**Big Data Visualization(CS-63016-001)**

**Group Survey**

**Reading List**

**Pavansai Pottimuthi – 811235129 Group No.30**

**Shiva Kumar Peddapuram - 811235874**

**Shirish Kumar Tallapelly - 811229216**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.No.** | **Reference** | **No. of Citations** | **Abstract** | **Link** |
| 1. | M. Banane and M. Ezziyyani, "A Survey on RDF Data Store Based on NoSQL Systems for the Semantic Web Applications" in Advanced Intelligent Systems for Sustainable Development (AI2SD'2018). AI2SD 2018. Advances in Intelligent Systems and Computing, Cham:Springer, vol. 915, 2019 | 7 | The evolution and democratization of technologies have created a veritable explosion of information, this, of course, gives rise to an urgent need to analyze and deal with the huge masses of data. In fact, the problems raised by the accumulation of data (storage, processing time, heterogeneity, capture rate/generation, etc.) become stronger because the data are massive, complex and varied. It is clear that the representation of information has the capacity to synthesize and condense data, and constitutes an efficient approach for analysis. Nevertheless, it remains ineffective and does not solve these problems. Besides that, conventional visualization techniques are rarely adapted to manage and process this mass of information. To face the complex big data challenges, various types of technologies have been developed. This paper talks about the visualization layer. This layer is located just above the Data Sources, Ingestion, Hadoop storage and Hadoop Platform Management layers for which we have already proposed a meta-modeling. It has a very important role at the level of Big Data Structure. In a continuous effort, we shall present in this paper a universal meta-modeling for the visualization layer and its relationship with the other layers of a Big Data system. | <https://ieeexplore.ieee.org/document/8880276/references#references> |
| 2. | Shelke N. An efficient low complexity compression based optimal homomorphic encryption for secure fiber optic communication. Optik. 2022;252: 168545 | 6 | In recent years, the researchers have perceived the modifications or transformations motivated by the presence of big data on the definition, complexity, and future direction of the real world optimization problems. Big Data visualization is mainly based on the efficient computer system for ingesting actual data and producing graphical representation for understanding large quantity of data in a fraction of seconds. At the same time, clustering is an effective data mining tool used to analyze big data and computational intelligence (CI) techniques can be employed to solve big data classification process. In this aspect, this study develops a novel Computational Intelligence based Clustering with Classification Model for Big Data Visualization on Map Reduce Environment, named CICC-BDVMR technique. The proposed CICC-BDVMR technique intends to perform effective BDV using the clustering and data classification processes on the Map Reduce environment. For clustering process, a grasshopper optimization algorithm (GOA) with kernelized fuzzy c-means (KFCM) technique is used to cluster the big data and the GOA is mainly utilized to determine the initial cluster centers of the KFCM technique. GOA is a recently proposed metaheuristic algorithm inspired by the swarming behaviour of grasshoppers. This algorithm has been shown to be efficient in tackling global unconstrained and constrained optimization problems. Based on the modified GOA, an effective kernel extreme learning machine model for financial stress prediction was created. Besides, big data classification process takes place using the Ridge Regression (RR) and the parameter optimization of the RR model is carried out via the Red Colobuses Monkey (RCM) algorithm. The design of GOA and RCM algorithms for parameter optimization processes for big data classification shows the novelty of the study. A wide ranging simulation analysis is carried out using benchmark big datasets and the comparative results reported the enhanced outcomes of the CICC-BDVMR technique over the recent state of art approaches. The broad comparison research illustrates the CICC-BDVMR approach’s promising performance against contemporary state-of-the-art techniques. As a result, the CICC-BDVMR technique has been demonstrated to be an effective technique for visualising and classifying large amounts of data. | <https://link.springer.com/article/10.1007/s43926-022-00022-1> |
| 3. | C.K. Leung, M.A.F. Mateo and D.A. Brajczuk, "A tree-based approach for frequent pattern mining from uncertain data", PAKDD, pp. 653-661, 2008 | 15 | In the current era of big data, a huge amount of data has been generated and collected from a wide variety of rich data sources. Embedded in these big data are useful information and valuable knowledge. An example is healthcare and epidemiological data such as data related to patients who suffered from epidemic diseases like the coronavirus disease 2019 (COVID-19). Knowledge discovered from these epidemiological data helps researchers, epidemiologists and policy makers to get a better understanding of the disease, which may inspire them to come up ways to detect, control and combat the disease. As “a picture is worth a thousand words”, having methods to visualize and visually analyze these big data makes it easily to comprehend the data and the discovered knowledge. In this paper, we present a big data visualization and visual analytics tool for visualizing and analyzing COVID-19 epidemiological data. The tool helps users to get a better understanding of information about the confirmed cases of COVID-19. Although this tool is designed for visualization and visual analytics of epidemiological data, it is applicable to visualization and visual analytics of big data from many other real-life applications and services. | [https://ieeexplore.ieee.org/document/9373130](16.%09https:/ieeexplore.ieee.org/document/9373130) |
