Bus Route Design with a Bayesian Network Analysis of Bus Service Revenues

Yi Liu, Yuanhua Jia, Xuesong Feng and Wu Jiang

MOE Key Laboratory for Urban Transportation Complex System Theory & Technology, Beijing Jiaotong University, No.3 Shangyuancun, Haidian District, Beijing 100044, P.R. China.

Abstract: A Bayesian network is used to estimate revenues of bus services in consideration of the effect of bus travel demands, passenger transport distances, and so on. In this research, a newly urbanized region of city in China has been selected as the study area because of its relatively high bus travel demand and, on the contrary, unsatisfactory bus services. It is suggested that the proposed Bayesian network approach is able to rationally predict the probabilities of different revenues of various route services, from the perspectives of both satisfying passenger demand and decreasing bus operation cost. In this way, the existing bus routes in the studied area can be optimized for their most probable high revenues.

Keyword: bus route design; revenue probability forecast; Bayesian network; probabilistic inference

1 Introduction

The bus network design problem has become one of commonly hard-to-solve issues of many cities in the world today. Unsatisfying bus services resulted from irrational bus routes make their revenues decreased. Therefore, a good design of a bus route is essential to satisfy the bus travel demand and improve the bus service revenue. Many methods for rationally designing a bus route or network have been continually developed. Ceder and Israeli (1998) combined mathematical programming approaches and decision-making techniques to solve the transit network design problem. Chien and Spasovic (2002) studied a grid bus transit system to optimize route spacing, station spacing, headway, and fare with the objectives of maximum total operator profit and social welfare. Lee and Vuchic (2005) offered an iterative approach to solve the relationship between the variable transit trip demand and the transit network design, under a given fixed total demand. Wirasinghe and Vandebona (2011) considered express route planning problem based on both grid and non-grid road networks to minimize operating costs, passenger access costs, waiting time and traveling time. Tirachini *et al*. (2014) introduced a social welfare maximization model with the interplay between congestion and crowding externalities, with the aim of optimizing the design of urban bus routes. Ceder *et al*. (2015) used the bi-level optimization approach realistically models to design of stop placement along a single bus route which allows the effects of uneven topography to be explicitly considered, with considering user and operational costs of a bus service.

The majority of researchers tried to minimize the total travel time, or the generalized cost. Genetic algorithm (GA), Tabu search, simulated annealing methods and so forth have all played important roles in recent research on bus network design. Zhao and Zeng (2006) combined GA and simulated annealing to minimize transfers with reasonable route directness while maximizing service coverage. Fan and Machemehl (2008) provided a multi-objective model for considering the design of public transportation networks in the case of variable demand, and the solution methodology was based on the Tabu search method. Pacheco *et al*. (2009) proposed an approach based on a local search strategy to solve route design and bus assignment problem under the effect of different demand. Szeto and Wu (2011) combined GA and a neighborhood search heuristic to simultaneously perform the suburban route design. Nayeem *et al*. (2014) have developed two versions of GA based meta-heuristic to discuss the transit routing problem, with the aim of minimizing the travel time and the number of transfers simultaneously.

Inherently, the bus service revenues are usually sensitive to passengers, operating costs and so on. Therefore, there is a strong need to develop a new approach from a comprehensive perspective that takes into account the dependency and uncertainty of the correlated factors of bus routes. A Bayesian Network (BN) can be used to discover the overall dependency structure of a large number of variables under uncertainty and incomplete information, and to predict future events such as traffic accidents in comparatively accurate manners (Ozbay and Noyan 2006; Pearl 2014; Sun and Sun 2015). As a result, this research has newly proposed an approach based on BN to predict the probabilities of different revenues of various route services to satisfy passenger demand and decrease bus operation cost. In this way, the existing bus routes in the studied area can be optimized for their most probable high revenues.

This paper is organized as follows. Section 2 introduces the study area represented as X here. The proposed process of BN learning and of probabilistic inference is given in section 3 and section 4, respectively. Section 5 presents the application of the methodology to the optimization of bus routes in X. Section 6 summarizes conclusions and comments.

