Logical Framework of Pattern Warehouse

Neha Gupta[1], Neha Bhardwaj[2]

Research Scholar, Assistant Professor

[1,2]Department of CSE & IT, Madhav Institute of Technology & Science, Gwalior (M.P), India.

Abstract

Pattern warehouse and data warehouse are the type of repository that main aim is to store data. Data warehouse store tremendous data get from disparate sources; on that huge amount of data, mining process is performed on data and patterns are extracted which store in pattern warehouse. These patterns can be directly access from pattern warehouse and no need to perform data mining process repeatedly. In this paper the main purpose is to propose logical architecture of pattern warehouse and describe end user communication also.

Keywords –

Pattern warehouse and pattern mining, Association rule mining, Data warehouse (DW), Genetic algorithm.

1. Introduction

In recent time, tremendous amount of data present in information industry. Data is of no benefit till modified it towards appropriate data. It is crucial to figure out the tremendous amount of data and to fetch appropriate content from it. Data warehouse is an area where data stocks in tremendous amount. According to Bill Inmon data warehouse can also be justifying as "it is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision-making process" [1]. Data warehouse is recent concept that of the database management system. Tremendous amount of data in which Extraction, Transformation and Loading (ETL) of blended data from diverse corner and then stock in DW [8]. Data mining related to gold mining is the mechanism of obtain value from the data stock in the DW. Data mining is the scrutiny of enormous data to detect consequential rules. Data mining approaches supports in the prediction of consumer nature in purchase of certain things. Extraction of information is not the only mechanism; DM also associates other mechanisms such as data cleaning, data integration, data transformation, data mining, pattern evaluation and data presentation [10].

Pattern warehouse and Pattern mining

Due to a few issues in data warehouse, Pattern warehouse came into existence. Pattern warehouse is different store mechanism which is adequate to store and supervise patterns and show compatible knowledge in the form of pattern [6]. In pattern warehouse, pattern mining procedure performed upon the patterns which gain analytical consequence for business. For extraction of pattern, no need to execute data mining process continuously. Patterns can be precisely and regularly access in pattern warehouse [2].
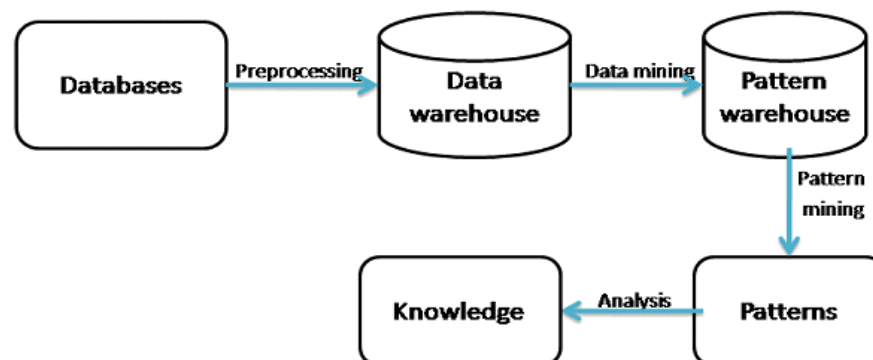

Figure 1. Work flow of Pattern warehouse [3]

2. Literature review

In 2014, Vivek Tiwari et.al, proposed a conceptual design of pattern schema, in this research paper author construct a snowflake schema for pattern storage and new concept introduce context and knowledge type for arrangement of pattern. Author also tried to describe about the pattern warehouse and how data warehouse is differing from pattern warehouse and introduce four quality forms for superior pattern warehouse design and reduce maintenance cost.

In 2016, Vishakha Agarwal et.al, proposed a progress from data warehouse to pattern warehouse, in this research paper author described the concept of data management like data warehouse and pattern warehouse or other types of repositories used for data storage.

In 2016, Vishakha Agarwal et.al, proposed an algorithm for finding the relevant and efficient patterns. In the proposed approach author is proposing an algorithm which works upon the optimization component for obtaining efficient and frequent patterns.

In 2017, V. Tiwari et al., described the important issues of pattern retrieval in pattern warehouse, and proposed an elementary view of pattern retrieval through classification. Pattern classification is not a single or independent process. But patterns are created through various data mining techniques, so it is very hard to develop pattern classification systems to classify patterns. The classification system should be able to identify the class boundary of attributes by only giving items and frequency count.

