

Knowledge Extraction and Analysis to Evaluate the Financial Performance of an Organization Using OLAM

Mahtab Ebrahimi

Department of Electrical and Computer Engineering, Islamic Azad University Damavand Branch Tehran, Iran
Email: mahtab.ebrahimi@gmail.com

Received: 23 September 2018; Accepted: 17 October 2018; Published: 08 December 2018

Abstract—Data mining or the discovery of knowledge out of databases extracts patterns, and useful and non-substantial, implicit, and unknown information from large databases. Searching for associative rules is a data mining method where relationships and dependencies Interactions between a large set of data items are shown. In large organizations, most data are created at the passage of time. Relational tables in a variety of business or scientific domains have rich information, with high quantitative and nominal data types. Therefore, in order to gather information, measure performance and increase business efficiency, it is necessary to isolate relational database and store data in a data warehouse. Using the multidimensional data mining that integrates online analytical processing with data mining, knowledge can be found in multidimensional databases.

In this article, to avoid inappropriate rules in exploring associative rules, the online analytical processing technique is combined with the Apriori algorithm and explores associative rules on multi-dimensional and multilevel data using the data cubic and Apriori algorithm. In addition, the criteria of accuracy, tool performance, and runtime are defined for comparing techniques for exploring associative rules. The metrics mentioned in the scope of this paper have been compared, and data from the Central Insurance Agency have been analyzed using data cubes and the Apriori algorithm, and useful association rules have been generated. Results of the research displays improvement of the response time compared to the other methods.

Index Terms—Data mining, data warehouse, online analytical processing, associative rules, data cube, OLAM.

I. INTRODUCTION

The discovery of frequent patterns, partnerships, and dependency relationships between very large data in selective marketing, decision management, and business management are of great importance, and if the particular features or attributes contained in an associative law are only of one dimension, then It is possible to use algorithms such as Apriori to find repeated sets of items and then strong

association rules, but if the number of dimensions and issues increases and each of the encoders themselves has different levels, then finding frequent patterns is problematic. So, in this article we try to discover the frequent patterns by creation of relationship between data mining and using OLAP¹ (the so-called OLAM), using the algorithms of data mining, such as Apriority and using OLAP operators.

To calculate the coefficient of assurance and backup of association rules, we use the data cubes that store the aggregates, and then we compare and evaluate the extracted rules using data cubes and without using them. Also, in order to implement the proposed method, the real data of the Central Insurance Agency were analyzed using data cubes and the Apriori algorithm, and useful association rules were generated.

II. SCIENTIFIC BACKGROUND

Firstly, we will briefly discuss the concepts of data warehouse, online analytical processing, associative rules, and exploration of the template in multidimensional and multilevel spaces and ultimately OLAM.

A. Analytical online processing

Services that use analytical databases, databases and data warehouses to quickly respond to analytic queries of users are called online analytical processing (OLAP).

OLAP is a collection of software that is used to quickly explore and analyze data based on a multi-dimensional approach with multiple levels of aggregation, and accelerates and facilitates decision-making. Using OLAP operations, such as roll up, drill down, slice and dice, data can be viewed from a variety of perspectives.

B. Multi-dimensional and multilevel associative rules

If we consider each dimension of a warehouse as a rig, we can explore the rules of an association that has multiple rings, then the rules of the association, in which there are more than one dimension or motility, are called the multi-dimensional association rule are known. The association

¹ Online analytical processing

rules derived from the exploration of data at various levels of abstraction are known as multilevel associative rules. These rules can be effectively investigated with the help of conceptual hierarchy, as well as support and confidence criteria.

C. On-line Analytical Mining (OLAM)

Data mining and OLAP can be considered analytical technologies in the BI6 family, which are called OLAM or OLAP mining. Data mining algorithms can use OLAP cubes to create data mining models. So OLAM has integrated OLAP data mining to explore knowledge in multi-dimensional databases.[2]

III. RESEARCH BACKGROUND

In recent research, a process for exploring associative rules on a data cube to obtain hidden patterns among the summarized data is proposed. In [3], Kamber et al an algorithm called mD-Slicing on a data cube for finding a collection of items in every dimension with the least amount of support and confidence has been offered.

The results show that the proposed algorithm performs better than the original Apriori algorithm. In [4], the exploration of associative rules in multidimensional environments is divided into three categories: the rules of an inter-dimensional association (containing only one dimension), extra-dimensional association rules (including multi-dimensional), and association associations rules. In this paper, this process is created by transforming the data cube into a tabular form. Other research in this area is [5] creating a framework for exploring rules in a multi-dimensional context.

