INTEGRATED LOGISTICS NETWORKS MODELING AND GENETIC ALGORITHM DEVELOPING

Yong PENG
Associate Professor
Traffic & Transportation School
Chongqing Jiaotong Univ.
Xuefu Road 66, Nan’an Distract, Chongqing
400074 China
Fax: +86-23-6265-2674
E-mail: pyepeng@hotmail.com

Abstract: This paper discusses how to integrate forward and reverse logistics to decrease the total cost of third party logistics service providers. A mathematic model is provided without the assumption of straight-line distance. Demand for product and return amount of product can vary randomly in this model. We also allow that the demand of customers can exceed the capacity of their suppliers. Then we develop a Genetic Algorithm to solve this NP-hard problem and discuss its convergence. At last a numerical example is available.

Keywords: integrated networks, forward and reverse logistics, genetic algorithm

1. INTRODUCTION

Third party logistics service providers (3PLs) are playing an increasing role especially in transportation and warehousing operations (Modern Material Handling, 2000). Meanwhile, increasing opportunities for cost savings or customer satisfaction from returned products prompt 3PLs to get involved in reverse logistics operations (Kirkke, 1999). In the past, the logistics networks have been independently designed with respect to either forward or reverse flows. Many mathematic models for forward logistics networks and reverse logistics networks respectively have been provided to balance the transportation cost and the storage cost, and to minimize the total cost (Andreas, 2005; Kokkinaki, 2000; Louwers, 1999). Hyun-Jeung KO (2005) provides an integrated solution on the logistics networks of 3PLs and the experimentation results show that selecting an integrated network approach definitely can be right decision for 3PLs when aiming to construct an efficient integrated forward and reverse logistics network. The models mentioned above always assume that the transportation or delivery routes between any two points are straight-line (Euclidean Metric or Metropolitan Metric), and that the demand and the return amount of product are constant. Hyun-Jeung KO’s integrated model also assumes that the demand of customers can be contented. However, the capacity of suppliers is sometimes larger than the demand of their customers on actual market. This paper proposes a mathematic model for the distribution network that integrates forward and reverse flows without the assumption of straight-line distance. And the demand and the amount of returns of product can vary randomly in this mathematic model. The paper also allows that the demand of customers is larger than the capacity of their suppliers. Then we develop a Genetic Algorithm (GA) to solve this NP-hard problem and discuss its convergence. At last a numerical example is available.

2. INTEGRATED LOGISTICS NETWORKS MODELING
We assume some 3PL provides forward and reverse logistics for some plants simultaneously. The 3PL must find the suitable locations of warehouses and collection centers to balance transportation cost and storage cost of products and return goods and minimize the total logistics cost.

Let $P$ be the set of plants’ forward/reverse product types, $I$ be the set of plant locations, $J$ be the set of possible sites for warehouses, $L$ be the set of possible collection centers, $S$ be the set of possible sites for hybrid warehouse-collection centers (warehouse and collection center are open at the same location), and $K$ be the set of customer locations.

Let $M_i^p$ be the maximum production capacity of plant $i$ for product $p$, $M_j$ be the maximum capacity of warehouse $j$, $M_l$ be the maximum capacity of $r_{pk}$ be the return amount of product $p$ of plant $i$ at customer $k$, $\alpha_p^j$ be the weight factor of product $p$ based on characteristics of the product type at warehouse, $\alpha_p^2$ be the weight factor of product $p$ based on characteristics of the product type at collection center, $w_j$ be fixed cost of opening warehouse $j$, $\nu_j$ be unit variable cost for warehouse $j$, $r_i$ be fixed cost of opening collection center $l$, $u_l$ be unit variable cost for collection center $l$, $h_s$ be cost savings from opening hybrid warehouse-collection center $s$ (cost saving because of storing sharing material handling equipment, infrastructure, and so on), $c_{pj}^f$ and $c_{pk}^f$ be unit transportation cost of product $p$ from plant $i$ to warehouse $j$ and from warehouse $j$ to customer $k$ respectively, $c_{pl}^p$ and $c_{pl}^r$ be unit transportation cost of return product $p$ from customer $k$ to collection center $l$ and from collection center $l$ to plant $i$ respectively, $D_{ij}$, $D_{jk}$, $D_{il}$ and $D_{li}$ be straight-line distance, and $\beta_j$, $\beta_{jk}$, $\beta_{ul}$ and $\beta_{li}$ be curve factors of two locations (the ratio of real distance to straight-line distance).

