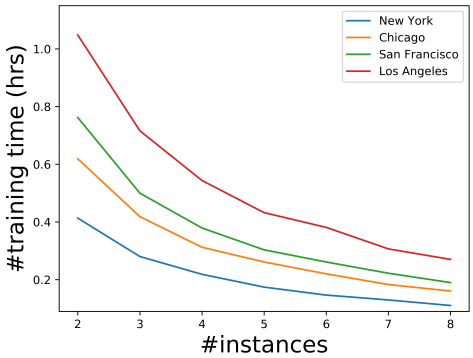


Figure 4: Training Time Performance Results on AWS with c5.18xlarge Xeon CPU instances.

Running times for sampling ride requests from the trained models and stitching the images of all the blocks together are significantly less than the training times, and are not included in these results.

We also conduct our GANs scaling experiments using GPU instances on AWS. In our initial experiments we observe no real performance improvements using GPUs. Scaling on GPUs requires more investigation. In this work we show that it is possible to achieve very nice scalability of our GANs workload using only CPU cores supported by Intel's MKL library and Berkeley's Ray framework.

5 Validation of Synthetic Data Sets

5.1 Spatial Variation

The correlation fractal dimension (D_2) gives a bound on the number of ride requests within a geographical region. This is an essential characteristic to match for the data set we are generating using GANs. In Table 2, we provide the fractal range for each city within which the fractal dimension remains constant. It is important to note that the fractal range for each city differs. The fractal range provides the ϵ range for which the data exhibits statistical self-similariy (Belussi & Faloutsos, 1998). The variation in the fractal ranges for the different cities can be attributed to the geographical shape of the city for which the ride requests are generated. We hypothesize that due to Los Angeles's sprawling nature, a larger ϵ is needed to observe self-similar patterns in comparison to the other three cities, which have a more corridor-like geographical region.

One may also interpret D_2 as a way to measure the fidelity of generated images to that from real data. Comparison of the ranges of values of D_2, in terms of min, max, and mean values, for the real and the synthetic data sets are fairly close although not identical. In most instances the mean value for D_2 is lower for the synthetic data sets in comparison to the real data sets. We believe this discrepancy in the values of D_2 require further investigation. Recent works to improve capture learning of high-resolution details of an image (Karras et al., 2017) can potentially benefit the learning for our ride request images.

City	Fractal Range	Real Data Sets			Synthetic Data Sets		
		D_2 min.	D_2 max.	D_2 mean	D_2 min.	D_2 max.	D_2 mean
New York	(450, 2500)	1.441	1.753	1.647	1.415	1.663	1.541
Chicago	(600, 3000)	1.139	1.558	1.384	1.188	1.567	1.435
San Francisco	(450, 2500)	1.283	1.731	1.548	1.242	1.65	1.426
Los Angeles	(1500, 4000)	0.962	1.638	1.355	1.048	1.489	1.314

Table 2: Summary of measured correlation fractal dimensions (D_2) for four cities; computed over a day for every hour using real and synthetic data sets.

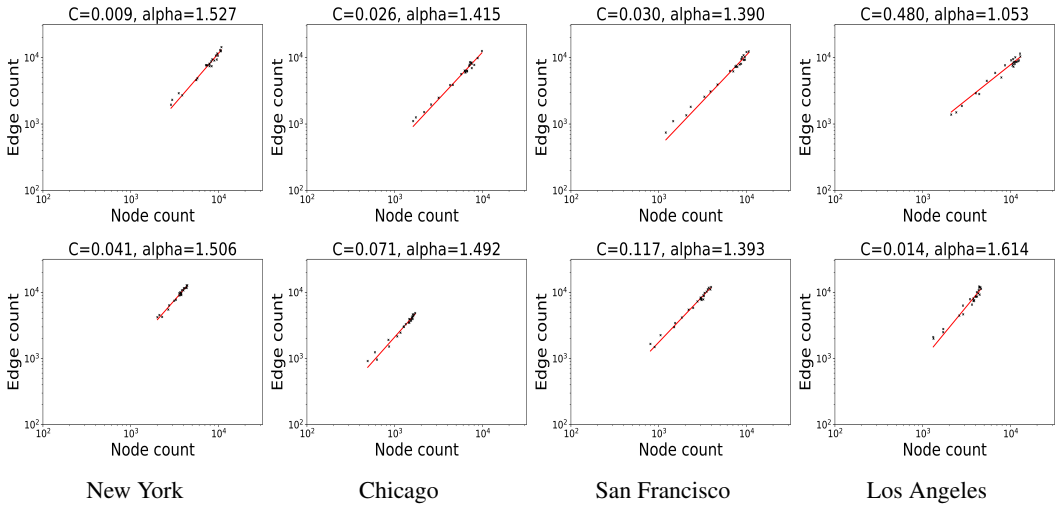


Figure 5: DPL plots from real data (top row) and synthetic data (bottom row) for four cities. The red line is the least square fit of the form $y = Cx^{\alpha}$, where y and x are number of edges and nodes respectively. $R^2 \approx 1.00$ for all of them.