2 Study Area

The proposed BN has been implemented on the real bus routes in X which is displayed in Figure 1. Data was obtained from the administrative statistics of the government department. The routes in Figure 1 are mainly concentrated in the north area of X. The uneven distribution of routes causes uneven distribution of bus passenger demands, and it also causes the traffic congestion which makes the traveling speed excessive low in the north area of X. It is found that the ratio of bus routes which their average load factor bellows 0.3 (Kuang 2005) reaches 25.00%. The lengths of most routes are relatively long and their daily passenger transport volumes are comparatively low, which makes their revenue per day often cannot meet with their costs. Merely about 50% of bus routes are very detours that cause the average nonlinear coefficient of those routes to exceed 1.4. Because of its relatively high bus travel demand and unsatisfactory bus services, X has been selected as the study area of this research to explain the proposed BN approach.

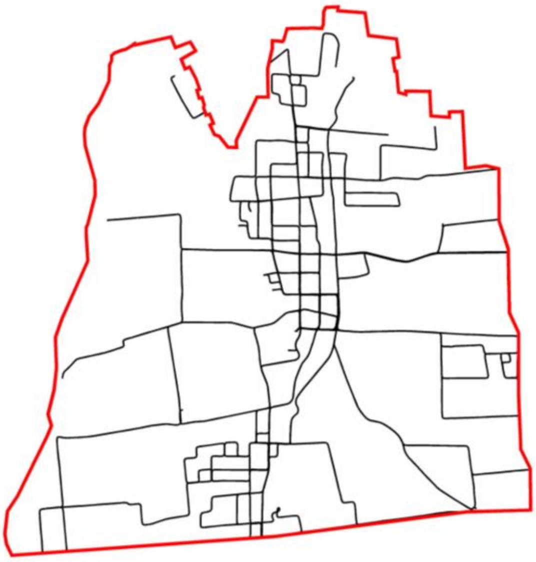

Figure 1. The existing layout of bus routes in X.

3 Methods

3.1 Bayesian Network

The model structure of a BN combining the principles from graph theory, probability theory and statistics is known as a directed acyclic graph *G* defined by the respective sets of nodes and directed edges (Jensen 2001; Feng *et al*. 2016). The nodes represent random variables, and the directed edges represent relationships between variables. Associated with each node of a BN is a conditional probability distribution (*CPtable*) quantifying how much a node depends on its parent node(s) (Pearl 1988; Cheng *et al*. 2002).

Formally, the directed acyclic graph *G* encodes conditional independence between the variables. Each variable X_i is independent of its non-descendants in the graph given the state of its parents. Thus, the joint distribution of *n* variables $X_1,...,X_n$ which can be described by Equation (1) is the product of the conditional distributions of each variable given its parent node(s).

$$P(X_1,..,X_n) = \prod_{i=1}^{n} p(x_i|\pi_i)\,, \tag{1}$$

where *n* represents the number of the all variables, π_i is an instantiation of the set of parent node(s) of X_i, and x_i denotes the state of X_i.

The conditional probabilities appear in the product form are stored in *CPtable* as the parameters. When the topology and *CPtable* have been completed, Bayes' theorem (Jensen 1996) can be used as a tool for performing probabilistic inference which can be thought of as a message (e.g. the available state of a certain variable) passing process in the BN. The theorem is shown in Equation (2).

$$P(x_i \mid x_j) = \frac{P(x_i, x_j)}{P(x_j)} = \frac{P(x_j \mid x_i) P(x_i)}{P(x_j)}, \tag{2}$$

where $P(x_i \mid x_j)$ is posterior probability distribution of x_i given x_j, $P(x_j \mid x_i)$ is prediction term for x_j given x_i, and $P(x_i)$ and $P(x_j)$ are the prior probabilities of x_i and x_j, respectively. A value that is assigned to a variable will be propagated through the network and will update the marginal posterior distributions of other nodes as explained above.

3.2 Determining variables

Before assessing the graph topology of a BN and the parameters of the *CPtable* explained in the BN, different variables represented by the nodes of the BN ought to be decided according to the characteristics of the bus routes in the research area and the availability of the data for the BN analysis. In this study, the revenue per unit distance (*RUD*) is used as the query variable to measure the operating quality of bus routes. The evaluation calculation is interpreted by Equation (3).

$$RUD = \frac{R_i}{LEN_i}, \tag{3}$$

where R_i denotes bus service revenues of route i, unit: Yuan RMB, and LEN_i represents the length of route i, unit: km.

