A Pattern Storage System using Pattern Warehouse along with Sources of Pattern Generation and Applications

Khedekar Vilas Baburao, Dharmendra Singh Rajput

Abstract*: Now a day different data mining algorithms are ready to create the specific set of data known as Pattern from a huge data repository, but there is no infrastructure or system to save it as persistent storage for the generated patterns. Pattern warehouse presents a foundation to make these patterns safe in the specific environment for long term use. Most organizations are excited to know the information or patterns rather than raw data or group of unprocessed data. Because extracted knowledge play a vital role to take right decision for the growth of an organization. We have examined the sources of patterns generated from large data sets. In this paper, we have presented little importance on the application area of pattern and idea of patter warehouse, the architecture of pattern warehouse then correlation between data warehouse and data mining, association between data mining and pattern warehouse, critical evaluation between existing approaches which theoretically published and more stress on association rule related review elements. In this paper, we analyze the patterns warehouse, data warehouse concerning various factors like storage space, type of storage unit, characteristics, and provide several research domains.

***Index Terms*: Data Warehouse, Data mining, Pattern Warehouse.**

.

I. INTRODUCTION

Today a massive amount of data routinely created by various sources. Users and no useful information cannot directly exploit it can be deduced simply by their observation [1]. This massive volume of data cannot be saved and handled by databases; therefore, the data warehouse used to save it for more processing and use. More detailed and high-level techniques are required to be examined to find the hidden information and make these information valuable [2]. Existing data mining techniques are found to obtain knowledge from raw data items [3]. The result of data mining techniques are association rules, decision tree, clusters, and other structures that explain features of the raw data [4]. Produced "pattern is compact and rich in the semantic representation of raw data" [5]. Generated patterns do not belong to any of the persistent storage like database or data warehouse. Traditional Database Management System is not powerful and flexible enough to manage this unique set of information; therefore, there is a need of a specific management system which is capable of modeling and storing patterns [1]. The comprehensive research in data mining produces complicated patterns, or simple patterns. We need a novel mechanism to manage patterns effectively so that patterns can be used for future analysis and application. We require a better warehouse system to the pattern which is named Pattern Warehouse; Pattern warehouse is a set of persistently collected patterns [4]. And "We need tools that will permit us to compare, query, and store the pattern to retrieve the information/patterns when we want" [4]. "Patterns should be modeled, stored, processed, and queried in a similar way to data in traditional DBMS" [2]. Semantically rich data can not handled by traditiobnal DBMS [6]; We require a specific database management system able to handle semantically significant data (Knows as Patterns) [7]. We need a unique management system to handle all patterns in pattern warehouse. Pattern base management system is presented logically inside the PANDA Project. "Patterns should be manipulated (e.g. extracted, synchronized, deleted) and queried through a Pattern Manipulation Language (PML) and a Pattern Query Language " [8].

II. PATTERN GENERATION SOURCES

Patterns are hidden continuously, and patterns are always identified or created by various data mining methods from fresh data item sets.

The following are some pattern generation sources are as follows.

A. Association Rules

In this rule-based approach of obtaining a relation among the variables in the more extensive database. Many algorithms have been introduced to produce association rules. Few of them are as follows:

a) *Apriori algorithm:* Works the breadth-first search system to estimate the support of data items and handles the creation of candidates which uses the advantage of the assets nearby to support [9].

b) *Eclat algorithm*: Used as a depth-first technique in this

Revised Manuscript Received on July 22, 2019.

Khedekar Vilas Baburao, Assistant Professor in Computer Engineering, JSPM's, JSCOE, Hadapsar, Pune, India.

