Journal of Theoretical and Applied Information Technology
15th October 2020. Vol.98. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

OPTIMISATION OF A PUBLIC BUS SCHEDULE USING A NEURAL NETWORK AND K-MEANS

YASUKI SHIMA[1], RABIAH ABDUL KADIR[2], ALI FATHELALEM[3], RIZA SULAIMAN[4]
[1,2,4]Institute of Visual Informatics, UKM, Bangi, Selangor, Malaysia
[3]Maio University, Nago-city, Okinawa, Japan
E-mail: [1]sssyasuki@gmail.com, [2]rabiahivi@ukm.edu.my, [3] ali@meio-u.ac.jp, [4] riza@ukm.edu.my

ABSTRACT

Public transport bus services are expected to evolve with the growth of a city. In urban areas in both developed and less developed countries, the demand for public transport is gradually increasing, and it is becoming necessary to increase the capacity of these services. In the case of public buses, operators tend to increase the number of vehicles and to extend bus routes to meet the requirements of passengers. However, waiting times for passengers are still critical, and passengers may spend longer waiting for public transport services. In this research work, we use a neural network and K-means to find the optimal scheduling of bus trips per day, and to minimise both the number of buses per day (i.e. the cost) and the waiting time throughout the day. A GPS-based dataset obtained from a public bus operator in Okinawa, Japan, is used. Three elements of the dataset (the number of buses, number of stops and dwell time per time zone) are extracted to identify the peak and off-peak times in each day's bus service using K-means. These three elements are also used as input to the ANN. As the output of K-means clustering, a moderate dwell time is used as a supervised dataset. A backpropagation neural network algorithm is used to optimise the bus schedule and to allocate vehicles per time zone in a way that minimises the operational cost and maintains reliability for passengers. Our research establishes a framework and methods for optimising operating costs while meeting the passenger demand for reliable services and minimal waiting times.

Keywords: *Public Bus, Bus Scheduling, K-Means, Neural Network.*

1. INTRODUCTION

The planning process for public transportation services can be divided into four activities: network route design, timetable development, vehicle scheduling, and crew scheduling. The first of these, network route design, focuses on the design and evaluation of the public transport network, while in the second phase, timetable development, the timetable is designed when the network route design activity is complete. The vehicle scheduling phase is carried out after the timetable is developed, and crew scheduling is then finally used to determine the number of crew members required [1].

In the case of higher demand for public bus services, operators tend to increase the number of vehicles and to extend their bus routes to meet the requirements of passengers. However, waiting times for passengers are still critical, and passengers may spend longer waiting for public transport services.

This situation is strongly affected by the scheduling of public transport [2]. The bus schedule is usually provided to passengers by the public bus operator in the form of a predetermined timetable. However, this fixed timetable often does not meet passenger demand. There is therefore a need for a method of improving bus scheduling that can provide a reliable, steady service at a reasonable operational cost. A good solution is expected to consider different working conditions, such as peak time and off-peak time demand conditions. An optimal solution allows a balance to be maintained between supply and demand [3][4].

Often, in order to strike a balance between operating costs and a reasonable response to passenger demand, public bus operations are supported by the government, to ensure a suitable service to the community. Passengers require good services, such as lower waiting times and more frequent bus trips [5], while the government and the bus operators seek to reduce the operating costs. However, controlling the operational cost while maintaining an improved bus transportation service is very challenging. Several authors have suggested

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

various approaches and algorithms to address the problem of optimising both the costs and public bus service, such as Wagale et al. [3] and Oort et al [6]. Wagale et al. attempted to calculate the operation costs based on actual vehicle location data provided by the bus operator, and to optimise the schedule based on a balance between operation cost and profit.

In their 2015 work, Oort et al. [6] concluded that the key elements affecting the provision of better and more efficient public transport were shorter and more reliable trip times. They found that by removing operational bottlenecks, costs could be reduced while quality was increased, thereby increasing ridership and revenues.

1.1. Research Motivation

The provision of a bus service with a high level of reliability for passengers at a reasonable cost to bus operators represents a significant challenge, despite the fact that several research and optimisation algorithms are available, and the public bus system for the city of the first author in Japan was no exception. Delays, interruptions to bus schedules and struggles by bus operators to offer a reliable service were well known in the public domain.

In a review of research on the analysis and optimisation of bus transport systems, we found widespread and diverse efforts that included work on methodologies, analysis, and optimisation algorithms. Moreover, using modern technologies and tools that are becoming affordable and available, data on timetables and real bus operations can easily be obtained by bus operators. Recent technologies such as automatic vehicle location (AVL) and automatic passenger counting (APC) systems that can automatically detect the locations of buses and tally the numbers of passengers, in conjunction with IoT technology, have made it possible to collect a variety of datasets for a range of events and environments [7][8].

Actual operational evidence can easily be collected using GPS and IoT technologies. In this paper, we aim to use actual data as evidence of bus operations. The following situations can be derived, and these form our motivation for this research.

