Data Mining and Visualization Techniques

Priti Vilasrao Bhagat[1], Rina Kisanrao Gadge[2]

Assistant Professor, Deptartment of CSE, DMIETR, Sawangi (M),Wardha, MH, India[1]

Assistant Professor, Department of CT, YCCE, Nagpur, MH, India[2]

Abstract: Visual Data Mining is a new approach for exploring very large data sets, combining traditional mining methods and information visualization techniques. The advantage of visual data exploration is that the user is directly involved in the data mining process, through analysis the results of the information visualization, user can integrate the specialist knowledge with the data mining algorithm. This paper presents current visualization methods applied in data mining.

Keywords: Visualization Techniques; Visual Data Mining; Data Mining.

I. INTRODUCTION

Data mining or knowledge discovery is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining tools predict behaviors and future trends, allowing businesses to make proactive, knowledge-driven decisions. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Data mining derives its name from the similarities between searching for valuable information in a large database and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find where the value resides.

Visualization is the process of transforming data, information and knowledge into a visual representation, and provides an interface between two information processing system of human and computer. Using effective visual interface can quickly and efficiently deal with large amounts of data to find hidden features, relations, patterns and trends, and can lead to new foresight and more efficient decision making.

Visual Data Mining presents the data in some visual form, allowing users to mine and gain insight into the data, draw conclusions and directly interact with the data. For data mining to be effective, it is important to combine the flexibility, creativity, and general knowledge of the user with the enormous storage capacity and computational performance of technology.

Visual Data Mining techniques have proven to be of high value in exploratory data analysis, and have a high potential for mining large databases. Visual Data Mining is especially useful when little is known about the data and the exploration goals are vague. Since the user is directly involved in the exploration and mining process, shifting and adjusting the exploration goals is automatically done if necessary.

- Visual Data Mining can be seen as a hypothesis generation process; the visualizations of the data allow the user to gain insight into the data and come up with new mining hypotheses. The verification of the hypotheses can also be done via data visualization, but may also be accomplished by automatic techniques from statistics, pattern recognition, or machine learning. Advantages of Visual Data Mining are: Easily interact with highly inhomogeneous and noisy data

- Users do not require an understanding of complex mathematical or statistical algorithms or parameters

- Visual data exploration allows for faster data mining and better results with a higher degree of confidence over automatic algorithms. This fact leads to a high demand for visual mining techniques and makes them indispensable in conjunction with automatic exploration techniques.

II. VISUALIZATION TECHNOLOGIES

The objective of visualization technology is to help individuals to improve intellectual capacity. PC based visualization technology is not just to manage the PC as an apparatus for mix of data, additionally as a correspondence medium between the clients and PC by utilizing PC representation and different advancements to consider more examples, variables and relations. Representation set up an input circle between subjective incitement and client perception, utilizing the learning of human comprehension, in the meantime keeping away from watched vague example, in order to stay away from oversight choice making and activity. Visualization technology of information mining is an ordinary intuitive. Visualization technology is generally separated into two classes relying upon regardless of whether including physical information: visualization in exploratory processing and data representation. The articles demonstrated in visualization in experimental processing

include diverse sorts of spatial information, for example, scalar, vector and tensor. The studies concentrate on how genuine, and rapidly showing the three-dimensional information field. Data representation is situated to multidimensional scalar information. The studies concentrate on planning and selecting suitable show approach to show vast multi-dimensional information and their common relations so that the client can get it.

Information mining innovation arranged principally on information visualization. Information visualization can be seen as directed mapping from the information data to visual structure and to individuals saw framework.

A. Traditional Visualization Methods

Conventional visualization strategies are utilized mostly as a part of low-dimensional information, including factual outline, bar graph, line graph, pie outline, dandy plot, q-q maps, diffuse chart, fractional relapse bend, form maps, timing diagram, the atomic bend, box outline, shading coding, information 3D cube. Data cube is a multi-dimensional structure shaped by sorting out information by various measurements. Clients can watch adaptability the information in the database at multi-edge, multi-dimensional through embracing different sorts' examination activity, for example, slicing, cutting, turning and boring, then look into the data and substance contained in information. Be that as it may, data cube needs instinctive in execution, particularly when the measurements past the three-dimensional, the information accumulation also, presentation are more troublesome.

B. Emerging Visualization Technology

1. Visualization Methods Based on Geometric Projection Technology [1], [2], [3]

The objective of visualization systems in light of the geometric projection innovation is establishing the fascinating projection of multi-dimensional information sets, consequently changing multidimensional information examination into investigation a little number of premium measurements information investigation, including disseminate network innovation, structure maps, overview maps, Andrews curve strategies, parallel direction perception innovation, radioactive visualization technology and detective statistics, for example, important part investigation, scaling measurement. Diffuse framework is the network of 2D disseminates focuses, which is the standard way that stretches out standard 2D scattered focuses to high-dimensional.

We can watch all conceivable two-way connection and relationship among measurements. Structure guide is additionally taking into account various paired pictures. It is settled for a set of particular variable to show, and after that draw a progression of sub chart by seeing one or more variables as condition. Sub diagram can utilize some other sorts of illustrations. Overview guide is a basic innovation of growing n-dimensional information tests in line map. Every measurement of tests is indicated on autonomous pivot. Andrews curve strategies draw every n-dimensional specimen as a curve. The favourable position is that it can express a considerable measure of measurement; the drawback is that the processing is drawn out. Parallel coordinate visualization technology is mapping k-dimensional space into two showcase measurements by k isometric pivot parallel to the presentation hub. Circular coordinate visualization technology is a basic change of parallel coordinate technology. The more extended the pivot emanated out the circle, the more prominent the estimation of the communicated information. Radioactive visualization technology is non-direct change of information.

