SD-seq2seq : A Deep Learning Model for Bus Bunching Prediction Based on Smart Card Data

[1,2]Zengyang Gong, [2,3]Bo Du*, [1]Zhidan Liu, [4]Wei Zeng, [2]Pascal Perez, [1]Kaishun Wu
[1]*College of Computer Science and Software Engineering, Shenzhen University, P. R. China*
[2]*SMART Infrastructure Facility, University of Wollongong, Australia*
[3]*School of Civil, Mining and Environmental Engineering, University of Wollongong, Australia*
[4]*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, P. R. China*
gongzengyang2017@email.szu.edu.cn, {bdu, pascal}@uow.edu.au, {liuzhidan, wu}@szu.edu.cn, wei.zeng@siat.ac.cn

***Abstract*—Bus bunching, a phenomenon due to the failure of headway or timetable adherence, often causes low level of public transit service with poor bus on-time performance and excessive passenger waiting time. To mitigate bus bunching, an accurate and real-time prediction method plays an important role. In this paper, we propose a supply-demand seq2seq model called SD-seq2seq to predict bus bunching using smart card data. Features from both supply and demand sides of bus service are taken into account, like bus stop type, dwelling time, passenger demand and type, and so on. Extensive experiments on multiple bus routes in real world demonstrate that our method outperforms other baseline methods. The proposed method is expected to provide useful online information of bus operation to both bus operators and passengers.**

***Index Terms*—Bus Bunching, seq2seq, Deep learning, Prediction, Smart card data, transportation**

I. Introduction

With fast growing population and limited urban living space, most megacities, like Tokyo, Singapore and Sydney, rely on public transport more and more. A reliable and efficient public transit system plays an important role in reducing traffic congestion, emission, and car dependency in urban environment. However, due to complex traffic condition, dynamic travel demand and heterogeneous bus driver behavior, it is difficult for buses to adhere to the pre-designed timetable or headway tightly under a stochastic traffic environment, which results in that two or more buses along the same bus route may arrive at the same bus stop simultaneously or closely. Such phenomenon is called bus bunching (BB), which usually increases passengers' waiting time, reduces efficiency of bus operation, and deteriorates the quality of public transit service. Therefore, it is important to investigate and understand the mechanics of bus operation and how BB happens.

It is challenging to tackle the BB problem under complex traffic condition. A significant body of research in the literature has investigated the BB problem from different perspectives, including BB identification and cause analysis, BB prediction, and corrective control strategy to prevent BB [1] [2] [3]. The majority of previous research focused on studying the statistical correlation between BB occurrence and potential influential factors (like bus dwelling time, bus trip headway, weather condition, and so on) or designing various corrective control strategies (like stop skipping, driving speed adjustment, cooperative bus operation, and so on). Limited studies have dedicated to the development of BB prediction models. To mitigate BB, an efficient and accurate prediction method plays a fundamental role in further online control strategy design. In this paper, we aim to predict bus trip headway and further BB occurrence along given bus routes in a real bus network. A supply-demand seq2seq model, named SD-seq2seq, is developed for BB prediction. To answer the question of when and where BB will happen along a specific bus route, information from both supply and demand sides of bus service is used to extract valuable features. On the supply side, bus stop ID, bus stop type (interchange stop or not), number of bus routes running through a specific bus stop, and bus dwelling time are adopted in the SD-seq2seq model; while the number of boarding and alighting passengers, and the corresponding proportions of heterogeneous passenger types (adult, senior and youth) at a specific bus stop are used on the demand side. Moreover, heterogeneous travel patterns at different time periods are often captured along the same bus route, hence temporal feature is considered as well by dividing 24 hours into four different time segments, A.M. Peak, Daytime Inter-peak, P.M. Peak, and Night Inter-peak, according to the time-varying distribution of travel demand. Those seven features aforementioned are encoded to the hidden state combined with historical headway to predict bus service headway and BB occurrence.

The main contributions of this paper are summarized as follows:

- A SD-seq2seq model is proposed to encode features from both supply and demand sides, and these features are combined with temporal feature and historical headway for prediction of bus trip headway and BB occurrence in advance.
- At the decoder side of the proposed SD-seq2seq model, long short-term memory (LSTM) and convolutional neural network (CNN) are utilized and integrated. CNN is applied to extract passenger demand features at bus stop level, which is fed together with supply features to LSTM to achieve more accurate BB prediction
- The threshold headway to define BB is dynamic rather

than fixed, which is determined based on the bus service frequency. During peak hours, the bus service frequency is usually higher with shorter headway, then we define BB with a small value of threshold; while BB is defined with a larger threshold at non-peak hours due to the lower bus service frequency and longer headway. In such way, we can identify BB more reasonably and accurately.
- The proposed SD-seq2seq model is able to predict BB occurrence at all bus stops simultaneously along a bus route. Multiple bus routes in a large-scale bus network in real world are used for numerical experiments with promising results.

The rest of this paper is organized as follows. Related literature studies are reviewed in Section II. The problem statement is presented in Section III. The development of SD-seq2seq model is elaborated in Section IV. Comprehensive numerical experiments on multiple bus routes in real world are conducted in Section V to test the proposed SD-seq2seq model. Finally, Section VI concludes this study and shows the future work.

II. Related Work

A. Bus bunching.

Relying on GPS data, regression prediction, time series model, K-nearest neighbor method, nearest-neighbor trajectory method, artificial neural network, and support vector machine have been extensively applied to investigating bus operation and BB problem [4] - [12]. However, most of those models built on GPS data usually suffered from low GPS data update frequency, which resulted in low prediction accuracy on bus arrival time, and further dwelling time at bus stops. With the increasing availability of various data sources, like General Transit Feed Specification (GTFS) data feed, Automated fare collection (AFC) data and automatic vehicle location (AVL) data, as well as rapid development of computing technique, data analytics and machine learning have been widely applied to BB research. Particularly, a diversity of data-driven approaches has been adopted to tackle the BB problem using various data sources. With aid of smart card data, [3] identified the occurrence of BB in a large-scale bus network from temporal-spatial-operational dimensions. Temporal and spatial analysis reflected the patterns and trends of BB at different time periods and locations, respectively; while in operational dimension, comparison of BB occurrence among different bus operators was conducted, which generated valuable information to assess bus operators' performance. [1] captured the stop-level headway irregularity based on public transit smart card data. However, the proposed model needed to be re-calibrated at different bus stops, which was not efficient and practical in a large-scale public transport network in reality. In the meanwhile, only limited features were considered in the proposed model, like number of passengers and travel time between stops. [2] used AVL data to identify and to model BB problem. Eleven independent variables were analyzed to identify their effects on the occurrence of BB, and schedule deviation was found the most influential factor. [13] proposed a methodological framework combining bus-following models and bus-to-bus cooperative control strategies to address the BB problem in public transport lines.

