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: 7.301
Volume 3, Issue 3, June 2023

Trimming the Fat: An Insightful Exploration of Feature Selection and Dimensionality Reduction

I.V. Dwaraka Srihith[1]
Alliance University, Bangalore, India[1]
L. Rajjitha[2], **K. Owdharya**[3], **A. David Donald**[4], **G. Thippana**[5]
Ashoka Women's Engineering College Dupadu, India [2,3,4,5]

Abstract: *Feature selection and dimensionality reduction are crucial techniques in the field of data analysis and machine learning. They aim to identify and retain the most informative and relevant features while discarding redundant or noisy ones. This short review delves into the concepts, methods, and benefits of feature selection and dimensionality reduction. It explores various approaches, such as filter, wrapper, and embedded methods, as well as popular dimensionality reduction techniques like Principal Component Analysis (PCA) and t-SNE. The review highlights the importance of these techniques in enhancing model performance, reducing computational complexity, and improving interpretability. By summarizing the key insights and challenges associated with feature selection and dimensionality reduction, this review aims to provide a comprehensive overview and serve as a foundation for further exploration in this field.*

Keywords: Feature selection, dimensionality reduction, Machine Learning(ML), data analysis, filter methods

REFERENCES

[1]. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157-1182.
[2]. Liu, H., & Motoda, H. (Eds.). (2007). Feature selection for knowledge discovery and data mining. Springer Science & Business Media.
[3]. Dash, M., & Liu, H. (1997). Feature selection for classification. Intelligent Data Analysis, 1(3), 131-156.
[4]. Jolliffe, I. T. (2011). Principal component analysis. Springer Science & Business Media.
[5]. Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9, 2579-2605.
[6]. Yu, L., Liu, H., & Lafferty, J. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), 856-863.
[7]. Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. Bioinformatics, 23(19), 2507-2517.
[8]. Liu, H., & Setiono, R. (1998). Feature selection and classification—A probabilistic wrapper approach. In Proceedings of the 9th International Conference on Tools with Artificial Intelligence (ICTAI-97), 380-385.
[9]. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.
[10]. Liu, F., Zhang, C., & Yin, J. (2020). A survey of dimensionality reduction techniques. arXiv preprint arXiv:2011.10254.
[11]. John, G. H., & Langley, P. (1995). Estimating continuous distributions in Bayesian classifiers. In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95), 338-345.
[12]. Dua, D., & Graff, C. (2017). UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences. Available at: http://archive.ics.uci.edu/ml.
[13]. Verleysen, M., & François, D. (Eds.). (2008). The curse of dimensionality: Data sampling, feature selection, and outlier detection. Springer Science & Business Media.

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: 7.301
Volume 3, Issue 3, June 2023

[14]. Inza, I., Larrañaga, P., Blanco, R., & Cerrolaza, A. J. (2001). Filter versus wrapper gene selection approaches in DNA microarray domains. Artificial Intelligence in Medicine, 31(2), 91-103.

[15]. Alshamlan, H., Badr, G., & Alohali, Y. (2014). A review of dimensionality reduction techniques for high-dimensional data processing. Journal of Data Mining and Bioinformatics, 8(3), 121-154.

[16]. Li, Y., & Wang, S. (2019). A survey on dimensionality reduction techniques. Journal of Computer Science and Technology, 34(2), 363-388.

[17]. Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. Artificial Intelligence, 97(1-2), 273-324.

[18]. Maldonado, S., & Weber, R. (2020). Feature selection techniques: A survey and experimental analysis. Information Sciences, 513, 243-269.

[19]. Wang, L., & Dong, Z. Y. (2018). A comprehensive survey on correlation-based feature selection for machine learning. Neurocomputing, 275, 1-12.

[20]. Srinivas, T. Aditya Sai, M. Monika, N. Aparna, Keshav Kumar, and J. Ramprabhu. "A Methodology to Predict the Lung Cancer and its Adverse Effects on Patients from an Advanced Correlation Analysis Method." In 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 964-970. IEEE, 2023.

