IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

Evaluating Clustering Algorithms for the Management of Extensive Datasets

S Archana[1], Dr. Harsh Lohiya[2], Dr. Pradosh Chandra Patnaik[3]
Research Scholar, Department of Computer Science and Engineering
Research Guide, Department of Computer Science and Engineering
Sri SatyaSai University of Technology and Medical Sciences, Sehore (M.P.), India[1,2]
Research Co-Guide, Professor & Principal, Department of Computer Science and Engineering
Aurora's PG College (MCA), Hyderabad, India[3]

Abstract: *The exploratory nature of data analysis and data mining makes clustering one of the most usual tasks in many applications like biology, text analysis, signal analysis, etc that involve huge amount of datasets . Traditional Clustering methods like K-means or hierarchical clustering are beginning to reach its maximum capability to cope with this increase of dataset size. The limitation for these algorithms come either from the need of storing all the huge data in memory or because of their computational time complexity. These have been opened an area for research of algorithms that able to reduce this overhead. In one perspective the solutions can be in the stage of data pre-processing by transforming the data to a lower dimensionality manifold that represents the structure of the data or at the last stage of summarizing the dataset by obtaining a smaller subset of examples that represent an equivalent information. A Second perspective is to modify the Traditional clustering algorithms or to derive other ones that are able to cluster larger datasets. This perspective depends on many different approaches. An approaches such as sampling techniques, on-line processing, summarization, and efficient data structures have being applied to the problem of scaling clustering algorithms. This paper presents a review of different approaches and clustering algorithms that apply these techniques. The aim is to cover various methodologies applied for clustering data and how they can be scaled.*

Keywords: Clustering, Data mining, bigdata, Scalability, distributed computing

REFERENCES

[1]. B. Bozdemir, S. Canard, O. Ermis, H. Möllering, M. Önen, and T. Schneider,"Privacy-preserving density-based clustering", 2021.Viewat: Google Scholar
[2]. L. Wang, H. Wang, W. Zhou, and X. Han, "A novel adaptive density-based spatial clustering of application with noise based on bird swarm optimization algorithm", Computer Communications, vol. 6,2021.View at: Google Scholar
[3]. Karwan Qader, Mo Addaand Mouhammd Alkasassbeh (2017), "Comparative Analysis of Clustering Algorithms in Network Traffic Faults Classification", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, No.4, pp.6551-6563.
[4]. Yu Wang, Yang Xiang, Jun Zhang, Wan lei Zhou and Bailin Xie(2016), "Internet traffic clustering with side information", Journal of Computer and System Sciences, Elsevier, Vol.80, No.5, pp.1021–1036.
[5]. M. J. Reddy and B. Kavitha, "Clustering the mixed numerical and categorical dataset using similarity weight and filter method," International Journal of Database Theory and Application, vol. 5, pp. 121–134, 2012.
[6]. Y. Wei, X. Zhang, Y. Shi et al., "A review of data-driven approaches for prediction and classification of building energy consumption," Renewable and Sustainable Energy Reviews, vol.82, pp. 1027–1047,2018.
[7]. M. Kumar, P. Chhabra, and N. K. Garg, "An efficient content based image retrieval system using Bayes Net and K-NN," Multimedia Tools and Applications, vol. 77,no. 16, pp. 21557–21570, 2018.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

[8]. A. Onan, S. Korukoğlu, and H. Bulut, "A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification," Information Processing & Management, vol. 53, no. 4, pp. 814–833, 2017.

[9]. Luiz Fernando Carvalhoa, Sylvio Barbona, Leonardo de Souza Mendes and Mario Lemes Proença(2016),"Unsupervised learning clustering and self-organized agents applied to help network management", Expert Systems with Applications, Elsevier, Vol. 54, pp.29-47.

[10]. P.DahiyaandD.K.Srivastava,"Acomparativeevolutionofunsupervisedtechniques for effective network intrusion detection in hadoop," in Proceedings of the International Conference on Advances in Computing and Data Sciences, pp.279–287,Dehradun, India, April 2018.

[11]. Mohiuddin Ahmedand Abdun Naser Mahmood (2015),"Novel Approach for Network Traffic Pattern Analysis using Clustering based Collective Anomaly Detection", Annals of DataScience,Springer,Vol.2,No.1,pp.111–130.

[12]. S. Mehrotra and S. Kohli, "Comparative analysis of K-means with other clustering algorithms to improve search result," in Proceedings of the 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), pp. 309–313,Delhi,India, October 2015.

[13]. Shital Salve and Sanchika Bajpai (2014), "Online stream mining approach forclusteringnetworktraffic",IJRET:InternationalJournalofResearchinEngineeringand Technology, Vol. 3 No.2, pp.300- 304.X. Zheng and N. Liu, "Color recognition of clothes based on K-means and mean shift," in Proceedings of the Intelligent Control, Automatic Detection and High-End Equipment (ICADE), pp. 49–53, Beijing, China, July2012.

[14]. Lin Guan-zhou, XIN Yang, NIU Xin-xin and JIANG Hui-bai (2010), "Networktraffic classification based on semi-supervised clustering", The Journal of ChinaUniversities of Posts and Telecommunications, Elsevier, Vol. 17, Supplement 2,pp.84-88.

[15]. Lin Guan-zhou, XIN Yang, NIU Xin-xin and JIANG Hui-bai (2010), "Network traffic classification based on semi-supervised clustering", The Journal of China Universities of Posts and Telecommunications, Elsevier, Vol. 17, Supplement 2, pp.84-88.