

PRIMARY RESEARCH

The bus arrival time prediction using LSTM neural network and location analysis

Yerkezhan Seitbekova ^{1*}, Bakhytzhan Assilbekov ², Iskander Beisembetov ³, Alibek Kuljabekov ⁴^{1, 2, 3, 4} Satbayev University, Almaty, Kazakhstan

Keywords

GPS data processing
Bus arrival
Time prediction
Traffic analysis
Bus dwelling time
LSTM NN
Location analysis

Abstract

The accurate bus arrival time information is crucial to passengers for reducing waiting times at the bus stop and improve the attractiveness of public transport. GPS-equipped buses can be considered as mobile sensors showing traffic flows on road surfaces. In this paper, we present an approach that predicts bus arrival time using historical bus GPS information and real-time situation on the road. In this study, we divide bus arrival time into bus dwelling time at bus stops and bus travel time between stations and predict each of them separately. The clustering approach used to predict the travel time between stations, and then for each cluster, we apply LSTM NN to predict walking time between stations. The latency at each bus stop we evaluate by historical dwelling time and using location analysis to find the importance of the bus stop as a point of interest during prediction time. The study is trained and tested on GPS data collected from 1200 buses in a period of 3 months. According to tests results our method show small mean absolute error for buses that not far from departure station. The outcomes of this work can be used as an additional information for bus passengers to know possible bus coming time and to estimate possible travel time in bus journey. The method for arrival time prediction proposed in this research has several advantages. It considers historical bus travel time information, real time information, bus dwelling time, riding time, traffic lights and city facilities.

Received: 6 April 2020**Accepted:** 13 July 2020**Published:** 16 November 2020

© 2020 The Author(s). Published by TAF Publishing.

1. INTRODUCTION

In contemporary life with high technologies, analyzing and extending data of human mobility have become significant topic of research. Generating human mobility patterns helps for such surveys as traffic prediction, migration movements, epidemic distribution. Active mobility of megapolis inhabitants is one of the crucial points for developing and modern conglomerations. Mobility system in urban areas is ways of individuals travel around the city, for example walking, private vehicles and public transport.

In huge urban areas, transit vehicle, such as metro, buses, trams, trolleybuses, has many advantages in various aspects. It is beneficial to reduce traffic congestion, decrease air pollution, positive influence on nation welfare and health etc. Government of Kazakhstan strongly encourages using public transport by adding road lanes for buses and trolleybuses, providing facilities for pupils, seniors, and

disabled people. It results in increasing popularity of public transit, especially in Almaty. Such demand pushed developing applications and systems for making using public transport more convenient. However, local infrastructure is not structured and developed enough to provide population with fixed schedule and smooth transport feed. Local buses' arrival to stops is reorganized and depends on several external factors as distance, traffic lights and jams on the road. Therefore, evaluating arrival time is possible only by retrieving actual real time data. Existing bus tracking systems, such as City Bus, Almaty Bus etc. do not provide users with information about waiting time.

The aim of paper is to identify main factors of delays on the roads of Almaty and design logical system for predicting bus arrival time. The object of research Almaty city bus transportation system. Study subject reasons of buses delays and their forecast.

*Corresponding author: Yerkezhan Seitbekova

†email: y.seitbekova@satbayev.university



To estimate arrival time of bus to exact bus station, the following steps should be taken:

- Find the nearest upcoming bus to the station, identify its location;
- Count number of bus stops, where that bus will make a stop. And estimate how many times a bus will delay at these stops;
- Count the number of lines bus should go, and estimate travel time between stations for each bus line.

Data for this research was collected and tested on data provided by bus tracing company citybus.kz of Almaty city. From this site we can take information about current location of buses, their speed, and direction. Using GPS information, the first step can be done without problems. But for the second and third steps we need to build models to predict bus dwell time and bus travel time between stations.

II. LITERATURE REVIEW

Bus arrival time prediction is an important component in urban public transportation system. This problem is attracted many researchers around the world. The main popular approaches are Artificial Neural Network (ANN), Kalman filtering, Support Vector Machines (SVM), regression analysis models and time series models.

The most popular approach is to use ANN for arrival time prediction [1, 2, 3, 4, 5, 6]. This approach use different kind of neural network structure to travel time forecasting. This process consists of two process: training model using huge amount of historical data, then validation on testing data. In these works, it is considered that the current bus arrival time has a certain connection with that of buses in front of it. Meanwhile, bus arrival time is affected by road, environment, and other factors.

The Kalman filter [7, 8] is a model-based estimation technology that has the ability to accommodate traffic fluctuations with their time-dependent parameters adequately. It takes into account the stochastic nature of process disturbances and measurement noise.

SVM [9, 10, 11] has been proposed in time series forecasting. By integrating multiple sources of information, previous works proved that support vector machine can accurately predict bus arrival times. Although SVM also depends on the similarity between historic and real-time traffic patterns, it still shows many breakthroughs and plausible performance in forecasting of bus arrival time.

Regression analysis models [12] determine a dependent variable with a complex mathematical function formed by a set of independent variables.

Time series models [13] are also used for predicting bus ar-

rival time. It is assumed that all external factors of the system either are accounted for or are constant.