| 4. | M. Aparicio and C. J. Costa, "Data visualization", Commun. Design Quart. Rev., vol. 3, no. 1, pp. 7-11, Jan. 2015, [online] Available: https://doi.org/10.1145/2721882.2721883 | 26 | With the explosion of Big Data, visualizing statistical data became a challenging topic that has involved many research efforts over the last years. Interpreting Big Data and efficiently showing information for good understanding are difficult tasks, especially in healthcare scenarios, where different types of data have to been managed and cross-related. Some models and techniques for health data visualization have been presented in literature. However, they do not satisfy the visualization needs of physicians and medical personnel. In this paper, we present a new graphical tool for the visualization of health data, that can be easily used for monitoring health status of patients remotely. The tool is very user friendly, and allows physician to quickly understand the current status of a person by looking at colored circles. From a technical point of view, the proposed solution adopts the geoJSON standard to classify data into different circles. | <https://ieeexplore.ieee.org/document/8370134> |
| 5. | J. Lee, H.-A. Kao and S. Yang, "Service innovation and smart analytics for industry 4.0 and big data environment", Procedia Cirp, vol. 16, pp. 3-8, 2014 | 12 | Information delivery in a visual format is always a better way of communication. Even with many data visualization techniques available, visualizing enormous amounts of data has always been a challenge. With recent advancements in technology, many new visualization techniques unfold, one of which is visualizing data through Augmented reality(AR). AR and big data have always gone together as AR requires large data sets to render information virtually in a real-time environment, and big data provides the same. In this paper, we explore some of the conventional visualization techniques and discuss the scope and possibilities for AR data visualizations. We also explore the areas implementing the technique of visualizing big data with AR. The advantages and limitations are also discussed. | <https://ieeexplore.ieee.org/document/9071328> |
| 6. | Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing, STOC ’96, pages 20–29, New York, NY, USA, 1996. ACM | 6 | Data exploration systems that provide differential privacy must manage a privacy budget that measures the amount of privacy lost across multiple queries. One effective strategy to manage the privacy budget is to compute a one-time private synopsis of the data, to which users can make an unlimited number of queries. However, existing systems using synopses are built for offline use cases, where a set of queries is known ahead of time and the system carefully optimizes a synopsis for it. The synopses that these systems build are costly to compute and may also be costly to store. We introduce Overlook, a system that enables private data exploration at interactive latencies for both data analysts and data curators. The key idea in Overlook is a virtual synopsis that can be evaluated incrementally, without extra space storage or expensive precomputation. Overlook simply executes queries using an existing engine, such as a SQL DBMS, and adds noise to their results. Because Overlook's synopses do not require costly precomputation or storage, data curators can also use Overlook to explore the impact of privacy parameters interactively. Overlook offers a rich visual query interface based on the open source Hillview system. Overlook achieves accuracy comparable to existing synopsis-based systems, while offering better performance and removing the need for extra storage. | <https://arxiv.org/abs/2006.12018> |
| 7. | B. Boehmke, B. Hazen, C.A. Boone, J.L. Robinson  A data science and open source software approach to analytics for strategic sourcing | 6 | In a [big data](https://www.sciencedirect.com/topics/computer-science/big-data) model (or database as interchangeably called), the data analysis can be eased out by extracting a smaller set of the data of interest, called subset, from the mammoth original dataset. Thus, a subset helps enhance the performance of a system by avoiding the unwanted iterations through the huge parent data in further analysis.  In this study, the data model of interest is the Parametric Big Data Model (ParaDB), which is well-known for handling multidimensional big data. Unlike the other classical data models, where subset provides additional strength to the system and the ParaDB completely lacks this potential functionality and consequently could not become equally efficient and effective comparatively. Therefore, in this research, we implement the subset capability in ParaDB to further strengthen its robustness and to ensure the relational outcome instead of streamed-out plain text.  Furthermore, to perform the preliminary investigation, the exploratory visual analysis is an important aspect in any, especially spatio-temporal, big data model. Unfortunately, the ParaDB does not offer any visualization support earlier in any format. Therefore, some comprehensive steps are taken to implement the visualization functionality in ParaDB in a way that is conducive to its structures. Additionally, the GIS-visual richness is integrated and implemented to further strengthen the visual maturity of ParaDB, where it offers the queryable-visualization. | <https://www.sciencedirect.com/science/article/pii/S2667096820300033> |
