Lemma and Proof for the Paper under Review - iEDeaL: A Deep Learning Framework for Detecting Highly Imbalanced Interictal Epileptiform Discharges [SDS Track]

Qitong Wang
LIPADE, Université Paris Cité
qitong.wang@etu.u-paris.fr

Stephen Whitmarsh
Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, APHP, Pitié-Salpêtrière Hospital
stephen.whitmarsh@gmail.com

Vincent Navarro
Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, APHP, Pitié-Salpêtrière Hospital, Epilepsy Unit and Reference Center for Rare Epilepsies, F-75013
vincent.navarro@aphp.fr

Themis Palpanas
LIPADE, Université Paris Cité &
French University Institute (IUF)
themis@mi.parisdescartes.fr

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at https://github.com/qtwang/sEEG.

Lemma 1. *Given an i.i.d. (independent and identically distributed) subset of negative instances $X_{0,s}$ from X_0, where $X_{0,s}$ and X_0 share the same underlying distribution $P(x|y=0,\Theta)$, L_{SaSu} on the negative-sampled training set $\{X_{0,s}, X_1\}$ approximately shares the same condition of first-order stationary points with F_β-score on the whole training set.*

Proof. Under the assumption that $X_{0,s}$ and X_0 share the same underlying distribution $P(x|y=0,\Theta)$, limits of the sample proportions on $\{X_{0,s}, X_1\}$ are changed to Equation 1 with the linearity of expectation.

$$p_{1*} = \frac{n_s}{n} p_{1*,s}, \quad p_{10} = p_{10,s}, \quad p_{01} = p_{01,s} \tag{1}$$

Taking Equation 1 into the limit of F_β-score [2], we can derive $\tilde{F}_\beta$ for the whole dataset based on $p_{1*,s}$, $p_{10,s}$ and $p_{01,s}$ as the following equation.

$$\tilde{F}_\beta = \frac{(1+\beta^2) \cdot \frac{n_s}{n} p_{1*,s} \cdot (1 - p_{01,s})}{\frac{n_s}{n} p_{1*,s} \cdot (\beta^2 + 1 - p_{01,s} - p_{10,s}) + p_{10,s}}$$

Taking the gradient of the logarithm of $\tilde{F}_\beta$ yields the condition at the first-order stationary point(s) $\partial \tilde{F}_\beta / \partial \Theta = \mathbf{0}$ in Equation 2.

$$\begin{aligned} -\frac{\nabla p_{10,s}}{1 - p_{10,s}} &= \frac{\nabla p_{01,s}}{\beta^2 \cdot \frac{n_s}{n} p_{1*,s} / (1 - \frac{n_s}{n} p_{1*,s}) + p_{01,s}} \\ &= \frac{\nabla p_{01,s}}{\beta^2 \cdot \frac{p_{1*,s}}{\alpha(1-p_{1*,s})} + p_{01,s}} \end{aligned} \tag{2}$$

where $\nabla p_{10,s}$ and $\nabla p_{01,s}$ are their gradients with respect to Θ.

On the other hand, using the smooth functions [1], the first order Taylor expansion around $\mathbb{E}(*)$, and the linearity of expectation, we approximate the expectation of L_{SaSu} as follows.

$$\begin{aligned} &\mathbb{E}(L_{SaSu}(f(x;\Theta), y)) \\ &= \mathbb{E}[-\log f(x;\Theta) \,|\, y=1] + \mathbb{E}[\log(\beta^2 \frac{\bar{p}_{1*,s}}{\alpha(1-\bar{p}_{1*,s})} + f(x;\Theta)) \,|\, y=0] \\ &\approx -\log(1 - \tilde{p}_{10,s}) + \log(\beta^2 \frac{\bar{p}_{1*,s}}{\alpha(1-\bar{p}_{1*,s})} + \tilde{p}_{01,s}) \end{aligned}$$

Examining $\partial/\partial\Theta\, \mathbb{E}(L_{SaSu}(f(x;\Theta), y)) = \mathbf{0}$, we can derive its condition at the first-order stationary point(s) in Equation 3.

$$-\frac{\nabla \tilde{p}_{10,s}}{1 - \tilde{p}_{10,s}} = \frac{\nabla \tilde{p}_{01,s}}{\beta^2 \cdot \frac{p_{1*,s}}{\alpha(1-p_{1*,s})} + \tilde{p}_{01,s}} \tag{3}$$

The gradient property in Equation 3 is equivalent to Equation 2. Hence, Lemma 1 holds that L_{SaSu} approximately shares the same condition of first-order stationary points with F_β-score on the whole training set. □

Precisely, the first order Taylor expansion-based approximation in Equation is an upper bound for $\mathbb{E}(L_{SaSu})$. This could be verified by examining larger order Taylor expansions [3] of $\mathbb{E}(L_{SaSu})$. Intuitively, this bound is tighter when the dataset is more imbalanced, hence well serving real-world IED detection. We will conduct the regret bound analysis with regard to the imbalance ratio, i.e., $\forall \epsilon > 0, \mathrm{P}(\mathbb{E}(|F_\beta(\Theta^*_{SaSu}) - \tilde{F}_\beta(\Theta^*)| < \epsilon)) > 1 - g(\epsilon, r_{im})$, in our future studies.

REFERENCES

[1] Martin Jansche. 2005. Maximum expected F-measure training of logistic regression models. In *HLT-EMNLP*.
[2] Namgil Lee, Heejung Yang, and Hojin Yoo. 2021. A surrogate loss function for optimization of F_β score in binary classification with imbalanced data. *arXiv* 2104.01459 (2021).
[3] Yee Whye Teh, David Newman, and Max Welling. 2006. A Collapsed Variational Bayesian Inference Algorithm for Latent Dirichlet Allocation. In *NIPS*.