Besides the *RUD*, the nodes of the BN developed in this research also represent fitted out vehicles (*FOV*), average daily passengers (*ADP*), traveling speed (*TRS*), average load factor (*ALF*), length (*LEN*), nonlinear coefficient (*NOC*), and average station spacing (*ASS*) for the study area. The *FOV* are the fitted out vehicles of route i, unit: vehicles, the *ADP* represents the average daily passengers of route i, unit: passengers, and the *ASS* denotes the average station spacing of route i, unit: m. Other variables are explained by Equation (4), Equation (5) and Equation (6), respectively.

$$TRS = \frac{LEN_i}{TT_i}, \tag{4}$$

$$ALF = \frac{Q_{id}}{Q_i \times run_i}, \tag{5}$$

$$NOC = \frac{LEN_i}{SL_i}, \tag{6}$$

where LEN_i represents the length of route i, unit: km, TT_i represents the entire traveling time of route i, unit: h, Q_{id} denotes the average daily passenger transport volumes of route i, unit: passengers, Q_i denotes the rated passenger transport volumes of a bus of route i, unit: passengers, run_i represents the daily schedules of route i, unit: trips, and SL_i indicates the linear distance of route i, unit: km.

In this research, the values of all the variables represented by the nodes of the proposed BN for X are calculated from January to May of 2016 according to the administrative statistics of the government department. The criteria of various discrete states of different variables are listed in Table 1. The discrete thresholds of the *RUD*, the *ADP* and the *FOV* are based on the respective average values of the historical data. According to Sun *et al*. (2003), 22.00km/h and 30.00km/h are selected as the discrete thresholds of the *TRS*. The discrete thresholds of the *ALF* are 0.30, 0.60 and 1.20 (Kuang 2005). According to CBIP (1995), the discrete thresholds of the *NOC* are 1.10, 1.40 and 2.00 and of ASS are 500.00m and 800.00m. The discrete thresholds of the *LEN* are 36.23km and 27.17km (Zhang and Yan 2007).

Table 1. The criteria of various states of different variables.

Variable	States
RUD	Low, [0, 575.00]; Medium, (575.00, 1000.00]; High, (1000.00, +∞)
FOV (vehicles)	Low, [0, 8]; High, (8, +∞);
ADP (passengers)	Small [0, 2125]; Medium (2125, 3718]; Large (3718, +∞)
TRS (km/h)	Low [0, 22.00]; Reasonable (22.00, 30.00]; High (30.00, +∞)
ALF	Small [0, 0.30]; Medium (0.30, 0.60]; Large (0.60, 1.20]
LEN (km)	Short [0, 27.17]; Reasonable (27.17, 36.23]; Long (36.23, +∞)
NOC	Reasonable [1.10, 1.40]; Large (1.40, 2.00]; Overlarge (2.00, +∞)
ASS (m)	Reasonable [500.00, 800.00]; Unreasonable, otherwise

3.3 Estimating graph topology and parameters

After determining all the variables represented by the nodes of a BN, the structure of the BN (i.e. the dependencies between different variables) and the parameters of the *CPtable* interpreted in the BN (i.e. the strengths of various dependencies as encoded by the entries in *CPtable*) can be estimated according to specific issues. A common and simple approach to the BN structure learning is to rank graph structures via a search-and-score method that measures how well each model structure fits the data to find the global optimization BN structure (Silander and Myllymaki 2006). Nevertheless, the search-and-score method requires prior information about an ordering of the nodes to reduce the search space (Heckerman *et al*. 1995; Wang 2010). Unfortunately, this prior information about the node ordering is not always given in advance. In order to make accurately efficient BN learning based on training data without adequate prior information, a directional dependence analysis (DDA) algorithm based on the dependency analysis of variables is proposed to learn a BN structure (Wang 2010). The calculation steps of the DDA algorithm including the sequential steps of building undirected BN, orienting edges and thinning BN are as follows.

