

Conformal confidence sets for binary image segmentation

Samuel Davenport

August 12, 2024

Abstract

1 Introduction

2 Introduction

3 Theory

Let $\mathcal{V} \subset \mathbb{R}^m$ be finite set corresponding to the domain, where $m \in \mathbb{N}$. In our context \mathcal{V} will represent the pixels/voxels at which we observe imaging data. Given a distribution \mathcal{D} on $\mathcal{X} \times \mathcal{Y}$ suppose that we observe i.i.d. from pairs $(X_i, Y_i)_{i=1}^n$ on \mathcal{D} . Here $X_i : \mathcal{V} \rightarrow \mathbb{R}$ represents the observed image and $Y_i : \mathcal{V} \rightarrow \{0, 1\}$ outputs labels at each $v \in \mathcal{V}$ giving 1s at the true location of the objects in the image X_i that we wish to identify and 0s elsewhere.

Let $\mathcal{X} = \{g : \mathcal{V} \rightarrow \mathbb{R}\}$. Let $s : \mathcal{X} \times \mathcal{V} \rightarrow \mathbb{R}$ be a score function - trained on an independent dataset - such that given an image pair $(X, Y) \in \mathcal{X} \times \mathcal{Y}$, $s(X, v)$ is intended to be higher at the $v \in \mathcal{V}$ for which $Y(v) = 1$. The score function can for instance be derived from a deep neural network image segmentation method such as U-net CITE.

In what follows, given an error rate α we will construct a classifier $c : \mathcal{X} \times \mathcal{V} \rightarrow \mathbb{R}$ such that given a new image pair $(X, Y) \sim \mathcal{D}$, and letting $L(X) = \{v \in \mathcal{V} : c(X, v) = 1\}$, we have

$$\mathbb{P}(Y(v) = 1 \text{ for all } v \in L(X)) \geq 1 - \alpha.$$

This corresponds to controlling $\mathbb{P}(Y(v) = 0 \text{ for all } v \in L(X))$, i.e. the probability of making an error, to a level α . This error rate is analogous to the familywise error rate from the multiple testing setting, an observation that allows us to control it using the distribution of the maximum in the spirit of Westphal-Young.

To do so, let $T_i = \max_{v \in \mathcal{V}: Y_i(v)=0} s(X_i, v)$ and define

$$\lambda_\alpha = \inf \left\{ \lambda : \frac{1}{n} \sum_{i=1}^n 1[T_i \leq \lambda] \geq \alpha \right\}.$$

be the upper α -quantile of the distribution of the maximum of the score function of the observed image over the areas at which the true label is equal to 0. Then define the classifier, $c : \mathcal{X} \times \mathcal{V}$ such that

$$c(X, v) = 1[s(X, v) > \lambda_\alpha].$$

Theorem 3.1. *Given $(X, Y) \sim \mathcal{D}$ independent of $(X_i, Y_i)_{i=1}^n$, we have*

$$\mathbb{P}(Y(v) = 1 \text{ for all } v \in L(X)) \geq 1 - \alpha.$$

Proof. Suppose that $Y(v) = 0$ for some $v \in L(X)$, then it follows that $s(X, v) > \lambda_\alpha$ and conversely. Thus the event $\{Y(v) = 0 \text{ for some } v \in L(X)\}$ occurs if and only if $\max_{v \in \mathcal{V}: Y(v)=0} s(X, v) > \lambda_\alpha$. Let $T_{n+1} = \max_{v \in \mathcal{V}: Y(v)=0} s(X, v)$. Then the vector (T_1, \dots, T_{n+1}) is exchangeable, so arguing as in XXX, it follows that T_{n+1} is equally to lie between (or before/after) the values T_1, \dots, T_n . As such $\mathbb{P}(T_{n+1} > \lambda_\alpha) \leq \alpha$ and the result follows. \square

4 Results

5 Applications

5.1 Polyps Tumor Segmentation

5.2 Brain Mask Segmentation

6 Acknowledgements

I'm grateful to Habib Ganjgahi at the Big Data Institute at the University of Oxford for useful conversations on this topic.

7 Discussion

Acknowledgements

References