Scopus

Documents

Pozi, M.S.M.[a][b], Azhar, N.A.[a], Raziff, A.R.A.[c], Ajrina, L.H.[a]

SVGPM: evolving SVM decision function by using genetic programming to solve imbalanced classification problem
(2022) *Progress in Artificial Intelligence*, 11 (1), pp. 65-77. Cited 3 times.

[a] School of Computing, Universiti Utara Malaysia, Changlun, Malaysia
[b] Institute of IR 4.0, Universiti Kebangsaan Malaysia, Selangor, Malaysia
[c] Kulliyyah of Information and Communication Technology, International Islamic University of Malaysia, Selangor, Malaysia

Abstract
In supervised learning, imbalanced class dataset is a state where the class distribution is not uniform among the classes. Most standard classifiers fail to properly identify pattern that belongs to minority class because most of those classifiers are built to minimize the error rate. As a result, a biased classification model is highly anticipated, as higher accuracy metrics can solely be represented by the majority class. In order to tackle this problem, several methods have been proposed, mainly to reduce the classifier's bias, such as performing resampling on the dataset, modification on a classifier optimization problem, or introducing a new optimization task on top of the classifier. Our proposal is based on a new optimization task on top of a classifier, combined as a part of the learning process. Specifically, a hybrid classifier based on genetic programming and support vector machines is proposed. Our classifier has shown to be competitive enough against several variations of support vector machines in solving imbalanced classification problem from the experimentation carried out. © 2021, Springer-Verlag GmbH Germany, part of Springer Nature.

Index Keywords
Classification (of information), Genetic algorithms, Genetic programming; Class distributions, Classification models, Decision functions, Hybrid classifier, Imbalanced class, Imbalanced classification, Optimization problems, Optimization task; Support vector machines

References

- He, H., Garcia, E.A.
  Learning from imbalanced data
  (2009) *IEEE Trans. Knowl. Data Eng.*, 21 (9), pp. 1263-1284.

- Zheng, B., Myint, S.W., Thenkabail, P.S., Aggarwal, R.M.
  A support vector machine to identify irrigated crop types using time-series landsat NDVI data
  (2015) *Int. J. Appl. Earth Obs. Geoinf.*, 34, pp. 103-112.

- Geiß, C., Pelizari, P.A., Marconcini, M., Sengara, W., Edwards, M., Lakes, T., Taubenböck, H.
  Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques
  (2015) *ISPRS J. Photogramm. Remote. Sens.*, 104, pp. 175-188.

- Yu, L., Zhou, R., Tang, L., Chen, R.
  A dbn-based resampling svm ensemble learning paradigm for credit classification with imbalanced data
  (2018) *Appl. Soft Comput.*, 69, pp. 192-202.

- Lameski, P., Zdravevski, E., Mingov, R., Kulakov, A.
  Svm parameter tuning with grid search and its impact on reduction of model over-fitting

(2015) *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pp. 464-474. Springer

- Mease, D., Wyner, A.J., Buja, A.
  Boosted classification trees and class probability/quantile estimation
  (2007) *J. Mach. Learn. Res.*, 8, pp. 409-439.

- Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.
  Smote: synthetic minority over-sampling technique
  (2002) *J. Artif. Intell. Res.*, 16, pp. 321-357.

- Iranmehr, A., Masnadi-Shirazi, H., Vasconcelos, N.
  Cost-sensitive support vector machines
  (2019) *Neurocomputing*, 343, pp. 50-64.

- Tanveer, M., Gautam, C., Suganthan, P.N.
  Comprehensive evaluation of twin SVM based classifiers on UCI datasets
  (2019) *Appl. Soft Comput.*, 83, pp. 105-617.

- Gonzalez-Abril, L., Nuñez, H., Angulo, C., Velasco, F.
  Gsvm: An svm for handling imbalanced accuracy between classes inbi-classification problems
  (2014) *Appl. Soft Comput.*, 17, pp. 23-31.

- Imam, T., Ting, K.M., Kamruzzaman, J.
  Z-SVM: An SVM for improved classification of imbalanced data
  (2006) *Advances in Artificial Intelligence*, pp. 264-273. Springer

- Hsu, C.W., Lin, C.J.
  A comparison of methods for multiclass support vector machines
  (2002) *IEEE Trans. Neural Networks*, 13 (2), pp. 415-425.

- Vapnik, V.N.
  The nature of statistical learning theory, Springer-Verlag

- Cristianini, N., Shawe-Taylor, J.
  (2000) *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press

- Chang, C.C., Lin, C.J.
  Libsvm: a library for support vector machines
  (2011) *ACM Trans. Intell. Syst. Technol. (TIST)*, 2 (3), p. 27.

