Data Mining for Customer Relationship Management in an Airline

Fariborz Ghahremani[1], Behroz Minaei[2], M.Reza Davari[3], Bahram Kazempour[4]

[1,3,4] MSc. in Information Technology Engineering- eCommerce, Shiraz University
[2] Assistant Professor, Department of Computer Engineering, Iran University of Science and Technology

{Fariborz. Ghahremani, Mreza.Davari}@gmail.com, b_minaei@iust.ac.ir, kazempour_b@yahoo.com

Abstract. *In present decade, customer orientation has been one of the major concerns of commercial companies. Discovering the needs of customers and scheduling to meet those needs, or in other words customer relationship management (CRM), is an undeniable fact that can have an influence in customers' attraction and changing them into regular customers, which consequently leads to an increase in the company's interests. Data mining is one of the ways of discovering [1] the customers" needs. The purpose of this research which has been done on the data base of an airline's passengers reservation system, is to patterning the behaviors of domestic flight's customers per flight, in time intervals, using the time series model. Discovering these patterns and with exact scheduling (and by taking into account the number of planes in its fleet) an airline will be able to meet the demands of the customers properly. In this case the customer (passenger) is changed to a regular customer and will increase the company's interests in the long term. It is also possible to have the least number of open seats in each flight which lessens the company's detriment.*

Keywords

Data Mining, Customer Relationship Management (CRM), Flight Scheduling, Clustering, Time Series.

1. Introduction

The main focuses of the most industries have been turned on customer services. Discovering the real and main needs of customers and providing them with proper services according to their needs forms the basic rules of customer relationship management. Quick and in-time attention to the main needs of the customers is the most important factor in reducing the costs and increasing the economic benefits [2].
In the aviation industry, due to numerous number of flights in days, weeks and years in various destination, a great load of data which indicates traveling preference of people (including business or trips) in different dates (concerning seasons, climate, school holidays, national or religion holidays), is a good source in discovering the rules and the relationships between flight capacities and these preferences. By knowing these relationships, providing a more appropriate schedule and considering the future needs, it is possible both to satisfy the customers and also to reduce the airline's costs. On the other side, terms of attracting the customers [3] in the first and second cycle of market's lifespan become operational. In this research we will try to focus on those parts which relates to the customer relationship management to be relevant with the reservation and offering flight routes. The current mechanized system of offering tickets is over 5 years old which contains a great bulk of flight information, including flight lists, flight time and date, lists of passengers, description and information of passengers and number of open and booked seats per flight.
Nowadays, considering the emulating aspects of the markets, reducing the costs is of great importance in an airline, while it is not simple due to nature of such companies. So reaching a minimum of detriment per flight is exigent. In the entire airlines world wide, the companies don't use the maximum physical capacity of each flight for various reasons such as referring back the tickets, opening seats due to some policies, lack of accurate scheduling in flight routes or in choosing an airline with appropriate capacity per flight and such reasons incur a big loss for the company. In addition, exceeding number of passengers in some flight routes will lead to a decrease in the airline profits. These problems cause an inefficiency of customer relationship management in an airline [4] because the first problem increase the ticket's cost and the second displease the customer who receives no service from the company. In this research, we try to data mining through studying data of reservation's system. Then through exercising different models of data mining such as clustering and time series try to realize:

- Relationship between the time, flight routes and rate of demand for open seats per flight.
- Relationship between the time, flight routes and the open seats per flight.
- Relation between the time, flight routes and the probable passengers more than the capacity of the aircraft.

And offering an effective and practicable model of customer relationship management's system to conform with knowledge of real data.

2. Problems and research objectives

Concerning the scheduling of domestic flights and estimating the number of passengers or number of open seats, there is no special method for each flight in weekdays. This is only based on marketing or statistical methods and mathematical models for optimization of the flight.

The research objectives are as follows:

- Through data mining on data base related to passengers and canceling the flights, we can offer flight models with the
  minimum open seats,
- Through data mining of data bases related to passengers, it is possible to have flight models in which all passengers
  receive good services,
- Offering models for finding suitable flight routes for those times with more passengers.

3. Research Method

The required data for data mining are those with more than 5 years existence in reservation system of data base in an airline. The data is saved in 5 fields define with 5000 characters. Concerning the definition of 5 class compartments and their combination for each aircraft, flight data follow 7 different ways and subsequently engender new lines in each record which leads to variable length of each record and saved data.
To this end, we try to extracting data through transferring them from DB2 to SQL server 2005 and writing the query, those data include information about flight number, date, route and total passengers.