The majority of the aforementioned approaches focused on identification of BB on bus network and analysis of the influential factors causing BB. Although [1] endeavoured to predict BB occurrence with aid of smart card data, their prediction model required re-calibration at each bus stop, which caused high computational cost and difficulty of extending its application to multiple bus routes or a city-scale large network. To fill the gap, in this paper, we develop a high-performance deep learning model with capability of predicting BB occurrence at all stops along the bus route simultaneously, and the proposed SD-seq2seq model is applicable to a large-scale bus network in real life. To the best of our knowledge, it is the first work to use SD-seq2seq framework to predict BB occurrence at all bus stops simultaneously along a bus route, with capability of application to large-scale bus network.

B. Spatial-temporal modeling.

Spatial-temporal modeling has attracted a lot of attention during past decades, including transport research. [14] used sensing resources, like cell tower signal and movement status generated from passengers' phones, to infer the surrounding environmental context, and then to predict the arrival time of next bus. [15] and [16] treated buses and taxis in road network as dynamic nodes on the graph, and their work aimed to propagate messages effectively under different spatial-temporal conditions. With adoption of a non-negative matrix factorization model, [17] showed the spatial-temporal characteristics of crowd flows at train stations in Sydney based on smart card data. [18] aimed to solve the routing problem for electric vehicles considering dynamic traffic flow in the road network in both spatial and temporal dimensions.

Recent advances in deep learning techniques and GPU computing created new opportunities for spatial-temporal modeling. [19] and [21] modified the RNN model and ResNet model, with fusion of some external features like weather, to predict traffic speed and crowd flow, respectively. [20] applied multi-level attention networks to developing a spatial and temporal attention mechanism to model the dynamic spatio-temporal correlations. All the research work aforementioned indicated that transport related data, like traffic speed, traffic flow and traffic trajectory, showed obvious and recurrent patterns in spatial and temporal dimensions.

C. Trajectory data mining.

Trajectory data mining is highly related to this research. In the literature, several studies focused on applying data mining approaches to extracting useful information from various trajectory data in the city. [22] and [24] used the trajectory data of cars in city-scale to extract features that could reflect drivers' driving behavior. [23] and [25] investigated the problem of representation learning for pedestrian trajectory. Their research findings showed a strong relationship between the movement trajectory of pedestrians and the location information in the city. Different from research on trajectories generated by cars and pedestrians, limited studies have been done on applying data mining techniques to analyzing bus trajectories to explore bus operation patterns with spatial and temporal features.

III. Problem Statement

AFC system is widely used in multimodal transportation system worldwide. Usually, passengers are required to tap on and tap off their smart cards when they get on and get off public transit, like bus, metro and train. In such way, the AFC system records bus operation information and passengers' trip details in both spatial and temporal dimensions, and generates a mass of smart card data daily accordingly. With the aid of tap-on and tap-off stamps and geographic information at stop level, the spatial-temporal bus trip trajectory can be re-constructed, and the corresponding headway between successive bus trips at each bus stop along the same bus route can be calculated to predict the occurrence of BB.

In the following subsections, relevant definitions, notations, features, re-construction of bus trip trajectory, and the description of BB prediction problem are introduced.

A. Definitions and Notations

Passenger Trip: In this paper, a passenger trip record in smart card data is represented by $r =< r_{type}, b_{id}, r_{dir}, t_{on}, s_{on}, t_{off}, s_{off} >$, where r_{type} indicates passenger type (adult, senior or youth); b_{id} is the bus route number; r_{dir} indicates the bus operational direction, and generally each bus route has two operational directions; t_{on} and s_{on} represent the time stamp and bus stop this passenger get on the bus, separately; and t_{off} and s_{off} denote the time stamp and bus stop this passenger get off the bus, respectively. The whole dataset of all passenger trips is denoted as $\mathcal{R}$.

Bus Stop: A bus stop is defined as a node $s(s_{id}, s_{type}, s_{num})$ in public transport network, where s_{id} is the sequence of stops along the bus route, and usually each bus route has fixed stop sequence in operation; s_{type} indicates whether this bus stop is an interchange stop connecting other travel modes such as train; and s_{num} presents the number of different bus routes running through this bus stop. All the bus stops along a bus route is denoted as $\mathcal{S}$, and the total number of stops along this bus route is represented by $|\mathcal{S}|$. Usually, interchange bus stops and stops with multiple bus service routes have certain effects on BB occurrence, therefore, in this study, we take both features into consideration.

Bus Trip: On a specific bus route, a series of bus trips are often fulfilled by a bus fleet with multiple buses. In the meanwhile, bus timetable is usually unchanged within a certain period, therefore we can expect that a series of bus service trips on a specific bus route on different days occur following the same time series. For example, following a stable timetable of a specific bus route, if the third trip happens at 6:05am at a specific day, then the third trip on the other days within a certain period should happen around 6:05am as well. In this study, we use the trip generation time t to represent the corresponding bus service trip sequence within one day, then a complete trip with bus arrival information at each stop is denoted as $X_t = [x_t^{s_0}, x_t^{s_1}, \cdots, x_t^{s_{|S|-1}}]$, where $x_t^{s_i}$ indicates the bus arrival time at stop s_i in trip t, and t also denotes its corresponding trip generation time. The total number of trips within one day is denoted as N. In Fig.1, four bus trip trajectories are shown as illustration.

