

An Approach to Construct the Fuzzy Data Mart using Soft Computing

A.Prema

Ph.D Scholar, Research and development,
Bharathiar Bharathiar University,
Coimbatore, TamilNadu, India

A.Pethalakshmi, PhD.

HOD, Department of Computer Science,
MVM Government Arts College(W),
Dindigul. TamilNadu, India.

ABSTRACT

Data Mart is emerging as one of the hottest areas of growth in global business and the critical components and tool for business intelligence. A data mart is a local data warehouse used for storing business oriented information for future analysis and decision-making. In business scenarios, where some of the data are fuzzy, it may be effective to construct a warehouse that can strengthen the analysis of fuzzy data. This paper presents a Fuzzy Data Mart model that imparts the exile interface to the users and also extends the DWs for storing and managing the fuzzy data along with the crisp data records. We have proposed, an algorithm to design data mart, which improves the decision making processes. To do so, we use Extraction, Transformation and Load (ETL) tools for better performance. In addition to that, the membership function of fuzzy is used for summarization.

Key Fields: Data Mart, Data repository, ETL, Fuzzy logic, aggregation, summarization and decision making.

1. INTRODUCTION

This paper proposes the model of a fuzzy data mart, which allows integration of fuzzy concepts directly into the data warehouse model. By using this approach, the concept of summarizability is not affected in dimensions, as the fuzzy concepts are rolled out in a meta-table structure. But a **data** mart collects data on a specific subject area such as sales or production or customer information. It can be a subset of company data warehouse and it is proposed to meet the desires of a single department. An operational data store (ODS) is an updatable set of integrated data used for enterprise-wide strategic decision making. It consists of live data, not snapshots, and has least history retained. The most significant profit of data mart is a single view of data, whose result is more accurate information and it enables better decision making. A data mart is a persistent physical store of operational and aggregated data statistically processed data that supports businesspeople in making decisions based primarily on analyses of past activities and results. A data mart contains a predefined subset of enterprise data organized for rapid analysis and reporting. In this paper we have proposed a new algorithm (named as quick algorithm) for fuzzy data mart in order to improve the decision making in sales department in a company.

Section 2 of this paper deals with related work done in the data mart, fuzzy and ETL. Section 3, explains the steps of proposed work to build fuzzy data mart, in section 4, Experimental analysis and result are given, and finally, section 5 presents a conclusion of this paper.

2. RELATED WORKS

Different approaches for the integration of fuzzy concepts in data warehouses have been proposed. However, if fuzzy concepts are not normalized in such approaches, inconsistent results on aggregation may be produced and can harm the concept of summarizability [1]. A data warehouse gives a set of numeric values (called facts) that are based on a set of input values in the form of dimensions [2]. It [3] introduces fuzzy concepts in the highest granularity of dimensions [Delgado et. al]. Over the years data warehouse technology has been used for analysis and decision making in enterprises [4]. Fuzzy dimensions are introduced in which a dimension attribute is composed of a fuzzy concept [5]. Numeric values of a classical data warehouse can be difficult to understand for business users, or may be interpreted incorrectly. Therefore, for a more accurate interpretation of numeric values, business users require an interpretation in meaningful non-numeric terms. However, if the transition between terms is crisp, true values cannot be measured and smooth, transition between classes cannot take place [6]. To address this problem, fuzzy concepts as linguistic variables can be integrated with a data warehouse for the interpretation of numeric values [7]. A data mart contains data from a particular business area and multiple data – marts can form a data warehouse. The following are the reasons for building a data mart[8].

Easy access to frequently needed data

- ✓ creates collective view by a group of users
- ✓ improves end-user response time
- ✓ provides ease of creation
- ✓ brings about lower cost than implementing a full Data warehouse

Data Mart can hold information which addresses both strategic and tactical information needs and provides information which allows key operating functions to effectively manage performance. It unifies information from various databases into a single database.