4. Preparation of data

Model is a small sample of real world [5], or model is explanation of real world [6]. Therefore one would say that logical knowledge is a kind of cognition from a series of data based on a series of distinguished regulation is usually obtained by induction method [7].
The data is never prepared to direct data mining [8]. So before modeling, they should be changed into the suitable format. The data on data base are saved in strings, regarding the intended model, they will change to the numeric format and the new data will be generated as follows:

- Aircraft capacity: numeric,
- Open seats: numeric,
- Flight route: string,
- Number of unanswered request: numeric,
- Total passengers: numeric,
- Percentage of open flights: numeric,
- Percentage of unanswered request: numeric,
- Type of aircraft: nominal.

To calculate the number of request more than the current capacity per flight the following formula was used:

$$Over\ Passenger\ Booked = -(To\_number\ (open) + To\_number\ (waiting\ list) + To\_number\ (request))$$

To calculate the percentage of open seats per flight:

$$Percent\ of\ Open\ Seats = (Open / Capacity) * 100$$

To calculate the percentage of unanswered request per flight:

$$Percent\ of\ Request\ Seats = (Over\text{-}passenger / Capacity) * 100$$

Finally the field's structure changed to the matrix format to carry out the time series model.

5. Suggested Model

After extracting and preparing data related to the domestic flights, through using the clustering model "TwoSteps", we classify flights based on routes and assign the percentage of unanswered request and open seats per flight. To clustering the flight route based on total passengers, open seats and unanswered request per flight, the "TwoStep" model was used.

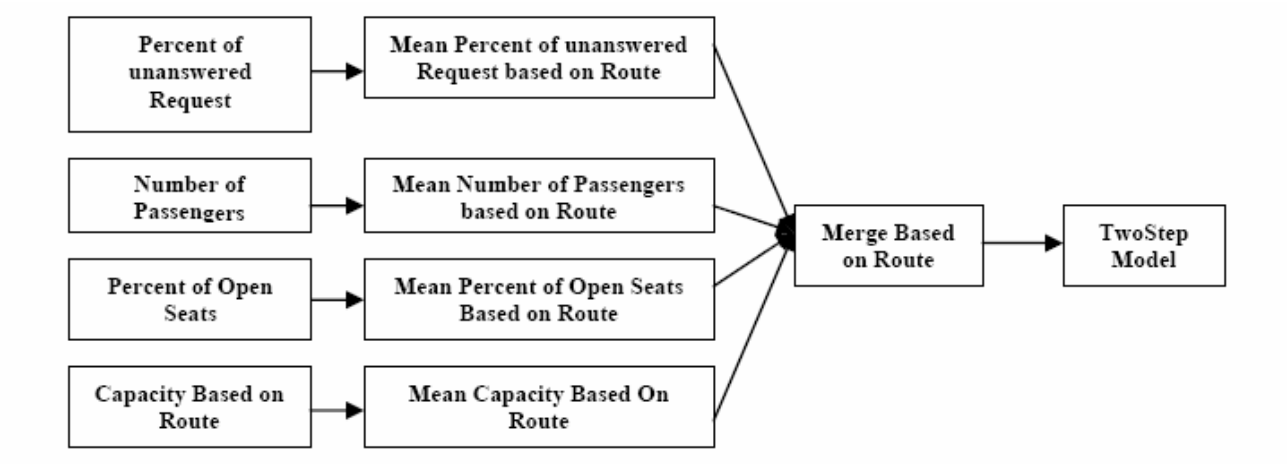

Fig.1. Model for clustering

According the above model, the data was clustered in 2 groups. (In Fig.2, visualization of 2 clustered to understand better [9]).

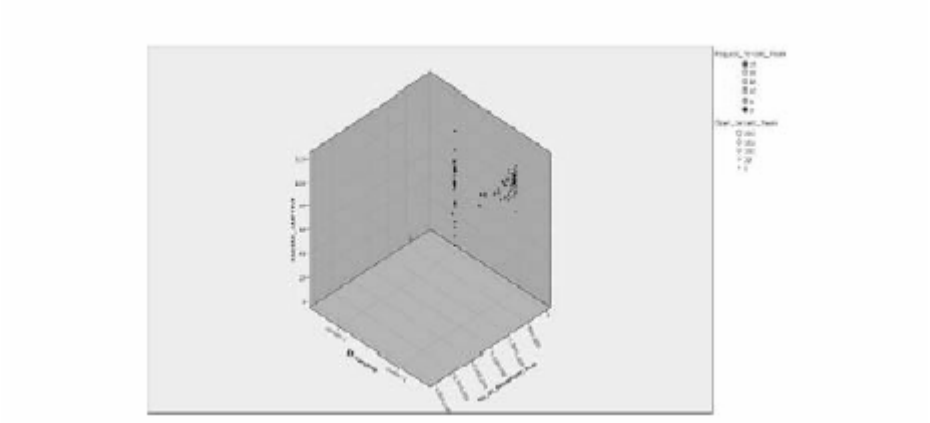

Fig.2. Visualization of clustering

After carrying out the above model, the rules of clustering were identified through the C&RT model and the confidence rate was studied.