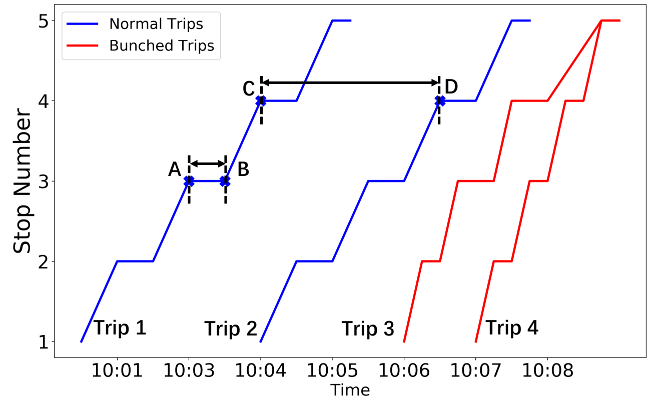

Fig. 1. Illustration of normal and bunched bus strips

Dwelling Time: Upon arrival at a bus stop, buses usually need to stop for a while to allow passengers to get on and off the bus. $D_t = [d_t^{s_0}, d_t^{s_1}, \cdots, d_t^{s_{|S|-1}}]$ presents the time duration a bus stops at each stop during the bus trip t. For illustration purpose, in Fig.1, segment $\mathbf{AB}$ represents the dwelling time at bus stop 3 in bus trip 1.

Headway: Headway can be defined as the time difference between two buses (serving two different bus trips on the same bus route) arriving at the same bus stop on a specific bus route, which is denoted by $H_t = [h_t^{s_0}, h_t^{s_1}, \cdots, h_t^{s_{|S|-1}}]$, and $h_t^{s_i} = x_t^{s_i} - x_{t-1}^{s_i}$, where $h_t^{s_i}$ represents the headway at stop s_i between bus service trips t and $t-1$.

As an instance, segment **CD** denotes headway at stop 4 between bus trip 1 and trip 2, as shown in Fig.1. In this study, historical headway is categorized into two groups, trend headway $H_{trend} = [H_{t-m*24}, H_{t-(m-1)*24}, \cdots, H_{t-1*24}]$ and closeness headway $H_{closeness} = [H_{t-n}, H_{t-(n-1)}, \cdots, H_{t-1}]$. Trend headway reflects the long-term pattern of historical headway, which consists of the headway information in trip t in the previous m days. Closeness headway shows the short-term dynamics of headway in the past n trips within one day.

Bus Bunching (BB): In this study, BB is defined as an event when two buses serving the same bus route arrive at the same bus stop with headway less than a certain threshold. For example, in Fig.1, BB happens at bus stop 5 between bus trip 3 and trip 4 (two bus trip trajectories in red color).

Travel Demand Matrix: Four travel demand matrices, $M_t^Y, M_t^A, M_t^S, M_t^P \in \mathbf{R}^{|S|\times|S|}$, are constructed to capture the number of young passenger, adult passenger, senior passenger, and the total number of all kinds of passengers, respectively in bus trip t. For example, the element $m_{t,i,j}^{P}$ in matrix M_t^P indicates the total number of passengers boarding at bus stop s_i and alighting at bus stop s_j in bus trip t.

B. Re-construction of Bus Trip Trajectory

With aid of GTFS data and geographic information from Google map, we can easily obtain the sequence of stops along a bus route, stop ID, number of bus routes running through the stop, and stop type (whether the stop is an interchange stop connecting other travel modes such as train). Although AVL data is not available in this study, bus arrival time and dwelling time based on passengers' boarding and alighting information could be estimated. By checking the first and the last tap-on and tap-off time stamps at each bus stop, we can re-construct bus service trips with bus arrival information and the corresponding dwelling duration at each bus stop. Sometimes there is no passenger boarding or alighting at some bus stops, hence the dwelling time at these stops is 0, and the relevant bus arrival information is missing. To overcome this obstacle, Algorithm 1 is proposed to fill the missed information with historical information, and then re-construct bus trip trajectory with bus arrival, dwelling and headway information.

C. Problem Definition

In order to predict BB occurrence, multiple steps are included from extracting useful information from raw smart card data to final prediction of headway and BB occurrence. Firstly, various useful features from both supply and demand sides need to be extracted from the smart card data, and those features have been introduced in Section I Paragraph 2. Next, using Algorithm 1, bus trip trajectory $\{X_i | i = 1, \cdots, t-1\}$ for a specific bus route is re-constructed given a set of smart card data $\mathcal{R}$. The corresponding bus dwelling time $D_t = [d_t^{s_0}, d_t^{s_1}, \cdots, d_t^{s_{|S|-1}}]$ and historical headway $\{H_i | i = 1, \cdots, t-1\}$ can be calculated. Finally, a SD-seq2seq model is developed to forecast the headway $\{H_j | j = t, \cdots, t+z\}$ in upcoming time periods at all bus stops, where z is the number of time intervals ahead to be predicted. Then BB occurrence can be predicted based on the threshold headway defined.

IV. SD-seq2seq System Architecture

In this section, we present the system architecture of the proposed SD-seq2seq model and details of its each component, as shown in Fig.2.

A. Overview

As illustrated in Fig.2, the system architecture of the proposed SD-Seq2seq model consists of three major components: Supply Learning, Demand Learning and Decoder. Supply Learning and Demand Learning compose the Encoder part of the SD-seq2seq model. The function of Encoder is to encode features from both supply and demand sides, and then the encoded hidden state is used to initialize Decoder. Along with the historical headway and temporal feature, the initialized Decoder is applied to predicting headway in the upcoming time periods.

