LITERATURE SURVEY

Project Title : Global Sales Data Analytics
Team Leader : AZHAR AYYASH. A.
Team Member 1 : MUHAMMED THAHIR. A.
Team Member 2 : ABDUS SHAKOOR. H.H.
Team Member 3 : ABDULLAH

ABSTRACT

In the information era, enormous amounts of data have become available on hand to decision makers. Big data refers to data sets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Due to the rapid growth of such data, solutions need to be studied and provided in order to handle and extract value and knowledge from these data sets. Furthermore, decision makers need to be able togain valuable insights from such varied and rapidly changing data, ranging from daily transactions to customer interactions and social network data. Such value can be provided using big data analytic, which is the application of advanced analytic techniques on big data. This paper aims to analyze some of the different analytic methods and tools which can be applied to big data, as well asthe opportunities provided by the application of big data analytic in Global sales domains.

INTRODUCTION

Current advances in information technology (IT) and the rising trend of social media have changed the way salespersons perform daily routine activities. Most often, the Salesforce is equipped with a Salesforce automation (SFA) system to enhance customer relationship management (CRM) capabilities and sales performance . SFA systems are a set of tools that facilitate organization by

providing analyzed information from available data to manage customer relationships and sales-related activities . An SFA system provides information regarding customer interactions, inventory control, sales forecasting, sales, communication history, and pipeline opportunities to efficiently achieve

day-to-day goals . Organizations annually invest millions of dollars in the implementation of SFA systems to achieve excellent customer relations and sales progress .

Imagine a world without data storage; a place where every detail about a person or organization, every transaction performed, or every aspect which can be documented is lost directly after use. Organizations would thus lose the ability to extract valuable information and knowledge, perform detailed analyses, as well as provide new opportunities and advantages. Anything ranging from customer names and addresses, to products available, to purchases made, to employees hired, etc. has become essential for day-to-day continuity. Data is the building block upon which any organization thrives. Now think of the extent of details and the surge of data and information provided nowadays through the advancements in technologies and the internet. With the increase in storage capabilities and methods of data collection, huge amounts of data have become easily available. Every second, more and more data is being created and needs to be stored and analyzed in order to extract value. Furthermore, data has become cheaper to store, so organizations need to get as much value as possible from the huge amounts of stored data. The size, variety, and rapid change of such data require a new type of big data analytic, as well as different storage and analysis methods. Such sheer amounts of big data need to be properly analyzed, and pertaining information should be extracted.

The contribution of this paper is to provide an analysis of the available literature on big data analytic. Accordingly, some of the various big data tools, methods, and technologies which can be applied are discussed, and their applications and opportunities provided in several decision domains are portrayed. The literature was selected based on its novelty and discussion of important topics related to big data, in order to serve the purpose of our research. The publication years range from 2008-2013, with most of the literature focusing on big data ranging from 2011-2013. This is due to big data being a recently focused upon topic.

Furthermore, our corpus mostly includes research from some of the top journals, conferences, and white papers by leading corporations in the industry. Due to long review process of journals, most of the papers discussing big data analytic, its tools and methods, and its applications were found to be conference papers,

and white papers. While big data analytic is being researched in academia, several of the industrial advancements and new technologies provided were mostly discussed in industry papers.

USE CASES

1. Predicting Sales

Predicting sales is of immense importance to organizations as its effects trickle down to critical business processes like inventory management, logistics, production and manpower planning. For instance, buying raw material and maintaining finished goods inventory is fundamentally driven by sales forecast. Accurately predicting sales helps organizations to make better decisions and ensure the smooth running of processes.

Sales forecasting algorithms use a large amount of diverse data to look for patterns and relationships among various factors that affect sales under changing circumstances, thus predicting sales with a high level of accuracy.

2. Improve lead generation

Analytic has proven to be a great tool to improve lead generation and automate pre-sales processes. Companies are leveraging a vast resource of data to identify the right customers at the right time. Enterprises use a wide array of historical data to get a holistic picture of their prospective sales and many companies are pushing the limit by deploying lead-scoring algorithms which are fueled by granular and segmented data about each of their prospects. A complete 360-degree view of the customer is generated by combining in-house customer data and external data from news reports and social media posts.

