Optimised one-class classification performance (abstract)*

Oliver Urs Lenz[1], Daniel Peralta[2], and Chris Cornelis[1]

[1] Department of Applied Mathematics, Computer Science and Statistics, Ghent University {oliver.lenz,chris.cornelis}@ugent.be http://www.cwi.ugent.be
[2] IDLab, Department of Information Technology, Ghent University – imec
daniel.peralta@ugent.be

1 Introduction

The goal of one-class classification is to predict membership of a target class purely on the basis of positive training records. Nonetheless, for some applications a certain number of negative records may be available that can be used to optimise the hyperparameter values of a given data descriptor (or *one-class classifier*). This is true in particular when data descriptors are used as building blocks in a multi-class classification ensemble.

The performance of data descriptors with hyperparameter optimisation has previously only been evaluated using grid search [4, 15]. In our paper, we evaluate a number of different optimisation algorithms, employ a large number of 246 one-class classification problems drawn from 50 datasets, and include Average Localised Proximity (ALP), a data descriptor recently proposed by us that has good performance with default hyperparameter values [8].

2 Data descriptors

In addition to ALP, we compare the Nearest Neighbour Distance (NND) [6], Localised Nearest Neighbour Distance (LNND) [13], Local Outlier Factor (LOF) [2] and Support Vector Machine (SVM) [16, 14] data descriptors. For NND, LNND and LOF, we optimise a single hyperparameter, while for ALP and SVM we optimise two hyperparameters. The optimisation goal is to maximise the area under the receiver operator characteristic (AUROC). For NND, LNND and ALP, we can perform efficient leave-one-out validation, while for SVM and LOF, we use (nested) five-fold cross-validation. For NND, LNND, ALP and LOF, we can reuse nearest neighbour queries, greatly reducing the run time of each optimisation step.

3 Optimisation algorithms

The algorithm that performs best overall is Malherbe-Powell optimisation [5], which alternates global and local optimisation steps. The global optimisation

* This is an extended abstract of [9].

steps, based on the AdaLIPO algorithm [10], operate with the assumption that the problem function is k-Lipschitz, for some positive k, indicating essentially the maximum steepness of the problem function. Given this assumption, the algorithm explores those parts of the function space with the highest potential improvement over the current optimum, and increases k if a newly evaluated value violates the working assumption. The local optimisation steps exploit the neighbourhood of the current optimal value, using the BOBYQA algorithm [12].

The other optimisers that we compare are the global (Bayesian) algorithms of Kushner-Sittler [7] and Bergstra-Bardenet [1], the local algorithms of Nelder-Mead [11] and Hooke-Jeeves [3], as well as random search.

4 Experiments and results

For all data descriptors, Malherbe-Powell optimisation produces the highest validation AUROC, except NND, for which it is virtually tied with Bergstra-Bardenet optimisation. NND is relatively easy to optimise, while LNND and SVM are relatively hard. When we compare validation and test AUROC, we find that all data descriptors display a degree of overfitting, especially LNND and LOF, but this reduces with dataset size. Looking at test AUROC obtained with Malherbe-Powell optimisation, the data descriptors approach their final scores (after 50 evaluations) to within 0.001 points after respectively 5 (NND), 10 (LNND and LOF), 13 (ALP) and 37 (SVM) evaluations.

After a handful of optimisation steps, test AUROC of all data descriptors is significantly higher than with their default hyperparameter values. After 50 optimisation steps, ALP and SVM significantly outperform LOF, NND and LNND, and LOF and NND outperform LNND. Further analysis reveals that ALP has a slight advantage over SVM with problems that admit a good solution, while SVM performs relatively better with problems that do not.

5 Conclusion

The performance of ALP and SVM is comparable, but ALP can be optimised more efficiently and so constitutes a good default choice. Alternatively, using validation AUROC as a selection criterion between ALP or SVM gives even higher performance, while NND is a less computationally demanding option. We thus end up with a clear trade-off between three options, allowing practitioners to make an informed decision.

Acknowledgements The research reported in this paper was conducted with the financial support of the Odysseus programme of the Research Foundation – Flanders (FWO).

References

1. Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B.: Algorithms for hyper-parameter optimization. In: NIPS 2011: Proceedings of the 25th Annual Conference on Neural Information Processing Systems. Advances in Neural Information Processing Systems, vol. 24, pp. 2546–2554. NIPS (2011)
2. Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: LOF: Identifying density-based local outliers. In: SIGMOD 2000: Proceedings of the ACM international conference on Management of data. vol. 29, pp. 93–104. ACM (2000)
3. Hooke, R., Jeeves, T.A.: "direct search" solution of numerical and statistical problems. Journal of the ACM **8**(2), 212–229 (1961)
4. Janssens, J.H.M., Flesch, I., Postma, E.O.: Outlier detection with one-class classifiers from ML and KDD. In: ICMLA 2009: Proceedings of the Eighth International Conference on Machine Learning and Applications. pp. 147–153. IEEE (2009)
5. King, D.E.: A global optimization algorithm worth using. `http://blog.dlib.net/2017/12/a-global-optimization-algorithm-worth.html` (Dec 2017), last accessed 6 Jan 2021
6. Knorr, E.M., Ng, R.T.: A unified notion of outliers: Properties and computation. In: KDD-97: Proceedings of the Third International Conference on Knowledge Discovery and Data Mining. pp. 219–222. AAAI (1997)
7. Kushner, H.J.: A versatile stochastic model of a function of unknown and time varying form. Journal of Mathematical Analysis and Applications **5**(1), 150–167 (1962)
8. Lenz, O.U., Peralta, D., Cornelis, C.: Average Localised Proximity: A new data descriptor with good default one-class classification performance. Pattern Recognition **118**, 107991 (2021)
9. Lenz, O.U., Peralta, D., Cornelis, C.: Optimised one-class classification performance. Machine Learning **111**(8), 2863–2883 (2022)
10. Malherbe, C., Vayatis, N.: Global optimization of Lipschitz functions. In: ICML 2017: Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 70, pp. 2314–2323 (2017)
11. Nelder, J.A., Mead, R.: A simplex method for function minimization. The Computer Journal **7**(4), 308–313 (1965)
12. Powell, M.J.D.: The BOBYQA algorithm for bound constrained optimization without derivatives. Tech. Rep. NA2009/06, University of Cambridge, Department of Applied Mathematics and Theoretical Physics (2009), `http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf`
13. de Ridder, D., Tax, D.M.J., Duin, R.P.W.: An experimental comparison of one-class classification methods. In: ASCI'98: Proceedings of the Fourth Annual Conference of the Advanced School for Computing and Imaging. pp. 213–218. ASCI (1998)
14. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Tech. Rep. MSR-TR-99-87, Microsoft Research, Redmond, Washington (1999)
15. Swersky, L., Marques, H.O., Sander, J., Campello, R.J.G.B., Zimek, A.: On the evaluation of outlier detection and one-class classification methods. In: DSAA 2016: Proceedings of the 3rd IEEE International Conference on Data Science and Advanced Analytics. pp. 1–10. IEEE (2016)
16. Tax, D.M.J., Duin, R.P.W.: Data domain description using support vectors. In: ESANN 1999: Proceedings of the Seventh European Symposium on Artificial Neural Networks. pp. 251–256. D-Facto (1999)