Extract, Transform and Load (ETL):

All tables except the reference table are transferred to the Data Mart using an ETL process [9]. Extraction is the process of reading data from a specified source database and extracting a desired subset of data. Transformation is the transformation phase which applies a series of rules or functions to the extracted data to derive the data to be loaded.

Three types of transformations are used, namely, Subsets of tables, Formatting Data and Primary Keys and Indexes. Subsets are created to remove personally identifiable

information so that users with lower level access can query the data and split the tables into smaller subsets to expedite queries. ICD9 codes are transformed into the industry standard format. Primary keys are created to ensure uniqueness within a table and to facilitate the joining of tables. Indexes are created to expedite queries. Loading is the process of writing the data into the target database.

The ETL process[10] includes designing a target, mapping sources to target, extracting data from sources, transforming data for the target, scheduling and monitoring processes, and managing the over all BI environment. This ETL Tool is used to simplify the process of migrating data, standardize the method of data migration, store all data transformation logic/rules as Meta data which enable the users, managers and architects to understand, review, and modify the various interfaces and reduce the cost and effort associated with building interfaces. The purpose of using ETL Tools is to save time and make the whole process more reliable. Weaknesses of traditional Extract, Transform and Load (ETL) tools' architecture of its openness and repeated developments are analyzed and proposes a three-layer architecture based on metadata. That makes ETL process more efficient, multi purposeful and flexible. Through intensive studies of ETL theory, the drilling data warehouse ETL tools were designed and implemented [18]. The ETL tools are customized to provide the functionality to meet the enterprise requirements. Hence, many of them choose to build their own data warehouse themselves. To reduce the risk and framework for building a data warehouse can be used. This approach involves less time, cost, and risk than building your own data warehouse (Guerra et al., 2007). A framework can be customized to meet certain requirements. Therefore, obtaining a framework is the safety approach as it acts as a guide throughout the building process to compare and build the data warehouse from a scratch. Fuzzy sets are one proposed for managing vagueness and can be used to help make analyses, perhaps ironically, less fuzzy because vagueness is managed formally [16]. The membership function is the fundamental quantity [18] necessary to use fuzzy sets. [17] Aims at the problem of measuring the steam consumption in the Dyeing process. A multiple neutral network soft sensing modeling of Dyeing steam consumption based on adaptive fuzzy C-Means clustering (FCM) is presented. That method is used for separating a whole real-time training data set into several clusters with different centers, and the clustering centers can be modified by an adaptive fuzzy clustering algorithm. Each sub-set is trained by radial base function networks (RBFN), then by combining the outputs of sub models to obtain the final result. After designing the basic scheme of data mart, by using this approach, the concept of summarizability [11] is not affected in dimensions as fuzzy concept are rolled out in a Meta table structure. The ETL [12] are used to extract the business oriented data from the organization's transaction oriented operational database and load into the data mart.

However the decision making [13] plays a vital role in business in order to earn more profit. Fuzzy membership logic uses the whole interval between 0 and 1 to describe human reasoning. Petrovic states that fuzzy logic can be used to improve decision making and this can be used to meet the multi-objectives of put-away activity in warehouse operation[14]. The rules use the input membership values [15] as weighing factors to determine their influence on the fuzzy output sets of the final output conclusion. Once the functions are inferred, scaled, and combined, they are defuzzified into a crisp output which drives the system.

3. PROPOSED WORK

Step I:

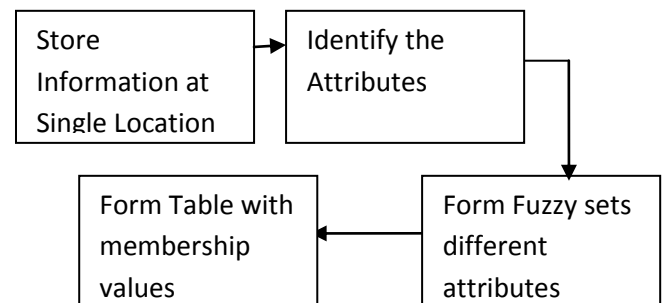
Input the data from heterogeneous source, which it may appear with mal value and redundancy and null.

