A Data Warehouse/Online Analytic Processing Framework For Web Usage Mining And Business Intelligence With Fuzzy Technique

Dr. R.P. Bhagat*, Akhilesh Kumar**

ABSTRACT

There is less research literature about the business intelligence technology applied in airports and airline companies. This paper mainly summarized and commented on the business intelligence technology applied in domestic and international airports and airlines companies especially focus on the data warehouse and data mining. In the data warehouse application, radio frequency identification technology enhanced fuzzy neural network technology and online analytical processing technology and other key technology have been applied in the daily management of business data of airports and airline companies. And in the data mining application, clustering, classification, decision tree, regression analysis and fuzzy detection analysis have been used in the management of airports and airlines. In recent years, the application of business intelligence technology in airports and airline companies has achieved satisfactory results. The summaries and comments in this paper aim to provide the reference and help to relevant field of researchers

Keywords: Business intelligence; Data warehouse; Data mining; Airlines; Airports; Fuzzy neural network.

INTRODUCTION

Business intelligence (BI) refers to the use of data warehouse technology to store and manage operational data, and through a variety of statistical analysis tools and data mining techniques to analyze operational data to provide a variety of analysis reports which can offer the support information for a variety of business decision-making activities.

Business intelligence is the enterprise use of modern information technology to collect, manage and analyze structured and unstructured data and business information to create and cumulate business knowledge and insights, to improve

* Dept. of Mathematics A. M. College Gaya(M.U. Bodh Gaya)

** Research Scholar (M.U. Bodh Gaya), akhilesh2gaya@rediffmail.com

business decision making, to carry out effective business operations, to improve various business processes, to promote the business performance, and to enhance the overall competitiveness of the wisdom and ability. Typically enterprise business intelligence analysis, including business dashboards, scorecards, ad hoc query, reporting and operational analysis and forecasting, business strategy implementation monitoring and other topics. Through the data warehouse modeling and centralized management of KPI, it can solve the problems of performance of business operations, to monitor the development of enterprises, to present a complex reports in a simple way. In the existing business intelligence systems, statistical analysis capabilities needed to help analyze customer segments, predict customer behavior, predict trends in customer service, and identify fraudulent behavior. From the perspective of aviation and airport industry, business trends is the aviation-related business revenue growth is less than business revenue growth, and how to analyze these business trends and the relationship between aviation-related business revenue and business revenue, is the major research contents in business intelligence analysis.

DATA WAREHOUSE AND INFORMATION COLLECTION

Data warehouse is a kind of data collection which is subject oriented, integrated, relatively stable, changing with time (different time) to support management in decision-making process, to help the data be geared to the needs of the subject, and to correspond to traditional database application. Subject will be a higher level of data classification standards for each subject area corresponds to a macro-analysis: integrated nature of the data warehouse is that before the data entering into the data warehouse, the data must be processed and integrated, which is a key step in the establishment of data warehouse. The contradictions in the raw data should be unified, but also the raw data should be changed from the application-oriented structure to the subject-oriented. Stability of the data warehouse is that data warehouse reflects the historical data, rather than data generated in daily transaction. After processed and integrated data entering into the data warehouse, there is little or no modification. Data warehouse is a collection of data at different times. It requires that retention of the data in data warehouse must meet the needs of decision analysis, and data must indicate the historical period of the data.

With the data warehouse technology application deep going in recent years, data warehouse technology develops rapidly. Many companies already well accepted 'the integration of data, find knowledge from the data, use of data, and explanation with the data' and other new concept related to the improvement of all aspects of

production, improvement of production efficiency and the development of the productive forces. Both domestic and foreign experts have done some research about data warehouse applications in airports and airlines. Harry K.H. Chow, King Lun Choy, W.B. Lee, K.C. Lau (2006) proposed the management of the airport's cargo through the application of RFID-based resource management system. By using Radio frequency identification, case-based reasoning technology to improve resource management efficiency, the practice prove that using data warehouse technology can improve the management efficiency, save time and reduce the cost. T. C. Poon, K. L. Choy, Harry K. H. Chowa, Henry C. W. Lau, Felix T.S. Chan, K. C. Ho (2009) used radio frequency identification technology based on logistics and resource management system to manage data. Radio frequency identification technology is used to get the airport cargo data to test the performance of RF equipment. Three objectives are achieved: (i) a simplication of RFID adoption procedure, (ii) an improvement in the visibility of warehouse operations and (iii) an enhancement of the productivity of the warehouse. At the end of the article, author gave the technology high evaluation and he thinks that the technology is available in practice. S G Li, X. Kuo (2008) proposed an enhanced fuzzy neural network (EFNN) technology based on the central warehouse management reserve inventory decision support system. Traditional neural networks are found to be suffered from the problem of low accuracy of forecasting unseen examples. Therefore, in this EFNN, the following improvement is made: First, it assigns connection weights based on the fuzzy analytic hierarchy process (AHP) method without painstakingly turning them. Second, by generating and refining activation functions according to genetic algorithm, EFNN can provide comprehensive and accurate activation functions and fit a wider range of nonlinear models. Last, but not least, an adaptive input variable is introduced to decrease the impact of the bullwhip effect on the forecasting accuracy.