Algorithm 1 Re-construction of Bus Trip Trajectory

Input: Smart card dataset $\mathcal{R}$, Bus stop sequence $\mathcal{S}$
Parameter: Optional list of parameters
Output: Dwelling Time D, Headway H

for $t \leftarrow 1$ to N **do**
  for $i \leftarrow 0$ to $|S|-1$ **do**
    Create set $\mathcal{R}_t^{s_i}$ for passengers taking trip X_t and tapping on/off at stop s_i, where $\mathcal{R}_t^{s_i} \in \mathcal{R}$.
    if $\mathcal{R}_t^{s_i} \neq \emptyset$ **then**
      Identify the first tap-on/off passenger $r_t^{'} \in \mathcal{R}_t^{s_i}$.
      Identify the last tap-on/off passenger $r_t^{''} \in \mathcal{R}_t^{s_i}$.
      $d_t{}^{s_i} = r_t^{''} \to t_{on} - r_t^{'} \to t_{off}$
      $D_t + = d_t{}^{s_i}$
      $h_t{}^{s_i} = r_t^{'} \to t_{off} - r_{(t-1)}^{'} \to t_{off}$
      $H_t + = h_t{}^{s_i}$
    else
      $d_t{}^{s_i} = 0$
      $D_t + = d_t{}^{s_i}$
      $h_t{}^{s_i} = Avg(h_t^{'}{}^{s_i} \in H_{Trend})$
      $H_t + = h_t{}^{s_i}$
    end if
  end for
end for
return D, H

The encoder-decoder architecture model has been successful applied in many fields such as natural language processing. In recent years, this kind of model has been widely used in learning and predicting time series patterns in urban transportation system, like traffic speed and traffic flow. The headway of a specific bus route is kind of sequential data with the length as the number of bus stops along this bus route. The encoder part of the designed SD-seq2seq model captures the time series patterns like historical headway and passenger travel demand; while the outputs of the decoder part include the predicted headway between the next bus trip and the current one at each bus stop.

B. Demand Learning

In the component of Demand Learning, we consider the number of passengers boarding and alighting at each bus stop along a given bus route, and the corresponding passenger types (adult, senior and youth) as features from the perspective of passenger demand.

1) Convolution: The CNN has been utilized in a diversity of research fields such as image data analysis. It has shown satisfactory performance on capturing spatial features from different grids of tensor. Passengers travel from stop to stop in a city, and the dwelling time can vary from stop to stop since the amounts of passengers boarding and alighting at different bus stops can be significantly different. Besides the number of passengers, the type of passengers presents certain influence on dwelling time as well. For example, at the stops with a large number of senior passengers boarding and alighting, the

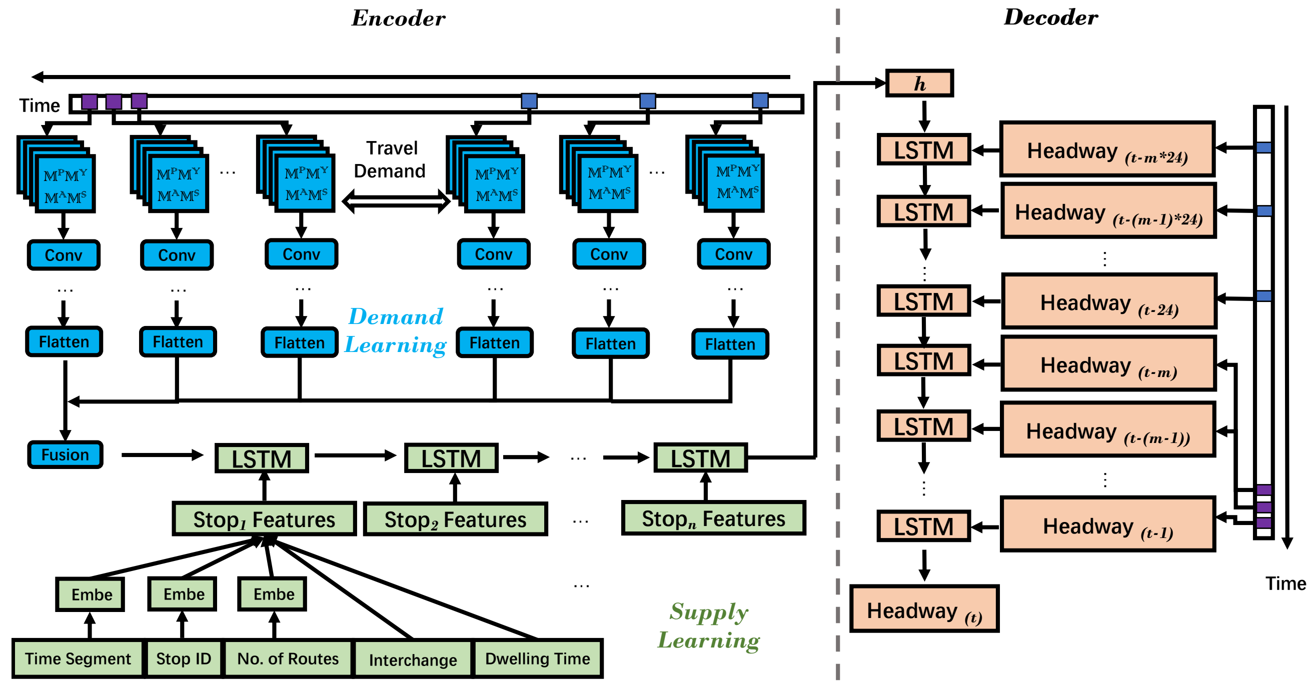

Fig. 2. System architecture of the SD-Seq2seq model

corresponding dwelling time is usually longer. In consequence, the probability of BB occurrence at this stop increases. To incorporate such passenger demand information, a convolution layer is designed at each time stamp to capture features of travel demand distribution and passenger type in spatial-temporal dimensions.

As defined in Section III A, for each bus trip t, we can get the corresponding set of travel demand matrices $[M_t^P, M_t^Y, M_t^A, M_t^S]$. We first concatenate them along with the first axis as one tensor $M_t^{(0)} \in \mathcal{R}^{4\times|\mathcal{S}|\times|\mathcal{S}|}$, which is followed by a convolution, a pooling and a flatten, respectively:

$$M_t^{(1)} = f(W_c^{(1)} * M_t^{(0)} + b_c^{(1)}) \tag{1}$$

$$M_t^{(2)} = MaxPooling(M_t^{(1)}) \tag{2}$$

$$M_t^{(3)} = Flatten(M_t^{(2)}) \tag{3}$$

where $*$ denotes the convolution process; f is an activation like $ReLu$; $W_c^{(1)}$ and $b_c^{(1)}$ are the learnable parameters in the first layer.

2) Fusion: The amount of passengers usually varies periodically, with higher travel demand during peak hours while lower demand at non-peak hours. Such periodical pattern of travel demand is more significant at stops located in city central area. Taking a bus stop at Bondi Junction in Sydney as example, Fig.3(a) depicts the distribution of trip generation from Monday to Sunday, which shows recurrent patterns.