1) There is high demand and a real need for better practical solutions that can be developed easily and applied by operators, which would give improvements in the service in terms of responding to passenger demand and ensuring comfort and reliability.
2) Vital data related to real bus operations can easily be obtained and can be used to assess the current service and develop a better alternative.

1.2. Problem Statement

Based on the analysis in previous works [9], passenger comfort can be directly connected to the reliability of the service, which can in turn be related to reductions in passenger waiting times and reliable estimates of trip times.

This research aims to optimise both the reliability of the service offered to passengers and the operational cost of public transportation. Figure 1 illustrates the research aim of this study.

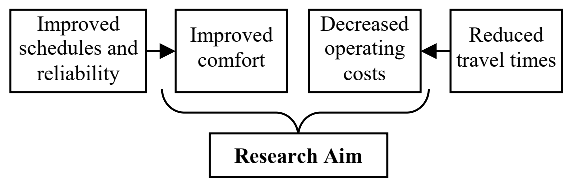


Figure 1: Aim Of The Research: Enhancing Passenger Comfort While Reducing Operational Cost

In the most general case, the objective of this optimisation is to minimise the total travel time, while a constraint is imposed in terms of a maximum fleet size, which is the main factor limiting increases in the frequency of trips. This is desirable from the viewpoint of the passengers, since it reduces the waiting time [10], which is essential for passenger comfort and satisfaction. One challenging task is that the passenger demand for service improvement and reductions in the operating costs of public buses are not directly proportional issues.

Global optimisation methods are suitable for optimising complex situations such as public transportation, and clustering is a popular method in this field. Of the existing clustering methods, K-means is particularly suitable, as it is a simple and efficient classification method even for a large amount of data [11].

Neural networks are widely used in learning and optimisation problems. Pekel et al. (2017) noted that neural networks are most commonly used in learning, while Ibarra-Rojas et al. (2015) stated that GAs were widely used in general optimisation problems [12][13]

Pekel and Kara (2017) presented a comprehensive review of the applications of ANNs in transportation [12], and found more than 70 papers on this topic. The authors showed that 69% of the papers in this area focused on forecasting and prediction, including forecasting of bus travel/arrival times, routing, passenger flow and waiting times. A further 25% focused on evaluation and analysis.

In our own review of previous work on the use of ANNs in public transportation, we found that ANNs are rarely used in vehicle scheduling and timetable optimisation problems.

In this work, we present a carefully designed learning criterion for the neural network, which is based on K-means clustering of the bus operation dataset. Prior to clustering, we extract three elements from the operation dataset.

1.3. Research Objectives

This study has the following objectives:

1) To propose a framework for the optimisation of public transport scheduling using K-means clustering and an ANN.
2) To optimise the bus schedule to reduce the operating costs while maintaining a reliable service and passenger comfort.
3) To evaluate the effectiveness of the proposed framework in optimising public transport scheduling.

In our experiments, we used a GPS log dataset generated from actual bus trips. From this dataset, the following elements were extracted:

1. The bus departure time zones (*DTZ*);
2. The number of stops at each bus stop per time zone (NS_t);
3. The dwell time per time zone (DT_t).

The results for the optimised bus schedule were then assessed based on the cost (i.e. the number of bus vehicles used per day) and the reliability of the service.

2. LITTERATURE REVIEW

In general, parameters such as the number of passengers and average travel time are important in public transport planning. To determine the frequency of a bus service using these accumulated data, it is necessary to infer the demand for passengers using an optimisation algorithm, as described in the previous section. For example, using an optimisation method, Hadas and Shnaiderman attempted to minimise the total cost based on empty seats rather than the served demand [14]. These authors recommended the use of the Global Positioning System (GPS), an APC system, and an AVL system as tools for providing valuable datasets for the optimisation of public transportation and to define probability distributions for trip times and passenger demand. Using this approach, they developed an analytical optimisation approach that determined the frequencies of bus trips and the vehicle sizes. The implementation of this approach in example cases showed that the most significant cost reduction was obtained in cases with low levels of service.

Li et al. also considered stochastic parameters such as passenger demand, arrival times, boarding and alighting times and trip times [15]. These authors defined a stochastic optimisation approach to find the optimal frequency that both maximised the profit to the bus company and minimised the passenger waiting time cost. Improvements in the overall quality of various aspects of public transport services mean that these services become more convenient for the passenger. The main aim of public transportation services is therefore to meet passenger expectations by continuously working to meet passenger demand [16], which usually involves a reliable timetable and minimal waiting times [9].

