Review for
"Conformal prediction beyond exchangeability"

Jiaxin Hu

Conformal prediction is a powerful tool to quantify the prediction uncertainty of "black box" algorithms through the empirical performance on testing data. However, many conformal methods heavily rely on the i.i.d. or exchangeability assumption on the training and testing data. Also, classical conformal methods focus on the symmetric algorithm; i.e., prediction results are not affected by the order of the data points. These exchangeability assumptions on data and algorithm are unrealistic in many applications with streaming data, time sensitive data, and so on. The prediction interval coverage may also decrease when these assumptions are violated.

This work extends previous conformal methods to the non-exchangeable scenario and quantify the coverage gap by the total variation distance between the permuted data. The key idea to address the non-exchangeability is to introduce the weight for each data. The data more "trusted" is assigned with a larger weight; e.g., in streaming data, the data with a larger index has a larger weight than previous data. The prediction intervals thus involve the pre-determined weights, and the coverage gap is upper bounded by the weighted sum of total variation distance between permuted data.

Since regular conformal prediction is the special case of the weighted method with weights all equal to 1, the new theorems for coverage gap also indicates the robustness of regular method against the mild violation of the exchangeability.

Random thoughts

1. One of the limitations of this work lies in the weight selection. In this work, the weights are pre-selected rather than automatically adjusted by the data. An interesting question would be: what is the optimal choice of the weight? The optimality can refer as the shortest length of the prediction interval while keeping a high coverage rate. My intuition is: the optimal weight should be a decreasing function of the total variation distance of the permuted data.

2. Notice that i.i.d. assumption is a stronger and a special case of the exchangeability assumption. As far as I am concerned, many works on exchangeability focus on the vector data; i.e., let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ denote a random vector, and then $\boldsymbol{x}$ is exchangeable if $\pi \circ \boldsymbol{x} \sim_d \boldsymbol{x}$ for all permutations π on $[n]$.

   How to define the exchangeability for matrix or tensor data? Let $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ denote a random tensor, and $\Pi_k \in \{0,1\}^{n_k \times n_k}$ denote the permutation matrix on the k-the mode. We call the tensor $\mathcal{X}$ is exchangeable if

   $$\mathcal{X} \times_1 \Pi_1 \times_2 \cdots \times_K \Pi_K \sim_d \mathcal{X}.$$

   In previous work, we usually assume the noise in the "signal+noise" model is i.i.d.. Therefore, we may ask whether we can extend our previous works on tensor regression and clustering to the case with exchangeable noise? The exchangeable assumption allows the dependence among the noise entries. Also, it is also curious to know the real applications about the exchangeable noise.

References