Bayesian Algorithm for The Construction of Logistics Node Delay Model and Its Impact on Subsequent Nodes in Supply Chain

Abstract: The logistics nodes delay of supply chain has become a focus of customer complaints at the State Post Bureau of China in recent years. Nodes delay usually also brings uncertain business loss. The purpose of this paper is to explore propagation of nodes delay and make it precisely controllable. To address this, Netica software is used and a hybrid strategy of genetic algorithm (GA) and tabu search algorithm (TS) validated by datasets is utilized to optimize the Bayesian algorithm model. Bayesian algorithm model of logistics nodes delay of supply chain is constructed through parameter learning. The empirical analysis is conducted on the basis of three types of nodes delay with 2041 sets of data for cargo departures and arrivals, upstream nodes delay, and nodes delay of supply chain. The results show that the nodes delay of supply chain gradually decreases from upstream to downstream; there are strong correlations between propagation probability and transfer time of nodes delay; the effects of initial node delay are the maximum, after 3 transitions, still having 30% probability to affect the succeeding nodes. Therefore, some suggestions, such as controlling the transfer points and transfer time and strengthening the management of initial node delay, are given.

Keywords: Nodes Delay, Delay Propagation, Genetic Algorithm, Tabu Search, Bayesian Network

I. INTRODUCTION

Timeliness is a critical issue throughout the whole process of supply chain operation. Good cooperation of supply chain nodes always plays a key role in the successful operation. Although enterprises have deliberate plans and information methods to deal with logistics tracking and controlling, some problems of cooperation of supply chain nodes may still occur, and thereby causing nodes delay by reasons of various contingencies, for example, typically nodes connections during transportation process of logistics, time wasting in transfer and delays caused by emergencies. The problem of logistics nodes delay has become a common fault of the operations and focus of customer complaint at the State Post Bureau of China in recent years (China International E-Commerce Network, 2020) [1]. As logistics delays have become severe challenges to the enterprises’ established operation plans and having significant effects on the subsequent management of supply chain nodes, it is meaningful to analyze and explore the rules of nodes delay in scientific way, and of practical significance in case that the results applied to precise control.

II. LITERATURE REVIEW

A. The Issue of Delay about Timeliness.

The research regarding this issue is mainly fourfold. One is the influences of timeliness of logistics supply chain. Hong and Liu [2] pointed out that logistics companies should value the logistics timeliness and engaging to reduce the logistics delay, in this case, the economic benefits can be improved and the cost lowered. Erceg and Sekuloška [3] claimed that the delivery time control is crucial for companies to achieve competitiveness, regardless of the traditional or e-commerce logistics. Zhang et al. [4] found that for fresh logistics distribution, especially for the cold-chain, the logistics time could directly affect the product quality. As cold-chain products have a high requirement for the delivery time, it is suggested that logistics companies should think over how to ensure on-time delivery.

Another is the delay effect on consumer decision making. Fan and Ye [5] concluded that the possibility of customer buying certain items online is affected by the timeliness of e-commerce logistics. Wei [6] found there is a significant difference between the real delivery time of online shopping and the expected delivery time, which implies that consumers may place a high value on logistics time when making purchase decisions. Besides, the ideal delivery time is usually less than the real delivery time. According to Lin [7], as customers pay increasing attention to logistics time, ensuring the timeliness of logistics delivery can effectively improve the recognition of customers to the entire logistics process. Wu and Yan [8] believed that customer satisfaction has strong

1 GuangZhou College of Technology and Business, Foshan, Guangdong, China
*Corresponding author: Pengliang Qiao
Copyright © JES 2024 on-line : journal.esrgroups.org
relationship with speed of logistics delivery. In further, customer satisfaction is the key factor that affects e-commerce enterprises and logistics industry. Li [9] found customers’ evaluations of e-commerce logistics are influenced by delivery time, the longer the delivery time, the more negative reviews and feedbacks the customers would give.

