Hyper-parameter Optimized Semi-Supervised Learning of Physiological Signal Anomaly Detection for Cardiac Health Analytics

Arijit Ukil, Soma Bandyopadhyay, Arpan Pal
TCS Research and Innovation
Tata Consultancy services, India
{arijit.ukil, soma.bandyopadhyay, arpan.pal}@tcs.com

Abstract

The class imbalance issue impairs the robustness of the practical clinical analytics solution. The key idea of our approach is learning by optimal hyper-parameters through metaheuristic search. The proposed method constructs smooth, unbiased decision boundary for robust semi-supervised learning. We choose two critical physiological signals related to cardiac condition: Heart sound or Phonocardiogram (PCG) and Electrocardiogram (ECG) to validate our claim of robust anomaly detection of clinical events. We perform extensive experiments on publicly available expert-labelled MIT-Physionet PCG and ECG databases to establish the performance merit of the proposed model in comparison with relevant state-of-the-arts.

1 Introduction

We demonstrate the efficacy of our proposed method of addressing the class-imbalance problem by 1. Constructing hyper-parameter (kernel co-efficient γ, rejection rate hyper-parameter ν) optimized one-class support vector machine (OC-SVM) with radial basis function (RBF) kernel. 2. Showing empirical evidences of robust anomaly detection from PCG and ECG signals on publicly available MIT-Physionet Datasets [Phy. 2016] [Phy. 2017] of better performance than state-of-the-art algorithms. In fact, class augmentation seems to be beneficial in analyzing physiological signals [Ukil et al. 2018].

.2 Cardio-vascular Management System

We introduce in Figure 1, a closed-loop architecture of cardio-vascular health management system. The analytics engine, hosted in edge devices like smartphones, analyzes the sensor signals (ECG, PCG) for taking various actions like anomaly detection (e.g. noisy signal removal), clinical inferencing and related others.

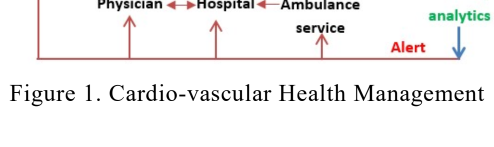

Figure 1. Cardio-vascular Health Management

3 Hyper-parameter Optimized Semi-supervised Learning

Let, $\mathcal{X} = (\mathbb{X}_+, \mathbb{X}_-)$, where $\mathbb{X}_+$ be the available majority class training set and $\mathbb{X}_-$ be the minority class training set, $\mathbb{X}_+ = \{\mathbb{x}_i^+\}_{i=1}^{\Pi}$, $\mathbb{X}_- = \{\mathbb{x}_i^-\}_{i=1}^{\pi}$, where $\Pi \gg \pi$; $\mathbb{x}_i^+, \mathbb{x}_i^- \in \mathbb{R}^d$; and $y_i \in \{+1, -1\}$ be the corresponding class labels. The optimization function with Radial Basis Function (RBF) kernel of OC-SVM for $(\alpha_1, \dots \alpha_n)$ non-negative Lagrange multipliers and n denotes the number of training vectors is solved using Lagrange dual problem [Schölkopf et al. 1999]:

$$minimize \left\{\frac{1}{2}\sum_{i,j}\alpha_i\alpha_j e^{\left(-\gamma\|\mathbb{x}_i - \mathbb{x}_j\|^2\right)}\right\} \quad (1)$$

subject to: $0 \leq \alpha_i \leq \frac{1}{\nu n}$, $\sum_{i=1}^{n}\alpha_i = 1$

From equation (1), we observe that the kernel co-efficient γ and rejection rate hyper-parameter ν are not independent. Our approach is to jointly optimize γ, ν with the criteria validation performance maximization (or validation loss minimization) while ensuring some-defined stability function.