Table:1

Icode	Name	Quantity	Price	Place
223	Paper	289	25	Erode
222	Pen	500	20	Trichy
A11	Marker	100	5	Dindigul
223	Paper	289	25	Erode
223	Sketch	100	20	Theni
113	Sharpener	300	5	Madurai
1A9	Scale	110	10	Madurai
555	Color pencil	700	20	Dindigul
61C	Painting box	200	Null	Chennai
210	Grinder	100	10	Kovai
Null	Book	300	50	Madurai
414	Brown cover	1000	5	Dindigul
523	Reynolds	700	15	Theni
127	Null	100	10	Erode
113	Sharpener	100	20	Kovai

Step 2:

Block diagram of pre-processing stage



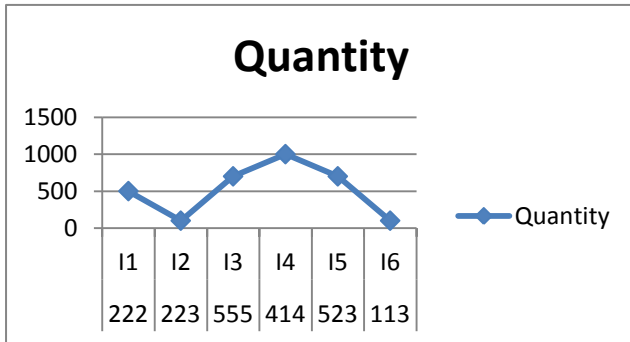
After preprocessing,

Table 2.

ICode	Name	Quantity	Price	Places
222	Pen	500	20	Trichy
223	Sketch	100	20	Theni
555	Color pencil	700	20	Dindigul
414	Brown cover	1000	5	Dindigul
523	Reynolds	700	15	Theni
113	Sharpener	100	20	Kovai

After preprocessing, data looks like

Figure.1.

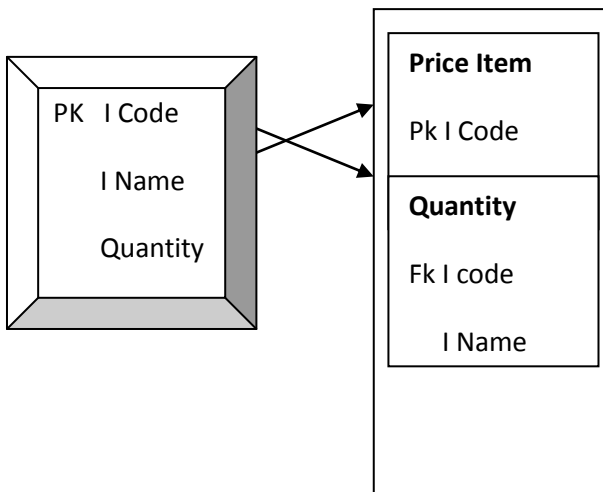


Step 3:

Based on step 2, create a fact table which is also set the dimension of the source and put the chart aggregates.

Fact table

Dimension Table



After aggregation, extract the data from the source and apply the transformation. Then load into the data mart.

Step 4:

Calculate to fuzzy membership μ_{ij} using Quantity.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m-1}}}$$

where, $d_{ij} = \|x_i - v_j\|$,

$\forall i = 1, \dots, n, \forall j = 1, \dots, c$

Rule 1: if S is less then Q is low.

Rule 2: if S is medium then Q is average.

Rule 3: if S is more then Q is high.