A third is about the real logistics time and the planned logistics time. Li et al. [10] concluded that customer keeps an ideal delivery time in mind, yet the real delivery time can deviate from the ideal planned time, that is to say, almost all the real delivery time is more than the ideal planned, the larger deviation between the real delivery time and the ideal planned, the lower rating the customer would give to the quality of logistics. Nan and Liu [11] contended that customers are concerned with not only the delivery time, but also the entire waiting time, that is the time from order placement to receipt, compared to the delivery time, which as well includes the e-commerce companies and couriers dispatching time. Wang et al. [12] held the viewpoint that more emphases should be put on the process of product preparation phase, so as to fulfill the order with a minimum time, and optimize the coordination of business units, and minimize order setup and delivery time to the largest degree.

The fourth is about the delay effects on other logistics parts. Liang [13] concluded that the delivery speed is closely correlated with the location of distribution center. Yu [14] confirmed that customers are concerned with not only the entire logistics time, but also the partial logistics time, especially the time of delivery after receiving order. For the sake of providing good online consumer experiences, it requires couriers to upgrade their comprehensive abilities and service level, especially for the last one-kilometer delivery. Chen et al. [15] suggested that after online shopping, instead of redirecting to the logistics company websites, consumers could directly check and track their packages from the website they shop, where already included is function of logistics information checking. Compared to the traditional way of package tracking and logistics information checking from the logistics company websites, this can provide a better customer experience on package tracking and attaining to higher rating for both the delivery time and service.

B. Delay Taken as Main Research Objective.

First, modelling delay and corresponding treatments. (1) Analysis of delay causes: Chen and Schonfeld [16] built a dispatching model to analyze the delay causes and delay propagation, and providing corresponding measures to for controlling delay. Song and Rong [17] summarized and analyzed different types of uncertain events and factors that can possibly cause delay. Geng et al. [18] found that many factors could cause air freight delay but not many treatments countering such delay. The main methods to prevent delay, based on flight path optimization, has long been the most effective and with the lowest cost. Yao [19] conducted the backward reasoning of the delay propagation and relevant phenomenon, and in further explored the corresponding causes, according to which the improvement plans are proposed by relevant logistics companies. Zhu et al. [20] proposed a double objective network flow robust optimization model for integrated aircraft scheduling so as to minimize the delay propagation and reduce the airline operational cost caused by delayed flight. (2) Optimization of delay-related model: for minimizing the delay time of logistics companies, Potvin et al. [21] built an objective function with respects to a weighted summation of three variables, i) travel time, ii) sum of lateness at customer locations, and iii) lateness at the depot. Chen et al. [22] developed three optimization models to analyze the issues related to intermodal logistics network delays.

Second, the effects caused by delay. Grout [23] found that exactly 100 per cent on-time delivery is non-optimal and only feasible under specific conditions. Ahmad [24] found that delivery delay could directly affect customer experience of online shopping, so logistics companies should take account of controlling and reducing delay so as to get better customer comments. Marimon and Vidgen et al. [25] found that the probability of a user's second purchase is closely related to whether the logistics is delayed. Chen et al. [26] claimed that customer evaluation on logistics quality is greatly determined by logistics delivery time, the longer delivery time, the worse reviews left by customers. Guo and Tan [27] believe that air transportation is an important mode of transportation for express logistics companies, and may affect consumers’ online shopping experience and evaluation indicators.

Third, dealing with delay. Logistics delay can be caused by many factors, such as mechanical failure, accident and traffic congestion [28]. To cope with delay, numerous measures are proposed and integrated to analyze the relevant factors, and hence are employed to minimize the delay. Hu et al. [29] analyzed the delivery delay and defined key factors with the stage division method of disruption management. As the causes of delivery delay are various and complex, it might be better to separately carry on causal analysis of delayed delivery rather than in integration. Traffic is a critical factor that can cause delay, thereby different means of transport may be considered to cope with varied traffic situations [30].
C. Literature Summary

From the above, plenty of research has been done on issues of delay, and the main research is concentrated on the timeliness of supply chain, the delay effects and optimization of delay model. Most of current research on delay are about air express delay, which has yielded abundant research results, however, not a lot of studies are about common logistics delay in the field. Besides, as there has been no unified model concerned with supply chain delay, there is a gap for the future research. In addition, many of current studies focus only on model per se, but lack of optimizing model on the basis of data for logistics operations and relevant empirical analysis. Therefore, this research is to build a novel logistics delay model on the basis of Bayesian Network and conduct empirical analysis with collected data, and able to provide a reference for the future research on dealing with delay related issues in the logistics field.