Other research on OLAP integration and various data mining techniques is in [6], [Lee et al., provided a software in order to improve the quality of the garment industry and obtain patterns among the defects in products.

To do this, the output of the data extracted from the data cube is given as input to the Apriori algorithm and, finally, the patterns are intended to improve the quality of the garment. In [7], Namchul Do and his collaborators have been analyzed and evaluated by presenting an OLAM-based approach to product development processes. In this regard, using the data extracted from database of PDM (Product Data Management), firstly the defect Factors and functions' index are identified and after fulfilling the ETL process and creating OLAP cubes, the data is provided to enter the data mining and knowledge mining models.

Among other tasks is the provision of a registration system based on analysis Multi-dimensional by E Abuelyaman at [8]. The purpose of this study is to create a framework for proposing appropriate courses for students when they go to college. Using part of the OLAP operation, including Dicing, a portion of the data has been selected and rules and templates have been extracted. In [9], H Dev and SK Mishra used the integration of decision tree in data

mining and online analytical processing to explore the online banking system in order to improve the quality.

In [10], Dr. Gupta and Giannan have used the online analytical and data processing relationship in the decision support system in a medical clinic to discover a number of diseases. [11] Professor Mahash and Jegna Jadaw state that data mining reveals knowledge and hidden patterns of data, and a consider online analytic process as a tool for data mining and data warehouse, which performs different operations on the data stored in multi-dimensional databases.

Here, the associative rules apply to cubes derived from a university structure and finally different rules end results are extracted. In planning the organization's resources, Tamer et al. Used online analytical processing techniques along with data mining to generate analytical reports on ERP data to predict business prospects and respond to different decision makers' questions.

In the Behavioral model Of Subscribers, in [13], [Fox and Wang], in an article under the title of " Behavioral model of Subscribers" have used "online analytic explore" to design "internet marketing".

Dr. Walid provided a model for the OLAP-based clinical decision support system and data mining to solve the data correlation problem.

IV. SUGGESTED METHOD

In this paper, for the purpose of exploratory analysis ,the online analytical processing and its operation are integrated with data mining techniques, including associative rules, so that organizations can better organize their data and take better strategic decisions.

In the proposed method, we have used a data warehouse which actually includes data cleansing and integration, and can be considered as one of the most important processing steps in data mining.

Without the use of data mining techniques, OLAP analyzes can not predict and extract knowledge, providing only a few simple reports for managers and analysts.

To improve the process of exploring associative rules and reducing the set of duplicate items, we used OLAP operations like Slice, Drilldown and Rollup to reduce the number of items.

In addition, in order to create more useful and concise rules and increase the speed of creating associative rules and reduce costs by using the hierarchical concept, we made a leveling the dimensions and placed dimensions having quantitative categories in qualitative categories.

Then, we used the process of exploring the inter-dimensional and multi-level associative rules that were proposed in previous studies, and we plotted the algorithm with the Java programming language. In order to improve this algorithm and to reduce the set of candidate items created at each stage of the Apriori algorithm property, as well as maintaining the set of items and creating a new set of new candidate items from the HashMap class in Java.

The steps for doing the job are summarized in Fig.1 .

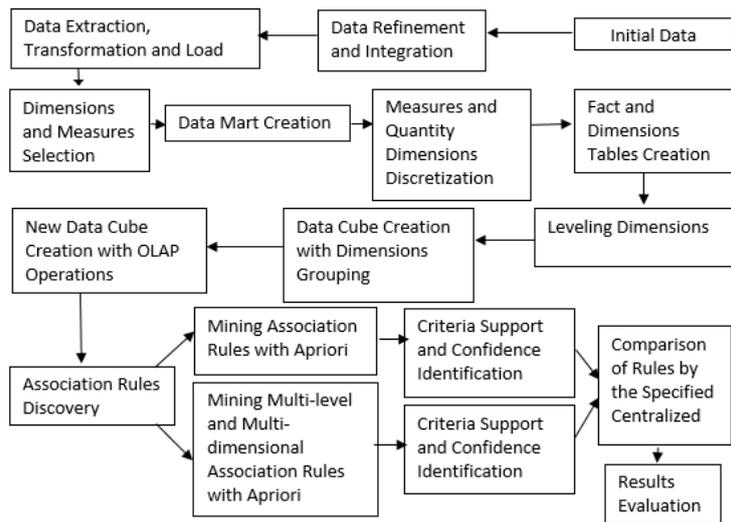


Fig.1. Procedures for doing the proposed method

V. CASE STUDY

In order to implement the proposed method, we use data from the Central Insurance data mart. These data are sent to the Tax Administration by various insurance agents in different parts of the country at specific timescales.

