A novel dynamically adjusted regressor chain for taxi demand prediction

1st Zipeng Wu
School of Computer Science and Information Security
Guilin University of Electronic Technology
Guilin, China
ylwzp0922@aliyun.com

2rd Guan Lian
School of Architecture and Transportation Engineering
Guilin University of Electronic Technology
Guilin, China
lianguan@guet.edu.cn

***Abstract*—Taxi is an essential part of urban traffic, accurately predicts the taxi demand, which not only facilitates people's travel but also promotes the further development of the entire smart city. The gap between demand and the actual amount for taxi causes trouble for travelers. Forecasts for taxi demand do not take into account the possible interactions of taxi demand between areas, which can lead to a decrease in the accuracy of the forecast. In further exploiting the interaction of taxi demand in each area, We propose An extended Maximum Correlation Regressor Chain method (MCRC) and a new MCRC-based Dynamically Adjusted Regressor Chain method. MCRC uses the various relationships existing among the targets, which are evaluated using Spearman's rank correlation coefficient, feature importance matrix, and maximal information coefficient, respectively, to form the maximum correlation chain with higher prediction accuracy. Based on MCRC, DARC dynamically adjusts the base-regressor of the regressor chain. A set of predictive approaches are implemented to compare the performances, and the results show that the maximal information coefficient DARC (DARC_MIC) achieves the best accurate rate by 91.80%. DARC_MIC is not only can provide managers a more rational taxi operation approach but also more proper for dealing with multi-target regression problems with Lots of targets. This idea of first measuring the degree of interaction between targets and then combining algorithms to further exploit this degree of interaction between targets can also be attempted to improve many other multi-target regression algorithms.**

Keywords—Traffic prediction; taxi demand; regressor chain; multi-target regression

I. INTRODUCTION

Taxi is one of the significant ways for citizens to travel. However, the problem of unbalanced supply and demand for taxis still occurs [1], which not only leads to the waste of fuel but also bothers travelers. Managers can schedule the idle taxis that according to the forecast information by predicting the taxis demand accurately in each area and at various periods. The benefits will follow with it, such as reducing the no-load rate and fuel consumption, optimizing the service of taxis, curtailing emissions. At the same time, taxi demand prediction plays an essential role in building a smart city.

The taxi demand prediction is a complex process because it is affected by many factors(Section III) such as the weather, the month, the day of the week, the hour, the temperature, and the traffic, which cause a high variability in the taxi demand over time [2]. In order to bring more convenience to travelers, taxi demand prediction is always with a large number of small areas, which means that there are a large number of prediction targets, and each area acts as a node of an urban transport network and interacts with each other. Many current researches don't consider the taxis demand in different areas that may have correlation and similarity and don't use these correlations or similarities between taxi demands in different areas to improve model performance. For example, there are two near locations of A and B. If most people think that the location A is easy for taking a taxi, more people will take taxis at the location A than at the location B. Nevertheless, the current researches (Section II.A) do not consider the taxis demand in different areas that have possible correlation and similarity and did not use these correlations between taxi demands in different areas to improve model performance. Regressor chain (RC)(Section IV.A) or Chain methods(Section II.B) is a series of multi-target regression methods which are suitable for solving this problem due to its ability to make good use of the interaction among targets. We will show a novel regressor chain method with additionally exploring how to make good use of interactions among targets to improve the performance of the multi-target regression model.

The main contributions of this paper are:

(1) It proposed an extended Maximum Correlation Regressor Chain method (MCRC) to explore the interaction among target variables(taxi demand of the different area) through Spearman's rank Correlation Coefficient(SC), Feature Importance matrix(FI), and Maximal Information Coefficient(MIC), respectively(Section IV.B). There is a linear monotonic relationship between the demand for taxis in several areas, and this relationship will have a good effect on the performance of the multi-target regression model. There is little linear or non-linear relationship between some areas, which, if not addressed, will have a potentially harmful effect on the performance of the multi-target regression model.

(2)It proposed a new MCRC-based dynamically adjusted regressor chain method, which can take better advantage of the correlation among targets in multi-target regression and improve the prediction than MCRC, which could be called DARC, is proposed. Furthermore, the DARC_MIC, which means through the MIC to measure interaction among targets to generate models, achieves the best predictive accuracy of taxi demand,91.80%(Section V).

The remainder of this paper is organized as follows. In Section II, we review the related work of taxi demand prediction. Section III is data preprocess and analysis. Section

This work was partially supported by the Guangxi Natural Science Foundation (AD19245021).

IV describes existing regressor chains and proposes our new proposal. Along with the experimental configuration, the experimental results are depicted and discussed in Section V. And in Section VI, we draw some conclusions of this research.

II. Related Works

A. Taxi demand predictions

Researchers have made many studies on taxi demand prediction. These predictive strategies and models can be divided into the following two categories:

Category I is characterized by the full exploitation of the patterns of change in taxi demand over time so that the ARIMA model and its variants are used as the main forecasting model. These models had a relatively good performance in the context of the time and were also very forward-looking. Perhaps because researchers in the early days used far less computer equipment than they do now, such models tended to be simpler, did not rely on the advanced computing ability of computers, so these studies did not use large complex models and did not add the impact of factors other than time on taxi demand into features of the prediction model, which is the fundamental difference between Category I and Category II.

Shekhar proposed an autoregressive integrated moving average (ARIMA) model for traffic demand and real-time traveler information prediction through combining with three filtering techniques including Kalman filter, recursive least squares, and least mean squares. This proposed method showed better performance than the Maximum likelihood estimated models [3]. The data used in their research has an entire year from eight different sites for the entire year of 2002, three- and four-lane urban and rural motorway sections in the United Kingdom. Their research built up a significant basis for studies on taxi demand prediction. Moreira-Matias et al. proposed a combination of time-varying Poisson model and ARIMA model for prediction in order to predict the demand of each taxi station in the short-term (within 30 minutes) for taxi drivers. Their models achieved the predictive accuracy by 78% with the 63 taxi stands of 6-month historical data in Porto city [4]. Similar taxi demand work is studied by Moreira-Matias et al. [5,6]. Li et al. used an improved ARIMA based prediction method to forecast the spatial-temporal variation of passengers and taxis in a hotspot. By Evaluating with a large-scale real-world data set of 4 000 taxis' GPS traces over one year, the predictive results show high predictive accuracy [7]. Pan et al. proposed an H-ARIMA model to predict traffic conditions in the next 30 minutes. And a time-series mining technique is used in this model. Their study showed it could reach a higher accuracy than the ARIMA model during traffic jams and accidents. Their dataset is a very large-scale and high resolution (both spatial and temporal) traffic loop detector dataset collected from entire LA County highways and arterial streets from November 1st to December 7th, 2011, which include traffic event data [8]. Rodrigues et al. proposed two deep learning architectures for combining time-series and textual data for taxi demand prediction in essential areas. These approaches were tested for taxi demand forecasting in event areas by using a large-scale public dataset of 1.1 billion taxi trips from New York. Comparing to Long Short-Term Memory (LSTM) model, the proposed approaches were able to reduce the error in the prediction significantly [9]. Yu et al. applied the deep LSTM recurrent neural network to traffic prediction for the first time and evaluated it on the large-scale traffic dataset in Los Angeles. The implemented results for prediction in 3-hour peak hours showed the approach achieved 30%-50% improvement than Linear regression and Ridge regression on the data located on the highways and arterial streets of Los Angeles County from May 19th, 2012, to June 30th, 2012 [10].