III. RESEARCH METHODOLOGY

A. Basic Model: Bayesian Network

The correlations between logistics nodes in supply chain have apparent sequential order, and there may apparently exist cause-and-effect in the issue of nodes delay, yet the nodes delay propagation also belongs to probability event. The objective of this research is to analyze the historical upstream nodes delay to predict the effects of delay on nodes, and thereby coming up with corresponding strategy for controlling the delay. Considering the correlations of nodes, and the main objective of our research, this paper built nodes delay network model on the basis of Bayesian Network. In general, there are three steps for building Bayesian Network, and with repeating the three steps, a final validated and revised Bayesian Model can be obtained.

B. Optimized Algorithm: Hybrid Strategy of Genetic Algorithm (GA) and Tabu Search (TS)

The learning outcome of Bayesian Network is to attain the optimal structure and match it with the actual data. The implementation process can be divided into two parts: model selection and optimization. The simplest way to optimize the model is to select the one with highest score through observing each enumerated structure. This method, however, is not easy to directly implement in practice, so we combine it with Tabu Search Algorithm. Glover et al. [31] combined the genetic algorithm with tabu search to improve the effectiveness of the outcome. This paper is to attain the crossover operators and mutation operators of genetic algorithm with tabu search, producing a new hybrid strategy of genetic algorithm and tabu search.

Validating hybrid strategy: this paper, employing the classic data set, the Iris flower, contact lenses and weather and corresponding indicators such as simulation time, mean deviation, data error, Kappa statistic and accuracy, made a comparison of K2 algorithm (K2), Hill climbing algorithm (HC), Genetic Algorithm (GA) and Hybrid strategy of Genetic Algorithm and Tabu Search (GATS). The results shown as Table 1:

<table>
<thead>
<tr>
<th>Data set</th>
<th>Algorithm</th>
<th>Simulation time (sec.)</th>
<th>MAD</th>
<th>RMSD</th>
<th>RAE (%)</th>
<th>RRSE (%)</th>
<th>Kappa</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris Flower</td>
<td>K2</td>
<td>0.02</td>
<td>0.04</td>
<td>0.16</td>
<td>6.42</td>
<td>29.17</td>
<td>0.8784</td>
<td>90.72</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.02</td>
<td>0.04</td>
<td>0.17</td>
<td>6.64</td>
<td>30.69</td>
<td>0.8684</td>
<td>90.05</td>
</tr>
<tr>
<td></td>
<td>GA</td>
<td>0.13</td>
<td>0.04</td>
<td>0.14</td>
<td>6.21</td>
<td>24.72</td>
<td>0.8984</td>
<td>92.05</td>
</tr>
<tr>
<td></td>
<td>GATS</td>
<td>0.27</td>
<td>0.03</td>
<td>0.14</td>
<td>4.07</td>
<td>24.12</td>
<td>0.9184</td>
<td>93.38</td>
</tr>
<tr>
<td>Contact Lenses</td>
<td>K2</td>
<td>0.01</td>
<td>0.22</td>
<td>0.31</td>
<td>56.62</td>
<td>66.08</td>
<td>0.4265</td>
<td>68.88</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.01</td>
<td>0.20</td>
<td>0.29</td>
<td>50.70</td>
<td>61.96</td>
<td>0.6984</td>
<td>81.38</td>
</tr>
<tr>
<td></td>
<td>GA</td>
<td>0.10</td>
<td>0.20</td>
<td>0.29</td>
<td>51.70</td>
<td>61.96</td>
<td>0.6984</td>
<td>81.38</td>
</tr>
<tr>
<td></td>
<td>GATS</td>
<td>0.10</td>
<td>0.20</td>
<td>0.28</td>
<td>49.94</td>
<td>61.00</td>
<td>0.6984</td>
<td>81.38</td>
</tr>
<tr>
<td>Weather</td>
<td>K2</td>
<td>0.01</td>
<td>0.46</td>
<td>0.50</td>
<td>93.23</td>
<td>95.46</td>
<td>-0.036</td>
<td>55.19</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.01</td>
<td>0.51</td>
<td>0.54</td>
<td>103.56</td>
<td>104.31</td>
<td>-0.3775</td>
<td>40.91</td>
</tr>
<tr>
<td></td>
<td>GA</td>
<td>0.13</td>
<td>0.39</td>
<td>0.43</td>
<td>78.36</td>
<td>82.90</td>
<td>0.1744</td>
<td>62.33</td>
</tr>
<tr>
<td></td>
<td>GATS</td>
<td>0.04</td>
<td>0.31</td>
<td>0.34</td>
<td>64.01</td>
<td>64.52</td>
<td>0.5049</td>
<td>76.62</td>
</tr>
</tbody>
</table>

