

Sakai, Tomoya; Niu, Gang; Sugiyama, Masashi

Semi-supervised AUC optimization based on positive-unlabeled learning. (English)

Zbl 1458.68178

Mach. Learn. 107, No. 4, 767-794 (2018); correction *ibid.* 107, No. 4, 795 (2018).

Summary: Maximizing the area under the receiver operating characteristic curve (AUC) is a standard approach to imbalanced classification. So far, various supervised AUC optimization methods have been developed and they are also extended to semi-supervised scenarios to cope with small sample problems. However, existing semi-supervised AUC optimization methods rely on strong distributional assumptions, which are rarely satisfied in real-world problems. In this paper, we propose a novel semi-supervised AUC optimization method that does not require such restrictive assumptions. We first develop an AUC optimization method based only on positive and unlabeled data and then extend it to semi-supervised learning by combining it with a supervised AUC optimization method. We theoretically prove that, without the restrictive distributional assumptions, unlabeled data contribute to improving the generalization performance in PU and semi-supervised AUC optimization methods. Finally, we demonstrate the practical usefulness of the proposed methods through experiments.

MSC:

68T05 Learning and adaptive systems in artificial intelligence

62H30 Classification and discrimination; cluster analysis (statistical aspects)

Cited in **1** Review
Cited in **1** Document

Keywords:

AUC optimization; learning from positive and unlabeled data; semi-supervised learning

Software:

RCV1; LIBSVM; UCI-ml

Full Text: [DOI](#)

References:

- [1] Amini, M. R., Truong, T. V., & Goutte, C. (2008). A boosting algorithm for learning bipartite ranking functions with partially labeled data. In *Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval* (pp. 99-106).
- [2] Bartlett, PL; Jordan, MI; McAuliffe, JD, Convexity, classification, and risk bounds, *Journal of the American Statistical Association*, 101, 138-156, (2006) · Zbl 1118.62330 · doi:10.1198/016214505000000907
- [3] Blondel, M; Seki, K; Uehara, K, Block coordinate descent algorithms for large-scale sparse multiclass classification, *Machine Learning*, 93, 31-52, (2013) · Zbl 1293.68216 · doi:10.1007/s10994-013-5367-2
- [4] Chang, CC; Lin, CJ, LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology*, 2, 27, (2011) · doi:10.1145/1961189.1961199
- [5] Chapelle, O., Schölkopf, B., & Zien, A. (Eds.). (2006). *Semi-supervised learning*. Cambridge: MIT Press.
- [6] Cortes, C; Mohri, M, AUC optimization vs. error rate minimization, *Advances in Neural Information Processing Systems*, 16, 313-320, (2004)
- [7] Cozman, F. G., Cohen, I., & Cirelo, M. C. (2003). Semi-supervised learning of mixture models. In *Proceedings of the 20th international conference on machine learning* (pp. 99-106).
- [8] Dredze, M., Crammer, K., & Pereira, F. (2008). Confidence-weighted linear classification. In *Proceedings of the 25th international conference on machine learning* (pp. 264-271). · Zbl 1432.68382
- [9] Plessis, MC; Niu, G; Sugiyama, M, Analysis of learning from positive and unlabeled data, *Advances in Neural Information Processing Systems*, 27, 703-711, (2014)
- [10] Plessis, MC; Niu, G; Sugiyama, M, Class-prior estimation for learning from positive and unlabeled data, *Machine Learning*, 106, 463-492, (2017) · Zbl 1412.68181 · doi:10.1007/s10994-016-5604-6
- [11] du Plessis, M. C., Niu, G., & Sugiyama, M. (2015). Convex formulation for learning from positive and unlabeled data. In *Proceedings of 32nd international conference on machine learning, JMLR workshop and conference proceedings* (Vol. 37, pp. 1386-1394).