Frequency setting problems such as bus scheduling can be solved or improved using an optimisation model. Shrivastava and Dhingra developed a nonlinear integer formulation for determining the frequency of feeder lines connecting trunk lines, in order to minimise both the transfer times at the connecting station and the operational cost [17]. They developed a GA that took into account the load factor as a measure for optimising the quality of public transportation. Their proposed algorithm involved a trade-off between the load factor and the transfer dwell time. Similar approaches have also been implemented by Verma, Shrivastava and O'Mahony in sequential approaches that generated lines heuristically and sought to optimise coordination [18][19]. Sivakumaran et al. proposed a continuous approximation approach that determined the frequency of feeder lines and minimised the weighted sum of the dwell times at station stops, transfer waits at each stop, and the operating cost [19]. Passenger demand was modelled using time-independent values that gradually changed with the distance travelled. They found that by dispatching vehicles in coordination with the given bus schedule, no increase was seen in the driver costs of the bus lines. Research by Wu et al. aimed to develop a bus operating model in which the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

operation cost was formulated and optimised using a GA [20]. Bus operations were also taken into account when building the adjusted operation model. The results showed that their proposed bus operation model could effectively reduce the total cost of the transportation system and the travel times for passengers.

2.1. Neural Network

Artificial neural networks (ANN) are widely used in learning and optimisation problems. Figure 2 shows the neural network model used in the proposed framework for bus schedule optimisation.

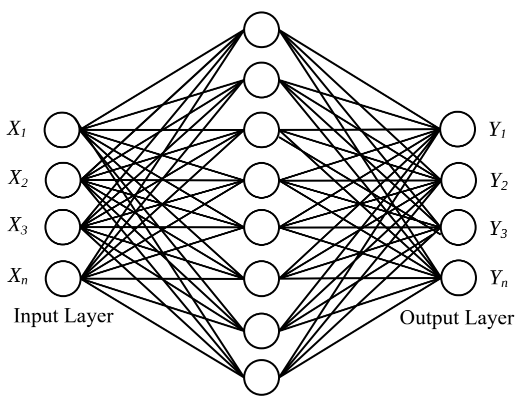


Figure 2: The Neural Network Model

As shown in Figure 2, the neural network consists of an input layer, a hidden layer and an output layer. The input layer has three components, while the output layer has five.

Neural networks are usually trained so that target outputs are achieved. The network is then adjusted based on a comparison between the output and the target until a match is found, as shown in Figure 3.

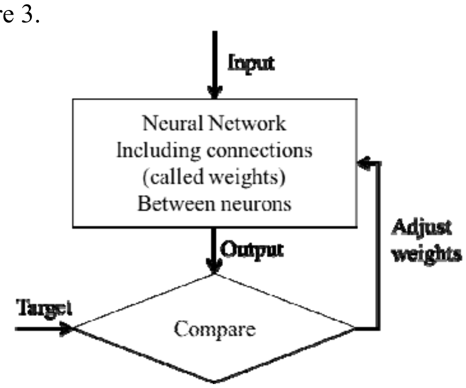
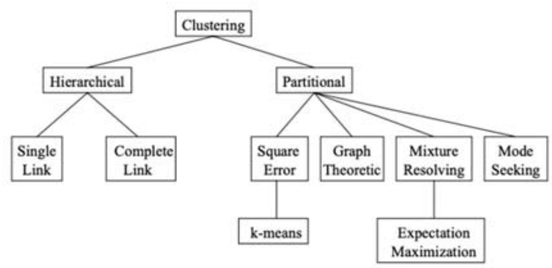


Figure 3: Flow Diagram Of Weight Adjustment In a Neural Network Model

2.2. K-Means Clustering

Clustering is used to discover groups and to identify interesting distributions in the underlying dataset. The clustering steps can be divided into several layers. This creates a basis for merging or splitting clusters based on similarity, and generates a series of partitions. These approaches and the method of layered clustering is illustrated in Figure 4 below. The upper layer of the clustering approach can be divided into a hierarchical algorithm and a partitioning algorithm



[21].

Figure 4. Taxonomy of Clustering Approaches.

K-means is the simplest and most commonly used algorithm employing the squared error algorithm [22]. It starts with a random initial partition and continues reassigning patterns to clusters based on the similarities between the pattern and the cluster centres, until a convergence criterion is met (e.g., there is no reassignment of any pattern from one cluster to another, or the squared error ceases to decrease significantly after a certain number of iterations). This is a popular algorithm as it is easy to implement.

The K-means algorithm consists of the following steps:

1) Initialisation: Suppose we decide to form K clusters for the given dataset. We randomly choose K distinct points (patterns), which represent the centroids of the initial groups. As these centroids will change after each iteration before the clusters are fixed, there is no need to spend time on the selection of the centroids.

2) Assign each object to the group with the closest centroid.

3) When all objects have been assigned, recalculate the positions of the K centroids.

Journal of Theoretical and Applied Information Technology
15th October 2020. Vol.98. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4) Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups based on which the metric to be minimised can be calculated.

The K-means algorithm is most successful for large datasets. It is simple to implement and computationally attractive due to its linear time complexity. However, it should be judiciously used in clustering to achieve meaningful results [22].

3. METHODOLOGY

3.1. GPS Dataset

The primary dataset contains the GPS location coordinates of buses and their time stamps, at one second intervals. In the first step, these data were supplemented by calculating the number of stops made by the bus during each trip. The duration of each stop was derived and added to the dataset. Table 1 shows a sample record representing a bus stopping and the average total stopping time, taken from our pilot study. The GPS data were set to one second intervals for the AVL, and the distance moved could then be measured. The distance travelled was calculated using the spherical law of cosines [23].