From the above, the simulation time of hybrid strategy (GATS) fall into the middle, but its accuracy is the highest, thereby guaranteeing the stability. In this regard, this paper also uses the hybrid strategy of genetic algorithms (GA) and tabu search (TS) to optimize the Bayesian Network.
IV. MODEL DESCRIPTION

A. Data Description

depdelay: the delayed time the ith logistics node departs, \( i = 1, 2, 3 \).
arrdelay: the delayed time the ith logistics node arrives, \( i = 1, 2, 3 \).

From \( t_1 \) to \( t_2 \): the delayed time of logistics is \( t_1 \) to \( t_2 \) minutes.

Less than & more than n: the delayed time less than 1 minute while the real delivery time is more than the guaranteed time. Considering the delivery time guaranteed by different transportation companies may be varied, we collected the logistics data from a representative company in China for study, that is, China Postal Express & Logistics.
difference: the difference between real transfer time and the planned time. \( i = 1, 2 \).

B. Modeling Process

1) Bits of parents set was adopted in coding: Original sample and new samples generated both use this method. The directed acyclic graph (DAG) that comprised of 8 nodes was applied to model the delay propagation. Each of nodes corresponds to a different parameter, the length of array set as 361, and subscripts should be included in 0–360.
2) Construction of initial group: Start with 10 random original sample sets and learning through network. Size = 10.
3) Score of bayesian network structure: BDe was adopted as scoring function. As the compossibility of this scoring function, the score of the whole network can be obtained by the superposition of partial network scores, and therefore the trend of the score of each node can also be regarded as the micro expression of the whole trend.
4) Assessing fitness: Judge the validity of Bayesian network structure through the results of scoring function. The higher score, the better fitness.
5) Crossover operation of tabu search: Initial solution: crossover operation is conducted between two individual samples, i.e. sample [i1] and Sample [i2], generating new sample data. Tabu list: the length of array is 3. Ti: the tabu length is set as 3. Tabu object: Let i be the mark of storage, if a certain individual sample [i] meets the tabu rule, then the sample corresponding to i is processed by tabu. Evaluation function: scoring function of Bayesian network structure. BSF: the highest score is processed in certain way. Candidate solution: 3 optional individual samples. Termination condition: the number of iterations is 10.
6) Mutation operation of tabu search: Initial solution: select a random individual sample Sample [i] and its random location to complete the mutation. And store the generated results as initial solution in DesendantSample [j]. Ti: the tabu length is set as 3. Tabu object: mark the solution of mutation at certain location and put it into the tabu list, so as to avoid next mutation operation on the same location of the same sample. Evaluation function: scoring function of Bayesian network structure; BSF: the highest score is processed in certain way. Candidate solution: 3 optional samples; Termination condition: the number of iterations is 10.
7) Following crossover and mutation, the new individual sample generated.
8) Selection (copy) operations: The copy operations of tabu search can be accomplished by two algorithms, Imperialist Competitive Algorithm (ICA) and Best Fit Algorithm (BFA). ICA is employed in this paper.
9) Other settings of initialization parameter of gats algorithm: Crossover probability \( P_c = 0.85 \); Mutation probability = 0.45; Maximum number of iterations nRuns = 10.