| 8. | Russom P. Big data analytics. TDWI Best Practices Report (Technical Report), TDWI, Renton, WA, 2011 | 53 | In big data analytics, advanced analytic techniques operate on big datasets aimed at complementing the role of traditional OLAP for decision making. To enable companies to take benefit of these techniques despite the lack of in-house technical skills, the H2020 TOREADOR Project adopts a model-driven architecture for streamlining analysis processes, from data preparation to their visualization. In this article, we propose a new approach named SkyViz focused on the visualization area, in particular on (1) how to specify the user’s objectives and describe the dataset to be visualized, (2) how to translate this specification into a platformindependent visualization type, and (3) how to concretely implement this visualization type on the target execution platform. To support step (1), we define a visualization context based on seven prioritizable coordinates for assessing the user’s objectives and conceptually describing the data to be visualized. To automate step (2), we propose a skyline-based technique that translates a visualization context into a set of most suitable visualization types. Finally, to automate step (3), we propose a skyline-based technique that, with reference to a specific platform, finds the best bindings between the columns of the dataset and the graphical coordinates used by the visualization type chosen by the user. SkyViz can be transparently extended to include more visualization types on one hand, more visualization coordinates on the other. The article is completed by an evaluation of SkyViz based on a case study excerpted from the pilot applications of the TOREADOR Project. | <https://journals.sagepub.com/doi/pdf/10.1177/1473871619858933> |
| 9. | Satyanarayan, A., Moritz, D., Wongsuphasawat, K., Heer, J.: Vega-lite: a grammar of interactive graphics. TVCG 23(1), 341–350 (2016) | 58 | Data visualization is crucial in today’s data-driven business world, which has been widely used for helping decision making that is closely related to major revenues of many industrial companies. However, due to the high demand of data processing w.r.t. the volume, velocity, and veracity of data, there is an emerging need for database experts to help for efficient and effective data visualization. In response to this demand, this article surveys techniques that make data visualization more efficient and effective. (1) Visualization specifications define how the users can specify their requirements for generating visualizations. (2) Efficient approaches for data visualization process the data and a given visualization specification, which then produce visualizations with the primary target to be efficient and scalable at an interactive speed. (3) Data visualization recommendation is to auto-complete an incomplete specification, or to discover more interesting visualizations based on a reference visualization. | <https://link.springer.com/article/10.1007/s00778-019-00588-3> |
| 10. | G. Andrienko and N. Andrienko, Visual Analytics of Mobility and Transportation: State of the Art and Further Research Directions, vol.18, p.2017 | 36 | In the context of data visualization and analytics, this report outlines some of the challenges and emerging applications that arise in the Big Data era. In particularly, fourteen distinguished scientists from academia and industry, and diverse related communities, i.e., Information Visualization, Human-Computer Interaction, Machine Learning, Data management & Mining, and Computer Graphics have been invited to express their opinions. | <https://hal.inria.fr/hal-02568845/> |
| 11. | Frank D. McSherry. Privacy integrated queries: An extensible platform for privacy-preserving data analysis. In Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, SIGMOD ’09, pages 19–30, New York, NY, USA, 2009. ACM | 13 | Data visualization is one of the interactive ways that lead to new innovation and discovery. It is a dynamic tool that opens new ways of research which facilitate the scientific process. With extensive use of the Internet and Web, a large amount of data is generated every day. There is a need to understand large and complex data. When the data is available in large volume, it has to be processed by using various data processing methods and need to present it with different types of techniques and methods. Data visualization is a key to the success of any enterprise as it helps enterprises to control the data in an effective manner and make the best utilization of that data to convert it into knowledge. It is a process of converting data and numbers into visual form. Data visualization techniques use different effects of computer graphics. It helps the stake holders to make an effective and fast decision making. It also provides the better understanding for pattern recognition, analysis of trends, and to extract the appropriate information from the visuals. Visualizing data may be a challenge but it is much easier to understand data in the visual form rather than in the form of text, numbers, and large tables with lots of row and columns. One can choose the data visualization technique wisely by understanding data and its composition. | <https://link.springer.com/chapter/10.1007/978-981-15-2282-6_4> |
