Increasing Taxi Business Revenue From Predicted Demand By Applying Recurrent Neural Networks

M. Theresa [1], K. Sreenidhi [2], S. Janani Gowsalya [3]
[1]Associate Professor, Dept of Computer Science and Engineering
[2, 3]Dept of Computer Science and Engineering
[1, 2] Panimalar institute of technology, Chennai, India.

Abstract- *Forecasting the demand of taxi through the entire city can be efficiently used to manage the fleet of taxi control drivers. By forecasting the demand, we can reduce the waiting time for all passengers and can serve them in a better way. In this paper, we recommend a deep learning model which could predict the demand in next timely hours in every part of the city where the taxi fleet is available [1], [2].Usually in machine learning models, retention of past memory is not available mostly. So we implement a deep learning model known as LSTM (Long short term memory). LSTM model has a memory unit which comprises of input gate, forget gate and output gate. These gates are used to recollect recent customer's taxi booking information which is an efficient way. By using this predicted demand, if the demand is higher at a particular place then taxi owners can increase the fare, thereby they can increase their revenue. Also if the taxis are dispatched at correct time to a nearby request location, fuel cost can be cut down which in turn increases the revenue of the taxi company.*

Keywords- Prediction of taxi demand, time series forecasting, recurrent neural networks, LSTM memory cell.

I. INTRODUCTION

With the ongoing improvement of mobile computing and internet technology, most of the people practice the usage of mobile application to book a taxi cab for commuting.Therefore, predicting the passengers' commuting demand becomes much more complicated. Taxi drivers' decision of picking up passengers should be accurate as much as possible in order for better customer satisfaction. Travellersalso desire to find the taxi as soon as possible without waiting for long hours. As a result effective taxi dispatching is needed for better improvement of service. Most of the studies in deep learning have shown that it is possible to predict the future demand by learning through past data [5], [8].

Using this framework of recurrent neural networks, it is possible to train and test the neural network effectively to predict the taxi demand at a particular place. To figure an intelligent taxi fleet system, that can forecast the forthcoming demand over the entire city we need the efficient dataset. The dataset is built by gathering historic global positioning system data collected from GPS enabled taxis [3], [4].Foreseeing taxi demand is puzzling because it is inter-related with many portions of core information.

In this paper, we put forward a real-time way for predicting taxi demands in various areas in a city. We split a large city into reduced areas and cumulate the number of taxi requests in each area during a small period (e.g. 30 minutes). In this approach, past taxi data becomes a sequence of data for the number of taxi requests in an area. We train a Long Short Term Memory (LSTM) recurrent neural network (RNN) by passing this data which is sequential in nature [9]. The input to the network is the current taxi demand and other significant information such as time required for travel, estimated distance etc. and the output is the demandin the next time-step. The motivation we choose a LSTM recurrent neural network is that it is the only cell in RNN that can be trained to accumulate all the important information in a sequence to predict future outcomes. LSTM are highly capable of remembering some past information because of the incorporated gating mechanisms. By forecasting the demand, the taxi owners can increase the fare due to high demand.

The structure of this paper is arranged as follows:

Section II introduces mathematical model behind the LSTM and data encoding features. Section III describes the proposed sequential learning model, training and testing procedures. In Section IV, we list out the performance metrics of the prediction model. Lastly, in Section V, the conclusion of the paper is stated.

II. MATHEMATICAL MODEL:

In this division, we present how can we transform the highly resolution GPS data into the number of taxi requests in every part of the city using mathematical structure of recurrent neural networks. After transformation of data, we concisely describe the recurrent neural network in the form of mathematical method.

In this paper, the Geohash library [10] is used which can divide a geographical area into smaller areas with subjective precision. Geohash is a system for geocoding which utilizes a concept of the hierarchical spatial data structure that separates space into buckets of grid shape. Then the data is preprocessed by eliminating some missing values using pandas library.

