

Conformal Prediction

UMA AVALIAÇÃO COMPARATIVA

**Ana Carolina Erthal, Bernardo Vargas,
Felipe Lamarca e Guilherme de Melo**

Escola de Matemática Aplicada,
Fundação Getulio Vargas

ESTRUTURA DA APRESENTAÇÃO

1	Introdução	2
1.1	Objetivo da pesquisa	2
1.2	Conformal Prediction	3
1.3	Tipos de Conformal Prediction	4
2	Introdução à Split Conformal Prediction	5
2.1	Calibration step	5
2.2	Geração de prediction sets	6
3	Metodologia	7
3.1	Principais referências	7
3.2	Como tornar Conformal Prediction comparável?	8
3.3	Limitações	9
3.4	Métrica de comparação	10
4	Resultados	11
4.1	Replicação de Guo et al. (2017)	11
4.2	Comparações	12
5	Referências bibliográficas	15

INTRODUÇÃO

OBJETIVO DA PESQUISA

Comparar empiricamente a capacidade de quantificar de incerteza para a Classificação Multiclasse supervisionada da **Conformal Prediction** — prática que vem ganhando popularidade na área de Machine Learning — em relação a métodos de calibração.

O objetivo é determinar, de acordo com algumas métricas de avaliação, se o uso da Conformal Prediction se justifica frente às abordagens mais bem consolidadas na literatura.

🔗 Repositório do projeto

INTRODUÇÃO

CONFORMAL PREDICTION

Objetivo: estabelecer prediction sets — conjuntos de *labels* possíveis para algum *input* dada a confiança desejada.

A predição conforme se propõe a estabelecer garantias probabilísticas sobre modelos arbitrários com o mínimo possível de suposições.

$$1 - \alpha \leq \mathbb{P}(Y_{\text{test}} \in \mathcal{C}(X_{\text{test}})) \leq 1 - \alpha + \epsilon$$

$(X_{\text{test}}, Y_{\text{test}})$ novo ponto de dado sob a mesma distribuição

$\mathcal{C}(X_{\text{test}})$ conjunto de possíveis *labels*

α taxa de erro escolhida¹

INTRODUÇÃO

TIPOS DE CONFORMAL PREDICTION

Full Conformal Prediction utiliza os dados completos para fazer subsequentes treinamentos do modelo.

Split Conformal Prediction realiza-se um split dos dados para obter um **conjunto de calibração**. Ele é utilizado para definir \hat{q} , a linha de corte que separa os $1 - \alpha$ melhores scores de calibração dos piores, ou seja, \hat{q} é o quantil $1 - \alpha$ de s_1, s_2, \dots, s_n .

E por que utilizamos o **Split Conformal Prediction**?

1. Partimos de dados já particionados;
2. Partimos de modelos pré-treinados;
3. O custo computacional bastante baixo, já que só é necessário treinar o modelo uma vez.

Embora a Full Conformal Prediction tenha sido desenvolvida primeiro, a Split Conformal Prediction é considerada um caso especial importante e utilizada pela comunidade de Predição Conforme.

INTRODUÇÃO À SPLIT CONFORMAL PREDICTION

CALIBRATION STEP

Calibration step

- ▶ Dado um modelo para classificação treinado $\hat{f} : \mathcal{X} \rightarrow \mathcal{Y}$, produzimos uma matriz X com os *outputs softmax*, onde cada entrada x_{jk} representa a probabilidade de a imagem j pertencer à classe k .
- ▶ Criamos scores de não-conformidade:

$$s_i = 1 - \hat{f}(X_i^{\text{cal}})_{Y_i^{\text{cal}}}$$

- ▶ • Definimos \hat{q} , quantil $1 - \alpha$ de $s_1, s_2, \dots, s_{n^{\text{cal}}}$.

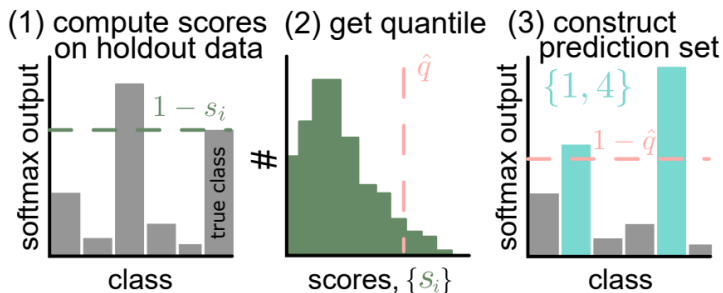


Figure. Calibration step. Fonte: Angelopoulos e Bates (2022, p.4).

