Learning a logistic regression with the help of unknown features at prediction stage

1st Mario Martínez-García
Basque Center for Applied Mathematics
Bilbao, Spain
0000-0002-6849-6239

2th Iñaki Inza
University of the Basque Country UPV/EHU
San Sebastián, Spain
0000-0003-4674-1755

3th Jose A. Lozano
Basque Center for Applied Mathematics
University of the Basque Country UPV/EHU
Bilbao & San Sebastián, Spain
0000-0002-4683-8111

***Abstract*—The use of features available at training time, but not at prediction time, as additional information for training models is known as learning using privileged information paradigm. In this paper, the handling of privileged features is addressed from the logistic regression perspective, commonly used in the clinical setting. Two new proposals, LOGIT+ and LRPROB+, learned with the influence of privileged features and preserving the interpretability of conventional logistic regression, are proposed. Experimental results on datasets report improvements of our proposals over the performance of traditional logistic regression learned without privileged information.**

***Index Terms*—Privileged Information, Logistic Regression**

I. Introduction

The world is widely driven by data, nevertheless much of this information remains unused due to the lack of methods to deal with it. The Learning Using Privileged Information (LUPI) paradigm [5] proposes to harness part of this information. In addition to the use of regular features, it pursues to learn accurate classifiers with the help of privileged features that are available at training but not at prediction time. Specifically, in clinical environments, every step of the patient's clinical life is monitored, promoting the appearance of features with a privileged nature whose proper use can contribute to improve the performance of clinical models. Although LUPI paradigm has been implemented in different classification methods [1], [4], [5], it has not yet been developed in logistic regression, widely used in the clinical setting. Hence, owing to the usability of logistic regression and the presence of features with privileged nature in different domains, both concepts are merged in this paper. Two privileged logistic regression methods, LOGIT+ and LRPROB+, are developed.

II. Learning logistic regression models with privileged information

In a classical supervised classification problem a sample $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ with n instances is provided. Specifically, $\boldsymbol{x}_i \in \mathbb{R}^l$ represents the i-th vector $\boldsymbol{x}_i = (x_{i1}, ..., x_{il})$ that contains information about the l regular features in our dataset and $y_i \in \{0,1\}$ describes the corresponding output for the i-th instance. In the privileged information paradigm our set of data $\{(\boldsymbol{x}_i, \boldsymbol{x}_i^*, y_i)\}_{i=1}^n$ is formed by regular features $\boldsymbol{x}_i \in \mathbb{R}^l$ and a privileged vector $\boldsymbol{x}_i^* \in \mathbb{R}^{l^*}$ composed of the $l*$ privileged features for the i-th instance: $\boldsymbol{x}_i^* = (x_{i1}^*, ..., x_{il^*}^*)$. In order to simplify the notation, in some cases the set of data is redefined as $\{(\boldsymbol{x}_i', y_i)\}_{i=1}^n$ where $\boldsymbol{x}_i' \in \mathbb{R}^{l+l^*}$ contains privileged and regular features $\boldsymbol{x}_i' = (x_{i1}, ..., x_{il}, x_{i1}^*, ..., x_{il^*}^*)$.

A. Logistic Regression

The posterior probability of the conventional logistic regression is described as follows:

$$P(y=1|\boldsymbol{x}, \boldsymbol{\theta}) = \frac{1}{1+e^{-(\boldsymbol{\theta}^T \cdot \boldsymbol{x})}}$$

where $\boldsymbol{\theta} \in \mathbb{R}^{l+1}$ are the parameters associated to regular features plus the intercept, which is also encoded in $\boldsymbol{x}$ by the addition of a constant feature.

B. Development of LOGIT+ and LRPROB+

The proposed LUPI paradigm approach within logistic regression has the challenge of improving the performance of the classifier learned without privileged information while maintaining its interpretability. The first step of our proposal is to learn the upper bound logistic regression ($\mathrm{LR_{UB}}$) by considering all the features (regular and privileged). It is represented under the $\boldsymbol{\theta}' \in \mathbb{R}^{l+l^*+1}$ parameter space. As privileged features commonly offer relevant information to the outcome, $\mathrm{LR_{UB}}$ should outperform the traditional logistic regression classifier learned exclusively with regular features.

The key idea for the elicitation of LOGIT+ and LRPROB+ classifiers resides in being as close as possible to the output of the $\mathrm{LR_{UB}}$ classifier but, in a reduced parameter space, $\mathbb{R}^{l+1} \subset \mathbb{R}^{l+l^*+1}$. Thus, $\mathrm{LR_{UB}}$ classifier is projected onto the parameter space associated to regular features. The projection, inspired by Krithje and Loog [3], to obtain the model parameters $\boldsymbol{\theta} \in \mathbb{R}^{l+1}$ is performed by the minimization of two different convex functions for obtaining LOGIT+ and LRPROB+. Therefore, two different squared loss functions have been chosen in order to simplify the learning process of our approaches.

C. LOGIT+

LOGIT+ introduces the privileged information into the learning process of logistic regression by using the logit function.

$$\boldsymbol{\theta}^+ = \underset{\boldsymbol{\theta}}{argmin} \sum_{i=1}^{n} \left(\boldsymbol{\theta}^T \cdot \boldsymbol{x}_i - \boldsymbol{\theta}'^T \cdot \boldsymbol{x}_i' \right)^2$$

$\boldsymbol{\theta}^+$ optimal parameters of LOGIT+ are obtained by the projection of the logit from $\mathrm{LR_{UB}}$ onto $\mathbb{R}^{l+1}$ space.