Table 1: Time, Distance Moved, Speed, And Bus Move Status, Based On GPS Dataset.

Time	Distance	Speed (km/h)	Bus Status (Stop/Move)
0:11:16	0.000657731	2.4	Move
0:11:17	0.000519982	1.9	Move
0:11:18	0.000413813	1.5	Move
0:11:19	0.001078258	3.9	Move
0:11:20	0	0.0	Stop
0:11:21	0	0.0	Stop
0:11:22	0	0.0	Stop
0:11:23	0	0.0	Stop
0:11:24	0	0.0	Stop
0:11:25	0	0.0	Stop
0:11:26	0	0.0	Stop
0:11:27	0	0.0	Stop
0:11:28	0.001905813	6.9	Move
0:11:29	0.001189536	4.3	Move
0:11:30	0.002112177	7.6	Move

As can be seen from Table 1, if a distance of zero has been travelled over a certain period time (8 seconds or more), this indicates that the bus has stopped, typically for passengers to get on or off.

3.2. Optimisation Framework for Public Bus Scheduling

This study proposes an optimised solution for public bus scheduling and aims to increase the reliability of services with reference to passenger convenience, while reducing the bus operation cost. This section presents our framework for the optimisation of public bus scheduling, which involves classification, clustering and optimisation processes (Figure 5). The following sub-section describes each of the processes in the proposed framework, and a description is given of the experiments carried out to validate the proposed framework.

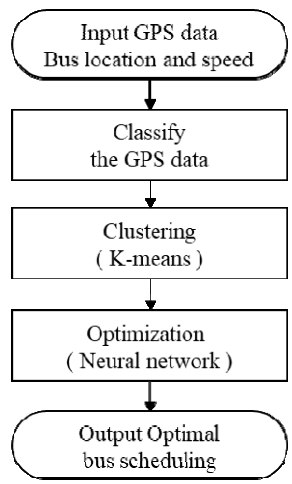


Figure 5: Proposed Optimisation Framework For Public Bus Scheduling

As mentioned above, this research aims to reduce the operational cost while maintaining the number of public bus passengers served, in order to improve the current bus schedule. If the bus provider increases the frequency of bus trips, passengers are able to select their preferred time zone, but the operation cost increases; conversely, if the bus provider reduces the frequency of bus trips to reduce the operation cost, the number of passengers per cycle and travel times will increase, but the service reliability will also be compromised. Finding the optimal balance between demand and supply is challenging, and this can be described as a multi-objective optimisation problem that requires careful selection and application of the optimisation algorithm.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. EXPERIMENTS AND RESULTS

4.1. Classification

Figure 6 shows the proposed model for the classification process. The implementation of this model involves a two-step process: extracting the bus location information, and extracting the status of the bus, i.e. whether it has stopped or is moving.

