IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 5, Ver. IV (Sep.- Oct. 2017), PP 01-04
www.iosrjournals.org

Data Analysis Using Apriori Algorithm

Nikita Chandekar [1], Priyanka Ashtikar [2], Rahul Motghare [3]
Department Of Information Technology in Smt. Radhikatai Pandav College of Engineering Nagpur.
Corresponding Author: Nikita Chandekar

Abstract*- Development of several case projects throughoffering a graduate level course on Data Mining. It then outlines a particular case project that describes the process of data extracting, data cleansing, data transfer, data warehouse design, and development. It also outlines the development of a data cube as well as application to understand business intelligence. The results can be beneficial to an instructor who wants to develop a practical course or a practitioner venturing into the data warehousing and data mining area. Apriori algorithm has been improved and applied to the substation data mining process. Ant colony algorithm is applied to get the optimal solution of reactive power allocation in substations. The state transition probability formula is amended and parameters are dynamically adjusted in this ant colony algorithm. The choice of the ant's path to the next node is determined by the table formulated according to the confidence level of the data mining. The switching strategy of the capacitor sets is given by algorithm.*

Keyword*: apriori, warehousing, datamining, OLAP, pattern evaluation, business intelligence (BI).*

Date of Submission: 05-09-2017 Date of acceptance: 07-10-2017

I. Introduction

The historical data in the warehouse play an important role in providing Business Intelligence (BI) that helps companies to streamline workflows, provide better customer services, and target market their products and services. Software development companies are also focused on developing new tools and technologies for data warehousing engines, providing data transfer services from traditional sources to data warehouses, performing analysis for business intelligence, generating reports and ad-hoc queries, and executing data mining algorithms. Other companies having significant market share of the data warehousing Many IT consulting companies help large companies develop and maintain their data warehousing Demand for personnel with specific IT skills in the data warehousing and BI technologies have also been growing.

The need of project:

Many IT consulting companies help large companies develop and maintain their data warehousing Demand for personnel with specific IT skills in the data warehousing and BI technologies have also been growing. Today, a search for data warehousing, BI data mining returns thousands of jobs scattered across the nation. The use of data warehousing and BI technology span sectors such as retail, airline, banking, health, government, investment, insurance, manufacturing,telecommunication, transportation, hospitality, pharmaceutical, and entertainment. Due to increasingly stringent budgets, rising operational costs, and competition from online universities, many educational institutions recently adopted data warehousing and BI technology to improve their business processes. Universities are using BI tools in areas such as academics, enrollment, financial aid, alumni, development, finance, and human resources. Discuss many important issues upon which universities can focus their data warehousing efforts. While almost all business sectors, government agencies, and academia moved into adopting data warehousing and BI tools, and there are significant demands for skilled personnel in these areas, the faculty members in computing and programs that are expected to teach the knowledge and skills necessary to prepare their students for the rising job market are lagging behind. Although a systematic research has not been done to find out how many Universities offer such a course, a simple search of the web or journal databases reveals very few course offerings or papers in relation to teaching data warehousing and/or data mining.

1. Situation before the initiative

The day-to-day operations of the company rely heavily upon the system. Everyone from the purchasing department to branch managers to accounts receivable relies on it for current information to make normal operational decisions. For the most part, it does what it is supposed to do – it tells users what is currently happening in the company. Yet the system is painfully inadequate when it comes to strategic decision support. These types of information requests from management must be dealt with individually by the information technology (IT) staff. Data aggregations are programmed into reports, but any comparisons across time or

products must be done manually. Data history in the system typically goes back 2 years, even though the company has been generating computerized data for over 20 years. A data warehouse is a solution. It will provide a central repository for historical data. It will provide an integrated platform for historical analysis of sales data. It will allow the application of**2.** online analytical processing (OLAP) techniques by users themselves. With a data warehouse and, we expect to empower users to perform their own roll-up and drill-down operations to analyze sales across product categories, subcategories, store regions, individual stores, or any combination desired. They will have the flexibility to view data and immediately look at data in another form without sending a request to IT for a new report. They will enjoy a true decision support system that will provide strategic analysis in a user-friendly format.

II. PROPOSED SYSTEM

Calculate the support level of the candidate set on the database scanning and pattern matching. It can be included that the candidate set is too large and the database is scanned repeatedly in the Apriori algorithm.
An improved method without these two drawbacks is applied to the data mining in the historical database of the substations. It is described as follow:

(i) Preprocess the original data based on partition. It divides the database of the substation into 9-zones according to the requirement of reactive power and bus voltages. Then it focuses on the data in the area except for the normal running area. So it is time-saving and fast-accessing because it only scans the corresponding area in the database without scanning the whole database.
(ii) Classify with similarity search, according to central substation operation conditions. The association level of the selected data is improved to meet the requirements of practical operation.
The improvement of the ant colony includes:
(i) Selection of parameters: The parameters are dynamically adjusted. At the beginning, the parameters are set at a small value, to avoid "false positive feedback" and "solution loss". When the calculation is running after a certain number of cycles, the parameters are increased to improve the solution quality.
(ii) Modification of the parameters: The state transition probability in (2) is modified according to the results of data mining. The higher the confidence level and the pheromone concentration are, the greater the probability that ants choose.