D. LRPROB+

The posterior probability is the cornerstone of the LRPROB+ method for exploiting the privileged information in logistic regression. The projection of $\mathrm{LR_{UB}}$ to obtain the optimal parameters $\boldsymbol{\theta}^+ \in \mathbb{R}^{l+1}$ are estimated by:

$$\boldsymbol{\theta}^+ = \underset{\boldsymbol{\theta}}{argmin} \sum_{i=1}^{n} \left(\frac{1}{1+e^{-(\boldsymbol{\theta}^T \cdot \boldsymbol{x}_i)}} - \frac{1}{1+e^{-(\boldsymbol{\theta}'^T \cdot \boldsymbol{x}'_i)}} \right)^2$$

III. Results

One of the main objectives pursued when introducing the privileged information on logistic regression is the enhancement of our methods beyond the performance of the traditional logistic regression learned exclusively with regular features ($\mathrm{LR_{LB}}$). The proposed methods are evaluated on MNIST+, Mackey-Glass time series and three UCI [2] datasets.

A. MNIST+

MNIST+ is a modification of the popular handwritten digits database MNIST and commonly used to evaluate novel LUPI contributions [5]. Only the numbers 5 and 8 are collected and the dimensions of the images are changed from 28x28 to 10x10 to make the problem more challenging. In addition, 21 features extracted from a poetic description of the numbers are adopted as privileged features.

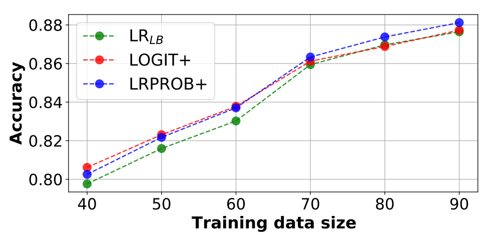

Fig. 1. Accuracy score for different percentage of instances considered for training the model with MNIST+ dataset.

Results show that both proposed models achieve a slight improvement over the standard logistic regression $\mathrm{LR_{LB}}$ for almost all training sizes. The degree of improvement obtained by both proposed methods is similar to the pioneering proposal of Vapnik and Vashist [5].

B. Mackey-Glass time series

Mackey-Glass time series is considered a benchmark dataset in the privileged information paradigm. Vapnik and Vashist [5] experiment is replicated.

TABLE I
ESTIMATED ACCURACY FOR MACKEY-GLASS IN DIFFERENT TIME STEPS (T) AHEAD PREDICTION. IN BOLD, THE BEST MODEL BUILT IN THE REGULAR PARAMETER SPACE.

Method	***T* = 3**	***T* = 5**	***T* = 8**
$\mathrm{LR_{LB}}$	0.810±0.002	0.764±0.001	0.726±0.001
LOGIT+	0.932±0.001	0.900±0.001	0.826±0.001
LRPROB+	**0.943±0.001**	**0.909±0.001**	**0.919±0.001**
$\mathrm{LR_{UB}}$	0.957±0.002	0.953±0.001	0.958±0.001

For a 2000 instances dataset and different steps ahead prediction (T), LOGIT+ and LRPROB+ outperform $\mathrm{LR_{LB}}$ in every situation.

C. UCI dataset

A set of three open UCI [2] datasets are used to evaluate the performance of LOGIT+ and LRPROB+. However, privileged features does not appear naturally. Hence, the two features with higher parameter values of the logistic regression trained with all the original features are selected as privileged. The main reason to select the most influential features is the usual property of privileged features for being relevant to the outcome.

TABLE II
ESTIMATED ACCURACY FOR UCI DATASETS WITH TWO PRIVILEGED FEATURES. IN BOLD, THE BEST MODEL BUILT IN THE REGULAR PARAMETER SPACE.

Method	**Breast Cancer**	**Obesity**	**Wine**
$\mathrm{LR_{LB}}$	0.959±0.003	0.822±0.002	0.663±0.001
LOGIT+	**0.966±0.002**	0.811±0.002	0.700±0.001
LRPROB+	0.965±0.002	**0.837±0.002**	**0.701±0.001**
$\mathrm{LR_{UB}}$	0.966±0.002	0.945±0.002	0.743±0.001

Except for LOGIT+ in the Obesity domain, the enhancement of proposed privileged methods over the base classifier ($\mathrm{LR_{LB}}$) is accomplished for every study.

IV. Conclusion

Although both algorithms achieve competitive performances, LRPROB+ outperforms LOGIT+ in almost every case of study. LRPROB+ incorporates properly the positive influence of privileged information features and outperforms the traditional logistic regression. The proposed joint treatment of LUPI paradigm within Logistic Regression, underpinned by the results, makes the contribution useful and broadly applicable in the clinical setting and other fields.

Acknowledgments

This research is supported by the Basque Government through the BERC 2022-2025 program, IT1504-22 and BMTF project, and by the Ministry of Science and Innovation: BCAM Severo Ochoa accreditation CEX2021-001142-S / MICIN / AEI / 10.13039/501100011033 and PID2019-104966GB-I00.

References

[1] Chevalier, M., Thome, N., Hénaff, G., Cord, M.: Classifying low-resolution images by integrating privileged information in deep CNNs. Pattern Recognition Letters 116, 29–35 (2018)

[2] Dua, D., Graff, C.: UCI machine learning repository (2017), http://archive.ics.uci.edu/ml

[3] Krijthe, J.H., Loog, M.: Robust semi-supervised least squares classification by implicit constraints. Pattern Recognition 63, 115–126 (2017)

[4] Lu, J., Ding, J.: A novel stochastic configuration network with iterative learning using privileged information and its application. Information Sciences 613, 953–965 (2022)

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