ISCE25: Poster 35 technical details

Juan José García

March 24th 2025

1 Introduction

Conformal model combination (CMC) [1] proposes a way to combine multiple risk stratification scores obtained with [2] to produce a single score that controls the false positive rate (FPR) and the false negative rate (FNR). The hope is that the combination will produce a risk score with smaller intermediate risk region (i.e. larger coverage) than all of the individual scores. Note this notion of FPR and FNR is not complementary to sensitivity and specificity. To be complementary, all cases in the intermediate risk category must be correctly classified as high-risk or low-risk.

2 ACS risk stratification scores

To risk stratify ACS in the prehospital setting we consider: the FasterRisk score [3], the HEAR score [4], the GBDT score [5, 6] and a DL-ECG score [7]. Each score takes as input different measurements of a patient. A list of inputs for GBDT, FasterRisk is in Table 1 of [6]. A list of inputs to the HEAR score is in Figure 2 of [2] and the input to DL-ECG is a prehospital ECG trace. We refer to the space of all measurements of a patient (i.e. a n-tuple of signs, symptoms, ECG trace and ECG interpretations) as $\mathcal{X}$. We apply the method in [2] on an i.i.d. sample D_n of our deployment site to estimate cutoffs l_m, h_m for each score f_m. Please review the **abstract** for the number of: calibration samples to estimate l_m, h_m; training samples to estimate scores f_m; and external validation samples to evaluate the method.

3 Conformal model combination (CMC)

[2] proposes an algorithm that leverages a deployment site sample D_n to control the FPR and FNR of explainable scores using class conditional conformal estimation. More concretely, given two constants α_-, α_+ and a score function f_m, it determines thresholds l_m, h_m which are functions of a random calibration sample $D_n = ((X_i, Y_i)_{i=1}^n)$. Assuming a new sample (X, Y) and D_n are i.i.d., it follows $P(f_m(X) < l_m | Y = +) \leq \alpha_+$ and that $P(f_m(X) > h_m | Y = -) \leq \alpha_-$

where the probability is over $(X,\, D_n)$. Note in [2], f_m is an explainable score but the result holds for non-explainable scores (e.g. the output of a probabilistic classifier $f_m(X) = \hat{P}_m(Y = +|X)$). Accordingly, we propose algorithm 1 to risk stratify a patient given the a collection of models, with corresponding CC stratification cutoffs $\{(f_m, l_m, h_m)\}_{i=1}^{M}$, measurements from a patient $x \in \mathcal{X}$ and a weight vector w in the probability simplex Δ_M.

Algorithm 1 Risk stratification through CMC

Require: $\{(f_m, l_m, h_m)\}_{i=1}^{M}$, $w \in \Delta_M$, $x \in \mathcal{X}$
Ensure: $\forall_{m\in[M]}(P\{f_m(X) < l_m|Y = +\} \leq \alpha_+ \wedge P\{f_m(X) > h_m|Y = -\} \leq \alpha_-)$

```
h(x) ← (1/M) Σ_{m=1}^{M} 1{f_m(x) > h_m}
l(x) ← (1/M) Σ_{m=1}^{M} 1{f_m(x) < l_m}
if h(x) > 1/2 then
    f(x) ← High risk
else if l(x) > 1/2 then
    f(x) ← Low risk
else
    f(x) ← Intermediate risk
end if
return f(x)
```

By proposition 1, it follows that $P\{h(X) > 1/2|Y = -\} \leq 2\alpha_-$ and $P\{l(X) > 1/2|Y = +\} \leq 2\alpha_+$. We emphasize the probability is w.r.t the calibration dataset D_n and patient measurements X. Accordingly, this algorithm preserves control of the FPR and FNR on the new sample.

Proposition 1. *Given a collection of (score, low-risk cutoff, high-risk cutoff) triplets (i.e. $\{(f_m, l_m, h_m)\}_{m=1}^{M}$), defined as above, such that $P(f_m(X) < l_m|Y = +) \leq \alpha_+$ and that $P(f_m(X) > h_m|Y = -) \leq \alpha_-$. It follows $P(\frac{1}{M}\sum_{m=1}^{M} 1\{f_m(X) < l_m\} > 1/2|Y = +) \leq 2\alpha_+$ and $P(\frac{1}{M}\sum_{m=1}^{M} 1\{f_m(X) > h_m\} > 1/2|Y = -) \leq 2\alpha_-$*

Proof. $P(\frac{1}{M}\sum_{m=1}^{M} 1\{f_m(X) < l_m\} > 1/2|Y = +) \leq 2E[\frac{1}{M}\sum_{m=1}^{M} 1\{f_m(X) < l_m\}|Y = +] = 2\frac{1}{M}\sum_{m=1}^{M} P\{f_m(X) < l_m|Y = +\} \leq 2\alpha_+$. Where the first inequality follows from Markov's inequality, the second inequality from linearity of expectation and the third inequality by construction of prediction intervals with CC. The proof for $P(\frac{1}{M}\sum_{m=1}^{M} 1\{f_m(X) > h_m\} > 1/2|Y = -) \leq 2\alpha_-$ is analogous. $\square$

References

[1] Matteo Gasparin and Aaditya Ramdas. Merging uncertainty sets via majority vote, March 2024. arXiv:2401.09379 [stat].

[2] Juan Jose Garcıa, Nikhil Sarin, Rebecca R Kitzmiller, Ashok Krishnamurthy, and Jessica K Zegre-Hemsey. Risk stratification through class-conditional conformal estimation: A strategy that improves the rule-out performance of MACE in the prehospital setting. *Proceedings of Machine Learning Research*, 252:1-15(Machine Learning for Healthcare), July 2024.

[3] Jiachang Liu, Chudi Zhong, Boxuan Li, Margo Seltzer, and Cynthia Rudin. FasterRisk: Fast and Accurate Interpretable Risk Scores, October 2022. arXiv:2210.05846 [cs].

[4] Jason P. Stopyra, William S. Harper, Tyson J. Higgins, Julia V. Prokesova, James E. Winslow, Robert D. Nelson, Roy L. Alson, Christopher A. Davis, Gregory B. Russell, Chadwick D. Miller, and Simon A. Mahler. Prehospital Modified HEART Score Predictive of 30-Day Adverse Cardiac Events. *Prehospital and Disaster Medicine*, 33(1):58–62, February 2018.

[5] Andrey Malinin, Liudmila Prokhorenkova, and Aleksei Ustimenko. Uncertainty in Gradient Boosting via Ensembles, April 2021. arXiv:2006.10562 [cs].

[6] Juan Jose Garcia, Rebecca Kitzmiller, Ashok Krishnamurthy, and Jessica K. Zègre-Hemsey. Selective classification with machine learning uncertainty estimates improves ACS prediction: A retrospective study in the prehospital setting., June 2024.

[7] Ran Xiao, Cheng Ding, Xiao Hu, and Jessica Zègre-Hemsey. ML for MI - Integrating Multimodal Information in Machine Learning for Predicting Acute Myocardial Infarction, October 2022.