| 12. | Srinivasan U. Arunasalam B.: ‘Leveraging big data analytics to reduce healthcare costs’, IT Prof., 2013, 15, (6), pp. 21– 28 | 43 | Big data has potential to unlock novel groundbreaking opportunities in power grid that enhances a multitude of technical, social, and economic gains. As power grid technologies evolve in conjunction with measurement and communication technologies, this results in unprecedented amount of heterogeneous big data. In particular, computational complexity, data security, and operational integration of big data into power system planning and operational frameworks are the key challenges to transform the heterogeneous large dataset into actionable outcomes. In this context, suitable big data analytics combined with visualization can lead to better situational awareness and predictive decisions. This paper presents a comprehensive state-of-the-art review of big data analytics and its applications in power grids, and also identifies challenges and opportunities from utility, industry, and research perspectives. The paper analyzes research gaps and presents insights on future research directions to integrate big data analytics into power system planning and operational frameworks. Detailed information for utilities looking to apply big data analytics and insights on how utilities can enhance revenue streams and bring disruptive innovation are discussed. General guidelines for utilities to make the right investment in the adoption of big data analytics by unveiling interdependencies among critical infrastructures and operations are also provided. | <https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/iet-stg.2018.0261> |
| 13. | * J.H. Elliott   [Living systematic review: 1. Introduction-the why, what, when, and how](https://www.sciencedirect.com/science/article/pii/S0895435617306364)  J. Clin. Epidemiol.  (2017) | 89 | We propose a new framework for research synthesis of both evidence and influence, named research weaving. It summarizes and visualizes information content, history, and networks among a collection of documents on any given topic. Research weaving achieves this feat by combining the power of two methods: systematic mapping and [bibliometrics](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/bibliometrics" \o "Learn more about bibliometrics from ScienceDirect's AI-generated Topic Pages). Systematic mapping provides a snapshot of the current state of knowledge, identifying areas needing more research attention and those ready for full synthesis. Bibliometrics enables researchers to see how pieces of evidence are connected, revealing the structure and development of a field. We explain how researchers can use some or all of these tools to gain a deeper, more nuanced understanding of the scientific literature. | <https://www.sciencedirect.com/science/article/pii/S0169534718302787> |
| 14. | F. F. Accurate, “MANUFACTURING Digital Twin Manufacturing bridges the past, present and future to drive business and manufacturing outcomes through actionable intelligence throughout the production lifecycle.” | 99 | With the wave of Industry 4.0, Digital Twin is attracting more and more attention world-wide. The term might have been coined some time ago, today the concept is increasingly being used in the field of smart manufacturing. Digital Twin provides advantages in different fields of manufacturing, such as production and design, remote diagnostics and service. Digital Twin relies on the continuously accumulated data and real-time presentation of the collected data to simultaneously update and modify with its physical counterpart. However, presenting a huge amount of collected data and information in a Digital Twin in an intuitive manner remains a challenge. Currently, augmented reality (AR) has been widely implemented in the manufacturing environment, such as product design, data management, assembly instructions, and equipment maintenance. By integrating graphics, audios and real-world objects, AR allows the users to visualise and interact with Digital Twin data at a new level. It gives the opportunity to provide intuitive and continual visualisation of the Digital Twin data. In this paper, an AR application that uses Microsoft HoloLens to visualise the Digital Twin data of a CNC milling machine in a real manufacturing environment is presented. The developed application allows the operator to monitor and control the machine tool at the same time, but also enables to interact and manage the Digital Twin data simultaneously, which provides an intuitive and consistent human machine interface to improve the efficiency during the machining process. The proposed application paves the way for further development of intelligent control process through AR devices in the future. | <https://www.sciencedirect.com/science/article/pii/S2212827119305281> |
