Risk transportation via a clique number problem formulation

G. Mavrommatis *, J.-C. Panayiotopoulos

Department of Informatics, University of Piraeus, Karaoli & Dimitriou 80, 185 34 Piraeus, Greece

Received 5 February 2001; accepted 24 April 2002

Abstract

This paper deals with the risk transportation programming problem. A new approach based on similarity coefficients, clique number and taxonomy theory, is proposed. The method can be applied in real-world transportation problems, in the case where the data area is a dynamic irregular one.

© 2003 Published by Elsevier Science B.V.

Keywords: Transportation; Risk management; Combinatorial optimization; Similarity coefficients

1. Introduction

The standard single-commodity transportation problem is that of determining the lowest cost to ship objects from some set of sources (m supply points) to some set of destinations (n demand points). In the balanced version of the problem, total supply and demand are equal. The transportation cost from producer i to consumer j is proportional to the number of units $x_{ij}$ which are transported, with a corresponding cost $c_{ij}$.

Since transportation problems are a special class of Linear Programming problems, the classical single-commodity, balanced version can be solved using well-known methods, under the condition that the costs remain constant within a given planning horizon $\Delta t$ (Regular Data Spaces). But this is not our case. In many real-world instances the transportation area we are dealing with, is not a regular one. Our transportation schedules may be affected by conditions like weather, war, strikes, accidents etc., changing at least one cost within $\Delta t$ and therefore the optimal solution might finally not be optimal at all. Additionally, in the case of Irregular Dynamic Data Spaces (IDDS) it is impossible to know a priori how data change although one can expect to know what can change them.

Broadly speaking, consider that transportation routes might be affected within $\Delta t$ under the following conditions:

* Corresponding author.

E-mail addresses: gmav@unipi.gr (G. Mavrommatis), jcp@unipi.gr (J.-C. Panayiotopoulos).
1. Transportation costs may change during a given planning horizon.
2. It is impossible to know how costs change.
3. It is known what can change the costs (“forces”).

It is obvious that in such cases we seek to pick a set of routes (paths) and amounts of goods, respectively, with the following specific properties:

(I) Routes are affected as less as possible by the same forces (dissimilar).
(II) We are willing to pay an additional cost in order to minimize risk.

Variations of the problem described above, can be applied in many real world cases. Two characteristic examples are the following ones.

• **Supplying of military units under war conditions.** The main effort of the Logistics Command is to ensure that friendly forces will be supplied with adequate goods (fuel, food, ammunition etc.), while cost becomes of second interest. The main priority is choosing proper routes ensuring that a minimum quantity of goods will eventually reach our fighting forces.

• **Transportation of hazardous materials** is a growing problem worldwide. For many reasons (for example, spreading the risk), it is strongly desired to have dissimilar paths chosen in route planning [1].

The rest of the paper is organized as follows: Section 2 presents a description and Section 3 a mathematical model/framework of the problem; Section 4 presents an approach for solving the problem based on clique; in Section 5 an easy to follow example is shown; finally, in Section 6 some conclusions are drawn.

### 2. Problem description

Let \( C \) denote the data space and \( F \) the given set of forces that can act upon our transport possibilities. We get the catastrophe system \((C, F, \Delta t)\) and the following catastrophe transportation problem [11]:

“Find the optimal solution of the transportation problem, according to \((C, F, \Delta t)\).”

Consider the given period of time \( \Delta t \) and the costs matrix

\[
C = (c_{ij}), \quad i \in I = \{1, 2, \ldots, m\}, \quad j \in J = \{1, 2, \ldots, n\},
\]

\(c_{ij}\) is the transportation cost from \(i\) to \(j\).