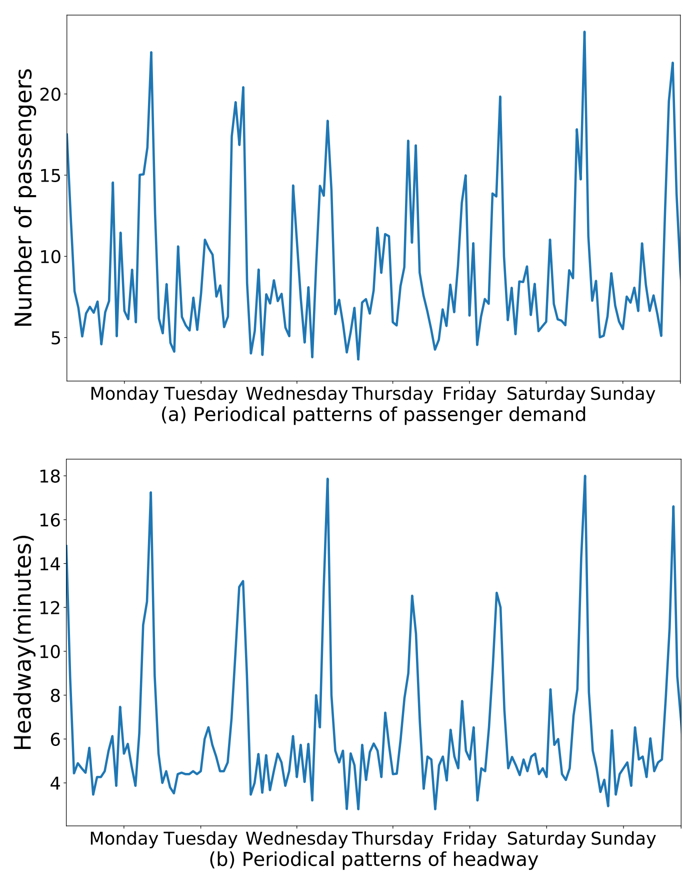

Fig. 3. Periodical patterns of passenger demand and headway

To extract the periodical pattern of passenger demand at bus stop level, we fuse the convolution results generated from travel demand matrices.

$$M_{fcn} = tanh(\sum_{x=1}^{m} M_x^{(3)} \circ W_x + \sum_{y=1}^{n} M_y^{(3)} \circ W_y) \quad (4)$$

where $\circ$ is the element-wise multiplication; W_x and W_y are the learnable parameters; and m and n represent Demand Learning fused with trend travel demand feature from the previous m days and closeness travel demand feature from the last n trips, respectively.

C. Supply Learning

The component of Supply Learning aims to extract features of bus stops along a given bus route, including fixed features like bus stop ID, number of bus routes running through the stop, and stop type (whether the stop is an interchange stop connecting other travel modes such as train), and dwelling time as dynamic feature in both spatial and temporal dimensions. To encode such features from the perspective of public transit service supplier, we adopt LSTM to extract the above-mentioned supply features at each bus stop along a given bus route, and the LSTM is initiated by M_{fcn} from Demand Learning in Section IV B.

$$f_t = \sigma_g(W_f H_t + U_f h_{t-1} + b_f) \quad (5)$$

$$i_t = \sigma_g(W_i H_t + U_i h_{t-1} + b_i) \quad (6)$$

$$o_t = \sigma_g(W_o H_t + U_o h_{t-1} + b_o) \quad (7)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_g(W_c H_t + U_c h_{t-1} + b_c) \quad (8)$$

$$h_t = o_t \circ \sigma_h(c_t) \quad (9)$$

D. Decoder

In the Decoder component, historical headway is used as input. Fig.3(b) shows weekly historical headway at a bus stop at Bondi Junction, which reflects time-varying but repeated headway pattern. In this study, LSTM is applied to extracting features from historical headway information, which consists of both trend headway H_{trend} and closeness headway $H_{closeness}$. In each bus trip t, LSTM derives the vector representation of hidden state h_t as follows:

where H_t is the input headway in each bus trip t; f_t, i_t and o_t are forget gate, update gate and output gate, respectively; W, U and b are all learnable parameters; σ_g is sigmod function and σ_h is tanh function; h_t represents the hidden state in each trip; and in the first trip, h_0 is initialized by the output h_e generated from Encoder, which consists of Demand Learning and Supply Learning components in Section IV B and C.

V. Performance Evaluation

In this section, multiple numerical experiments are conducted to evaluate the proposed SD-seq2seq model using multiple bus routes in the public transit network in Sydney.

A. Experiment Setup

Dataset. Two-month smart card data in Sydney is used in this study for various experiments, which includes more than 10,000 bus routes (including route variants). Considering people's travel pattern and passengers' demand on weekend and holiday are quite different from those during working days, and BB happens much more during working days, we only use the data between Monday and Friday. Moreover, data during public holidays are removed as well in the following experiments.

To include temporal feature, we split one day (24 hours) into four time segments, namely A.M. Peak from 7:00 am to 9:00 am, Daytime Inter-peak from 9:00 am to 3:00 pm, P.M. Peak from 3:00 pm to 6:00 pm, and Night Inter-peak from 6:00 pm to 7:00 am the next day, based on the distribution of trip generation, as shown in Fig.4. Fig.4 shows the distribution of trip generation based on the average information of two-month smart card data covering the whole bus network in Sydney.

