Visual Data Mining: A Review

Deepa Boinwad, Santosh Khamitkar , Parag Bhalchandra and Satishkumar Mekewad

SRTM University, Nanded (MS) INDIA.

ABSTRACT

Visual Data mining (VDM) is extensively used and being studied field of research area now a days. We surveyed in this paper the basic idea behind the visual data mining. How and why we have to classify the VDM techniques and corresponding classification criteria's. We also presented some of its advantages.

Introduction

As we are experiencing huge information around us being stored and monitored by today's high performing computers, we will be needing this data should be analyzed by data science people. So, then the KDD/Data Mining comes into picture to retrieve potential patterns, hidden knowledge etc. from that stored historical data. Visual data mining is another powerful tool with which we can make data science people can interactively participate in each and every step to get useful information in visualization form for humans. Humans can easily understand and grasp the information by means of visualization very easily.

The 21st century, we humans in this information age are experiencing and contributing to the explosion of data. Recent developments in computing hardware and its contribution towards growth rate of data collection and storage are increased day by day. On the other side, rate of growth of analysis techniques and its approach is very slow [3].There is almost every field is computerized, that means every digital media we use will be regulating and monitoring our online transaction history by means of database management systems and the transactions are like online shopping, telephone or cell phone usage etc. The Cambrian Explosion...of Data as you can see the visualization in bellow figure, how data growing tremendously around us [5]. By 2020 humans will have 50 thousand plus billions of GB (Giga Bytes) data to deal with.

Visual Data Mining

Data mining is a process of discovering various models, summaries, and derived values from a given collection of data [6]. If we consider this standard definition you are given a standard data, for data mining we call it as raw data, so the basic format is in tabular form and no matter whichever technology is being used to manage/store/handle such data (raw data). Ultimately the data mining people start working with pre-processing the same. After all this efforts like, selection of domain specific or application oriented data then cleaning process (dealing with missing values, aggregation, smoothing etc.)
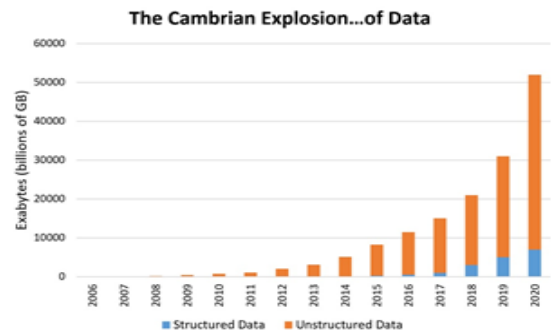

Figure1. The growth of structured versus unstructured data over the past decade shows that unstructured data accounts for more than 90% of all data (Source: http://www.eetimes.com/author.asp?section_id=36&doc_id =1330462)

After that data mining algorithms may be applied, however, every time you do a KDD/DM step you have to review it, it is iterative process. Finally the results you have got should be in the form of charts, graph, patterns represented in visual form etc. ultimately we use to the visualization of things, which can easily be understood by human brain. So the Visual data mining (VDM) is the approach to deal with such large datasets. We can achieve our desired goal by combining the traditional data mining methods with visualization techniques. And there are large number of visualization application are developed to support different data mining techniques/KDD process.

"A picture is worth a thousand words" is an English idiom [1]. Same way if we able to get insight into data to extract hidden knowledge or potentially useful patterns with the help of KDD or data mining techniques along with visualization techniques (e.g. graphs, charts etc.), it will help make human life simpler. The term Visual Data mining (VDM) is extensively used and being studied field of research area. Humans can grasp much easily with the help of visual representations of things rather than theoretical, verbal, or numerical approach.

Tele: 02462-229251
E-mail address: deepaboinwad@gmail.com

Due to its ability to merge the cognitive ability and contextual awareness of humans with the increasing computational power of data mining systems, VDM is undoubtedly not just a future trend but destined to be one of the main themes for data mining for many years to come [2].

VDM Classification

Many researchers have came up with new visualisation techniques, however some have added feature to already available techniques. Various tools and methods have been developed in last few years. With the help of such powerful methods and tools the users can quickly switch between the data automation and visualization methods. In [4], the authors have stated that application data and the methods we apply on the data should be tightly integrated with visualization methods. Therefore, when we get the results from data mining techniques one should expect informative results from visual data mining techniques rather that attractive visualization of those results. The models or patterns received should be understood by mere visual experience, human brain can easily interpret the knowledge hidden in the data. As suggested in [8], VDM and machine learning algorithms should be tightly integrated to get the maximum benefit from both the techniques.
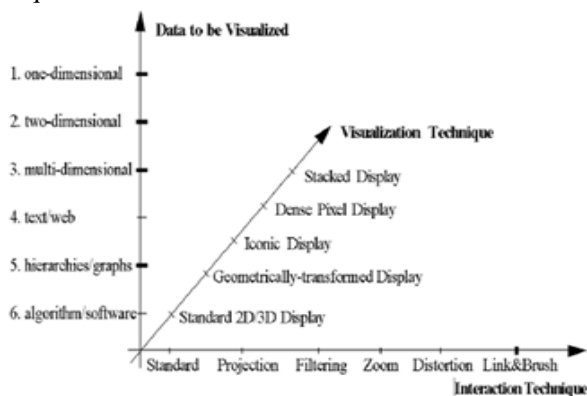

Figure 2. Classification of Information Visualization Techniques.