3. Proposed methodology

In this paper proposing the generalized architecture of pattern warehouse because still misses the proper working of pattern warehouse and also misses the components of logical architecture [3]. In past work author proposed a contextual snowflake schema but did not describe about the meta data and end user communication [2]. Meta data describe data in pattern warehouse in more efficient way and end user communication is necessary. Therefore, end user communication must be involved in the design of pattern warehouse.
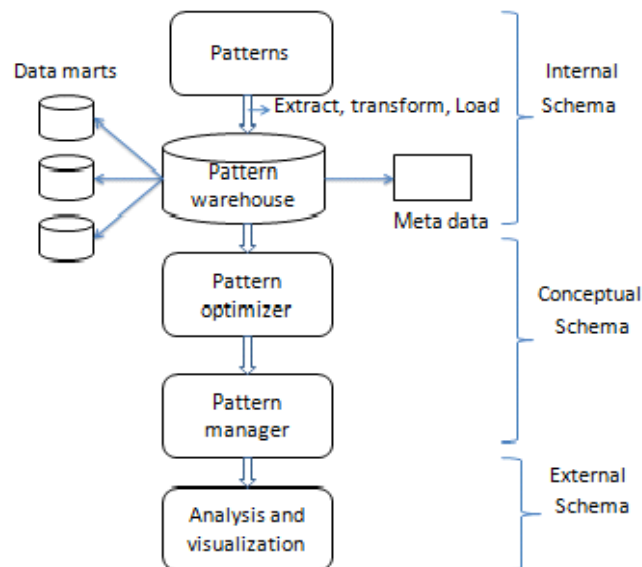

Figure2. Logical architecture of Pattern warehouse

Internal schema – All type of patterns comes out from data mining process and on that patterns some operation performs like extract, load and refresh and then store in pattern warehouse. In pattern warehouse store different patterns will become data mart and meta data of patterns.

Conceptual schema – In conceptual schema, pattern optimizer has imposed and extract out optimal pattern in pattern warehouse. Those optimal patterns managed by the pattern manager.

External schema –This schema is user friendly. Analysis, representation and visualization of the patterns.

Components of logical architecture –

I. Patterns

Special type of data is called Patterns [5]. Homogeneous data store in data warehouse, on that data we will execute the data mining process by which some patterns will come out.

In this paper set up few transactional datasets out of the data warehouse and perform some operation on them to get some patterns because pattern is more impactful than data on user and represents knowledge.

- **Association Rule mining (ARM)**

ARM is a method which intension to detect regularly patterns and rules from datasets found in numerous kinds of databases such as relational databases, transactional databases. It can be used not only to upgrade customer service but also enhance the company's revenue [6]. It is depending upon support and confidence parameters. For generating association rules, FP growth algorithm is used.

- **FP growth**

FP- growth is an algorithm in ARM, FP growth algorithm used for discovery of regular itemset in a transactional database without applicant generation. FP growth algorithm is an improvement of Apriori algorithm.

 Transactional data on which apply FP growth algorithm and after applying this algorithm few patterns are generated are given below –

Association Rules

```
[Curd] --> [Fresh] (confidence: 1.000)
[Fresh] --> [Curd] (confidence: 1.000)
[Curd] --> [Frozen] (confidence: 1.000)
[Frozen] --> [Curd] (confidence: 1.000)
[Curd] --> [Grocery] (confidence: 1.000)
[Grocery] --> [Curd] (confidence: 1.000)
[Curd] --> [Milk] (confidence: 1.000)
[Milk] --> [Curd] (confidence: 1.000)
[Curd] --> [Vegetables] (confidence: 1.000)
[Vegetables] --> [Curd] (confidence: 1.000)
[Fresh] --> [Frozen] (confidence: 1.000)
[Frozen] --> [Fresh] (confidence: 1.000)
[Fresh] --> [Grocery] (confidence: 1.000)
[Grocery] --> [Fresh] (confidence: 1.000)
[Fresh] --> [Milk] (confidence: 1.000)
[Milk] --> [Fresh] (confidence: 1.000)
[Fresh] --> [Vegetables] (confidence: 1.000)
[Vegetables] --> [Fresh] (confidence: 1.000)
[Frozen] --> [Grocery] (confidence: 1.000)
[Grocery] --> [Frozen] (confidence: 1.000)
[Frozen] --> [Milk] (confidence: 1.000)
[Milk] --> [Frozen] (confidence: 1.000)
```

II. Pattern warehouse

These patterns will be store in pattern warehouse in the form of snowflake schema [2]. Schema is a characterization of the unified database. The snowflake schema is a development of the star schema. Snowflake schema consists of a fact table enclosed by multiple dimension tables and dimension table act as a fact table which can be linked to another dimension tables.
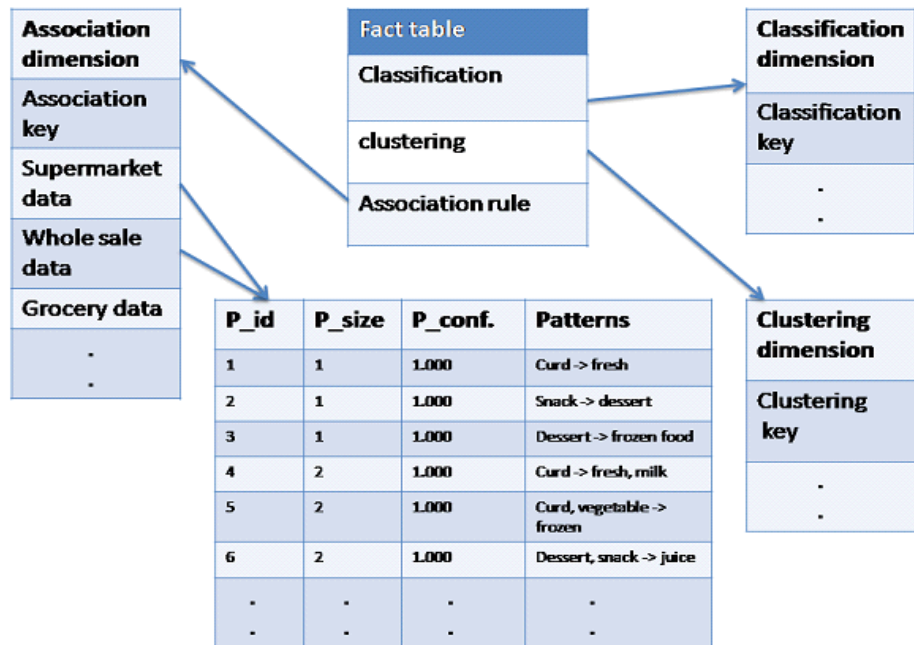
Figure 3. Transactional Patterns stored in snowflake schema [2]

