Data Visualization For Wi-Fi Hotspots Through Data Mining Using D3

Suraj Bansode[1], Swapnil Bhalerao[2], Nitin Mhadik[3]

Computer Engineering, D.Y. Patil Institute Of Engineering And Technology Ambi / Savitribai Phule University Pune, India

Abstract :- Data visualization is an effective tool to explore data that is stored in huge dynamic Databases. The government, corporate and industries face a problem of increasing no of databases. Data visualization for Wi-Fi hotspot that is the application will produce the standard graphical representation of hotspot data. D3 is a dynamic library that will be mainly used for manipulating documents based on data. The data represented will be retrieved after the data mining phase and given as an input to the visualization graphs. The representation will include all the details like categorization of data collected at the hotspot according to date on yearly, weekly and daily basis. It also includes number of users at particular hotspot (paid or unpaid),which websites are being visited or mostly preferred by the users, user's system configuration, Billing information visualization. Technologies mainly used are PHP,D3 library, SON data format,MySQL,HTML and CSS.

Keywords :- Hotspot ,Data Acquisition, Data Mining, Data Visualization, Data Optimization,

I.   INTRODUCTION

Electronic media that stores the large amount of data is growing exponentially fast. That is particularly data related to management of networks like customer log, network traffic details etc. These databases that are large need not to be managed, but it also has to be explored. Making sense of such huge type data is becoming harder and more challenging.[1] So the integration of data visualization and data mining can employ a way to explore and identify the structure and patterns in these huge amount of databases. An exploratory tools requires for supporting the identification of domain and mission critical elements such as patterns in data access (e.g., security breach determinations), patterns in the data (e.g., marketing and clustering), or for patterns in transactions (e.g. data compression), to site a few Knowledge Discovery in Databases is a relatively new research area that employs a variety of tools to explore and identify structure and patterns in these large databases.

1.1 KEY TERMS

Following are the important terms that sholud be known to the user of the project they mainly include hotspot, data acquisition, data optimization, data mining, and data visualization.

1.1.1   Hotspot

A hotspot is a site that offering Internet access over a wireless local area network (WLAN) through the use of a router.The router connected to a link to an Internet service provider. These hotspots typically use Wi-Fi technology.Special application software require for some hotspots that to be installed on the client of Wi-Fi, mainly for the security and billing purposes, but some others not require any configuration other than knowledge of the network name (SSID). The Wireless service providers like T-Mobile generally own and maintain hotspots.The majority of hotspots require hourly, daily, monthly payment or other subscription fees. Hotspot providers strive to make connecting Wi-Fi clients as simple and being secure to the users as possible. Generally public hotspots provide less secure Internet connections than do other wireless business networks.

1.1.2   Data Acquisition

Data acquisition systems are the products and or processes that are used to collect information ,document or analyze some phenomenon. It is a set of processes and programs that extracts the data for the database, and operational data store from the operational systems. Cleansing as well as the integration of the data and transformation into an enterprise format is performed by the data acquisition programs. This enterprise format reflects an integrated set of enterprise business rules. That this enterprise format usually causes the data acquisition layer that to be the most complex component. In addition to programs that transform and clean up data,audit and control processes and programs included in the data acquisition layer to ensure the integrity of the data as it enters the database or operational data store.

1.1.3   Data Mining

The Data Mining is the process used by companies to convert or turn the raw data into useful information. Software is used to look for the patterns in large batches or amount of data, businesses can learn more about their customers So they can develop more effective marketing strategies as well as increase sales and decrease costs. As the name implies Data mining depends on effective data collection and ware housing and also processing of a computer. The well-known users of data Mining are Grocery stores. Data mining can be a cause for concern when only selected information, which is not representative of the all sample group is used to prove a some hypothesis. Data Mining (DM) is the mathematical core of the KDD process.

1.1.4   Data Optimization

Preparing the logical schema from the data view schema is done by the Data Optimization process. Data Optimization is the counterpart of data de-optimization. So Data optimization is an important aspect in management of database, in dataware house it is also the important aspect. Data optimizations the data optimization is nonspecific technique used by several applications in fetching data from a data sources. Therefor these data could used in data view tools and applications such as that would used in statistical reporting. Data optimization can be achieved by the data mapping.

1.1.5   Data Visualization

Data.Information that has been abstracted in some schematic form to study the visual representation of the Data visualization is necessary including attributes or variables for the units of information. Data visualization is related to visualization of information, graphics realted to information, visualization of scientific terms and statistical graphics. [2] As data mining techniques mature, it will be important to integrate them with visualization techniques.The approach is to use visualization techniques to present the results that are obtained from mining the data in the databases. These results may be in the form of clusters or they could specify correlations between the data in the databases. For example, there may be a correlation between house prices and apartment buildings, such as the neighbor hoods which have high priced houses and do not have any apartment buildings. Therefore, when one queries for houses prices in a particular region, the data mining could also extract the information about correlations between house prices and apartment building. This information could be displayed in visual form by showing the house prices and graphically displaying the apartment buildings in a map. In case Wi-Fi hotspot management there are various hotspots that are geographically distributed. Hotspot offers internet access over wireless network. It typically uses Wi-Fi technology. Hotspots found in public places like coffee shops, hospitals, hotels etc. In the process of providing internet access to customers data related to network, its traffic, customer details, their accounting and billing increases rapidly. Data has to be managed as it increases for better utilization. In this case Data Mining and Data visualization plays a vital role and provides a way to analyze complex data in efficient way. Data Mining is used for identifying new patterns and in sighting in data. The database grows ,because the volume of data collected and

stored in databases, identify important patterns and trends there is a growing need to provide data summarization.Whereas Data visualization is the process of representing data as a visual image in which an image is created using a combination of points, lines, coordinate systems, numbers, symbols, words, shadings, and colors to represent different measured quantities.