Objective:
Extracting useful data for effective decision-making of reactive power optimization. It describes the concepts and improvements of association rules algorithm - Apriori algorithm and ant colony algorithm. The improved Apriori algorithm is applied to extract.

Scope:
A large number of running data in the substation operation process. The overall model based on Apriori algorithm and ant colony algorithm is established for reactive power optimization. An example power substation is used to illustrate the application of the proposed models in the voltage and reactive power automatic control system. Based on historical data, the proposed method is used to get the optimal operating conditions of the optimal solution to guide the practical operation.

Proposed Approach:

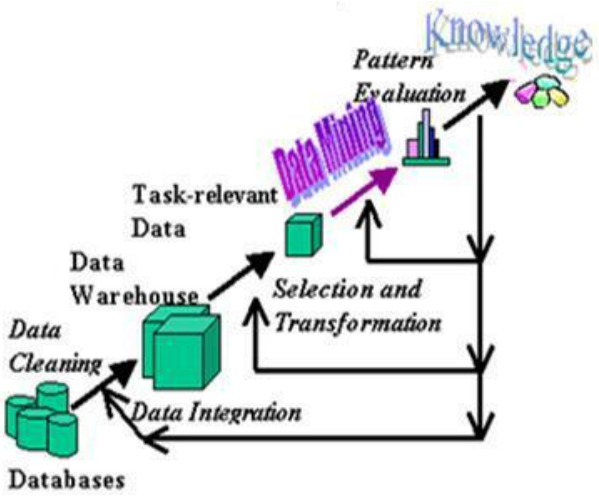

Fig.: Dataset Analysis System

1. Overview:

Customers are included in three hierarchies within the customer dimension, it is important to note that each customer will be found in each of the three hierarchies. This provides three different ways to look at summarized customer data. Looking at the Time dimension should be straightforward until you come to the "Season" attribute. Because of the nature of the business, seasonal sales differences can be an important analytical topic.

2. Services Provided By System

The improved Apriori algorithm is applied to extract the useful information for the ACA from a large number of running data in the substation operation process. The overall model based on Apriori algorithm and ant colony algorithm is established for reactive power optimization. An example power substation is used to illustrate the application of the proposed models in the voltage and reactive power automatic control system. Based on historical data, the proposed method is used to get the optimal operating conditions of the optimal solution to guide the practical operation.

3. Tools and Technologies used:

The purpose of the data collection and any data mining projects, how the data will be used, who will be able to mine the data and use them, the security surrounding access to the data, and in addition, how collected data can be updated.

III. Literature Survey

Supporting literature exists focusing on banknote recognition software intended for a variety of applications such as assisting the visually impaired [2], banknote sorting and Automatic Teller Machine (ATM) software [3], and banknote fatigue detection [4]. A trend observed is that an image is acquired, it is pre-processed then classified and finally the result is output. Various methods are employed at each stage; the correct combination to use is subjective to the currency in question. The same workflow is employed by our system to classify the note. Image acquisition and image preprocessing techniques are employed; either the entire note or distinct Regions of Interest (ROI) are acquired and compared independently. The serial number uniquely identifies an individual banknote, in some cases, the location of manufacture can be pinpointed [1]. Therefore this adds a layer of security and can be computed [5]. If a note is found to have an incorrect or duplicate serial number, it is not authentic. Image acquisition techniques are explored, the aggregation of RGB color is with ultra-violet information [6]. Under ultraviolet light a different visual appearance is observed, specific here, the aim of edge detection is basically to localize the currency note that is the region of interest.

Reactive power plays an important role in supporting the real power flow by maintaining voltage stability and system reliability. The available reactive power capabilities of the system have to be optimally deployed so that bus voltage are kept within specified limits. The purpose of reactive power dispatch is to determine the proper amount and location of reactive support with several constraints.

The paper focuses on the voltage/reactive power problem keeping the real power flows fixed to values determined from a base case load flow analysis. In paper [14], optimal power dispatch is solved by time varying acceleration coefficients particle swarm optimization (TVAC-PSO).It proposes a comprehensive model for reactive power pricing in an ancillary services market. Paper [15] presents an efficient Genetic Algorithm (GA) based reactive power optimization approach to minimize the total support cost from generators and reactive compensators. This paper focuses on the problem of extracting useful data for effective decision-making of reactive power optimization. It describes the concepts and improvements of the association rules algorithm-Apriori algorithm.

