

3.1 Data&knowledge mining, relationships detection using inductive modeling

Making Use of Data Mining for Presenting Behavior Model of Airline Agencies Ticketing in the supply chain of airline companies

MohammadReza Davari¹, Fariborz Gaharemani², Bahram Kazempour³, Behrooz Minaei⁴

^{1,2,3} M.Sc. in Information Technology – e-Commerce/Shiraz University,

⁴ Assistant Professor/Iran Industrial & Science University

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

Abstract. Every strategic planning requires enough information regarding the subject and the appropriate knowledge of making use of that information. Nowadays the technique of data mining is recognized in the highly developed countries as a principle way for knowledge discovering on the basis of the data accumulated through several years. The need for applying this technique is more remarkable in such companies as airline agencies (because of having low interest and need to be subsidized by the government) than the other ones. The aim of this paper is to extract the existing knowledge as well as the behavior model between an Iranian airline company and the ticketing agencies as its obverse of contracts, using methods of business intelligence, in order to optimizing the relationship between them and producing further and stronger motivations among personals of ticket seller agencies and encourage them to operate tours.

The common way of paying commission to the ticketing agencies by the airline companies is to pay a fix percent of the price of the ticket, irrespective of how many tickets each agency has sold. The authors of the present article suggest a new method of paying commission: dispensing a variable coefficient according to the amount of selling the tickets. Interesting to note, in this new method, the sum of commission paid would be finally equal with the whole amount paid in the current method, but the further attempt an agency makes to sell the tickets, the further percent of the price of the ticket it would takes for itself. Considering the variable nature of aforementioned coefficient, this method of paying can be called the dynamic method.

Being different in population, number of flight, number of ticketing offices and the sum of tickets sold, cities are primarily divided into four groups by applying the clustering method with the data existing in the system of ticket selling through five recent years. Then, in every group including cities with similar conditions, we will separately obtain to the knowledge of the behavior patterns of agencies in selling tickets according to the index of each agency, five years of selling, twelve months in every solar year, the percent of selling tickets by each agency per month in proportion to whole selling of the same group and the maximum of ticket reservation at every time of selling by each agency per month. Having studied these behavior patterns, we will finally suggest a new dynamic model in the way of paying the commission to the agencies. By comparing the obtained consequences of the knowledge acquired in every different kind of clustering and classification methods⁵, it is possible to prepare an optimum model of knowledge discovering from every sort of these data (which is the same in all countries because of using a same standard frame) for the system of selling ticket of airline companies for countries all over the world.

Keywords

Data Mining, Business Intelligence (BI), Supply Chain Management (SCM), Clustering, Classification.

1. Introduction

by Computer technology acquisition and using of database systems in many ways, In many of countries in last decades, a lot of information in different sections whether Industrial or social service and government or private sections has saved in data ware houses. This information can be a valuable source for knowledge discovery. This knowledge helps organization elevation , particularly in making competitive advantages for organization in target markets. In orientation of this purpose we start research on an Iranian airline ticketing reservation. Generally aviation systems consist huge chain of various suppliers (among airplane producer companies, spare accessory producers, services companies, fuel producer companies, airport companies, airplane companies, ticketing agents, and end users). But in this paper we zoom on B2B ring between airplane companies and ticketing agencies. reservation system in this company whose under our research, include last 6 years information. This database consist many fields same as passenger information(name, sex, address, and tel. no.), number of tickets purchased, source and destination, flight No., ticketing office code, ticket cancellation status, and etc. in an traditional procedure, the airplane companies are paying commission in static way.(a regular percent of tickets fee). We think this method of commission paying is not help to making motivation between agencies for rising efficiency, efficient supply chain management, and CRM⁶ improvement. So reviewing this method to making dynamic method to reaching competition advantages in comparison with the other competitor by using hidden knowledge in reservation data base, is seems necessary.