A. Create data warehouse and data cube

In this proposed method, we are forced to create an inventory for analysis, extraction of knowledge and decision making from the historical information.

After creating the data warehouse and saving of the historical data in an integrated and refined form, we first disconnect the data and convert all existing values that are numerically into these nominal values. For example, here we convert the values for the insurance cover, which is numerically, after the discrete creation and grouping into nominal values. Then make a data cube with four dimensions of time, insurance representative and insurance policy, and the insurance policy cover as shown in Fig. 2.

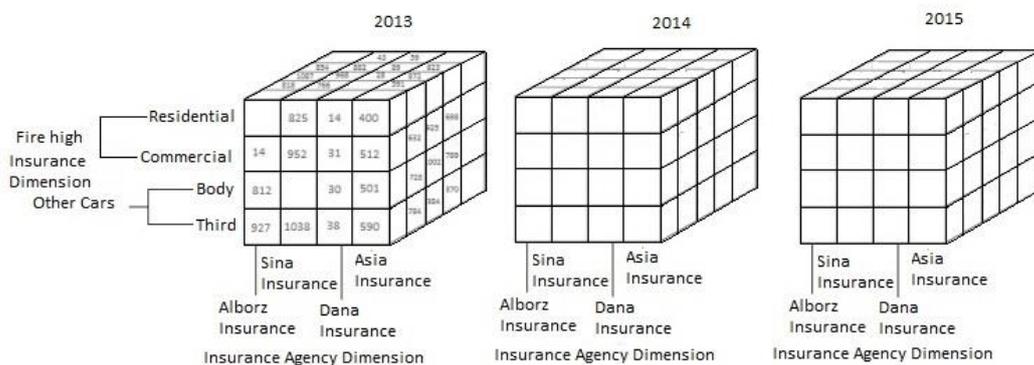


Fig.2. Displaying the four-dimensional data cubes of the insurance center

B. Finding a duplicate item set

Now, after creating data cubes, we present a collection of duplicate items with the aim of discovering associative rules. In this paper, we examine three different methods, and finally, we compare the results with the proposed method.

1. Using the Apriori algorithm

In the first method, we explore the associative rules using the original Apriori algorithm without using the

data cube, for this we use the weka data mining tool version 3.6. In order to compare the results of the runtime of this algorithm we calculate the algorithm's execution speed for number of records, and also for t different values for the amount of the backup factor. The results show that with increasing number of records, the duration of the algorithm has also increased. Also, in the event that the minimum backup factor is low, the number of repetitive and non-useful rules is created.

2. Using the cube of the Apriori data and attribute

In the second method, we use the data cube made without leveling in Oracle Olap, the Slice operation and the Apriori attribute. The Apriori feature is used to create sets of the same set of initial duplicate items at each step, and the number of candidate items in the set is reduced, resulting in a reduction in the number of repetitions and the time it finds duplicate items.

The second method is similar to the Apriori algorithm, only the amount of support with the data cube is calculated. To calculate the amount of support, we use the next-generation associative rules. These rules only include one dimension, to which we will tell the items

later, grouping this dimension with another dimension to which I will refer to the transaction later, generally two dimensions will be introduced into the next associative rules. So we will create a two-dimensional data cube using OLAP. To do this, we use Slice operations in OLAP, then we cut the multi-dimensional data cube and select one after that to explore the duplicate items. As shown in Fig. 3, here we have cut the data onto a 3D data cube and have segmented the dimensions of the insurance and insurance representation. Then we calculate the support value with the data cube, and the details of the algorithm are summarized in Fig. 4.

