2012 IEEE Second Conference on Healthcare Informatics, Imaging and Systems Biology

Predicting Complications of Percutaneous Coronary Intervention using a Novel Support Vector Method

Gyemin Lee*, Hitinder S. Gurm† and Zeeshan Syed*
*Department of Electrical Engineering and Computer Science
†Department of Internal Medicine
University of Michigan, Ann Arbor, MI, USA
Email: gyemin@umich.edu, hgurm@med.umich.edu, zhs@umich.edu

***Abstract*—Clinical tools to identify patients at risk of complications during percutaneous coronary intervention (PCI) are important to determine care at the bedside and to assess quality and outcomes. We address the growing need for such tools by proposing a novel support vector machine (SVM) approach to stratify PCI patients. Our approach simultaneously leverages properties of both one-class and two-class SVM classification to address the diminished prevalence of many important PCI complications. When studied on the Blue Cross Blue Shield of Michigan Cardiovascular Consortium (BMC2) multi-center cardiology registry data, our SVM method provided moderate to high levels of discrimination for different PCI endpoints, and improved model performance in many cases relative to both traditional one-class and two-class SVMs.**

I. Introduction

Despite a decline in the rate of complications during percutaneous coronary intervention (PCI), the mortality and morbidity associated with these events remains constant and high. Algorithms that can accurately risk stratify PCI patients are essential to reduce this burden, to better evaluate patients at the bedside and to make more meaningful assessments of quality and outcomes. However, predicting PCI complications remains challenging due to small model derivation data sets and severe class imbalance. These factors adversely affect the traditional approach of developing PCI models within a supervised learning framework. In particular, its discriminative ability suffers from a lack of sufficient positive (i.e., event) examples for model training. While collecting additional data offers one way to address this situation, it is often infeasible because of delays, expenses, and burden to both patients and caregivers. As an alternative, unsupervised learning has recently been explored to develop clinical models in the presence of small datasets with few positive examples. While this approach is promising, it does not consistently improve performance over supervised learning approaches. In this setting, we introduce a novel support vector machine (SVM) classification approach, called the one-plus-class SVM (OP-SVM), that incorporates aspects of both supervised (two-class SVM) and unsupervised (one-class SVM) learning. We explore this approach to develop models for PCI outcomes using data from the multi-center Blue Cross Blue Shield of Michigan Cardiovascular Consortium (BMC2) registry.

II. Method

The OP-SVM solves the following optimization problem:

$$\min_{\mathbf{w},\boldsymbol{\xi}} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i$$
$$\text{s.t.} \quad y_i\langle \mathbf{w}, \Phi(\mathbf{x}_i)\rangle \leq y_i + \xi_i, \quad \xi_i \geq 0, \quad \forall i.$$

The OP-SVM finds a hyperplane that separates data points while considering their labels. We studied this approach to predict thirteen in-laboratory PCI complications in the BMC2 registry. We compared the OP-SVM to the following existing methods: (i) logistic regression, (ii) one-class SVM on the entire/non-event patients, (iii) two-class SVM, We also compared cost-sensitive (CS) variants of the TC-SVM and OP-SVM. All models were trained on data from 22,023 patients treated in 2008, and tested on data from 20,289 patients from 2009.

III. Results

The proposed OP-SVM(CS) performed best on seven PCI endpoints and second best on three PCI endpoints. On the other hand, the next best approach TC-SVM(CS) was best only for two PCI endpoints. Table I presents the average AUROC and rank of each algorithm over thirteen in-laboratory PCI outcomes. The OP-SVMs provided improved risk discrimination relative to all the other approaches.

IV. Conclusion

We present a novel approach to develop models for PCI complications. Our approach simultaneously leverages properties of two-class and one-class SVM, and achieves higher levels of discrimination than a variety of existing two-class and one-class methods. We believe this allows for PCI models that have increased clinical utility at the bedside, and for objective assessments of quality and outcomes.

Table I. The AUROC and rank of clinical models averaged over thirteen PCI outcomes.

	LR	OC	OC (Neg)	TC	TC (CS)	OP	OP (CS)
Average AUC	0.669	0.700	0.702	0.711	0.743	0.707	0.750
Average Rank	5.23	4.92	4.92	4.00	2.54	4.23	2.15

978-0-7695-4921-7/12 $26.00 © 2012 IEEE
DOI 10.1109/HISB.2012.14