| 15. | Chen, Y., Guan, Z., Zhang, R., Du, X., Wang, Y.: A survey on visualization approaches for exploring association relationships in graph data. J. Vis. 22(3), 625–639 (2019) | 28 | The World Health Organisation estimates that 92 per cent of the world’s population does not have access to clean air. The World Bank in 2013 estimated that only air pollution (AP) was responsible for a $225bn cost in lost productivity. The purpose of this paper is to contribute to the current scholarly debate on the value of Big Data for effective healthcare management. Its focus on cardiovascular disease (CVD) in developing countries, a major cause of disability and premature death and a subject of increasing research in recent years, makes this research particularly valuable. | <https://www.emerald.com/insight/content/doi/10.1108/MD-07-2018-0835/full/html> |
| 16. | K. ElDahshan, E. Elsayed and H. Mancy, "Enhancement Semantic Prediction Big Data Method for COVID-19: Onto-NoSQL", IAENG International Journal of Computer Science, vol. 47, no. 4, 2020 | 6 | Due to the rapid development in our world today, multiple data sources from various sources such as sensors, the internet of things, healthcare data, etc., have been increased. So, it is necessary to analyze these data to obtain valuable data and visualize it with visualization tools. Big data visualization plays an essential role in today's world for decision-makers. This paper surveys big data concepts, big data visualization concepts, and big data visualization tools and techniques. Moreover, this paper presents an up-to-date comprehensive comparison among numerous big data visualization tools based on the characteristics and criteria of each visualization tool. | <https://ieeexplore.ieee.org/abstract/document/9799819> |
| 17. | Agrawal, R., Dai, X., Andres, F.: Challenges and opportunities with big data visualization (2015) | 9 | The continual growth of big data necessitates efficient ways of analysing these large datasets. Data visualisation and visual analytics has been identified as a key tool in big data analysis because they draw on the human visual and cognitive capabilities to analyse data quickly, intuitively and interactively. However, current visualisation tools and visual analytical systems fall short of providing a seamless user experience and several improvements could be made to current commercially available visualisation tools. By conducting a systematic literature review, requirements of visualisation tools were identified and categorised into six groups: dimensionality reduction, data reduction, scalability and readability, interactivity, fast retrieval of results, and user assistance. The most common themes found in the literature were dimensionality reduction and interactive data exploration. | <https://link.springer.com/chapter/10.1007/978-3-030-44999-5_39> |
| 18. | Agrawal, Rajeev, et al. "Challenges and opportunities with big data visualization." Proceedings of the 7th International Conference on Management of computational and collective intElligence in Digital EcoSystems. 2015 | 6 | Today, almost everything is recorded digitally, from our browsing history to our health records in hospitals, we make and process billions of data every day. In this era of big data, large amounts of data are continuously obtained for different purposes. But, just processing and analyzing the data isn’t enough. If data is displayed visually, humans always search for patterns more effectively. Visualization and interpretation of the data are very critical tasks in making choices in various industries. This also guides us to new ways to find innovative ideas through visualization to solve big-data problems. In this paper, we will discuss the problems, challenges, and potential of Visualization in Big Data | <http://bright-journal.org/Journal/index.php/JADS/article/view/8/3> |
| 19. | M. Agarwal, A. Srinivasan and J. Stasko, "VisWall: Visual data exploration using direct combination on large touch displays", VIS’, pp. 26-30, 2019 | 32 | In this work we propose the combination of large interactive displays with personal head-mounted Augmented Reality (AR) for information visualization to facilitate data exploration and analysis. Even though large displays provide more display space, they are challenging with regard to perception, effective multi-user support, and managing data density and complexity. To address these issues and illustrate our proposed setup, we contribute an extensive design space comprising first, the spatial alignment of display, visualizations, and objects in AR space. Next, we discuss which parts of a visualization can be augmented. Finally, we analyze how AR can be used to display personal views in order to show additional information and to minimize the mutual disturbance of data analysts. Based on this conceptual foundation, we present a number of exemplary techniques for extending visualizations with AR and discuss their relation to our design space. We further describe how these techniques address typical visualization problems that we have identified during our literature research. To examine our concepts, we introduce a generic AR visualization framework as well as a prototype implementing several example techniques. In order to demonstrate their potential, we further present a use case walkthrough in which we analyze a movie data set. From these experiences, we conclude that the contributed techniques can be useful in exploring and understanding multivariate data. We are convinced that the extension of large displays with AR for information visualization has a great potential for data analysis and sense-making. | <https://ieeexplore.ieee.org/abstract/document/9223669> |