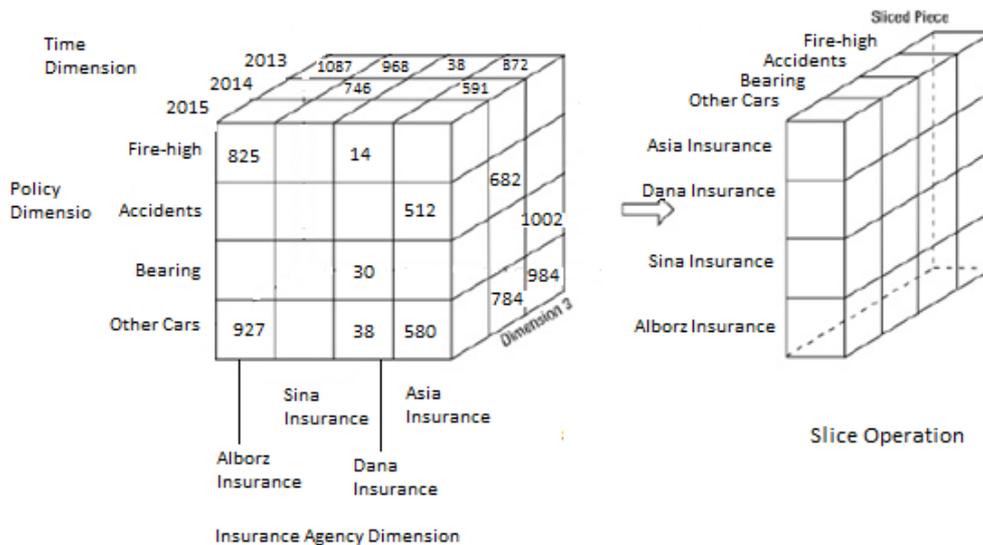


Fig.3.Slice operation on Data cube

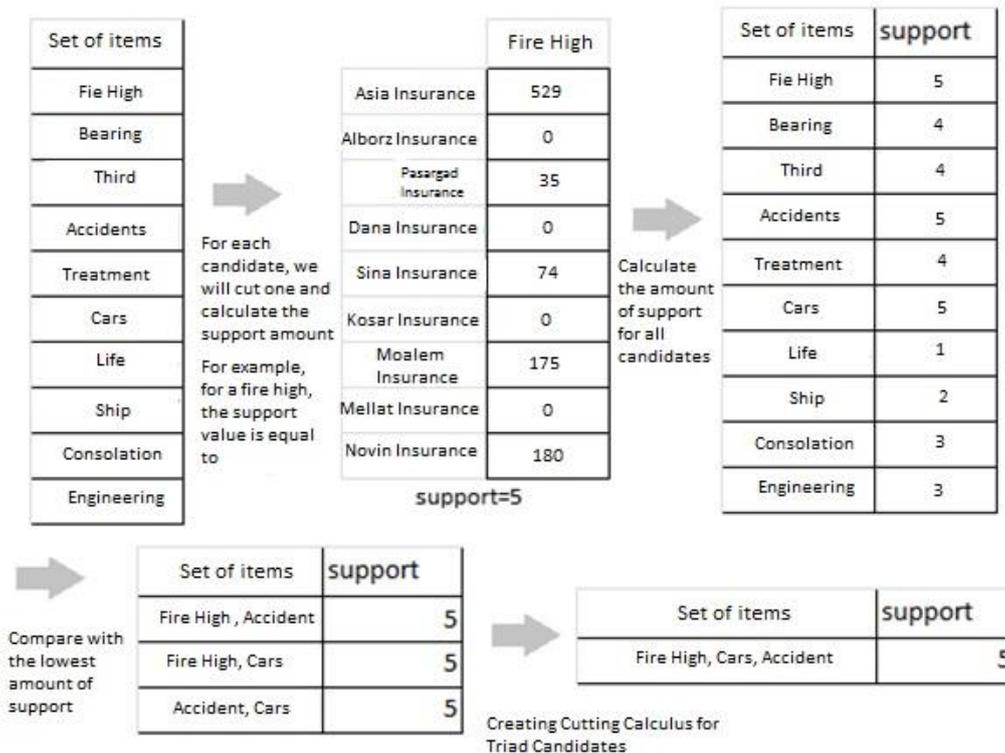


Fig.4. The steps for doing inter-dimensional associative rules

3. Using multi-dimensional and multi-level data cubes and the Apriori attribute

Finally, in the third method, we use the cube and the RollUp and DrillDown operations in OLAP, along with the Apriori feature. In this method, using the dimensions we designed and taking into account the different levels for the dimension of the time dimension and the dimension of the insurance policy in this paper, the data cube with different dimensions using Analytic

Workspace Manager software (AWM) Make up. In order to compare the methods, we first consider a 3D data packet with three dimensions of insurance agents, insurance policies and insurance coverage. We will record the results of these three dimensions. An example of this cube is shown in Fig. 5. It is clear:

The steps involved in this algorithm are summarized in Fig. 6 using the multidimensional and multi-level data cube and the Apriori attribute.