(1) Building undirected BN. Make a list L to display every pair of distinct nodes (X,Y) with sufficient mutual information $I(X,Y)$ calculated by Equation (7) over a threshold value θ. Sort (X,Y) in L according to the order of $I(X,Y)$ from large to small. Connect a pair of nodes by an edge. Remove the edge e_{XY} with the maximal $I(X,Y)$ from L to list E and then remove the edge e_{XZ} including one of the nodes of the edge e_{XY} from L to E with the conditional mutual information $I(Y,Z \mid X)$ computed by Equation (8) over a threshold value θ. Iteratively examine the edges in the L.

$$I(X,Y) = \sum_{x,y} P(x,y)\ log\ \frac{P(x,y)}{P(x)P(y)} \tag{7}$$

$$I(Y,Z|X) = \sum_{x,y,z} P(y,z,x)\ log\ \frac{P(y,z \mid x)}{P(y \mid x)P(z \mid x)} \tag{8}$$

Add an edge between a pair of nodes belonging to (X,Y) remaining in L when they cannot be separated by a set of relevant conditional independence tests based on their conditional mutual information $I(X,Y \mid C)$ with respect to the set of 'evidence' variables (i.e. condition-set) C.

(2) Orienting edges. After building the undirected BN, the direction of an undirected edge e_{XY} is oriented using conditional relative ability to predict (CRAP) calculated in Equation (9) given the union of domain set $Z = y(X) \cup y(Y)$, where $y(X)$ is the domain of X, $y(Y)$ is the domain of Y, x, y and z denote an instantiation of X, Y and Z, respectively.

$$\text{CRAP}(Z, X \rightarrow Y) = \frac{\sum_{z} P(z) max[P(x, y \mid z)]}{\sum_{z} P(z) max[P(x \mid z)]} - 1 \tag{9}$$

According to causal semantic theory (Cheng *et al*. 2002), if

$$\text{CRAP}(Z, X \rightarrow Y)\text{-}\text{CRAP}(Z, X \leftarrow Y) > \lambda \quad (\lambda > 0),$$

and no loops exists in causal direction $X \rightarrow Y$, X is chosen as a parent node to Y. Similarly, if

$$\text{CRAP}(Z, X \leftarrow Y)\text{-}\text{CRAP}(Z, X \rightarrow Y) > \lambda \quad (\lambda > 0),$$

and no loops exists in causal direction $X \leftarrow Y$, Y is chosen as a parent node to X. The proposed method based on causal semantic theory will not be able to orient all the edges in a network because of the week causal relationship. In this research, Minimum Description Length (MDL) principle (Suzuki 1993, 1996) is used to deal with the week causal semantic of arc $X — Y$.

(3) Thinning BN. Find a minimal cut-set $M(X,Y)$ separating X from Y in the BN graph using heuristic approaches (Cheng *et al.* 2002). Remove an edge $X \rightarrow Y$ ($X \leftarrow Y$) if the $I(X,Y \mid M(X,Y))$ is less than a certain threshold value. Iteratively examine the edges in the BN graph.

Received by virtue of applying the DDA algorithm with entering the statistical data of the variables represented by the nodes of the BN in this research, the structure of the BN newly developed for the estimation of bus service revenues of X is shown in Figure 2. In this study, the mean absolute percentage error (MAPE) is performed to validate the accuracy of the developed BN model by calculating the average percentage difference between predicted values and observed ones given validation data (Zong *et al*. 2013). The result of the MAPE of each variable in the BN is almost less than 0.10 which demonstrates the good robustness of the BN model.

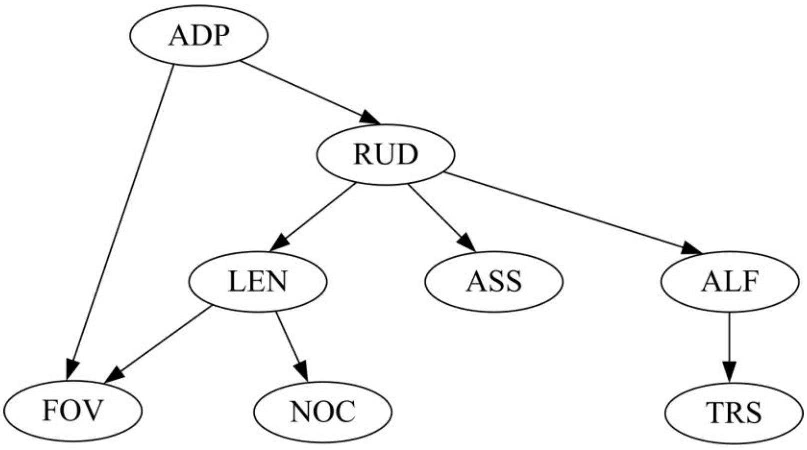

Figure 2. The structure of the BN developed for the study area.