Most studies that analyze bus travel time considers only link running time. Despite that the dwell time plays a key role in the bus travel time, some models do not predict it separately from the running time. However, running time in links and dwell time at stops are affected by different issues. According to study done by Padmanaban and Vanajakshi [14] the dwell time takes up to 26% of the bus's total travel time, especially at busy bus stops or during the rush hour. In this research the dwell time associated with the total travel times of the buses under heterogeneous traffic conditions. But only the number of passengers alighting, and boarding was considered in the dwell time prediction. Number of alighting and boarding passengers, fare payment method, passenger's movement and capacity of bus stop are key factors of bus dwelling time according to Transit Capacity and Quality of Service Manual [15]. Dueker [16] found other factors as time of day, type of route, and type of vehicle as additional components of bus dwell time. The connection between bus dwelling time and the existence of steps at doors, the age of passengers, and the possible friction between users boarding, alighting, and standing passengers are studied in work [17]. Recent studies by Chioni [18] studied factors affecting bus bunching by a geographically weighted regression approach, and found that number of lanes, number of routes per stop, the distance to the nearest metro station influence to bus bunching. Which is according to conclusion of Verbich et al., [19] increases both dwell time at stops and bus travel time, because that the longer a bus has been servicing a stop, more time is added to the dwell time of the subsequent bus.

In this paper we predict bus arrival time using real time information about traffic condition and city facilities.

III. METHODOLOGY FOR EVALUATING BUS ARRIVAL TIME

A. Data Collection

For outlined goals necessary data about public transport bus services was obtained from public API provided by City-Bus with two kind of links in json format. First link consisted list of all bus stops and their details such as: ids, names, location and what buses are serving the station. Second type contained individual route information: route number, list of served stations, route coordinates, serving buses specifications such as ids, manufacturers, their current locations, azimuths, and speed. Location is measured in geographic coordinates latitude and longitude.

For this research we collected from 1200 buses in a period

of 3 months. In Almaty city there are 136 bus lines and approximately 8 to 10 buses on each line. To train and test models we collect every minute GPS data from 7:00 to 19:00 for May, June and July 2020 year.

Route is represented as a multiple consistent coordinate with a distance around 10 meters. There are 2 types of functioning routes in Almaty one way and circular with additional loops. Circular routes can be identified with a special key in provided json data. To collect and parse all these data was used Python programming language.

Another data was retrieved from openstreetmap.org which is free open world map data. OpenStreetMap provides an API with raw geodata and sophisticated database. It was used to gather all traffic lights locations which were necessary for evaluating bus arrival time and delay. Since these data is static it should have been collected once was used overpass turbo - OpenStreetMap interface with its own query.

In OpenStreetMap has its own system and structure, for example there are different types of data most common are nodes, ways and relations which are connected to each other. In these occasion traffic signals were represented as

a node which belong to the ways which are features of the map in defined area. Same way from overpass turbo was retrieved location and amenity or buildings. For example, amenities can represent universities and school, types can detect malls, marketplaces, pharmacies ed cetera.

B. Distance Estimation

As one of the main features of application was recommending nearest bus and predicting its arrival time of the bus to the bus stop. For achieving this goal firstly, it was required to estimate distance between bus and users starting station. Distance was necessary to assess time required for bus to reach the station for following calculations. Solution of this problem was to use formula adapted for spherical surfaces or as it called in spherical trigonometry great circle distance. For this purpose, was decided to use Haversine formula to determine distance from one point to another [20]. It requires as an input data two point coordinates and radius of the Earth which is common knowledge. All coordinates should be polar, thus latitude and longitude are converted into radians.

Haversine formula is derived and given below (2):

$$a = \left(\sin \frac{\text{lat}2 - \text{lat}1}{2} \right)^2 + \cos(\text{lat}1) \cos(\text{lat}2) \left(\sin \frac{\text{lon}2 - \text{lon}1}{2} \right)^2 \quad (1)$$

$$d = 2R * \arcsin(\min(1, \sqrt{a})) \quad (2)$$

Here lat1, lon2 are latitude and longitude of the first point and lat2, lon2 of the second, respectively. R is a Earth radius that is assumed as 6373 kilometers, since radius units are in kilometers same would be for the final distance [21].

Haversine formula solves distance problem, however it was still not applicable to compute how far apart bus from bus stop, for the reason that haversine formula connects two points, but buses do not approach stations straight though following prescribed route. From the collected data described in previous chapter there were station, current bus location coordinates and route. How many kilometers bus should pass can be derived by suggested method? Since all route is represented as consistent points it was divided into multiple segments.

Using haversine formula from all route points computing distance to the bus, to find closest point to the bus. Then important step is to identify whether previous or following points are closer to the bus to define in what order bus station would be in list of route points. Same procedure should be done with buses location to detect closest point to the

bus. Next haversine formula is applied from closest point to the bus to the following point in route towards to the bus stop, which is recognized simply if the bus closest points index is less than stations closest points index then station would be after the bus relatively to the list of route points. This distance should be saved, after that from next point to another calculated haversine formula and result is summed with previous one. Similarly, it is done from one point to the following as a small segment until it reaches the station closest point. All segments are summed up in total. Last step is found distance between bus closest point to the bus by itself and the same operation for station with its closest point respectively. After what those distances are summed with total distance obtaining final distance which bus should pass from its current location to the bus station.

IV. BUS STATION QUANTITY AND DWELL TIME ESTIMATION

During the travel bus makes stops at stations to load and unload passengers. Determining number of stops bus would

proceed from start and final destination it is necessary to distinguish the time it would spend on that what would be called dwell time.

The list with stations of bus route is iterated to get index of the closest station to bus location ignoring opposite located station with same name but different id. It is accomplished by haversine formula, which was introduced in previous chapters. It is important to diagnose whether the closest bus stop was passed or not, since it can be also taken into consideration.

Shown example on the Figure 1 is illustrating bus with number 015DE02 moves from “Esentai Mall” to “Ramstore” station, where user is expecting it. As it can be noticed bus is close to “Shashkina” station, but algorithm would take it into account and begin count from next one. In given case, is “Zhamakaeva” station.

Computations compare distances between the closest station and start station with distance between bus and start

station. If first distance would be less than second one, iterator would start from the closest station and quantity would be estimated by module of subtraction of two stations indexes.