- [12] Fujino, A., Ueda, N. (2016). A semi-supervised AUC optimization method with generative models. In \textit{IEEE 16th international conference on data mining} (pp. 883-888).
- [13] Gao, W., & Zhou, Z. H. (2015). On the consistency of AUC pairwise optimization. In \textit{International joint conference on artificial intelligence} (pp. 939-945). · [Zbl 1412.68181](#)
- [14] Gao, W; Wang, L; Jin, R; Zhu, S; Zhou, ZH, One-pass AUC optimization, *Artificial Intelligence*, 236, 1-29, (2016) · [Zbl 1357.68168](#) · [doi:10.1016/j.artint.2016.03.003](#)
- [15] Hanley, JA; McNeil, BJ, The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, 143, 29-36, (1982) · [doi:10.1148/radiology.143.1.7063747](#)
- [16] Herschtal, A., & Raskutti, B. (2004). Optimising area under the ROC curve using gradient descent. In \textit{Proceedings of the 21st international conference on machine learning}.
- [17] Joachims, T. (2002). Optimizing search engines using clickthrough data. In \textit{Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining} (pp. 133-142).
- [18] Kawakubo, H; du Plessis, MC; Sugiyama, M, Computationally efficient class-prior estimation under class balance change using energy distance, *IEICE Transactions on Information and Systems*, E99-D, 176-186, (2016) · [doi:10.1587/transinf.2015EDP7212](#)
- [19] Kotlowski, W., Dembczynski, K. J., & Huellermeier, E. (2011). Bipartite ranking through minimization of univariate loss. In \textit{Proceedings of the 28th international conference on machine learning} (pp. 1113-1120). · [Zbl 1357.68168](#)
- [20] Krijthe, JH; Loog, M, Robust semi-supervised least squares classification by implicit constraints, *Pattern Recognition*, 63, 115-126, (2017) · [Zbl 1455.68172](#) · [doi:10.1016/j.patcog.2016.09.009](#)
- [21] Lang, K. (1995). Newsweder: Learning to filter netnews. In \textit{Proceedings of the 12th international machine learning conference}.
- [22] Lewis, DD; Yang, Y; Rose, TG; Li, F, RCV1: A new benchmark collection for text categorization research, *Journal of Machine Learning Research*, 5, 361-397, (2004)
- [23] Li, YF; Zhou, ZH, Towards making unlabeled data never hurt, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37, 175-188, (2015) · [doi:10.1109/TPAMI.2014.2299812](#)
- [24] Lichman, M. (2013). UCI machine learning repository. <http://archive.ics.uci.edu/ml>.
- [25] Mendelson, S, Lower bounds for the empirical minimization algorithm, *IEEE Transactions on Information Theory*, 54, 3797-3803, (2008) · [Zbl 1328.68332](#) · [doi:10.1109/TIT.2008.926323](#)
- [26] Niu, G., du Plessis, M. C., Sakai, T., Ma Y., & Sugiyama, M. (2016). Theoretical comparisons of positive-unlabeled learning against positive-negative learning. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., & Garnett, R. (Eds.) \textit{Advances in neural information processing systems} (Vol. 29, pp. 1199-1207)
- [27] Rakhlin, A., Shamir, O., & Sridharan, K. (2012). Making gradient descent optimal for strongly convex stochastic optimization. In \textit{Proceedings of the 29th international conference on machine learning} (pp. 449-456).
- [28] Rätsch, G; Onoda, T; Müller, KR, Soft margins for adaboost, *Machine Learning*, 42, 287-320, (2001) · [Zbl 0969.68128](#) · [doi:10.1023/A:1007618119488](#)
- [29] Sakai, T., du Plessis, M. C., Niu, G., & Sugiyama, M. (2017). Semi-supervised classification based on classification from positive and unlabeled data. In \textit{Proceedings of the 34th international conference on machine learning}. · [Zbl 1118.62330](#)
- [30] Sokolovska, N., Cappé, O., & Yvon, F. (2008). The asymptotics of semi-supervised learning in discriminative probabilistic models. In \textit{Proceedings of the 25th international conference on machine learning} (pp. 984-991).
- [31] Sundararajan, S., Priyanka, G., & Selvaraj, S. S. K. (2011). A pairwise ranking based approach to learning with positive and unlabeled examples. In \textit{Proceedings of the 20th ACM international conference on information and knowledge management} (pp. 663-672).
- [32] Usunier, N., Amini, M., & Patrick, G. (2006). Generalization error bounds for classifiers trained with interdependent data. In: Weiss, Y., Schölkopf, P. B., & Platt J. C. (Eds.) \textit{Advances in neural information processing systems} (Vol. 18, pp. 1369-1376). La Jolla, CA: Neural Information Processing Systems Foundation Inc.
- [33] Vapnik, V. N. (1998). \textit{Statistical learning theory}. London: Wiley. · [Zbl 0935.62007](#)
- [34] Ying, Y., Wen, L., & Lyu, S. (2016). Stochastic online AUC maximization. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., & Garnett, R. (Eds.) \textit{Advances in neural information processing systems} (Vol. 29, pp. 451-459). La Jolla, CA: Neural Information Processing Systems Foundation Inc.
- [35] Zhao, P., Jin, R., Yang, T., & Hoi, S. C. (2011). Online AUC maximization. In \textit{Proceedings of the 28th international conference on machine learning} (pp. 233-240).

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.