Application Rough Set Theory and Decision Network as a New Approach to Simplify Airline Hubs Network Location

Shaaban M. Shaaban

Department of Engineering Basic science, Faculty of Engineering,
Menofiya University, Menofiya, Egypt
*Corresponding author's Email: shabaan27@gmail.com

Abstract: Airline hub location selection problems have become one of the most popular and important issues not only in the truck transportation and the marine transportation, but also in the air transportation. There are different methods for selecting hub location, however, they are mostly dependent on engineer decision making and need for high cost. In this paper, a method of rules extraction for site hub location based on rough set theory and decision network is proposed. The information system attributes are firstly reduced and a decision network with different reduced levels is constructed. The network's node is initialized with the attribute reduction sets and the decision rule sets are extracted according to the node of the decision network. In addition, the coverage degree and certainty factor are applied to filter noise and evaluate the extraction rules. A practical case of local 15 cities track network designing for airlines is given. The results are found to be comparable with the previous practical methods. Furthermore, the obtained results indicate that the proposed method saves in the time and cost of the calculation.

Keywords: Hub location; Rough-set theory; Decision Network; Decision making; Rule Extraction.

1. Introduction

Sixty years ago, air transport was a very small industry. Air travel was costly and was limited to a small portion of the inhabitance. However, in 1999, more than 1.5 billion passengers were carried on schedule airline services, equivalent to approximately 25% of world inhabitance as against an equivalent of 0.5 % of world inhabitance carried on scheduled carriers in 1945[1]. Facility location decision is the crucial side in strategic logistics planning. Nowadays the location of the facilities i.e. logistics hubs/centers, warehouses etc. is the major interest of the companies related to this business. The success of a logistic hub depends on four essential factors, such as; location, financial sustainability, efficiency and level of services. Between these the location of hubs is the most critical success factor and needs to be taken into account accurately since it has direct and indirect effects on different stakeholders including investors, policy makers, infrastructure suppliers, hub operators, hub users and the

community [2]. A hub is a point in a network where flows from different provenance to specific destinations are unified in order to serve passengers traveling between partitions and destinations for which the request is not suitable to reside direct non-stop flights. Numerous works dedicated for solving the problem of site selection of airline hubs have been reported in literature. Reference [3] is considered the first known published work classifying the airline hub location problem, however [4,5] presented the first known mathematical model for this problem by studying airline passenger networks. Recent advances in the formulations for this problem have been presented in [6,7] which used integer programming formulations for a variety of single and multiple allocation hub location, and introduced hub center and hub covering problems. Several integer programming models have been developed for single and multiple allocation [8], which include fixed costs for establishing hubs. The proposed solution includes enumeration algorithms and greedy interchange heuristics, along with Lagrangian relaxation based lower bounds. Another

1

method based on tight linear programming relations has been proposed in [9] which provided mixed integer formulations for single and multiple allocation problems, and computational results for linear programming relaxations of them. A different linear integer programming formulation has been proposed [10]. It requires fewer variables and constraints in an attempt to solve larger problems. Another technique called ant colony optimization (ACO) algorithm has been proposed [11] to solve the capacitated single allocation hub location problem (CSAHLP). An algorithm which uses ACO to solve the single allocation p-hub center problem is proposed [12]. A heuristic methods based on Tabu Search (TS) in order to solve the capacitated single allocation hub location problem has been proposed [13]. In that method three variants of a simple and efficient multi-start TS heuristic were proposed in order to explore several different initial solutions. Also, a two-stage integrated TS heuristic is proposed to improve the location and the allocation components of a solution to the problem. An algorithm for hub covering location problem, under crowdedness, with constraints and random variables are considered such as the transportation time and rate of arrived trucks has been proposed [14]. Another technique used to solve the incapacitated single allocation p-hub center problem with stochastic travel times using cutting planes and Benders decomposition [15]. It should be noted that the common difficulty with almost all the previous methods is the use many variables which make the problem complicated, tedious and time consuming. Besides, all these methods have a slow convergence rate. In fact, not all variables have the same importance and priority for final decision, therefore an attribute-based reduction procedure is proposed here as an alternative solution. The present work introduces a decision location method for airline hubs based on rough set theory and decision network. In this method the site hub location attributes are reduced, and then a decision network with different reduced levels is built. The network's node is initialized with the attribute reduction sets and the decision rule sets are extracted according to the node of the decision network. In addition, the coverage degree and the certainty factor were applied to filter out noise and evaluate the extraction rules. The proposed method has the advantage of removing redundancy from the historical data. It simplifies the design and implementation of the actual pattern classifier by determining what features should be made available to the system. In addition, the reduced input dimensionality increases the classifier processing speed, leading to better time response.