Information dimensional is viewed as point, and dispersed at the outline in normal interim, and utilize a spring model to show point. Gradient visualization (Gradviz) is a basic augmentation of radioactive visualization. The measurements stamping of Gradviz is more troublesome, however contrast with radioactive maps, the showcase measurements are noteworthy expanded. Important segment examination speaks information to new variables. Anticipating multi-information to information can make examination information perception at the states of slightest cost data. On the other hand, this system is substantial just for information inside of two-dimensional straight sub-space of the locale crosswise over by measured variables.

2. Image-Based Visualization Technology [1], [2], [3]

Image based visualization technology maps each multidimensional information to a image, for example, line maps, symbols, and shading guide. Line maps outline measurements to the show measurements, the remaining measurements are mapped into edge and the length of segment. As far as possible the quantity of measurements that can be envisioned. Symbols are some little maps; the extent of its diverse attributes is chosen by the estimation of specific variable, for example, needle-formed symbol, star symbol. In star symbol, the diverse headings with respect to the cause compares to distinctive variables; the length of the range anticipated on these bearings relates to the estimation of variables.

3. Pixel-Oriented Visualization Methods

Pixel-oriented visualization techniques outline information worth to shaded pixels, and express the information esteem that fits in with every property in partitioned windows. The upside of this strategy is that monstrous data can be depicted one-time and won't have cover.

Not just can viably hold a little portion area of clients intrigued, additionally review the information. On the off chance that a pixel speaks to an information esteem, this system can envision greatest information of current showed (up to 1, 000, 000 information values).

Visualization methods, for example, recursive mode technology, radio division of technology, winding technology and Z-request methods fit in with such technology [4].

C. New Development of Visualization Technology

1. Distortion Techniques

Distortion techniques show a piece of information in abnormal state points of interest and other information in low and numerous subtle elements levels. The innovation gives an engaged methodology while keeping up a review of the information in the meantime, which is helpful for the procedure of intelligent investigation. There are run of the mill mutilation advancements: fish-eye view [5], compression distortion technology (hyperbolic program) [5].

2. Interactive Technologies and Collaborative Technologies

Interactive visualization technology permits representation of element changes as per location protests, and makes blend of assorted, free representation conceivable, for example, intuitive mapping, projection, and separating, scaling, intelligent connections and brushing. The client can look at much models, make full utilize the benefits of diverse visualization technology and illustrative methods for distinctive models. At the point when accentuating some piece of a model, graphic methods for distinctive related models will be shown consequently in various autonomous windows in the meantime. General utilization of such technology can get more data than the autonomous thought of these visual segments. Collective technologies permits creating portrayal of the crude information and a wide range of depiction methods for the relating model, which is helpful for similar examination of the model.

3. Drill-Through Technology

Drill-through technology implies that we can see a piece of the model are extricated from which unique information and can get to them when selecting a piece of model. Case in point, decision tree visualization techniques permit selecting and drill-through the branches of the decision tree, with the goal that clients can get to the information significant to the structure of the branches, while overlooking other information depiction [6].

4. Virtual Technologies

Virtual technologies can yield model results to a virtual gadget or virtual visualization environment, which empowers clients to remain. Clients can hunt intriguing data through route, and get a more natural comprehension of information and examination. This technology can join the intellectual capacities of individuals, and make individuals completely coordinated into the procedure of information mining. Virtual technologies which have been proposed are information 3D cube [7] and head protector show.

III. CONCLUSION

It has been demonstrated that visualization of data mining has compelling quality and potential in the exploratory data analysis and treatment of huge databases. The present routines for visualization of data mining can be generally partitioned into the accompanying gatherings. The primary gathering is made by conventional methods out of visualization or algorithm independent of data mining. The second gathering portrays the method of extraction through data mining algorithms, in this manner adding to a superior comprehension of the model. The third gathering interfaces intently with a mixed bag of visualization strategies in the running procedure of data mining algorithms.

A definitive objective is to make computer based visual data mining framework has the straightforwardness, client self-governance, unwavering quality, reusability, and security, with better, speedier and more natural limit of investigation of extensive data assets.

REFERENCES

[1]. Mehmed Kantardzic. Data Mining Concepts, Models, Methods and Algorithms [M]. IEEE Press, 2002

[2]. David Hand, Heikki Mannila, Padhraic Smyth, Principles of Data Mining [M].China Machine Press, 2003.

[3]. Shi zhongzhi. Knowledge Discovery [M]. first Edition. Beijing: Tsinghua University Press,2002.

[4]. http://ww.dbs.informatik. uni-muenchen.de /~ankers/po DM.K dd 2001.pdf.

[5]. Tubao Ho, Trongdung Nguyen, etc. Visualization Support For User- Centered Model Selection In Knowledge Discovery and DataMining. International Journal On Artificial Intelligence Tools, Vo.l 10, No. 4(2001): 691~713.

[6]. Kurt Thearling, Barry Becker, etc. Visualizing Data Mining Model Information Visualization in Data Mining and Knowledge Discovery, Morgan Kaufman Publishers, 2001.

[7]. http:‖www. cs. ualberta. ca/~ayman/published/wcgs2000. pdf.

[8]. Zhang xueying. Analysis and Comparison of Overseas Leading Data Mining Tools. Computer Engineering, 2003, 29(16): 1~3.

[9]. Qian xiaolu, Zhu jianqiu, Zhu yangyong. DM Visua Miner: A Visual Data Mining Analysis Platform. Computer Engineering, 2003, 29(supplementary): 148~150.