It is quite often situations when buses stay on stops more than 10 minutes to load more passengers, what causes irregularity in bus schedule and provokes other long stops to maintain bus's arriving. It may seem like impossible to stay at station almost 10 minutes, but for Almaty bus drivers it is common. Buses may stay at popular stations, near markets, universities and cityside bus stations, or at starting point to wait and collect passengers. And this situation effects on public transport responsibility. The example shown in Figure 2. In this figure we illustrated average dwelling time of buses in line 79, where circle radius corresponds to average dwelling time. As you can see, buses makes a long stop at the first and last stations, and makes longer stops near universities and markets.



Fig. 1. Station counting problem example

According to the most valuable studies in this area, there are many aspects that influence at time waiting. There are several studies done to estimate dwell time, for example, formulas (3) suggested for estimation dwell time, which was

suggested by Levinson [9].

$$t_d = t_0 + n \cdot t_{ba} \quad (3)$$

where, t_d is dwell time at bus stop; t_0 is time required to open and close bus doors; n is number of boarding and alighting passengers; t_{ba} is the average time required by

each passenger to board or alight a bus. This formula is suitable to the stations, where buses do not make long stops to collect passengers. To estimate dwelling time, we make

location analysis to identify the attractiveness of stops at each time period. For most attractive stations we predict bus dwell time by historical average.

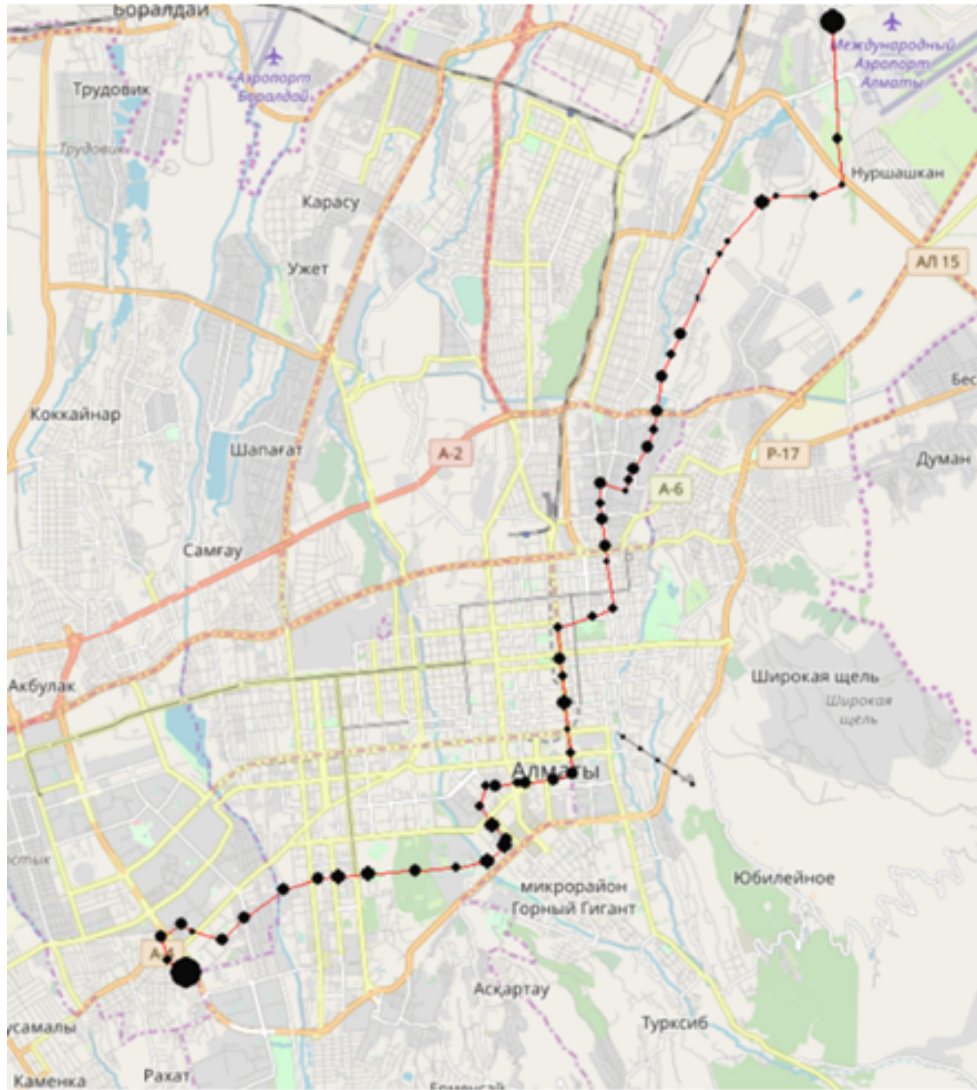


Fig. 2. Bus line 79

Historical average - is used in most studies in prediction bus dwell time [22]. This model takes average dwell time at a bus stop at the same time in previous weeks. The Historical Average (HA) method is a data-driven approach. Thus, we need to collect data about dwelling time previous buses. The historical average dwell time at the bus stop p at time t_0 can be calculated as:

$$HA(p, t_0) = \frac{\sum_{i=0}^d t_{i,p,t_0}}{d} \quad (4)$$

Where t_i is the observed dwelling time of the bus at station p during the period t_0 in i the day of the dataset. Each location has different conditions. There is no doubt that dwell time is affected by a lot of factors, also in huge cities, where

many accidents happen, and city dwellers are extremely active.