3.4 Probabilistic inference

Given the structure of the BN and the parameters of the *CPtable*, inference queries can be evaluated by making full use of conditional independent information between different variables. Predictive inference (also called top-down reasoning) and diagnostic inference (also called bottom-up reasoning) are considered as two main types of probabilistic inference. They are based on the evidence nodes connected to the queried node through its parent and children nodes, respectively. In consideration of the predictive property of the probabilistic inference in this study, the Clique tree (CT) (Lauritzen and Spiegelhalter 1988; Pearl 2014), which is one of the major approaches to inference in multiply connected BN, is applied for the top-down reasoning work in this study because of its accuracy inference computation and high computational efficiency.

The CT is an undirected tree. Each tree node consists of a set of nodes from the BN which is called Clique. The CT takes the mechanism of information propagation with the steps of distributing evidence and collecting evidence. According to the difference of the information propagation programmes, the CT method can be further categorised into the Lauritzen–Spiegelhalter method, the HUGIN method and the Shenoy–Shafer method. The HUGIN algorithm (Andersen *et al*. 1989) is employed in this study because of its computationally efficient. The process of the HUGIN algorithm, which includes the sequential steps of clustering and propagation, is shown in Figure 3.

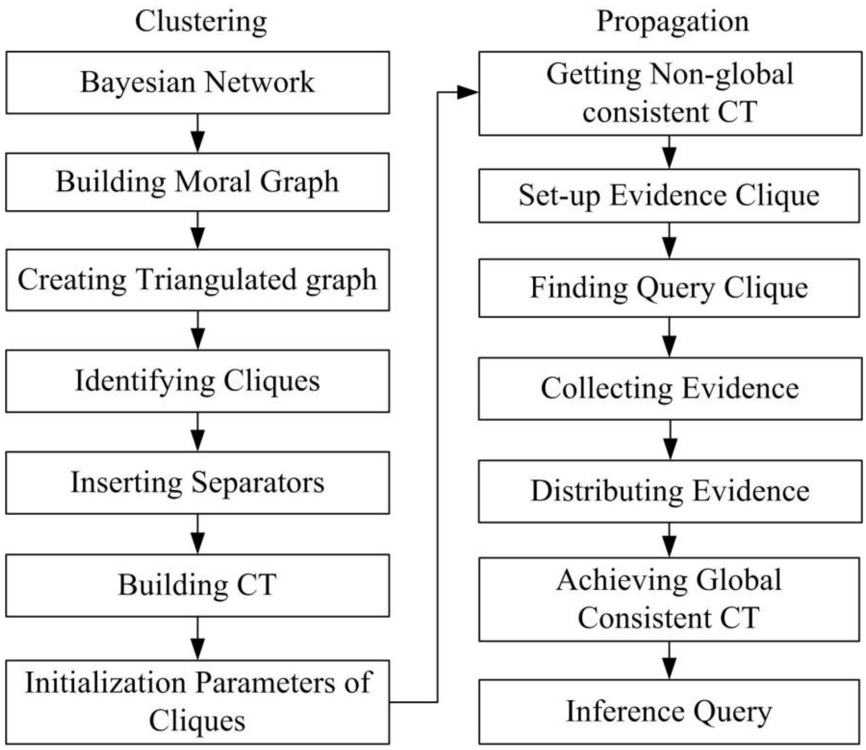

Figure 3. The process of CT.

(1) Clustering. First, an initial moral graph G_M is constructed by making an undirected copy of the BN and then augmenting it as follows. Let *X* systematically ranges over all nodes in G_M. For each node *X*, HUGIN adds to G_M an edge between each pair of nodes in parent nodes of *X* if no such edge already exists in G_M. Second, HUGIN triangulates the moral graph G_M creating a triangulated graph G'_M. The achieved triangulated graphs are proven to be the Cliques by Kjærulff (1990). Third, a CT is created from the Cliques by inserting separators. Last, the *CPtable* of the BN translate into potential functions of the CT by initialization.