The most widely used and efficient model is recurrent neural networkwhich can handle sequential data in a timely manner. Zhang *et al.*[5] proposes a recommendation system for taxi drivers using hotspots. A hotness score is assigned to each hotspot by analyzing the historical taxi data. This hotness score will be predicted in each time-step and combined with the driver's location, the *top−k* hot-spots would be recommended. But this hotspot framework can only perform short term prediction because hotspots are updated after a particular time interval.De Brébisson *et al.* [11] proposed a method of using recurrent neural networks for taxi demand prediction using the traces of taxi trip GPS data networks to predict taxi destination given the beginning of taxi trip GPS traces. The problems with the above proposed models are that these models don't fit for time series data, because time series data has to be trained separately.

We elongate this concept of RNN by using LSTM which has gate mechanisms.The concept behind RNNs is to remember relevant sequences of the input and makes use of this information for predicting the future output. Feed-forward neural networks predict the output based on the current output. But RNNs are capable of containing memory in which mostrelevant information from the past inputs can be stored. For example, when we train RNNs on a language translationtask in which it is important to store previous words, since the next word depends on the past word only. RNNs are called recurrent because they achieve the same working out on every component of a sequence, with the output being trained on the previous working outs.

The most widely used type of RNNs are LongShort Term Memory networks (LSTMs). LSTMs are an important kind of RNN, proficient of learning long-term dependencies because of their gating mechanism. They were introduced by Hochreiter& Schmidhuber [9] and were advanced and popularized by many researchers in the following years.

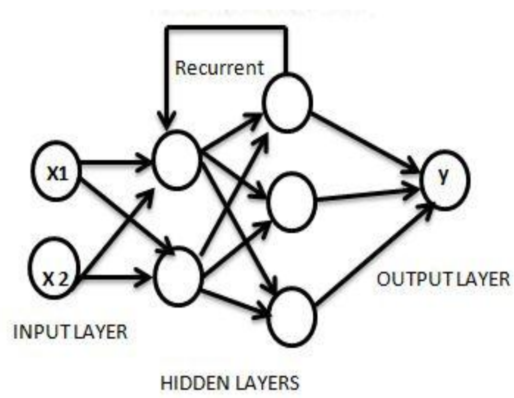

Fig. 1. A basic structure of recurrent neural network.

In the figure 1, we can decode the information that, the RNN processes the series of inputs*x*, stores the activation function in the hidden state *h* and outputs targetsat each time-step *t*. A feedback is provided back to the previous layer which imposes the importance of recurrence in neural networks.The weights are shared among different time steps. For the purpose of training weights among different time-steps, we need to unroll the recurrent neural networks for infinite number of times over number of time-steps.

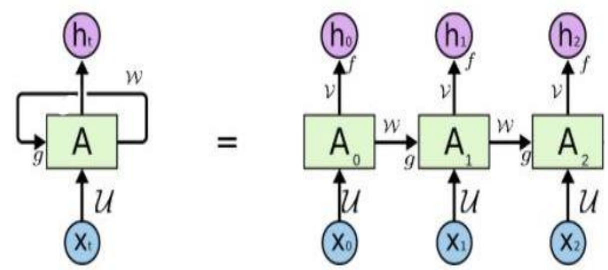

Fig 2. Unrolling of RNN network

As shown in the Fig 2, unrolling the RNN network makes sense how long the information is passed to the future sequence. The mathematical computation is formulated at each time step as follows:

ht = tanh(Wxh xt + Whhht−1 + bh)

-*xt* refers to input given at time-step *t*.
- *ht* refers to hidden state at time-step *t*.

The hidden state is calculated based on the inputs from previous layer and the current input which totally forms non-linearity. The network performs the same functionality at each step but with different inputs which accounts for recurrence.

III. PREDICTION MODEL FOR TAXI DEMAND

In this section, we confer about the most frequently used model which is known as sequential learning model. The model may be inaccurate or deviate to some extent from original prediction due to various other external factors. Therefore the number of taxi requests may also be inaccurate.

