Weekly Study Report

Wang Ma

2024-12-31

Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute

Outline

1. Credal Deep Ensembles for Uncertainty Quantification

Credal Sets. Suppose we have 2 possible outcomes: $\Omega = \{A, B\}$, a credal set over these outcomes might be:

$$\mathbb{C} = \{P : P(A) + P(B) = 1, 0.2 \leq P(A) \leq 0.6, 0.3 \leq P(B)\} \leq 0.5\}$$

.

This is another way to model the distributions of output Probability.

Main Challenges:

- How to get the bounds of the predictive probability?
- Different ways may result in different bounds, what the bounds represent in real-world terms?

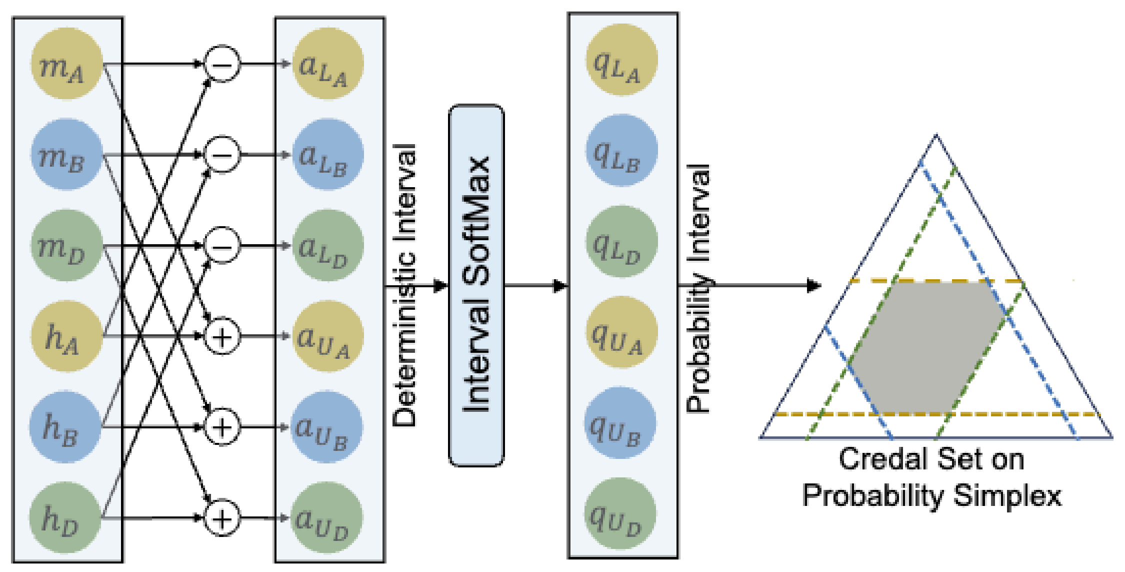

Figure 2: CreNet final layer structure for three classes.

1. Get intervals for logits (each logit has a lower bound and an upper bound)
2. From logits' interval, compute an "interval SoftMax" to get the output Credal Sets

1.2.1 Interval SoftMax

$$q_{L_i} = \frac{\exp(a_{L_i})}{\exp(a_{L_i}) + \sum_{k \neq i} \exp(\frac{a_{U_k} + a_{L_k}}{2})}, \quad q_{U_i} = \frac{\exp(a_{U_i})}{\exp(a_{U_i}) + \sum_{k \neq i} \exp(\frac{a_{U_k} + a_{L_k}}{2})},$$

q_{L_i} means i-th class's probability lower bound and q_{U_i} means i-th class's probability upper bound.

Vanilla SoftMax cannot satisfy the requirements if credal sets of probability:

1. $q_{L_i} \leq q_{U_i}$
2. $\sum q_{L_i} \leq 1 \leq \sum q_{U_i}$

Note. here t is the label of x.

