TECH-CHRONICLE
AN INTERNATIONAL E-JOURNAL ON
EMERGING TRENDS IN SCIENCE, TECHNOLOGY AND MANAGEMENT

Review on Information Visualization and Visual Data Mining

Miss.Ankita V. Suramwar
Department of Information Technology,
Tulsiramji Gaikwad-Patil College of Engg. & Tech.,
Nagpur, Maharashtra, India

Prof. Anup H. Gade
Department of Information Technology,
Tulsiramji Gaikwad-Patil College of Engg. & Tech.,
Nagpur, Maharashtra, India

Abstract— Data visualization is the graphical display of abstract information for two purpose namely sense-making also called data analysis and communication. Important stories live in our data and data visualization is a powerful means to discover and understand these stories, and then to present them to others. In this paper, we propose a classification of information visualization and visual data mining techniques which is based on the data type to be visualized, the visualization technique and the interaction and distortion technique. We have presented comprehensive reviews on the classification using a few examples, most of them referring to techniques and systems..

Keywords— Data visualization, Data mining, Triangle Map, Uncertainty, Motion pictures.

I. Introduction

Now a days advancement in digital imaging, storage and communication have ushered an era of exceptional growth of digital image and video data. Providing the large amount of image and video contents, for example, many website provides sharing and accessing of videos such as youtube and flicker. even the World Wide Web (WWW), image and video data mining becomes an increasingly important problem. Although a lot of research has been done in developing data mining tools that can discover knowledge from business and scientific data, it is much more difficult to understand and extract useful patterns from images and videos. Diff erent from the business and scientific data, such as transactions, graphs or texts, image and video data usually can convey much more information[1].

Data visualization is the graphical display of abstract information for two purposes: sense-making (also called data analysis) and communication. Important stories live in our data and data visualization is a powerful means to discover and understand these stories, and then to present them to others[2]. The information is abstract in that it describes things that are not physical. Statistical information is abstract. Whether it concerns sales, incidences of disease, athletic performance, or anything else, even though it doesn't pertain to the physical world, we can still display it visually, but to do this we must find a way to give form to that which has none.

Encouraged by previous successful research in mining business data, such as frequent item set mining and association rule learning, we have make clear about the discover frequent visual patterns from images and videos. Secondly need to discover whether these patterns be interesting to identify meaningfulness of the data or information [8].

Never before in history data has been generated at such high volumes as it is today. Exploring and analyzing the vast volumes of data becomes increasingly difficult. Information visualization and visual data mining can help to deal with the flood of information. The advantage of visual data exploration is that the user is directly involved in the data mining process [4][5]. There is a large number of information visualization techniques which have been developed over the last decade to support the exploration of large data sets.

Instead of using data-centric clustering, propose methodology perform visual clustering by geometrically deforming and grouping polylines while they are being plotted. Proposed clustering is achieved by analyzing the geometric relationship between polylines rather than the data itself. An optimization scheme is designed to minimize the curvature of the edges and maximize the parallelism of adjacent edges through an energy function[9].

In addition to the large growing volume of information which is explicitly accessible in the World Wide Web (WWW), there is tremendous increase in rich and comple metadata. The hyperlink connections and Structure induced by those connections provide another valuable information source, but the dynamic information resulting nfrom the use of that structure and retrieval of information within that structure is an even richer knowledge resource. But that knowledge resource does not give up its Insight easily, for it is not well-studied or well-understood, and there are few

methods or Measures that help provide insight into how the WWW is organized, how it is changing, or how it is used [14].

II. LITERATURE REVIEW

Andreea Griparis, Daniela Faur, Mihai Datcu(2016) focuses on visualization based approach to mining the EO data. Main aims of this approach is to map the existing data correlations in the multidimensional information space to the spatial correlations revealed by the 3D space which involves the number of the intrusions, extrusion as well to reveal the performance of dimensionality reduction methods.[18].

Ali Asadi, Anoop Verma,Yang, Ben (2016) proposed aeration process optimization which aim is used to minimize energy usage without sacrificing water quality. Presently available technique such as a data-driven approach has been applied, which is one large scale wastewater in Midwest. Models developed by data mining algorithms are useful in developing a clear and concise relationship among input and output variables [21].

Filippo Maria Bianchi,,Antonello Rizzi, Alireza Sadeghian(2016) proposed novel system for clusters two and knowledge discovery called LD-ABCD implementations of the data mining procedure. The knowledge discovery called LD-ABCD approach instead is based on PROCLUS, the well-know sub clustering algorithm. The dataset under analysis contains records characterized only by few features [23].