Category II takes into account traffic and other non-traffic factors that may affect the demand for taxis as features and builds complex forecasting models, and the accuracy of the taxi demand forecasts in this stage has improved considerably over the Category I and has reached a high level of accuracy.

Chang et al. proposed a context-aware taxi demand method by mining historical data for 60 days. This method, including data filtering, clustering, semantic annotation, and hotness calculation, could improve the predictive accuracy in rush hours of the next day in different areas [11]. Wang et al. presented an end-to-end framework called Deep Supply-Demand (DeepSD) using a novel deep neural network structure, which can automatically discover complicated supply-demand patterns from the car-hailing service data while only requires minimal amount hand-crafted features. By training data of 24 days to test data of the next 28 days, which is 58 square areas in Hangzhou, China, their mothed can achieve higher predictive accuracy compared with other approaches such as LASSO and RF [12]. Tong et al. proposed a linear regression model with super-high-dimensional multi-features and a series of optimization techniques for efficient model training and updating. They predicted the taxi demand in a specific same time period in the test set by training the same time periods in the historical dataset that contains 75 days of taxi order data from Beijing and Hangzhou, as well as corresponding POI data, data sets for meteorological data [13]. These predictive methods consider many features that can affect the taxi operation and travelers to predict taxi demand. However, they just predicted the taxi demand within the overall area, and these methods could not accurately predict the distribution of taxi demand in each small area.

B. Multi-target regression with Chain methods

The problem above can be regarded as a multi-target regression process. The most challenging challenge for multi-objective regression is to identify the relationship between the input data and its corresponding output values. The development of the regressor chain (RC) made up for the shortcomings in this aspect. Spyromitros-Xioufis introduced the RC to the predictive problem by multi-target regression. Through evaluating the six multi-target regression data sets, it showed all the proposed multi-target methods were able to improve the accuracy of the other approaches, which performed a separate regression for each target [14]. Mastelini et al. used the Ensemble of Regressor Chain (ERC) to prove this method performed well when producing the multi-target prediction problem comparing with traditional single-target methods [15]. Mastelini et al. also applied the ERC method to 18 benchmark datasets to solve the Multi-target regression problem. Their predictive results showed the performance was superior to the single-target approach [16]. Santana et al. [17,18], Tsoumakas et al. [19], and Spyromitros-Xioufis et al. [20] used RC and its variants to deal with multiple-target predictive problems in different fields. Based on this work, Melki et al. adopted a support vector regressor correlation chain (SVRCC) method on 24 multi-target datasets. Their results showed that the maximum correlation SVR approach improved the performance of using ensembles of random chains [21].

In the previous work, the number of targets of their datasets was between 2 and 16, and the data set of this paper has 37 targets. The prediction error is likely to propagate and increase along a chain when making predictions for a new test instance, which means the longer the length of the regression chain, the more unfavorable the performance of the regressor chain. Thus, how to use the chain methods to predict dataset with a larger number of targets become a challenge. This also illustrates the need for the approach proposed in this paper (Section IV).

III. Problem and Data preprocessing

Predicting taxi demand in each area is a complex multi-targeted regression problem. We generate the multi-target regression dataset through data preprocessing. Specifically, the number of taxi orders per hour in each area is the target. The month, day of the week, hour, weather, and temperature are used as features:

1. Month: 1-12 represent January to December; 2. Day of the week: 1-7 represent Monday to Sunday, and 0 represents holiday; 3. Hour: 0-23 represents from 0 am to 11 pm; 4. Weather: 1-3, represent sunny, rainy, and snowy; 5. Temperature: degree Celsius. This paper uses 3911 hours of data from March 1st to August 10th in 2014 as a training set and 168 hours of August 11th-17th as a test set to predict the grid of each grid in hours.

A. Dataset resource

This paper obtained the taxi transaction data, between January 1st 2014 and December 31st 2015, from the data collection official website published by the New York Taxi and Car Committee [22]. Each order includes boarding time, drop-off time, boarding longitude, boarding latitude, getting off longitude, getting off latitude, number of passengers, driving distance, type of charge, fare, etc. The size of data in 2014 is 25.9G, with an average of 452,368 orders per month, while the size of data in 2015 is 21G, with an average of 400,309 orders per month. The original data file is not arranged in the order of time. Hence we rearrange all the data into daily files (730 days total) and hourly files (17520 hours in total) for subsequent data analysis.

B. Influencing factors analysis and features selection

(1) Time factors: In order to analyze what can influence taxi demands in New York, two groups of experiments are conducted.

The first group is to compare the change patterns of taxi demand every day in a week. The number of taxi orders in 20 weeks in different months is computed. Like in Fig.2 and Fig.3, the relation between the taxi orders and the hours every day in a week in October and August 2015 are drawn respectively. Overall, taxi demands are different every day from Monday to Sunday, the taxi demands in working days are different from the taxi demands on weekdays. In detail, on working days (from Monday to Friday), the change patterns of taxi demand are similar within 5:00-19:00, but with differences within 0:00-5:00 or 19:00-23:00. While on weekdays (from Saturday to Sunday) the patterns of taxi demand are similar between 0:00-19:00, but with differences between 19:00-23:00.

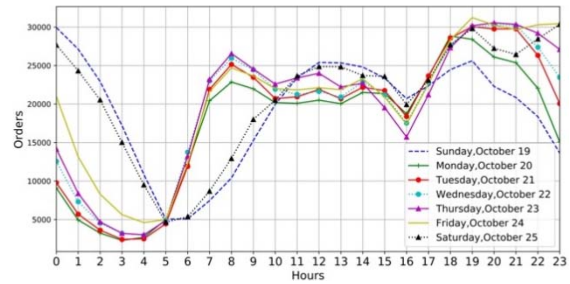

Fig. 1. The total number of taxi order varies with hours in a week in October 2015

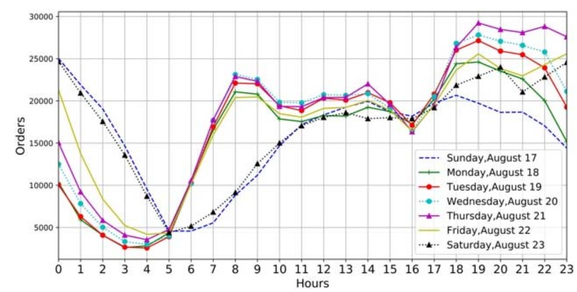

Fig. 2. The total number of taxi order varies with hours in a week in August 2015

The second group is to compare the change patterns of taxi demand in holidays and non-holidays. The number of taxi orders in 10 weeks in different months is computed. Like in Fig.4 and Fig.5, the relation between the taxi orders and the hours in four Mondays in May and September 2015 are drawn respectively. To be noted is that May 25th is American Memorial Day, and September 7th is American Labor Day. From Fig.4 and Fig.5, it can be observed that the change patterns of taxi demand in holidays are different from non-holidays.

