Using Graph Algorithm and Classification Technique for Finding an Optimal Bus Route in Time-Dependent Travel Times

Satjayod Deelertmongkolchai, Todsanai Chumwatana, and Choonhapong Thaiupathump
Rangsit University, Thailand, satjayod.deelertmongkolchai@th.rhenus.com, {todsanai.c; choonhapong}@rsu.ac.th

ABSTRACT
In the last decade, traffic jam has been regarded as a main problems for Bangkok. Most people selects a bus option for traveling because it is cheap and cover every area in Bangkok. However, they are suffering from the long hours in traffic jam especially in rush hour. They also cannot avoid this such jam as bus routes are fixed by Bangkok Mass Transit Authority (BMTA). This paper aims to propose a technique for finding an Optimal Bus Route in Time-Dependent Travel Times by using graph algorithm and data mining technique. The proposed technique is able to find a least spent travel time path between two nodes in a bus network with time-dependence. Graph algorithm is used to generate all possible paths to reveal the distances. Classification technique is then used to analysis traffic situation in different period of the time. By analysis traffic situation, date, time, week, month, location are used as a main factor for training process in classification technique. From the experimental studies, the proposed technique is able to show the best route from any given node to the final destination depending on the different period of the time. The proposed technique provides significant benefit for traveler to select the best bus route, which is short distance and fast, among generated routes

Keywords: Optimal bus route, Decision support system, Graph algorithm, Classification, Data mining.

I INTRODUCTION

At present, Bangkok Mass Transit System has various mode of transportation such as taxi, public bus, sky train (BTS) or subway (MRT) which an optimal transportation depends on people's final destination. A survey shows that public bus transportation is the most popular transportation mode in Bangkok because of its convenience to find a bus stop nearby and route coverage. Considering a fare, bus transportation has cheapest fare among others. The present fare for open-air bus is 6.5 - 9.00 THB and for air-conditioned bus is 11.0 - 26.0 THB. Due to the bus routes are fixed by Bangkok Mass Transit Authority (BMTA) which some routes pass through traffic congested area, people have to face the unexpected time spent in travelling by public bus. In order to save the travelling time of the metropolitan, this paper aims to find an optimal bus route with least time spending by integration of two techniques which are graph algorithm and classification technique. As the result, the proposed technique provides significant benefit for traveler to select the best bus route, which is short distance and time saving, among the generated routes.

II RELATED WORK

Graph Algorithm (Alfred, 1983) is a main technique used to find an optimal route. The Graph Algorithm emphasizes a connection between two vertices which are called "Adjacent Vertices" that have connecting line between two vertices as shown in Figure 1.

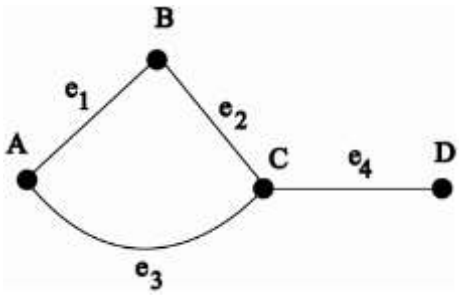


Figure 1. The Example of Graph

Description of the above Figure 1:

Vertex "A" and "B" are "Adjacent Vertices"
Vertex "A" and "C" are "Adjacent Vertices"
Vertex "B" and "C" are "Adjacent Vertices"
Vertex "C" and "D" are "Adjacent Vertices"

BUT Vertex "A" and "D" are not "Adjacent Vertices"
Vertex "B" and "D" are not "Adjacent Vertices"

From above definition, Graph Algorithm is used to find the shortest path from point to point by using Minimum Spanning Tree Algorithm or shortest path of the graph to find the shortest path which connects between each point.
Several researches are proposed for finding an optimal route by using Graph Algorithm such as

"Finding Safe and Optimal Route for Trip Itinerary"(Suwannapruek, 2011) which mentions a finding of the shortest path by using Graph Algorithm (Alfred, 1983) and "Considering Safety from an Severity Index of Junctions for Proposing Safe Route for Travelling" or "Finding Transport Routes for Bulk Goods by using Dynamic Programming" (Manee-ngam, 2013) which mentions mathematical techniques used for finding transport routes that are minimum total cost of transport by comparing with an original transport which uses the shortest path, the result illustrates that the total cost of transport can be reduced to 17% from the original path selection method.