| 20. | C. K. Leung, "Big data analysis and mining", Encyclopedia of Information Science and Technology, vol. 4e, pp. 338-348, 2018 | 18 | As high volumes of a wide variety of valuable data of different veracities can be easily generated or collected at a high velocity nowadays, big data visualisation and visual analytics are in demand in various real-life applications. Musical data are examples of big data. Embedded in these big data are useful information and valuable knowledge. Many existing big data mining algorithms return useful information and valuable knowledge in textual or tabular forms. Knowing that "a picture is worth a thousand words", big data visualisation and visual analytics are also in demand. In this paper, we present a system for visualising and analysing big data. In particular, our system focuses on the big data science task of the discovery and exploration of frequent patterns (i.e., collections of items that frequently occurring together) from musical data. Evaluation results show the applicability of our system in big data visualisation and visual analytics for music data mining. | <https://ieeexplore.ieee.org/abstract/document/8564166> |
| 21. | T. Reznik, V. Lukas, K. Charvat, K. Charvat, Z. Krivanek, M. Kepka, et al., "Disaster Risk Reduction in Agriculture through Geospatial (Big) Data Processing", ISPRS International Journal of Geo-Information, vol. 8, no. 6, pp. 11, 2017 | 13 | There is an increasing tension in agriculture between the requirements to assure full safety on the one hand and keep costs under control on the other hand, both with respect to (inter)national strategies. Farmers need to measure and understand the impact of huge amount and variety of data which drive overall quality and yield in their fields. Among others, those are local weather data, Global Navigation System of Systems data, orthophotos and satellite imagery, data on soil specifics etc. A strong need to secure Big Data arises due to various repositories and heterogeneous sources. Data storage and visualisation requirements are in some cases competing as they are a common interest as well as a threat that helps one part of a value chain to gain a higher profit. As demonstrated in this paper, handling (Big) data is therefore a sensitive topic, where trust of producers on data security is essential. | <https://ieeexplore.ieee.org/abstract/document/8517556> |
| 22. | J. Aycock, “Computer Viruses and Malware,” in Advances in Information Security, Springer-Verlag, New York, NY, USA, 1st edition, 2006 | 110 | With the explosion of Internet of Tings (IoT) worldwide, there is an increasing threat from malicious sofware (malware) attackers that calls for efcient monitoring of vulnerable systems. Large amounts of data collected from computer networks, servers, and mobile devices need to be analysed for malware proliferation. Efective analysis methods are needed to match with the scale and complexity of such a data-intensive environment. In today’s Big Data contexts, visualisation techniques can support malware analysts going through the time-consuming process of analysing suspicious activities thoroughly. Tis paper takes a step further in contributing to the evolving realm of visualisation techniques used in the information security feld. Te aim of the paper is twofold: (1) to provide a comprehensive overview of the existing visualisation techniques for detecting suspicious behaviour of systems and (2) to design a novel visualisation using similarity matrix method for establishing malware classifcation accurately. Te prime motivation of our proposal is to identify obfuscated malware using visualisation of the extended x86 IA-32 (opcode) similarity patterns, which are hard to detect with the existing approaches. Our approach uses hybrid models wherein static and dynamic malware analysis techniques are combined efectively along with visualisation of similarity matrices in order to detect and classify zero-day malware efciently. Overall, the high accuracy of classifcation achieved with our proposed method can be visually observed since diferent malware families exhibit signifcantly dissimilar behaviour patterns. | <https://downloads.hindawi.com/journals/scn/2018/1728303.pdf> |
| 23. | Aiello G (2007) The appearance of diversity: Visual design and the public communication of EU identity. In: Bain J, Holland M (eds) European Union Identity: Perceptions from Asia and Europe. baden-Baden: Nomos Verlagsgesellschaft, 147–181 | 205 | This article highlights the role that emotions play in engagements with data and their visualisation. To date, the relationship between data and emotions has rarely been noted, in part because data studies have not attended to everyday engagements with data. We draw on an empirical study to show a wide range of emotional engagements with diverse aspects of data and their visualisation, and so demonstrate the importance of emotions as vital components of making sense of data. We nuance the argument that regimes of datafication, in which numbers, metrics and statistics dominate, are characterised by a renewed faith in objectivity and rationality, arguing that in datafied times, it is not only numbers but also the feeling of numbers that is important. We build on the sociology of (a) emotions and (b) the everyday to do this, and in so doing, we contribute to the development of a sociology of data. | <https://journals.sagepub.com/doi/pdf/10.1177/0038038516674675> |