Tab.1. Results for output field TwoStep

Correct	119	99.17%
Wrong	**1**	0.83%
Total	120	

Through classifying the routes in 2 clusters, the first one shows the numerous request and passengers in routes, but the second cluster shows no request for open seats and total passengers are few.

Number of open seats and unanswered request for each route were calculated in daily time span.

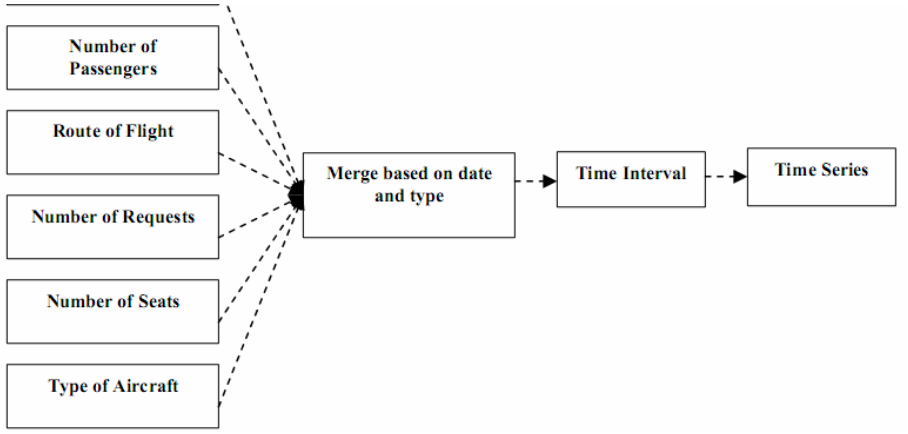

Fig.3. Model for Time Series

There are different models in time series [10] and with due attention to the conformity rate of model with the diagram, the simple season model in the exponential smoothing method was selected. The simple season model was used because of flight's seasonal behavior and stationary R- square.

Tab.2. Analysis of Time Series model

Target	**Model**	**Stationary** R**2	Q
SUM of Passengers	Simple seasonal	0.587993	50.978
Mean of Open Seats	Simple seasonal	0.605825	124.14
Mean Of Request Seats	Simple seasonal	0.727991	140.87

Through this model, the number of request and open seats was predicted in the first cluster for the following 7 days. (Fig.2)

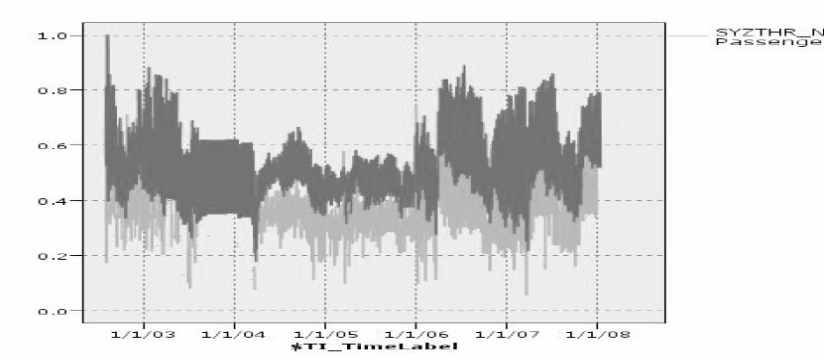

Fig.4. Prediction of passengers for next 7 days

The obtained gain from the first cluster is based on the diagram. (Fig.5)

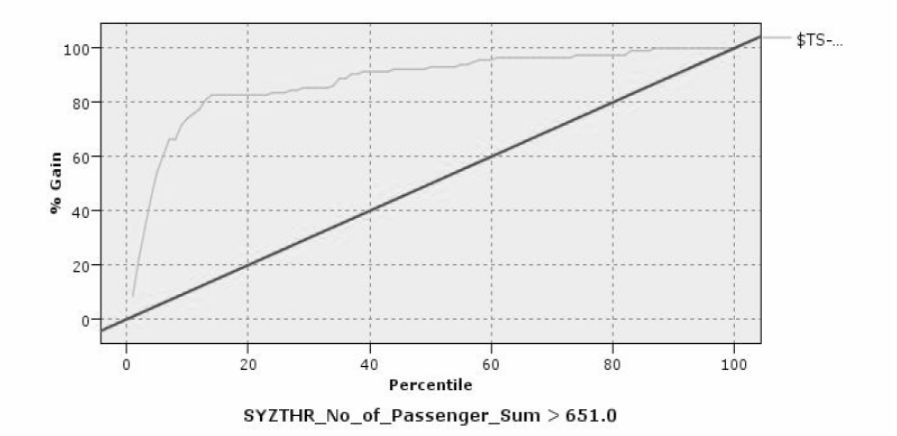

Fig.5. Gain of Prediction

Concerning the few number of passengers in the second cluster, it is not possible to estimate daily request and only the total passengers per month can be predicted.