C. M. Velu , K. R. Kashwan (2013), proposed Clustering technique is quite often used by many researchers, it uses Expectation-Maximization (EM) algorithm for sampling. The study of classification of diabetic patients was main focus of this research work[2].

Laura P. R. Rivera , Sergio V. Chapa Vergara, Amilcar Meneses Viveros (2012),explains how it was implemented a method of visual data mining deals to maximize the advantages provided by the use of a distributed object over a video wall, to facilitate the assessment of large amounts of data[3].

V. Vijayakumar, R. Nedunchezhian (2012), This has objective of video data mining is to discover and describe interesting patterns from the huge amount of video data as it is one of the core problem areas of the data-mining research community[4].

Jiyang Chen , Tong Zheng , William Thorne(2007) This approach presented visual data mining system, WebViz, which allows interactive investigation of web usage data within their structure context, as well as ad-hoc knowledge pattern discovery on web navigational behavior[6].

K.-K. Tan and C.-W. Ngo(2005), it applies color histogram to characterize image segments for common pattern discovery through finding max flows between matching images. In order to consider the spatial relations among these visual primitives, graph-based model is applied (e.g., Attribute Relational Graph), where each visual primitive denotes a vertex of the graph and spatial relation among visual primitives are represented as edges of the graph [10][11].

H. Cheng, X. Yan, J. Han, and C.-W. Hsu(2007) proposed ,a systematic study is performed in using frequent item sets for pattern classification. Although frequent patterns can be efficiently discovered, they cannot be discriminative features. It is not uncommon that a frequent pattern appears in both positive and negative training samples, and thus of limited discrimination power[12].

T. Zheng, Y. Niu, and R. Goebel. (2002) , the web domain, there are many possible data layers that can be extracted from the web access log dataset. We apply the methods of our previous work to generate the usage data record and user sessions from web access logs. As a default, WebViz uses node size to represent the page visit count, node colour to indicate the average page view time, edge thickness to show the hyperlink usage count, and edge colour to represent the usage count percentage of a hyperlink [14].

Ed Chi et al. [15] proposed the Disktree representation to display the usage data and structure information. The graph has been used to visualize web site evolution, web usage trends over time, and evaluation of information .foraging. [14].

III. PROPOSE WORK

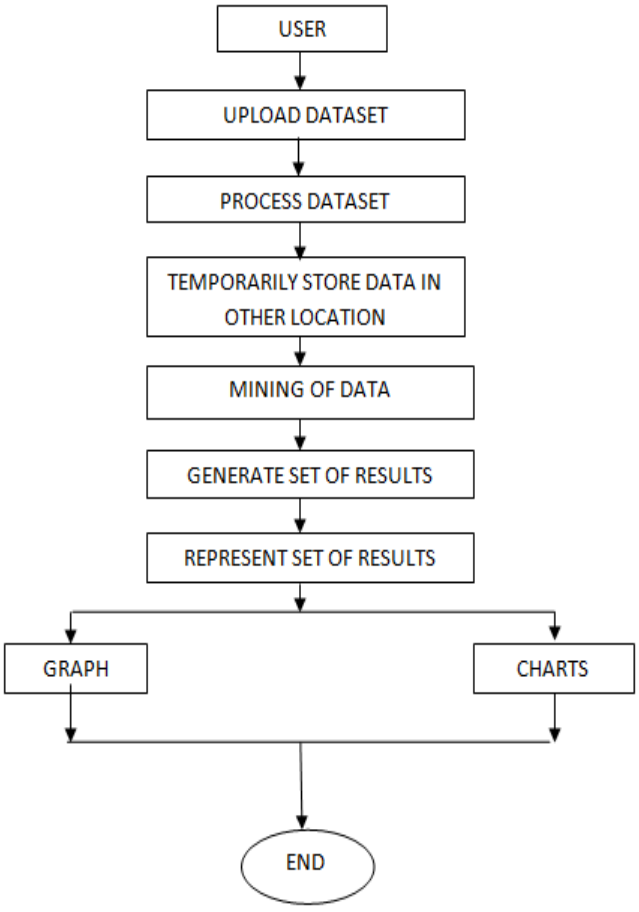

Figure: Flow of proposed work

IV. Conclusion

We have presented comprehensive review on visual data mining approaches to interpret the data we extracted from web access logs. We extend our visual data mining prototype framework to visualize and mine web usage data to understand web page access behaviours .We have described the concepts of web image, information layers and web graphs, and present the idea of mapping data and pattern to visual cues as distinct layers on top of a radial tree representation of the web structure, which allows incontext information display. We will be extracting the required data to analyze via the data parsers available in java. Data visualization will be done using Charts and graphs and it will be drawn according to the data analysis.