II.   RELATED WORK

In what way can we convey data to users effectively? Data Visualization aims to communicate data clearly and effectively through graphical representation. Data visualization has been used extensively in many applications. g., work for reporting, managing business operations, and tracking progress of tasks. The same way in proposed system data visualization would be mapped to the raw data collected by the system through the data mining and knowledge discovery process. Data Visualization for Wi-Fi Hotspots through Data Mining using D3 intends to the integration of data mining and data visualization in order to visualize the data stored in the database tables into a graphical format that will be easily understandable to the user mainly for analysis purpose.
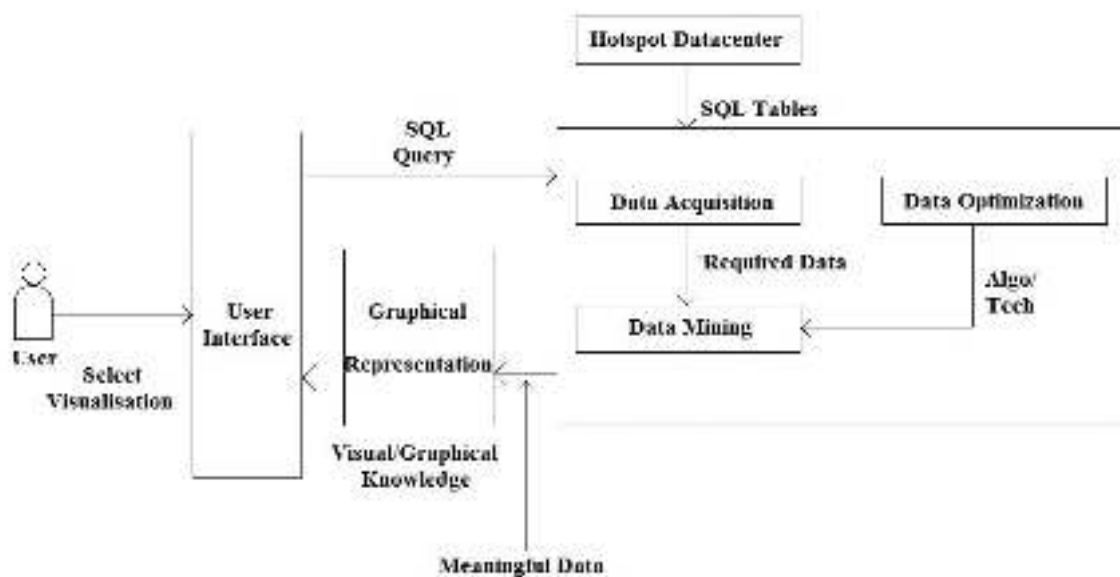
III.   PROPOSED SYSTEM ARCHITECTURE

Fig 1: Overall System Architecture

The proposed system consists of a web-based user interface with multiple menu choices available data interpretation and information visualization. Above figure shows the overall system architecture for the proposed system with the abstract view of its major components. As shown in the figure, user and data at hotspot datacenter are the two external entities for the system. Firstly user will initiate the session by selecting the menu choice from the interface provided. Respectively data will be requested from the database (hotspot datacenter) with the help of database queries and at this phase required raw data is fetched from the database.In next phase data mining techniques (prediction, pattern extraction) are applied to this raw data along with the optimization techniques, if needed. It is the second last phase of the process where raw data is transformed into meaningful form. Lastly the visualization phase comes into picture where meaningful data will get represented in the graphical form, such huge processed data which we are getting after the data mining phase from distributed Wi-Fi hotspot.

IV.   CONCLUSION

The various hotspots that are geographically distributed can be managed effectively using the visualized graphs. Visualization provides a broader view of the data that is being stored into the SQL tables. The graphs will play a vital role and provide a way to analyze complex data in efficient way and also publish the information about the various regions like the amount of data usage in the region according to the specified dates, active users and the browser and os types used by the particular user, and the revenue gained by the company from the various plans provided.

V.   ACKNOWLEDGEMENTS

REFERENCES

[1] Neelamadhab Padhy1, Dr. Pragnyaban Mishra 2, and Rasmita Panigrahi *"The Survey of Data Mining Applications And Feature Scope*" (IJCEIT)vol 2 june 2012

[2] Dr. Georges Grinstein, Dr. Bhavani Thuraisingham*," Data Mining and Data Visualization: Position Paper for the Second IEEE Workshop on Databases Issues for Data Visualization"*

[3] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth,   "From Data Mining to Knowledge Discovery Databases"

[4] *Dataviz: Improving Data Visualization for the public sector*.

[5] *Scaling up data mining algorithms: review and taxonomy* (13 january 2012)

[6] Patrick Ozer," *Data Mining Algorithms for Classification"*

[7] Jiawei Han , Micheline Kamber," *Data Mining:Concepts and Techniques"*

[8] XindongWu , Vipin Kumar , J. Ross Quinlan , Joydeep Ghosh," *Top 10 algorithms in data mining*"

[9] Mohammed J. Zaki ,Wagner Meira Jr.," *Data Mining and Analysis:Fundamental Concepts and Algorithms*"

[10] Carlos Roberto Valêncio1*, Fernando Takeshi Oyama1, Paulo Scarpelini Neto1, Angelo Cesar Colombini2, Adriano Mauro Cansian1, Rogéria Cristiane Gratão de Souza1 and Pedro Luiz Pizzigatti Corrêa3 *"MR-Radix: a multi-relational data mining algorithm"*