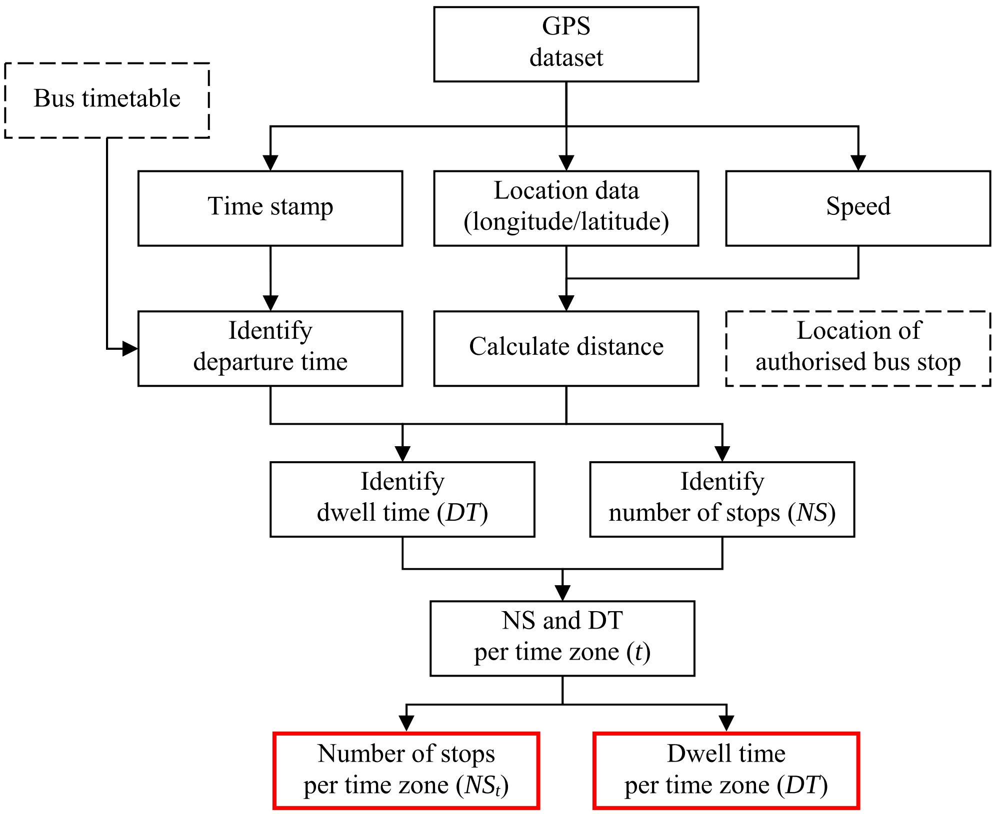


Figure 6: The Data Classification Model

Information on the location and speed of the bus was used to determine its status. Buses stopping due to traffic jams or other traffic conditions were excluded by referring to the information on the locations of bus stops along the specific route provided by the bus operator, which made it reasonably possible to identify whether the bus had stopped at a bus stop and its dwell time. The number of times the bus stopped and the dwell times at these bus stops could then be identified for each bus trip, and these were accumulated for each time zone of operation throughout the day.

Figure 6 shows a flow diagram of the steps involved in data classification.

The parameters used in the above diagram are as follows:

- GPS dataset: This was obtained from a GPS recording, and typically included a series of timestamps and the locations of vehicles in terms of their latitude and longitude.
- Bus timetable: This contained bus departure times for the hours of operation throughout the day. Timetables are usually planned and provided by the bus operator.
- Locations of authorised bus stops: This is a map or location information (GPS data) on

the authorised stops along the route of the bus.

- Departure times: These are the times of departure from the bus terminal, as identified from the GPS timestamp and the bus timetable.

- Distance: This is the Euclidean distance (i.e. an ordinary straight line) connecting the location of the bus with the nearest authorised bus stop.

- Number of stops (*NS*): This is the number of stops made by the bus at authorised bus stops along a specific bus route, where passengers get on or off the bus. The total *NS* per time zone is referred to as NS_t in the following.

- Dwell time (*DT*): This is the time in seconds for which the bus stops at an authorised stop along a specific bus route, to allow passengers to get on or off the bus. The total dwell time per time zone *t* is referred to as DT_t in the following.

- Time zone (*t*): This refers to the hour-long slots of bus service operation, such as 5:00, 6:00, 22:00. These are denoted as *t1* for the first hour (5.00), *t2* for the second hour (6:00), and so on.

The bus timetable and bus stop location data were external information that were not included in the GPS dataset. These are usually made available by the bus operator. From the bus timetable data, it is possible to determine how many vehicles are operating per time zone. Bus stop location data are used to measure the distance between the bus vehicle and the nearest stop. In this way, it is possible to extract the values of NS_t and DT_t when each bus stops at the authorised stops. A dwell time above a certain distance from the nearest authorised bus stop is interpreted as a non-passenger-service dwell time, typically related to road conditions such as traffic jams or street traffic control signals. These non-passenger dwell times are filtered out and are not used in the framework processes.

4.2. Clustering

In the clustering phase, the total number of bus stops (NS_t) and the total dwell time (DT_t) per time zone are used as parameters. The values of these parameters are calculated as part of a data classification step, which is carried out prior to the clustering step. Our clustering model uses the bus stop location information provided by the bus operator. The clustering operation identifies the number of stops per trip and the dwell times at designated bus stops and estimates the number of passengers. Figure 7 shows the operations involved in the clustering phase.

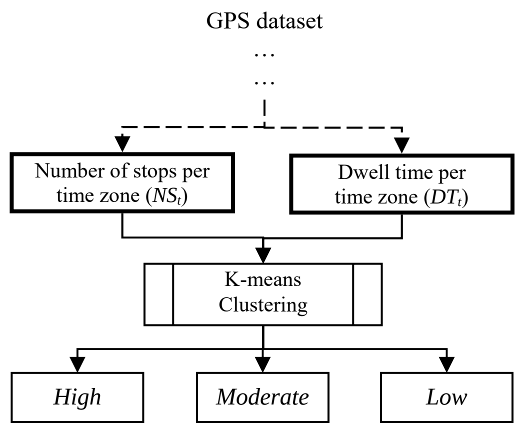


Figure 7: Data Clustering Model (Applied After The Dataset Classification Step)

The clustering model described above uses Number of stops (*NS*) and Dwell time (*DT*) from the classification model described in the previous section (Figure 6). Since this classification output data is classified based on time zones, it is possible to identify how many times a bus stopped at authorised bus stops, and how much time was spent during these brief stops. Accordingly, as per *DT* and *NS*, three clusters of *High*, *Moderate*, and *Low* are identified. This makes it possible to identify and estimate the peak and off-peak hours.