We use a stochastic model which predicts the entire probability distribution of taxi demands throughout entire city instead of predicting the deterministic taxi demand value. Then this probability distribution is used for deciding taxi demand for each area.

A. MIXTURE DENSITY NETWORKS:

The choice of network structure when predicting real valued data.The idea of mixture density networks (MDNs) [12] is to utilize the outputs of a neural network to parameterize a distribution which is mixture in nature. MDN importance lies in the stochastic behavior. They are widely used in prediction applications where there may be multiple outcomes.

In this application, instead of directly predicting the number of taxi requests, the parameters of amixture model are outputted by the neural network. The parameters are the mean and variance of each Gaussian kernel and also the probability mixing of each kernel. These parameters are sampled and then used in the final prediction.

B. LSTM-MDN SEQUENCE LEARNING MODEL

From the given prediction, it is easy to compute the price based on the demand.Here two LSTM layers are used where layer consists of activation neurons of about 1500 neurons for a particular training and testing set. The neurons that correspond to the variances **σk (x)**are passedvia an exponential function and the neurons corresponding to the means **μk (x)**are used directly without any changes. The weighted sum of *M* Gaussian kernels gives the probability density of the next output.

The multiclass classification methods uses the softmax function to “squash” a vector of *n* subjective real values *z* into a fixed values that sums up to 1, and which can be interpreted as probabilities. The Gaussian kernels[13] were given thecondition of the complete history of the inputs till current time-step $x = \{x1 \ . \ . \ . \ \mathrm{xt}\}$.This form of Gaussian mixture model is very much enough for approximating a mixed density function [13]. The error function is defined in terms of negative likelihood function.

After training the model, the prediction can be made for time-step $t + 1$ by giving the input as taxi demand at time-step *t*. The output from the mixture density parameters are used to parameterize the function. A sample is created from this distribution and this sample will be the prediction at the next time step. This prediction can be repeated in a loop to predict taxi demand for multiple time-steps.

C. LSTM MDN CONDITIONAL MODEL:

This LSTM-MDN-Conditional model learns patterns for taxi demand at a particular place from the previous taxi data, but also remembers the unrolled LSTM-MDN-Conditional model for a single time-step.This approach of using conditional model has been implemented in other works such as in Pixel RNNs [14] and Neural Autoregressive Distribution Estimator (NADE) [15].The probability distributions are predicted at the same time-step for the allocated taxi area. This shows that demand predicted for each area is being conditioned for all theareas of previous time-steps. But, the taxi demand in an area might not only be related to the past, but also to the current time-step which falls for taxi demand prediction. So instead of resulting a joint distribution for all areas in a single timestep, we formulate the network to predict the conditional distribution at each area at a single time-step.Training the LSTM-MDN model takes much time because the LSTM wants to be unrolled for longer time-steps. If the city has N areas then, the model should be unrolled *N* times.This approach is referred to as LSTM-MDN-Conditional model because its output is similar to the input of LSTM-MDN mode, but theinput only points out to output of single area taxi demand.

IV. EXPERIMENTAL STUDY

In this section, the proposed LSTM-MDN and LSTM-MDN-Conditional models which are applied on a preprocessed dataset of taxi requests are evaluated and see how they can perform in prediction for taxi demand in the future.

A. SETUP FOR EXPERIMENTATION

The two kinds of cabs in New York City are the yellow cabs, which will be operatingin Manhattan, and the green cabs, which will be operating in the suburbs. The dataset contains past taxi trips from January 2009to June 2017 for both the kind of cabs. The dataset also includes the passenger pickup and drop location details. The data preprocessing is done by removing null values and unwanted columns [16]. About 500 million taxi trips data were used for training and testing phase.80% of the data were used for training the model and the remaining 20% of the data are used for testing the model.

The network model is implemented in Blocks framework which is built on top of Theano [3]. The training is stopped when the error ofvalidation does not change for the batch of 20 epochs.Subsequentlyfollowing the training, the network takes less than a second to predict the demand. Than the training time, prediction time is more important.Prediction time is important because once the model is deployed it needs to forecast the demand in a round to provide the demand prediction information in real-time.