- Fernández, A., López, V., Galar, M., Del Jesus, M.J., Herrera, F.
  Analysing the classification of imbalanced data-sets with multiple classes: binarization techniques and ad-hoc approaches
  (2013) *Knowl.-Based Syst.*, 42, pp. 97-110.

- Barua, S., Islam, M.M., Yao, X., Murase, K.
  Mwmote-majority weighted minority oversampling technique for imbalanced data set learning
  (2012) *IEEE Trans. Knowl. Data Eng.*, 26 (2), pp. 405-425.

- Mathew, J., Pang, C.K., Luo, M., Leong, W.H.
  Classification of imbalanced data by oversampling in kernel space of support vector machines
  (2017) *IEEE Trans. Neural Netw. Learn. Syst.*, 29 (9), pp. 4065-4076.

- Douzas, G., Bacao, F.
  Self-organizing map oversampling (somo) for imbalanced data set learning
  (2017) *Expert Syst. Appl.*, 82, pp. 40-52.

- Koziarski, M., Krawczyk, B., Woźniak, M.
  Radial-based oversampling for noisy imbalanced data classification
  (2019) *Neurocomputing*, 343, pp. 19-33.

- Liu, X.Y., Wu, J., Zhou, Z.H.
  Exploratory undersampling for class-imbalance learning
  (2009) *IEEE Trans. Syst. Man Cybern. Part B: Cybern.*, 39 (2), pp. 539-550.

- Mani, I., Zhang, I.
  KNN approach to unbalanced data distributions: A case study involving information extraction
  (2003) *Proceedings of Workshop on Learning from Imbalanced Datasets*,

- Galar, M., Fernández, A., Barrenechea, E., Herrera, F.
  Eusboost: enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling
  (2013) *Pattern Recogn.*, 46 (12), pp. 3460-3471.

- Kang, Q., Chen, X., Li, S., Zhou, M.
  A noise-filtered under-sampling scheme for imbalanced classification
  (2016) *IEEE Trans. Cybern.*, 47 (12), pp. 4263-4274.

- Koziarski, M.
  Radial-based undersampling for imbalanced data classification
  (2020) *Pattern Recognit*, 102, pp. 107-262.

- Barua, S., Islam, M., Yao, X., Murase, K.
  Mwmote-majority weighted minority oversampling technique for imbalanced data set learning
  (2014) *IEEE Trans. Knowl. Data Eng.*, 26 (2), pp. 405-425.

- Lu, W., Li, Z., Chu, J.
  Adaptive ensemble undersampling-boost: a novel learning framework for imbalanced data
  (2017) *J. Syst. Softw.*, 132, pp. 272-282.

- Batuwita, R., Palade, V.
  Fsvm-cil: fuzzy support vector machines for class imbalance learning
  (2010) *IEEE Trans. Fuzzy Syst.*, 18 (3), pp. 558-571.

- Khemchandani, R., Chandra, S.
  Twin support vector machines for pattern classification
  (2007) *IEEE Trans. Pattern Anal. Mach. Intell.*, 29 (5), pp. 905-910.

- Tomar, D., Agarwal, S.
  Twin support vector machine: a review from 2007 to 2014
  (2015) *Egypt. Inf. J.*, 16 (1), pp. 55-69.

- Ji, W., Liu, D., Meng, Y., Xue, Y.
  A review of genetic-based evolutionary algorithms in SVM parameters optimization
  (2020) *Evolutionary Intelligence*, pp. 1-26.

- Xuefeng, L., Fang, L.
  Choosing multiple parameters for SVM based on genetic algorithm
  (2002) *6Th International Conference on Signal Processing, 2002*, 1, pp. 117-119.

- Gupta, P., Mehlawat, M.K., Mittal, G.
  Asset portfolio optimization using support vector machines and real-coded genetic

algorithm
(2012) *J. Glob. Optim.*, 53 (2), pp. 297-315.

- Kalyani, S., Swarup, K.
  Static security assessment in power systems using multi-class SVM with parameter selection methods
  (2013) *Int. J. Comput. Theory Eng.*, 5 (3), p. 465.

- Mishra, S., Ahirwar, A.
  An analysis on feature selection method using real coded genetic algorithm (RCGA)
  (2018) *J. Softw. Eng. Tools & Technol. Trends*, 5 (1), pp. 23-30.

- Rai, P., Barman, A.G.
  Design optimization of spur gear using SA and RCGA
  (2018) *J. Braz. Soc. Mech. Sci. Eng.*, 40 (5), pp. 1-8.

- Yin, Z.Y., Jin, Y.F., Shen, S.L., Huang, H.W.
  An efficient optimization method for identifying parameters of soft structured clay by an enhanced genetic algorithm and elastic-viscoplastic model
  (2017) *Acta Geotech.*, 12 (4), pp. 849-867.