A. Traffic Signals and Crossroads

Predicting time delay cannot be fulfilled without considering travel time spent on passing traffic lights what can be quite challenging [23]. Traffic signals in Almaty generally equipped on heavy traffic road with two more lines. Since traffic light system is mechanical and cannot cover all factors of traffic flow most of the traffic congestion happens due to improper time management. Alejandro Tirachini evaluated the time bus spend on road intersection with traffic light in average 11.3% of all time travelling along route

[24].

Traffic light cycle is full period of all light stages, for example if the traffic light is consisting of red, yellow green light, cycle would be equal to the sum of all phases. Interval of time spent on crossroad can be equal to the full traffic light cycle if a vehicle approaches to the start of red light. According to the Websters method, calculation of traffic signal timing adequate time for the green light for vehicles or time required for traffic flow to pass the road is an effective green period [25]. Can be estimated by formula (5).

$$g = y(c - L)/Y \quad (5)$$

Where:

g = effective green period

y = flow factor

c = cycle time

L = total lost time

Y = sum of y flow factor

$c-L$ = total effective green time

While pedestrian green light should be sufficient to pass by walk from the beginning of green light which should be calculated by suggested formula (6) by Vladimir Gorodkin, Zlata Almetova and Vladimir Shepelev stated in Procedure for Calculating On-Time Duration of the Main Cycle of a Set of Coordinated Traffic Lights [26].

$$t = 5 + S/V \quad (6)$$

Where: t is sufficient time for crossing road, S is the width of the road which should be passed by pedestrians, V average velocity of pedestrians.

Unfortunately, theoretically time delay on traffic signals cannot be measured due to the insufficiency of data and individual conditions on each intersection like number of lines, traffic flow, traffic signal cycle. Thus, the empirical method would be applied to the study as a solution.

Obtained data from OpenStreetMap allows to define location for all traffic signals. Methodically process work was based on counting number of road intersections, which recommended bus by algorithm described in previous chapters should pass to reach users start station. To achieve this goal all Almaty traffic signals are appended to the list. After that it is required to select from that list all traffic lights which belong to the recommended bus route. For this purpose, distance from each traffic light from list containing all traffic lights to each point of bus route is measured by Haversine method to identify which crossroad is passed by bus. Therefore, from route list should be identified closest point to the bus. To determine whether bus station would

be in list before or after bus it is necessary to compare the neighboring point to the obtained one.

From the closest point of bus and closest point to station looped through distance from traffic light and route points is determined to discover how many crossroads left to pass by bus. Then for each crossroad would be added 45 seconds for the reason it is 120 seconds in average for full cycle of traffic light.

B. Traffic Congestion Affection

Traffic congestion is integral part of every urban areas and Almaty especially due to the high rate of agglomeration. Traffic flow is dynamic parameter relying on geo spatial conditions including period of time. Most of the methods of traffic probe in real time estimation apply sophisticated mathematical approaches using variety of data including GPS and data from video registrars. Moreover, such kind of work demands a lot of expenditures and equipment's. Nevertheless, it is possible to predict traffic jam empirically using attractors concept.

Attractors concept is based on predicting places of mass visiting by people. According to the article in Journal of Geographic Information System attractors identified as point causing traffic fluxes, namely places attracting certain number of drivers in a particular period of time [27]. For example, school with 600 students would generate a concentration of students, employees and parents in the morning and in the evening, what makes it an attractor as it provokes a lot of passengers to reach that site in particular period of the day.

Caused probes which indicating by red and yellow colour as very heavy and heavy condition respectively. Traffic jam fluxes surround two schools: one public, one private, and three universities. Such kind of calculation gives quite rough result. but allows to produce predictions of probes in area.

All attractors should be assigned parameter weight indicating number of people associated with the site. Value of attractors equals to following formula (7):

$$V = \sqrt{P * T} \quad (7)$$

Where p is weight of attractors with the scale from 1 to 10, T is period of time correlated to place.

For previous example with the school assuming that time of concentration would be from 6 A.M. to 8 A.M and 6 P.M to 7 P.M., T would be equal to 3 hours. Thus, the value of attractor would equal to 4.2 since traffic flux weight equal to 6 out of 600 of students.

By this method value of probes would be measured on

scale from 1-10, where 1-3 level is classified as “acceptable” meaning road conditions are quite free to move, from 4-7 is “mild congestion” and 7-10 “heavy congestion”.

From obtained data OpenStreetMap such buildings categorized as markets and malls, amenities as restaurants and schools would be retrieved to create database of attractors. Since it is not available to assess number of associated people to attractors, it would be roughly assigned to every type of amenity individual level of congestion. For example, schools’ value of attractors would be evaluated as 4. Each level would be divided by 10, multiplied by estimated time of bus arriving and added to that time as example in formula (8).

$$T = V/10 * t + t \quad (8)$$

Where, t is absolute time of arrival considering congestion level, V -average value of attractors, t estimated time of arrival without congestion level.

C. Travel Time Estimation

There are a lot of factors affecting on travel time between stations. The traffic lights, buildings, weather conditions were taken into account. To forecast travel time, we use LSTM NN to build model. Collected historical GPS data, weather condition and information about city facilities is used in training process. To optimize the process of training, we divide bus line segments into clusters according to distance and point of interests of each line segments. Each cluster is trained separately, and for each cluster we have different LSTM model for estimation travel time. This process is illustrated in Figure 3.