Furthermore, the input data is always sufficient for its implementation without the need of initial guess, boundary conditions or any other constraints. On the basis of a given confidence level, the concept of rule coverage was introduced. So the noises were effectively filtered-out and the extraction efficiency of rules was improved.

This paper is organized as follows. Section 2 gives an introduction to the rough set theory and decision network. Section 3 introduces the proposed rough set and decision network model along with its application to the airline hub location selection problem. Comparison with stepwise linear regression method is given in Section 4. Finally, the conclusions of the present work are given in Section 5.

2. Review of Rough Set Theory

Rough set theory (RST) is first introduced by Z. Pawlak in 1982 [16-18]. It presents a mathematical tool for dealing with imperfect and vague knowledge and uncertainty information; its methodology is interested with the analysis and classification of uncertain, imprecise, or incomplete information .it is appreciated one of the first non-statistical techniques to data analysis. Vagueness and uncertainty information has become a critical issue for computer scientists, and there are many approaches to recognition and affecting imperfect knowledge. The most successful approach is based on the fuzzy set notion. The main idea of (RST) has been successfully applied in many real-life applications in engineering, remedy, banking, pharmacology, fiscal and market evaluation . The main advantage of rough set theory is that it does not need any introductory or additional knowledge about data, such as probability in statistics or degree of membership in the fuzzy set theory.

2.1 Decision Tables

Definition 1): In rough set theory, an information system can be considered as system $S = (U, A, V, f)$, where U is the universe; $A = C \cup D$ is the sets of condition attribute, the subset C and D are disjoint sets of condition symptoms attribute and decision attributes respectively; $V = \bigcup_{r \in R} V_a$, where Va is the value set of fault symptoms attribute a, is named the domain of attribute a .Each attribute $a \in A$; f is an information function $f : UxA \rightarrow V$, and $f(x, a) \in V_a$, in which $x \in U$.

2.2 Equivalent relations

Definition 2): In decision system S = (U, A, V, f), every attributes subset, an indiscernible relation (or equivalence relation) IND(B) defined in the following way:

$$IND(B) = \{(x,y) \in U \times U \,|\, \forall a \in B, f(x,a) = f(y,a)\} \quad (1)$$

The family of all equivalence relation of IND(B), a partition determined by B, denoted by U/IND(B),[x]B can be considered as equivalence classes, and defined as follows:

$$[x]_B = \{y \in U \,|\, \forall a \in B, f(x,a) = f(y,a)\} \quad (2)$$

And

$$[x]_{IND(B)} = \cap [x]_B \quad (3)$$

2.3 Reduction and Core

Definition 3): In decision system S=(U, A,V ,f),Let $b \in B$ and $B \subseteq A$, if posB(D) = posB $-\{b\}$(D), attribute b is superfluous to B, which relatives to D, otherwise the attribute b is essential.

If IND(B) = IND(A) and POSB (D) $\neq$ POSB $-\{b\}$ (D) , then B is called a reduction for information system S , are denoted as RED(A) ; the intersection of these reduction sets is called core, denoted as CORE = $\cap$ RED(A) .

2.4 Decision Networks and Decision Rules

Discovering patterns is considered the main problem in data mining. The data patterns are usually presented in the form of decision rules, which are logical formulas in the form "if X then Y" where X and Y are referred to (conditions) and (decisions) of the rule, respectively. Knowledge extraction is implemented through representing hidden relationships between data in a form of decision algorithms which is the decision rules. In certain circumstances, a set of decision rules is not sufficient for describing relationships in the database. Therefore, a knowledge of relationship between decision rules is necessary in order to understand better data structures. A decision algorithm provided with the relationship between decision rules will be called a decision network. It is described as a finite, directed acyclic graph, nodes of which represent logical formulas, whereas branches are interpreted as decision rules. Every path in the graph represent a chain of decisions rules, which are used to describe compound decisions.