III.   Pattern optimizer

Some patterns are taken in pattern warehouse and apply any optimization algorithm towards frequent pattern of pattern warehouse. The optimization algorithm may be genetic algorithm, particle swarm optimization, Ant colony, Honey Bee, spider monkey etc. Those frequent pattern scan access by users and analysis for business decision making process.

In this work genetic algorithm is used for extracting optimal patterns from pattern warehouse.

- **Genetic algorithm**

Genetic algorithm is a heuristic algorithm and it is situated on the principles of natural selection. It is used to discover optimal solutions to tough problems and solve optimization problems [4]. Optimization is the mechanism of taking the input values and we get the "best" output values. Three operators basically use in genetic algorithm: - Selection, Crossover and Mutation.

In pattern warehouse, optimization algorithm perform upon pattern in following steps are [4] –

- Select pattern-set in pattern warehouse.

- Assign the fitness value.

- In pattern-set, randomly select the two patterns.

- Crossover operator applies upon the two selected patterns.

- If pattern cross the fitness amount then go to the mutation operator for optimal pattern otherwise, go to the selection operator.

- All patterns are check then process is terminated otherwise, go to the selection operator and repeat the whole process continuously.

After applying genetic algorithm on sample dataset following optimal patterns comes out.

Tid	Patterns
T2	Vegetable, milk
T3	Curd, milk, frozen
T4	Vegetable, frozen, milk, curd

Table 3.Optimal pattern of wholesale pattern

IV. Pattern manager

There are some optimal patterns extracting out after applying some optimization algorithm. That's why pattern manager is necessary to hold those patterns in an efficient way. It is important component of architecture.

V. Analysis and visualization

This component is used for analysis and visualization of patterns.

4. Conclusion and future work

In this research paper, logical architecture of Pattern warehouse is proposed and described processing and execution of all components of this architecture. In future work missing data structure which can hold all types of pattern and describe another approach for extracting most frequent patterns in the pattern warehouse can be proposed.

References

[1] Inmon, W.H., Hackathorn, and R.D (1994) Using the data warehouse. Wiley-QED Publishing, Somerset, NJ, USA.

[2] Vivek Tiwari, Ramjeevan Singh Thakur "Contextual snowflake modelling for pattern warehouse logical design" Maulana azad national institute of technology (MA-NIT), September 2014.

[3] Vishakha Agarwal, Akhilesh Tiwari "From data warehouse to pattern warehouse: A progressive step" international journal of engineering research, pp 249-252, April 2016.

[4] Vishakha Agarwal, Akhilesh Tiwari "A novel optimal pattern mining algorithm using genetic algorithm" international journal of computer applications, June 2016.

[5] Vivek Tiwari and Ramjeevan Singh Thakur "Towards important issues of pattern retrieval: pattern warehouse" Int. J. Data Science, 2017.

[6] Harshita jain, Akhilesh Tiwari "Sequential step towards pattern warehousing" international journal on recent and innovation trends in computer and communication, pp 665-670.

[7] Dr. Ajay Kukshal, Susheel Chandra Bhatt "Changing Pattern of Warehousing: Data Warehouse to Pattern Warehouse" International Journal of Advanced Research in Computer Science and Software Engineering, January 2014.

[8] A.R. Arunachalam, S. Srigowthem "A STUDY ON DATA WAREHOUSE ARCHITECTURE" International Journal of Pure and Applied Mathematics, pp 273-275, 2017.

[9] Richa Pandey, Lalit Mohan, Sanjeev Bisht and Janmejay Pant "Data Mining and Data Warehouse" International Journal on Emerging Technologies (Special Issue NCETST-2017).

[10] Ramandeep Kaur1 , Amanpreet Kaur2 , Sarabjeet Kaur3 , Amandeep Kaur4 , Ranbir Kaur "An Overview of Database management System, Data warehousing and Data Mining" International journal of advanced research in computer and communication engineering, July 2013.