In [7][11] (Fig. 2) the authors has stated that the data visualization classification are done on the basis of their respective criteria's, and the criteria's are, data to be visualized, the visualization technique, and the interaction technique used.

● **Data to be visualized:**

It depends on what type of data you are going to analyze, like one-dimensional data for example stock price data. Two-dimensional data could be geographical data, multidimensional data is relational tables in data bases, and text/web data can be textual data and lastly algorithm/software data, every software development life cycle contains huge amount of data for future references so one has to analyze it for future similar projects to be undertake by analyzing with the help of VDM [12][13].

● **Visualization Techniques:**

There are different subfields of Information Visualization [11][12][13] as given in (Fig. 1) are classified as staked display, dense pixel display, iconic display, geometrically-transformed display and standard 2D/3D display.

● **Interaction and Distortion Techniques:**

Interaction and distortion techniques allows analyst efficiently and directly interact with visualization techniques. Interaction techniques like standard, projection, filtering, zoom, distortion, and link & brush to efficiently use with different visualization techniques [13].

In [13] Keim stated that, there is orthogonality is available in between these three classes of VDM. This means, for any data dimension you can use any visualization techniques in conjunction with any interaction and distortion techniques. And it is also possible that there may be some new or specific technique can also be used with any of these three types. A formal VDM process follows three-steps: Overview first, zoom and filter, and then details-on-demand. At the beginning of every data mining process, user has to overview the data with the help of VDM to decide the track of the analysis process and plan methodology required as par the database application area. It also helps in data cleaning process and takes interesting subset of data. Then analyst has to choose subset of interesting parameters from it. Further he has to drill-down the subset to observe details about data [3]. The benefits of using VDM techniques [3], [9] are:

● Large volume data can be explored very easily.

● As there are large numbers of integrated applications available to date you don't need to be expert in mathematical and statistical algorithms.

● Visualization can provide a qualitative overview of the data, allowing data phenomena to be isolated for further quantitative analysis.

● Current developments in hardware technologies will also help in save time for projecting data mining algorithm results with visualization.

So ultimately data analyst can use VDM in either of two approaches were presented: the first one is visualization technique as itself a tool for knowledge discovery or data mining, and the second one is visualization techniques applied on results of data mining process [10].

Conclusion

This is the very short survey about what the Visual data mining is? And why we need it what are the different techniques we use while applying VDM techniques on data to be analyzed. We also stated that some advantages of VDM. Basically it will be used with traditional data mining techniques to get interactive and highly qualitative results from both the techniques. There should be tight integration between VDM and Data mining (machine learning) algorithms.

References

1. Wikipedia the Free Encyclopedia (2001, January 10). Retrieved April 15, 2017 from, https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words.
2. Simoff, S.J., Böhlen, M.H., Mazeika, A. (Eds.), (2008). Visual Data Mining Theory, Techniques and Tools for Visual Analytics.
3. Keim D A, Mansmann F, Schneidewind J, Ziegler H. Challenges in visual data analysis. In Proc. the IEEE Conference on Information Visualization, Oct. 2006, pp.9-16.
4. Schulz, H., Nocke, T., Schumann, H. A Framework for Visual Data Mining of Structures. Twenty-Ninth Australasian Computer Science Conference (ACSC 2006) Australia.
5. EE Times Connecting the Global Electronics Community (2016, September 14) Retrieved April 16, 2017 from http://www.eetimes.com/author.asp?section_id=36&doc_id=1330462.
6. Kantardzic, M. (2011). Data Mining Concepts, Models, Methods and Algorithms Second Edition.

7. Keim, D.A., Sips, M., Ankerst, M. Visual Data-Mining Techniques. In C. Hansen and C. R. Johnson (Eds.),Visualization Handbook (pp. 831-843). Elsevier Academic Press.

8. Rossi, F., Visual Data Mining and Machine Learning. In proceedings (ESANN 2006) European Symposium on Artificial Networks. Bruges (Belgium), April 2006.

9. Qiang, X., Wei, Y., Hanfei, Z., Application of Visualization Technology in Spatial Data Mining. In International Conference on Computing, Control and Industrial Engineering 2010.

10. Dias, M.M., Yamaguchi, J.K., Rabelo, E., Franco, C., Visualization echniques: Which is the Most Appropriate in the Process of Knowledge Discovery in Data Base? From Advances in Data Mining Knowledge Discovery and Applications 2012 INTECH Publishing.

11. Keim, D., Ward, M., Visual Data Mining Techniques. In Intelligent Data Analysis an Introduction, Berthold, M. R., Hand, D., (Eds.) 2003, Springer-Verlag Berlin Heidelberg.

12. Kucher, K., Kerren, A., Text Visualization Techniques: Taxonomy, Visual Survey, and Community Insights. In Proceedings of the 8th Pacific Visualization Symposium (PacificVis 2015), pp. 117-121. Hangzhou, China 2015. IEEE Computer Society Press.

13. Keim, D.A., Information Visualization and Visual Data Mining. IEEE Transactions On Visualization And Computer Graphics, Vol. 7, No. 1, 2002