In the way of knowledge discovery, attention to this items by using different data mining techniques, such as clustering, and classification are proposed:

- The count of sold tickets in any agencies in generally.
- The count of sold tickets in any agencies in yearly and monthly periods.
- The count of sold tickets in any ticketing area.
- The count of sold tickets in any agencies I any ticketing area in yearly and monthly periods.

After relation and knowledge discovery in this huge database, the airline commercial section can present new dynamic method for commission paying to agencies for reaching to this objects:

- making advance programming in one layer of SCM, consist of airplane company and ticketing agencies.
- Making competition atmosphere in ticketing agencies for selling this airline tickets.
- Making collaboration motivation for new entrance agencies to this market.
- Making better atmosphere for setting modern CRM in agencies to increasing the ticket selling.
- Making better atmosphere for seat chartering and tour operating by ticketing agencies.

And the object of this research is presenting the new hybrid models for data mining in any airplane reservation systems that likes to reach objects same to this paper objects in each countries.

2. Problem statements and research objects

The airplane companies in many countries seems with low income and in many times needy to governmental soubcide . this situation daily is going to worse, cos of increasing fuel price, maintenance price, and etc. so these companies try to decreasing the expense and presenting newer and better services. Maybe until last years, the ticketing agencies have not important roles in this market, but we think now cos of intense competition among airplane companies, they are playing basic roles. So presenting new attraction packages to them are so important.

One of these attraction offers is presenting new dynamic commission paying on the base of count of tickets sold in agencies and number of tickets booked in each reservation(group booking.

Now this company is paying 5 percent of ticket prices in domestic flight, and 9 percent of ticket prices in ongoing flight for commission to agencies. Execution of new dynamic commissioning is important factor for making attraction among ticketing agencies for rising investment on selling of this company tickets. This help to full the empty seats on many flights and increasing number of flights in some routes.

For reaching to these targets, we should answer to these questions:

- Are the agencies in special date in aspect of selling ticket have special behavior models.
- Are the agencies locality situations making the special selling behavior models.

⁶ Customer Relationship Management

- Are the count of tickets booked in each reservations has role in cities segmentation.
- Are the count of tickets booked in each reservations has role in agencies segmentation.

3. Research methods

The reservation data are saving in many tables in DB2 data bases that data patterns are unknown for our. We convert it to SQL database format, that is useable for our project in data mining software.

This paper act in two main steps:

- cities behavior modeling
- agencies behavior modeling in each groups of same cities.

3.1. Data preparation for cities behavior modeling

The data ware house include amount to 14 millions records . each record consist of many fields and each fields consist of many subfields. Each record is dedicate to one ticketing process by a customer.. that can include one or many passengers. 8 fields is useable for this research that present in Tab.3.1

Tab3.1 Input data

Field	Type
CODE	NUMERIC
OFFICE NUM	STRING
NUM OF PASSENGER	NUMERIC
AGENT NUM	NUMERIC
DATE	DATE
FLIGHT NUM	STRING
ROUTE	STRING

Likewise by researched we understood that 44 cities in Iran consist of offices for selling this company flights ticket. The name and uni code of those present in Tab3.2.

Tab. 3.2 Ticketing cities and theirs code

City Code	City Name	City Code	City Name	City Code	City Name
RJN	Rafsanjan	RAS	Rasht	THR	Tehran
KHY	Khoy	ZAH	Zahedan	ADU	Ardabil
GZW	Qazvin	SRY	Sari	ABD	Abadan
SDG	Sanandadj	CQD	Shahr-e-Kord	OMH	Uromieh
QUM	Qum	SYZ	Shiraz	IFN	Isfahan
BJB	Bojnurd	RZR	Ramsar	AWZ	Ahwaz
MRX	Mahshahr	YEH	Asaluyeh	BND	Bandar-e-abbas
ACP	Maraghe	GSM	Qeshm	BDH	Bandar-e-lengeh
SYJ	Sirjan	KER	Kerman	BUZ	Bushehr
GBT	Gorgan	KSH	Kermanshah	XBJ	Birjand
IKA	Imam Khomeyni	KIH	Kish	TBZ	Tabriz
YES	Yasouj	LRR	Lar	MHD	Mashhad
IHR	Iranshahr	ACZ	Zabol	AZD	Yazd
ILL	Ilam	JWN	Zanjan	ZBR	Chabahar
		AFZ	Sabzevar	KHD	Khoramabad