In year 2014, the research "An Application of the Travelling Salesman Problem Case Study: Routing for Streetcar Tour of the Chiang Rai Municipality" (Seksan, 2014) proposes a finding of shortest path by means of linear programming in comparison with an existing route which the electric car tour runs through all the tourist locations, the experimental result finds a shorter path to 9.075% in comparison with an existing route and proposes a factor of traffic condition for the best result for route management.

Furthermore, classification theory is applied to this research by using K-NN Classification method (Pang-Ning, 2005) which classifies data by measuring the distance between predicted data and nearby data that defined the amount as "K", the result from prediction shows a group of most data which is close to the "K" amount. A method of measuring a distance according to an equation below:

$$D_{Euclidean} = \sqrt{(F1_{test} - F1_{train})^2 + (F2_{test} - F2_{train})^2}$$

This method is applied to several researches , for example "A Comparison of Credit Card Approval Prediction Efficiency between Four Classification Methods of Thai Commercial Banks" (Sinsomboonthong, 2014) which proposes to compare an efficiency of each algorithm (Ian, 2005), for example : K-NN, Decision Tree, Neural Network and Support Vector Machine. As a result, a neural network method shows the best accuracy at 70.37% in comparison with an efficiency of other algorithms.

III USING GRAPH ALGORITHM AND CLASSIFICATION TECHNIQUE FOR FINDING AN OPTIMAL BUS ROUTE IN TIME DEPENDENT TRAVEL TIMES

This research proposes an integrated technique between Graph Algorithm and Classification Technique for finding an optimal bus route. A working process is shown in Figure 2

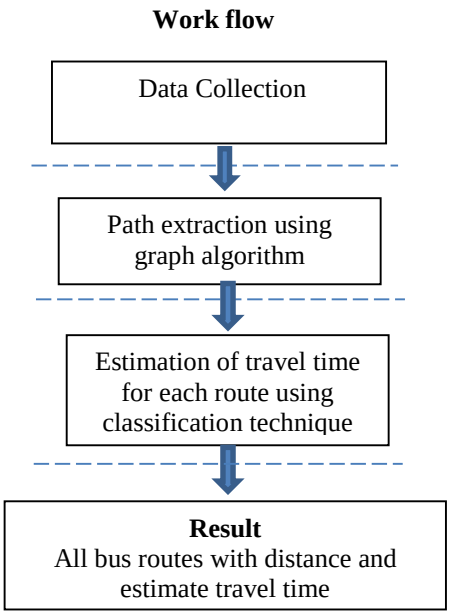


Figure 2. Process of the proposed Technique

From Figure 2, a proposed working technique consists of two main steps which are A) using graph algorithm for finding route from place to place and B) using classification technique for bus route time calculation by predicting a traffic condition from traffic density data. The result from the proposed technique is an optimal bus route from place to place which takes less time and short path. The detail of each processes will be described as following:

A. Using graph algorithm for finding route from place to place

This step begins with the data collection from a distance of each bus stop regarding to a fare table of a bus which passes the studied route as shown in Figure 3.

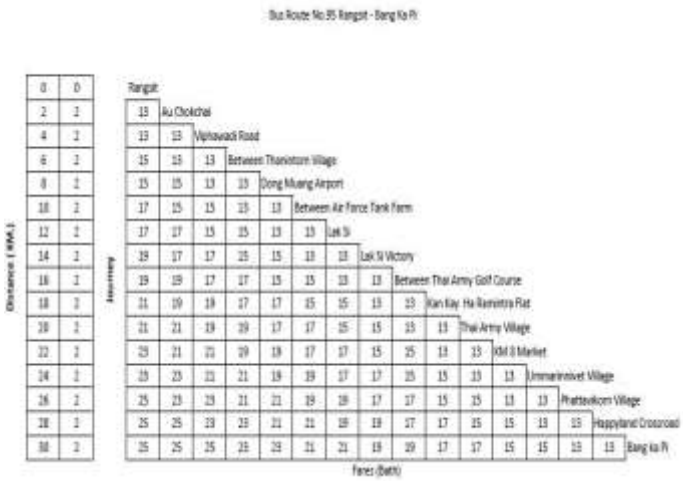


Figure 3. Fare of bus No.95

From Figure 3. the above fare table is used for defining the places on map (refer to: https://map.longdo.com/) and finding a distance of each bus stop in order to collect data of a distance and a bus which passes each place as shown in Figure 4 and Table 1.