Let $x_{p_{ijk}}$ be the amount of demand for product $p$ at customer $k$ served from plant $i$ and warehouse $j$, and $y_{p{kl}}$ be the amount of product $p$ returned from customer $k$ to plant $i$ via collection center $l$. $A_j$ equals 1 if warehouse $j$ is open, otherwise 0. $B_l$ equals 1 if collection center $l$ is open, otherwise 0.

Now we can formulate this problem as follow.

$$\min z = \sum_{j \in J} [w_j A_j + v_j \sum_{p \in P} \sum_{i \in I} \sum_{k \in K} \alpha_p^j x_{p_{ijk}}] + \sum_{i \in I} [r_i B_i + u_l \sum_{p \in P} \sum_{k \in K} \sum_{j \in J} \alpha_p^2 y_{p_{kl}}] - \sum_{s \in J} h_j A_j + \sum_{p \in P} \sum_{i \in I} \sum_{k \in K} (c_{pj}^f \beta_{ij} D_{ij} + c_{pk}^f \beta_{jk} D_{jk}) x_{p_{ijk}} + \sum_{p \in P} \sum_{k \in K} \sum_{l \in L} \sum_{i \in I} (c_{pl}^f \beta_{il} D_{il} + c_{pl}^r \beta_{li} D_{li}) y_{p{kl}}$$

$$s.t.$
\[
\begin{align*}
\sum_{j \in J} x_{pikj} & \leq E(d_{pk}), \forall \in K,&& \text{if } M_i^p < \sum_{k \in K} E(d_{pk}) \\
\sum_{k \in K} \sum_{j \in J} x_{pikj} & = M_i^p, \forall i \in I, p \in P \\
\sum_{j \in J} x_{pikj} & \geq E(d_{pk}), \forall \in K,&& \text{if } M_i^p \geq \sum_{k \in K} E(d_{pk})
\end{align*}
\]
\[\sum_{p \in P} \sum_{l \in L} \sum_{k \in K} \alpha_p^l x_{pikj} \leq M_j A_j, \forall j \in J \]  
\[\sum_{l \in L} y_{pkli} \geq E(r_{pk}), \forall i \in I, p \in P, k \in K \]  
\[\sum_{p \in P} \sum_{k \in K} \sum_{i \in I} \alpha_p^k y_{pkli} \leq M_i B_i, \forall l \in L \]  
\[x_{pikj}, y_{pkli} \geq 0, \text{ and } x_{pikj}, y_{pkli} \in N, \forall p \in P, i \in I, j \in J, k \in K, l \in L \]  
\[A_j, B_i = 0, or, 1, \forall j \in J, l \in L \]  

Constraint 2 demonstrates that plant \( i \) should provide enough product \( p \) if the capacity of plant \( i \) is enough, otherwise plant \( i \) should try its best to provide product \( p \) and the amount of product \( p \) to customer \( k \) should not exceed its demand. Constraint 3 assures that the total volume of products shipped to customers cannot exceed the capacity of the warehouse serving them. Constraint 4 guarantees that all the returned product \( p \) of plant \( i \) from customer \( k \) should be collected. Constraint 5 assures that the total volume of returned products cannot exceed the capacity of the collection center serving them. Constraint 6 assures that variable \( x \) and \( y \) should be nonnegative integers.

It is easy to transform the non-linear objective 1 to linear one. Through simply analyzing, we find our problem is NP-hard. A GA will be designed to solve it.

3. GA FOR OUR PROBLEM

3.1 Encoding
Let \( g \) be the number of plants, \( k_i \) be customer number of plant \( i \), \( n \) be the number of possible warehouses, and \( m \) be the number of possible collection centers. The gene number of the chromosome is \((\sum_{i=1}^{g} k_i) \times (m + n)\). The chromosome is:
For example, the gene $a_{\sum_{i=1}^{k_{1}}}^{(1)}$ means the product amount of customer 1 transported from plant $g$ via warehouse 1. And $a_{\sum_{i=1}^{k_{1}}}^{(1)}$ means the product amount of customer 1 returned to plant $g$ via collection center 1.

