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

**Survey**

**Reading List**

Sreenath Reddy Kurukunda – 811211580

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Reference** | **Number of Citations (From Google Scholar)** | **Abstract** | Link |
| 1 | R. Cattell, Scalable SQL and NoSQL data stores, *ACM SIGMOD Record*, vol. 39, no. 4, pp. 12–27, 2011. | 25 | Computer clusters with the shared-nothing architecture are the major computing platforms for big data processing and analysis. In cluster computing, data partitioning and sampling are two fundamental strategies to speed up the computation of big data and increase scalability. In this paper, we present a comprehensive survey of the methods and techniques of data partitioning and sampling with respect to big data processing and analysis. We start with an overview of the mainstream big data frameworks on Hadoop clusters. The basic methods of data partitioning are then discussed including three classical horizontal partitioning schemes: range, hash, and random partitioning. | <https://ieeexplore.ieee.org/document/9007871> |
| 2 | F. Schneider and C. Weillner, "Big data and artificial intelligence", Nervenarzt, vol. 89, no. 8, pp. 859-860, 2018. | 15 | Application of big data and artificial intelligence has become one influence factor of English teaching, which have broken the balance of the teaching Eco-environment for English. In this article, the artificial intelligence and big data are introduced into English teaching to propose a new teaching Eco-environment construction method to meet the needs of the social development and international communication in English. In the proposed method, the characteristics of English teaching under big data environment are analyzed in detail. Then the big data technology is used to construct a new Eco-environment of English teaching to improve the teaching and learning quality. | <https://ieeexplore.ieee.org/document/9235580> |
| 3 | A. Alim, X. Zhao, J. Cho and F. Chen, "Uncertainty-aware opinion inference under adversarial attacks", IEEE BigData, pp. 6-15, 2019. | 15 | In the current technological era, huge amounts of big data are generated and collected from a wide variety of rich data sources. These big data can be of different levels of veracity in the sense that some of them are precise while some others are imprecise and uncertain. Embedded in these big data are useful information and valuable knowledge to be discovered. An example of these big data 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-via data science techniques such as machine learning, data mining, and online analytical processing (OLAP)-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. | <https://ieeexplore.ieee.org/document/9378407> |
| 4 | Alamanda MS. Aspect-based sentiment analysis search engine for social media data. CSI Trans ICT. 2020;8(2):193–7. | 10 | There is an exponential growth in textual content generation every day in today's world. In-app messaging such as Telegram and WhatsApp, social media websites such as Instagram and Facebook, e-commerce websites like Amazon, Google searches, news publishing websites, and a variety of additional sources are the possible suppliers. Every instant, all these sources produce massive amounts of text data. The interpretation of such data can help business owners analyze the social outlook of their product, brand, or service and take necessary steps. The development of a consumer review summarization model using Natural Language Processing (NLP) techniques and Long short-term memory (LSTM) to present summarized data and help businesses obtain substantial insights into their consumers' behavior and choices is the topic of this research. A hybrid approach for analyzing sentiments is presented in this paper. | <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00680-6> |
| 5 | P. Buneman, S. Khanna and W.-C. Tan, "Why and where: A characterization of data provenance", Proc. Int. Conf. Database Theory (ICDT), pp. 316-330, 2001. | 13 | Nowadays, big data has become a hot research topic. It gives fresh impetus to the economic and social development. However, the huge value of big data also makes it the focus of attacks. Big data security incidents occur frequently in recent years. The security supervision capacities for big data do not match its important role. Data provenance which describes the origins of data and the process by which it arrived the current state, is an effective approach for data supervision. For the full use of provenance in big data supervision, a provenance model which defines the concepts used to represent the provenance types and relations is required to be built in advance, but current provenance models do not adapt to big data scenarios well. | <https://ieeexplore.ieee.org/document/9007438> |