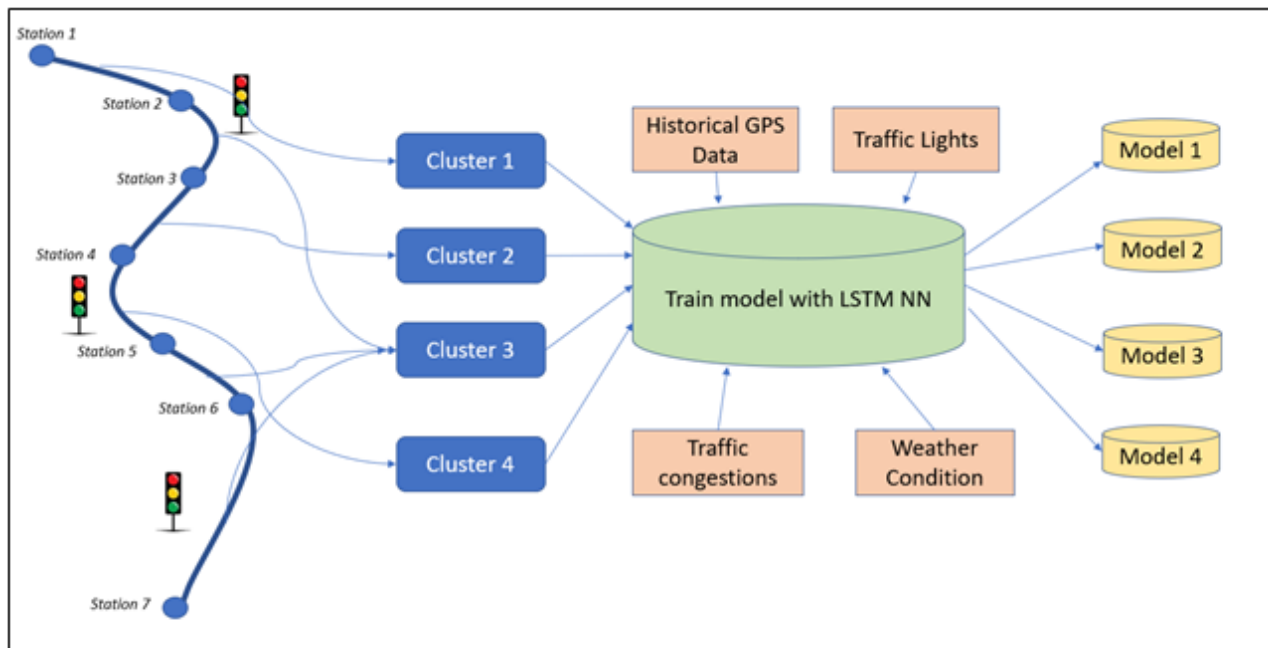


Fig. 3. Model for prediction travel time between stations

LSTM based neural networks have been used for travel time prediction and have achieved better performance in recent years [28].

The input matrix is fed into two stacked LSTM layers, where each LSTM layer has 128 neurons. The LSTM memory cell

can be described with the following equations:

$$\begin{aligned} i_t &= \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \\ f_t &= \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \\ o_t &= \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \\ \tilde{C}_t &= \tanh(W_{Cx}x_t + W_{Ch}h_{t-1} + b_c) \\ C_t &= i_t * \tilde{C}_t + f_t * C_{t-1} \\ h_t &= o_t * \tanh(C_t) \end{aligned} \quad (9)$$

where t indicates the t th timestamp, i_t, f_t, o_t refer to the output of the input gate, forget gate and output gate respectively. x_t, c_t, h_t are the input vector, state vector and

hidden vector, respectively, and h_{t-1} is the former output of h_t . \bar{C}_t and C_t are the input state and output state of the memory cell, and C_{t-1} is the former state of C_t . gis , a sigmoid function. $W_{ix}, W_{fx}, W_{ox}, W_{cx}$ are the weight matrices connecting x_t to the three gates and the cell input, $W_{ih}, W_{fh}, W_{oh}, W_{Ch}$ are the weight matrices connecting x_{t-1} to the three gates and the cell input, b_i, b_f, b_o, b_c are the bias terms of the three gates and the cell gates. All the parameters mentioned above are initialized randomly and learned automatically through backpropagation during the learning stage.

The output of the LSTM layers goes into several fully-connected layers, where each layer is of size 128. The fully-connected layers are connected with residual connections, which is shown to be effective for training a very deep neural network [29]. The residual connection adds shortcuts between different layers, thus previous information flow can skip one or more non-linear layers through the shortcut and the skipped layers just need to learn the "residual" of the non-linear mapping. For the first fully connected layer, its input is the output of the last LSTM layer. Let σ_{fi} be the i th residual fully-connected layer, then the output of the first layer is $\sigma_{fi}(o_z)$, where o_z is the output of the LSTM layer. For the rest of the residual layers, let o_{fi} be the output of the i th layer, then the output of the $(i+1)$ th layer can be represented as $o_{fi+1} = o_{fi} \oplus \sigma_{fi+1}(o_{fi})$, where \oplus is an element-wise add operation.

Finally, we apply a tanh activation function and obtain the prediction results. In order to prevent overfitting, two widely used regularization techniques are employed: dropout and L_2 regularization. The dropout mechanism is applied to each hidden layer, where the rate of dropout is set to 0.5 [30]. Moreover, we apply L_2 regularization on model weights to prevent possible overfitting. Formally, the loss function used for training the model is:

$$L_{\text{loss}} = \sum_{i=1}^F (\hat{y}_i - y_i)^2 + \lambda \|W\|^2 \quad (10)$$

where λ is a hyper-parameter to control the regularization strength and W denotes all weights in the network. The Adam optimizer is utilized as the gradient descent optimization algorithm. The training process repeats for 50 epochs.