### 3.2 Crossover
We use two-point crossover in our GA (PENG Yong, 2004).

### 3.3 Mutation
Each gene undergoes mutation at probability $p_{m}$. For any gene, if the probability generated randomly is less than $p_{m}$, the gene will get a nonnegative integer value not larger than maximum value $e$ (customer demand or returned amount) at the same probability. That is, the gene will get a value different from itself at the probability $\frac{e}{e+1}p_{m}$ and will keep changeless at the probability $(1-\frac{e}{e+1}p_{m})$.

### 3.4 Fitness function
The fitness function is given based on penalty function methods.

### 3.5 Population selection
Here we use normalized geometric ranking method to map the individuals to a partially ordered set (PENG Yong, 2004).

### 3.6 Rule of termination
The criterion we use stopping the GA is the maximum number of generations we specify. We need write down the best individual of each step to get the best solution of our problem. Before the crossover, we write down the best individual of this generation and insert it into the successive generation instead of the worst individual of the successive generation.
4. CONVERGENCE OF OUR GA

Optimization theory thinks that the algorithm is convergence if the solution generated in searching process is the global optimization value of the problem (PENG Yong, 2004; X.F.Qi, 1994). We need to analyze the convergence of our GA to make sure that we can use our algorithm to get the best solution of our problem.

Assumption 1: In any generation $t$, for any individual $x(x \in P(t))$ of population $P(t)$ and any $y \in \Psi$ ($\Psi$ is the searching space), if $x \neq y$, $x$ will be switched to $y$ at the probability more than $p(t)$ after one step of mutation. Here we assume that $p(t)$ is a constant more than 0 and it may have relation to generation $t$.

Assumption 2: In assumption 1, we can find a constant $p_{con} > 0$ that makes $p(t) \geq p_{con}$ correct to any generation $t$.

These two assumptions describe the fact that all the states of Markov chain can be accessed each other.

Now we can get the following conclusion.

Theorem 1: If the GA contents assumption 1 and the best value of each generation is written down, such GA will convergence to the optimization solution of the problem at some probability and the convergence has no relation to start population. Further, if the GA contents assumption 2, it will converge completely.

Theorem 2: If the best individual of any generation is kept and mutation probability is positive constant, the GA for our problem based on our encoding method is convergence completely and the convergence has no relation to start population.

Prove: For any two individuals $x, y$, let $H(x, y)$ be the distance of the two individuals. The value of $H(x, y)$ is the number counting the different genes at the same loci. For example, if there are two individuals $a = (1,2,3,2,1,1)$ and $b = (1,1,3,2,2,1)$, their distance $H(a, b) = 2$.

When we do mutation operation to any gene of $x$ at probability $p_m$, the probability of this gene’s value switching to another value is $\frac{1}{e+1}p_m$ ($e$ is the maximum value this gene can access). $e$ maybe change with gene, but we can find the maximum value $E$ among them. The probability $p(x \rightarrow y)$ of $x$ switching to $y$ can be depicted as follow ($\lambda$ is the gene number of individuals):

$$p(x \rightarrow y) \geq (\frac{1}{E+1}p_m)^{H(x,y)} \times (1-p_m)^{\lambda-H(x,y)}$$

Generally $p_m < 0.5$ to not let GA become a random researching method. Because $p_m < 1 - p_m$, we have:

$$p(x \rightarrow y) \geq (\frac{1}{E+1}p_m)^{H(x,y)} \times p_m^{\lambda-H(x,y)} = (\frac{1}{E+1})^{H(x,y)} p_m^\lambda \geq (\frac{1}{E+1})^\lambda p_m^\lambda = (\frac{1}{E+1}p_m)^\lambda$$

According to formula 9, any individual at generation $t$ will be switched to any other individual in searching space at probability more than $p(t) = (\frac{1}{E}p_m)^\lambda$. We let $p_m(t) = p_m$, then assumption 2 is contented. According to theorem 1, we can conclude that theorem 2 is contented. Actually, we can get the same conclusion when $p_m \geq 0.5$.