4.3. Neural Network Optimisation

The input layer contains three components, *x1*, *x2* and *x3*, which correspond to the original schedule and the number of buses per time zone (NB_t), the observed number of stops at authorised bus stops (*NS*), and the observed dwell time (*DT*), as per the dataset obtained earlier for the bus operation. Table 2 below shows a sample of the input data for the proposed ANN.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: NB, NS And DT, Used As Input Data For The ANN (Per Day and Time Zone)

Date	*DTZ*	*NB*(x_1)	*NS*(x_2)	*DT*(x_3)
1/7	5:00	1	22	278
1/7	6:00	4	37	1233
1/7	7:00	4	38	2203
1/7	8:00	3	31	833
1/7	9:00	3	28	629
1/7	10:00	4	28	762
1/7	11:00	4	26	749
1/7	12:00	3	22	463
1/7	13:00	4	29	888
1/7	14:00	4	30	1398
1/7	15:00	4	24	910
1/7	16:00	5	39	2043
1/7	17:00	3	29	1309
1/7	18:00	3	33	1923
1/7	19:00	3	30	786
1/7	20:00	4	21	818
1/7	21:00	1	9	133
…	…	…	…	…
31/7	21:00	2	12	196

The ANN output is associated with an array of *correction elements*. The *correction element* structure is set to five, with values: +2, +1, 0, –1, and –2, corresponding to output pattern *y1*, *y2*, *y3*, *y4*, and *y5*, respectively (Table 3). As a correction scheme, the first element (+2) represents adding 2 bus vehicles to the current assigned number per specific time zone; the third value (0) represents a decision to keep the number of bus vehicles as same as in the original plan, and so on.

Based on the ANN output, for each output element (*y1, y2, y3, y4, y5*), a new number of buses (new *NB* [*y1*, *y2*, *y3*, *y4*, and *y5*]) and corresponding estimated dwell time (*eDT* [*y1, y2, y3, y4, y5*]) are calculated per time zone. After this step, each *eDT* is compared to the *moderate dwell time* (*mDT*) derived from the clustering step (Figure 7). The element with the nearest dwell time to the moderate value is used as a target element for that specific time zone set of input set of parameters (Table 3).

MATLAB was used to implement the proposed ANN model. A dataset of 527 points (a set of 17 time zones over 31 days) was used in the experiment. Of the full dataset, 75% was used for training, and the remainder was used for validation and testing. Figure 4 shows the performance of the ANN. The neural network model input and corresponding output patterns that are shown in the table, resulted from the training of the datasets using the MATLAB Neural Network Toolbox.

Table 3: Input, Output Patterns And Correction Elements Of The Proposed ANN

NB **(x_1)**	***NS*** **(x_2)**	***DT*** **(x_3)**	**Output patterns and correction elements**				
			+2	**+1**	**0**	**−1**	**−2**
1	22	278	0	0	1	0	0
4	37	1233	0	1	0	0	0
4	38	2203	1	0	0	0	0
3	31	833	0	0	0	1	0
3	28	629	0	0	0	1	0
4	28	762	0	0	0	1	0
4	26	749	0	0	0	1	0
3	22	463	0	0	0	0	1
4	29	888	0	0	0	1	0
4	30	1398	0	1	0	0	0
4	24	910	0	0	0	1	0
5	39	2043	1	0	0	0	0
3	29	1309	0	1	0	0	0
3	33	1923	1	0	0	0	0
3	30	786	0	0	0	1	0
4	21	818	0	0	0	1	0
1	9	133	0	0	1	0	0
…	…	…	…	…	…	…	…
2	12	196	0	0	0	1	0

A Levenburg–Marquardt backpropagation scheme was applied to train the data, and a neural network with 10 hidden layers was used. As input, data for 17 time zones over 31 days were used. The input data were organised into five elements: date, *DTZ*, *NB*, *NS*, and *DT*, and 70% of the data were used for training, 15% for validation and 15% for testing.

The parameter settings for the neural network algorithm were as follows:

- Data division: random.
- Training: scaled conjugate gradient.
- Performance: cross-entropy.
- Calculations: MEX.

Figure 8 shows the results of testing the trained ANN using weekday and weekend data, each representing a single day. The operation rate is shown for both the ANN schedule and the original operating schedule.

The results show improved cost and dwell time for the ANN schedule. The average operation rate increasing from 26% to 28% for weekdays, and from 43% to 48% per weekends, while maintaining a decrease in number of bus vehicles used per day. 3 and 4 bus vehicles are decreased per weekdays and weekends, respectively (Table 4).

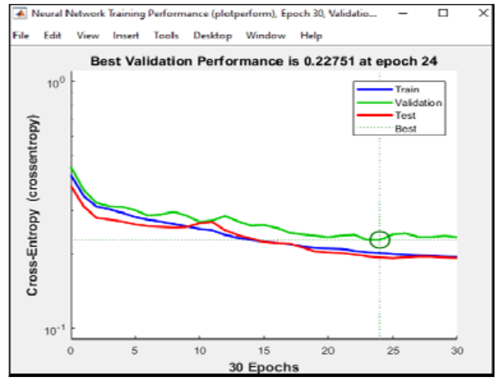


Figure 8: Performance Of The ANN

Table 4: Comparison of Original And ANN Bus Schedules (Line B)