5.2 Temporal Variation

DPL provides a characterization of the temporal evolution of ride requests. In the top row of Figure 5 we observe the plot of the DPL exponents α (slop of the line) based on the temporal patterns of the real data sets. For the ride request graph to obey DPL properties, we use graph generator proposed by (Jauhri et al., 2017a) to connect source and destination points. In the bottom row of Figure 5 we see the same results based on the synthetic data sets. We can see that the DPL exponent values α correleated quite nicely with that from the real data sets for New York, Chicago, and San Francisco. For Los Angeles, the synthetic exponent is higher than the real observed value; the geographical region for LA is much larger and due to many prominent regions of high request density, the model could likely suffer from bias towards generating more requests in prominent regions leading to a faster increase of the number of edges connecting nodes present in high density regions.

Another validation of our GANs approach is provided in Figure 6. Here we observe temporal variation of ride requests in terms of the volume of ride requests generated for each hour of a typical weekday. We see that for all four cities, the temporal variation of the synthetic data sets match quite well the temporal variation exhibited by the actual data set.

6 Conclusion

The emergence of ride sharing services and the availability of extensive data sets from such services are creating unprecedented opportunities for: 1) doing city-scale data analytics on urban transportation for supporting Intelligent Transportation Systems (ITS); 2) improving the efficiency of ride sharing services; 3) facilitating real-time traffic congestion prediction; and 4) providing new public services for societal benefit. Moreover, the power of neural networks for machine learning has

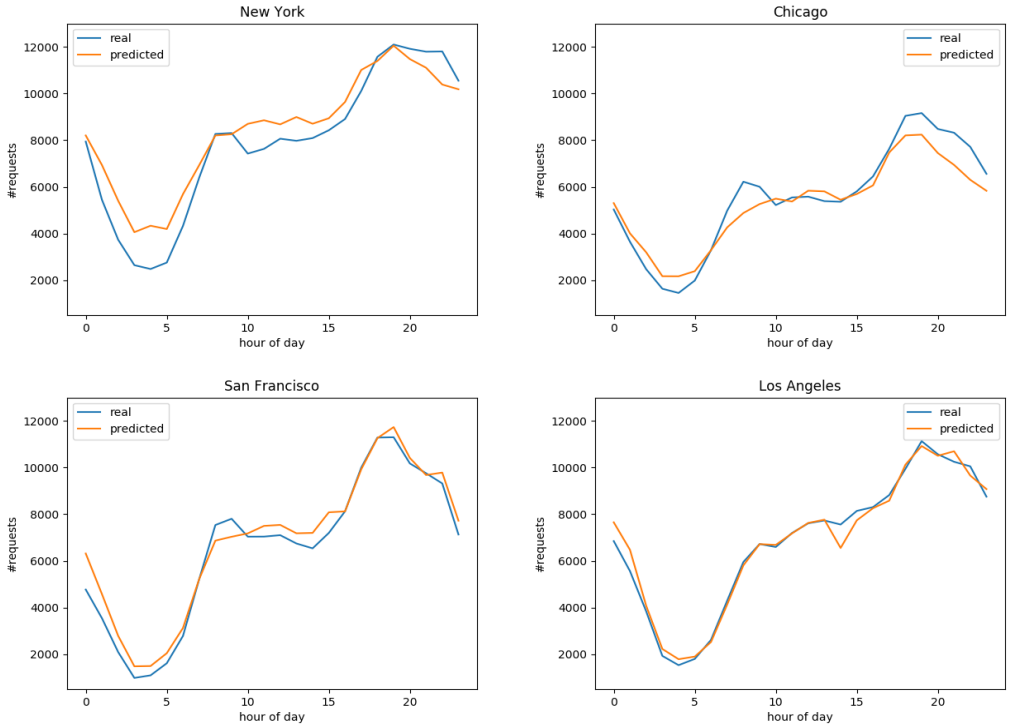


Figure 6: Plots for four cities highlighting the temporal variability of ride requests visible in both real and our model (predicted) for ride request generation. The pattern is representative of any typical day of week.

allowed the creation of useful models which can capture human behavior and dynamic real-world scenarios. The key contributions of this paper include:

- We map the ride requests of ride sharing services into a time sequence of images that capture both the temporal and spatial attributes of ride request patterns for a city.
- Based on extensive real world ride request data, we introduce a GANs based workflow for modeling and generating synthetic and realistic ride request data sets for a city.
- We further show that our GANs workload can be effectively scaled using Xeon CPU clusters on AWS, in reducing training times from hours to minutes for each city.
- Using previous work on modelling urban mobility patterns, we validate our GANs generated data sets for ride requests for four major US cities, by comparing the spatial and temporal properties of the GANs generated data sets against that of the real data sets.