(2) Propagation. The evidence information is introduced into potential functions of the CT. Find a query Clique C_Q which includes the query variable (i.e. the revenue per unit distance) as the root cluster. In the collecting evidence step, information is propagated from the furthest Clique to the query Clique C_Q until all the information reaches C_Q. On the contrary, in the distributing evidence step, information is propagated from the query Clique C_Q to the furthest Clique until all the information reaches each Clique. In this way, the CT built on the basis of the BN achieves globally consistent. Thereafter, the posterior probability distributions of the variables represented by the nodes of the BN are available from the prior *CPtable* explained in each Clique of the CT. Then the conditional probability of the query variable X_Q given the values of the evidence variables X_E in the BN is equal to Equation (10).

$$P(X_Q|X_E) = \frac{P(x_E, x_Q)}{\sum_{x_Q} P(x_E, x_Q)}, \tag{10}$$

where $\sum_{x_Q} P(x_E, x_Q)$ is the marginal probability of X_E.

4 Results and Discussions

Three types of route services in the study area X are considered as examples here by the inference queries. The dotted line, the solid line and the dot chain line shown in Figure 4 stand for route 26, route 28 and route 36, respectively. The red nodes represent the location of terminal stations of bus routes in X. In this research, the locations of terminal stations in X are assumed to be fixed. The proposed BN enables us to achieve the maximum predicted probabilities of high bus service revenues corresponding to the certain states of the variables of route 26, route 28 and route 36, respectively. According to these states, the optimized layouts of the three routes can be designed, shown in Figure 5. The dotted line, the solid line and the dot chain line stand for the optimized layouts of route 26, route 28 and route 36, respectively. The red nodes represent the location of terminal stations of bus routes in X.

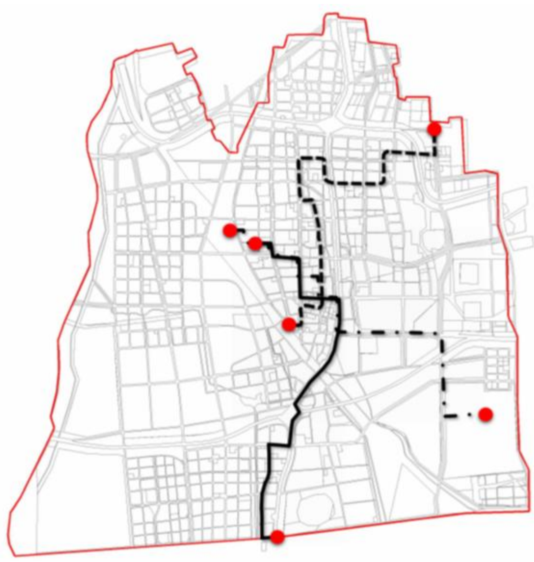

Figure 4. The existing layouts of route 26, route 28 and route 36 in X.

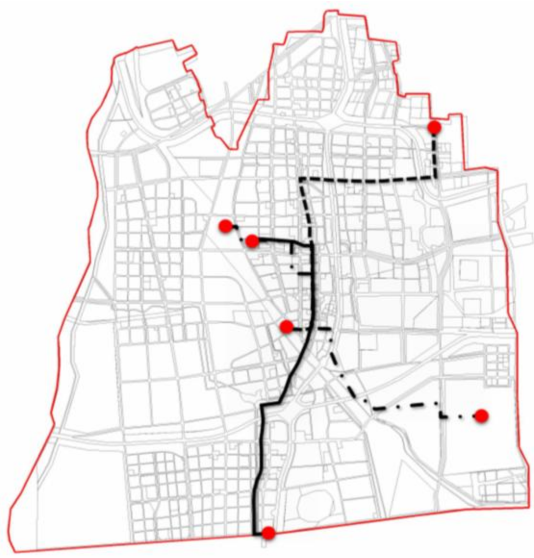

Figure 5. The optimized layouts of route 26, route 28 and route 36 in X.

Table 2 provides the maximum predicted probabilities of high bus service revenues and the values of variables for the existing routes and the optimized routes. The *LEN*, the *NOC* and the *ASS* are calculated directly by the configuration of the optimized routes. The *ADP* is obtained according to the split-flow of passenger transport demands of the existing bus routes which partly overlap the optimized routes. Then the *FOV* is given based on the passenger transport demands of the optimized routes. The *ALF* is calculated by simultaneously considering the passenger demands and the fitted out vehicles. The *TRS* is calculated by means of the traveling speed of the existing bus routes which partly overlap the optimized routes.

