Learning Under Weak Supervision: Some Insights From Decision Theory

Julian Rodemann

Department of Statistics, LMU Munich
julian@stat.uni-muenchen.de

Labeled data are scarce in a good deal of applied classification settings. This has given rise to the paradigm of semi-supervised learning (SSL), where information from unlabeled data is (partly) taken into account to improve inference drawn from labeled data in a supervised learning framework. Within SSL, an intuitive and widely used approach is referred to as self-training or pseudo-labeling [16, 5, 8, 9]. The idea is to fit an initial model to labeled data and iteratively assign pseudo-labels to some of the unlabeled data according to the model's predictions. This process requires a criterion for pseudo-label selection (PLS), that is, the selection of unlabeled instances to be pseudo-labeled and added to the training data.

In my talk as part of the Young Statisticians Lecture Series, I will argue that PLS is nothing but a decision problem. This perspective clears the way for deploying several decision-theoretic approaches – first and foremost, finding Bayes optimal actions (selections of pseudo-labels) under common loss/utility functions. With the joint likelihood as utility, the Bayes optimal criterion turns out to be the posterior predictive of pseudo-samples [13]. Since the latter requires computing a possibly intractable integral, I will spotlight some analytical approximations based on Laplace's method that circumvent expensive sampling-based evaluations of the posterior predictive [13]. Empirical evidence suggests that such a Bayesian approach to PLS can mitigate the confirmation bias in self-training that results from overfitting initial models [1]. Notably, the decision-theoretic embedding of PLS also yields the framework of optimistic/pessimistic superset learning [3, 4, 15] as max-max-/min-max-actions.

In the second part of my talk, I will discuss some extensions of Bayesian PLS based on generalized Bayesian decision theory [14]. They aim at robustifying PLS w.r.t. model selection, accumulation of errors and covariate shift [12]. What is more, I present Bayesian PLS with convex sets of prior and a regret-based updating rule. The set of priors can reflect uncertainty regarding prior information, but might as well represent priors near ignorance, see e.g. [2, 7, 6, 10, 11]. The talk will conclude with several applications of (generalized) Bayesian PLS on real-world classification problems under weak supervision.

Keywords: Semi-Supervised Learning · Self-Training · Approximate Inference · Bayesian Decision Theory · Generalized Bayes · Robustness

References

1. Arazo, E., Ortego, D., Albert, P., O'Connor, N.E., McGuinness, K.: Pseudo-labeling and confirmation bias in deep semi-supervised learning. In: 2020 International Joint Conference on Neural Networks. pp. 1–8. IEEE (2020)
2. Benavoli, A., Zaffalon, M.: Prior near ignorance for inferences in the k-parameter exponential family. Statistics **49**(5), 1104–1140 (2015)
3. Hüllermeier, E.: Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization. International Journal of Approximate Reasoning **55**, 1519–1534 (2014)
4. Hüllermeier, E., Destercke, S., Couso, I.: Learning from imprecise data: adjustments of optimistic and pessimistic variants. In: International Conference on Scalable Uncertainty Management (SUM). pp. 266–279. Springer (2019)
5. Lee, D.H.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, International Conference on Machine Learning. p. 896. No. 2 (2013)
6. Mangili, F.: A prior near-ignorance Gaussian process model for nonparametric regression. In: International Symposium on Imprecise Probabilities (ISIPTA) 2015
7. Mangili, F., Benavoli, A.: New prior near-ignorance models on the simplex. International Journal of Approximate Reasoning **56**, 278–306 (2015)
8. McClosky, D., Charniak, E., Johnson, M.: Effective self-training for parsing. In: Proceedings of the Human Language Technology Conference of the NAACL, Main Conference. pp. 152–159 (2006)
9. Rizve, M.N., Duarte, K., Rawat, Y.S., Shah, M.: In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. In: International Conference on Learning Representations (ICLR) (2020)
10. Rodemann, J., Augustin, T.: Accounting for imprecision of model specification in bayesian optimization (2021), *Poster presented at International Symposium on Imprecise Probabilities (ISIPTA)*
11. Rodemann, J., Augustin, T.: Accounting for Gaussian process imprecision in Bayesian optimization. In: International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making (IUKM). pp. 92–104. Springer (2022)
12. Rodemann, J., Fischer, S., Schneider, L., Nalenz, M., Augustin, T.: Not all data are created equal: Lessons from sampling theory for adaptive machine learning (2022), *Poster presented at International Conference on Statistics and Data Science (ICSDS), Institute of Mathematical Statistics (IMS)*
13. Rodemann, J., Goschenhofer, J., Dorigatti, E., Nagler, T., Augustin, T.: Bayesian PLS! Approximate Bayes optimal pseudo-label selection (PLS). arXiv preprint (2023), under review
14. Rodemann, J., Jansen, C., Schollmeyer, G., Augustin, T.: In all likelihoods: How to reliably select pseudo-labeled data for self-training in semi-supervised learning. arXiv preprint (2023), under review
15. Rodemann, J., Kreiss, D., Hüllermeier, E., Augustin, T.: Levelwise data disambiguation by cautious superset learning. In: International Conference on Scalable Uncertainty Management (SUM). p. 263–276. Springer (2022)
16. Shi, W., Gong, Y., Ding, C., Tao, Z.M., Zheng, N.: Transductive semi-supervised deep learning using min-max features. In: European Conference on Computer Vision. pp. 299–315 (2018)