Let U be a non-empty finite set, called the universe and let X, Y be logical expressions. The

meaning of X in U, denoted by $|x|$ is the set of all elements of U, that satisfies X in U. The truth value of X denoted $val(X)$ is defined as $card|X|/card(U)$ where card(X) denotes cardinality of X. The number $Supp(X,Y) = Card(|X \wedge Y|)$ will be called a support of the rule [16-18].

With every decision rule $X \rightarrow Y$ we associate its strength defined as:

$$str(X,Y) = supp(X,Y)/card(U) \quad (4)$$

Moreover, the certainty factor defined as:

$$cer(X,Y) = str(X,Y)/val(X) \quad (5)$$

And the coverage factor defined as:

$$cov(X,Y) = str(X,Y)/val(Y) \quad (6)$$

3. Application Of Rough Set Theory And Decision Network In Airlines Hub Location Selection

In rough set theory and decision network data modeling all the information should be collected in a table, in which each column represents a characteristic or a property and the last column indicates the decision level. All the information available about the location is the identical row in the table. Information tables are usually complicated to appreciate. They may memorize a vast quantity of data, which is difficult to manage for decision making. One of the main objectives of Rough Sets data analysis is to reduce data size. In order to show the applicability of the Rough Sets theory, the information table was constructed, according to standards that are prepared by experts in airline industry [19]. Thus, Table 2 can be obtained with 10 conditional attributes, such as (a) Passenger volume; (b) Cargo volume, and (c) GDP per Capita. The collected data is classified as a table in which each row indicates the specifications of a particular site and each column of the table indicates one of the characteristics considered for the location and the last column shows the suitability of the site for the project. For application of Rough Sets theory to analyze the information, the data should be classified. Consequently, each conditional attribute is provided with 3 classes, which show high, medium, and low and the decision parameter (attribute) is classified by two levels, which describes high suitability condition (H) and (N) which indicates none suitable condition. The classification of all attributes has been carried

out by defining the specific levels and assigning a code to each specified attribute in the rows of the table. Table 1 shows the class numbers of conditional attributes and decision levels for 15 selected sites. For instance, site L1 is classified into class number 2 of conditional attribute (a), class number 3 of conditional attribute (b) and class number 3 of conditional attribute (c) and its decision level is detected as "H". This table shows the relationship between the classes numbers of the conditional attributes of each location and its decision attribute. This relationship is named "decision rules" and such a table is called "a decision table".

Table 1. Decision table

Locations	Conditional Attributes										Decision Levels
	a	b	c	d	e	f	g	h	i	j	
L1	2	3	3	3	2	3	3	2	3	1	H
L2	2	2	2	3	1	2	3	1	2	2	N
L3	1	2	2	3	1	2	3	1	1	1	H
L4	2	2	2	3	1	2	3	1	2	2	N
L5	1	2	2	3	1	2	3	1	1	1	H
L6	2	3	3	3	2	3	3	2	2	3	H
L7	1	2	1	3	1	1	3	1	2	1	N
L8	1	2	1	3	1	1	3	1	2	1	N
L9	2	1	2	1	1	2	3	1	2	1	N
L10	1	2	1	3	1	1	3	1	2	1	N
L11	2	2	2	3	1	2	3	1	2	2	H
L12	1	1	1	3	1	1	1	1	1	1	N
L13	1	1	2	3	1	2	3	1	2	1	H
L14	2	2	1	3	1	2	3	2	3	3	N
L15	1	1	1	3	1	1	1	1	1	1	N

The determination of site hub location method based on rough set theory and decision network as follows:

Step 1. Check whether the decision levels are compatible with 10 conditional attributes in Table 1, which shows the summary of diagnostic results by experts. The decision rules of all locations were examined to find non-deterministic rules; i.e., locations which were classified into one and the same class under every conditional attribute but were classified different decision levels.

Table 2. The hub model conditional attributes

Conditional Attributes	units
(a) GDP per Capita	Dollar
(b) International influence	Route
(c) Passenger volume	10000 person
(d) Infrastructure of railway	(bureau, section and station)
(e) Cargo volume	10000 ton
(f) National influence	Route
(g) Airport facility	(3C-4F)
(h) Infrastructure of Road	Kilometer
(i) GDP in region	Dollar
(j) Infrastructure of Harbor	10000 ton

Step2. Implement the Reduction of decision Table 1. Decision Table 1 is reduced by using software called ROSE2 (a Rough Set tool kit for analysis of data). The GA is adopted in Reduction process. Thirteen reductions are produced; they are shown in Table 3.