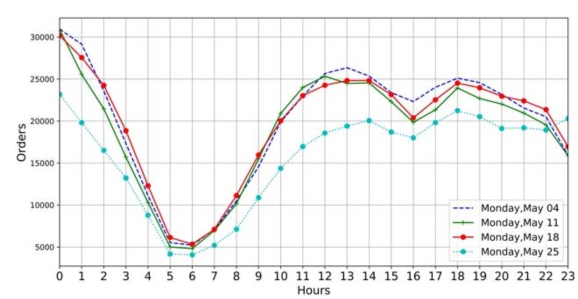

Fig. 3. The taxi demand of NewYork varies with hours for fout Monday in May

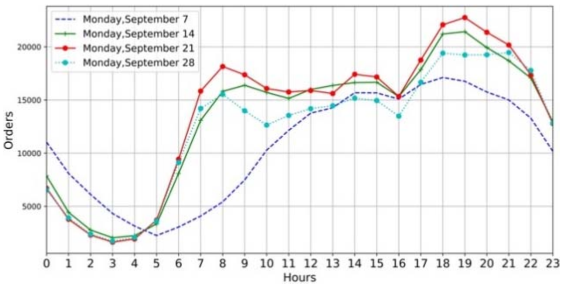

Fig. 4. The taxi demand of NewYork varies with hours for four Mondays in September

(2) Climatic factors: It is simply to be understood that the orders of taxi demands are influenced by the weather. Fig.6 and Fig.7 show the total number of taxi orders in New York from 12:00 pm to 1:00 am and 1:00 am to 2:00 am in all days in April respectively. It can be observed from them that the total number of taxi orders in 4th and 5th April (red circled) are smaller than in 11th, 12th, 18th, 19th, 25th, and 26th April, because it was raining from 12:00 pm to 2:00 am in 4th and 5th April, while it is sunny from 12:00 pm to 2:00, am in 11th, 12th, 18th, 19th, 25th, and 26th April.

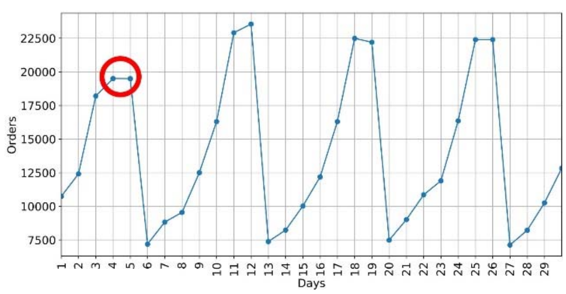

Fig. 5. Taxi demand changes from 12:00 am to 1:00 am for each day in April

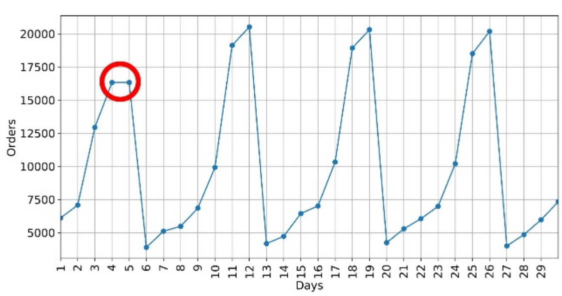

Fig. 6. Taxi demand changes from 1:00 am to 2:00 am for each day in April

Fig.8 shows temperature varies with days in 2014 and 2015 in New York, while Fig.9 shows The average number of taxi orders per hour in a day varies with days in 2014 and 2015 in New York. It can be observed from them that the number of taxi orders is reduced when the temperature is above 20℃.

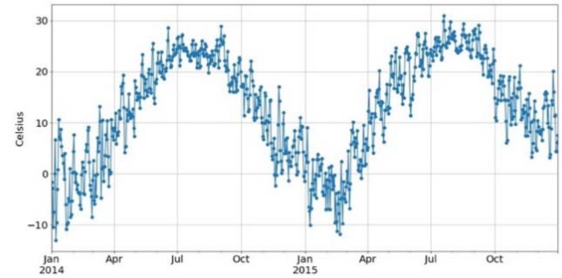

Fig. 7. Temperature varies in 2014 and 2015

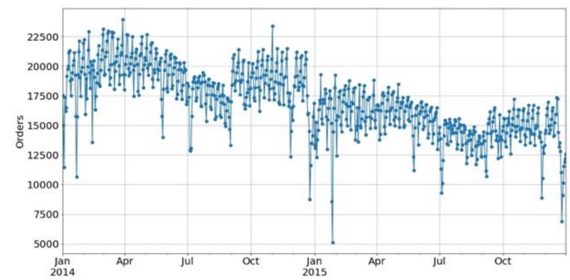

Fig. 8. New York Taxi demand per hour in a day varies in 2014 and 2015

C. Grid division

To predict the taxi demands in New York accurately, this paper chooses to divide New York into many grids, each of which covers a subregion with latitude and longitude of 0.01°×0.01°. Geographical latitude or longitude 1 degree is about 111km on the ground. Then the taxi transaction data in 2014 and 2015 are used to calculate the hourly taxi demands in each grid. The grids, in which the average taxi demand per hour is less than 100, is eliminated in this study. Thus, thirty-five grids are determined from Manhattan Island. In addition, two airports of JFK and La Guardia are added as two grids, and hence 37 grids in total are obtained. These grids are shown in Fig.9. The number of taxi orders per hour in a grid is defined as a target.

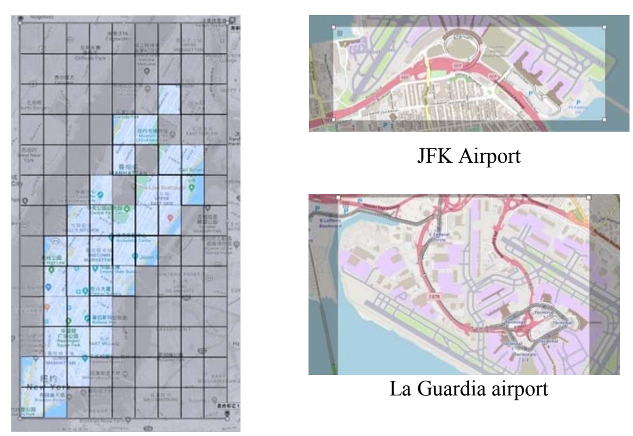

Fig. 9. The meshing of Manhattan Island, JKF Airport and La Guardia Airport, New York

IV. METHODS

Briefly, both MCRC and DARC were designed with the idea of further improving the level of utilization of correlations between targets, by different measures. MCRC is an extension of SVRCC, both of which are essentially about finding some order of the regressor chain that maximizes the correlation among the targets, thereby improving the model's performance. SVRCC uses linear correlation coefficients to measure correlations among targets, and it seems to be very effective when there is a strong linear correlation among targets. In this section, I extend this to generate MCRC using linear relations (Spearman correlation coefficient), non-linear relations (feature importance), and composite relations (MIC), respectively. The essence of DARC is to improve model performance by removing the weakly correlated targets from each submodel in the regressor chain according to the strength of the correlation between the targets. Clearly, this improvement builds on the chain of maximal correlation regressors(MCRCs), that have been formed.