V. SOFTWARE IMPLEMENTATION AND TESTING

The original purpose of the project is manufacturing beneficial program for bus passengers with user-friendly application interface. This study fully discloses the main principles of bus circulation in the city and how external circumstances can influence on them. Because of all limita-

tions and drawbacks of the public transit system, Almaty occupants do not have a positive impression about public transport, especially busses. It is common knowledge that cars and other fuel-based vehicles are the main suppliers of carbon dioxide. Furthermore, this kind of transport negatively affects to the environment and causes air pollution. Developing more facilities for public transport passengers can be crucial and significant step for saving clean air. In addition, there is personal advantages for everyone. Using public transport, instead of personal cars, citizens can not worry about such problems, as parking and permanent traffic jams, which are usual chores of mega polices. That is why one of the primary aspirations of this project is promoting usage of public vehicles to city inhabitants.

Another purpose was creating application that satisfies all user requirements. Existing tracking systems have several serious drawbacks, that were fixed and modified in the Bus-Tracker app. Minimalistic and user-friendly interface with a feature arrival time estimation make this system actual and functional. Calculated results by proposed algorithm were quite accurate and cover a lot of factors affecting on arrival time. Undoubtedly, empirical methods to evaluate delay time factors could not be precise, but provide clear idea of approach, including minimum amount of data available.

To test model in real time, we implement mobile application to track bus and see the estimated arrival time. iOS mobile application was chosen as a platform to test model of bus arrival time. The layout of it can be seen on Figure 4.

There are 3 pages, where the first one is starting view with only one functional button. By pressing it, user opens main view with two inputs, where he types start and end stations. Table view with all bus stops in Almaty would be presented while typing and updating by every typed character to suit user input. After filling in both inputs, user gets information about the nearest bus of every route with these two stations. Information about buses would be demonstrated as table view which can be seen on Figure 3. Every cell would contain name, specific serial number of the bus and approximate arrival time estimated by model.

We test model in real time during 3 -day period. The performance metrics used in the testing Mean Absolute Error (MAE). The formula is the following.

$$MAE = \frac{\sum_{t=1}^N |F_t - A_t|}{N} \quad (11)$$

Where t is observed arrival time and a predicted arrival time during time interval t . N is the size of testing cases. We see that the values are differentiated during the testing,

then we decided to divide test cases according to the number of stops that the waiting bus should pass. The results on Table 1. According to the results, if we are waiting for the upcoming bus the mean absolute error is about 20 seconds.

But the mean absolute error for the buses that far from departure station, then the mean absolute error is 3.77 minutes. It causes to error in prediction of dwell time in stops and travel time of each line segments.

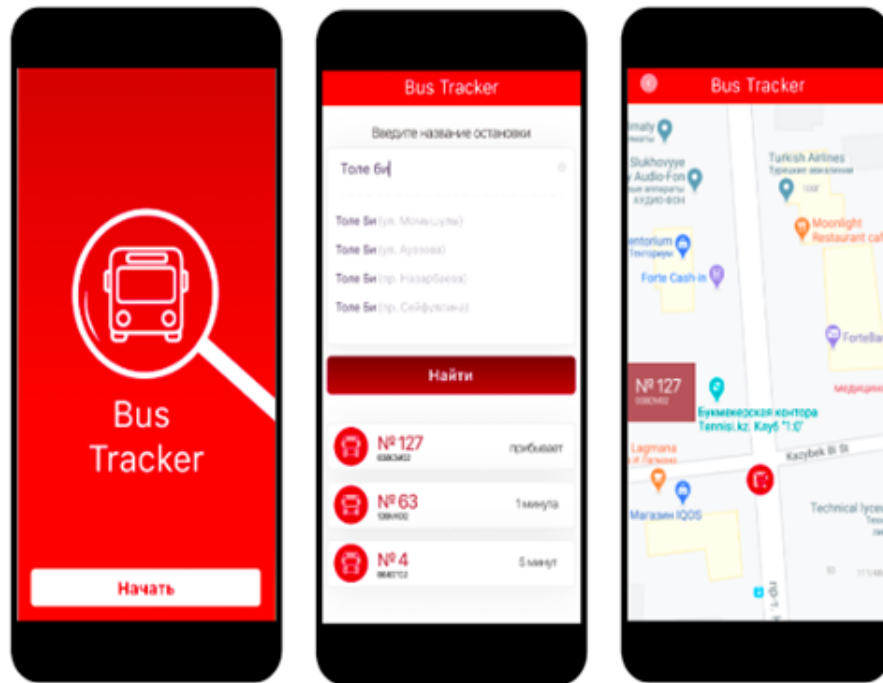


Fig. 4. Mobile application design

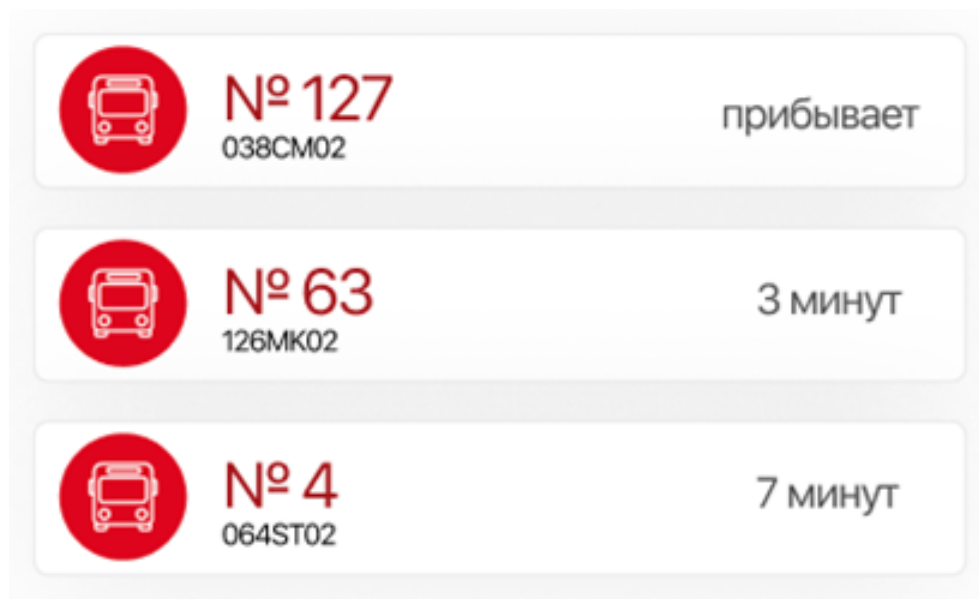


Fig. 5. Layout of information displaying

TABLE 1
THE RESULTS OF TESTING

| Bus Location | MAE (Minutes) |
|-------------------|---------------|
| 0-1 stops far | 0.31 |
| 2-3 stops far | 0.71 |
| 4-5 stops far | 1.22 |
| More than 6 stops | 3.77 |