Table 3. Reducts of Decision table 1

	Reducts	Length
1	{f, d, j}	3
2	{c, a, d}	3
3	{c, a, b}	3
4	{c, d, j}	3
5	{a, i, b}	3
6	{c, a, j}	3
7	{a, f, j}	3
8	{a, i, f, d}	4
9	{c, i, b, d}	4
10	{a, b, f, h}	4
11	{i, b, d, j}	4
12	{a, f, h, d}	4
13	{i, b, f, d}	4

Step3. Construct hub location decision network for decision Table 1. As shown from Table 3, that there is thirteen reducts for decision Table 1. So if we construct the decision network it will be complicated.

In order to get a simple decision, network the frequency of appearance of each conditional attribute in Table 3 is shown in Figure 1, and by ranking conditional attributes in the descending order of appearance frequency, obtain for each reduct set the sum of frequency of attributes corresponding to it and length equal four, one can see that the most significant reduced set is: {a, b, d, f}.
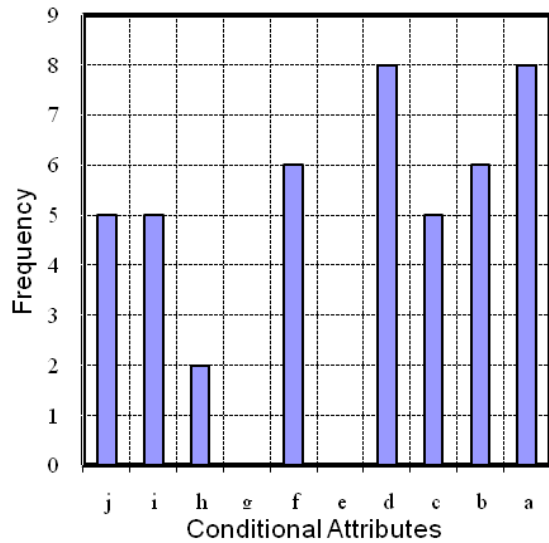
Figure 1. The frequency of appearance of conditional

attributes in Table 1

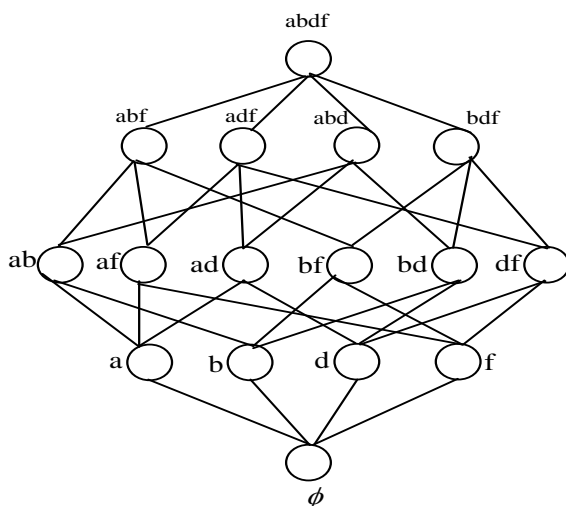Hence the fault decision network for Table 1 is constructed as shown in Figure2.

Figure 2. The decision network for decision table1

Step4. Extracting the fault diagnosis rules. From the decision network shown in Figure 2, it can be seen that all the nodes with decision rules except the bottom empty node, which includes extracting effective diagnosis rules from the incomplete diagnostic information. For each node of the decision network, evaluate the rule's accuracy and coverage degree. Extract decision rules whose accuracy is greater than the threshold value $\mu 0$. Set $\mu 0=0.8$ if all the confidence degrees are less than 0.8 put the biggest two confidence degree items in rule set. The corresponding decision rules of Table 1 are shown in table 4.