Chuan Lu, Wenli Yan, Xinjie Hu (2010) pointed that passengers should be divided based on customer value under the analysis of domestic frequent passengers being insufficient. The aviation customer relationship management data warehouse is designed and decision support is provided to customers through SPSS data mining. A homogeneous source data is provided for airline customer relationship management and marketing. With data warehousing and data mining technology to analyze the data on passengers, this information can promote the competitive advantage of the airline company. Weibin Li, Jiafeng Yang, Baobao Wang (2010) proposed a aviation maintenance information analysis system model. This model can be applied to aviation maintenance information on-line analytical processing, and use the decision tree and

association rules, data mining, aviation maintenance information to achieve the intelligent analysis.

DATA MINING AND INFORMATION ANALYSIS

Data mining is the use of mathematics, statistics, artificial intelligence and machine learning, to extract implicit, unknown and potentially useful knowledge from large, incomplete, noisy, fuzzy and random data. Data warehouse technology is a series of new independent application technology based on the needs of business information systems and the development of database systems. The essential difference between data mining and traditional data analysis methods (queries, reports, statistics and online analytical processing OLAP) is that data mining digs and finds that knowledge in the absence of clear information of assumptions. There are three characteristics in data mining model which are implicit, unpredictable, potential value. In most cases, data should be first put in the data mining database or data mart.

The data obtained directly from the data warehouse for data mining has many advantages. The data cleaning in data warehouse and in data mining is similar. If the data has been cleaned up in the import data warehouse, it's probably no need to clean up once in doing data mining and all the data inconsistencies have been resolved. Data mining database may be a subset of logic, and not necessarily have been physically independent database. However, if the data warehouse computing resources already strained, it is best to create a separate data mining database. Of course, data mining also do not have to build a data warehouse, data warehouse is not required. To build a huge data warehouse, unify the different sources of data together to resolve all data conflicts, and to guide all the data into a data warehouse is a huge project, which may take several years and cost millions of money to complete. If only for data mining, one or more read-only transaction database can be guided into a database. It can be regard as a data mart and data mining can be carried out in the data mart.

Some journals about data mining applications in airports and airlines in the domestic and oversea have been published in the literature. Viet Van Pham, Jiangjun Tang, Sameer Alam, Chris Lokan, Hussein A. Abbass (2010) found that the CO2 concentration in some parts of Australia is much higher than other parts, especially in some majorities from the temporal and spatial analysis of fuel usage rate and emissions data. The emission results also show that NOx emission of aviation may have a significant impact on the ozone layer in the upper troposphere, but not in the stratosphere. It is expected that with the availability of this real time aviation emission

database, environmental analysts and aviation experts will have an indispensable source of information for making timely decisions regarding expansion of runways, building new airports, applying route charges based on environmentally congested airways, and restructuring air traffic flow to achieve sustainable air traffic growth. FeyzaGürbüz, LaleÖzbakir, HüseyinYapici (2009) applied the data mining application on the incident data reports of the Federal Aviation Administration Accident Data System database. The categorization tools and decision trees are used to find the relations and rules about the incidents from 2000 to 2006 resulted in fatality. The accuracy and reliability of this method are valid in practice through analysis of the data. Shijin Wang, Dong Sui, Bin Hu (2010) used gray mixed model to predict air flow. With the historical data analysis of the national-wide air traffic from the year 1985 to 2008, the forecasting applicability is examined with three different methods, such as the time series forecasting, regression forecasting and neural-network forecasting. The result shown that predict accuracy of gray mixed model was better than the other three models.

FeyzaGürbüz, LaleÖzbakir, HüseyinYapici (2011) explored the use of different pretreatment techniques in handling airline data. This paper focuses on different preprocessing and feature selection techniques applied on the 15 component reports of an airline company in Turkey to understand and process the dataset. Regression analysis, anomaly detection analysis, find dependencies and rough sets are used in this study in order to reduce the dataset. Also the classification techniques of data mining are used to predict the warning level of the component as the class attribute. For this purpose, Polyanalyst, SPSS Clementine, Minitab and Rosetta software tools are used.