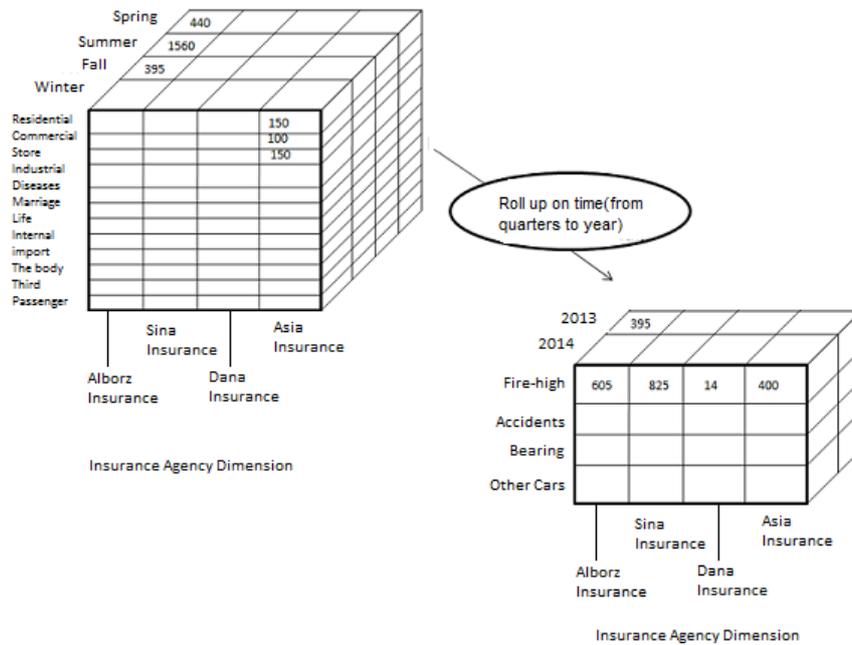


Fig.5. DrillDown operations on insurance policy data cube

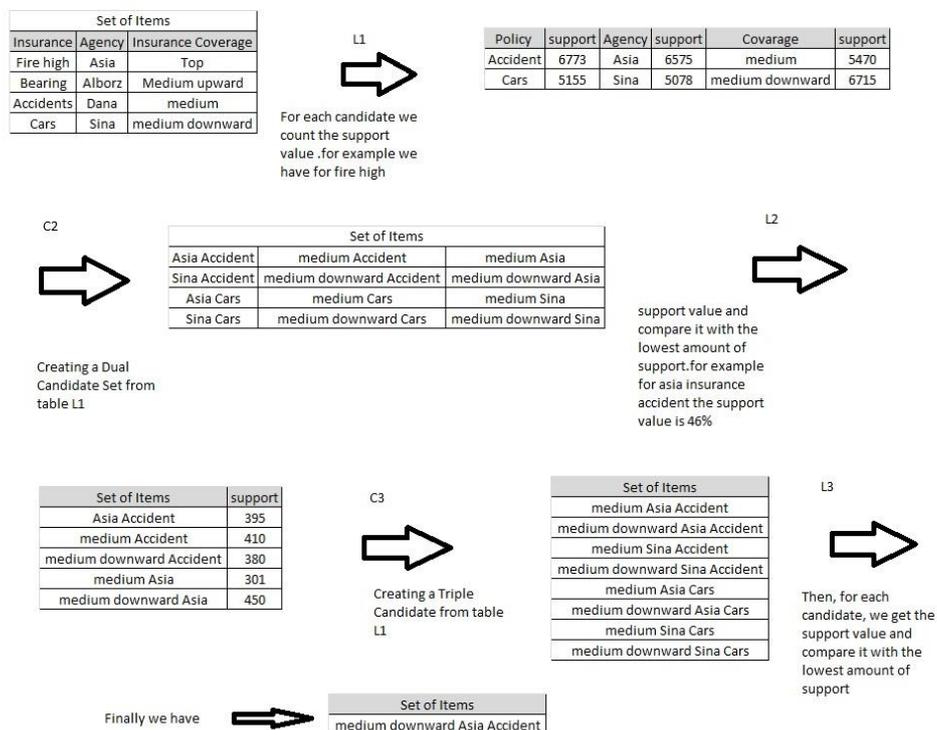


Fig.6. The steps for doing multidimensional and multi-level associative rules with Apriori attribute

If the time dimension of three levels of the year, season, and month is considered as the fourth dimension in this

data cube, the results of the runtime of this algorithm will be shown in Table 1.