Our method uses a part of the positive or majority class instances for training purpose: $\mathcal{X}_{Train} = \mathbb{X}_+^{Train}$. In validation set, both positive and negative class instances are present: $\mathcal{X}_{Validation} = \{\mathbb{X}_+^{validation}, \mathbb{X}_-\}$, and $\mathbb{X}_+ = \{\mathbb{X}_+^{Train}, \mathbb{X}_+^{validation}\}$. Let Ψ_ν, Ψ_γ be the total search space of hyper-parameters $\mathbb{S}_{\nu,\gamma} = \{(\nu_1, \gamma_1), \dots, (\nu_{\Psi_\nu}, \gamma_1), (\nu_2, \gamma_1), \dots, (\nu_{\Psi_\nu}, \gamma_{\Psi_\gamma})\}$ with the range of $\nu = (0,1)$ and $\gamma = (2^{-4}, 2^5)$. We first train the classifier OC-SVM over all the hyper-parameters in $\mathbb{S}_{\nu,\gamma}$ and $\Psi_\nu \times \Psi\gamma$ trained models are generated. Each of the trained models is validated against $\mathcal{X}_{Validation}$. Let the corresponding response space (response space is the outcome, in terms of accuracy, sensitivity or other related performances metrics) be $\mathbb{P} = \left\{\rho_{\theta_\nu, \theta_\gamma}\right\}, \theta_\nu = 1,2, \dots, \Psi_\nu;\ \theta_\gamma = 1,2, \dots, \Psi_\gamma$. Typical grid search approach performs poorly on generalization (overfits on the training data). In this work, we propose metaheuristic optimization by stable region searching, where we search over a stable response region $\mathbb{P}_{stable}$ based on a defined stability criteria.

Input: $\mathbb{P} = \left\{\rho_{\theta_\nu, \theta_\gamma}\right\}, \theta_\nu = 1,2, \dots, \Psi_\nu;\ \theta_\gamma = 1,2, \dots, \Psi_\gamma$.

Output: $k_{opt} = \{\gamma_{optimal}, \nu_{optimal}\}$

Procedure:

1. **Search region identification:** We find out the outliers or inconsistent points $\mathbb{P}_{outliers}$ by DBSCAN algorithm [Ester et al. 1996] in the complete response space $\mathbb{P}$ and discard those outliers and form $\mathbb{P}_{stable} = \mathbb{P} \setminus \mathbb{P}_{outliers}$.
2. **Stability parameter:** We compute stability parameter $\mathcal{s}_{stable} = \left(\frac{max(\mathbb{P}_{stable})}{standard\ deviation(\mathbb{P}_{stable})}\right)$, $\mathcal{s}_{stable}$ ensures that the maximum reported performance reported in the $\mathbb{P}_{stable}$ is high simultaneously *not* vastly deviated from the mean performance of $\mathbb{P}_{stable}$. Higher $max(\mathbb{P}_{stable})$ and lower $standard\ deviation(\mathbb{P}_{stable})$ is intended.
3. **Stable search region identification:** $\mathbb{P}_{stable}$ matrix is divided into $\mathcal{L}$ equal quadrants, typically $\mathcal{L} = 4$. $\mathbb{P}_{stable}^{1,2,3,4}$ are formed. Again remove the outliers in each of the $\mathbb{P}_{stable}^{1,2,3,4}$. Next, $\mathcal{s}_{stable}^{1,2,3,4}$ is computed for outlier eliminated $\mathbb{P}_{stable}^{1,2,3,4}$. $\mathbb{P}_{stable}^{1,2,3,4}$ are the children of $\mathbb{P}_{stable}$.
4. **Stable region selection:** Our notion is that when the stability parameter $\mathcal{s}_{stable}$ of the mother region (e.g. $\mathbb{P}_{stable}$) is more than any of the child regions, the mother region achieves stability. Hence, we compute the stability parameter of the mother region $\mathcal{s}_{stable}|_{mother}$ and the children regions (part of the mother region, illustrated in Figure 2) $\mathcal{s}_{stable}|_{child}$ and $\eta = \mathcal{s}_{stable}|_{mother} - \mathcal{s}_{stable}|_{child}$. Stop if $\eta \geq 0, \forall child$ the mother region $\mathbb{P}_{stable}^{\text{mother}}$ is selected.
5. Else, Goto step 1 with $\mathbb{P} = \mathbb{P}_{stable}^{\text{mother}}$.
6. $\mathcal{k}_{opt} = \{\gamma_{optimal}, \nu_{optimal}\}$: the selected value of $\{\gamma, \nu\}$ corresponds to: $\underset{\rho}{\text{argmax}}\, \mathbb{P}_{stable}^{\text{mother}}$.