VI. RELATED WORKS AND DISCUSSION

There are a lot of studies done on bus travel time estimation and prediction. For example, the study [22] investigated the problem of predicting bus journey time for passengers. They take into account the bus riding time and the waiting times at transfer points. But for the bus riding time they didn't consider bus dwelling time separately. Also, the next bus's waiting time at transfer point takes only historical average bus coming time. Petersen and et al., [31] has presented a multi-model deep neural network prediction framework for bus arrival time using Convolutional and Long short-term memory. This paper provides link travel time prediction as in our research. As the previous work, Petersen and et.al. didn't model delay time at stops. As we said before, for Almaty city dwelling time is important. Another paper [32] predict a bus arrival time by real-time data of the bus location data. The method proposed by authors estimate link travel time by predicting road speed with BP neural network. The main limitations of this work are that they calculate buses' riding time as for normal cars, but the travelling behavior of buses is quite different from cars because buses make a lot of stops and collaborate with passengers.

The method for arrival time prediction proposed in this research has several advantages. It considers:

- Historical bus travel time information;
- Real time information;
- Estimate bus dwelling time and riding time separately;
- Take into account traffic lights;

- Considers city facilities.

All those points make proposed method more reliable.

VII. CONCLUSION

In this paper, we examine the issue of public bus arrival time with real time traffic data and historical GPS data. To settle such transport travel time expectation problem, a novel methodology dependent on real time information is introduced regarding the information learned from historical bus GPS data and information about city infrastructure. We separately estimate bus dwelling time at stations and link riding time between stations separately in our method. To predict riding time, we firstly apply clustering approach to bus line links, then for each cluster we train LSTM NN model with city infrastructure data and weather condition data. According to tests results our method show small mean absolute error for buses that not far from departure station. To deal with continuous traffic data, we need to increment the models' forecast precision dependent on transport GPS information under both normal and abnormal traffic conditions. Based on the results, our future work will be on utilizing existing taxi or different sorts of traffic information to gauge the transport travel time of recently planned or with sparse records.

VIII. ACKNOWLEDGEMENT

This work has done under the project NAP05134776 Location Analytics Techniques for Prediction of Mobility Patterns of the Ministry of Education and Sciences of the Republic of Kazakhstan

REFERENCES

- [1] S. I.-J. Chien, Y. Ding, and C. Wei, "Dynamic bus arrival time prediction with artificial neural networks," *Journal of Transportation Engineering*, vol. 128, no. 5, pp. 429-438, 2002. doi: [https://doi.org/10.1061/\(ASCE\)0733-947X\(2002\)128:5\(429\)](https://doi.org/10.1061/(ASCE)0733-947X(2002)128:5(429))
- [2] F. Cathey and D. J. Dailey, "A prescription for transit arrival/departure prediction using automatic vehicle location data," *Transportation Research Part C: Emerging Technologies*, vol. 11, no. 3-4, pp. 241-264, 2003. doi: [https://doi.org/10.1016/S0968-090X\(03\)00023-8](https://doi.org/10.1016/S0968-090X(03)00023-8)
- [3] C. I. van Hinsbergen, J. Van Lint, and H. Van Zuylen, "Bayesian committee of neural networks to predict travel times with confidence intervals," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 5, pp. 498-509, 2009. doi: <https://doi.org/10.1016/j.trc.2009.04.007>
- [4] J. Liu, X. Yu, Z. Xu, K.-K. R. Choo, L. Hong, and X. Cui, "A cloud-based taxi trace mining framework for smart city," *Software: Practice and Experience*, vol. 47, no. 8, pp. 1081-1094, 2017. doi: <https://doi.org/10.1002/spe.2435>