For convenience, the following notations are defined.

TABLE I. NOTATIONS

Notations	Definition
m	number of targets(number of taxi demand of areas)
d	number of input features
N	number of timestamps for the training dataset
$\boldsymbol{X}_t = \left(x_1^t, x_2^t, ..., ..x_d^t\right)^{\mathrm{T}}$, $1 \le t \le N$	a training instance at time t
$\boldsymbol{X} = \left(\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_N\right) \in \mathbb{R}^{d \times N}$	training space
$\boldsymbol{y}_j = \left(y_j^1, ..., y_j^N\right)^{\mathrm{T}}, 1 \le j \le m$	taxi demand in the j th target from time 1 to N
$\boldsymbol{Y}_t = \left(y_1^t, y_2^t, ..., y_m^t\right)^{\mathrm{T}}, 1 \le t \le N$	taxi demand in different targets at time t
$\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_N) \in \mathbb{R}^{m \times N}$	outputs of training sample space
$\boldsymbol{D} = \left\{(\boldsymbol{X}_1, \boldsymbol{Y}_1), (\boldsymbol{X}_2, \boldsymbol{Y}_2), ..., (\boldsymbol{X}_N, \boldsymbol{Y}_N)\right\}$	multi-target training dataset
M	number of timestamps for the test data (number of hours)
$\hat{\boldsymbol{X}} = (\boldsymbol{X}_{N+1}, \boldsymbol{X}_{N+2}, ..., \boldsymbol{X}_{N+M}) \in \mathbb{R}^{d \times M}$	test sample space
$\hat{\boldsymbol{Y}} = (\boldsymbol{Y}_{N+1}, \boldsymbol{Y}_{N+2}, ..., \boldsymbol{Y}_{N+M}) \in \mathbb{R}^{m \times M}$	outputs of test sample space
$\hat{\boldsymbol{y}}_j = \left(\hat{y}_j^{N+1}, \hat{y}_j^{N+2}, ..., \hat{y}_j^{N+M}\right)$, $1 \le j \le m$	taxi demand predictions in the jth target from time N+1 to N+M

According to the dataset used in this paper, it can be observed that m=37, d=5, N=3911, and M=168.

A. Regressor chain methods

The regressor chain method has received extensive attention due to its simple concept and excellent performance. The training process of RC is shown in Figure 10. The first step is to choose a random chain order for the targets. Once a chain order is selected, a regression model is built for each target according to the order of the chain selected. When j >1, all of the raw input spaces have been expanded by all previous target values, the model h_j is trained on the transformed training set D_j.

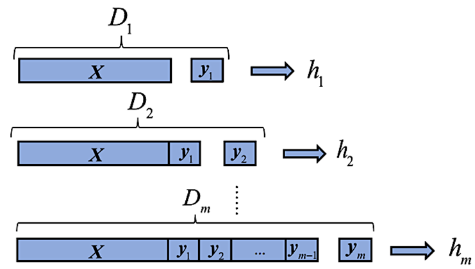

Fig. 10. RC training process

The prediction process of RC is shown in Figure 11. During the RC prediction process, since the true value of the target variable is not available, the previously obtained predicted values $\hat{y}_1, ..., \hat{y}_{j-1}$ (1<j<m) are continuously used as additional input variables to extend the input space.

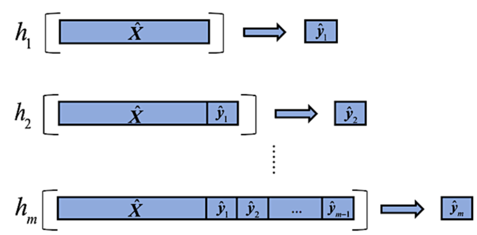

Fig. 11. RC prediction process

B. Maximum Correlation Regressor Chain

The inherent problem of RC lies in that the discrepancy of the values of the additional input variables between training and prediction. Many previous studies analyzed the factors which affect the accuracy of RC as follows: (1) the order and length of the regression chain; (2) the correlation among the targets; (3) the discrepancy. The main reason behind these factors is the degree of correlation among the targets.

ERC was proposed to improve the performance and the stability of RC. In ERC, a set of k RC models with different random chains are built on samples of the training set, and the final predictions come from majority voting. While SVRCC takes a completely different approach to deal with the problem of chain order, it first calculates the linear correlation coefficient of the targets and forms a correlation coefficient matrix of the targets. SVRCC uses a chain of targets with descending order to train the regression model for each target.

SVRCC illustrated that the chain order would affect the performance of chain methods. So, we can guess that the correlation between targets and the chain order sorting by correlation are both significant. However, SVRCC only takes advantage of the linear relationship among the targets. Only linear correlation is not sufficient to measure the relationship among targets in the real world. There are many ways to measure the relationship among targets. So, inspired by SVRCC, we try to adopt different methods for measuring the correlation of targets and take it to sort the target variables. This proposed method, called Maximum Correlation Regressor Chain(MCRC). The first step of the MCRC training process is to take a function *coefficient*() to measure the correlation among target variables $\boldsymbol{Y}$. Then MCRC calculates the correlation between each two target variables, and a correlation matrix is formed. Then each column is summed respectively, and the cumulative correlation coefficient for each target is gotten. The targets are sorted in descending order according to the cumulative correlation coefficients of the targets, and the chain order C is obtained. All subsequent steps are the same as RC. The pseudocode of the MCRC algorithm is as follows.

Algorithm 1: MCRC training algorithm

Input: Training dataset D

Output: Chained model $h_j, j=\{1,\ldots,m\}$

1: $COE_{m\times m} = coefficient(\boldsymbol{Y})$;

2: $c_l = \text{sum}(COE(:,l)), l=\{1,\ldots,m\}$;

3: $C = \text{sort}(\{c_1,\ldots,c_m\},\text{descending})$;

4: $D'_{C[1]} = \{(\boldsymbol{X}_1, y^1_{C[1]}),\ldots,(\boldsymbol{X}_N, y^N_{C[1]})\}$;

5: $h_1 : D'_{C[1]} \to \mathbb{R}$;

6: *for* $j=1$ *to* $m-1$ do:

7: $D'_{C[j+1]} = \{\}, \boldsymbol{Y}^p = \{\}$;

8: $\boldsymbol{Y}^p = \left((y^1_{C[1]},\ldots,y^1_{C[j]})^T,\ldots,(y^N_{C[1]},\ldots,y^N_{C[j]})^T\right)$;

9: $\boldsymbol{X}' = \begin{pmatrix}\boldsymbol{X}\\ \boldsymbol{Y}^p\end{pmatrix}$;

10: $D'_{C[j+1]} = \{(\boldsymbol{X}'(:,1), y^1_{C[j+1]}),\ldots,(\boldsymbol{X}'(:,N), y^N_{C[j+1]})\}$;

11: $h_{j+1} : D'_{C[j+1]} \to \mathbb{R}$;

12:end for

13:Return Chained model $h_j, j=\{1,\ldots,m\}$;

In order to measure the degree of correlation among the targets, we use three methods to calculate them for taxi demand in all areas.