sounce	Destination	Distance (m)	Bus number
AA	BB	900	29,34,39,59,95n,185,503,510,520,522
BB	CC	1500	29,59,95n,510
CC	DD	3100	29,59,95n,510
DD	EE	1900	29,59,95n,510
EE	FF	2600	29,59,95n,510
FF	GG	2800	29,510
GG	HH	3100	29,510
HH	II	1700	29,510
II	JJ	1900	29,510
JJ	KK	2400	29,510
KK	LL	3300	34,39,59,503
LL	MM	600	34,39,59,503
MM	NN	1600	34,39,59,185,503
NN	OO	2200	34,39,59,185,503,522
NN	HH	1600	522
OO	PP	4500	34,39,59,185,503,522

Figure 4. Points Of Bus Stops On A Map

Table 1. List of bus number pass from source to destination

From the above data, a distance between source to destination can be calculated by distributing all possible routes from a graph and calculating total distance from place to place as shown in Figure 5 below which KK is defined to be a source and OO is defined to be a destination and a total distance of each route is also presented.

Route 1	KK--> LL--> MM--> NN--> OO-->	7700	m.
Route 2	KK--> JJ--> II--> HH--> NN--> OO-->	9800	m.
Route 3	KK--> JJ--> II--> HH--> GG--> FF--> QQ--> PP--> OO-->	18250	m.
Route 4	KK--> JJ--> II--> HH--> GG--> FF--> EE--> DD--> CC--> BB--> UU--> TT--> SS--> RR--> PP--> OO-->	35300	m.

Figure 5. Showing all routes and total distance of each route from KK to OO

The routes and total distance of each route are used to calculate traffic density of each route in order to predict travelling time as described below.

B. Using Classification Technique for Bus Route Time Calculation by Predicting a Traffic Condition from Traffic Density Data

This step begins with a collection of online traffic condition data on map (refer to: https://map.longdo.com/) in each period of time, starting from 6.00 am. to 8.00 pm (working hour) as shown some example in Figure 6. With data collection of retroactive date and time, this is applied to training process and leading to predictive ability of future traffic condition.

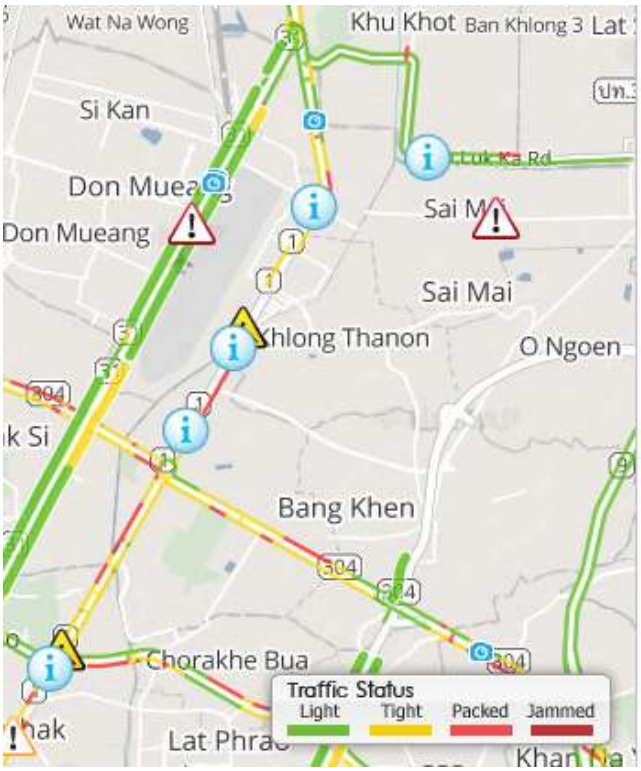


Figure 6. Online Traffic Condition

From the above figure, traffic condition is classified to four levels as follows (refer to: www.traffic.longdo.com):

- Green line is used to define a good traffic
- Yellow line is used to define a traffic congestion
- Orange line is used to define a traffic jam

- Red line is used to define a really bad traffic

These traffic conditions are applied as the factor of travelling time calculation on each route which is extracted in the first step as shown in Figure 5. In this paper, K-NN Classification Technique (Ian, 2005) is applied to traffic condition predictive analytic by calculating from a retroactive data as follows:

Souce	AA	UU	MON	SUN	06.00-07.00	19.00-20.00	WEEK1	WEEK4	STATUS
AA	N	N	Y	N	Y	N	Y	N	GREEN
AA	N	N	Y	N	N	N	Y	N	GREEN
AA	N	N	Y	N	N	Y	Y	N	YELLOW
BB	N	N	Y	N	Y	N	Y	N	GREEN
BB	N	N	Y	N	N	N	Y	N	GREEN
CC	N	N	Y	N	N	N	Y	N	GREEN
CC	N	N	Y	N	N	N	Y	N	GREEN
DD	N	N	Y	N	N	N	Y	N	YELLOW
DD	N	N	Y	N	N	N	Y	N	YELLOW
DD	N	N	Y	N	N	Y	Y	N	GREEN
FF	N	N	Y	N	Y	N	Y	N	ORANGE
GG	N	N	Y	N	N	N	Y	N	GREEN
GG	N	N	Y	N	N	N	Y	N	GREEN

Figure 7. Sample Of Data Collection Which Is Used For Analysis

When collecting the results from above two steps, it founds that the techniques show a result of all bus routes from a source to a destination, including a distance and an average time for travelling as shown in Table 2.

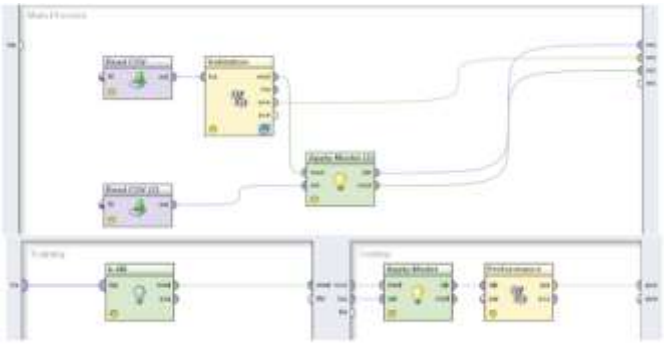
Table 2. Example Of Result From The Propose Technique

Route	Bus stop station in the each place	Distance (m)	Total Times (Min)	Suggestion For Travelling	
				Source - Destination	Bus Number
Route 1	KK LL MM NN OO	7700	25.30	KK - OO	34,39,59,305
Route 2	KK JJ II HH NN OO	9800	22.00	KK - HH	29,510
				HH - OO	522
Route 3	KK JJ II HH GG FF QQ PP OO	18250	40.68	KK - FF	29,510
				FF - OO	59
Route 4	KK JJ II HH GG FF EE DD CC BB UU TT SS RR PP OO	35300	84.20	KK - BB	29,510
				BB - UU	34,39,305

From Table 2, the result of the proposed technique can supports people to make a smart decision when travelling by bus. The decision can be made into different ways. For example, some people select long distance but spend faster time while the others may select short distance but spend long time. The decision is depending on individual acceptable and situation.

IV EXPERIMENTAL STUDIES

This section describes an experiment for finding the optimal Bus Route with travelling time by using efficiency technique which is proposed in this research. Example of an experiment for finding an optimal bus route starts from KK (Mochit Skytrain Station) and finishes at OO (Bangbua Temple). An experiment of the travel is made on Wednesday of week 4 of a month from 6.00 pm. to 7.00 pm and the results show as follows:



Route 1	KK--> LL--> MM--> NN--> OO-->	7700	m.
Route 2	KK--> JJ--> II--> HH--> NN--> OO-->	9800	m.
Route 3	KK--> JJ--> II--> HH--> GG--> FF--> QQ--> PP--> OO-->	18250	m.
Route 4	KK--> JJ--> II--> HH--> GG--> FF--> EE--> DD--> CC--> BB--> UU--> TT--> SS--> RR--> PP--> OO-->	35300	m.

Figure 8. Result of an Applicable Route for Travelling

The next section is a prediction of traffic condition by using Rapid Minor Software for testing as shown in Figure 9.