Table 1. Runtime of this algorithm with number of different dimensions

Data cube	Number of dimension	Cube size	Support	Runtime
two dimensions	2	4140	0.5	116240
Three dimensions	3	20700	0.5	581200
four dimensions	4	1097100	0.5	3080360

C. Analysis and evaluation of explored rules

According to the results of the experiments, the number of sets of items, regardless of the level of data for the data is high, so the more time it takes to find a set of repetitive items. Among the extracted rules are:

1. Fire-high ==> Asia Insurance (Confidence: 100%, Support: 0.37)
2. Asia Insurance ==> Third-Low (Confidence: 93%, Support: 0.94)

In accordance with applicable laws, if the customer claims a fire insurance policy with high insurance coverage, with the possibility of 100, this insurance policy will be provided by Asia Insurance to customers. Also, 37% of the insurance companies have issued a fire insurance with high insurance coverage from Asia insurance. The law can provide 94% of the total customer with 93% third-party insurance, with low insurance

coverage from Asia. These rules for managers and analysts of the insurance organization can be very constructive and effective in determining future trends and plans.

In the next experiments to test the size of the data, first, the 10,000 records are considered as the initial data for creating the data cube, and then by considering the different values for the value of the support, also the default value of the two dimensions of the insurance policy and The insurance representative for the cube got different times, with the results shown in Fig. 7. As you can see, the duration of the multi-dimensional and multidimensional data cube has decreased, and with the increase in the amount of data, as shown in the graph, the time period has not changed much. This is contrary to the results of the Apriori algorithm found in Chart 5.1, which has greatly increased and increased with increasing data volumes.

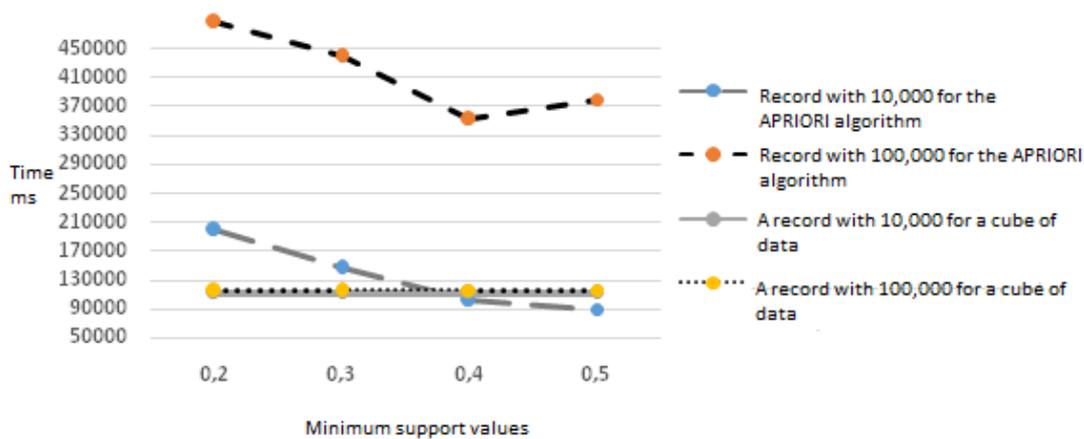


Fig.7. Comparison graph of time - amount of support for a number of different records

Fig. 8 shows the comparison of the proposed method with other methods and the algorithm written in the source [15]. As shown in the diagram, the duration of implementation in the proposed method, which is less

than the other two methods, is obtained from the results of our experiments. Of course, with the assumption that in all three methods the data cube is three-dimensional.

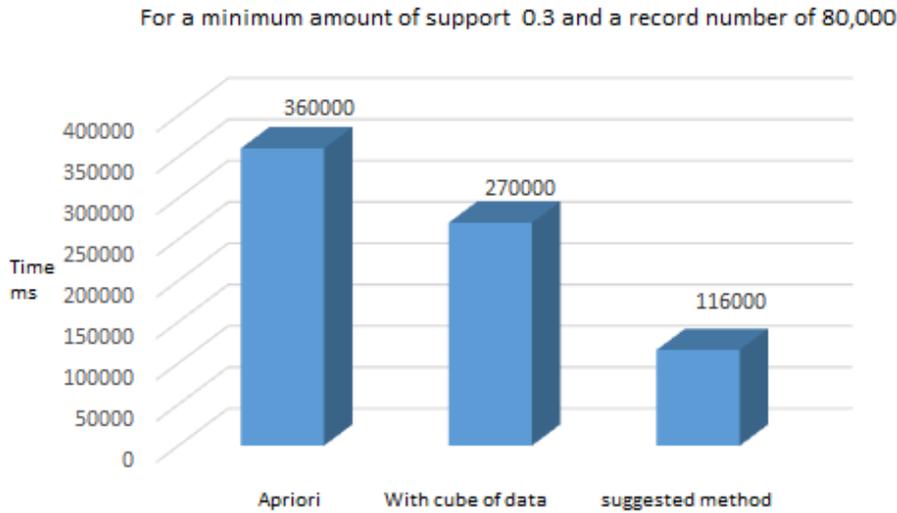


Fig.8. Compare the proposed method with other methods

As you can see, using the cubic technique of multidimensional and multilevel data as well as using the Apriori feature to reduce repetitive items at each stage reduces the number of repetitive items and reduces the duration of the algorithm, as well as increases the accuracy and optimization of the extracted rules.