Table 2. Maximum probabilities of bus service revenues.

Variables	Route 26		Route 28		Route 36	
	Existing	Optimized	Existing	Optimized	Existing	Optimized
	0	96.32%	0	99.86%	13.22%	98.89%
FOV (vehicles)	7	15	8	16	5	10
ADP (passengers)	924	3853	1560	4027	842	2698
TRS (km/h)	28.40	28.40	30.00	24.17	17.16	23.00
ALF	0.50	0.42	0.65	0.65	0.17	0.65
LEN (km)	35.50	33.30	45.00	42.90	14.30	10.70
NOC	1.68	1.58	1.46	1.39	2.07	1.55
ASS (m)	845.20	740.00	1022.70	794.44	621.70	509.52

The passenger demands of the west lines of route 26 are influenced by split-flow effect, and the trend of the east lines of route 26 is inconsistent with the primary passenger flow. The comparatively low daily passenger transport volumes and the unreasonable layouts of route 26 make the probability of the existing high bus service revenues of route 26 is 0. Route 26 is adjusted according to select the roads with the comparatively few bus routes and the stops with relatively large demands in order to increase the passenger transport volumes, meanwhile, increase the number of operating vehicles to satisfy the increasing amount of passenger demands. As a result, the revenues of optimized route 26 are the most probable high revenues.

Route 28 is mainly along the expressway which concentrates more than 40 bus routes. The concentrated routes split the passenger transport volumes of route 28. Moreover, Route 28 is a relatively long and circuitous route. As a result, the probability of the existing high bus service revenues of route 28 is 0. In order to achieve the highest revenues, route 28 is optimized by selecting the roads with the comparatively few bus routes and the stops with relatively large demands to ensure larger passenger demands, meanwhile, slightly improving route directness to ensure lower nonlinear coefficient, and increasing the fitted out vehicles to satisfy the increasing amount of passenger demands and control the average load factor in a reasonable range.

The concentrated routes cause low passenger demands of route 36 by split-flow effect and make the traveling speed excessive low. Moreover, the route is more detour which increases bus operation cost. The unreasonable layout of route 26 leads the probability of the existing high bus service revenues of route 36 is 13.22%. Route 36 is modified by selecting the roads with the comparatively few bus routes to increase the passenger demands and improve traveling speed, meanwhile, increasing the number of operating vehicles to satisfy the increasing amount of passenger demands and control the average load factor in a reasonable range. Moreover, eliminating the circuitous alignment is going to ensure lower nonlinear coefficient. Then this modification can be successfully provided the most probable high revenues.

5 Conclusions

By using the urbanized region of city in China as the study area, a BN approach has been newly developed and applied to forecast the probabilities of different revenues of various route services, from the perspectives of both satisfying passenger demand and decreasing bus operation cost. In this way, the existing bus routes in the studied area can be optimized for their most probable high revenues. In the future, combining other methodologies with the BN is worthy of the exploration for more rational estimation of the bus operating revenues. And extending the proposed solution methodology to solve other transport route design problems can be another future research direction.

Competing Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research is supported by National Natural Science Foundation of China (71571011) and Fundamental Research Funds for the Central Universities (2016JBM028; 2017YJS110).

References

Anderson, R. D., Mackoy, R. D., Thompson, V. B., and Harrell, G. (2004). A Bayesian Network Estimation of the Service-Profit Chain for Transport Service Satisfaction. *Decision Sciences*, 35(4), 665-689.

Andersen, S. K., Olesen, K. G., Jensen, F. V., and Jensen, F. (1989, August). HUGIN-A Shell for Building Bayesian Belief Universes for Expert Systems. *In IJCAI* (pp. 1080-1085).

CBIP (China Building Industry Press). (1995). *Urban public transport specification*. Beijing: China Building Industry Press.

Ceder, A. A., Butcher, M., and Wang, L. (2015). Optimization of bus stop placement for routes on uneven topography. *Transportation Research Part B: Methodological*, 74, 40-61.

Ceder, A., and Israeli, Y. (1998). User and operator perspectives in transit network design. *Transportation Research Record: Journal of the Transportation Research Board*, (1623), 3-7.

