

An Enhanced K-Nearest Neighbor Predictive Model through Metaheuristic Optimization

Allemar Jhone P. Delima*

College of Computing Education, University of Mindanao, Philippines

Received 06 August 2019; received in revised form 02 October 2019; accepted 16 December 2019

DOI: <https://doi.org/10.46604/ijeti.2020.4646>

Abstract

The k-nearest neighbor (KNN) algorithm is vulnerable to noise, which is rooted in the dataset and has negative effects on its accuracy. Hence, various researchers employ variable minimization techniques before predicting the KNN in the quest so as to improve its predictive capability.

The genetic algorithm (GA) is the most widely used metaheuristics for such purpose; however, the GA suffers a problem that its mating scheme is bounded on its crossover operator. Thus, the use of the novel inversed bi-segmented average crossover (IBAX) is observed. In the present work, the crossover improved genetic algorithm (CIGAL) is instrumental in the enhancement of KNN's prediction accuracy. The use of the unmodified genetic algorithm has removed 13 variables, while the CIGAL then further removes 20 variables from the 30 total variables in the faculty evaluation dataset.

Consequently, the integration of the CIGAL to the KNN (CIGAL-KNN) prediction model improves the KNN prediction accuracy to 95.53%. In contrast to the model of having the unmodified genetic algorithm (GA-KNN), the use of the lone KNN algorithm and the prediction accuracy is only at 89.94% and 87.15%, respectively. To validate the accuracy of the models, the use of the 10-folds cross-validation technique reveals 93.13%, 89.27%, and 87.77% prediction accuracy of the CIGAL-KNN, GA-KNN, and KNN prediction models, respectively. As the result, the CIGAL carried out an optimized GA performance and increased the accuracy of the KNN algorithm as a prediction model.

Keywords: CIGAL-KNN, GA-KNN, IBAX operator, KNN algorithm, prediction models

References

- [1] A. J. P. Delima, "Predicting scholarship grants using data mining techniques," *International Journal of Machine Learning and Computing*, vol. 9, no. 4, pp. 513-519, August 2019.
- [2] A. J. P. Delima, "Applying data mining techniques in predicting index and non-index crimes," *International Journal of Machine Learning and Computing*, vol. 9, no. 4, pp. 533-538, August 2019.
- [3] M. J. Rezaee, M. Jozmaleki, and M. Valipour, "Integrating dynamic fuzzy C-means, data envelopment analysis and artificial neural network to online prediction performance of companies in stock exchange," *Physica A: Statistical Mechanics and its Applications*, vol. 489, pp. 78-93, January 2018.
- [4] U. O. Cagas, A. J. P. Delima, and T. L. Toledo, "PreFIC: predictability of faculty instructional performance through hybrid prediction model," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 7, pp. 22-25, May 2019.