By using the data in Tab. 3.1, and using some procedure same as Derive, Select, Aggregate, and merge, the necessary fields for next step that is clustering to purpose of segmentation of 44 cities to 4 segment, that each cities of each segment has same behavior to others cities, is extracting. These fields present in tab3.3.

Tab. 3.3 Fields name and types for cities clustering

Type	Field
SET	City
RANGE	Number of ticketing offices in each cities
RANGE	Number of Flights in each cities
RANGE	Number of agents in each cities
RANGE	Number of passengers in each cities
SET	Population of each cities

3.2. Cities behavior modeling

In this step by using K-Means clustering modeling and data in Tab. 3.3 , the modeling is executing, and then the clustering results is sending to C&RT classifier model for rule induction. The result of rule induction represent in Tab. 3.4

Tab. 3.4 The result of cities behavior clustering

Cities code	Cluster
ABD, ADU, AZD, BDH, BND, BUZ, CQD, GBT, GSM, ILL, IKA, KER, KHD, KIH, KSH, OMH, RAS, SRY, TBZ, XBJ, ZAH, ZBR, LRR, SDG, YEH	1
THR	2
RJN, QUM, ACP, ACZ, AFZ, BJB, GZW, IHR, JWN, KHY, MRX, RZR, SYJ, YES	3
SYZ, AWZ, IFN, MHD	4

3.3. Data preparation for ticketing agencies behavior modeling in each cities segment

By using the data in Tab. 3.1, and using some procedure same as Derive, Select, Aggregate, and merge, the necessary fields for next step that is clustering to purpose of discovery ticketing agencies behavior in each groups of cities (4 groups) ,is extracting. These fields present in tab3.5.

Tab. 3.5 Fields for discovery ticketing agencies behavior in each groups of cities

نوع	فيلد
SET	Ticketing agencies code ⁷
RANGE	Sum of the sold tickets in each agencies per months ⁸
RANGE	Sum of the sold tickets in each segments per months ⁹
RANGE	Sum of the agents in each segments ¹⁰
RANGE	Maximum passenger in each tickets sold in each agencies per months ¹¹
RANGE	Maximum passenger in each tickets sold in each segment per months ¹²
RANGE	Year ¹³
RANGE	Month ¹⁴
RANGE	Proportion of each agencies tickets sold per months ¹⁵

3.4. Ticketing agencies behavior modeling in each cities segments

In this step by using K-Means, Kohonen, and TwoStep clustering modeling and data in Tab. 3.5 , the modeling is executing, and then the clustering results is sending to C&RT classifier model for rule induction. The results of rule induction represent that 2 clustering modeling (K-Means and Kohonen) are not useable for clustering of these data, cos the rules inducted by those are based on year and months, that are not compatible by this research objects. But the results of TwoStep modeling are suitable. The result of TwoStep clustering and rule induction by C&RT classifier in 4 cities groups presents in Tab. 3.6.