C. Initial Model Building

1) Network model of logistics nodes delay: The initial logistics network model is shown in Figure 1. The directed arc goes from depdelay to arrdelay, indicating that in the same continuous logistics, the delay caused by the logistics node in transfer area leads to the delay of next node at arrival. Embodied in Bayesian network structure, arrdelay, and transfer time difference difference, constitute parent node of depdelay. There are three types of random variables in the model, depdelay, arrdelay, and difference. The departure delay is depdelay, as the value of parent node does not change regularly with the changes of parameter value of arrival delay, there is no correlation between the both.
2) **Parameter learning**: On accounting of data incompleteness, Expectation Maximization (EM) and Gradient Descent (GD) are adopted to complete parameter learning process. Three steps are as follows.

   a) **When units missing in data sample, process the missing item with maximum likelihood estimation (MLE) in calculation.**

   b) **Missing value taken as prior data, the approximation obtained through maximum likelihood estimation, is viewed as ideal value, that is, the ideal value can be considered as an approximation alternative of missing value. Although the ideal value is not exactly equal to, or to say, hardly equal to the missing value, the difference between the both is considered as eventually within ideal permission range, thereby reasonable.**

   c) **Use the ideal value to cover the missing item**: Repeat the above steps i to iii till the difference between the ideal value and missing value is standard, where the ideal value is considered as covering the missing value, and ensuring the integrity of the whole data.

V. **EMPIRICAL ANALYSIS**

A. **Data Collection**

Data are actual logistics data, which were collected from China Postal Express & Logistics between 26th June 2019 and 27th December 2020. Data were preprocessed, and the original data are from all around the country, we selected Guangxi Nanning as the transfer point and urban area of Liuzhou as the terminal. Each of data contains the arrival and departure time of the same place, delivery and receipt time of the package. 2135 sets of data in total were collected, where 94 sets were invalid, 2041 sets of data were valid which could be applied to the experiment.

B. **Basic Model**

Figure 2 is a topological structure generated from software. On the basis of this Bayesian structure, the parameter learning was completed. Figure 2 also shows the results of parameter learning after inputting training data, which reflect the delay propagation. In Figure 2, logistics delay starts from the first node to next several nodes, where the delayed time of arrival and departure shows a decreasing trend. For example, departure delay, arrival delay, less than 300 minutes respectively are depdelay, arrdelay, and lessthan300, the proportions of departure and arrival delays less than 300 minutes gradually increase from the 1st node to the 4th node, respectively from 5.91% to 59.9% and 7.08% to 72.9%, which in turn means that the delay propagation of the same goods decreases with the time and order of logistics nodes. The results of difference imply that the real time for the same goods arrival and departure at adjacent nodes more often is less than the planned.

C. **Different Situations of Logistics Delay**

1) **Cargo departure delay effects on its arrival delay at next node**: For this situation, we specifically analyze effects of departure delay of goods at node i on its arrival delay at node i+1. As is shown in Figure 2, for example, the continuous logistics departure delay between 600 and 900 minutes at the second node is 27.2%, whereas the arrival delay at the third node decreased to 23.3%; meanwhile the delay less than 300 minutes increased from 12.2% to 21.0%. Such is also the case for logistics delays of other nodes.

2) **Upstream-node arrival delay effects on the downstream nodes**: Figure 3 presents the delay effects on departure delay of the last node when the arrival delays of the 2nd node and 3rd node in continuous logistics are between 600 and 900 minutes. The results show that between 600 and 900 minutes, 100% of the arrival delay at the 2nd node and 3rd node decline to only 4.54% of arrival delay at the 4th node; and most of arrival delays at the 4th node, accounting for 68.4%, decreases to less than 300 minutes. 33.5% departure delay at the 1st node fall between 900 and 1200 minutes, however, through counteraction and reduction in transition, decreased to 2.75% at the 4th node. The departure delays at the 2nd node, through the counteraction and reduction, are controlled less than 300 minutes with 48.3% probability and between 300 and 600 minutes with 27.8%
probability. The results show that transfer time of logistics has a significant influence on delay propagation. Therefore, reducing real transfer time plays a key role in reduction and elimination of the delay of entire logistics.