* Corresponding author. E-mail address: allemarjpd@yahoo.com

Tel.: +639661549000

- [5] A. J. P. Delima and M. T. Q. Lumintac, "Application of time series analysis for philippines' inflation prediction," *International Journal of Recent Technology and Engineering*, vol. 8, no. 1, pp. 1761-1765, May 2019.
- [6] S. Fei, "The hybrid method of VMD-PSR-SVD and improved binary PSO-KNN for fault diagnosis of bearing," *Shock and Vibration*, vol. 2019, pp. 1-7, January 2019.
- [7] V. Vishnupriya and M. Valarmathi, "An effective data mining techniques for analyzing crime patterns," *IOSR Journal of Computing Engineering*, vol. 1, pp. 26-30, 2017.
- [8] M. Kumar, A. J. Singh, and D. Handa, "Literature survey on student's performance prediction in education using data mining techniques," *International Journal of Education and Management Engineering*, vol. 7, no. 6, pp. 40-49, November 2017.
- [9] A. Rairikar, V. Kulkarni, V. Sabale, H. Kale, and A. Lamgunde, "Heart disease prediction using data mining techniques," *International Conference on Intelligent Computing and Control (I2C2)*, June 2017, pp. 1-8.
- [10] D. García-gil, J. Luengo, S. García, and F. Herrera, "Enabling smart data: noise filtering in big data classification," *Information Sciences*, vol. 479, pp. 135-152, April 2019.
- [11] R. N. Patil and S. C. Tamane, "Upgrading the performance of KNN and naïve bayes in diabetes detection with genetic algorithm for feature selection," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 3, no. 1, pp. 1371-1381, 2018.
- [12] A. J. P. Delima, "An experimental comparison of hybrid modified genetic algorithm-based prediction models," *International Journal of Recent Technology and Engineering*, vol. 8, no. 1, pp. 1756-1760, May 2019.
- [13] M. Y. Orong, A. M. Sison, and R. P. Medina, "A hybrid prediction model integrating a modified genetic algorithm to K-means segmentation and C4.5," *TENCON 2018 - 2018 IEEE Region 10 Conference*, October 2018, pp. 1853-1858.
- [14] M. Mafarja, I. Aljarah, A. A. Heidari, A. I. Hammouri, H. Faris, A. M. Al-zoubi, et al., "Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems," *Knowledge-Based Systems*, vol. 145, pp. 25-45, April 2018.
- [15] A. J. P. Delima, A. M. Sison, and R. P. Medina, "A modified genetic algorithm with a new crossover mating scheme," *Indonesian Journal of Electrical Engineering and Informatics*, vol. 7, no. 2, pp. 165-181, June 2019.
- [16] A. J. P. Delima, A. M. Sison, and R. P. Medina, "Variable reduction-based prediction through modified genetic algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, pp. 356-363, 2019.
- [17] J. Gou, H. Ma, W. Ou, S. Zeng, Y. Rao, and H. Yang, "A generalized mean distance-based K-nearest neighbor classifier," *Expert Systems with Applications*, vol. 115, pp. 356-372, January 2019.
- [18] J. Gou, W. Qiu, Z. Yi, Y. Xu, Q. Mao, and Y. Zhan, "A local mean representation-based K-nearest neighbor classifier," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 3, pp. 29:1-29:5, April 2019.
- [19] Y. Mitani and Y. Hamamoto, "A local mean-based nonparametric classifier," *Pattern Recognition Letters*, vol. 27, no. 10, pp. 1151-1159, July 2006.
- [20] W. Li, Q. Du, F. Zhang, and W. Hu, "Collaborative-representation-based nearest neighbor classifier for hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 2, pp. 389-393, February 2015.
- [21] J. Gou, W. Qiu, Q. Mao, Y. Zhan, X. Shen, and Y. Rao, "A multi-local means based nearest neighbor classifier," *2017 IEEE 29th International Conference on Tools for Artificial Intelligence (ICTAI)*, June 2018, pp. 448-452.
- [22] J. Gou, W. Qiu, Z. Yi, X. Shen, Y. Zhan, and W. Ou, "Locality constrained representation-based K-nearest neighbor classification," *Knowledge-Based Systems*, vol. 167, pp. 38-52, March 2019.
- [23] F. Gieseke, J. Heinermann, C. Oancea, and C. Igel, "Buffer k-d trees: processing massive nearest neighbor queries on GPUs," *Proc. 31st International Conference on Machine Learning, ICML 2014*, January 2014, pp. 172-180.
- [24] Y. Chen, L. Zhou, Y. Tang, J. P. Singh, N. Bouguila, C. Wang, et al., "Fast neighbor search by using revised k-d tree," *Information Sciences*, vol. 472, pp. 145-162, January 2019.
- [25] M. Muja and D. G. Lowe, "Scalable nearest neighbor algorithms for high dimensional data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2227-2240, November 2014.
- [26] K. Li and J. Malik, "Fast K-nearest neighbour search via dynamic continuous indexing," *International Conference on Machine Learning*, June 2016, pp. 671-679.
- [27] K. Li and J. Malik, "Fast K-nearest neighbour search via prioritized DCI," *International Conference on Machine Learning*, 2017, pp. 2081-2090.
- [28] Y. Chen, L. Zhou, N. Bouguila, B. Zhong, F. Wu, Z. Lei, et al., "Semi-convex hull tree: fast nearest neighbor queries for large scale data on GPUs," *2018 IEEE International Conference on Data Mining (ICDM)*, November 2018, pp. 911-916.
- [29] K. Baskaran, R. Malathi, and P. Thirusakthimurugan, "Feature fusion for FDG-PET and MRI for automated extra skeletal bone sarcoma classification," *Materials Today: Proceedings*, vol. 5, no.1, 2018, pp. 1879-1889.

- [30] Y. Li, M. Y. A. Khan, Y. Jiang, F. Tian, W. Liao, S. Fu, et al., "CART and PSO + KNN algorithms to estimate the impact of water level change on water quality in poyang lake, China," *Arabian Journal of Geosciences*, vol. 12, no. 9, pp. 1-12, April 2019.
- [31] R. S. El-Sayed, "Linear discriminant analysis for an efficient diagnosis of heart disease via attribute filtering based on genetic algorithm," *Journal of Computers*, vol. 13, no. 11, pp. 1290-1299, July 2018.
- [32] S. Nagpal, S. Arora, S. Dey, and S. Shreya, "Feature selection using gravitational search algorithm for biomedical data," *Procedia Computer Science*, vol. 115, pp. 258-265, 2017.
- [33] C. Gunavathi and K. Premalatha, "Performance analysis of genetic algorithm with KNN and SVM for feature selection in tumor classification," *International Journal of Computer and Information Engineering*, vol. 8, no. 8, pp. 1490-1497, 2014.
- [34] E. Sugiyarti, K. A. Jasmi, B. Basiron, M. Huda, S. K, and A. Maselena, "Decision support system for scholarship grantee selection using data mining," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 15, pp. 2239-2249, 2018.
- [35] H. Rao, X. Shi, A. K. Rodrigue, J. Feng, Y. Xia, M. Elhoseny, et al., "Feature selection based on artificial bee colony and gradient boosting decision tree," *Applied Soft Computing*, vol. 74, pp. 634-642, January 2019.



Copyright© by the authors. Licensee TAETI, Taiwan. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).