Fig. 9 shows the results of the comparison of the increase in the magnitude of the data and thus the increase in the number of dimensions, taking into account the minimum backup factor of 0.5 in these experiments. According to this diagram, it is clear that with increasing the number of dimensions Runtime is also increasing.

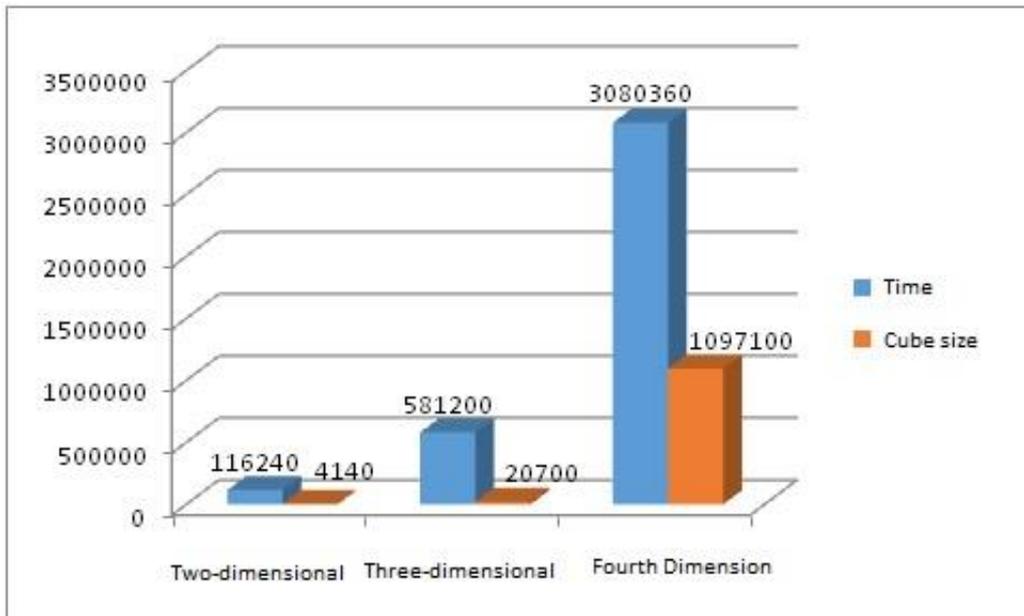


Fig.9. Compare the cube size and Runtime

VI. CONCLUSION AND FUTURE WORK

The results and evaluations of this method and other research have shown that the use of the Apriori feature is very effective. Also, the use of exploring multidimensional rules reduces the number of repetitive rules created. In addition, considering the dimensions of the cube of data in a hierarchical and multilevel manner, the rules created by the user are more interesting and

concise and more accurate. And the rules are created less and without repetition.

In the remainder of this paper, there are methods for automatic updating of data cubes, the use of other Olap operations to reduce the number of created candidate items, adding more dimensions to the sample DataSet, such as location dimension and optimization of written algorithms. To do the calculation of association rules.