In the aspect of linear relationships, The Spearman Correlation Coefficient(SC) is a standard and traditional way of measuring linear correlation.

In the aspect of non-linear relationships. Feature Importance(FI), also called variable importance, is a well-known method that can be used to measure how important a feature is to the prediction target. Feature Importance can describe the non-linear relationship among the targets.

The relationships among taxi demand in different areas are likely to be a very complex relationship. Using linear or non-linear approaches only to measure may ignore much useful information among the targets. Maximal Information Coefficient (MIC) is an excellent way to calculate data relevance [23]. Reshef et al. develop an approximation algorithm for computing MIC[23]. According to the nature of MIC, MIC has generality that when the sample size is large enough, it can capture a variety of interesting associations and is not limited to specific function types (such as a linear function, an exponential function, or a periodic function). The complex relationship of general variables can be modeled not only by a single function but also by a superposition function. MIC has the advantage of equitability, which allows it to pick out the strongest relationships in the target.

C. *Dynamically Adjusted Regressor Chain (DARC)*

In the previous studies that used the chained approach to solve the problem, the number of targets in their datasets did not exceed 16. Hence, the researchers have not considered that when treating targets as inputs, feature redundancy may occur. Feature redundancy or the existence of irrelevant features in the model may reduce the prediction accuracy of the model and increase the discrepancy. However, the number of targets in our dataset is 37, and we are more concerned with the degree of interaction among targets——further down to each pair of targets. Our hypothesis is that in the RC with a large number of targets, as the number of targets (which are added as features) increases, the possibility of feature redundancy in the base-regressor increases.

ERC reduces the impact of the order of the RC model on predictive performance by ensemble multiple sequential regression chains. SVRCC had begun to explore how to use the correlation among the targets to get the optimal chain ordering. There are several reasons for the predictive performance factors that affect the regression chain, as summarized above. SVRCC and ERC only consider the chain order to improve performance. By mitigating the discrepancy in each base-regressor, it is also an effective way to mitigate the overall discrepancy. This article also uses the correlation among the targets to improve the predictive performance of RC. Our approach not only sorts the chain ordering but also dynamically adjust base-regressor. As analyzed above, feature redundancy or the presence of irrelevant features in the model may result in reduced prediction performance. Irrelevant or redundant features may confuse the learning algorithm [24]. Hence, we eliminated the redundant features or irrelevant features of the base regressor in RC, which is equivalent to make a filter feature selection. We proposed algorithm2 called Dynamically Adjusted Regressor Chain.

The DARC was modified on the basis of the MCRC, so the first three steps of the DARC are identical to the MCRC. The difference is that the DARC obtains a target variable set $\boldsymbol{Y}_{c_j}^{MC_{m\times m}}$ that is highly correlated with the *j*th target according to the correlation matrix $COE_{m\times m}$. When $j>1$ in the jth model of RC, the model h_j is learned on the transformed training set D_j . All of the raw input spaces have been expanded by the values of the all previous target $\boldsymbol{Y}^p$ whose correlation coefficient with the output target is lower than θ has been deleted. To form an extended input space $\boldsymbol{X}'$. The pseudocode of the DARC method is shown in below Algorithm 2.

Algorithm 2: DARC training algorithm

Input: Training dataset D, *Threshold* θ

Output:Chained model $h_j, j=\{1,..,m\}$

1: $COE_{m\times m} = coefficient(\boldsymbol{Y})$;

2: $c_l = sum(COE(:,l)), l=\{1,\ldots,m\}$;

3: $C = \text{sort}(\{c_1,\ldots,c_m\},\text{decreasing})$;

4: $D'_{C[1]} = \{(\boldsymbol{X}_1, y^1_{C[1]}),\ldots,(\boldsymbol{X}_N, y^N_{C[1]})\}$;

$h_1 : D'_{C[1]} \to \mathbb{R}$;

for $j = 1$ *to* $m-1$ *do*:

$\boldsymbol{Y}^p = \left(\left(y^1_{C[1]}, \ldots, y^1_{C[j]}\right)^T, \ldots, \left(y^N_{C[1]}, \ldots, y^N_{C[j]}\right)^T \right)$;

for i=1 to C[j] ;

if $COE[C[j+1]][i] \le \theta$;

Delete the i^{th} row from $\boldsymbol{Y}^p$;

end for

$\boldsymbol{X}' = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y}^p \end{pmatrix}$;

$D'_{C[j+1]} = \left\{ (\boldsymbol{X}'(:,1), y^1_{C[j+1]}), \ldots, (\boldsymbol{X}'(:,N), y^N_{C[j+1]}) \right\}$;

$h_{j+1} : D'_{C[j+1]} \to \mathbb{R}$;

end for

Return Chained model h_j ,j={1,..,m}

V. EXPERIMENT

A. Performance measures

To compare the performance of each prediction algorithm, we use the following four evaluation indicators: root mean squared logarithmic error (RMLSE), accuracy rate (AR), root mean square error (RMSE) and mean absolute error (MAE).

$$\text{RMLSE}=\frac{1}{M}\sum_{t=N+1}^{N+M}\sqrt{\frac{1}{m}\sum\nolimits_{j=1}^{m}\left(\log\left(\widehat{y}_j^t+1\right)-\log\left(y_j^t+1\right)\right)^2} \quad (1)$$

$$\text{AR}=1-\frac{1}{M}\sum_{t=N+1}^{N+M}\frac{\sum\nolimits_{j=1}^{m}\left|\widehat{y}_j^t-y_j^t\right|}{\sum\nolimits_{j=1}^{m}y_j^t} \quad (2)$$

$$\text{RMSE}=\frac{1}{M}\sum_{t=N+1}^{N+M}\sqrt{\frac{1}{m}\sum\nolimits_{j=1}^{m}\left(\widehat{y}_j^t-y_j^t\right)^2} \quad (3)$$

$$\text{MAE}=\frac{1}{M}\sum_{t=N+1}^{N+M}\sqrt{\frac{1}{m}\sum\nolimits_{j=1}^{m}\left|\widehat{y}_j^t-y_j^t\right|} \quad (4)$$

B. Baseline

In order to confirm the validity of our model, we conduct experiments to compare our methods with 13 baselines. We can divide our baseline algorithms into three groups.