Figure 9. Experimental Analysis on Rapid Minor Software

From this experiment which is made by collecting data for 2 months, a result shows as follows:

Table 3. Data collection for 2 months

Souce	Prediction	Con G	Con Y	Con Or	Con R	AA	UU	WED	18.00-1	WEEK1
AA	YELLOW	0	1	0	0 N	N	Y	Y	Y	Y
BB	YELLOW	0	1	0	0 N	N	Y	Y	Y	Y
CC	GREEN	1	0	0	0 N	N	Y	Y	Y	Y
DD	GREEN	1	0	0	0 N	N	Y	Y	Y	Y
EE	YELLOW	0	1	0	0 N	N	Y	Y	Y	Y
FF	GREEN	1	0	0	0 N	N	Y	Y	Y	Y
GG	GREEN	1	0	0	0 N	N	Y	Y	Y	Y
HH	ORANGE	0	0	1	0 N	N	Y	Y	Y	Y
II	YELLOW	0	1	0	0 N	N	Y	Y	Y	Y
JJ	ORANGE	0	0	1	0 N	N	Y	Y	Y	Y
BB	YELLOW	0	1	0	0 N	Y	Y	Y	Y	Y
UU	YELLOW	0	1	0	0 N	N	Y	Y	Y	Y
TT	YELLOW	0	1	0	0 N	N	Y	Y	Y	Y
SS	ORANGE	0	0	1	0 N	N	Y	Y	Y	Y
RR	ORANGE	0	0	1	0 N	N	Y	Y	Y	Y
QQ	ORANGE	0	0	1	0 N	N	Y	Y	Y	Y

From Figure 8 and Table 3, the results of two steps is used for calculating an average time by calculating average speed from www.traffic.longdo.com as shown in Figure 10.

Green	Average Speed 40 Km / Hour
Yellow	Average Speed 20 Km / Hour
Oreange	Average Speed 15 Km / Hour
Red	Average Speed 10 Km / Hour

Figure 10. Average Speed According To The Highlight

The average speed in different time period shown in the above figure is used to find a time of each route, a calculation result is shown in Table 4.

Table 4. A Prediction Result In The Experimental Route

Source		Destination		Distance (m)	Bus Number	Prediction	Total Time (min)
KK	Mochit skytrain Station	LL	Sahmrananbengkoen School	3300	34,39,59,503	yellow	9.9
LL	Sahmrananbengkoen School	MM	Ratchayothin Intersection	600	34,39,59,503	yellow	2.4
MM	Ratchayothin Intersection	NN	Kasetsart Intersection (Phahon Rd.)	1600	34,39,59,385,503	orange	6.4
NN	Kasetsart Intersection (Phahon Rd.)	OO	Bangbua Temple	2200	34,39,59,385,503,522	yellow	6.6
KK	Mochit skytrain Station	JJ	Ho Wang School	3400	29,530	orange	3.6
JJ	Ho Wang School	II	Mass Newspaper	1900	29,530	yellow	2.85
II	Mass Newspaper	HH	Kasetsart Intersection (Vibhavadi Rd.)	1700	29,530	orange	2.55
HH	Kasetsart Intersection (Vibhavadi Rd.)	NN	Kasetsart Intersection (Phahon Rd.)	1600	522	orange	6.4
HH	Kasetsart Intersection (Vibhavadi Rd.)	GG	Police Academy School	3100	29,530	green	9.3
GG	Police Academy School	FF	Lak si Vibhavadi Rd.	2800	29,530	green	4.2
FF	Lak si Vibhavadi Rd.	QQ	Phra Nakhorn Rajabhat University	850	59,954	orange	3.4
QQ	Phra Nakhorn Rajabhat University	PP	Laksi Monument	850	59,954	green	1.275
PP	Laksi Monument	OO	Bangbua Temple	4500	34,39,59,385,503,522	yellow	13.5
FF	Lak si Vibhavadi Rd.	EE	Bangkok Aviation Fuel Services	2600	29,59,954,530	yellow	7.8
EE	Bangkok Aviation Fuel Services	DD	Don Mueang International Airport	2900	29,59,954,530	yellow	2.85
DD	Don Mueang International Airport	CC	Thanin Thani Village	3100	29,59,954,530	green	4.65
CC	Thanin Thani Village	BB	Vibhavadi Rangsit Intersection	1500	29,59,954,530	yellow	4.5
BB	Vibhavadi Rangsit Intersection	UU	Phahon Yothin 62	1600	34,39,385,503,530,522	yellow	4.8
UU	Phahon Yothin 62	TT	Royal Thai Air Force Academy	2200	34,39,385,503,530,522	yellow	6.6
TT	Royal Thai Air Force Academy	SS	Bhumibhol Adulyadej Hospital	1600	34,39,385,503,530,522	yellow	6.4
SS	Bhumibhol Adulyadej Hospital	RR	Ying Charoen Market	1600	34,39,385,503,530,522	orange	6.4
RR	Ying Charoen Market	PP	Laksi Monument	2800	34,39,385,503,530,522	green	4.2

An above result can be summarized an experiment for finding route which starts from KK (Mochit Skytrain Station) and finishes at OO (Bangbua Temple), the result is presented in Table 5 and Figure 11.