Key Idea. Train the lower bound and upper bound separately.

$$\mathcal{L}_{\text{CreNet}} := \underbrace{\frac{1}{N}\sum_{n=1}^{N} \text{CE}(\boldsymbol{q}_{U_n}, \boldsymbol{t}_n)}_{\text{Vanilla Component}} + \underbrace{\underset{\boldsymbol{w}\in\mathbb{S}}{\text{maximize}} \frac{1}{N}\sum_{n=1}^{N} w_n \cdot \text{CE}(\boldsymbol{q}_{L_n}, \boldsymbol{t}_n)}_{\text{DRO Component}},$$

1. Vanilla CE often yields optimistic estimates
2. DRO (Distributionally Robust Optimization) can result in overly pessimistic results, because they focus on the worst cases.

1.3 How to trian the Model

Algorithm 1 CreNet Training Procedure

Input: Training dataset $\mathbb{D} = \{\boldsymbol{x}_n, \boldsymbol{t}_n\}_{n=1}^{N}$; Portion of samples per batch $\delta \in [0.5, 1)$; Batch size η

while *enable training* **do**

 1. Compute $\mathrm{CE}(\boldsymbol{q}_{U_n}, \boldsymbol{t}_n)$ and $\mathrm{CE}(\boldsymbol{q}_{L_n}, \boldsymbol{t}_n)$ for each sample

 2. Sort the sample indices $(m_1, \ldots, m_\eta)$ in descending order of $\mathrm{CE}(\boldsymbol{q}_{L_n}, \boldsymbol{t}_n)$

 3. Define $\eta_\delta = \lfloor \delta \eta \rfloor$

 4. Minimize $\mathcal{L}_{\text{CreNet}} = \frac{1}{\eta} \sum_{n=1}^{\eta} \mathrm{CE}(\boldsymbol{q}_{U_n}, \boldsymbol{t}_n) + \frac{1}{\eta_\delta} \sum_{j=1}^{\eta_\delta} \mathrm{CE}(\boldsymbol{q}_{L_{m_j}}, \boldsymbol{t}_{m_j})$

end while

Class Prediction. Two strateges:

$$i_{\min} = \text{argmax}_i q_{L_i} \; ; \qquad i_{\max} = \text{argmax}_i q_{U_i} .$$

Uncertainty Quantification.

Total Uncertainty: $\overline{\mathbb{H}}(\mathbb{Q}) = \text{maxmize} \sum_{i=1}^{C} -q_i \log q_i$

Aleatoric Uncertainty: $\underline{\mathbb{H}}(\mathbb{Q}) = \text{minimize} \sum_{i=1}^{C} -q_i \log q_i$

Epistemic Uncertainty: $\text{EU} = \overline{\mathbb{H}}(\mathbb{Q}) - \underline{\mathbb{H}}(\mathbb{Q})$

Deep Ensemble:

$$\tilde{q}_L^* = \frac{1}{M}\sum_{m=1}^{M} q_{L_m}^*, \;\; \tilde{q}_U^* = \frac{1}{M}\sum_{m=1}^{M} q_{U_m}^*$$

UQ in Deep Ensemble:

$$\mathrm{TU} := \overline{\mathbb{H}}(\mathbb{Q}) = \text{maxmize} \sum_{i=1}^{C} -\tilde{q}_i \log \tilde{q}_i$$

$$\mathrm{AU} := \underline{\mathbb{H}}(\mathbb{Q}) = \text{minimize} \sum_{i=1}^{C} -\tilde{q}_i \log \tilde{q}_i$$

$$\mathrm{EU} := \mathrm{TU} - \mathrm{AU}$$

Strength.

1. This may be the first paper to combine Credal Sets with NN tasks. Credal Sets way is good since it has well-explored UQ methods. (Although the utilized method is not the only way/best way to quantify Credal Uncertainty).

Limitations.

1. The proposed way to construct credal sets is too straightforward. The training process may be not as stable as normal methods.
2. They use Deep Ensembles to confirm the results, this actually means their proposed Credal set NN is not very stable. The usage of Ensemble introduces more computation, even more expensive the normal DEs.
3. As an original paper, this paper leaves many improvement points.