First group: These methods don't consider non-traffic factors or the interactions of taxis between different areas(targets), including the ARIMA, HH, HA, HP_MSL, and ST.

(1) HH: The average value of the demand quantity that needs to be predicted by the historical data is used as the predicted value of the corresponding future hour; (2) HA: Average value of historical taxi demands in the corresponding periods. E.g., for 1:00 pm-2:00 pm on Friday, its corresponding periods are all the historical time intervals from 1:00 pm to 2:00 pm on weekdays; (3) ARIMA: Auto-Regressive Integrated Moving Average is also the most classic method for time series prediction. In related work, many people use the ARIMA model and its variants to predict taxis demands [3]; (4) HP_MSI: The state-of-the-art in predicting the number of bikes to be rented from or returned to each bike station [25]; (5) ST(XGBoost): Extreme Gradient Boosting Chen et al. described a scalable end to end tree boosting system named XGBoost and a novel sparsity-aware algorithm for sparse data is designed. Comparing with deep learning method, the advantage of this non-linear model is that the simple model structure avoids the need to redesign the model due to scene changes [26].;

Second group: These methods consider the interaction between targets and put all targets as the feature into prediction models, including the ERC, SVRCC, and MCRC_SC, MCRC_FI, MCRC_MIC.

(1) ERC: Ensemble of Regressor Chains. In this paper, we selected ten random chains, and the final prediction values are obtained by taking the mean values of the ten predicted values for each target [20]. (2) SVRCC：Support Vector Regressor Correlation Chains.[21] (3) MCRC_SC: Max-Spearman correlation order Regressor Chain based on Xgboost [21]; (4) MCRC_FI: Maximum Feature Importance Regressor Chain based on XGBoost; (5) MCRC_MIC: Maximum maximal information coefficient Correlation Regressor Chain;

Third group: These methods not only consider the interactions of taxis between different areas(targets) but also avoid the possible features redundancy by removing irrelevant features, including DARC_SC, DARC_FI, DARC_MIC.

(1) DARC_SC: Max-Spearman correlation Dynamically Adjusted Regressor Chain. (2)DARC_FI: Maximum Feature Importance Dynamically Adjusted Regressor Chain. (3)DARC_MIC: Maximum Maximal Information Coefficient Correlation Dynamically Adjusted Regressor Chain.

C. Experimental results and discussion

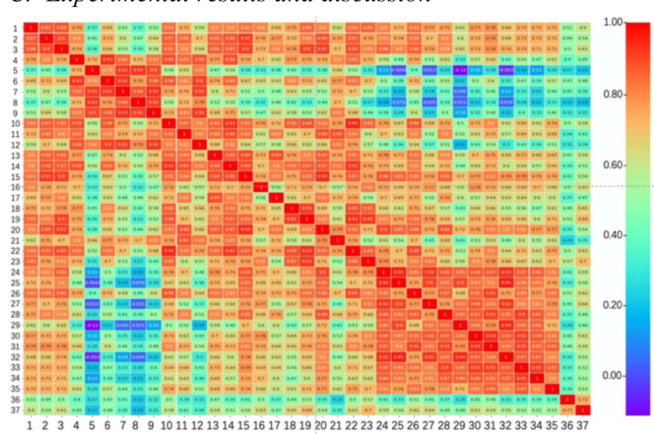

Fig. 12. Spearman Correlation Coefficient matrix for taxi demand in 37 areas

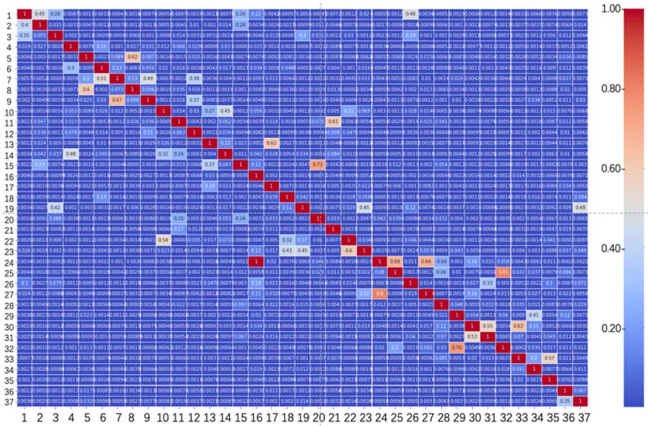

Fig. 13. Feature Importance matrix for taxi demand in 37 areas.

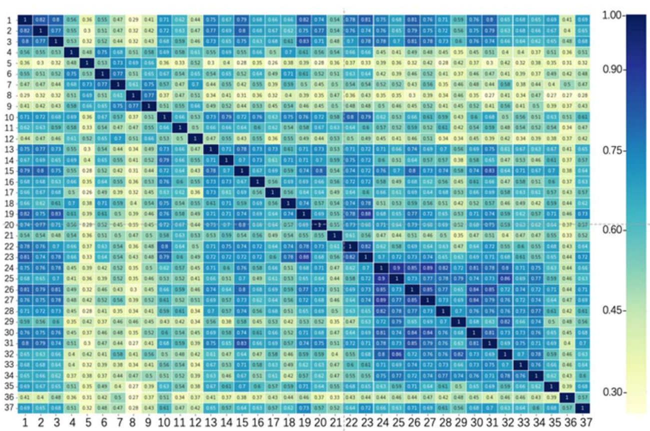

Fig. 14. Maximal Information Coefficient matrix for taxi demand in 37 areas.

The Fig.12, Fig.13, and Fig14 proved that the MIC measures the relationships better than the other two methods. As we can see from the fig.13. Features Importance always gives a high score to a few targets and gives other targets a low score due to FI's characteristic, which is clearly incorrect. For example, when there is a pair of very similar feature variables, the Feature Importance will give a high score to one of the feature variables and a low score to another. However, both variables deserve a high score. Thus, FI not as good as the other two methods.

As we can see from the Fig.12. Some variables present a strong linear relationship, Such as grid2 and grid3, grid6 and grid7,grid8 and grid9,grid14 and grid15,grid18 and grid19,grid24 and grid25,grid30 and grid31 and so on. MIC also gives a high score on these pairs of variables above. We can see that the correlation coefficients between MIC and Spearman are mostly similar, but there are still some differences. For example, the score of MIC for grid7 and grid27 is 0.56, while the score for SP for grid7 and grid27 is 0.24.

The Fig.15.is the statistics of the number of taxis from grid5 to other grid areas from June 2nd, 2014, to June 8th, 2014.

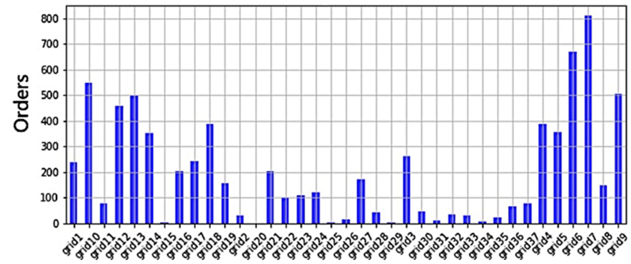

Fig. 15. The number of taxis from grid5 to other grid areas from June 2nd, 2014, to June 8th, 2014.

