

Application of Data mining in the Specifying the Ongoing Rate of Revenue Rise in Over-Flight Field

Bahram Kazempour¹, MohammadReza Davari², Fariborz Gahahremani³, Behrooz Minaei⁴
^{2,3} M.Sc. in Information Technology – e-Commerce/Shiraz University,
⁴ Assistant Professor/Iran Industrial & Science University

kazempour_b@yahoo.com, mreza.davari@gmail.com, fariborz.gahremani@gmail.com,
b_minaei@iust.ac.ir

Abstract. *Over-flight over I.R. of Iran airspace territory is one of the main sources of revenue for our country. In this regard, the appropriate planning in encouraging airlines to over-fly Iran, which is a monopoly in this country, is of a significant importance. On the other hand, the competitive atmosphere dominating the aviation industry particularly in Middle East and considering the presence of new rivals due to reopening the new air routes in our neighboring countries including Iraq is one of our country's chief concerns for importing foreign currencies. At present in many countries, including Iran the tariff for air navigation charges has a fixed formula and there has been a consistency in this regard. In many countries, including Iran a fixed discount like 30% has been applied to encourage more flights. In this research, having used the data mining techniques, we describe the data of over-flight over 15 years.*

By using clustering, the total flights are divided into 6 groups and we recommend a changing discount to the directors of the organization in order to increase the flights with an ongoing with an ongoing perspective. In the calculation of navigation charges and considering the competitive atmosphere in encouraging more over-flights and using data mining techniques, we use an ongoing and flexible pattern and the output will lead to a rise in the revenue. Application of this pattern in the whole country will be helpful as well, because economic and commercial planners will be familiarized with application of data mining and its impact on revenue increase and since many problems exist in other industries as well which can be eliminated using the data mining by an appropriate and effective plan in the national and industrial level.

Keywords:

Data mining, Clustering, Over-Flight, Tariff for air navigation, Changing discount

1. Introduction

Nowadays the Knowledge of data mining as one of man's modern technologies, not only has undergone tremendous changes, but also is speedily affecting life pattern, research method instruction, management, commerce, transportation, safety and security measures and other aspects of human life.

Technology of data mining is one of the useful ways in finding data identified helpful among enormous quantity of data.

Having discovered patterns and relationships between data, data mining reveals the hidden values. Discovery of these values can help us anticipate the values of other variables and decide.

This research simultaneously introduces this technology and assesses the advantages, foundation, processes, duties and performances of it and ultimately we have put our efforts in exploiting the use of techniques of decision tree in data and data mining and its application in achieving the concerned goals.

Undoubtedly, making wise use of data mining assists us through increasing revenues gained from over-flight charges in Iranian Airports Company and promotes commercial efficiency and also contributes us to be more active in the field of commerce and enhances this company in terms of regional and international markets.

2. Problem statements and research objects

Nowadays, despite the abundancy of data, we suffer from poor knowledge of internal data. In such circumstances, those organizations that discover and achieve more internal data will gain success.

Having the enormous quantity of transactions, aviation industry plays a chief role in economies of countries. Iran is not an exception. Air transportation industry, particularly Iranian Airports Holding Company, is one of the basic

elements of this industry. In our country this company is the forerunner and poses the monopoly for its activities. Its major concerns are globalization, entering international trade organization, increasing competition in this industry and also subjects such as air-routes in the neighboring countries including Iraq.

The above transformations ensure survival and sustainability of Iranian Airports Holding Company in the long run and its profitability in the short run. Considering the revenues gained from over-flight charges and in order to maintain its competitive advantage, and to keep and promote its share in Middle East, to increase profitability and customer satisfaction (that are usually airlines), the company should make use of this modern innovation of information era. Application of data mining in the developing countries including Iran is so obvious. Therefore doing researches in this field seems to be necessary through which more familiarity with the concept and dimensions of data mining is secured and that elements affecting it will be identified as well.

Hence we try to analyze the issue of over-flights using the territory of Iran airspace and recognize practicable and efficient solutions for increasing the number of flights in this research. This process will allow the management of Iranian Airports Holding Company to make proper and better decisions.

2-1. How This Research Is Done?

This project is done in several major phases:

- [1] data collection, assessing and analyzing it and finally choosing the appropriate data
- [2] necessary actions for recognition and preprocessing
- [3] application of various techniques by data mining soft wares on data and classification, clustering and discovering the existing association rules.
- [4] In the final phase our effort will be to provide a pattern to increase the revenue.

2-2. Definition of Investigation Concept

Iranian Airports Holding Company: it is a 100% governmental company Connected with Ministry of Transportation that is lawfully and financially independent and its duty is organizing, creating, managing, maintenance, completion, facilitating and exploitation of airports of our county and provision of services and fringe airport and aeronautical services.

Over-flights: it is referred to a non-stop flight overflying airspace of the country using air navigation services while having prior permission

3. Research Method

Websites are full of hidden data which can be used in making intelligent decisions. Airlines are permanently assessing their performance by collecting data related to the flights and creating websites. Analysis of data is often provided in the form of reports and statistical tables and ultimately comparing performances of previous months or years.