| 24. | C. Ordonez et al., "An intelligent visual big data analytics framework for supporting interactive exploration and visualization of big OLAP cubes", pp. 421-427, 2020 | 7 | In the current era of big data, huge volumes of valuable data have been generated and collected at a rapid velocity from a wide variety of rich data sources. In recent years, the initiates of open data also led to the willingness of many government, researchers, and organizations to share their data and make them publicly accessible. An example of open big data is healthcare, disease and epidemiological data such as privacy-preserving statistics on patients who suffered from epidemic diseases like the coronavirus disease 2019 (COVID-19). Analyzing these open big data can be for social good. For instance, analyzing and mining the disease statistics helps people to get a better understanding of the disease, which may inspire them to take part in preventing, detecting, controlling and combating the disease. As “a picture is worth a thousand words”, having the pictorial representation further enhances people’s understanding of the data and the corresponding results for the analysis and mining. Hence, in this paper, we present a visual data science solution for the visualization and visual analytics of big sequential data. We illustrate the ideas through the visualization and visual analytics of sequences of real-life COVID-19 epidemiological data. Our solution enables people to visualize COVID-19 epidemiological data and their temporal trends. It also allows people to visually analyze the data and discover relationships among popular features associated with the COVID-19 cases. Evaluation of these real-life sequential COVID-19 epidemiological data demonstrates the effectiveness of our visual data science solution in enhancing user experience in the visualization and visual analytics of big sequential data. | <https://ieeexplore.ieee.org/document/9582677> |
| 25. | G. Ventura, G. Vilardo, C. Terranova and E. B. Sessa, "Tracking and evolution of complex active landslides by multi-temporal airborne LiDAR data: the Montaguto landslide (Southern Italy)", Remote Sensing of Environment, vol. 115, no. 12, pp. 3237-3248, 2011 | 9 | Big data technologies have been growing up quickly during past years. New storage and computing solutions appear while those already established in the market are improved with new features and better performance. Along with this growth also rises the number of applications and fields where the inclusion of big data technologies provides a large number of benefits, from the reduction in computational costs and economic resources to the improvement in the quality of the services provided which has a direct impact on the customers satisfaction. LiDAR (Light Detection and Ranging) data processing is one of the topics that could benefit from the adoption of these kind of technologies due to the massive datasets that are being gathered nowadays, with applications in archaeology, geography, geology or forestry, among many others. An efficient management of this volume of data becomes a key point especially in visualization, computing and analytic processes. In this paper, we analyse how web applications for the visualization of LiDAR data can benefit from the adoption of big data storage technologies, as well as the advantages and disadvantages that may determine the choice of one of them. | <https://ieeexplore.ieee.org/document/8622589/references#references> |
| 26. | Y. Zhao, H. Zhang, L. An and Q. Liub, "Improving the approaches of traffic demand forecasting in the big data era", Cities, vol. 82, pp. 19-26, December 2018 | 7 | In recent years, the number of Internet of Things and Internet of Everything (IOT/IOE) paradigms has increased significantly. The large number of devices contributed to generate a huge amount of data (Big Data) inserted in Smart City solutions, which are experiencing an explosion of complexity, also due to the increment of protocols, formats and providers. In this perspective it becomes essential to create a data indexing infrastructure that can optimize the performance of the system itself, for creating the so called data shadowing on IOT and other data on cloud. Therefore, it is fundamental to study paradigms to manage the indexing and visual analytics a great variety of data including IOT/IOE. One of the important aspects to be addressed for managing data in the smart city context are: the uniform model, the performance and scalability, response times in research, and the possibilities of performing visual analytic such as data flow analysis and drill down. All these needs imply the creation of a Smart Solution capable of managing and analysing heterogeneous kinds of data, providing a multitude of final applications based on the type of user who requires a certain service. To this end, in this paper, a unified model for IOT/IOE and data ingestion is presented. In addition, two possible architectural solutions have been implemented and compared in terms of performance, resource consumption, reliability and visual analytic tools for data flow. The solutions proposed for data indexing and shadowing have been tested in the context of Snap4City pilot Helsinki and Antwerp for smart city of EC project Select4Cities | <https://ieeexplore.ieee.org/document/9060218> |