We can see from Fig. 15 that the number of taxis from grid5 to grid2, grid15, grid20, grid25, grid26, grid29, grid31, etc. is very small. However, there are still many taxis departing from the grid5 area to the grid27 area. At least it shows that there is a taxi relationship in the grid5 area to the grid27 area. So MIC is better than the SC and FI.

TABLE II. THE RESULTS OF EACH PREDICTIVE METHOD

Algorithm	RMLSE	MAE	RMSE	AR
DARC_MIC	**0.1470**	**29.3711**	**41.4155**	**91.80%**
MCRC_MIC	0.1592	29.6268	41.6308	91.68%
DARC_SC	0.1517	30.1231	42.5309	91.61%
MCRC_SC	0.1470	30.1549	42.7574	91.60%
DARC_FI	0.1507	30.5389	42.8765	91.59%
MCRC_FI	0.1508	29.6771	42.0429	91.70%
SVRCC	0.1851	39.7677	59.2328	88.88%
ERC	0.1521	30.17009	42.9159	91.59%
ST	0.1556	33.0816	46.5256	90.89%
ARIMA	0.6694	233.4744	178.4519	15.40%
HA	0.2567	95.2458	75.2585	81.01%
HH	0.4322	159.1984	122.2743	53.42%
HP_MSI	0.2147	44.6982	59.2368	77.18%

The critical point of this paper is that in a multi-target regression problem, other things being equal, the better the interaction between targets is utilized, the better the model performance will be. From Table 2, The DARC_MIC achieves the best performance comparing with the remaining models, and the maximum value of the accuracy is 91.80%, which proves that the DARC_MIC makes the best use of interactions of taxi demand between 37 areas in this study.

Predictions for taxi demand do not consider the possible interactions between taxi demand in different areas, which can reduce forecast accuracy. The poor predictive performance of the methods in the first group due to the lack of consideration of the non-traffic features or interactions of taxi demand in various areas.

For ERC and SVRCC in the second group, The predictive performance of MCRC_MIC, MCRC_FI, and MCRC_SC is better than ERC, which implies that MCRC methods perform better than ERC. The forming process of the ERC determines its performance to be inferior to that of the best performing regressor chain. SVRCC is very similar to MCRC_SC. SVRCC has a different base regressor with MCRC_SC cause their different performance, XGBoost performs better on this dataset than a Support Vector regressor, another reason may be that the method of SVRCC measures linear correlation to a lesser extent than SP.

For all the MCRC and DARC methods in the second group and third group, These methods have all taken interactions between targets into account. The reason for their differences in performance is that they differ in the accuracy with which they measure interactions between targets, this indirectly causes the differences in the targets' ordering in the three MCRCs and the differences in the removal of 'irrelevant features' in the three DARCs.

Further speaking to MCRC methods, MCRC_MICget the best performance in three MCRC methods proves that the taxi demand relationship between different areas is complex and diverse, and MIC is better than the SC and FI. The performance of the MCRC_FI, MCRC_MIC, MCRC_SC is better than the ST methods implies that the performance of RC could exceed ST as long as the interdependence between targets is carefully explored. Thus, ordering the targets makes the prediction model better, once estimating the interactional degrees accurately, and MCRC_MIC possesses this characteristic since it has an efficient ability to measure the dependence of variables.

Further speaking to DARC methods, the predictive performance of DARC_MIC is better than that of MCRC_MIC, indicating that our hypothesis is correct: the sub-model h_{cj} of the regression chain generated by data training has irrelevant features. We found the following feature redundancy caused by irrelevant features affects the base regressors appear later in the chain because not all targets have strong dependence with others, especially for the very long RC. The DARC is a method that carefully determines the order of the chain by dynamically adjusted regressor chain. It not only can reduce the computational cost, but also increase the stability of the predictive results, and it also possesses advantages with long-chain situations.

For RC and multi-target problems, through comparing these predictive methods, we found the interactional degree among targets significantly affects the RC performance when the dataset has a large number of targets. Estimating the interactional degree, and implementing prediction according to these interactional degrees, would improve the predictive ability effectively. Therefore, different approaches can be used to estimate the interactional degrees as the diversity of multi-target, e.g., linear, non-linear, complex, etc. After accurately measuring the degree of interaction between the goals, further exploitation of the relationships between the goals will lead to better performance of the RC model, such as sorting, removing irrelevant features, etc. Making good use of interdependencies between targets is also a meaningful way to address the multi-target regression problem.

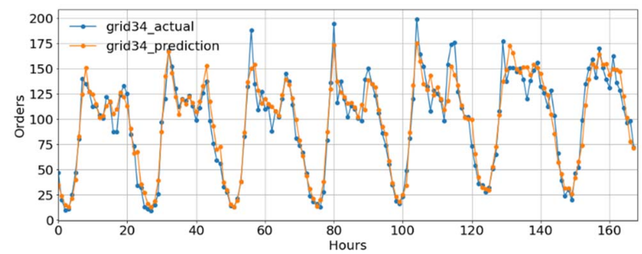

Fig. 16. The fitting curve of DARC_MIC with area grid34

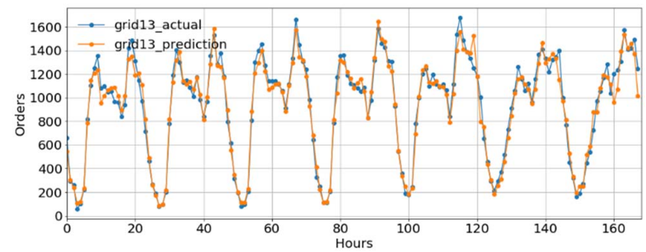

Fig. 17. The fitting curve of DARC_MIC with area grid13

Figure 16 depicts the excellent fit for most of the time in this study in RMSE. In contrast, Figure 17 depicts the bad fit for most of the time in this study in RMSE. We can see the high accuracy of DARC_MIC's predictions of taxi demand. (Only 1.11km *1.11km per areas except for the two airports).It is great for real-world applications

VI. CONCLUSION

In this paper, we proposed a novel method named dynamically adjusted order regressor chain for improving the prediction of taxi demands. The predictive area is divided as multiple subareas, and 37 of which are kept as the predictive targets. Several of the time and the climatic factors are chosen as the training features through analyzing the dataset. 3911 hours and the following 168 hours of NYC dataset were used to train and test, respectively. The comparable performance of 13 approaches shows chained methods could exceed the ST method on the premise of combining prior knowledge, even though in the case of a large number of targets (more than twice the target number of datasets used in previous regression chain applications). MCRC methods outperforming the ERC method implies that the order of RC is significant, which can be reflected by their interactional degree. The existing methods such as RC, ERC, and SVRCC are very likely to be unsuitable for solving this problem that the number of predicted targets is large, and DARC's dynamic adjustment ability is very advantageous. This proposed DARC_MIC method can provide an effective way for taxi demand predicting problems.