Table 4. Hub location model decision rules

Node	Diagnosis decision rule	Con. Degree	Cov. degree
a b d f	a=2, b=3, d=3 , f= 3 $\rightarrow$ Decision Level is H	1	0.1
a b f	a=1, b=2, f= 2 $\rightarrow$ Decision Level is H	1	0.375
	a=1, b=1, f= 3 $\rightarrow$ Decision Level is H	1	0.5
a b d	a=1, b=1, d=3 $\rightarrow$ Decision Level is N	1	0.1
b d f	b=1, d=3 , f= 1 $\rightarrow$ Decision Level is N	1	0.125
	b=1, d=1 , f= 2 $\rightarrow$ Decision Level is H	1	0.1
a b	a=2, b=1 $\rightarrow$ Decision Level is N	1	0.1
a f	a=2, f= 2 $\rightarrow$ Decision Level is N	1	0.273
b f	b=3, f= 3 $\rightarrow$ Decision Level is H	1	0.363
	b=3, f= 1 $\rightarrow$ Decision Level is H	1	0.125
b d	b=1, d=3 $\rightarrow$ Decision Level is L	1	0.1
d f	d=1 , f= 2 $\rightarrow$ Decision Level is N	1	0.2
d	d=1 $\rightarrow$ Decision Level is H	1	0.2
f	f= 1 $\rightarrow$ Decision Level is N	1	0.64

4. Comparison Between Rough Set Decision Network Model And Stepwise Linear Regression Method

Regression analysis is a statistical technique that can test the hypothesis that a variable is dependent upon one or more other variables [21]. The decision making algorithms obtained from the rough set decision network model and the algorithm resulted from the stepwise linear regression method are calculated in this paper. Therefore, as the first criterion, the R^2 is obtained from the multiple linear regression equations by various decision making algorithms. These are achieved by statistical software, the results of which are reflected in Table 5. In addition, the degree of difference within the mean values of R^2 obtained from the multiple linear regression of decision-making algorithms of a rough set decision network model and the stepwise method is about 1.8%. To evaluate the results of the stepwise linear regression method, the values of the accuracy and quality of approximation have been investigated. These values represent how carefully the defined independent parameters can predict the dependent parameter in the algorithm. The algorithm resulting from stepwise analysis It was observed that other algorithms of attributes have less values of accuracy

and quality of approximation in comparison to reducts resulting from rough set analysis. For example, it was observed that the algorithm resulting from stepwise analysis and the algorithm obtained from the most important attributes defined in different hub location models, have lesser values of accuracy and quality of approximation. For instance, the differences between these values with the algorithms concerning the rough-set analysis are26% and 14% respectively, which is indicative of the great difference between the results of the two methods.

Table 5. Comparison of correlation coefficient, quality and accuracy of approximation of different reducts

Algorithm	R^2	Accuracy of approx.	Quality of approx.
{f, d, j}[a]	68.2%	1	1
{c, a, d}[a]	73.5%	1	1
{c, a, b}[a]	70.5%	1	1
{c, d, j}[a]	65.4%	1	1
{c, a, j}[a]	73.9%	1	1
{a, f, j}[a]	76.9%	1	1
{a, i, f, d}[a]	78.8%	1	1
{c, i, b, d}[a]	65.7%	1	1
{a, b, f, h}[a]	79.7%	1	1
{i, b, f, d}[a]	65%	1	1
{a, f, h, d}[b]	81.5%	0.74	0.86

(a) The shortest decision-making Reducts resulted from Rough Sets Decision network analysis.
(b)The Reduct resulted from stepwise regression analysis. The R^2 value resulted from the linear regression of all parameters (all the 10 parameters) is 84.7%

Therefore, the value of accuracy and the quality of approximation were studied for different decision-making algorithms and consequently it was observed that other algorithms of attributes have less value of accuracy and quality of approximation in comparison to reductions resulting from Rough Sets analysis.

5. Conclusion

In this paper, rough set theory decision network is used to determine the optimal decision-making reductions to provide and analyze hub location selection models and compared with the results obtained by stepwise linear regression method. The correlation coefficients of the reductions obtained from the proposed method are compared with those determined from stepwise linear regression method. It was observed that the difference between the values of R^2 in reductions concerning rough set decision network analysis and the stepwise linear

regression method is negligible. The quality and accuracy of approximation values are also investigated for the reductions concerning both methods. It was observed that the reductions of attributes determined from stepwise linear regression method have a lower value of accuracy and quality of approximation, in comparison with those resulting from the proposed method. Thus, the proposed rough set decision network provides a significant reduction in the amount of information necessary for hub location selection modeling which is positively reflected on the speed and effectiveness of information processing. Besides, the proposed method has no need to compare parameters at the beginning besides all relevant parameters are equally entered in the calculation procedure, resulting a remarkable saving in time and cost of calculation. The future extension of the present work include design the hybrid intelligent algorithms by combining the proposed approach and multi-criteria decision making (MCDM) to ranking the hub locations and select the best location.