INTRODUÇÃO À SPLIT CONFORMAL PREDICTION

GERAÇÃO DE PREDICTION SETS

Por fim, para um x_i^{test} não visto, criamos o prediction set (\mathcal{C}) contendo os labels cujos valores softmax $\geq 1 - \hat{q}$, garantindo a cobertura desejada:

$$P(y^{\text{test}} \in \mathcal{C}(X^{\text{test}})) \geq 1 - \alpha$$



['barn spider', 'European garden spider', 'wolf spider']
0.82 0.12 0.02





['southern black widow']
0.99

Figure. Exemplos de prediction sets na ImageNet

METODOLOGIA

PRINCIPAIS REFERÊNCIAS

As duas principais referências que embasam esta análise são Angelopoulos e Bates (2022), com o artigo [A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification](#), e Guo et al. (2017), em [On Calibration of Modern Neural Networks](#).

- ▶ Escolhemos duas redes neurais presentes nos testes do artigo Guo et al. (2017)
- ▶ ResNet110 pré-treinada no CIFAR10, com base nesse repositório .
- ▶ ResNet152 pré-treinada na ImageNet disponibilizada pelo [PyTorch](#)
- ▶ Implementamos a predição conforme em `Python` a partir do repositório  disponibilizado por Angelopoulos e Bates (2022).

METODOLOGIA

COMO TORNAR CONFORMAL PREDICTION COMPARÁVEL?

- ▶ Para cada entrada x_i^{test} , iteramos possíveis valores para α no intervalo $[0, 1]$, atualizando seu valor até o momento em que o prediction set $\mathcal{C}(x_i)$ para aquela determinada entrada possua apenas uma *label* possível.
 - **Complicação prática:** não é possível escolher um α arbitrariamente pequeno, devido a uma limitação que inviabiliza o cálculo dos quantis na predição conforme.
 - Embora essa abordagem não seja usual em Conformal Prediction, o objetivo é torná-la comparável e poder calcular seu erro de calibração.

METODOLOGIA

LIMITAÇÕES

Durante o desenvolvimento do projeto, questionamos em partes nossas motivações. Queremos, é claro, comparar as abordagens e chegar a conclusões claras quanto à eficácia e relevância do CP.

Forçar a conformal prediction a escolher uma classe para as comparações é necessário. No entanto, não é isso que ela se propõe a fazer.

- Não é claro se as garantias teóricas da Conformal Prediction valem para o método proposto, mas no caso assintótico nos parece uma abordagem válida.

"Isto é comparar laranjas com maçãs."

- Daniel Csillag

"Podemos pintar as maçãs de laranja."

- Felipe Lamarca

METODOLOGIA

MÉTRICA DE COMPARAÇÃO

Para a comparação, seguimos os *papers* de referência, em particular o método utilizado por Guo et al. (2017), em *On Calibration of Modern Neural Networks*:

- ▶ Expected Calibration Error (ECE)
 - Média ponderada da diferença absoluta entre a acurácia e a confiança das i faixas de confiança geradas.
- ▶ Maximum Calibration Error (MCE)
 - ECE utilizando norma infinita.

RESULTADOS

REPLICAÇÃO DE GUO ET AL. (2017)

Após termos nosso modelo, conjunto de dados e técnicas de comparação, decidimos replicar os resultados do *paper* Guo et al. (2017) para validar nossos modelos e métricas, buscando atingir os mesmos resultados.

Table. Comparação de desempenho

Table. Desempenho do experimento

	ECE	NLL	Acc.
ResNet110	4.4%	0.299	93.6%
ResNet110 T.S.	2.5%	0.206	93.6%
ResNet152	4.9%	0.873	78.5%
ResNet152 T.S.	4.1%	0.861	78.5%

Table. Desempenho no paper

	ECE	NLL	Acc.
ResNet110	4.6%	0.328	93.8%
ResNet110 T.S.	0.83%	0.210	93.8%
ResNet152	5.48%	0.896	77.7%
ResNet152 T.S.	1.86%	0.865	77.7%

- ▶ A ResNet110 foi aplicada no CIFAR-10 e a ResNet152 foi aplicada na ImageNet
- ▶ T.S. abrevia Temperature Scaling