Figs and Tables

REFERENCES

- [1] J.W.Seifert, 2004, "Data Mining: An overview", Congressional Research Service of the Library of Congress.
- [2] J. Han, M. Kamber, and J. Pei, 2011,"Data Mining: Concepts and Techniques": Morgan Kaufmann Publishers.
- [3] J. Han, M. Kamber, and J. Chiang, 1997 "Mining multidimensional association rules using data cubes," *Technical Report, Database Systems Research Laboratory*.
- [4] H. Zhu, 1998, "On-line analytical mining of association rules," *Doctoral Dissertation, Department of Computer Science, Simon Fraser University*.
- [5] S. Nestorov and N. Jukic, 2003, "Ad-hoc association-rule mining within the data ware-house," in *International Conference on System Sciences*, Page. 232-242.
- [6] C.K.H. Lee , K.L. Choy,G.T.S. Ho,K.S. Chin,K.M.Y. Law,Y.K. Tsec. 2013 "A hybrid OLAP-association rule mining based quality management system forextracting defect patterns in the garment industry" *International Journal of Expert Systems with ApplicationsIssues* Vol.66, Issues 7, Pages 52–62
- [7] Namchul Do, Sungmoon Bae,Chulsoon Park. 2015 "Interactive Analysis Of Product Development Experiments Using On-Line Analytical Mining" *International Journal Of Computers In Industry* Vol.40, Issues 7, Pages 2435–2446
- [8] Eltayeb Abuelyaman,Atifa Elgimari. 2014 "A Prototype For A Data Mining Based Pathfinder To Sudanese Universities" *International Conference On Computer Modelling And Simulation* Vol.8, Pages 119-124
- [9] H Dev, SK Mishra 2011 "Design of Data Cubes and Mining for Online Banking System" *International Journal of Computer Applications* Vol.4.
- [10] A. Guille and H. Hacid. 2013."Decision support system" in medical science using olap & data mining. In *International Journal of Computer Science and Application*, Vol. 6, No.2, pages 248–254.
- [11] JJ Jadav, M Panchal.2012. "Association rule" mining method on olap cube. *International Journal of Engineering Research and Application*, Vol.2, Issue 2, pages 1147–1151.
- [12] TS Abdellatif, MA Elsoud, HA Ali. 2011."Online analytical processing" and data mining tasks in enterprise resource planning. *IJCSI International Journal of Computer Science Issues* Vol.8, Issues 6, No 2, pages 161–173.
- [13] ISY Kwan, J Fong, HK Wong. 2005. An "e-customer behavior" model with online analytical mining for internet marketing planning. *Decision Support Systems - Elsevier*, pages 101–123.
- [14] WQ Qwaid. 2012. Apply On-Line Analytical Processing (OLAP) With Data Mining For Clinical Decision Support. *International Journal of Managing Information Technology (IJMIT)* Vol.4, No.1, pages 25–37.
- [15] Y. Wang et al., 2013, "Improved multi-level association rule in mining algorithm based on a multidimensional data cube," in *International Conference on Consumer Electronics, Communications and Networks*, Page. 355-358 .
- [16] A. Silva and C. Antunes, 2014, "Multi-relational pattern mining over data streams," *Data Mining and Knowledge Discovery*, Page. 1-32.
- [17] R. Kimball,Margy Ross,2013, "The Data Warehouse Toolkit", Third Edition.
- [18] Z. Ma, Q. Qiu, and P. Feng, 2014, "CPM algorithm for mining association rules from databases of engineering design instances," *Journal of Information Science and Engineering*, Vol. 30, Paper. 463-481.
- [19] Sheila A. Abaya, July-2012, "Association Rule Mining Based On Apriori Algorithm In Minimizing Candidate Generation", *International Journal of Scientific & Engineering Research*, Volume 3, Issue 7 .
- [20] Ms Shwetal Dr. KanwalGarg, June 2013 ,"Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithms", *International journal of Advanced Research in Computer Science and Software Engineering*, Volume 3, Issue 6.
- [21] Jayshree Jhal and Leena Ragha2, Number 2013, "Educational Data Mining using Improved Apriori Algorithm", *International Journal of Information and Computation Technology. Department of Computer Engineering* ,Volume 3, Paper. 411-418.
- [22] Jiao Yabing January 2013, "Research of an Improved Apriori Algorithm in Data Mining Association Rules ", *International Journal of Co Communication Engineering*, Vol. 2, No. 1.
- [23] Jyoti Arora1, Nidhi Bhalla2, Sanjeev Rao3, July 2013, "A Review on Association Rule Mining Algorithms", *International Journal of Innovat in Computer and Communication Engineering*, Vol. 1, Issue 5.
- [24] Mohammed Al-Maolegi1, Bassam Arkok2, February 2014, "An Improved Apriori Algorithm For Association Rules", *International Journal on Natural Language Computing (IJNLC)*, Vol. 3.
- [25] Jaishree Singh, Hari Ram and Dr.J.S.Sodhi, January 2013, "Improving efficiency of Apriori algorithm using Transaction Reduction", *International Journal of Scientific and Research Publications*, Volume 3, Issue 1.

Author's Profile



Mahtab Ebrahimi: works currently in the field of web-based services in the tax organization.she graduate of a senior engineer degree in computer engineering.

She expert in java programming and business intelligence.

How to cite this paper: Mahtab Ebrahimi, "Knowledge Extraction and Analysis to Evaluate the Financial Performance of an Organization Using OLAM", *International Journal of Modern Education and Computer Science(IJMECS)*, Vol.10, No.12, pp. 20-27, 2018.DOI: 10.5815/ijmeecs.2018.12.03