References

[1] J.T.D. Jayalath, J.M.S.J. Bandara. Future of Colombo airbort (CMB) as an airline hub. *Journal of Air Transportation World Wide (IJSCE),* Vol. 6, No. 2,pp 117-128, 2001.

[2] Sheikh ArifulAlam. "Evaluation of the potential locations for logistics hubs: A case study for a logistics company". *Master Thesis, KTH Royal Institute of Technology at Stockholm, Sweden*, 2013.

[3] A. J. Goldman. Optimal locations for centers in a network. *Transportation Science*, 3(4):352–360, 1969.

[4] M. E. O'Kelly. Activity levels at hub facilities in interacting networks. *Geographical Analysis*,18(4):343–356, 1986.

[5] M. E. O'kelly. The location of interacting hub facilities. Transportation Science, 20(2):92–106,1986.

[6] J. F. Campbell. Integer programming formulations of discrete hub location problems. *European Journal of Operational Research*, 72(2):387–405, 1994.

[7] J. F. Campbell. Hub location and the p-hub median problem. *Operations Research*, 44(6):923–935, 1996.

[8] T. Aykin. Networking policies for hub-and-spoke systems with application to the air transportation system. *Transportation Science*, 29(3):201–221, 1995.

[9] D. Skorin-Kapov, J. Skorin-Kapov, and M. E. O'Kelly. Tight linear programming relaxations of uncapacitated p-hub median problems. *European Journal of Operational Research,* 94(3):582– 593, 1996.

[10] A. T. Ernst and M. Krishnamoorthy. Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research*, 86:141–159, 1999.

[11] M. Randall. Solution approaches for the capacitated

single allocation hub location problem using ant colony optimization. *Computational Optimization and Applications*, 39(2):239–261, 2008.

[12] T. Meyer, A. T. Ernst, and M. Krishnamoorthy. A 2-phase algorithm for solving the single allocation p-hub center problem. *Computers & Operations Research,* 36(12):3143–3151, 2009.

[13] M. R. Silva and C. B. Cunha. New simple and efficient heuristics for the uncapacitated single allocation hub location problem. *Computers & Operations Research*, 36(12):3152–3165, 2009.

[14] M. Mohammadi, F. Jolai, and H. Rostami. An m/m/c queue model for hub covering location problem. *Mathematical and Computer Modeling*, 54(11):2623–2638, 2011.

[15]Hult, E., Jiang, H. & Ralph, D., 'Exact computational approaches to a stochastic uncapacitated single allocation p-hub center problem', *Computational Optimization and Applications* 59(1-2), 185–200, 2014.

[16]Pawlak, Z. : Rough sets theoretical aspects of reasoning about data, Kluwer Academic Publishers, Boston, London, Dordrecht, 1991.

[17]Pawlak,Z; Skowron,A: Rough sets: Some extensions, *Information Sciences* 177 (2007) 28–40

[18]Pawlak,Z.;Skowron,A.: Rough sets and Boolean reasoning*Information Sciences* 177 , 41–73, 2007.

[19]Shaaban M. Shaaban and Hossam A. Nabwey, A probabilistic rough set approach for water reservoirs site location decision making, in*: B.Murgante et al. (Eds.), Proceedings of The 12th International Conference on Computational Science and Its Applications (ICCSA 2012), LNCS 7334*, Springer, pp. 358–372, 2012.

[20]Qhrn A., "Discernibility and Rough Sets in Medicine: Tools and Applications", Ph.D dissertation, Dep. Compu. Sci. Inform. Sci., Norwegian Univ. Sci. Technol., Trondheim, Norway, 2000.

[21] S. Chattefuee, A.S. Hadi, Regression Analysis by Example, 4th edition, John Wiley& Sons, 2006