Meta Learning for Tasks in Pragmatics and Social Meaning

Tamara Czinczoll, Christoph Hönes, Daniel Rodríguez Baleato and Ard Snijders

Motivation

The availability of large labeled corpora is crucial for the performance of neural networks. However, for many pragmatics tasks such corpora are small or not available at all, since annotation is costly. We consider two methods designed to adapt to new tasks more quickly:

- **Multi-task learning:** Simultaneously training on several (related) tasks and parameter sharing can improve performance on all or some of those tasks
- **Meta-learning:** *Learning to learn*, i.e. learning how to quickly adapt to new tasks, can improve performance in a *few-shot* scenario where little annotated data is available

Primary Research Question:
Can meta-learning adapt faster to new tasks related to pragmatics and social meaning than multi-task learning?

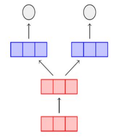
Tasks & Datasets

Name	Task	#C	Size
OLID	Offensive Language	2	13K
SE-Emotion	Emotion Detection	11	6.8K
FigLang	Sarcasm Detection	2	4K
SE-Sentiment	Sentiment Analysis	2	2.7K
WaseemHovy	Abusive Language	3	16K
SE-IronyA	Irony Detection	2	3.8K
SE-IronyB	Irony Detection	4	3.8K
Stanford	Politeness Detection	3	4.3k

Models

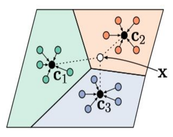
Multi-Task

Idea: Information learned for each task can improve learning for other tasks
How? Hard-shared encoder; separate output layers for each task



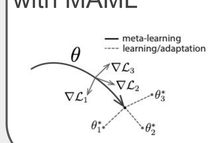
Proto-Network

Idea: Compute probability distribution over labels for all query examples, based on their similarity with the *prototypical* vectors



Proto-MAML

Intuition: learn internal representations with good generalization
Idea: Initialize output layer parameters with prototypes, train with MAML



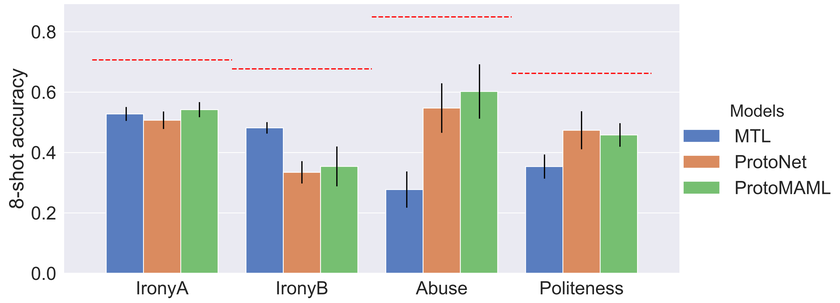
BERT for encoding + additional layer

- ***BERT-base***: 12 Transformer blocks (Only train top 2 blocks)
- Outputs a 768-d classification token
- CLS token is fed through single additional layer

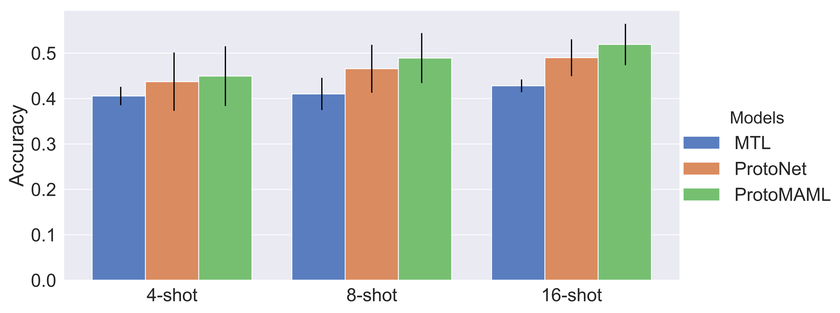
Experiments

- **K-shot performance**: Train each model on SE-Emotion, OLID, and FigLang tasks and evaluate 4-, 8- and 16-shot performance on remaining datasets.
- **Carousel tasks**: Train on all datasets except for one (leave one out) and compare the change in performance on Sentiment Analysis to identify tasks that are beneficial or harmful to performance.

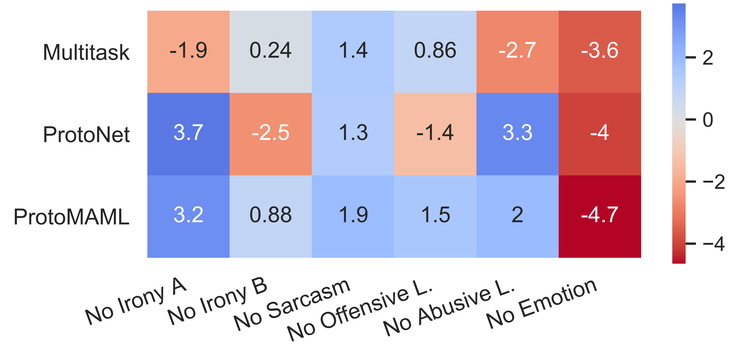
Results



Meta learning outperforms Multitask learning in most tasks but is inferior in some cases.
$\longrightarrow$ task dependent



Performance increases with more examples for all models. Difference especially in 4-shot case lower than expected.



The emotion task seems to be most beneficial to performance (redder = better).