| 27. | B. Hopkins, S. Sridharan, E. Cullen and J. Lee, "The forrester wave: Enterprise insight platform suites Q4 2016" in , Cambridge, MA:Forrester Research, Inc., Dec. 2016 | 32 | Five years after the first state-of-the-art report on Commercial Visual Analytics Systems we present a reevaluation of the Big Data Analytics field. We build on the success of the 2012 survey, which was influential even beyond the boundaries of the InfoVis and Visual Analytics (VA) community. While the field has matured significantly since the original survey, we find that innovation and research-driven development are increasingly sacrificed to satisfy a wide range of user groups. We evaluate new product versions on established evaluation criteria, such as available features, performance, and usability, to extend on and assure comparability with the previous survey. We also investigate previously unavailable products to paint a more complete picture of the commercial VA landscape. Furthermore, we introduce novel measures, like suitability for specific user groups and the ability to handle complex data types, and undertake a new case study to highlight innovative features. We explore the achievements in the commercial sector in addressing VA challenges and propose novel developments that should be on systems' roadmaps in the coming years. | <https://ieeexplore.ieee.org/document/8423105> |
| 28. | E. Kurilovas, "Advanced machine learning approaches to personalise learning: Learning analytics and decision making", Behaviour Inf. Technol., vol. 38, no. 4, pp. 410-421, Apr. 2019 | 10 | Huge amounts of educational data are being produced, and a common challenge that many educational organizations confront, is finding an effective method to harness and analyze this data for continuously delivering enhanced education. Nowadays, the educational data is evolving and has become large in volume, wide in variety and high in velocity. This produced data needs to be handled in an efficient manner to extract value and make informed decisions. For that, this paper confronts such data as a big data challenge and presents a comprehensive platform tailored to perform educational big data analytical applications. Further, present an effective environment for non-data scientists and people in the educational sector to apply their demanding educational big data applications. The implementation stages of the educational big data platform on a cloud computing platform and the organization of educational data in a data lake architecture are highlighted. Furthermore, two analytical applications are performed to test the feasibility of the presented platform in discovering knowledge that potentially promotes the educational institutions. | <https://ieeexplore.ieee.org/document/9393907> |
| 29. | A Mishra and M Mohapatro, "An IoT framework for bio-medical sensor data acquisition and machine learning for early detection", International Journal of Advanced Technology and Engineering Exploration, vol. 6, no. 54, pp. 112-125, 2019 | 6 | Data mining visualization is an important aspect of big data visualization and analysis. The impact of the nature-inspired algorithm along with the impact of computing traditions for the complete visualization of the storage and data communication needs have been studied. This paper also explores the possibilities of the hybridization of data mining in terms of association of cloud computing. It also explores the data analytical view in the exploration of these approaches in terms of data storage in big data. Based on these aspects the methodological advancement along with the problem statements has been analyzed. This will help in the exploration of computational capability along with the new insights in this domain. | <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&amp;arnumber=9142979> |
| 30. | G. Chawla, S. Bamal and R. Khatana, "Big Data Analytics for Data Visualisation: Review of Techniques", International Journal of Computer Application, vol. 182, no. 21, pp. 0975-8887 | 7 | The volume of big data has opened up great opportunities for prediction and analysis of different aspects of weather. Data Visualisation is common in day to day life. Various charts and graphs are used to illustrate the practical approach towards the classification of rainfall with the help of data visualisation methods. Since it was impossible to analyze the large datasets earlier, the data visualisation techniques has made easier to plot the graphs for the better understanding of the weather. With the help of data visualisation patterns such as the highest, lowest and average rainfall in the States/Union Territories the weather of India has been visualised. In this paper, the rainfall pattern in the States/Union Territories of India was successfully visualised. The pattern identifies drought prone region in India, decrease in the annual rainfall over the century and heavy rainfall in the coastal regions of India. | <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&amp;arnumber=9057928> |