On the other hand, because it is known what can change the costs, the following set of forces is known too [12]:

\[
F = \{f_1, f_2, \ldots, f_k\}
\]
as well as the catastrophe matrix \(B\) of the system \((C, F, \Delta t)\),

\[
B = (b_{rq}), \quad r \in R = \{1, 2, \ldots, nm\}, \quad q \in \{1, 2, \ldots, k\},
\]

\[
b_{rq} = \begin{cases} 
1, & \text{if } f_q \text{ is able to change cost } c_{ij}, \quad i = \mu + 1, j = \lambda, \quad \text{if } \lambda \neq 0, \\
0, & \text{otherwise,} \quad i = \mu, j = n, \quad \text{if } \lambda = 0,
\end{cases}
\]

where \(\mu, \lambda\) are non-negative integers such that \(r = n\mu + \lambda, \lambda < n\).

Each row of matrix \(B = (b_{rq})\) is a \(k\)-dimensional vector representing the behavior of the corresponding route subject to the forces field. Every two routes (and therefore costs \(c_{ij}\) and \(c_{kl}\)) can either be both affected by the same force \(f_p\) or not. Suppose that these two costs can be both affected by \(\partial\) forces. These two routes
can be considered as \( \partial \)-similar. Given that the set of forces is of cardinality \( k \), every couple of routes can be from 0-similar (totally dissimilar) up to \( k \)-similar (totally similar).

In order to cluster the routes we need a proper, normalized, dissimilarities coefficient. Several relevance factors can be found in literature; from Euclidean distance to vector inner product or cosine similarity measure and even more sophisticated ones [15].

### 3. The model

In the present paper a similarities coefficient was chosen from Group Technology (GT) field [5,9]. GT is based on the principle of grouping similar parts into families leading to economies through the manufacturing cycle. This coefficient had been reported performing well [2,10,14] and was adjusted here in order to measure dissimilarities rather than similarities among the behavioral vectors of routes instead of machines and parts.

This adjusted dissimilarities coefficient is

\[
\text{dis}(i, j) = \frac{x_i + y_j}{x_i + y_j + b_{ij}}
\]

where \( x_i \) is the number of forces affecting \( i \) route but not \( j \) route, \( y_j \) is the number of forces affecting \( j \) route but not \( i \) route, \( b_{ij} \) is the number of forces affecting both \( i \)-route and \( j \)-route.

Obviously \( \text{dis}(i, j) \in [0, 1] \). If \( \text{dis}(i, j) = 0 \) then routes \( i \) and \( j \) are totally similar and if \( \text{dis}(i, j) = 1 \) then corresponding routes are completely dissimilar.

The dissimilarities matrix \( S \) is therefore constructed as follows:

\[
S = (s_{ij}), \quad i, j \in R = \{1, 2, \ldots, nm\},
\]

\[
s_{ij} = \text{dis}(i, j).
\]

For convenience, the matrix elements are be multiplied by 100 and we refer to them as percentages.

By giving an arbitrary value to parameter \( \partial \in [0, 100] \), we can search for routes with dissimilarity greater than \( \partial \) and solve the classical transportation problem according these certain routes. The closer \( \partial \) is to 100\% the less the risk.

We need to satisfy the following:

- \( \partial \) is a risk (safety) measure according the force field described above and our aim is to maximize it.
- Keep the cost within predefined margins which means that we have to decide how much we are willing to pay in order to reduce risks.
- Supply demand points with adequate goods needed.

The problem can be formulated as follows:

Minimize the transportation risk by choosing a set of routes, defined by the supply points set \( I^* \) and the demand points set \( J^* \), subject to the following restrictions:

\[
\sum_{i \in J} x_{ij} = d_j, \quad j \in J,
\]

\[
\sum_{j \in J^*} x_{ij} = p_i, \quad i \in I,
\]
Cost of transport can be defined as:

\[ \text{Cost}(x) = \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} \leq \text{MinCost} + \text{Slack}, \]

Maximize \( \partial \),

where \( x_{ij} \) is the amount of items transported from supply point \( i \) to demand point \( j \), \( d_j \) is the total amount of items that demand point \( j \) requires, \( p_i \) is the maximum amount of items that supply point \( i \) can produce and distribute, MinCost is the minimum cost of the transportation assuming that the set of forces \( F \) remains inactive during \( \Delta t \), Slack is the minimum cost of the transportation assuming that the set of forces \( F \) remains inactive during \( \Delta t \), Slack is a non-negative number giving the information on how much cost we are willing to pay in order to minimize the risk of the transportation. The maximized value of \( \partial \) found, is the so-called risk transportation number (RTN).