- [5] V. Kumar, B. A. Kumar, L. Vanajakshi, and S. C. Subramanian, "Comparison of model based and machine learning approaches for bus arrival time prediction," in *Proceedings of the 93rd Annual Meeting*, California, CA, 2014.
- [6] R. Suryanita, H. Maizir, and H. Jingga, "Prediction of structural response based on ground acceleration using artificial neural networks," *International Journal of Technology and Engineering Studies*, vol. 3, no. 2, pp. 74-83, 2017. doi: <https://doi.org/10.20469/ijtes.3.40005-2>
- [7] S. I.-J. Chien and C. M. Kuchipudi, "Dynamic travel time prediction with real-time and historic data," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 608-616, 2003. doi: [https://doi.org/10.1061/\(ASCE\)0733-947X\(2003\)129:6\(608\)](https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(608))
- [8] M. Chen, X. Liu, J. Xia, and S. I. Chien, "A dynamic bus-arrival time prediction model based on APC data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 19, no. 5, pp. 364-376, 2004. doi: <https://doi.org/10.1111/j.1467-8667.2004.00363.x>
- [9] Y. Bin, Y. Zhongzhen, and Y. Baozhen, "Bus arrival time prediction using support vector machines," *Journal of Intelligent Transportation Systems*, vol. 10, no. 4, pp. 151-158, 2006. doi: <https://doi.org/10.1080/15472450600981009>
- [10] B. Yu, Z. Z. Yang, and J. Wang, "Bus travel-time prediction based on bus speed," in *Proceedings of the Institution of Civil Engineers-Transport*, London, UK. Thomas Telford Ltd, 2010.
- [11] C. N. M. and R. J. F. Locsin, "Neural networks application for water distribution demand-driven decision support system," *Journal of Advances in Technology and Engineering Studies*, vol. 4, no. 4, pp. 162-175, 2018. doi: <https://doi.org/10.20474/jater-4.4.3>
- [12] R. Jeong and L. R. Rilett, "Prediction model of bus arrival time for real-time applications," *Transportation Research Record*, vol. 1927, no. 1, pp. 195-204, 2005. doi: <https://doi.org/10.1177/0361198105192700123>
- [13] N. I. Sapankevych and R. Sankar, "Time series prediction using support vector machines: A survey," *IEEE Computational Intelligence Magazine*, vol. 4, no. 2, pp. 24-38, 2009.
- [14] R. Padmanaban, L. Vanajakshi, and S. C. Subramanian, "Estimation of bus travel time incorporating dwell time for apts applications," in *Conference Proceedings of IEEE Intelligent Vehicles Symposium*, Florida, FL, 2009.
- [15] A. Kittelson, *Transit Capacity and Quality of Service Manual*. Washington, D.C: WA: Transportation Research Board of the National Academies, 2003.
- [16] K. J. Dueker, T. J. Kimpel, J. G. Strathman, and S. Callas, "Determinants of bus dwell time," *Journal of Public Transportation*, vol. 7, no. 1, pp. 2-10, 2004. doi: <http://doi.org/10.5038/2375-0901.7.1.2>
- [17] A. Tirachini, "Bus dwell time: the effect of different fare collection systems, bus floor level and age of passengers," *Transportmetrica A: Transport Science*, vol. 9, no. 1, pp. 28-49, 2013. doi: <https://doi.org/10.1080/18128602.2010.520277>
- [18] E. Chioni, C. Iliopoulou, C. Milioti, and K. Kepaptsoglou, "Factors affecting bus bunching at the stop level: A geographically weighted regression approach," *International Journal of Transportation Science and Technology*, vol. 9, no. 3, pp. 207-217, 2020. doi: <https://doi.org/10.1016/j.ijtst.2020.04.001>
- [19] D. Verbich, E. Diab, and A. El-Geneidy, "Have they bunched yet? An exploratory study of the impacts of bus bunching on dwell and running times," *Public Transport*, vol. 8, no. 2, pp. 225-242, 2016. doi: <https://doi.org/10.1007/s12469-016-0126-y>
- [20] N. R. Chopde and M. Nichat, "Landmark based shortest path detection by using a* and haversine formula," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 1, no. 2, pp. 298-302, 2013.
- [21] R. Sinnott, "Virtues of the haversine," *Sky and Telescope*, vol. 6, no. 9, pp. 158-170, 1984.
- [22] P. He, G. Jiang, S.-K. Lam, and Y. Sun, "Learning heterogeneous traffic patterns for travel time prediction of bus journeys," *Information Sciences*, vol. 512, pp. 1394-1406, 2020. doi: <https://doi.org/10.1016/j.ins.2019.10.073>
- [23] W. Wang, M. Baumann, and X. Jiang, *Green, Smart and Connected Transportation Systems*. New York, NY: Springer Nature, 2020.
- [24] A. Tirachini, "Estimation of travel time and the benefits of upgrading the fare payment technology in urban bus services," *Transportation Research Part C: Emerging Technologies*, vol. 30, pp. 239-256, 2013. doi: <https://doi.org/10.1016/j.trc.2011.11.007>
- [25] S. Partha, H. I. MAHMUD, Q. S. HOSSAIN, and M. Z. Islam, "Passenger Car Equivalent (PCE) of through vehicles at signalized intersections in Dhaka Metropolitan City, Bangladesh," *International Association of Traffic and Safety Sciences*,

- vol. 33, no. 2, pp. 99-104, 2009. doi: [https://doi.org/10.1016/S0386-1112\(14\)60248-X](https://doi.org/10.1016/S0386-1112(14)60248-X)
- [26] V. Gorodokin, Z. Almetova, and V. Shepelev, "Procedure for calculating on-time duration of the main cycle of a set of coordinated traffic lights," *Transportation Research Procedia*, vol. 20, pp. 231-235, 2017. doi: <https://doi.org/10.1016/j.trpro.2017.01.060>
- [27] E. Loret, G. Gullotta, M. Fea, and F. Sarti, *Traffic Fluxes and Urban Congestion: A Simple Approach with the Attractors' Method*. London, UK: Scientific Research Publishing, 2012.
- [28] Y. Duan, L. Yisheng, and F.-Y. Wang, "Travel time prediction with lstm neural network," in *IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, New Dehli, India, 2016.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, London, UK, 2016.
- [30] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929-1958, 2014.
- [31] N. C. Petersen, F. Rodrigues, and F. C. Pereira, "Multi-output deep learning for bus arrival time predictions," *Transportation Research Procedia*, vol. 41, pp. 138-145, 2019. doi: <https://doi.org/10.1016/j.trpro.2019.09.025>
- [32] J. Fei, Y. Lu, Y. Guo, and H. Zhang, "Predicting bus arrival time using BP neural network and dynamic transfer," *Procedia Computer Science*, vol. 174, pp. 95-100, 2020. doi: <https://doi.org/10.1016/j.procs.2020.06.062>