⁷ CITY_OFFICE_NUM

⁸ NUM_OF_PASSENGER_SUM

⁹ NUM_OF_PASSENGER_SUM_MONTHS

¹⁰ AGENT_COUNT

¹¹ NUM_OF_PASSENGER_MAX

¹² NUM_OF_PASSENGER_MAX_MONTHS

¹³ YEAR

¹⁴ MONTH

¹⁵ OFFICE_SOLD_PER_MONTHS

Tab. 3.6 Agencies behavior models in each cities segments

Passengers Counts	Records Counts	Rules Inducted	Agencies Clusters	Cities Segment
4,577,827	2,833	OFFICE_SOLD_PER_MONTHS) <=(0,010 OR NUM_OF_PASSENGER_MAX) <=23(1 st Cluster	1 st cities group
951,981	1,242	<=0,009 OFFICE SOLD PER MONTHS <=0,005	2 nd Cluster	
843,584	3,710	<=0,004 OFFICE SOLD PER MONTHS	3 rd Cluster	
2,465,124	4.272	OFFICE SOLD PER MONTHS < =0,003	1 st Cluster	2 nd cities group
4,850,713	7.767	<=0,003 OFFICE SOLD PER MONTHS <=0,002	2 nd Cluster	
4,577,827	10,139	<=0,001 OFFICE SOLD PER MONTHS	3 rd Cluster	
		No Important Rule Inducted		3 rd cities group
417,660	3,611	<=0,002 OFFICE SOLD PER MONTHS	1 st Cluster	4 th cities group
1,464,655	3,339	OFFICE SOLD PER MONTHS <=0,005 < =0,003	2 nd Cluster	
4,104,435	3,731	< OFFICE SOLD PER MONTHS=0,006	3 rd Cluster	

3.5. New dynamic commission calculating methods proposal for 4 cities groups

In this section for summarization only proposed New dynamic commission calculating methods for 1st cities groups. The other groups can be research same as it.

3.5. 1. New dynamic commission calculating methods proposal for 1st cities group

In this group we suggest that combine 1st and 2nd clusters passengers counts, that sum of those equal with 1,795,565 passengers. According to next equation, the count of sold tickets in 3rd cluster is 2.5 as many as the others.

$$4,577,827 / 1,795,565 = 2,5$$

So the airline company can divide domestic flights commission (5 percent) and foreign flights commission (9 percent) to new methods:

- a. in domestic flights
 - if (OFFICE_SOLD_PER_MONTHS < 0,010) then (commission = 3,75 percent of tickets price).
 - if (OFFICE_SOLD_PER_MONTHS > 0,010 or NUM_OF_PASSENGER_MAX>=32) then (commission = 5.5 percent of tickets price).
- b. in foreign flights
 - if (OFFICE_SOLD_PER_MONTHS < 0,010) then (commission = 7,75 percent of tickets price)
 - if (OFFICE_SOLD_PER_MONTHS > 0,010 or NUM_OF_PASSENGER_MAX>=32) then (commission = 9.5 percent of tickets price).

3.6. Hybrid data mining model for discovery new commission calculation methods in airplane companies

In attention that many airplane companies in countries are using the system reservation same as the reservation system in this paper (that proposed by SITA), so we can suggest new hybrid data mining model for discovery new commission calculation methods in the other airplane companies in all of countries. this hybrid model presents in Fig.3.1.

