Learning using Privileged Time-Series

Rickard K.A. Karlsson, Martin Willbo, and Fredrik D. Johansson

Chalmers University of Technology

Keywords: Long-term prediction · privileged information · time-series

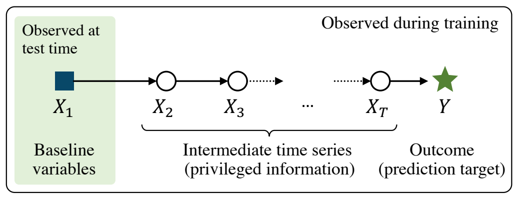

Fig. 1: Prediction with intermediate time-series privileged information. The goal at test time is to predict Y based only on X_1 but the learner has access to samples from the full series $(X_1, X_2, \ldots, X_T, Y)$ during training.

1 Introduction

Prediction of future outcomes is a central learning problem in many domains. For example, accurate prediction of the progression of chronic disease allows for identification of patients at higher risk and may be used to trigger interventions. Standard supervised learning algorithms for this task minimize the empirical risk in predicting the outcome using features collected at a baseline time point. When data is scarce, variance can plague this approach and reduce its potential impact. However, in practice, data is often collected not only at the time for prediction and the time of the outcome, but at multiple time points between them; in healthcare, disease markers or lab values of patients are recorded at regular intervals. This is data that could be used for more efficient learning but is only available during training and not at test time.

Related Work The described setting can be seen as an instance of learning using *privileged information* (LuPI) (7) or *side information* (3). In contrast to classical supervised learning, the privileged information is data that is only available during training and not at test time. We consider the case where the privileged information constitutes the intermediate part of a time series starting with baseline features and ending with the target outcome, see Figure 1. While the

Joint MSc thesis by RKAK and MW, RKAK is now affiliated with TU Delft. Supervised by FDJ. Link to full thesis here and a conference paper based on it here.

paradigm of LuPI have shown promise both theoretically (7) and empirically (2; 6), performance guarantees for practical algorithms remain elusive (5; 3). Our question is when using this privileged data leads to more sample-efficient learning of models that use only baseline data for predictions at test time. We call this setting *learning using privileged time-series* (LuPTS).

Contribution We give an algorithm for this setting and prove that when the time series are drawn from a non-stationary Gaussian-linear dynamical system of fixed horizon, learning with privileged information is more efficient than learning without it. On synthetic data, we test the limits of our algorithm and theory, both when our assumptions hold and when they are violated. On two real-world tasks – forecasting air pollution and predicting Alzheimer's disease progression – we show that our approach is generally preferable to classical learning, particularly when data is scarce.

2 Problem Setting

We learn models that use baseline variables $X_1 \in \mathbb{R}^d$ to predict outcomes $Y \in \mathbb{R}$, see Figure 1. For a given loss function, $\mathcal{L} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, our goal is to find a function $h \in \mathcal{H} \subseteq \{h : \mathbb{R}^d \to \mathbb{R}\}$ *of only the baseline variables* X_1, which minimizes the expected risk over the variables with respect to a distribution p, $R(h) \coloneqq \mathbb{E}_{X_1,Y\sim p}[\mathcal{L}(h(X_1), Y)]$. In our setting, the learner has access to privileged information in the form of time series sampled from states $X_2, ..., X_T$.

3 Theory and Experiments

Our theory states that, in the Gaussian-linear case, the estimator h_{LuPTS} that learns using privileged time-series is never worse on average than the best unbiased estimator h_{OLS} learning only from (X_1, Y), irrespective of the distribution of X_1, as we prove that $R(h_{\text{LuPTS}}) \leq R(h_{\text{OLS}})$ through usage of the Rao-Blackwell theorem (4; 1). And many times, we note that the LuPTS estimator also performs better meaning that *privileged information is provably useful in this case.*

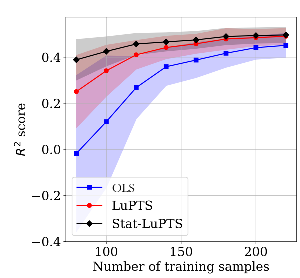

Fig. 2: Predicting cognitive test scores four years into the future.

In one experiment, we compare OLS and LuPTS on a real-world task of predicting the cognitive test score in 4 years from a Alzheimer's disease study, see Figure 2. The privileged information comprises data points at different time points after baseline that have been collected through follow-up meetings with subjects. We observe a noticeably improvement in performance and reduction in variance when using LuPTS, in particular when the sample size is small. More generally, we note a bias-variance trade-off with LuPTS as it often leads to variance reduction at the cost of being sensitive to bias.

References

[1] Blackwell, D.: Conditional expectation and unbiased sequential estimation. The Annals of Mathematical Statistics pp. 105–110 (1947)

[2] Hayashi, S., Tanimoto, A., Kashima, H.: Long-term prediction of small time-series data using generalized distillation. In: 2019 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2019)

[3] Jonschkowski, R., Höfer, S., Brock, O.: Patterns for learning with side information. arXiv preprint arXiv:1511.06429 (2015)

[4] Rao, R.C.: Information and accuracy attainable in the estimation of statistical parameters. Bulletin of the Calcutta Mathematical Society **37**(3), 81–91 (1945)

[5] Serra-Toro, C., Traver, V.J., Pla, F.: Exploring some practical issues of svm+: Is really privileged information that helps? Pattern Recognition Letters **42**, 40–46 (2014)

[6] Tang, F., Xiao, C., Wang, F., Zhou, J., Lehman, L.w.H.: Retaining privileged information for multi-task learning. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1369–1377 (2019)

[7] Vapnik, V., Vashist, A.: A new learning paradigm: Learning using privileged information. Neural networks **22**(5-6), 544–557 (2009)