Assign the value 1 to 0 for the highest priority data to least priority data in order to build an intelligent data mart. This shown by following figure

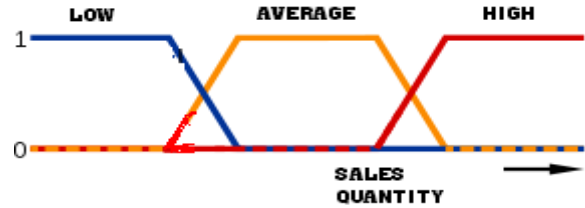
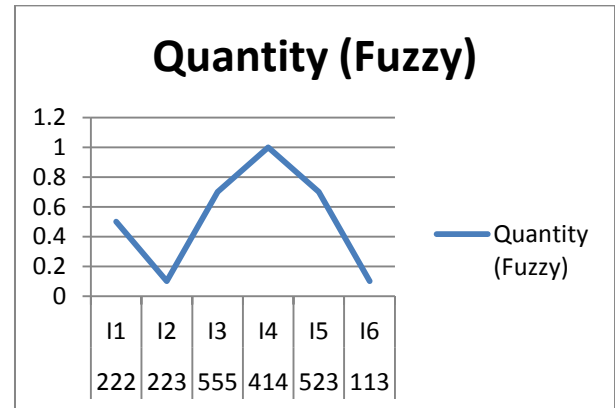


Figure.2. Shows fuzzy membership assigned to the products.



This paper proposes a Fuzzy data Mart, for sales department that allows integration of ETL and fuzzy concept into the data mart. Data come from different business systems. They should be initially extracted from raw data of various data sources of systems, then after a serial of filtering and converting, and finally be loaded to data warehouse. The membership function is a graphical representation of the magnitude of participation of each input. It associates a weighing with each of the inputs that are processed, defines functional overlap between inputs, and ultimately determines an output response.

4. Experimental Analysis and Results:

This experiment is implemented by using Operation Research technique (Assignment problem). Here, the quantity of item and places are taken into the account. The result of assignment problem shows that, the I1 is assigned to P3, I2 to P1, I3 to P4, I4 to P6, I5 to P2 and I6 to P5.

Histogram analysis

The result of assignment problem is given in the form of histogram analysis.

Table 3:

Before the membership of Fuzzy value:

Quantity of sales	No of Places
10	7
20	6
30	7
40	6
50	6
60	4

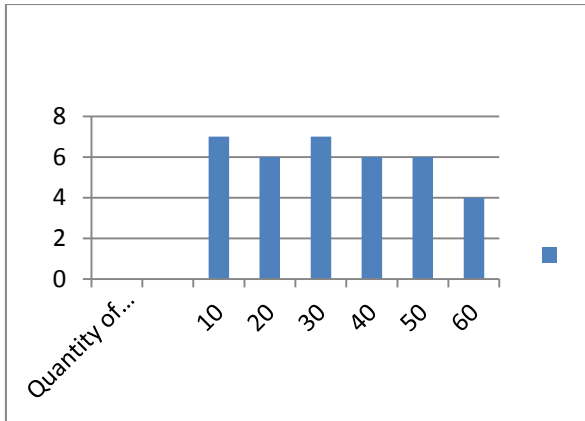
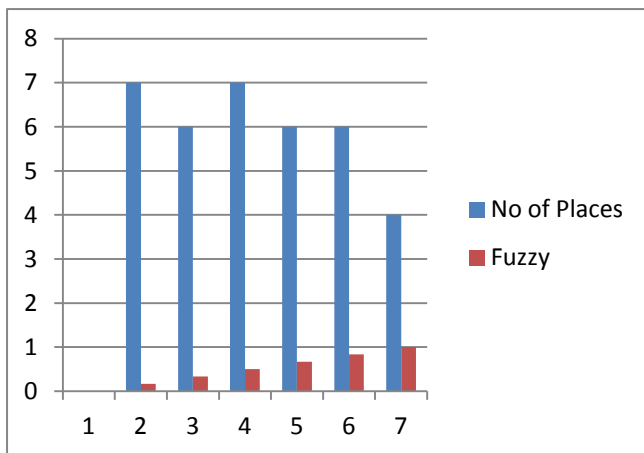