Dr. Dharmendra Singh Rajput , Associate Professor in the Software and Systems Engineering, SITE, VIT, Vellore, India

algorithm. The algorithm is a naturally attractive algorithm that is suitable for both parallel and sequential execution [10].

c) FP-growth algorithm: The meaning of FP is frequent pattern. The algorithm used for the finding of short as well as long patterns [11].

d) AprioriDP: An Effective way to search Frequent Pattern with Dynamic Programming Approach [12].

e) Node-set-based algorithms: Following four are algorithms based on node sets.

a) FIN Algorithm: FIN checks items frequently in the search tree directly [13].

b) Prepost algorithm: PrePost can immediately detect frequent itemsets without creating candidate itemsets. In any other situations by creating use of the unique path feature of N-list [14].

c) PPV algorithm: PPV is dependent on a node-list data structure, which comes from the coding prefix-tree named PPC-tree [15].

d) PrePost+: PrePost+ employs an efficient pruning of children-parent utility roping for reducing search space [16].

The result of all algorithms is a collection of rules which gives an association among item. There are following two other methods to create patterns.

B. Clustering

Clustering introduces to the grouping of related objects [17] in such a way that groups of objects in the same group.

C. Decision Tree

Is a mechanism to help the decision that handles a tree-like structure for an appropriate decision in minimum time [18] as well as its probable consequences, including accidental event outcomes, resource expenses, plus service.

III. PATTERN AND PATTERN WAREHOUSE

A. Basic Concept

Pattern is a special kind of information generated through different data mining methods [19]. Patterns are a piece of information generated by the massive knowledge container. The pattern is a distinct entity in the environment, reappeared general structure and determination of methods of data mining.

"Pattern Warehouse is a Repository for patterns created through various data mining techniques, Collection of persistently stored patterns" [4], "Software framework which provides the infrastructure for knowledge representation [20] and mining by allowing the pattern to be stored permanently." [19]

B. Evolution of Database Technologies

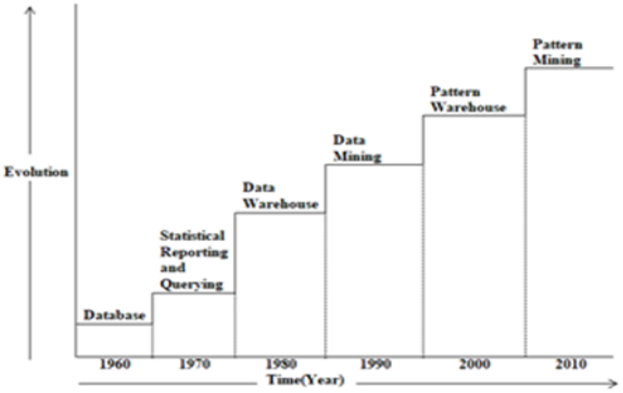


Fig. B.1

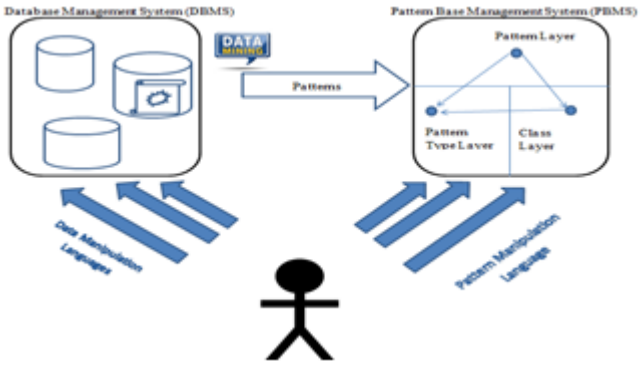
C. Architecture

a) PBMS Architecture

The complete architecture of the system is presented in Fig. C.1.

Fig. C.1

The left-hand side is associ



ated with collection of data, and right side associated with the management system for patterns. User can interact among those systems using unique management language.