IV. Conclusion

As mentioned earlier, many of the executives within this privately held company are very skeptical towards new technology and resist change. However, even someone with this mindset is impressed when they are given a new set of tools that give them the ability to make more informed decisions. The power of the OLAP tools alone implemented in this project would make a very persuasive argument for the implementation of a full-scale data warehouse. With management buy-in also comes new ideas for aggregation levels that can be added or modified to fit the user's analytical needs. This may be the only way to incorporate views of the data that have been buried in obscure reports or that have possibly never been implemented before due to their complexity. We would include more dimensional attributes and actual data to enable a meaningful big data Hadoop effort, which we have tried rather unsuccessfully with the current data.

References

[1]. WierschemD.Mc.Millen.J. And McBroom, R., (2003),
[2]. "What Academia Can Gain from Building a Data Warehouse," Educes Quarterly, Number 1, pp. 41-46.
[3]. Fang, R. and Tuladhar, S. "Teaching Data Warehousing and Data Mining in a Graduate Program in Information Technology, "Journal of Computing Sciences in Colleges, Vol.21, Issue 5, pp. 137-144.
[4]. Pierce, E. M., "Developing and Delivering a Data Warehousing and Data Mining Course," Communications of the AIS, Vol. 2, Article 16, pp. 1-22.
[5]. Slazinski, E. D., "Teaching Data Warehousing to Undergraduates – Tales from the Warehouse Floor," CITC'03.
[6]. Ponniah, P., Data Warehouse Fundamentals, a Comprehensive Guide for IT Professionals; John Wiley & Sons, New York.
[7]. Jacobson, R., Microsoft SQL Server 2000 Analysis Services, Step by Step; Microsoft Press, Redmond, Washington. 8. Inman, W. H., Wiley & Sons, New York.
[8]. R. Agrawal, C. Faloutsos, and A. Swami. Efficient similarity search in sequence databases. In Proc. of the Fourth International Conference on Foundations of Data Organization and Algorithms, Chicago, October 1993.
[9]. R. Agrawal, S. Ghosh, T. Imielinski, B. Ayer, and A. Swami. An interval classi_er for database mining applications. In Proc. of the VLDB Conference, pages 560{573, Vancouver, British Columbia, Canada, 1992.
[10]. R. Agrawal, T. Imielinski, and A. Swami.Database mining: A performance perspective. IEEE Transactions on Knowledge and Data Engineering, 5(6):914{925, December 1993. Special Issue on Learning and Discovery in Knowledge-Based Databases.
[11]. R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In Proc. of the ACM SIGMOD Conference on Management of Data, Washington, D.C., May 1993.
[12]. R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. Research Report RJ 9839, IBM Alma den Research Center, San Jose, California and June 1994.
[13]. D. S. Associates. The new direct marketing.Business One Irwin, Illinois, 1990.
[14]. R. Brachman et al. Integrated support for data archeology. In AAAI-93 Workshop on Knowledge Discovery in Databases, July 1993.
[15]. L. Breiman, J. H. Friedman, R. A. Olsten, and C. J. Stone. Classification and Regression Trees.Wadsworth, Belmont, 1984.
[16]. P. Cheeseman et al. Autoclass: Abayesian classification system. In 5th Int'l Conf. on Machine Learning. Morgan Kaufman, June 1988.Report CS-R9406, CWI, Netherlands, 1994.
[17]. M. Houtsma and A. Swami. Set-oriented mining of association rules. Research Report RJ 9567, IBM Alma den Research Center, San Jose, California, October 1993.
[18]. R. Krishnamurthy and T. Imielinski. Practitioner's problems in need of database research: Research directions in knowledge discovery. SIG-MOD RECORD, 20(3):76{78, September 1991.
[19]. P. Langley, H. Simon, G. Bradshaw, and J. Zytkow. Scientific Discovery: Computational Explorations of the Creative Process. MIT Press, 1987.
[20]. H. Manila and K.-J. Raiha. Dependency inference. In Proc. of the VLDB Conference, pages 155{158, Brighton, England, 1987.
[21]. H. Mannila, H. Toivonen, and A. I. Verkamo. Ancient algorithms for discovering association rules. In KDD-94: AAAI Workshop on Knowledge Discovery in Databases, July 1994.
[22]. S. Muggleton and C. Feng. Efficient induction of logic programs. In S. Muggleton, editor, Inductive Logic Programming. Academic Press, 1992
[23]. J. Pearl. Probabilistic reasoning in intelligent systems: Networks of plausible inference, 1992.
[24]. G. Piatestsky-Shapiro. Discovery, analysis, and presentation of strong rules. In G. Piatestsky-Shapiro, editor, Knowledge Discovery in Databases. AAAI/MIT Press, 1991.
[25]. G. Piatestsky-Shapiro, editor. Knowledge Discovery in Databases. AAAI/MIT Press, 1991.
[26]. J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufman, 1993.

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Nikita Chandekar . "Data Analysis Using Apriori Algorithm." IOSR Journal of Computer Engineering (IOSR-JCE) , vol. 19, no. 5, 2017, pp. 01–04.