Figure 2: Logistics Delay Propagation and Parameter Learning Results

Figure 3: Arrival Delay and Transfer Time Effects on Its Departure Delay

3) Initial node delay effects on subsequent delay propagation: In Figure 4, we consider the situation that 100% departure delay at the first node fall between 300 and 600 minutes, yet when logistics coming to the 4th node, turn to the delay less than 300 minutes with 82.3% probability. Nevertheless, 9.63% delay between 300 and 600 minutes and 8.07% delay over 600 minutes are still left, indicating that some new delays are created instead of being reduced. The transfer delays are increased when passing the 2nd and 3rd nodes, leading to the possibility that the succeeding nodes delay exceeds the initial delay. In Figure 4 and 5, we analyze the delay propagation of e-commerce logistics when all delays taken into account are between 900 and 1200 minutes, as shown in Figure 1. The network model of logistics nodes delay. Arriving at the 2nd node, 87.94% delays are over 900 minutes; after the first transfer, departing from the 2nd node, even though the delay is controlled, the probability of delay over 900 minutes are still more than 50%. After controlling delay in the 2nd transfer, still 11.81% delays left are over 900 minutes. And eventually, there are still 30% delays over 300 minutes arriving at 4th node.
VI. CONCLUSIONS

First, describing issues on nodes delay of supply chain, we built the relevant model on the basis of Bayesian network, using classic data sets to validate the hybrid strategy of genetic algorithm and tabu search algorithm, and collected real data to conduct empirical analysis. The results showed that Bayesian network model of logistics Nodes delay could exactly describe three situations of logistics nodes delay, where genetic algorithm and tabu search play a key role. It could provide a new insight and reference for modelling logistics delay.

Second, the empirical results showed that the effects of logistics nodes delay have the following rules: the nodes delay effects on subsequent nodes are gradually decreased. Once nodes delay occurs, the subsequent freight process tends to form weakened nodes delay, and even be back to the normal level. Based on this rules, if the freight planning at transfer nodes is qualified, the cause-and-effect of time and space of logistics nodes, that is “delay-propagation-delay weakened-back to normal”, are formed from node to node.

Third, the initial node delay had the most significant effect on the subsequent node. 1) Between the cargo departing at the node and arriving at the next node, the delay was mainly controlled through the transportation time, and that is why reducing transportation time can counteract and eliminate delay; 2) the delay between two adjacent nodes also depended on the transfer at nodes; 3) the initial node had 30% probability of delay to
influence the subsequent node after 3 transfers. From the above, the delay of initial node is the most significant and hence needs to be intensively controlled.

Fourth, empirical analyses were based on the model and data collection, behind the rules of nodes delay and propagation are the actual operations. To reduce the already occurred delay, always needed are speeding, working overtime and rescheduling, which, however, may lead to security risks and equipment overloading. In this case, this kinds of ways of reducing delay may not be sustainable in a long term. In the future research, more theoretical work may be developed with combining a broader data collection to analyze the cause of delay propagation and focusing on how to more effectively weaken the delay.

ACKNOWLEDGEMENT

This research was supported by the projects: Control of supply chain delay network system under large-scale COVID-19(Grant: 2021ZDJS122); China National Bureau of Statistics: Statistical monitoring, Dissemination Network and Impact Assessment of Manufacturing Supply Chain Node Delays under the impact of COVID-19(Grant:2022LZ); Research on dynamic mechanism, standard system and policy design of green supply chain under dual-carbon target(Grant:GD24CGL26);Low carbon system and GSCM driven by internal enterprise and its influence on enterprise performance(Grant:2023LS017C).

REFERENCES