We illustrate the metaheuristic notion of our proposed hyper-parameter optimization algorithm in Figure 2. Step 1 of the algorithm provides the heuristic guidance, whereas Step 2, 3,4,5 are the heuristic search method.

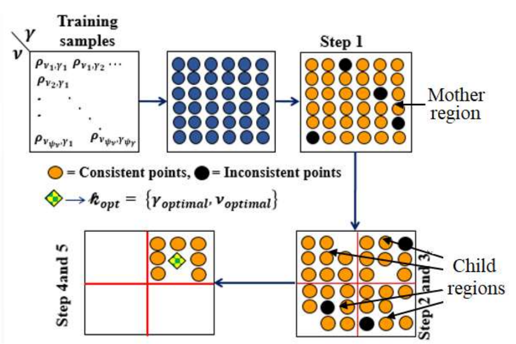

Figure 2. Illustration: Proposed Hyper-parameter Optimization

4 Experimental Results

Publicly available expert-labeled annotations are available at MIT-Physionet database [Phy. 2016] [Phy. 2017]. However, many of the PCG and ECG signals are noisy and corrupted with noise from different noise sources, device faults and motion artifacts. Computational analysis or machine learning on such anomalous data does not ensure correct or reliable clinical inference. The practical problem is that expert-annotated anomalous (noisy) PCG or ECG data are rare in availability. E.g. PCG (number of instances~ 3500, class imbalance ratio~ 1: 5) and ECG (number of instances~ ~ 8700, class imbalance ratio~ 1: 200) signals from MIT-Physionet database. We apply our hyper-parameter optimized method to identify noisy ECG and PCG signals. The results as shown in Figure 3 is the *K*-fold cross-validated (*K*=5) performance. Apart from regular performance metrics like sensitivity and specificity, we consider *GM* = Geometric Mean (sensitivity, specificity) as the measure of balanced performance score. We find that our method demonstrates better balanced performance than the state-of-the-art algorithms, B: [Haixiang et al. 2017], C: [Chung et al. 2011], D: [Muandet. et al. 2013], where **A**: **Our proposed method**. Under the class-imbalance scenario, we state that the proposed algorithm ensures consistent balanced performance (in terms of *GM*) for each of the physiological signals.

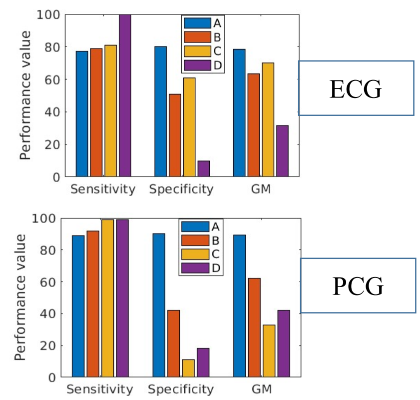

Figure 3. Noisy Signal Identification Performance Comparison

References

[Chung et al. 2011] Chih-Chung Chang, Chih-Jen Lin. LIBSVM: A library for support vector machines. ACM TIST, 2011.

[Ester et al. 1996] Martin Ester, et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In ACM KDD. 1996.

[Haixiang et al. 2017] Guo Haixiang, et al. Learning from class-imbalanced data. Journal Expert Systems with Applications, pp. 220 - 239, 2017.

[Muandet. et al. 2013] Krikamol Muandet, Bernhard Sch¨olkopf. One-Class Support Measure Machines for Group Anomaly Detection. One-class support measure machines for group anomaly detection. In UAI, 2013.

[Phy. 2016] https://physionet.org/challenge/2016/

[Phy. 2017] https://physionet.org/challenge/2017/

[Schölkopf et al. 1999] Bernhard Schölkopf, et al. Support vector method for novelty detection. In NIPS, 1999.

[Ukil et al. 2018] Arijit Ukil, et al. Class Augmented Semi-Supervised Learning for Practical Clinical Analytics on Physiological Signals. Machine Learning for Health (ML4H) Workshop at NeurIPS 2018, https://arxiv.org/abs/1812.07498.