5. NUMERICAL EXAMPLE
Provided that there’s one 3PL who provides forward and reverse logistics service for three plants. And each plant has 8 customers. We know the locations of plants and customers. The demands of the customers change randomly. Also, we know the potential locations of warehouses as well as collection centers. Opening a hybrid warehouse-collection center is meant when a warehouse and a collection center are open in the same location. This hybrid facility thus achieves cost savings by sharing infrastructure, material handling equipment, transportation cost, etc. All the information is shown in the tables as follow.

Table 1 Plant and customer data

<table>
<thead>
<tr>
<th>Index</th>
<th>Plant 1</th>
<th>Plant 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Location</td>
<td>Location</td>
</tr>
<tr>
<td>1</td>
<td>(187.06,104.83)</td>
<td>(116.35,120.27)</td>
</tr>
<tr>
<td>2</td>
<td>(107.86,101.13)</td>
<td>(15.95,161.82)</td>
</tr>
<tr>
<td>3</td>
<td>(185.40,88.89)</td>
<td>(196.37,63.03)</td>
</tr>
<tr>
<td>4</td>
<td>(12.56,28.98)</td>
<td>(132.30,160.52)</td>
</tr>
<tr>
<td>5</td>
<td>(140.64,66.80)</td>
<td>(179.12,18.17)</td>
</tr>
<tr>
<td>6</td>
<td>(128.58,125.84)</td>
<td>(77.42,22.55)</td>
</tr>
<tr>
<td>7</td>
<td>(51.47,66.56)</td>
<td>(170.46,16.85)</td>
</tr>
<tr>
<td>8</td>
<td>(34.18,139.27)</td>
<td>(176.24,28.51)</td>
</tr>
</tbody>
</table>

Table 2 Warehouse and collection center data

<table>
<thead>
<tr>
<th>Index</th>
<th>Warehouse</th>
<th>Collection Center</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Location</td>
<td>Location</td>
</tr>
<tr>
<td>1</td>
<td>(68.28,76.36)</td>
<td>(68.28,76.36)</td>
</tr>
<tr>
<td>2</td>
<td>(37.47,23.82)</td>
<td>(37.47,23.82)</td>
</tr>
<tr>
<td>3</td>
<td>(45.00,161.99)</td>
<td>(45.00,161.99)</td>
</tr>
<tr>
<td>4</td>
<td>(140.73,93.30)</td>
<td>(140.73,93.30)</td>
</tr>
<tr>
<td>5</td>
<td>(153.46,63.34)</td>
<td>(153.46,63.34)</td>
</tr>
<tr>
<td>6</td>
<td>(104.32,150.21)</td>
<td>(104.32,150.21)</td>
</tr>
<tr>
<td>7</td>
<td>(161.60,33.36)</td>
<td>(161.60,33.36)</td>
</tr>
<tr>
<td>8</td>
<td>(122.44,120.74)</td>
<td>(122.44,120.74)</td>
</tr>
</tbody>
</table>

Table 3 Curve factor

<table>
<thead>
<tr>
<th>Warehouse(Collection Center)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>
The start population of our GA is generated randomly. The population size is 80; the maximum number of the generations is 4000; the crossover probability is 1; \( p_m \) is 0.12 and the selection probability of the best individual is 0.15.

The result suggests that we can open hybrid warehouse-collection centers at (37.47,23.82), (45.00,161.99), (153.46,63.34), (161.60,33.36) and (122.44,120.74). The total cost of the optimal integrated network is $555285. Figure 1 shows the resulting optimal network, managing forward and returned products of three plants. For the sake of simplicity, we have omitted the flows from customers to collection centers and from warehouses to customers.
CONCLUSION

This paper discusses how to integrate forward and reverse logistics to decrease the total cost of 3PLs. We formulate a mathematic model without the assumption of straight-line distance. And we assume the demand and the return amount of product can vary randomly in this mathematic model. We also allow the demand of customers can exceed the capacity of their suppliers. Through simply analyzing, we can find this model is NP-hard. Then we develop a special GA and discuss its convergence. At last our GA for the proposed model is applied to an example problem.
REFERENCES