There are other promising avenues for further research. Some open research topics include:

- Using the GANs generated data sets for experiments on new algorithms for dynamic ride pooling, real-time pre-placement of vehicles, and real-time traffic congestion prediction.[2]
- Using the GANs generated data sets for conducting experiments on *what-if* scenarios related to traffic congestion prediction and mitigation, and planning for future development of transportation infrastructures.

We are currently pursuing these research topics. As our GANs generated data sets are used in our follow up research, we plan to further validate the synthetic data sets by comparing our research results with results from using the real data sets. We plan to continue to tune our GANs models and generate improved synthetic data sets that can be made available for other researchers.

[2]We plan to publically release the code and models of our GANs along with baseline algorithms for looking at transportation related problems.

REFERENCES

Alberto Belussi and Christos Faloutsos. Estimating the selectivity of spatial queries using thecorrelation'fractal dimension. Technical report, 1998.

Min Hao Chen, Abhinav Jauhri, and John Paul Shen. Data driven analysis of the potentials of dynamic ride pooling. In *Proceedings of the 10th ACM SIGSPATIAL Workshop on Computational Transportation Science*, pp. 7–12. ACM, 2017.

Boyd Cohen and Jan Kietzmann. Ride on! mobility business models for the sharing economy. *Organization & Environment*, 27(3):279–296, 2014.

Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, pp. 1486–1494, 2015.

Jon Gauthier. Conditional generative adversarial nets for convolutional face generation.

Ian Goodfellow. Introduction to GANs. `http://efrosgans.eecs.berkeley.edu/CVPR18_slides/Introduction_by_Goodfellow.pdf`, 2018. [Online; accessed 31-August-2018].

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT Press, 2016.

Robert Hahn and Robert Metcalfe. The ridesharing revolution: Economic survey and synthesis. Technical report, 2017.

Intel. Introducing DNN primitives in Intel Math Kernel Library. `https://software.intel.com/en-us/articles/introducing-dnn-primitives-in-intelr-mkl`, 2018. [Online; accessed 31-August-2018].

Abhinav Jauhri, Brian Foo, Jerome Berclaz, Chih Chi Hu, Radek Grzeszczuk, Vasu Parameswaran, and John Paul Shen. Space-time graph modeling of ride requests based on real-world data. *arXiv preprint arXiv:1701.06635*, 2017a.

Abhinav Jauhri, Carlee Joe-Wong, and John Paul Shen. On the real-time vehicle placement problem. *arXiv preprint arXiv:1712.01235*, 2017b.

Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, volume 2, pp. 4, 2017.

Minhyeok Lee and Junhee Seok. Controllable generative adversarial network. *arXiv preprint arXiv:1708.00598*, 2017.

Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pp. 177–187. ACM, 2005.

Lucas Maystre and Matthias Grossglauser. Choicerank: Identifying preferences from node traffic in networks. *arXiv preprint arXiv:1610.06525*, 2016.

Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

Philipp Moritz, Robert Nishihara, Stephanie Wang, Alexey Tumanov, Richard Liaw, Eric Liang, William Paul, Michael I Jordan, and Ion Stoica. Ray: A distributed framework for emerging ai applications. *arXiv preprint arXiv:1712.05889*, 2017.

Recode. How many Americans use ride-sharing services? https://is.gd/Ed14R1, 2018a. [Online; accessed 31-August-2018].

Recode. Uber powered four billion rides in 2017. https://is.gd/cgSZFe, 2018b. [Online; accessed 31-August-2018].

Manfred Schroeder. *Fractals, chaos, power laws: Minutes from an infinite paradise*. Courier Corporation, 2009.

Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

New York Times. Taxi Medallions, Once a Safe Investment, Now Drag Owners Into Debt. https://www.nytimes.com/2017/09/10/nyregion/new-york-taxi-medallions-uber.html, 2018. [Online; accessed 31-August-2018].

Caetano Traina Jr, Agma Traina, Leejay Wu, and Christos Faloutsos. Fast feature selection using fractal dimension. *Journal of Information and data Management*, 1(1):3, 2010.

Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng, Yanwei Yu, and Zhenhui Li. Modeling spatial-temporal dynamics for traffic prediction. *arXiv preprint arXiv:1803.01254*, 2018a.

Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, and Jieping Ye. Deep multi-view spatial-temporal network for taxi demand prediction. *arXiv preprint arXiv:1802.08714*, 2018b.

Xian Zhou, Yanyan Shen, Yanmin Zhu, and Linpeng Huang. Predicting multi-step citywide passenger demands using attention-based neural networks. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 736–744. ACM, 2018.