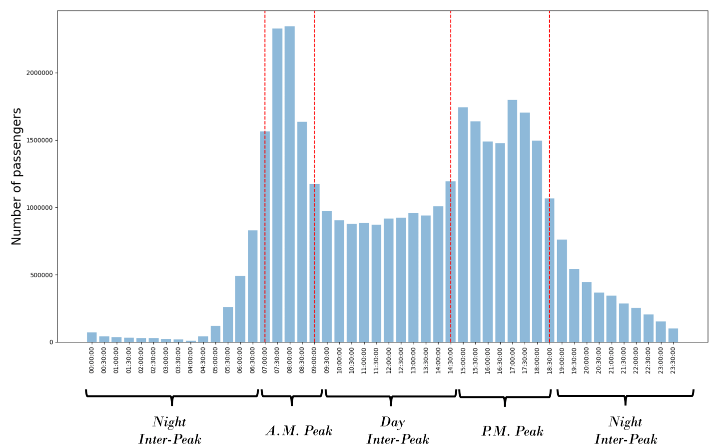

Fig. 4. Distribution of trip generation at different time segments

In this study, we use time-varying BB threshold rather than fixed considering dynamic travel demand and heterogeneous bus service frequencies at peak and non-peak hours. For illustration purpose, two BB threshold values are used in the following experiments for peak and non-peak periods, respectively. During peak hours (A.M. Peak and P.M. Peak), the threshold headway to identify BB occurrence is set as a quarter of the headway, while half of the headway is used during non-peak hours (Daytime Inter-peak and Night Inter-peak).

Bus routes. A series of numerical experiments have been done on different bus routes. For illustration purpose, two bus routes covering different areas in Sydney are chosen to demonstrate the results. To select the representative bus routes, BB occurrence at all the bus stops covering the whole bus network in Sydney has been visualized, and two bus routes (400 and 380) with observed BB issue but covering different areas have been chosen, as shown in Fig.5. In Fig. 5, each bar represents a bus stop, and the bar's color and height indicate the average daily number of BB occurrence.

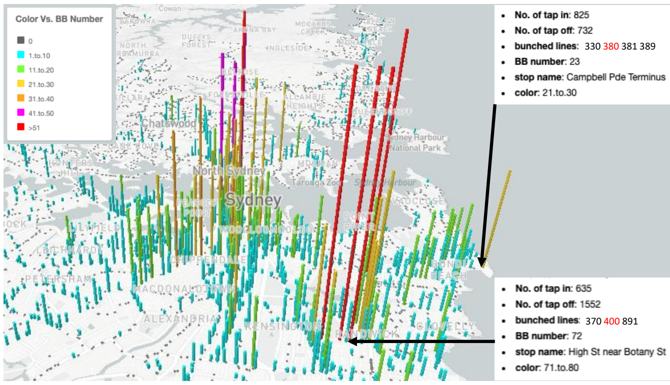

Fig. 5. BB identification with stop-level information

Two representative bus routes, 400 and 380, covering different areas of Sydney, are utilized for a series of numerical experiments in the following subsections.

- **Bus Route 400.** This bus route starts at a busy interchange bus stop Bondi Junction near the Centennial Park in Sydney CBD, and its last stop is the Sydney Airport. Within two months, around 79,000 passenger trips and 9,000 bus service trips have been identified along this bus route.
- **Bus Route 380.** Bus route 380 starts at the Watsons Bay in the east part of Sydney, running through the famous Bondi Beach, and then ends at Bondi Junction. Within two months, around 74,000 passenger trips and 6,500 bus service trips have been observed along this bus route.

Performance metrics. Accuracy and the root mean square error (RMSE) are used as performance metrics in the numerical section. Since BB prediction is modeled as a binary classification problem in this study, both precision and recall are evaluated.

B. Comparison with other methods

Methods. To validate the efficiency and accuracy of the proposed SD-seq2seq model, we compare the proposed method with the other four popular methods as follows:

1) Seq2seq [26]. Similar as the proposed SD-seq2seq model, this model also considers both supply and demand features, but it flattens and concatenates all features as input to the encoder of seq2seq model.
2) GRU [27] is a kind of variation of LSTM. To apply GRU approach, closeness and trend headways are used as input, and the output is the predicted headway.
3) DNN [28] is a basic model in deep learning. To apply DNN, we flatten the closeness and trend headways to a vector as an input to the model, and the output is the predicted headway.
4) ARIMA [29] is a well-known and classic model for forecasting in time series. To apply ARIMA, at each bus stop, we build an ARIMA model. The input is the historical headway at a stop, and output is the predicted headway.

Implementation. In the following experiments, we implement the SD-seq2seq model and other baselines in Python and TensorFlow. Our network has been trained with the following hyper-parameters setting: mini-batch size (128), learning rate (0.01) with adam optimizer. For each bus route, we randomly choose 80% trips at each time segment to make up a training set to train the model, and the remaining 20% trips are used to test the performance of the model at corresponding time segment. All experiments have been conducted on a server with one Tesla K40m and Intel Xeon E5. Each experimental setting has been repeated ten times and the average results are reported.

To check the performance of the proposed SD-seq2seq model, the results of comparison with other popular models are shown in Fig.6. In general, our model outperformes other approaches, no matter on bus route 400 or 380, peak hours or non-peak hours. On the contrary, ARIMA shows the worst performance according to its RMSE. In the meanwhile, the RMSE results on bus 380 performs slightly better than that on bus 400.

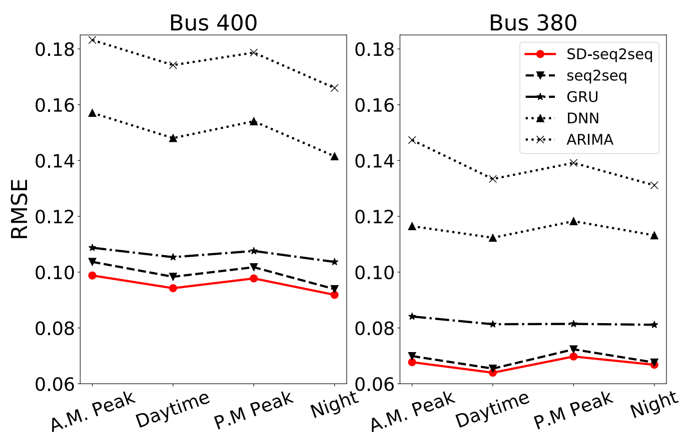

Fig. 6. Comparison of numerical results among different methods

As shown in Fig.5 in the previous subsection, much more BB has been identified on bus route 400 than that on bus route 380. It happens because bus route 400 is operating at city central area with much higher population density, and corresponding higher travel demand and more complex traffic condition. On the contrary, bus route 380 is operating in a relatively lower-density region with lower travel demand and better traffic condition, which makes the bus operation under a more stable condition with less BB occurrence. Consequently, the prediction of BB occurrence on bus route 380 shows better performance due to its relatively more stable traffic condition and lower travel demand. In the meanwhile, no matter in which area or with which method, we can expect much higher traffic volume, travel demand and more uncertain traffic environment during peak hours than non-peak hours. Therefore, better prediction results at non-peak hours than peak hours could be expected, which is evident from the comparison results shown in Fig.6 that the RMSE value at peak hours is higher than that at inter-peak hours for all methods and both bus routes.