Cheng, J., R. Greiner, J. Kelly, D. Bell, and W. Liu. (2002). "Learning Bayesian Networks from Data: An Information-theory Based Approach." *Artificial Intelligence* 137 (1-2): 43–90.

Feng, S. M., and Chen, H. R. (2006). The research of computational method of bus configurations. *Transportation Systems Engineering and Information Technology*, 6(3), 79-81.

Fan, W., and Machemehl, R. B. (2008). Tabu search strategies for the public transportation network optimizations with variable transit demand. *Computer-Aided Civil and Infrastructure Engineering*, 23(7), 502-520.

Feng, X., Saito, M., and Liu, Y. (2016). Improve urban passenger transport management by rationally forecasting traffic congestion probability. *International Journal of Production Research*, 54(12), 3465-3474.

Heckerman, D., Geiger, D., and Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20(3), 197-243.

Jensen, F. V. (1996). *An introduction to Bayesian networks* (Vol. 210, pp. 1-178). London: UCL press.

Jensen, F. V. (2001). Bayesian Networks and Decision Graphs. Series for Statistics for Engineering and Information Science.

Jensen, F. V., Olesen, K. G., and Andersen, S. K. (1990). An algebra of bayesian belief universes for knowledge-based systems. *Networks*, 20(5), 637-659.

Kjærulff, U. (1990). Triangulation of graphs--algorithms giving small total state space.

Kuang, X. (2005). *The evaluation research of urban public transport level of services*. Changchun: Jilin University.

Lauritzen, S. L., and D. J. Spiegelhalter. (1988). "Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems." *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2), 157-224.

Lee, Y. J., and Vuchic, V. R. (2005). Transit network design with variable demand. *Journal of Transportation Engineering*, 131(1), 1-10.

Nayeem, M. A., Rahman, M. K., and Rahman, M. S. (2014). Transit network design by genetic algorithm with elitism. *Transportation Research Part C: Emerging Technologies*, 46, 30-45.

Ozbay, K., and Noyan, N. (2006). Estimation of incident clearance times using Bayesian Networks approach. *Accident Analysis & Prevention*, 38(3), 542-555.

Pacheco, J., Alvarez, A., Casado, S., and González-Velarde, J. L. (2009). A tabu search approach to an urban transport problem in northern Spain. *Computers & Operations Research*, 36(3), 967-979.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.

Pearl, J. (2014). Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann.

Silander, T., and Myllymaki, P. (2006). A simple approach for finding the globally optimal Bayesian network structure. In *Proceedings of the Twenty-second Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, Dechter, R. and Richardson, T. (eds). AUAI Press, 445–452

Suzuki, J. (1993, July). A construction of Bayesian networks from databases based on an MDL principle. In *Proceedings of the Ninth international conference on Uncertainty in artificial intelligence* (pp. 266-273). Morgan Kaufmann Publishers Inc..

Suzuki, J. (1996). Learning Bayesian belief networks based on the MDL principle: an efficient algorithm using the branch and bound technique.

Sun, J., and Sun, J. (2015). A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data. *Transportation Research Part C: Emerging Technologies*, 54, 176-186.

Sun, F. C., Wang, Z. P., and Wang, J. (2003). Statistic analysis of the bus average speed of Beijing. *Automotive Engineering*, 25(3): 219-222.

Szeto, W. Y., and Wu, Y. (2011). A simultaneous bus route design and frequency setting problem for Tin Shui Wai, Hong Kong. *European Journal of Operational Research*, 209(2), 141-155.

Tirachini, A., Hensher, D. A., and Rose, J. M. (2014). Multimodal pricing and optimal design of urban public transport: The interplay between traffic congestion and bus crowding. *Transportation Research Part B: Methodological*, 61, 33-54.

Wirasinghe, S. C., and Vandebona, U. (2011). Route layout analysis for express buses. *Transportation Research Part C: Emerging Technologies*, 19(2), 374-385.

Zhang, R., and Yan, H. (2007). *The theory and practice of urban public transport planning*. Beijing: China Railway Publishing House.

Zhao, F., and Zeng, X. (2006). Simulated annealing–genetic algorithm for transit network optimization. *Journal of Computing in Civil Engineering*, 20(1), 57-68.

Zong, F., Xu, H., and Zhang, H. (2013). Prediction for traffic accident severity: comparing the Bayesian network and regression models. *Mathematical Problems in Engineering*, 2013.