Following are three tiers of PBMS:

a. *Class layer*, in which the pattern is defined as a the pattern expressed as a class, means a collection of patterns.

b. *Patterns layer*, which is associated with patterns;

c. *Patterns type layer*, user-defined and pre-defined types used to different and group the patterns;

End-users can interact directly with PBMS; Because of this, The robust methods are used in PBMS to save different patterns and evaluate patterns to reuse for further applications.

b) PWMS Architecture

The entire architecture of the system is presented in Fig. C.2

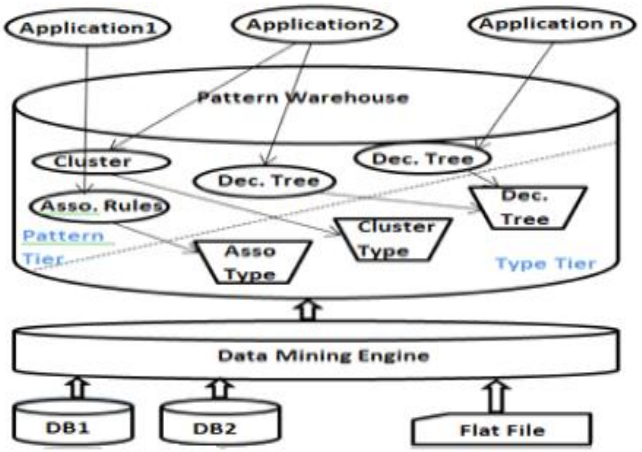


Fig. C.2

The head of the architecture includes an application layer which describes different machines, applications, and users. The next middle layer represents the working model of PWMS which consist of two layes. 1. Type-Tier, 2. Pattern-Tier. Last or bottom layer represents the input media and patterns are generated and then modified by using the data mining engine. **Pattern layer:** This is populated with patterns. They can be obtained by different data mining techniques [21]. **Type layer:** This layer keeps user-defined types and built-in types of patterns. It defines the language of the patterns. Type define pattern type of similar properties [21].

IV. APPLICATIONS

Patterns created from various applications and saved in the data warehouse for forthcoming analysis. Patterns selected from data warehouse by multiple data mining methods. Produced patterns can be complex or simple [22] and "patterns are not persistent by nature" [23]. "Pattern gets lost when it goes out of memory. There is long and complex process behind the pattern generation" [4]. Maximum of the business or industries want information or patterns for decision making. Correct decision improves the business growth as well as business intelligence system [24]. General applications of pattern warehouse is Healthcare Applications, Business analytics, Traffic Control, Weather forecasting, Agriculture related applications, Social media data analysis etc. PANDA Project describes following application area of Pattern Warehouse: Signal Processing, Data Mining, Information retrieval, Visualization, Mathematics. Visualization.

V. LITERATURE SURVEY

Many researchers have proposed the concept of Pattern Warehouse. Many of other have proposed Pattern Base Management System, Pattern Management Language as well as Pattern Query Language.

Table V.1

Paper No.	Techniques used	Parameters
1	PBMS, PQL	In this paper, Author dealings with pattern reproduction and storage. Also analyzes the DBMS approaches like Semi-structured (XML) Model for relational objects.
2	PBMS, PMML, PQL, PDL(Pattern Definition Language)	In this paper, the author contributes to the fundamentals of pattern association; As well as association among data and patterns. Also defines pattern query language and pattern definition language for pattern related operations.
3.	PWMS(Pattern Warehouse Management System)	Author focus on this work is: Explain the design model of PWMS. Explain the relationship between patterns generation and other model for storage concurrently.
4	DBMS, PWMS, PQL, PML(Pattern Manipulation language)	In this paper author explains the design architecture of a PWMS and provides method for pattern management.
5	PWMS, Knowledge warehousing, Pattern retrieval: Classification	In this paper author Decided that the use of distribution in pattern retrieval is well possible. Explained the difference in pattern distribution and data distribution. Explained the idea of information warehouse.
6	PBMS	In this paper author presents preliminary result of PANDA project. Present the reasonable grounds of a general structure based on the concepts of pattern classes and pattern sets. Introduce the critical problem in querying patterns.
7	Knowledge discovery and Data mining techniques	In this paper, the author does structure for the evaluation of correlation among complex and simple pattern.
8	RDBMS, XML, SQL, PBMS	In this paper author discusses Meaning of pattern. Generic model for pattern Queries and operations on pattern. How to manage patterns? Pattern management challenges.
9	Dimensional data models	In this paper author Introduces analysis pattern for dimensional data model. Describes decision support task and their solutions in the form of dimensional data model.