These algorithms guide sales strategies by predicting the factors that are pivotal to lead conversion . According to a report by McKinsey, big-data analytic can be used to predict leads that are most likely to close which is

useful in planning the allocation of resources to improve lead conversion rate.

By employing intelligent automation into the insight generation process, companies are seeing a significant leap in their ability to identify promising prospects and zero in on the right moment to target them. Enterprises are testing AI-enabled agents powered by predictive analytics and natural language processing to automate pre sales activities and early lead-generation activities.

3. Analyzing customer sentiment

Sentiment analysis proves useful in understanding the feedback by customers. It employs AI to discern the emotions conveyed by customers and the semantics of the conversation. This is beneficial for businesses to understand how customers perceive their brand.

Sentiment analysis uses text mining algorithms to extract insights from social media websites, blogs or review sites. Automated sentiment analysis tools can be useful in extracting real-time actionable insights.

4. Better Cross-selling and Up-selling

With data analytic, companies can have an understanding of how their up sell and cross-sell strategies will perform well beforehand and also identify important sales parameters like key value items, key value categories, popular products and high demand products that can affect the sales bottom line. Data science is also used to provide personalized cross-selling recommendations which suggest additional products that a customer would desire to buy along with an item already bought or intended to buy

EXISTING SYSTEM

The Data mining Algorithms can be categorized into the following :

- Association Algorithm

•Classification

•Clustering Algorithm

1 Classification:

The process of dividing a data set into mutually exclusive groups such that the members of each group are as ”close” as possible to one another, and different groups are as ”far” as possible from one another where distance is measured withrespect to species variable(s) you are trying to predict for example, atypical classification problem is to divide a database of companies into groups that are as homogeneous as possible with respect to a creditworthiness variable with values ”Good” and ”Bad.”

2 Clustering:

The process of dividing a data-set into mutually exclusive groups such that the members of each group are as ”close” as possible to one another, and different groups are as ”far” as possible from one another, where distance is measured with respect to all available variables given databases of sufficient size and quality, data mining technology can generate new business opportunities by providing these capabilities

- •Automated prediction of trends and behaviors
- •Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data quickly. A typical exampleof a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events.

- •Automated discovery of previously unknown patterns

Data mining tools sweep through databases and identify previously hidden patterns in one step. DARM discovers rules from various geography-ically distributed data sets. However, the network connection between those data sets isn’t as fast as in a parallel environment, so distributed mining usually aims to minimize communication costs.