| 6 | I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, The rise of ‘big data’ on cloud computing: Review and open research issues, *Inform. Syst.*, vol. 47, pp. 98–115, 2015 | 36 | With the recent advancements in computer technologies, the amount of data available is increasing day by day. However, excessive amounts of data create great challenges for users. Meanwhile, cloud computing services provide a powerful environment to store large volumes of data. They eliminate various requirements, such as dedicated space and maintenance of expensive computer hardware and software. Handling big data is a time-consuming task that requires large computational clusters to ensure successful data storage and processing. In this work, the definition, classification, and characteristics of big data are discussed, along with various cloud services, such as Microsoft Azure, Google Cloud, Amazon Web Services, International Business Machine cloud, Hortonworks, and MapR. A comparative analysis of various cloud-based big data frameworks is also performed. Various research challenges are defined in terms of distributed database storage, data security, heterogeneity, and data visualization. | <https://ieeexplore.ieee.org/document/9663258> |
| 7 | A. S. Alblawi and A. A. Alhamed, "Big data and learning analytics in higher education: Demystifying variety acquisition storage NLP and analytics", Proc. IEEE Conf. Big Data Analytics (ICBDA), pp. 124-129, Nov. 2017. | 14 | The proliferation of mobile devices and the rapid development of information and communication technologies (ICT) have seen increasingly large volume and variety of data being generated at an unprecedented pace. Big data have started to demonstrate significant values in higher education. This paper gives several contributions to the state-of-the-art for Big data in higher education and learning technologies research. Currently, there is no comprehensive survey or literature review for Big educational data. Most literature reviews from a few authors have focused on one of these fields: educational mining, learning analytics with discussions on one or two aspects such as Big data technologies without educational focus, social media data in education, etc. Most of these literature reviews are short and insufficient to provide more inclusive reviews for Big educational data. | <https://ieeexplore.ieee.org/document/9093868> |
| 8 | T. Devasia, V. T P and V. Hegde, "Prediction of students performance using educational data mining", Proc. Int. Conf. Data Mining Adv. Comput. (SAPIENCE), pp. 91-95, Mar. 2016. | 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. | <https://ieeexplore.ieee.org/document/9393907> |
| 9 | Y. Zhang, Y. Xu, Z. Y. Dong, Z. Xu and K. P. Wong, "Intelligent early warning of power system dynamic insecurity risk: Toward optimal accuracy-earliness tradeoff", IEEE Trans. Ind. Informat., vol. 13, pp. 2544-2554, Oct. 2017. | 8 | In order to keep the bottom line of systemic financial risks and prevent the mitigation of major risks, this work focuses on the investigation of multi-source heterogeneous data fusion algorithms and cleaning technologies to establish a suitable style for data analysis and big data computation frame. In this paper, according to the above method, we provide the basis for early analysis of economic security. Utilizing the big data analysis, an emerging information technology method, we can be able to explore new risk early-warning methods, build a risk monitoring and early-warning platform and achieve scientific economic decision-making, so that the sources of economic risk in national economic security can be traced. | <https://ieeexplore.ieee.org/document/8967014> |
| 10 | R. Punnoose and P. Ajit, "Prediction of employee turnover in organizations using machine learning algorithms", Int. J. Adv. Res. Artif. Intell., vol. 5, no. 9, pp. 5, 2016. | 10 | In the era of data science and big data analytics, people analytics help organizations and their human resources (HR) managers to reduce attrition by changing the way of attracting and retaining talent. In this context, employee attrition presents a critical problem and a big risk for organizations as it affects not only their productivity but also their planning continuity. In this context, the salient contributions of this research are as follows. Firstly, we propose a people analytics approach to predict employee attrition that shifts from a big data to a deep data context by focusing on data quality instead of its quantity. In fact, this deep data-driven approach is based on a mixed method to construct a relevant employee attrition model in order to identify key employee features influencing his/her attrition. | https://ieeexplore.ieee.org/document/9409047 |