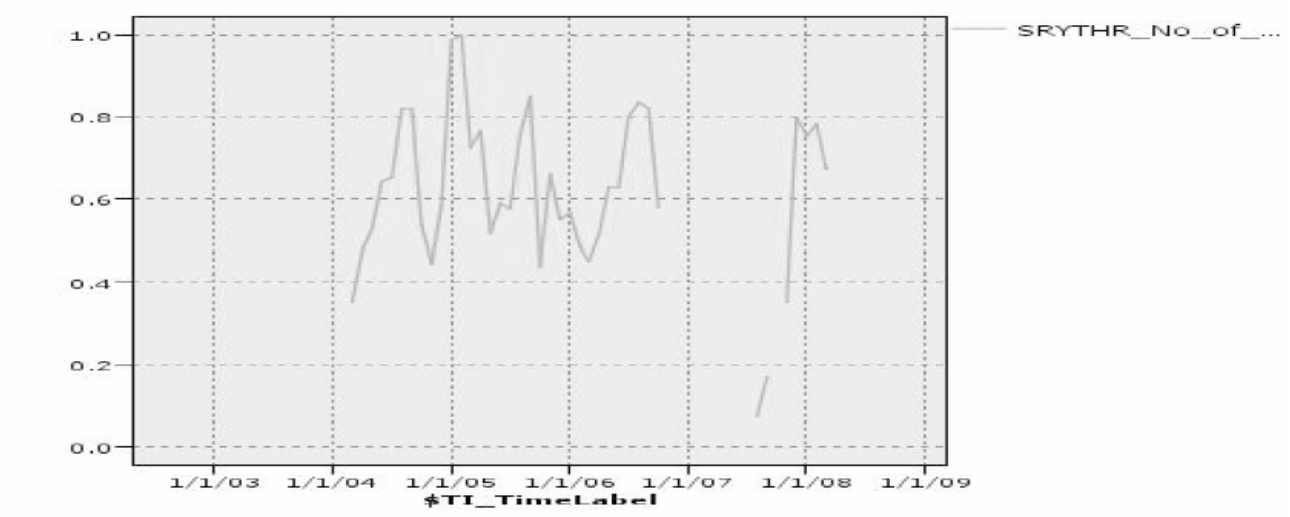

Fig.6. Number of passenger per month

6. Conclusion and Future research

Through clustering, 2 different clusters showing the request rate based on each flight routes in time intervals were obtained. In the first cluster which shows the great request rate for domestic flights, estimating the number of passengers and open seats per routes was carried out daily. But with request in weekdays in all routes, it is possible to predict the number of passengers monthly.

So according to the customer's behavior the request for each route and also number of passengers and open seats can be predicted.

For future studying, it is recommended to define over requested times through considering each flight times. Also through considering the type of predominant aircraft in each route, the connected flight's legs, flight time and estimating number of passengers, it is possible to reduce the experimental rate of manpower and have an automatic daily schedule.

References

[1] Pang-Ning Tan, Michael Steinbach and Vipin Kumar (2006)." *Introduction to Data Mining", Addison Wesley",* ISBN: 0-321-32136-7

[2] M.H. Nikbakhsh Tehrani, (2001). *"Electronic Commerce"*, Tehran: Isiran institute.

[3] Bose and Sugurmaran (2003), *"Application of Knowledge Management Techniques inCustomer relationship Management",* Journal of Knowledge and Process Management, Vol.10, No. 1. pp.3 - 17.

[4] Lisaand Pritscher and Hans Fey en, (2004). *"Data Mining And Strategic Marketing In TheAirline Industry",* Switzerland, CH-8058 Zurich-Airport.

[5] S.M.T. Roohani Rankoohi, (2004). *"Fundamental Concepts of Databases",* Tehran,Jelveh.

[6] Giudici. Paolo (2003). *"Applied Data Mining Statistical Methods for Business and Industries",* Wiley. London

[7,8] M. Kantardezik, (2006). "Data Mining", Babol: Uloome Rayane

[9] FROST, R.A. (1987). *"An Introduction to knowledge Base Systems",* Collins. London.

[10] A. Adel, (2001). *"Statistic and Application in Management",* Tehran: SAMT.

.