METHODOLOGY

The objective of the project is to develop inteligent methodologies for improving the accuracy of mobile positioning in the cellular communications system. The monintring system is trageted at a real time implementation using logitude and latitude techechlogy. The primary deciplines which one needs to Bridge Monitoring System (BMS) in order to develop a successful and an accurate mobile location are mobile communicaions and Fuzzy Logic and Fuzzy Control. Knownledgge of these discipline is need to analyse and develop algorithms capable of existing information about logitude and latitude. The structure of the controller will take a from of an adaptive learning for tracking the mobile movement. Such a controller will be developed using either or a combination of fuzzy logic and neural networks.

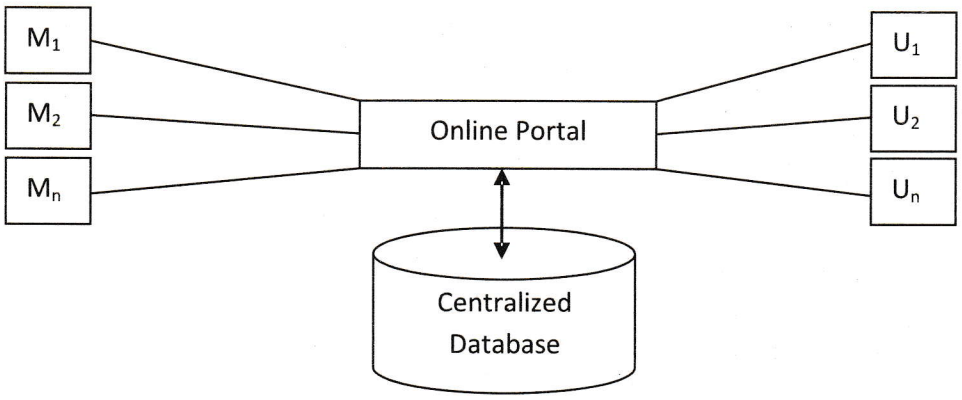


Fig-1 : Direct trust Model

Components

M_1, M_2, M_n : Mobile users

U_1, U_2, U_n : End users

Online Portal : A portal on which senior project engineer, Project engineer and online users.

Data representation:

Rating of x_j denoted by $case_{ij}$ is thus triple in the flowing from: $case_{ij} = <e_j, e_j, t>$ where i is the ith rating. In this triple, e_j =(a, b, x_j, v_j) denotes that user a expects x_j 's value is vj when assigning the where v_j ? [0, 1] t is time when the rating is recorded.

Interaction History:

A matrix I = H x n where H represents transcations while n represents attributes. I is a queue of the past H interactions. Each column (for each attribute x_j ? X) in the matrix is represented as

CB_j = { $case_{1j}$, $case_{2j}$ ....., $case_{nj}$,}

Direct Trust fuzzy logic:

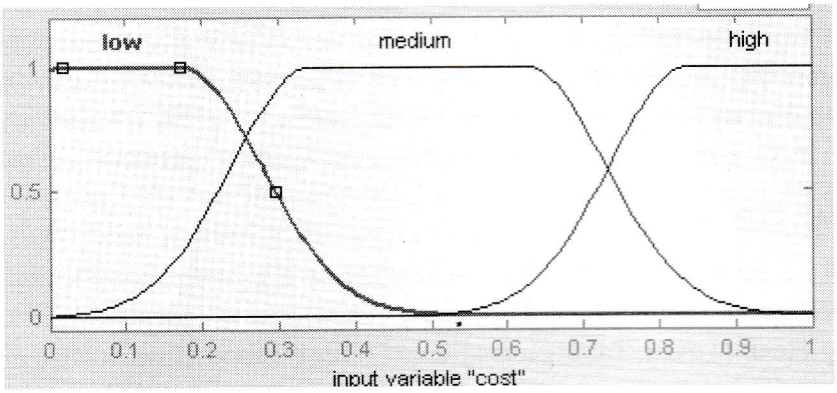
Let R(caseij) be the rating given by the senior project engineer for the jth interaction and ith attribute, Bij be the bridge and f(caseij) be effective rating.

$f(case_{ij}) = B_{ij} * R(case_{ij})$

The value of Bij is assigned depending on the transaction's and bridge updation. The assignment of Bij is done using fuzzy logic as explained below. Membership functions

are assigned for bridge progress.

Membership function plots



Input variable bride points

Fig-2 : Membership function for Bridge

The fuzzy rules are designed as follows:

- If(difference is low) and (bridge point is low) then (bridge_adjustment is positive)
- If (difference is medium) and (bridge point is medium) then (bridge_adjustment is constant)
- If(differnce is high) and (bridge point is high) then (bridge_adjstment is decrease)

Min() function is used for "and" operation

$$\mathbf{Ti(bridge,cost) = min\{\mu(bridge), \mu(cost)\}}$$

Where μ(bridge) is the membership degree of bridge of transaction monthly progress and μ(cost) is the membership degree of cost of transaction cost.