- Tao, M., Xinzhi, Z., Yinjie, L.
  A parameters optimization method for an SVM based on adaptive genetic algorithm
  (2016) *Comput. Measur. Control*, 24 (9), pp. 215-217.

- Tam, V.W., Cheng, K.Y., Lui, K.S.
  Using micro-genetic algorithms to improve localization in wireless sensor networks
  (2006) *JCM*, 1 (4), pp. 1-10.

- De Sampaio, W.B., Silva, A.C., de Paiva, A.C., Gattass, M.
  Detection of masses in mammograms with adaption to breast density using genetic algorithm, phylogenetic trees, lbp and svm
  (2015) *Expert Syst. Appl.*, 42 (22), pp. 8911-8928.

- Zhang, J., Zhou, X., Yang, J., Cao, C., Ma, J.
  Adaptive robust blind watermarking scheme improved by entropy-based svm and optimized quantum genetic algorithm
  (2019) *Mathematical Problems in Engineering*, 2019.

- Chen, P., Yuan, L., He, Y., Luo, S.
  An improved svm classifier based on double chains quantum genetic algorithm and its application in analogue circuit diagnosis
  (2016) *Neurocomputing*, 211, pp. 202-211.

- Devos, O., Downey, G., Duponchel, L.
  Simultaneous data pre-processing and svm classification model selection based on a parallel genetic algorithm applied to spectroscopic data of olive oils
  (2014) *Food Chem.*, 148, pp. 124-130.

- Li, X., Kong, W., Shi, W., Shen, Q.
  A combination of chemometrics methods and gc-ms for the classification of edible vegetable oils
  (2016) *Chemom. Intell. Lab. Syst.*, 155, pp. 145-150.

- Adankon, M.M., Cheriet, M.
  Genetic algorithm-based training for semi-supervised svm
  (2010) *Neural Comput. Appl.*, 19 (8), pp. 1197-1206.

- Ding, S., Zhu, Z., Zhang, X.
  An overview on semi-supervised support vector machine
  (2017) *Neural Comput. Appl.*, 28 (5), pp. 969-978.

- Corus, D., Oliveto, P.S.
  Standard steady state genetic algorithms can hillclimb faster than mutation-only evolutionary algorithms
  (2017) *IEEE Trans. Evol. Comput.*, 22 (5), pp. 720-732.

- Maratea, A., Petrosino, A., Manzo, M.
  Adjusted f-measure and kernel scaling for imbalanced data learning
  (2014) *Inf. Sci.*, 257, pp. 331-341.

- Ripley, B.
  Classification and regression trees
  (2005) *. R Package*, version pp, pp. 1-0.

- Bache, K., Lichman, M.
  (2013) *UCI Machine Learning Repository*,

- Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.
  Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework
  (2011) *J. Multiple-Valued Logic Soft Comput.*, 17, p. 1.

- Lemaître, G., Nogueira, F., Aridas, C.K.
  Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning
  (2017) *J. Mach. Learn. Res.*, 18 (17), pp. 1-5.

- Demšar, J.
  Statistical comparisons of classifiers over multiple data sets
  (2006) *J. Mach. Learn. Res.*, 7, pp. 1-30.

- Tahir, M.A.U.H., Aghar, S., Manzoor, A., Noor, M.A.
  Classification model for class imbalance dataset using genetic programming
  IEEE Access, 7, pp. 71013-71037.

- Lessmann, S., Stahlbock, R., Crone, S.F.
  Genetic algorithms for support vector machine model selection
  (2006) *International Joint Conference on Neural Networks, IJCNN'06*, pp. 3063-3069.

- Howley, T., Madden, M.G.
  The genetic evolution of kernels for support vector machine classifiers
  (2004) *15Th Irish Conference on Artificial Intelligence*, pp. 445-453.
  Citeseer

- Frohlich, H., Chapelle, O., Scholkopf, B.
  Feature selection for support vector machines by means of genetic algorithm
  (2003) *Proceedings of 15Th IEEE International Conference on Tools with Artificial Intelligence*, pp. 142-148.

- Shao, L., Liu, L., Li, X.
  Feature learning for image classification via multiobjective genetic programming
  (2014) *IEEE Trans. Neural Netw. Learn. Syst.*, 25 (7), pp. 1359-1371.

- Cervantes, J., Li, X., Yu, W.
  Using Genetic Algorithm to improvecassification accuracy on imbalanced data
  (2013) *IEEE Internationalconference on Systems, Man, and Cybernetics*, pp. 2659-2664.

Correspondence Address
Pozi M.S.M.; Institute of IR 4.0, Malaysia; email: syafiq.pozi@uum.edu.my

Abbreviated Source Title: Prog. Artif. Intell.
2-s2.0-85113146085
Document Type: Article
Publication Stage: Final
Source: Scopus