The density map for both real demand and the predicted demandis constructed based on the LSTM recurrent neural network where the red areas indicate high demand for taxis, yellow areas indicate lower demand, and the grey-white portions indicate there is no demand in that particular area.

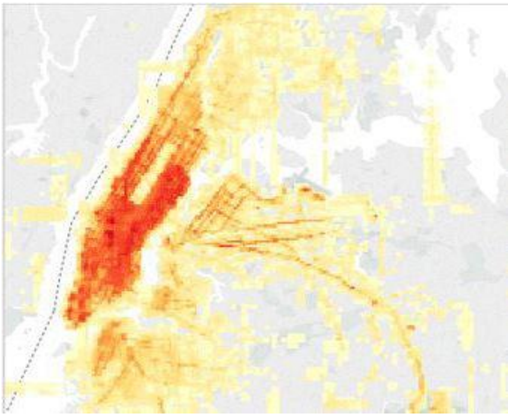

Fig.3.Density map of predicted demand at New York on 6.00 pm

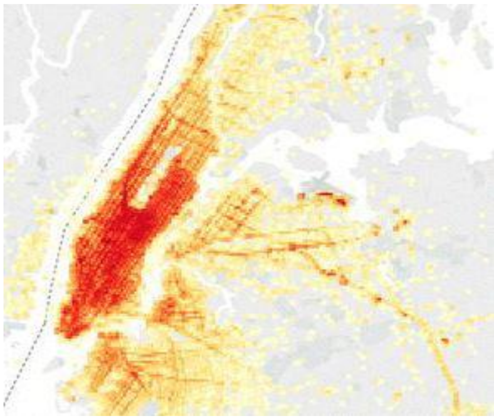

Fig.4. Density map of actual demand at New York in 6.00 pm

Hypothetically, LSTM networks are usually trained using subjective sequence lengths. As far as the computational power is considered as constraint, a single week data is used as a sequence and changed it into time-steps containing various lengths. For example, Fortime steplength of 80*mins*, the sequence length will be 24 × 7.Then for the time-step length of 40*mins*, the sequence length will be 24 × 2 × 7.

B. PERFORMANCE METRICS AND BASELINES

The performances of the proposed models are evaluated by comparing the outcomes with prediction approaches based on two strategies[17], [18]. One is fully connected feed-forward neural networks and another one is naive statistic average predictor.

Fully Connected Feed-Forward Neural Network Predictor:

For classification and regression problems, feed-forward neural networks are used. In feed-forward neurons have associations from their input to their outputs. The majordissimilarity between feed-forward neural networks and recurrent neural networks is that in RNNs, the recurrent connection from the output to the input at next time-step makes the network accomplishing of remembering the information.

Naive Statistic Average Predictor:

Inthis approach the prediction is based on the mean value of previous demands in a slidingwindow.

Performance Results:

On the overall training the experimental results indicates that LSTM overtakes the performance of other prediction approaches. The only reason that LSTMoutperforms is due to its capability of processing information in the previous time-steps. For example, if a group of people request taxis for going to shopping, it will remember this information and use this information to predict that after fewhours there will be the same number of requests in the shopping area. The FC network will find the accurate mapping from the time and geographicalinformation to the number of requests without the need for accessing to the demand in the previous time-steps. This is the only limitation that causes larger errors in its prediction. Overall for better prediction, we are in need of mathematical model which is very authoritative and properly outputs the target value based on the available information. The best prediction performance is attained when all the impacting factors considered in this work are available as input to the network.