[21]. Cai, Y., & Zhang, Y. (2021). Recent advances in embedded feature selection techniques: A survey. Knowledge-Based Systems, 225, 107117.

[22]. Cai, Y., Zhang, Y., Zhang, J., & Zhang, C. (2020). A comprehensive survey of wrapper feature selection in machine learning. Expert Systems with Applications, 152, 113363.

[23]. Fan, W., Zheng, X., & Zhang, X. (2020). A review of filter feature selection methods and their applications. Neurocomputing, 407, 1-19.

[24]. Srinivas, T., G. Aditya Sai, and R. Mahalaxmi. "A Comprehensive Survey of Techniques, Applications, and Challenges in Deep Learning: A Revolution in Machine Learning." International Journal of Mechanical Engineering 7, no. 5 (2022): 286-296.

[25]. Belanche, L. A., & Sánchez-Pérez, J. M. (2021). Wrapper feature selection in machine learning: A comprehensive survey. Artificial Intelligence Review, 54(2), 987-1040.

[26]. Guo, Y., Hastie, T., & Tibshirani, R. (2007). Regularized linear discriminant analysis and its application in microarrays. Biostatistics, 8(1), 86-100.

[27]. Zhang, Z. (2016). Dimensionality reduction: A comparative review. Statistical Analysis and Data Mining: The ASA Data Science Journal, 9(4), 283-297.

[28]. Saeed, F., & Mahmood, T. (2019). Feature selection and dimensionality reduction techniques for high-dimensional data analysis. Pattern Analysis and Applications, 22(3), 1037-1064.

[29]. Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8), 1226-1238.

[30]. Zhang, L., & Wang, J. (2021). An updated survey on feature selection methods for big data analytics. IEEE Access, 9, 105285-105303.

[31]. Srinivas, T. "Aditya Sai et MANIVANNAN, SS Prevention of hello flood attack in IoT using combination of deep learning with improved rider optimization algorithm." Computer Communications (2020).

[32]. Dash, D., Agrawal, R. K., & Naik, B. (2019). Feature selection and dimensionality reduction techniques for machine learning: A review. In Proceedings of the International Conference on Sustainable Computing and Intelligent Systems (pp. 131-140). Springer.

[33]. Li, C., Song, Y., Huang, W., & Li, X. (2021). Feature selection methods for machine learning: A review. Expert Systems with Applications, 167, 114165.

[34]. Liu, F., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. IEEE Transactions on Knowledge and Data Engineering, 17(4), 491-502.

[35]. Peng, J., Xie, J., & Hu, Y. (2018). A review of feature selection methods based on mutual information. Neural Computing and Applications, 29(1), 1-17.

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: 7.301
Volume 3, Issue 3, June 2023

[36]. Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. Data Classification: Algorithms and Applications, 37-64.

[37]. Chen, L., Zhang, H., & Lu, J. (2020). A survey on feature selection with deep learning: Toward deep feature selection. Information Fusion, 58, 149-166.

[38]. Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(5), 849-911.

[39]. Ramasubbareddy, Somula, T. A. S. Srinivas, K. Govinda, and E. Swetha. "Sales analysis on back friday using machine learning techniques." In Intelligent System Design: Proceedings of Intelligent System Design: INDIA 2019, pp. 313-319. Springer Singapore, 2021.

[40]. Liu, H., & Yu, L. (2005). Feature selection with dynamic mutual information. In Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence (pp. 295-301). IEEE.

[41]. Wang, D., Liu, H., & Wang, H. (2008). A novel feature selection method based on neighborhood rough set model. IEEE Transactions on Knowledge and Data Engineering, 20(9), 1136-1150.

[42]. Gu, Q., Chen, X., & Cao, B. (2020). Feature selection for high-dimensional data: A fast correlation-based filter solution revisited. Pattern Recognition Letters, 130, 396-403.

[43]. Zhang, J., & Zuo, M. J. (2021). A review on dimensionality reduction techniques for high-dimensional and big data. Big Data Mining and Analytics, 4(3), 153-176.

[44]. Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2007). Supervised machine learning: A review of classification techniques. Emerging Artificial Intelligence Applications in Computer Engineering, 160-240.