10	PBMS, TPML(Temporal pattern manipulation language), TPQL(Temporal pattern query language)	In this paper, the author shows an overall structure for pattern control. Specifies the services on the pattern using manipulation languages like PQL. Describes pattern validity, pattern predicates, query operators, temporal operators etc.
11	PBMS, XML, PMML, PQL	In this paper author describes Relation between raw data and patterns. Logical framework for modeling patterns. Types of relationship between patterns. Difference between DBMS and PBMS. PBMS architecture.
12	Concept of classification, its application and trends in data mining.	In this book, the authors describe data mining concepts as well as concepts of data warehouse, concepts of frequent pattern mining , classification, clustering analysis. Application and trends in data mining.

Table V.2

Characteristics	**Data-warehouse**	**Pattern-Warehouse**
Features(Processing)	Informational	Analytical
Processing System	Old records	Pattern
Volume of warehouse(MB)	100	0.001
Representation of Data	Old data	Patterns
Count of Users	100	Few(Only administrators)

Note: This table is partially based upon as given in [25].

Table V.3

Characteristics	**PW**	**PBMS**	**PQL**	**PMML**
Stands for	Pattern Warehouse	Pattern Base Management System	Pattern Query Language	Predictive Model Markup Language
Features	Permanent warehouse for patterns	Dedicated pattern storage manipulation system	Pattern extraction query language	the model for analytical and knowledge mining models
Developed by	PANDA-Project	Bartolini [1]	Evangelos Kotsifako, 2005 [5]	Data Mining Group ,2008
Properties	Pattern arrangement, Pattern Management,	1. Abstraction: Raw data to be directly analyzed and manipulated [4]. 2. Efficiency: Improves the efficiency of DBMS and PBMS [4].	Pattern fetching from pattern warehouse	XML oriented language. Represents models created by data mining algorithms..

With the extensive study on database, data warehouse, pattern database and pattern warehouse we got the results shown in following Fig. V1.

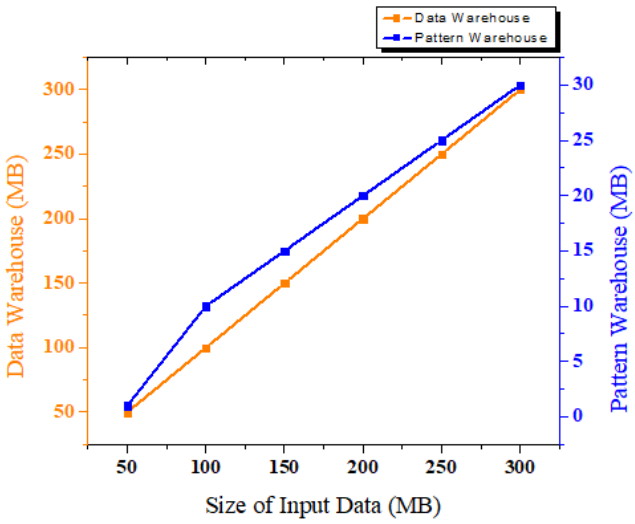


Fig. V1

As we all know, the data warehouse is a container for all raw data items. It simply means the volume of the data warehouse is directly proportional to the input data sets. As input datasets increase the volume of the data warehouse also increases in the same rate of data generation.

But as per users requirement, any user needs only useful information. Here patterns warehouse fulfills the need of user simply extracting useful information from pattern warehouse. This system provides two main benefits.