V. CONCLUSION

We recommend a recurrent neural network which is sequential learning model in combination with mixture density networks to predict taxi demand in different areas of a city. On learning the patterns from previous taxi demand patterns, the LSTM-MDN model can predict taxi demand predictions for overall the city. Four years taxi trips data covering entire area of New York City is used for training our model. Based on the experimental results, it is shown that the LSTM-MDN model can achieve a good accuracy of around 90% at the city level.The LSTM-MDN model is further extended to a conditional model. In conditional model, it takes past data as well as current data as input for prediction whose outcome is much more accurate. From the predicted demand, we can efficiently manage the fleet of taxis to particular area and based on that customer can eliminate waiting time for taxis. As a result, customer satisfaction also increases. In addition, we show that our model outperforms other models in terms of saving the fuel by distributing the fleet of taxis to nearest predicted area.

This prediction model can also be improved by adding more information to such as where businesses, shops, restaurants, etc. are located to the input network. Additionally, we can consolidate the taxis in a city and allocate them in real-time according to the demand prediction by our model. This model overcomes the problems in situations where in some areas there is large demand but the taxi drivers are competing with each other for having travellers in various other parts of the city. A centrally managed taxi dispatch system would be especially economical in the future self- driving cars which are in need to be organized automatically to respond to the taxi requests in a city. Such a system can save a lot of money for the managers of taxi-Company by saving fuel, increased customer satisfaction therefore increased business target for customers. Eventually, the business revenue increases as result of efficient taxi dispatching from the predicted demand.

REFERENCES

[1] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, “T-finder: A recommender system for finding passengers and vacant taxis,” *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 10, pp. 2390–2403, Oct. 2013.

[2] K. T. Seow, N. H. Dang, and D.-H. Lee, “A collaborative multiagent taxi-dispatch system,” *IEEE Trans. Autom. Sci. Eng.*, vol. 7, no. 3, pp. 607–616, Jul. 2010.

[3] P. Santi, G. Resta, M. Szell, S. Sobolevsky, S. H. Strogatz, and C. Ratti, “Quantifying the benefits of vehicle pooling with shareability networks,” *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 37, pp. 13290–13294, 2014.

[4] X. Ma, H. Yu, Y. Wang, and Y. Wang, “Large-scale transportation network congestion evolution prediction using deep learning theory,” *PLoS ONE*, vol. 10, no. 3, p. e0119044, 2015.

[5] K. Zhang, Z. Feng, S. Chen, K. Huang, and G. Wang, “A framework for passengers demand prediction and recommendation,” in *Proc. IEEE SCC*, Jun. 2016, pp. 340–347.

[6] F. Miao *et al.*, “Taxi dispatch with real-time sensing data in metropolitan areas: A receding horizon control approach,” *IEEE Trans. Autom. Sci.Eng.*, vol. 13, no. 2, pp. 463–478, Apr. 2016.

[7] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *NeuralComput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] G. Niemeyer. (2008). *Tips & Tricks about Geohash*. Online tool Available: http://geohash.org/site/tips.html

[9] A. de Brébisson, É. Simon, A. Auvolat, P. Vincent, and Y. Bengio. (2015). “Artificial neural networks applied to taxi destination prediction.” [Online]. Available: https://arxiv.org/abs/1508.00021

[10] C. M. Bishop, *Mixture Density Networks*. Birmingham, U.K.: Aston University, 1994

[11] G. J. McLachlan and K. E. Basford, *Mixture Models: Inference andApplications to Clustering*, vol. 84. New York, NY, USA: MarcelDekker, 1988.

[12] A. van der Oord, N. Kalchbrenner, and K. Kavukcuoglu, “Pixel recurrentneural networks,” in *Proc. ICML*, Jun. 2016, pp. 1747–1756.

[13] H. Larochelle and I. Murray, “The neural autoregressive distributionestimator,” in *Proc. AISTATS*, Jun. 2011, pp. 29–37.

[14] NYC Taxi & Limousine Commission. *Taxi and Limousine Commission(TLC) Trip Record Data*. Accessed: Dec. 2016. [Online]. Available: http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

[15] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, “Traffic flow predictionwith big data: A deep learning approach,” *IEEE Trans. Intell. Transp.Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

[16] M. Yang, Y. Liu, and Z. You, “The reliability of travel time forecasting,”*IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 162–171, Mar. 2010.