Experiments

Baselines

- **Few-shot learning models**
  - CNP(Conditional neural process, ICML'18): combines neural network and gaussian process by using a small set of input-output pairs as context to output prediction.
  - ANP(Attentive neural process, ICLR'19): outputs prediction based on concatenation of learned distribution of context, context features and given input.
  - MAML(ICML'17): learn a set of shared model parameters across different tasks which can rapidly learn novel task with a small set of labeled data.
  - Meta-SGD(arXiv'17): beside MAML, also learns step sizes and update direction during the training procedure.

Experiments

Performance Comparison: Twitter

Method	Twitter				Weibo			
	5-Shot		10-Shot		5-Shot		10-Shot	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
VQA	73.62 ± 1.83	76.69 ± 1.23	73.49 ± 2.61	74.69 ± 2.97	76.93 ± 0.71	75.88 ± 0.45	77.80 ± 1.43	76.36 ± 1.77
attRNN	63.04 ± 2.09	60.25 ± 4.63	63.14 ± 2.00	56.60 ± 5.25	76.07 ± 1.63	74.36 ± 2.96	78.09 ± 0.58	77.69 ± 0.35
EANN	70.01 ± 3.58	72.95 ± 2.86	70.56 ± 1.00	67.77 ± 0.80	76.43 ± 0.84	74.51 ± 0.56	77.49 ± 1.95	76.56 ± 1.28
CNP	71.42 ± 2.58	72.58 ± 3.57	72.47 ± 3.61	72.11 ± 5.74	77.47 ± 5.19	77.01 ± 4.66	78.81 ± 1.57	78.07 ± 1.98
ANP	77.08 ± 2.92	79.65 ± 3.81	74.25 ± 0.76	75.16 ± 1.27	77.85 ± 1.67	76.00 ± 3.61	76.52 ± 1.84	73.73 ± 2.78
MAML	82.24 ± 1.54	82.97 ± 1.76	85.22 ± 0.64	84.98 ± 1.70	74.68 ± 0.75	74.16 ± 0.33	75.87 ± 0.33	73.41 ± 0.86
Meta-SGD	74.13 ± 2.31	75.35 ± 2.56	74.63 ± 2.46	74.57 ± 2.74	71.73 ± 1.81	69.51 ± 2.28	73.34 ± 2.35	71.42 ± 2.80
MetaFEND (Improvement)	**86.45 ± 1.83 (↑5.12%)**	**86.21 ± 1.32 (↑3.91%)**	**88.79 ± 1.27 (↑4.19%)**	**88.66 ± 1.09 (↑4.33%)**	**81.28 ± 0.75 (↑4.41%)**	**80.19 ± 1.27 (↑4.13%)**	**82.92 ± 0.13 (↑5.22%)**	**82.37 ± 0.28 (↑5.51%)**

- In 5-shot setting, compared with CNP, ANP incorporates the attention mechanism and hence achieve more informative context information.
- In 10-shot setting, as the size of give support data increases, the soft-attention ANP unavoidably incorporates the irrelevant data points.