1. Keeps only useful information; Means minimum storage required to preserve it.
2. Results are always available to the user within minimum time compare to the data warehouse.

VI. FUTURE SCOPE

Pattern means the use of knowledge that is understood to be rich and rich in meaning. Massive odd fresh data related in any information-focused application. Data mining, knowledge retrieval, or image processing these are the examples for the same.

Building a pattern repository and saving a pattern is an obstacle, but using self-collected patterns is a significant problem. It is necessary to consider the modern query language also attempt to perform a pattern recovery. Currently used data mining techniques will be expanded to perform operations on pattern recovery. There are several parameters needed to be considered for design and development of pattern warehouse. The working or executed pattern manipulation system is required to produce all the difficulties and benefits related to the pattern management. In order to create more complete PBMS need to extend research in PQL, PML and PMML.

VII. CONCLUSION

This research paper shows the basics of pattern and pattern warehouse. Also shows the architecture and importance of pattern warehouse. The comparative study between Data warehouse and Pattern warehouse as well as comparative study between Pattern Warehouse, PBMS, PQL, PMML. Built a new way to define system structure that describes, supports and manages the patterns efficiently. This is the foundation of this unique method; it is an effective way to represent huge volumes of information. This work tries to study all related techniques differently. As different techniques are available; it inspires us to generate a unique management method for handling specific patterns.

APPENDIX

List of Figures

Sr. No	Figure No	Title
1	Fig. B.1	Evolution of database technology
2	Fig. C.1	PBMS Architecture [19]
3	Fig. C.2	PWMS Architecture [21]

List of Tables

Sr. No	Table No	Title
1	Table V.1	Survey of reference papers
2	Table V.2	Finding after survey; Comparison of, Data warehouse and Pattern Warehouse
3	Table V.3	Finding after survey; Comparison between PW, PBMS, PQL, PMML

