Problem

- Question answering (QA) is widespread, impactful NLP task
- Current QA models have trouble with domain adaptation
- Holy grail: a model that can generalize with only small amount of out-of-domain training data

Background + Methods

![Diagram of model-agnostic meta-learning algorithm (MAML)](image)

**Implemented Model-Agnostic Meta-Learning (MAML) training regime**
- Each in-domain dataset treated as a separate training task
- Leveraged learn2learn library to streamline second-order backpropagation

![Algorithm 1 MAML Training Loop](algorithm)

1. Figure from Chelsea Finn, Pieter Abbeel, and Sergey Levine: Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks, 2017

Experiments

- **FS-Base (“Few-Shot Baseline”):** finetune the baseline model with out-domain training data
- **Meta-base: train new baseline model on in-domain data using Algorithm 1, then finetune on out-domain data**

**Idea: Subdivide in-domain datasets**
- MAML shines when trained on many well-formed tasks
- Meta-NewsQA-Split: split NewsQA into two datasets to augment tasks, motivated by context length distribution (Figure 2)
- Meta-Even-Split: split all train datasets into small, medium, and large datasets to augment tasks

![Figure 2: Context Length Across QA Datasets](image)

Table 3: Model scores on validation set

<table>
<thead>
<tr>
<th>Model</th>
<th>EM</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>30.63</td>
<td>47.72</td>
</tr>
<tr>
<td>FS-base</td>
<td>33.51</td>
<td>49.83</td>
</tr>
<tr>
<td>Meta-base</td>
<td>33.51</td>
<td>47.37</td>
</tr>
<tr>
<td>Meta-NewsQA-Split</td>
<td>31.15</td>
<td>46.33</td>
</tr>
<tr>
<td>Meta-Even-Split</td>
<td>30.89</td>
<td>46.48</td>
</tr>
</tbody>
</table>

Analysis

- Challenging to beat FS-base, which uses classic training system
- Likely need more tasks for meta-learning to show its strength
- Potential to further optimize number of examples, meta learning rate, number of adaptation steps
- Task augmentation through dataset subdivision does not yield promising results

Conclusions

- **Meta-Learning seems promising for improving QA robustness**
- More well-formed tasks (e.g., all available QA datasets) and using the best hyperparameters could lead to increased performance over standard training regime (FS-base)