A Novel Distance-Based k-Nearest Neighbor Voting Classifier

Wen-Shin Lin¹ Chien-Pang Lee^{2*}

¹ Department of Plant Industry, National Pingtung University of Science and Technology

Pingtung 912, Taiwan, ROC

wslin@mail.npust.edu.tw

² Department of Information Management, Da-Yeh University

Changhua 515, Taiwan, ROC

chien.pang@gmail.com

Received 23 July 2012; Revised 7 October 2012; Accepted 18 October 2012

Abstract. Recently, many classification methods are widely used on real life data. *K*-nearest neighbor (KNN) is one of the popular classification methods. Although KNN is a simple and popular classifier, it still has two problems: including the classification accuracy is often worse than nonlinear classifiers such as support vector machine (SVM); the size of parameter k for KNN. To enhance the classification accuracy and to avoid the sensitivity influence of parameter k, we propose a novel modified KNN method, the distance-based k-nearest neighbor voting classifier (DBKNNV). In our study, the classification accuracy and the sensitivity of parameter k of DBKNNV are compared with KNN and two modified KNN methods. The experiment shows that DBKNNV often achieves higher and more stable classification accuracy. Moreover, the influence with the size of the parameter k of DBKNNV is not sensitivity. That means the classification accuracy of KNN and two modified KNN methods are affected with the different parameter k setting. In contrast, the classification accuracy of DBKNNV is more stable with different parameter k setting. Furthermore, the experiment also shows the classification accuracies of DBKNNV and SVM are similar to each other.

Keywords: distance-based *k*-nearest neighbor voting classifier (DBKNNV), classifier, sensitivity, weighted voting

References

- E. Fix, J.L Hodges, "Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties," Technical Report 4, US Air Force School of Aviation Medicine, 1951.
- [2] J. Gou, L. Du, Y. Zhang, T. Xiong, "A New Distance-Weighted k-Nearest Neighbor Classifier," Journal of Information and Computational Science, Vol. 9, pp. 1429-1436, 2012.
- [3] J. Zavrel, "An Empirical Re-examination of Weighted Voting for K-NN," in *Proceedings of the 7th Belgian-Dutch Conference on Machine Learning*, pp. 139-148, 1997.
- [4] I. Măndoiu, A. Zelikovsky, Bioinformatics Algorithms: Techniques and Applications, John Wiley, New York, 2008.
- [5] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, D. Steinberg, "Top 10 Algorithms in Data Mining," *Knowledge Information System*, Vol. 14, pp.1-37, 2008.
- [6] S. A. Dudani, "The Distance-Weighted k-Nearest Neighbor Rule," *IEEE Transactions on System, Man, and Cybernetics*, Vol. 6, pp. 325-327, 1976.
- [7] Y. Zeng, Y. Yang, L. Zhao, "Pseudo Nearest Neighbor Rule for Pattern Classification," *Expert Systems with Applications*, Vol. 36, pp. 3587-3595, 2009.
- [8] T. M. Cover, P. E. Hart, "Nearest Neighbor Pattern Classification," *IEEE Transactions on Information Theory*, Vol. 13, pp. 21-27, 1967.

*Correspondence author

- [9] M. L. Raymer, W. F. Punch, E. D. Goodman, P. C. Sanschagrin, L. A. Kuhn, "Simultaneous Feature Extraction and Selection Using A Masking Genetic Algorithm," in *Proceedings of the 7th International Conference on Genetic Algorithms*, pp. 561-567, 1997
- [12] X. Zhou, D.P. Tuck, "MSVM-RFE: Extensions of SVM-RFE for Multiclass Gene Selection on DNA Microarray Data," *Bioinformatics*, Vol. 23, pp. 1106-1114, 2007.
- [11] H.-L. Huang, F.-L. Chang, "ESVM: Evolutionary Support Vector Machine For Automatic Feature Selection and Classification of Microarray Data," *Biosystems*, Vol. 90, pp. 516-528, 2007.
- [10] B.E. Boser, I.M. Guyon, V.N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144-152, 1992
- [13] A. Frank, A. Asuncion, UCI Machine Learning Repository, http://archive.ics.uci.edu/ml, 2012.
- [14] A. Statnikov, C.F. Aliferis, I. Tsamardinos, Gene Expression Model Selector, http://www.gems-system.org/, 2012.