In Section 4, we describe an approach to solve the problem following with a simple example.

4. Risk transportation number

In what follows, the clique problem in graph theory [3] is used for solving the problem described. It must be noted that other models, related to the clique number problem, have already been used in transportation routing problems [1].

For some value of \( \partial \in [0, 100] \) we construct the matrix

\[ P(\partial) = (p_{ij}^\partial), \quad i, j = \{1, 2, \ldots, mn\} \]

where

\[ p_{ij}^\partial = \begin{cases} 0, & \text{if } (i = j) \quad \text{OR } (i \neq j \quad \text{AND } s_{ij} < \partial) \\ 1, & \text{otherwise} \end{cases} \]

Obviously, the matrix \( P(\partial) \) represents an undirected graph \( G \) where an arc exists if and only if the condition \( p_{ij}^\partial = 1 \) holds. Finding a maximum clique of this graph, say \( G^* \), provides us with the largest group of routes that are \( \partial\% \) dissimilar or more. It is therefore obvious that we have to start by giving \( \partial \) the largest possible value.

Let \( k_1 = \min_{i,j} \{s_{ij}\} \) and \( k_2 = \max_{i,j} \{s_{ij}\} \). The algorithmic approach to solve the problem is as follows:

**Algorithm Risk_transportation**

**Input:** Cost matrix \( C \), Demand & Supply Quantities \( d_j, p_i \), Slack cost, Catastrophe matrix \( B \)

1. Compute the dissimilarities matrix \( S \).
2. Compute optimal cost by solving the classical transportation problem.
3. Set RTN = \( k_2 \).
4. Set \( \epsilon_{\text{approx}} \) to a desired approximation value (i.e. \( \epsilon_{\text{approx}} = 1\% \)).
5. Construct Matrix \( P(\text{RTN}) \).
6. Find a maximum clique of the graph being represented by \( P(\text{RTN}) \).
7. Exclude every route that corresponds to a node not belonging to maximum clique.
8. Solve the generated transportation problem
9. If the solution is not feasible then set RTN = RTN - \( \epsilon_{\text{approx}} \) and go to step 5.

**Output:** RTN, Optimal Routes, Total Cost.

**End Risk_transportation**

Certainly, the problem is NP-Hard, since it contains as sub-problems the clique number problem, which is NP-Complete [6], and probably the hardest of the family [7]; but there are known methods for solving it...
distributed with good performance like [4,13], or more recently [8] where the problem size is close to graphs like 6000 nodes and approx. 1,800,000 edges, solved exactly, within a minute using just a network of 10 usual Pentium microcomputers.

5. Example

In this chapter an easy to follow example will clarify the details of the proposed method. The numeric data of the example are the same as in [11] for comparison purposes.

Suppose that $D = 3$ months, $m = 3$, $n = 3$ and the costs matrix $C$ is as follows:

$$
C = \begin{pmatrix}
2 & 2 & 8 \\
10 & 1 & 4 \\
6 & 9 & 11
\end{pmatrix}.
$$

Also suppose that $d_j = 8, 6, 3$, $j \in J = \{1, 2, 3\}$,

$$
p_i = 6, 1, 10, \quad i \in I = \{1, 2, 3\}.
$$

Solving the standard transportation problem yields to the following optimal solution:

$$
\text{MinCost} = 86, \quad x_{12} = 6, \quad x_{23} = 1, \quad x_{31} = 8, \quad x_{33} = 2.
$$

Now, suppose that $F = \{f_1, f_2, \ldots, f_8\}$

and the catastrophe matrix is

$$
B = \begin{pmatrix}
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\
1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 1 & 0 & 1 & 1 & 1 & 0
\end{pmatrix}.
$$