Time	Original (Outbound)				Optimised schedule using ANN (Outbound)			
	Weekday		Weekend		Weekday		Weekend	
	NB	*Or %*	*NB*	*Or %*	*NB*	*Or %*	*NB*	*Or %*
5:00	1	19	0	0	1	19	0	0
6:00	4	29	4	40	5	23	3	40
7:00	4	37	4	48	5	30	5	32
8:00	3	32	2	54	3	32	2	48
9:00	3	33	4	54	2	49	3	72
10:00	4	25	3	49	3	33	2	73
11:00	4	23	3	52	3	30	2	78
12:00	3	23	1	60	2	34	1	60
13:00	4	27	4	43	3	36	5	34
14:00	4	30	3	46	5	24	3	46
15:00	4	21	2	47	3	28	2	47
16:00	5	32	2	53	5	32	2	53
17:00	3	34	3	49	3	34	2	74
18:00	3	29	4	45	5	17	4	45
19:00	3	21	2	42	3	21	2	42
20:00	4	11	2	26	2	22	1	52
21:00	1	11	2	24	1	11	2	24
Sum / Avg	**57**	**26%**	**45**	**43%**	**54**	**28%**	**41**	**48%**

5. CONCLUSION

This paper puts forward a method of improving bus schedules. We began by analysing GPS data from a real-life public bus operator. Three elements from the dataset are identified and used in K-means classification: the number of *buses*, number of *stops*, and the *dwell time* at stops. This research aimed to generate an improved bus schedule by analysing the performance of this operating timetable, using K-means classification followed by the application of a backpropagation neural network algorithm to generate an optimised bus schedule. We have proposed a framework for analysing bus operation datasets and bus route information, and then finding an optimised bus schedule.

The paper has presented experimental results and a performance evaluation of our proposed optimisation framework for bus scheduling. The process is organised into three distinct parts, which correspond to the three steps used to optimise the bus schedule: Start with the process and results of extracting the relevant elements from the dataset, then K-means clustering; and third comes the application of an ANN for schedule optimisation. An evaluation and comparison of the experimental results were presented for each individual contribution to the proposed optimisation framework. We demonstrate that the proposed framework can achieve optimum scheduling without the need for expensive techniques such as APC.

The problem of transport scheduling optimisation using ANNs has not been widely addressed, and as our literature review shows, there have been limited numbers of proposals in this area compared with other fields of optimisation and scheduling. In this study, we have proposed a neural network method that uses a novel correction structure for the scheduled allocation of vehicles to reach a target output, and our experiments have proved the usefulness of our framework.

The experiments carried out here indicate that only a small proportion of the available data need to be used to implement scheduling optimisation. Our proposed methodology was found to reduce the operating costs of the Okinawa bus service and to improve the operation rate.

We have developed a framework for the use of datasets that are typically available or are easy to collect, and have demonstrated the optimisation and improvement of the vehicle scheduling of a bus operator in Okinawa, Japan. Our goal was to generate an optimal scheduling that would minimise the operation cost while offering an improved service to passengers. The framework contained two distinct and consecutive processes: K-means clustering, and the application of an ANN.

Our implementation and the experimental results address the objectives of this research as follows:

1) Theoretical foundations for the proposed framework and methods were explained, and their experimental application to bus operation datasets was demonstrated. The details of the structure and the implementation of the processing steps are discussed in Sections 3 and 4, respectively.

2) A schedule was generated that gave a significant reduction in the operation cost in terms of the size of the bus fleet. The operating rate was reduced by half, while maintaining comfort and a reliable service for passengers, as discussed in Section 4.

3) An ANN was applied with a carefully designed learning formula that made use of K-means clustering groups. Good optimisation results were for the bus schedule compared to the existing operating schedule. The improvements in the operation cost and operating rate (discussed in Section 2) demonstrated the effectiveness of our method.

A novel approach was used in regard with the dataset. Three elements were extracted and used in the process: bus departure time zones (DTZ); number of stops at bus stops per time zone (NS_t); and the dwell time in each time zone (DT_t).

The optimisation framework consisted of three distinct steps. In the first step, NS_t and DT_t were extracted, while in the second, K-means clustering was used to group and distinguish a moderate bus operation rate, based on a moderate dwell time. In the third step, the ANN was used to generate the vehicle schedule as a result of ANN optimisation. The moderate dwell time (mDT) (i.e. the outcome of the clustering step) was used as a criterion for reducing or increasing the number of vehicles in the ANN process.

Our experiments involving the proposed optimisation processes gave clear improvements in the generated schedule. The result was a lower operation cost in terms of the number of operating vehicles, while continuing to meet passenger demand. Our research has established a framework and methods for optimising operating costs while fulfilling passenger demand for a reliable service and minimal waiting times.

6. ACKNOWLEDGEMENTS

The data used in this research referred to public buses in Okinawa Prefecture. We would like to thank the Ryukyu, Okinawa, Naha and Toyo bus operators for providing data on their bus operations and routes. We also extend our gratitude to Okinawa Industry Promotion Corporation for providing partial funding for this research. Also we would like to our thanks to Universiti Kebangsaan Malaysia (under project code ZG-2018-001) and Meio University, Japan, for supporting this research project.