Exploiting the relationship between multiple objectives is the biggest challenge in multi-target regression problems, and this method of measuring the degree of interaction between objectives can be used to improve many other multi-target regression methods, this is a direction for future work.

ACKNOWLEDGMENT

We would like to thank Yimin Wen for writing advice.We would like to thank Ze-Xing Sun and Ji-Peng Wei for Writing

Refreshment at first. Thanks also go Jian Wen for the original work and people who helped this work in Machine Learning and Media Mining Lab.

.

REFERENCES

[1] Zhao K. Khryashchev D.; Freire J.; Silva C. and Vo H. Predicting taxi demand at high spatial resolution: approaching the limit of predictability. In Proceedings of 2016 IEEE International Conference on Big Data (Big Data), Washington DC, USA, 2016, pp. 833–842.

[2] Yao H. X.; Wu F.; Ke J. T.; Tang X. F.; Jia Y. T.; Lu S. Y.; Gong P. H. and Ye J. P. Multi-target regression via random linear target combinations. arXiv:1802.08714, http://arxiv.org/pdf/1404.5065, Feb. 2018.

[3] Shekhar S.; Williams, B. M. Adaptive seasonal time series models for forecasting short-term traffic flow. Transportation Research Record: Journal of the Transportation Research Board, 2008, 116–125.

[4] Moreira-Matias L.; Gama J.; Ferreira M.; Mendes-Moreira J.; and Damas L.; Online predictive model for taxi services. In Proceeding of Advances in Intelligent Data Analysis XI, vol. 7619Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 230–240.

[5] Moreira-Matias L.; Gama J.; Ferreira M.; Mendes-Moreira J. and Damas L. Predicting taxi–passenger demand using streaming data. IEEE Transactions on Intelligent Transportation Systems, 2013,14(3): 1393–1402.

[6] Moreira-Matias L.; Gama J.; Ferreira M. and Damas L. A predictive model for the passenger demand on a taxi network. In Proceedings of 2012 15th International IEEE Conference on Intelligent Transportation Systems, Anchorage, AK, USA, 2012, pp. 1014–1019.

[7] Li X.; Pan G.; Wu Z. H.; Qi G. D.; Li S. J.; Zhang D. Q.; Zhang W. S. and Wang Z. H. Prediction of urban human mobility using large-scale taxi traces and its applications. In Proceeding of Frontiers of Computer Science. 2012, pp. 111-121.

[8] Pan B.; Demiryurek U. and Shahabi C. Utilizing real-world transportation data for accurate traffic prediction. In Proceedings of 2012 IEEE 12th International Conference on Data Mining, IEEE, Brussels, Belgium, 2012, pp. 595–604.

[9] Rodrigues F.; Markou I.; Pereira F. C. Combining time-series and textual data for taxi demand prediction in event areas: a deep learning approachJ. Information Fusion, 2019,49:120-149.

[10] Yu R.; Li Y.; Shahabi C.; Demiryurek U. and Liu Y. Deep learning: a generic approach for extreme condition traffic forecasting. Society for Industrial and Applied Mathematics, 2017, pp. 777–785.

[11] Chang H.; Tai Y. and Hsu J. Context-aware taxi demand hotspots prediction. International Journal of Business Intelligence and Data Mining, 2010, 5(1): 3-18.

[12] Wang D.; Cao W.; Li J. and Ye J. DeepSD: supply-demand prediction for online car-hailing services using deep neural networks. In Proceeding of IEEE 33rd International Conference on Data Engineering (ICDE), San Diego, CA, USA, 2017, pp. 243–254.

[13] Tong Y.; Chen Y. Q.; Zhou Z. M.; Chen L.; Wang J. Yang Q.; Ye J. P. and Lv W. F. The simpler the better: a unified approach to predicting original taxi demands based on large-scale online platforms. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD'17, Halifax, NS, Canada, 2017, pp. 1653–1662.

[14] Spyromitros-Xioufis E.; Tsoumakas G. and Groves W. Multi-label classification methods for multi-target regression. Computer Science, 2012. 43: 1-9

[15] Mastelini S. M.; Santana E. J.; Cerri R. and Barbon S. DSTARS: a multi-target deep structure for tracking asynchronous eegressor stack. In Proceedings of Brazilian Conference on Intelligent Systems (BRACIS), Uberlandia, 2017, pp. 19–24.

[16] Mastelini S. M.; Da Costa V. G. T. and Santana E. J. Multi-output tree chaining: an interpretative modelling and lightweight multi-target approach. Journal of Signal Processing Systems, 2018, 91(2): 191–215.

[17] Santana E. J.; Geronimo B. C.; Mastelini S. M.; Carvalho R. H.; Barbin D. F.; Ida E. I. and Barbon S. Predicting poultry meat characteristics using an enhanced multi-target regression method. Biosystems Engineering, 2018 (171): 193–204.

[18] Santana E. J.; Mastelini S. M.; and Jr S. B. Deep regressor stacking for air ticket prices prediction. In Proceedings of XIII Brazilian Symposium on Information Systems, Lavras, Brasil, 2017, pp. 25-31.

[19] Tsoumakas G.; Spyromitros-Xioufis E. and Vlahavas I. Drawing parallesl between multi-label classification and multi-target regression. International Workshop on Multi-Target Prediction, Nancy, France, http://users.auth.gr/users/0/9/022090/public_html/publications/slides/spyromitrosECMLPKDD14MTPW.pdf 2014.

[20] Spyromitros-Xioufis E.; Tsoumakas G.; Groves W. and Vlahavas I. Multi-target regression via input space expansion: treating targets as inputs. Machine Learning, vol. 2016, 104(1):55–98.

[21] Melki G. Cano A.; Kecman V. and Ventura S. Multi-target support vector regression via correlation regressor chains. Information Sciences, 2017, vol. (415–416): 53–69.

[22] Data: http://www.nyc.gov/html/tlc/html/about/trip record_data.shtml

[23] Reshef D.N., Reshef Y.A., Finucane H.K., Gross man S.R., McVean G., Turnbaugh P.J., Lander E.S., Mitzenmacher M. and Sabeti P.C. Detecting novel associations in large datasets. Science. 2011,334(6062): 1518–1524.

[24] Yu L. and Liu H. Efficient feature selection via analysis of relevance and redundancy. Journal of Machine Learning Research, 2004, 5:1205-1224.

[25] Li Y.; Zheng Y.; Zhang H. and Chen L. Traffic prediction in a bike-sharing system. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems-GIS'15, Bellevue, Washington, 2015, pp. 1–10.

[26] Chen T. Q. and Guestrin C. XGBoost: a scalable tree Boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD'16, San Francisco, California, USA, 2016, pp. 785-794.