References

[1] Yu-Ru Lin , Nan Cao, David Gotz, "Visual Mining for Data with Uncertain Multi-labels via Triangle Map" , Data Mining (ICDM), 2014 IEEE International Conference on 29 January 2015.

[2] C. M. Velu , K. R. Kashwan, "Visual data mining techniques for classification of diabetic patients" , Advance Computing Conference (IACC), 2013 IEEE 3rd International, 13 May 2013

[3] Laura P. R. Rivera , Sergio V. Chapa Vergara, Amilcar Meneses Viveros, "Visual data mining over a video wall" , Electrical Communications and Computers (CONIELECOMP), 2012 22nd International Conference on , 26 April 2012

[4] V. Vijayakumar, R. Nedunchezhian,"A study on video data mining" International Journal of Multimedia Information Retrieval October 2012, Volume 1, Issue 3, pp 153–172 Year: August 2012.

[5] Ko Fujimura Shigeru Fujimura, Tatsushi Matsubayashi, Takeshi Yamada and Hidenori Okuda, "Topigraphy: Visualization for Large-scale Tag Clouds" in WWW 2008 / Poster Paper, pp. 1087-1088, 2008.

[6] Jiyang Chen , Tong Zheng , William Thorne,"Visual Data Mining of Web Navigational Data Purchase or Sign In ", Information Visualization, 2007. IV '07. 11th International Conference ,July 2007.

[7] Alfredo Cuzzocrea, Davood Zall ," Parallel Coordinates Technique in Visual Data Minining ", Information Visualisation (IV), 2013 17th International Conference December 2013.

[8] J. Han, H. Cheng, D. Xin, and X. Yan, "Frequent pattern mining: current status and future directions," in Data Mining and Knowledge Discovery, 2007.

[9] Hong Zhou, Xiaoru Yuan, Huamin Qu, Weiwei Cui1, Baoquan Che," Visual Clustering in Parallel Coordinates", Eurographics/ IEEE-VGTC Symposium on Visualization ,Volume(27) 2008.

[10] K.-K. Tan and C.-W. Ngo, "Common pattern discovery using earth mover's distance and local flow maximization," in Proc. IEEE Intl. Conf. on Computer Vision, 2005.

[11] P. Hong and T. S. Huang, "Spatial pattern discovery by learning a probabilistic parametric model from multiple attributed relational graphs," Discrete Applied Mathematics, pp. 113–135, 2004.

[12] H. Cheng, X. Yan, J. Han, and C.-W. Hsu, "Discriminative frequent pattern analysis for eff ective classification," in Proc. of Intl. Conf. on Data Engineering, 2007.

[13] T. Zheng, Y. Niu, and R. Goebel. Webframe: In pursuit of computationally and cognitively ef_cient web mining. In *PAKDD 2002*, pages 264.275.

[14] Ed H. Chi. Improving web usability through visualization. *IEEE Internet Computing*, 6(2):64.71, March/April 2002.

[15] Ed H. Chi, J. Pitkow, J. Mackinlay, P. Pirolli, R. Gossweiler, and S. K. Card. Visualizing the evolution of web ecologies. In *Proceeding of CHI*, 1998.

[16] C. Oosthuizen, J.Wesson, and C. Cilliers. Visual web mining of organizational web sites. In *IV '06: Proceedings* of the conference on Information Visualization,pages 395.401.

[17] [17] G. Robertson, J. Mackinlay, and S. Card. Cone trees: Animated 3d visualizations of hierarchical information.In Proc. of ACM SIGCHI conference on Human Factors in Computing Systems '91, pages 189.194.

[18] Andreea Griparis, Daniela Faur, Mihai Datcu," A dimensionality reduction approach to support visual data mining: Co-ranking-based evaluation", 2016.

[19] WebViz: http://kingman.cs.ualberta.ca/research/demos/ content/webviz/demo/src/front/WebVizClientApplet.html.

[20] A. Yousse_, D.Duke, M.Zaki, and E.Glinert. Toward visual web mining. In *Proceeding of Visual* Data Mining at IEEE Intl Conference on Data Mining (ICDM), Florida, 2003.

[21] Ali Asadi, Anoop Verma' ,Yang, Ben," Wastewater treatment aeration process optimization: A data mining approach", http://dx.doi.org/10.1016/j.jenvman.2016.07.047.

[22] K. Zhao, B. Liu, T. M. Tirpak, andW. Xiao. A visual data mining framework for convenient identi_cation of useful knowledge. In *ICDM '05*, pages 530.537.

[23] Filippo Maria Bianchi,,Antonello Rizzi, Alireza Sadeghian,"Identifying user habits through data mining on call data records", http://dx.doi.org/10.1016/j.engappai.2016.05.00.