Max() function is applied for each output mobile user M. with 3 rules the final membership degree T(M) will be as follows:

$$\mathbf{T(M) = max\{ \mu1(M), \mu2(M), \ldots\ldots\ldots, \mu n(M),\}}$$

BAYESIAN NETWORKS

Bayesian Networks are directed acyclic graphical models combined with probabilities that follow the rules of the probability theory. Probability theory establishes a set of cause-effect relationships where the nodes are connected by directional arcs, ensuring

that the system as a whole is consistent, and providing ways to interface models to data. The nodes represent random variables, and the relationships represent probabilistic dependencies between variables. These dependencies are quantified through a set of Conditional Probability Tables (CPTs). Each variable is assigned a CPT of the variables acting as its parents. For variables without parents, this is an unconditional distribution. The basic concept in Bayesian Networks is using Bayes' rule for conditional probabilistic inference. Equation [1] gives a basic description of Bayes' rule. If B1, B2, ⋯, B k are the possible scenarios with an effect on the event A,

EQUATION PUT HERE

where P (B j | A) is the conditional probability of B j given that all we know is A. Bayesian Networks are used for diagnosing real world problems where uncertainty and incomplete data exist. Taking advantage of this method, defect cause probabilistic analysis is applied in the proposed bridge inspection system.

CONCLUSION

With the continuous development of airports and airlines, airports and airlines will continue to increase in volume of business. The amount of various data of airports and airlines becomes very large, large amount of data can affect the speed of the decision maker. Business intelligence technology applied to airports and airlines will assist decision-makers to make more timely and effective, more scientific and reasonable decision-making. Decision-makers' decision-making will be on a higher level.

REFERENCES

[1] A.S. Brooke, R.E. Caves, D.E. Pitfield. (2013) Methodology for predicting European short-haul air transport demand from regional airports: An application to East Midlands International Airport. Journal of Air Transport Management, Volume 1, Issue 1, March 1994, Pages 37-46

[2] Feyza Gürbüz, Lale Özbakir, Hüseyin Yapici. (2011) Data mining and preprocessing application on component reports of an airline company in Turkey. Expert Systems with Applications, Volume 38, Issue 6, June 2011, Pages 6618-6626.

[3] Haytham Remawi, Paul Bates, Ian Dix. (2011) The relationship between the implementation of a Safety Management System and the attitudes of employees towards unsafe acts in aviation. Safety Science, Volume 49, Issue 5, June 2011, Pages 625-632.

[4] Viet Van Pham, Jiangjun Tang, Sameer Alam, Chris Lokan, Hussein A. Abbass. (2010) Aviation emission inventory development and analysis. Environmental Modelling &

Software, Volume 25, Issue 12, December 2010, Pages 1738-1753.

[5] Weibin Li, Jiafeng Yang, Baobao Wang. (2010) Data warehouse in aviation maintenance in the application of information analysis. Electronics Technology, 2010, (9).

[6] Shijin Wang, Dong Sui, Bin Hu. (2010) Forecasting Technology of National-Wide Civil Aviation Traffic. Journal of Transportation Systems Engineering and Information Technology, Volume 10, Issue 6, December 2010, Pages 95-102.

[7] Chuan Lu, Wenli Yan, Xinjie Hu. (2010) Based on data warehouse aviation customer relationship management. Microcomputer Applications, 2010, (4)

[8] Feyza Gürbüz, Lale Özbakir, Hüseyin Yapici. (2009) Classification rule discovery for the aviation incidents resulted in fatality. Knowledge-Based Systems, Volume 22, Issue 8, December 2009, Pages 622-632.

[9] Zhaobin Liu. (2009) Cargo data warehouse system design and implementation. Computer System, 2009, (7).

[10] T.C. Poon, K.L. Choy, Harry K.H. Chow, Henry C.W. Lau, Felix T.S. Chan, K.C. Ho. (2009) A RFID case-based logistics resource management system for managing order-picking operations in warehouses. Expert Systems with Applications, Volume 36, Issue 4, May 2009, Pages 8277-8301.

[11] J. A. D. Atkin, E. K. Burke, J. S. Greenwood, et al. (2008) On-line decision support for take-off runway scheduling with uncertain taxi times at London Heathrow airport. Journal of Scheduling, 2008, 11(5):323-346.

[12] Kewu Yan, Lei Zhang, Qiang Sun. (2008) ID3 decision tree classification algorithm in the aviation market in the customer value segmentation application. Business Studies, 2008, (3).

[13] K. W. Yan, J. F. Zhu, S. Qiang. (2007) The application of ID3 algorithm in aviation marketing. Nanjing, China, 2007:1284-1288.

[14] Harry K.H. Chow, King Lun Choy, W.B. Lee, K.C. Lau. (2006) Design of a RFID case-based resource management system for warehouse operations. Expert Systems with Applications, Volume 30, Issue 4, May 2006, Pages 561-576.

●