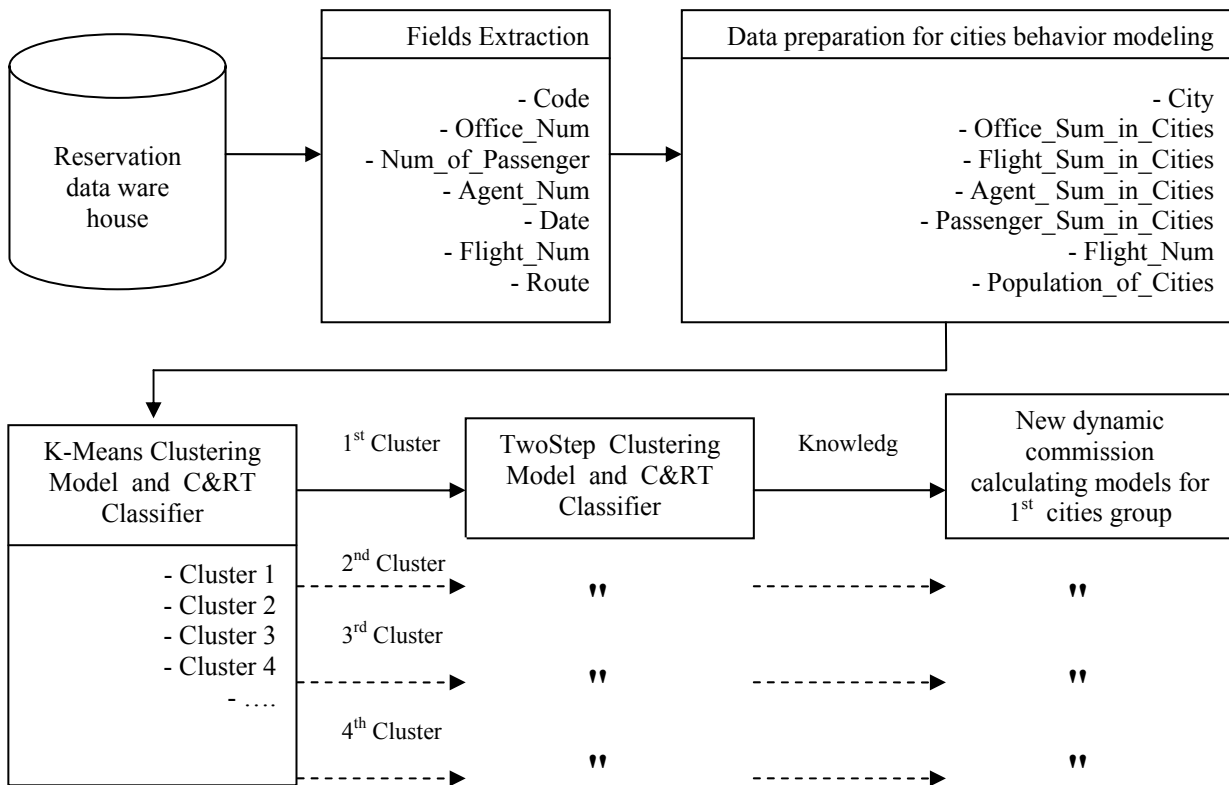


Fig.3.1. Hybrid data mining model

4. Conclusion

In this paper at first we investigated in ticketing reservation data ware house for presenting New dynamic commission calculating methods, that helps to Improving collaboration among an airplane company and ticketing agencies . and the final we suggest a hybrid data mining framework for discovering new commission calculation methods in airplane companies in the others countries.

References

1. Agrawal, R., Sivasubramaniam, N. & Simmons D. (2002). "Leveraging technology to improve field service". *"Service Industry Management"*, 13(1), 47-68.
2. Al-Hakim, L. (2003), "Web-Based Supply Chain Integration Model", in Mariga, J. (Eds) *"Managing E-Commerce and Mobile Computing Technologies"*, IRM Press, Melbourne, pp. 183-207.
3. 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
4. Cahlink, George (Oct. 2000). "Data Mining Taps the Trends," *"Government Executive Magazine"*.
5. 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.
6. Craig A. Knoblock, (2007). "Building software agents for planning, monitoring, and optimizing travel". *"Information Science Institute, University of Southern California, U.S.A"*.
7. Danskin & Etal , (2006). " Knowledge Management as Competitive advantage : Lessons From The Textile and Apparel Value Chain " , *" Journal of Knowledge Management"* . Vol.9 . No.3 . PP 91 - 102 .
8. Dholakia, Nikhilesh and Xiao, Jing J. and Dholakia, Ruby Roy and Mundorf, Norbert (2000). "The impact of retail ecommerce on transportation: A conceptual framework"
9. Ester M., Frommelt A., Kriegel H.P., & Sander J. (1998). "Algorithms for characterization and trend detection in spatial databases". *"Proceeding of 4th International Conference on Knowledge Discovery and Data Mining"*(pp, 44-50).
10. Fayyad, AI (1996). "Advances in knowledge discovery and data mining".
11. Fayyad, Usama and Piatetsky-Shapiro, Gregory and Padhraic, Smyth. (1996). *"From Data Mining to Knowledge Discovery in Databases"*.
12. Gunasekaran, A., Patel, C. & Tirtiroglu, E. (2001). "Performance measures and metrics in a supply chain environment". *"International Journal of Operation & Production Management"*. 21 (1/2), 71-87.
13. Hand, David and Heikki, Mannila and Padhraic, Smith (2001). "principles of Data Mining".
14. J.Han, and M.Kamber,(2001). *Data Mining: Concepts and Techniques*, San Diego Academic Press.
15. Krzysztof, J. Cios and Witold, Pedrycz. And Roman, W.Swiniaraski (2007). "Data Mining ,A Knowledge Discovery Approach". Springer.
16. 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).