Table 5. An Experimental Result Of A Suggestion For Travelling By Bus Starting From KK (Mochit Skytrain Station) And Finishing At OO (Bangbua Temple)

Route	Bus stop station in the each place																Distance (m)	Total Times (Min)	Suggestion For Traveling	
																			Source - Destination	Bus Number
Route 1	KK	LL	MM	NN	OO												7700	25.30	KK - OO	34,39,59,503
Route 2	KK	JJ	II	HH	NN	OO											9800	22.00	KK - HH	29,510
																			HH - OO	522
Route 3	KK	JJ	II	HH	GG	FF	QQ	PP	OO								18200	42.68	KK - FF	29,510
																			FF - OO	59
Route 4	KK	JJ	II	HH	GG	FF	EE	DD	CC	BB	UU	TT	SS	RR	PP	OO	35300	84.20	KK - BB	29,510
																			BB - UU	34,39,503

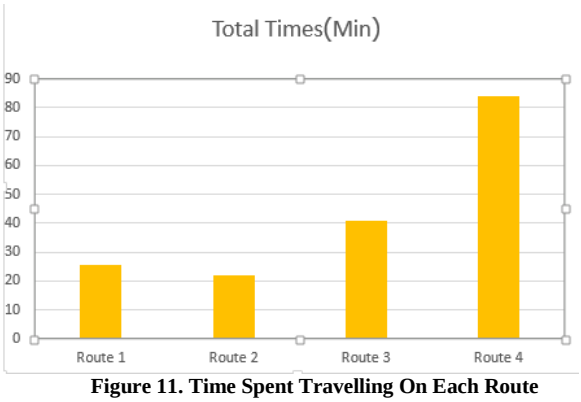


Figure 11. Time Spent Travelling On Each Route

From graph shown in Figure 11, an experiment for finding all applicable routes shows four routes which the shortest-distance route is route 1 and the longest-distance route is route 4. When a factor of traffic condition prediction is integrated to this experiment, the result shows that the distance of route 1 is 7,700 meters long that takes 25.30 minutes time for travelling, whereas the distance of route 2 is 9,800 meters long but it takes only 22 minutes time for travelling. As a result, it can be assumed that route 2 would be an optimal bus route for travelling from KK to OO place.

V CONCLUSION

This research proposes the integrated techniques between 2 theories in order to provide an optimal bus route by using Graph Algorithm to generate the possible bus routes and integrates a classification technique to this research by using K-NN Algorithm for traffic condition prediction. From the experimental study, the result of an integration of 2 theories reveals the best result which is an optimal bus route. In the comparison between route 1 and route 2, the result shows that although route 1 is the shortest distance route but takes 25.30 minutes for travelling, on the other hand, route 2 is 2,100 meters longer than route 1 but takes only 22 minutes for travelling. Hence, this research provide significant benefit for traveler to find the optimal bus route in time-dependent travel times in Bangkok.

REFERENCES

Alfred, V. Aho, J. Hopcroft, E and Jeffrey U. (1983). Data Structures and Algorithms (1st ed.). *Addison-Wesley Longman Publishing Co., Inc*., Boston, MA, USA.

Ian, H. Witten and Eibe, F. (2005). Data Mining: Practical Machine Learning Tools and Techniques, *Second Edition (Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann Publishers Inc*., San Francisco, CA, USA.

Manee-ngam, A. and Sripathamswat, K. (2013). Path Selection for Bulk Cargo Transportation using Dynamic Programming. vol. 1, pp. 52-62, 2013.

Pang-Ning, T. Michael, S. and Vipin, K. (2005) .Introduction to Data Mining, (First Edition). *Addison-Wesley Longman Publishing Co., Inc*., Boston, MA, USA.

Seksan, W. (2014). An Application of the Travelling Salesman Problem Case Study: Routing for Streetcar Tour of the Chiang Rai Municipality. *Industrial Technology Lampang Rajabhat University Journal*,.vol. 2, pp. 85-97.

Sinsomboonthong, S. (2014). A Comparison of Credit Card Approval Prediction Efficiency between Four Classification Methods of Thai Commercial Banks. *Journal of Industrial Education*.

Suwannapruek ,T. (2011). The development of an optimal safe path for trip planning. *IEEE Transactions on Communications*, pp. 105-111