Table 4:After the fuzzy membership values:

Quantity of sales	No of Places	Fuzzy
10	7	.1666
20	6	.3333
30	7	.5000
40	6	.6666
50	6	.8333
60	4	1



5. Conclusion

These sales fuzzy Data Mart provides a consistent means of business rules, information availability, consistent business views and sales item windows. This proposed work is implemented in a linear programming problem through an assignment problem in terms of quantity. Its application potential in general and business environments in particular, leads to decision support and analysis. The work presented in this paper is aimed at exploring an effective decision making in sales promotion.

6. REFERENCES

[1] H.-J. Lenz and A. Shoshani."Summarizability in olap and statistical data bases". Statistical and Scientific Database Management, 1997.

[2] R. Kimball and M. Ross. "The Data Warehouse Toolkit." WileyPublishing, Inc., 2002.

[3] D. P´erez, M. J. Somodevilla, and I. H. Pineda."Fuzzy spatial data warehouse: A multidimensional model." Eighth Mexican International Conference on Current Trends in Computer Science, 2007.

[4] W. H. Inmon." Building the Data Warehouse". Wiley Publishing,Inc., 4 edition, 2005.

[5] M. Delgado, C. Molina, D. Sanchez, A. Vila, andL. Rodriguez-Ariza. "A fuzzy multidimensional model for supporting imprecision in OLAP". 2004 IEEE International Conference on Fuzzy Systems, 3:1331 – 1336, 2004.

[6] D. Fasel and D. Zumstein. „A fuzzy data warehouse approach for web analytics. "In M. D. Lytras, E. Damiani, J. M. Carroll, R. D. Tennyson, D. Avison, A. Naeve, A. Dale, P. Lefrere,F. Tan, J. Sipior, and G. Vossen, editors, Visioning and Engineering the Knowledge Society – AWeb Science Perspective,volume 5736 of Lecture Notes in Computer Science, pages 276–285. Springer, 2009.

[7] H.-J. Zimmermann." Fuzzy Set Theory and its applications."Kluwer Academic Publishers, 1991.

[8] Inmon, William (2000-07-18). "Data Mart Does Not Equal Data Warehouse". DMReview.com.

[9] Li Jian and Xu Bihua-" ETL Tool Research and Implementation Based on Drilling Data Warehouse "by,Southwest Petroleum University,Chengdu, China

[10] Ilovici, Lior Sapir,Armin Shm "A methodology for the design of a Fuzzy Data Warehouse".

[11] Daniel Fasel,"A DataWarehouse Model for Integrating Fuzzy Concepts in Meta Table Structures" Information System Research Group,Department of Informatics,University of Fribourg , Switzerland .

[12] Powerful Techniques for High-Performance Data Warehousing

[13] G. T. S. Ho, K. L. Choy, T. C. Poon," Providing decision support functionality in warehouse management using the RFID-based fuzzy association rule mining approach "

[14] Petrovic,(2007) "Decision Support tool for multi-objective job shop scheduling problems with linguistically quantified decision function".

[15] N.Sujatha and K.Iyakutty "Improved fuzzy c-mean clustering of web usage data with genetic algorithm "

[16] JAY VERKUILEN "Assigning Membership in a Fuzzy Set Analysis" *University of Illinois, Urbana-Champaign*

[17] Huang GuoJun and Hao ping," An Approach for Detecting Approximately Duplicate Data Warehouse Records"College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310032, China,2010 International Conference on Computer Application and System Modeling (ICCAASM 2010)

[18] Teh Ying Wah, Ng Hooi Peng, and Ching Sue Hok "Building Data Warehouse "Department of Information Science,University Malaya Malaysia Proceedings of the 24th South East Asia Regional Computer Conference, November 18-19, 2007, Bangkok, Thailand.