REFRENCES:

[1] A. Ceder, "Public Transit Planning and Operation: Modeling, Practice and Behavior", CRC press, 2016.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[2] J. Wu, Y. Min, S. Rasouli, and C. Xu, "Exploring Passenger Assessments of Bus Service Quality Using Bayesian Networks", Journal of Public Transportation 19(3): 3, 2016.

[3] M. Wagale, A. P. Singh, A. K. Sarkar, and S. Arkatkar, "Real-Time Optimal Bus Scheduling for a City Using a Dtr Model" Procedia - Social and Behavioral Sciences 104, 2013, pp. 845-854.

[4] F. D. Wihartiko, A. Buono, and B. P. Silalahi, B. P. "Integer Programming Model for Optimizing Bus Timetable Using Genetic Algorithm" IOP Conference Series: Materials Science and Engineering 166, 2017.

[5] O. Cats, and Z. Gkioulou, "Modeling the Impacts of Public Transport Reliability and Travel Information on Passengers' Waiting-Time Uncertainty", EURO Journal on Transportation and Logistics 6(3), 2017, pp. 247-270.

[6] V. Oort, N., D. Sparing, T. Brands, and R.M.P. Goverde, "Optimizing Public Transport Planning and Operations Using Automatic Vehicle Location Data: The Dutch Example", Conference paper International Conference on Models and Technologies for Intelligent Transport Systems, MT-ITS, Dresden (Germany) 2-4 Dec. 2013.

[7] J. Patnaik, S. Chien, and A. Bladikas, "Using data mining techniques on apc data to develop effective bus scheduling plans", Journal of Systemics, Cybernetics and Informatics, 4(1), 86-90, 2006.

[8] M. Sandim, R. J. Rossetti, D. C. Moura, Z. Kokkinogenis, and T. R. Rúbio, "Using GPS-based AVL data to calculate and predict traffic network performance metrics: A systematic review", In 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC) (x). IEEE, 2016, pp. 1692-1699.

[9] P. Eklund, and C. Cook, "Toward real-time multi-criteria decision making for bus service reliability optimisation", International Symposium on Methodologies for Intelligent Systems, HLM, 2015, pp. 371-378.

[10] R. Giesen, H. Martínez, A. Mauttone, and M. E. Urquhart, "A Method for Solving the Multi - Objective Transit Frequency Optimization Problem", Journal of Advanced Transportation 50(8), 2016, pp. 2323-2337.

[11] A. Verma, and S. Dhingra, "Developing integrated schedules for urban rail and feeder bus operation", Journal of Urban Planning and Development, Vol. 132, 2006, pp. 138-146.

[12] E. Pekel, and S. Soner Kara, "A comprehensive review for artificial neural network application to public transportation", Sigma: Journal of Engineering & Natural Sciences/Mühendislik ve Fen Bilimleri Dergisi, Vol. 35, No. 1, 2017.

[13] O. J. Ibarra-Rojas, F. Delgado, R.Giesen, and J. C. Muñoz, "Planning, operation, and control of bus transport systems: A literature review", Transportation Research Part B: Methodological, 77, 2015, pp. 38-75.

[14] Y. Hadas, and M. Shnaiderman, "Public-transit frequency setting using minimum-cost approach with stochastic demand and travel time", Transportation Research B, Vol. 46, 2012, pp. 1068-1084.

[15] Y. Li, W. Xu, and S. He, "Expected value model for optimizing the multiple bus headways", Applied Mathematics and Computation, Vol. 219, 2013, pp. 5849–5861.

[16] D. Van Lierop, M. G. Badani, and A. M. El-Geneidy, "What influences satisfaction and loyalty in public transport? A review of the literature", Transport Reviews, Vol. 38, No. 1, 2018, pp. 52-72.

[17] P. Shrivastava, and S. Dhingra, "Development of coordinated schedules using genetic algorithms", Journal of Transportation Engineering, Vol. 128, 2002, pp. 89-96.

[18] P. Shrivastava, and M. O'Mahony, "Modeling an integrated public transportation system – a case study in Dublin, Ireland", European Transport, Vol. 41, 2009, pp. 1-19.

[19] K. Sivakumaran, Y. Li, M. Cassidy, and S. Madanat, "Cost-savings properties of schedule coordination in a simple trunk-and-feeder transit system", Transportation Research A, Vol. 46, 2012, pp. 131-139.

[20] J. Wu, R. Song, Y. Wang, F. Chen, and S. Li, "Modelling the coordinated operation between bus rapid transit and bus", Mathematical Problems in Engineering, Article ID 709389, in press, 2013.

[21] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data Clustering: A Review", ACM Computing Surveys 31(3), 1999, pp. 264-323.

[22] J. Mcqueen, "Some Methods for Classification and Analysis of Multivariate Observations", the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297.

[23] H. Miura, "Three distance calculation methods using latitude and longitude", Operations Research, Vol. 60, 2015.