With regard to data assessment, first we identify, preprocess and prepare data in order for data mining and then having applied various methods and techniques by data mining software, we classify, cluster and discover the existing rules. In this way the enormous existing data will be distinctive from the general data and then with studying this data we try to draw a conclusion existing in them. Consequently with assessing the discovered patterns, a new and an ongoing model of floating discounts for aircrafts passing airspace of Iran territory is provided. In general this research is done in 6 major parts:

- [5] comprehension of commerce
- [6] understanding the data
- [7] data preprocessing
- [8] pattern discovery
- [9] assessment
- [10] putting in order

3-1. Data Understanding

In Tab. 1, Data for mining and type of them is presenting

Tab.1. Data and data type

Field name	Type	Size
Type	Character	5
Weight	Numeric	8
Enter point	Character	5
Exit point	Character	5
From	Character	5
To	Character	5
Distance	Numeric	7
Charge	Numeric	10
Airline	Character	4

3-2. Founding TwoStep Model

After executing the TwoStep Cluster modeling on data (Tab.1), some rule founded by C&RT Classifier that present in Fig.1

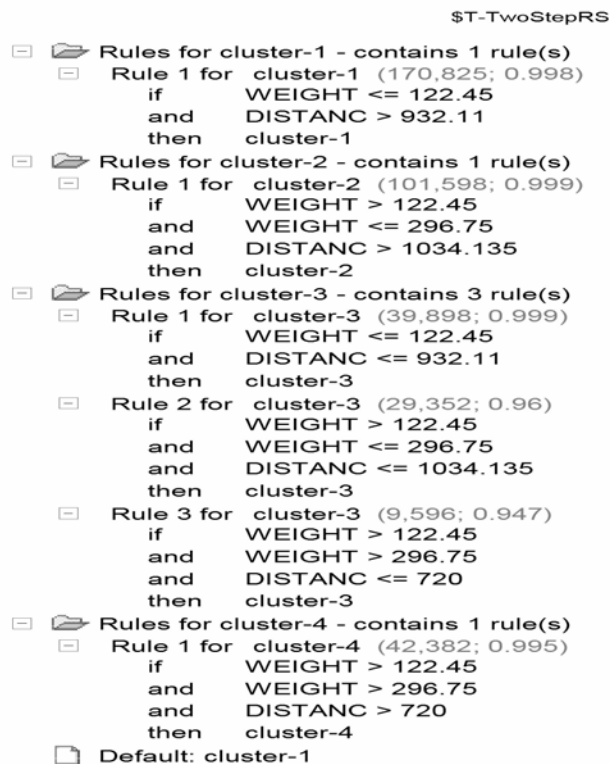


Fig.1. Rules founded by TwoStep Model

3-3. Analysis design on the results of C&RT Tree Node :

Analysis

☐ Results for output field \$T-TwoStep

☐ Comparing \$R-\$T-TwoStep with \$T-TwoStep

Correct	391,261	99.39%
Wrong	2,390	0.61%
Total	393,651	

the above analysis shows that, the obtaining results have a high verification of 99.39% .

3-4. Presenting new model for discount payment

1 Computation of incoming in the first cluster:

In this phase, by inserting the following formula in the selection node and then by making a sum on the foreign key field PK and the field of charge, the total revenue of this cluster will be as follows:

Weight <= 122.45 and Distance > 932.11

The resulting revenue of the first cluster : 74,835,894 \$

The total discount in this cluster = 0 \$

For computing the revenue portion percentage resulting from the first cluster of total income, we have:

$$74,835,894 / 656,434,505 * 100 = 11.4\%$$

Revenue Part of the first cluster 11.4 percentage

The rate of performing discount in this cluster due to flights with less than 150 tons, has been zero percent.

2 Computation of incoming in the second cluster:

Similar to first phase but with the following formula, in the selection node, the result will be:

WEIGHT > 122.45 and WEIGHT <= 296.75 and DISTANC > 1034.135

The revenue of the second cluster more than 615,000,000

The total discount in this cluster About 168,000,000

For computing the revenue portion percentage of the second cluster from the total revenue we will have:

$$615,000,000 / 1,656,000 * 100 = 37.1\%$$

The revenue portion of the second cluster 37.1 percent

The rate of performing discount in this cluster has reached to 30 percent, up to now (only belongs to the flights in this cluster with the weights more than 150 tons.) Regarding that there is no discount to the flights Less than 150 ton.

Note: the rate of dedicated discount of this group to the total revenue equivalent to 27.30 percent is obtained from the following formula:

$$168,000,000 / 615,000,000 * 100 = 27.30\%$$

3 Computation of incoming in the third cluster:

Similar to the first phase but with the following formula in the selection node:

(WEIGHT <= 122.45 and DISTANC <= 932.11) or (WEIGHT >122.45 and WEIGHT <= 296.75 and DISTANC <=1034.135) or (WEIGHT >296.75 and DISTANC <=720)

The incoming of the third cluster more than: 78,000,000

The rate of total discount in this cluster for computing the revenue portion percentage of that total revenue of the third cluster we will have: $78,000,000 / 1,656,424,505 * 100 = 4.7\%$

The incoming portion of the third cluster 4.7 percent the result of above research of data mining

4. Conclusion

Regarding to the different behavior of over fly flights in different distance and weights obtained in the first phases of this research the results will be presented in independent different clusters during data mining phases, so it is suggested for Iranian Airports Holding Company to notice to mentioned suggestions for its incoming plans.