RESULTADOS

COMPARAÇÕES

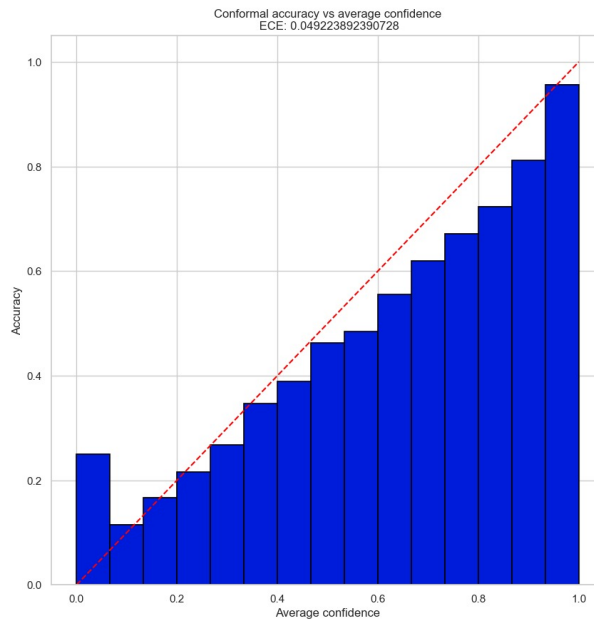


Figure. ResNet antes da calibração

RESULTADOS

COMPARAÇÕES

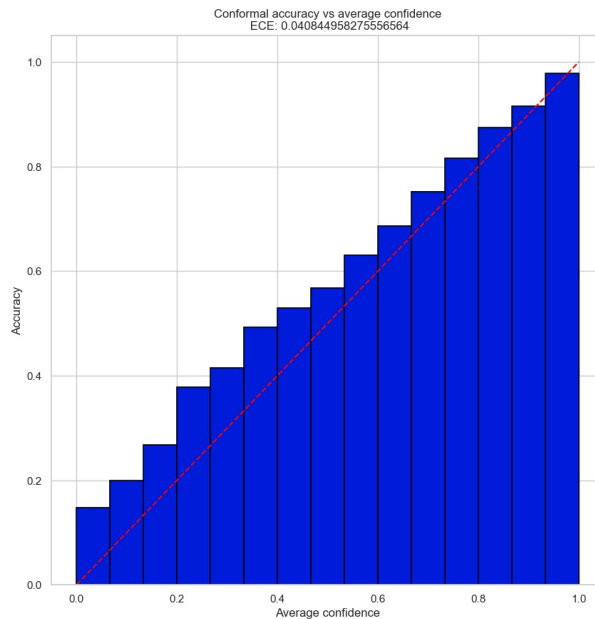


Figure. ResNet calibrada

RESULTADOS

COMPARAÇÕES

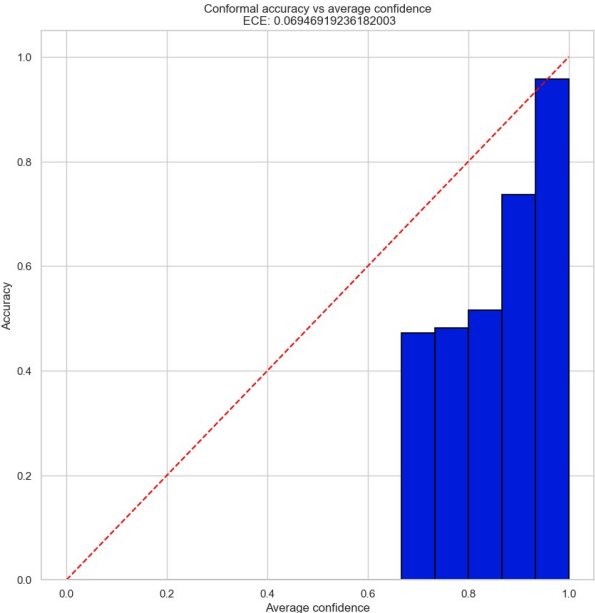


Figure. Conformal

REFERÊNCIAS BIBLIOGRÁFICAS

- 📄 ANGELOPOULOS, A. N.; BATES, S. A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification. *arXiv*, 2022.
- 📄 GUO, C. et al. On Calibration of Modern Neural Networks. *arXiv*, 2017.
- 📄 MUKHOTI, J. et al. Calibrating Deep Neural Networks using Focal Loss. *arXiv*, 2020.
- 📄 SHAFER, G.; VOVK, V. A Tutorial on Conformal Prediction. *arXiv*, 2007.