REFERENCES

1. I. Bartolini et al., "PAtterns for Next-generation DAtabase systems: preliminary results of the PANDA project," in Sebd, 2003, pp. 293–300.
2. M. Terrovitis et al., "Modeling and language support for the management of pattern-bases," Data Knowl. Eng., vol. 62, no. 2, pp. 368–397, 2007.
3. S. M. Basha, D. S. Rajput, N. C. S. N. Iyengar, and R. D. Caytiles, "A Novel Approach to Perform Analysis and Prediction on Breast Cancer Dataset using R," Int. J. Grid Distrib. Comput., vol. 11, no. 2, pp. 41–54, 2018.
4. J. Wang, "Encyclopedia of Business Analytics and Optimization," J. Clean. Prod., vol. 16, no. 15, pp. 1799–1808, 2014.
5. E. Kotsifakos, I. Ntoutsi, and Y. Theodoridis, "Database support for data mining patterns," in Pci, 2005, pp. 14–24.
6. S. M. Basha and D. S. Rajput, "Parsing Based Sarcasm Detection from Literal Language in Tweets," Recent Patents Comput. Sci., vol. 11, no. 1, pp. 62–69, 2018.
7. S. M. Basha and D. S. Rajput, "A supervised aspect level sentiment model to predict overall sentiment on tweeter documents," Int. J. Metadata, Semant. Ontol., vol. 13, no. 1, p. 33, 2018.
8. B. Catania, A. Maddalena, M. Mazza, E. Bertino, and S. Rizzi, "A framework for data mining pattern management," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 3202, pp. 87–98, 2004.
9. R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," J. Comput. Sci. Technol., vol. 15, no. 6, pp. 487–499, 1994.
10. M. J. Zaki, "Scalable algorithms for association mining," IEEE Trans. Knowl. Data Eng., vol. 12, no. 3, pp. 372–390, 2000.
11. Y. Y. Jiawei Han, Jian Pei, "Mining frequent ordered patterns without candidate generation," Proc. - Fourth Int. Conf. Fuzzy Syst. Knowl. Discov. FSKD 2007, vol. 1, pp. 402–406, 2007.
12. D. K. M. P. C. P. Bhalodiya, "An Efficient way to Find Frequent Pattern with Dynamic Programming Approach," pp. 2–6, 2013.
13. Z. H. Deng and S. L. Lv, "Fast mining frequent itemsets using Nodesets," Expert Syst. Appl., vol. 41, no. 10, pp. 4505–4512, 2014.
14. Z. Deng, Z. Wang, and J. Jiang, "A new algorithm for fast mining frequent itemsets using N-lists," Sci. China Inf. Sci., vol. 55, no. 9, pp. 2008–2030, 2012.
15. Z. Deng and Z. Wang, "A new fast vertical method for mining frequent patterns," Int. J. Comput. Intell. Syst., vol. 3, no. 6, pp. 733–744, 2010.
16. Z. Deng and S. Lv, "Expert Systems with Applications PrePost + : An efficient N-lists-based algorithm for mining frequent itemsets via Children – Parent Equivalence pruning," Expert Syst. Appl., vol. 42, no. 13, pp. 5424–5432, 2015.
17. S. M. Basha, V. Vandhan and D. S. Rajput, "Impact of Gradient ascent and Boosting algorithm in classification," Int. J. Intell. Eng. Syst., vol. 11, no. 1, pp. 41–49, 2018.
18. S. Kallam, S. M. Basha, D. Singh Rajput, R. Patan, B. Balamurugan, and S. A. Khalandar Basha, "Evaluating the Performance of Deep Learning Techniques on Classification Using Tensor Flow Application," Proc. 2018 Int. Conf. Adv. Comput. Commun. Eng. ICACCE 2018, no. June, pp. 331–335, 2018.
19. V. and R. S. T. Tiwari, "Toward important issues of pattern retrieval : pattern warehouse," Inderscience Enterprice ltd., vol. 2, no. 1, pp. 1–14, 2014.
20. S. M. Basha and D. S. Rajput, "A roadmap towards implementing parallel aspect level sentiment analysis," Multimed. Tools Appl., 2019.
21. V. Tiwari and R.S. Thakur, "P 2 MS : a phase-wise pattern management system for pattern warehouse Vivek Tiwari and Ramjeevan Singh Thakur," vol. 7, no. 4, pp. 331–350, 2015.
22. I. Bartolini, P. Ciaccia, I. Ntoutsi, M. Patella, and Y. Theodoridis, "A Unified and Flexible Framework for Comparing Simple and Complex Patterns," 2004, pp. 496–499.
23. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. 2012.
24. S. M. Basha, Y. Zhenning, D. S. Rajput, R. D. Caytiles, and N. C. S. . Iyengar, "Comparative Study on Performance Analysis of Time Series Predictive Models," Int. J. Grid Distrib. Comput., vol. 10, no. 8, pp. 37–48, 2017.
25. M. Han, Jiawei and Kamber, Data-Mining, ConceptsandTechniques, vol. 12. 2011.

AUTHORS PROFILE

Khedekar Vilas Baburao, pursuing Ph.D. from SCOPE, VIT, Vellore, India. Working as Assistant Professor in Computer Engineering, JSPM's, JSCOE, Hadapsar, Pune, India. Completed M.Tech (July 2014) in Computer Science from JNTU, Hyderabad, India. Completed B.E. (May 2009) in Information Technology from Pune University, India. Area of research work is Data mining, Pattern Mining, Pattern Database. Membership: ACM.

Dr. Dharmendra Singh Rajput working as Associate Professor in the Software and Systems Engineering, SITE, VIT, Vellore, India. Completed PhD (Jan 2015) from NIT, Bhopal, India. His research areas are Data Mining and BigData Predictive Analytics. He has published 10+ reputed Journal Papers; 12 papers presented in international conference. He has received various awards from Indian Government like DST, CSIR Travel Grant and MPCST Young Scientist Fellowship.