As the resulting clustered executive model twoStep and K-means indicates, most of the results are based on clustering on weights and distance factor. (It seems that the results obtained are useful and suitable for the goal of this research since the goal of this research is declared as follows: **computation of the logical and floating discount which not only is emotional in over flights attraction, but also increases the revenue for the country.**

In results of the average K-means Model, division of clusters are in such a way that the most of flights are located in one cluster and little percentage of them are located in another. (this result in not suitable for deduction of Airlines behavior pattern which is the goal of this research). But the results of twoStep clustering are very practicable because the they are located between the different clusters with different distance and weights more logically and a more efficient pattern for presenting discounts can be provided.

With investigating the rules of twoStep clustering based on distance traversed and the weights of over flying airplanes traversed and the weights of over flying air planes over Iran, the following applicable results will be achieved:

- In the fourth cluster including flights with the weights more than 296.75 tons and distances more than 720 kilometers with no discount up to now, 29 % discount is suggested (which is more than 8800000 Dollars from economizing in the fourth cluster)
- In the first cluster including flights with the weights less than 122.45 tons and distances more than 932.11 kilometers with no discount up to now, 12% discount is suggested. (Which is equal to 8800000 \$ from economizing in the fourth cluster).
- In the second cluster including flights with the weights between 122.45 and 296.75 Tons and distances more than 1034.135 Kilometers, 27.30% discount is suggested to the total of cluster.
- In the third cluster including flights with weights less than or equal to 122.45 tons and distances less than or equal to 932.11 kilometers or weights between 122.45 and 296.75 Tons and distances less than or equal to 1034.135 Kilometers or with weights more than 296.75 tons and distances less than or equal to 720 kilometers, 24.2 % discount to the total of cluster is suggested.

5. References

1. Bose & Sugurmaran ,(2005). " Application of Knowledge Management Techniques in Customer relationship Management " , "*Journal of Knowledge and Process Management*" . Vol.10 , No.1 . PP.3 - 17
2. Cahlink, George (Oct. 2000). "Data Mining Taps the Trends," "*Government Executive Magazine*".
3. Chen, M.S., Han, J., & Yu, P.S.(1996). "Data Mining : an overview from a database perspective ". "*IEEE Transactions on Knowledge and data Engineering*". 8(6). 866-882.
4. Dholakia, Nikhilesh and Xiao, Jing J. and Dholakia, Ruby Roy and Mundorf, Norbert (2000). "The impact of retail ecommerce on transportation: A conceptual framework"
5. Fayyad, Al (1996). "Advances in knowledge discovery and data mining".
6. Fayyad, Usama and Piatetsky-Shapiro, Gregory and Padhraic, Smyth. (1996). "From Data Mining to Knowledge Discovery in Databases
7. J.Han, and M.Kamber,(2001). *Data Mining: Concepts and Techniques*, San Diego Academic Press.
8. Krzysztof, J. Cios and Witold, Pedrycz. And Roman, W.Swiniaraski (2007). "Data Mining ,A Knowledge Discovery Approach". Springer.
9. Koperski, K., & Han, J. (1995). " Discovery of spatial association rules in Geographic information databases". "Proceeding of 4th International Symposium on Large Spatial Databases". (pp. 47-66).
10. Lummus, R.R., and Vokurka, R.J. (1999), "Defining supply chain management: a historical perspective and practical guidelines". "*Industrial Management & Data Systems Journal*". Vol. 1999, No. 1, pp.11-17
11. Meade,L.M., & Sarkis, J. (1999). "Analyzing organizational project alternatives for agile manufacturing processes: an analytical network approach". "*International Journal of Production Research*", 37(2), 241-261.
12. Nazeri, Zohreh(2003), "Application of aviation safety data mining and workbench at American airlines, Center for advance aviation system development (CAASD)", The MITRI Corporation.
13. Ralf H. Mayer, (2007). " Estimating operational benefits of aircraft navigation and air traffic control procedure using an integrated aviation modeling and evaluation platform, Center for advance aviation system development (CAASD)" , The MITRI corporation, 7515 Colshire Drive, Mclean, VA22102, U.S.A.
14. Shekhar, S., & Schrater, P. & Vatsavai, W.r., & Wu, W., & Chawla, S., (2002)." Spatial contextual classification & prediction models for mining geospatial data". "*IEEE Transactions on Multimedia*, 2(4)", 174-188.
15. Two Crows Corporation,) 1999" .(Introduction to Data Mining and Knowledge Discovery".
16. U.S. General Accounting Office, (May 2004). " Data Mining: Federal Efforts Cover a Wide Range of Uses". " GAO Report GAO-04-548 (Washington)"
17. W. Seifert, Jeffrey.(2004) "Data Mining: An Overview, CRS Report for Congress".