C. Accuracy analysis

Since BB prediction is a binary classification problem, besides RMSE, other important metrics such as accuracy, precision and recall are also evaluated for performance measurement. Due to the space limitation, in this subsection we only compare the proposed SD-seq2seq model and the classic seq2seq model (with the second best performance in the previous comparison).

TABLE I shows the results generated from the proposed SD-seq2seq model on two different bus routes at different time segments. Generally speaking, the SD-seq2seq model performs better on bus route 380 than on 400, which indicates that predicting BB at crowded areas with higher travel demand is more difficult than at suburb areas. In the meanwhile, the prediction accuracy fluctuates at different time segments, which is due to the complexity of traffic condition and uncertainty of travel demand. During peak hours, traffic condition is more complex, hence we get lower prediction accuracy, while higher accuracy can be expected during non-peak hours. The prediction results on bus route 380 at Night Inter-Peak (no traffic congestion and low travel demand) shows highest accuracy around 92.56%, and the worst case is on bus route 400 during A.M. Peak (severe traffic congestion and high travel demand), which is 84.31%.

TABLE I
COMPARISON AMONG DIFFERENT TIME SEGMENTS AND BUS ROUTES USING SD-SEQ2SEQ MODEL

Bus Route	Time Segment	**Accuracy**	**Precision**	**Recall**
400	A.M. Peak	84.31%	76.86%	82.48%
	Daytime Inter-Peak	87.01%	82.34%	87.62%
	P.M. Peak	86.83%	81.11%	85.96%
	Night Inter-Peak	89.74%	81.83%	89.32%
380	A.M. Peak	88.67%	77.73%	86.42%
	Daytime Inter-Peak	90.65%	83.99%	89.03%
	P.M. Peak	89.35%	81.74%	87.64%
	Night Inter-Peak	92.56%	87.14%	91.83%

TABLE II
COMPARISON AMONG DIFFERENT TIME SEGMENTS AND BUS ROUTES USING SEQ2SEQ MODEL

Bus Route	Time Segment	**Accuracy**	**Precision**	**Recall**
400	A.M. Peak	80.04%	71.22%	79.93%
	Daytime Inter-Peak	84.80%	77.09%	85.12%
	P.M. Peak	81.91%	75.17%	83.14%
	Night Inter-Peak	83.75%	76.68%	84.24%
380	A.M. Peak	85.94%	73.72%	83.38%
	Daytime Inter-Peak	87.71%	78.61%	86.36%
	P.M. Peak	86.53%	75.18%	84.11%
	Night Inter-Peak	89.70%	80.83%	87.41%

TABLE II shows the results generated from the baseline seq2seq model on the same bus routes with same time segment division. By comparing the results between TABLE I and TABLE II, we can find that the proposed SD-seq2seq model performs better on both bus routes, and during all different time segments.

D. Feature analysis

To figure out how different features influence the prediction results, and to what extent, we conduct various experiments with inclusion of different features. In Fig.7, “All Features” represents the complete SD-seq2seq model including all features such as historical headway, temporal feature, supply and demand features; “Demand Only” represents a modified SD-seq2seq model without considering supply features; “Supply Only” denotes the one removing demand features; and “Headway Only” means only historical headway is used in the model.

According to the comparison results in Fig.7, demand features such as passenger demand and proportions of different passenger types are more influential than supply features in BB prediction. As for the model with historical headway only, it is difficult to reduce RMSE. Therefore, it is more reasonable and accurate to include features from both supply and demand sides into the prediction model.

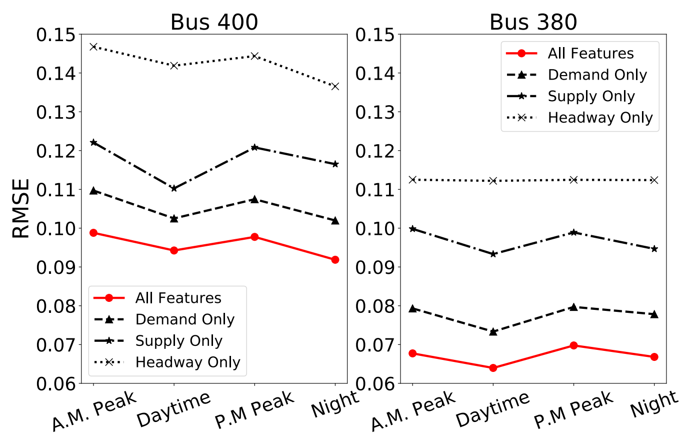

Fig. 7. Comparison of results considering different features

VI. Conclusion and Future Work

In this paper, we proposed a SD-Seq2seq model to predict bus trip headway and corresponding BB occurrence based on smart card data. The SD-seq2seq2 modelling framework was designed to take the advantages of CNN and LSTM to extract features from supply and demand sides, respectively. Such supply and demand features were further combined with historical headway, including both trend headway and closeness headway, to predict bus trip headway and BB occurrence in the upcoming time periods. To define BB in a more reasonable and accurate way, time-varying headway threshold was adopted to detect BB occurrence. Two-month smart card data in Sydney was used to train and test the proposed SD-seq2seq model. Experiment results showed that the SD-seq2seq model could achieve more than 85% prediction accuracy on different bus routes and during various time segments in most cases. Moreover, the proposed SD-seq2seq model outperformed other baseline methods in various comparisons. The proposed SD-seq2seq model is able to predict BB occurrence at all bus stops simultaneously along a bus route, which is applicable to a large-scale bus network in real life.