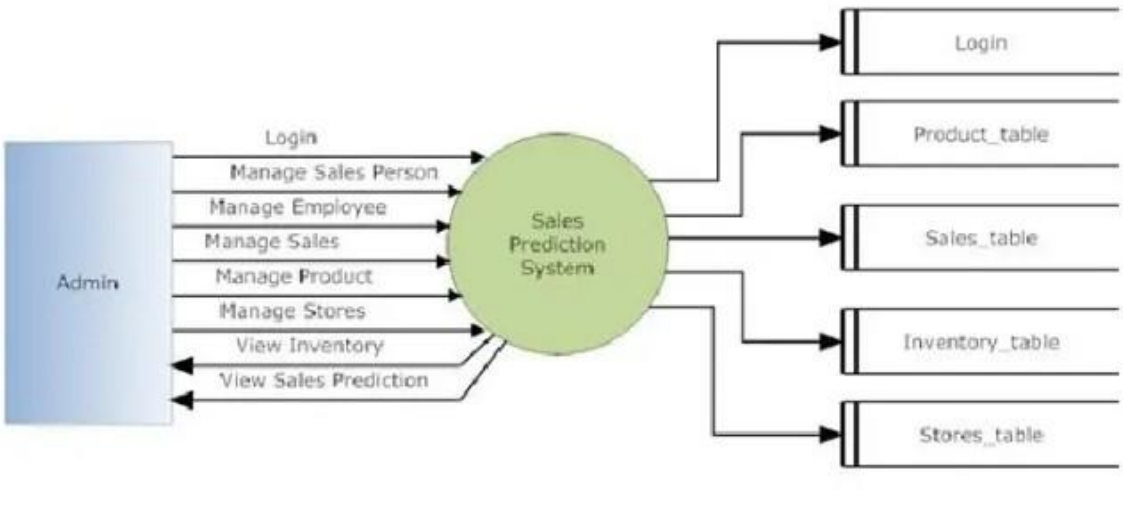
2.3 PROPOSED SYSTEM

•Unlike other algorithms, ODAM offers better performance by minimizing candidate item set generation costs. It achieves this by focusing on two major DARM issues communication and synchronization. Communication is one of the most important DARM objectives. DARM algorithms will perform better if we can reduce communication (for example, message exchange size) costs. Synchronization forces

•Each participating site to wait a certain period until globally frequent itemset generation completes.Each site will wait longer if computing support counts takes more time. Hence, we reduce the computation time of candidate item sets' support counts.

• Each method has different aims, expectations, advantages, and disadvantages. For example, the first method exchanges each candidate item set's support count to generate globally frequent item sets of that pass (CD and FDM are examples ofthis approach)

Data Flow Diagram



CONCLUSION

Distributed ARM algorithms must reduce communication costs so that generating global association rules costs less than combining the participating sites' data sets into a centralized site. We have developed an efficient algorithm for mining association rules in distributed databases.

•Reduces the size of message exchanges by novel local and global pruning.
•Reduces the time of scan partition databases to get support counts by using a compressed matrix-C Matrix, which is very effective in increasing the performance.

•Founds a center site to manage every the message exchanges to obtain all globally frequent item-sets, only O(n) messages are needed for support count exchange. This is much less than a straight adaptation of Apriori, which requires O(n2) messages for support count exchange.

References

[1] "A Fast Distributed Algorithm for Mining Association Rules",Proc. Parallel and Distributed Infor-mation Systems; D.W. Cheung, et al., IEEE CS Press, 1996,pp. 31-42

[2]"Introduction: Recent Developments in Parallel and Distributed Data Mining",J. Distributed andParallel Databases; M.J. Zaki and Y. Pin, vol. 11, no. 2, 2002,pp. 123-127

[3]"Efficient Mining of Association Rules in Distributed Databases",IEEE Trans. Knowledge andData Eng.; D.W. Cheung , et al., vol. 8, no. 6, 1996,pp. 911-922

[4] "Communication-Efficient Distributed Mining of Association Rules"; A. Schuster and R. Wolff,Proc. ACM SIGMOD Int'l Conf. Management of Data, ACM Press, 2001,pp. 473-484

[5] "Mining Association Rules Between Sets of Items in Large Databases"; R. Agrawal, T. Imielinski,and A. Swami, Proc. ACMSIGMOD Int'l Conf. Management of Data, , May 1993

[6] "An Optimized Distributed Association Rule Mining Algorithm"; M.Z Ashrafi,

Monash UniversityODAM, IEEE DISTRIBUTED SYSTEMS ONLINE 1541-4922
2004

[7]”The Data Warehouse Toolkit, The Complete Guide to Dimensional Modeling”,2nd edn. John Wi-ley & Sons; Kimball, R., Ross, M., New York (2002)

[8] ”Web for Data Mining: Organizing and Interpreting the Discovered Rules Using theWeb”,SIGKDD Explorations; Ma, Y., Liu, B., Wong, C.K., Vol. 2 (1). ACM Press, (2000) 16-23.

[9]”New Algorithm for Fast Discovery of Association Rules”,Technical ReportNo. 261; Zaky, M.J., Parthasarathy, S., Ogihara, M., Li, W., University of Rochester(1997),http://cs.aue.aau.dk/contribution/projects/datamining/papers/tr651.pdf