MAML-Based Models

Algorithm 1 Model-Agnostic Meta-Learning

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters
1: randomly initialize θ
2: **while** not done **do**
3: Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
4: **for all** $\mathcal{T}_i$ **do**
5: Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ with respect to K examples
6: Compute adapted parameters with gradient descent: $\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$
7: **end for**
8: Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'})$
9: **end while**

Parallel MAML

- ➔ Ensemble baseline with a MAML model at eval time
- ➔ Pick the result with the highest confidence

Version	F1	EM
Baseline	49.881	34.555
Vanilla-MAML	45.021	28.092
ParallelMeta	47.858	31.675

Exploration of Techniques for Robust Question Answering

Bhavik Shah + Sudeep Narala

Introduction

With the rise in availability of high quality language data, as well as the constant improvements made in machine learning models, large scale question-answering problems have been pushed further and further into the forefront of the natural language processing field. Now, extremely expressive transformer-based models such as BERT are considered state of the art in these types of tasks. However, we do notice that these models require large quantities of data in order to be effectively trained in some question-answering task. Moreover, we notice that these models often struggle to generalize to datasets of questions that are out-of-domain, which often have smaller quantities of - and more niche - information. As such, we explore techniques in this paper that can be useful for the adaptation of larger models

One field that has become increasingly popular in reinforcement learning and classification domains in recent years is meta-learning, which is the process of "learning how to learn". More concretely, we are able to learn more general purpose parameters such that we can encourage the model to learn how to quickly tune itself to a new, unseen task. These methods can be particularly useful in few-shot learning situations, as a model can quickly learn how to adapt itself when given only a few examples.

Additionally, we try other methods for encouraging the model to perform well on the out-of-domain sets. Particularly, we try forms of data augmentation, where we modify the original, larger datasets to be more "challenging", which can allow us to then perform better when evaluating on the harder domains. Moreover, we try methods that involve modifying the final layers and/or loss function of the model.

Sigmoid Loss

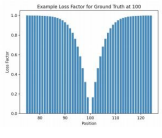
Idea

Penalize based on distance from ground truth answer (more akin to F1 than EM)

$$loss = -log(y[i]) - \alpha \sum_{n=0}^{max\_query\_len} f(n) \cdot log(1 - y[n])$$

where

$$f(x) = 2 \cdot |\sigma(\frac{x-i}{\gamma}) - 0.5|$$



Performance

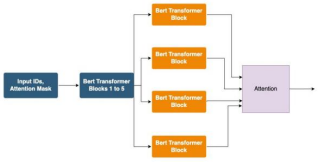
Data Augmentation

Answer Masking

Context Masking

Transformer Level Attention (Clustering)

Architecture



Model Interpretation

Results

Model Type	EM	F1
Baseline	49.881	34.56
Baseline + OOD Finetune	51.24	35.86
Baseline + Answer_Masking	50.71	35.60
Clustering	49.56	34.29
Clustering + OOD Finetune	50.01	35.34
Clustering + Context_Masking(0.15)	49.22	34.29
Clustering + Context_Masking(0.25)	49.00	33.77
Clustering + Answer_Masking	51.32	35.34

Table 1: Model Results (Dev Set)

Conclusions:

- ➔ MAML seems to underperform due to the distribution of tasks not being as clear when compared to something like omniglot.
- ➔ Sigmoid loss made F1 scores higher on the indomain task.
- ➔ Having organized code is very very important for running tests efficiently

Next Steps:

- ➔ Tune α and γ and experiment with γ being based on ground truth answer length in sigmoid loss method
- ➔ Try to make task distribution for MAML
- ➔ Try synonym replacement for data augmentation
- ➔ Combine these approaches (eg: MAML w/ data augmentation, sigmoid loss w/ anything other than baseline)