As future work, data fusion among multiple data sources such as smart card data, GTFS data and AVL data will be taken into account to improve the prediction accuracy. Based on the BB prediction outputs, corresponding dynamic and adaptive control strategies will be investigated and tested in bus operation.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Transport for NSW, Australia for data access and support. The first author is thankful to the SMART Infrastructure Facility in the University of Wollongong, Australia for providing internship opportunity, and the College of Computer Science and Software Engineering in Shenzhen University, China for providing international visiting funding support.

REFERENCES

[1] Dongwei Chen, Haiyang Yu and Zhihai Wu. Headway-based bus bunching prediction using transit smart card data, *Transportation Research Part C*, 72:45–59, November 2016.

[2] Prakash Ranjitkar, Soroush Rashidi and Orosz Csaba. Using automatic vehicle location data to model and identify determinations of bus bunching, *Transportation Research Procedia*, 25:1444–1456, November 2017

[3] Bo Du and Paul-Antonin Dublanche. Bus bunching identification using smart card data, In *Proceedings of the IEEE 24th International Conference on Parallel and Distributed Systems*, pages 1087–1092, Singapore, December 2018.

[4] Steven I-Jy Chien, Yuqing Ding and Chienhung Wei. Dynamic Bus Arrival Time Prediction with Artificial Neural Networks, *Journal of Transportation Engineering*, vol.128 no. 5, September 2002.

[5] J. Patnaik, S. Chien and A. Bladikas. Estimation of bus arrival times using APC data, *Journal of Public Transportation*, March 2004.

[6] R. Jeong, The prediction of bus arrival time using automatic vehicle location systems data, *Dissertation Abstracts International*, vol. 65–12, Section: B, page: 6530, 2004

[7] Dalia Tiesyte and Christian Jensen. Similarity-based prediction of travel times for vehicles traveling on known routes, In *Proceedings of the 16th ACM Sigspatial international conference on Advances in geographic information systems*, pages 1–10, Irvine California, November 2008

[8] Hyunho Chang, Dongjoo Park and Seungjae Lee. Dynamic multi-interval bus travel time prediction using bus transit data, *Transportmetrica A: Transport Science*, vol. 6, pages 1–10, 2010

[9] R.P.S. Padmanaban, K. Divakar and L. Vanajakshi. Development of a real-time bus arrival prediction system for Indian traffic conditions, *IET Intelligent Transport Systems*, vol. 4 Issue. 3, pages 189-200, September 2010

[10] C. Coffey, A. Pozdnoukhov and F. Calabrese. Time of arrival predictability horizons for public bus routes, *ACM Sigspatial International Workshop on Computational Transportation Science*, pages 1-5, 2011

[11] Mathieu Sinn, Ji Won Yoon and Francesco Calabrese. Predicting arrival times of buses using real-time GPS measurements, *15th International IEEE Conference on Intelligent Transportation Systems*,pages 1227–1232, 2012

[12] Yongjie Lin, Xianfeng Yang, and Lei Jia. Real-Time Bus Arrival Time Prediction: Case Study for Jinan, China. *Journal of Transportation Engineering*, vol. 139 Issue. 11, pages 1133-1140, November 2013

[13] Konstantinos Ampountolas, Malcolm Kring. Mitigating Bunching with Bus-following Models and Bus-to-bus Cooperation, *IEEE Transactions on Intelligent Transportation Systems*, February 2020

[14] Yuanqing Zheng, Pengfei Zhou and Mo Li. How long to wait? predicting bus arrival time with mobile phone based participatory sensing, *IEEE Transactions On Mobile Computing*, 13:1228–1241, June 2014.

[15] Jianping Pan, Lei Zhang, Boyang Yu. GeoMob: A mobility-aware geocast scheme in metropolitans via taxicabs and buses, In *IEEE Conference on Computer Communications*, pages 1779–1787, July 2014.

[16] Lin Cai, Jianping He and Jianping Pan. Delay minimization for data dissemination in large-scale vanets with buses and taxis, *IEEE Transactions on Mobile Computing*, 15:1939–1950, August 2016.

[17] Zhibin Li,Yongshun Gong, Yu Zheng. Network-wide crowd flow prediction of sydney trains via customized online non-negative matrix factorization, In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1243–1252, New York, United States, October 2018.

[18] Ankur Sarker, Haiying Shen, John A.Stankovic. MORP: Data-Driven Multi-Objective Route Planning and Optimization for Electric Vehicles, In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, January 2018.

[19] Yu Zheng Junbao Zhang and Dekang Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction, In *Proceedings of the 31th AAAI Conference on Artificial Intelligence*, pages 1655–1661, February 2017.

[20] Songyu Ke, Yuxuan Liang and Yu Zheng. Geoman: Multi-level attention networks for geo-sensory time series prediction. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 3248–3434, July 2018.

[21] Jiajie Xu, Zhongjian Lv and Hongzhi Yin. Lc-rnn: A deep learning model for traffic speed predictions. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 3470–3476, July 2018.

[22] Weishan Dong, Ting Yuan and Changsheng Li. Autoencoder Regularized Network For Driving Style Representation Learning. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 1603–1609, Melbourne, Australia, August 2017.

[23] Qiang Gao, Fan Zhou and Kunpeng Zhang. Identifying Human Mobility via Trajectory Embeddings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*,Melbourne, Australia, August 2017.

[24] Tung Kieu, Bin Yang and Chenjuan Guo. Distinguishing Trajectories from Different Drivers using Incompletely Labeled Trajectories. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages:863-872, Torino Italy, October 2018

[25] Ka-Ho Chow, Anish Hiranandani and Yifeng Zhang. Representation Learning of Pedestrian Trajectories Using Actor-Critic Sequence-to-Sequence Autoencoder. *arXiv preprint arXiv:1811.08069*, 2018

[26] Quov V.Le llya Sutskever, Oriol Vinyals. Sequence to sequence learning with neural networks, In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, pages 3104–3112, December 2014

[27] Dzmitry Bahdanau, Kyunghyun Cho and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation, In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1724–1734, October 2014.

[28] Yoshua Bengio, Yann LeCun and Geoffrey Hinton. Deep learning, *Nature*, 521:436–444, May 2015.

[29] G.E.P. Box and A.P. David. Distribution of residual autocorrelations in autoregressive integrated moving average time series models, *Journal of the American Statistical Association*, 65:1509–1526, December 1970.