The dissimilarities matrix $S$ is computed

$$
S = \begin{pmatrix}
0.0 & 83.3 & 100.0 & 40.0 & 66.7 & 66.7 & 80.0 & 80.0 & 71.4 \\
83.3 & 0.0 & 80.0 & 60.0 & 60.0 & 25.0 & 75.0 & 100.0 & 85.7 \\
100.0 & 80.0 & 0.0 & 100.0 & 83.3 & 83.3 & 75.0 & 75.0 & 66.7 \\
40.0 & 60.0 & 100.0 & 0.0 & 66.7 & 66.7 & 80.0 & 80.0 & 71.4 \\
66.7 & 60.0 & 83.3 & 66.7 & 0.0 & 40.0 & 100.0 & 100.0 & 50.0 \\
66.7 & 25.0 & 83.3 & 66.7 & 40.0 & 0.0 & 80.0 & 100.0 & 71.4 \\
80.0 & 75.0 & 75.0 & 80.0 & 100.0 & 80.0 & 0.0 & 100.0 & 83.3 \\
80.0 & 100.0 & 75.0 & 80.0 & 100.0 & 100.0 & 0.0 & 83.3 & 0.0 \\
71.4 & 85.7 & 66.7 & 71.4 & 50.0 & 71.4 & 83.3 & 83.3 & 0.0
\end{pmatrix}.
$$
By setting $RTN = 100.0$ we construct the graph adjacency matrix

$$P(100) = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}. $$

This graph has clique number $= 3$ and a maximum clique is $\{5, 7, 8\}$. This clique corresponds to routes $x_{22}, x_{31}, x_{32}$ therefore we are solving the transportation problem taking under consideration only these routes. There is no feasible solution no matter of the extra cost we are willing to pay. However, Table 1 presents the whole procedure.

The clique number problem was solved using the method described in [8]. As we can see, if one sets slack cost to 18 is then able to use a solution with minimum dissimilarities about 66%, while setting the slack cost to 2 gives a solution with minimum dissimilarities 50% among the paths.

### 6. Conclusions

In previous sections, we presented an approach to solve a case of risk transportation programming, when the data area we are dealing with, is not a regular one.

Transportation paths are subject to forces that may affect them. The target is to find a set of dissimilar routes, in order to spread the risk and make sure that not all our effort and goods will be lost. The problem is NP-Hard; therefore we presented a heuristic method in order to solve it, based on similarities and the clique.

The selection of dissimilarities coefficient is an under consideration problem. Various relevance factors should be tested on real problem data.

Another open question is the selection of $\epsilon_{aprox}$. Choosing a large enough value might lead us to loss of a better solution and adoption of a worse one, while choosing a small value for $\epsilon_{aprox}$ increases the number of generated problems and therefore the computational effort needed to solve the problem. This comes because the proposed method is heuristic and cannot afford big jumps in the search space. Therefore, in order to attack the problem successfully, a distributed computation system is strongly advised. At this point we remind the fact that in order to solve the Clique problem a distributed system is already required. It is

<table>
<thead>
<tr>
<th>Risk transportation number, RTN (%)</th>
<th>Clique number</th>
<th>Clique</th>
<th>Solution</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>3</td>
<td>5,7,8</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>3</td>
<td>7,8,9</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>3</td>
<td>1,7,8</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>5</td>
<td>1,2,3,7,8</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>5</td>
<td>1,2,3,7,8</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>6</td>
<td>1,3,6,7,8,9</td>
<td>$x_{11} = 6, x_{23} = 1, x_{31} = 2, x_{32} = 6, x_{33} = 2$</td>
<td>104</td>
</tr>
<tr>
<td>60</td>
<td>6</td>
<td>1,2,3,7,8,9</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>7</td>
<td>1,2,3,5,7,8,9</td>
<td>$x_{11} = 1, x_{12} = 5, x_{22} = 1, x_{31} = 7, x_{33} = 3$</td>
<td>88</td>
</tr>
</tbody>
</table>
possible to use this particular distributed system in order to share workload among processors and speed up the computation.

References