17. Kopreski, K., & Han, J., & Stefanovic, N. (1998). "An efficient Two-Step method for classification and spatial data". *"Proceedings of the International Symposium on Spatial Data Handling"*. (320-328)
18. Kopreski, K. (1998). "A progressive refinement approach to spatial data mining. Ph.D. thesis, Simon Fraser University.
19. 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
20. M. Brian Blake ,(2006). A Component-based Data Management and Knowledge Discovery Framework for Aviation Studies ,Georgetown University and center for advance Aviation System Development , the MITRE Corporation ,USA.
21. 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.
22. Michael K. Painter , Brian Beachkofski , (2007). "Using Simulation ,Data Mining and Knowledge discovery Techniques for optimized Aircraft engine Fleet management ,Knowledge Based Systems ",Inc. , Air Force Research Laboratory USA.
23. Nazeri, Zohreh ,(2006). "Data mining of Air Traffic Control operational errors", Georgia Mason University, 4400 university Drive, Fairfax, VA 22030
24. Nazeri, Zohreh(2003), "Application of aviation safety data mining and workbench at American airlines, Center for advance aviation system development (CAASD)", The MITRI Corporation.
25. Ng, R., & Han, J. (2002), "CLARANACE: A method for clustering objects for spatial data mining" . *"IEEE Transactions on knoeledge and Data Engineering"*. 14(5), 1003-1017.
26. Peppard, J. (1995), "Broadening Visions of BPR: the Imperative of Strategic Integration", Cranfield University, SWP 11/95, Cranfield.
27. Poirier, C.C. and Bauer, M.J. (2000), "E-Supply Chain: Using the internet to revolutionize you business", Berrett-Koehler Publishers, San Francisco.
28. 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.
29. 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.
30. Stewart, G. (1997), "Supply-chain operations reference model (SCOR): the first cross-industry framework for integrated supply-chain management", *"Logistics Information Management Journal"*, Vol. 10, No. 2, pp. 62-67.
31. Tan, Pang-Ning & Steinbach, Michael & Kumar, Vipin. (2006), Introduction to Data Mining, Addison Wesley, ISBN: 0-321-32136-7
32. Towill, D.R. (1997). "The seamless supply chain: the predator's strategic advantage". *"International Journal of Technology Management"*, 13(1), 37-56.
33. Two Crows Corporation, (1999). "Introduction to Data Mining and Knowledge Discovery".
34. U.S. General Accounting Office, (May 2004). " Data Mining: Federal Efforts Cover a Wide Range of Uses". " *GAO Report GAO-04-548 (Washington)"*.
35. W. Seifert, Jeffrey.(2004) "Data Mining: An Overview, CRS Report for Congress"
36. Williams, L.R. Esper, T.L. and Ozment, J. (2002), "The electronic supply chain: Its impact on the current and future structure of strategic alliances, partnerships and logistics leadership", *"International Journal of Physical Distribution & Logistics Management"*, Vol. 32 No. 8, pp. 703-719.
37. Xu R., & Wunsch, D.II. (2005). "Survey of clustering algorithms". *"IEEE Transactions on Neural Networks"*, 16(3), 645-678.