Meta-Learning for Cross-lingual COVID-19 Fake News Detection in Low-Resource Languages

Giacomo Bugli
gb2572@nyu.edu

Luigi Noto
ln2205@nyu.edu

Anirudh Nistala
apn7823@nyu.edu

1 Introduction

As we went through a COVID-19 pandemic that imposed many challenges to the communities around the world and their health systems, serious questions arose surrounding the credibility of news about medical and political information, such as prevention, treatment, causes, and consequences of the disease on people's well-being. This posed the need for frameworks able to detect and flag unreliable articles before their spreading. In this period of information uncertainty, most of the automatic fake-news detection applications focused on news written in English. The reason for this is that, being English the principal language used for communication on the internet, the abundance of labelled data made it possible to implement deep learning (DL) models for text classification. However, this models are notorious for being data- and computation-intensive, making the great majority of languages in the world under-resourced for the successful application of such methods.

To overcome the limitations imposed by these approaches, a solution to the problem of automatic fake-news detection in low-resource languages could be resorting to meta-learning (Bengio et al., 1991), or 'Learning to Learn', a learning paradigm used to achieve high performance when the available labels are lacking. The meta-learning framework has been shown to allow faster fine-tuning, converge to better performance, and achieve outstanding results for few-shot learning in many applications. For this reason, we decided to replicate the paper by van der Heijden et al. (2021), proposing a meta-learning framework for few-shot cross-lingual adaptation and multilingual joint-learning for document classification tasks in different domains (namely, news stories and product reviews). The experiments demonstrate that meta-learning methods generally improve performance on both cross-lingual adaptation and multilingual joint-training.

Motivated by the substantial improvements and state-of-the-art achieved through the experiments on document classification in limited-resource settings, we propose to apply this approach for few-shot cross-lingual COVID-19 fake news detection in low resource languages.

2 Background Knowledge

2.1 Meta-Learning

Meta-learning refers to the idea of using previous knowledge experiences to guide efficient new tasks learning (learning to learn) by optimizing performance on the task distribution. Each learning task is associated with a dataset $\mathcal{D}$ containing both features and labels, and is split into a support set (used for fast adaptation) and a query set (used to evaluate performance after adaptation). In the meta-training part, the goal is to optimize for the best performance on the task distribution. So given training tasks $\mathcal{T}_i$ with labelled datasets $\mathcal{D}_i$, find a common meta parameter (often model parameter initialization) that plays a role in the task distribution

$$\omega^* = \underset{\omega}{\operatorname{argmin}} \sum_i \mathcal{L}_\omega(\mathcal{D}_i),$$

where $\mathcal{L}_\omega$ is the loss function of the learning algorithm parameterized by ω. A cycle of fast adaptation on a support set and evaluation on a query set for each task in the sampled batch is called an *episode*. In the meta-testing part, given an unseen target task $\mathcal{T}_j$, the goal is to use the learned meta-knowledge ω^* to obtain optimal task parameters with few support samples

$$\theta_j^* = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}_\theta(\mathcal{D}_j \mid \omega^*)$$

Different meta-learning approaches focus on learning different components. Three common approaches are: *metric-based* meta-learning (e.g. Prototypical Networks), which aim at learning good feature extractors and similarity measures over inputs; *model-based* meta-learning (e.g. Meta Networks), where the architecture is specifically designed for fast adaptation; *optimization-based* meta-learning (e.g. MAML), which aim at adjusting the optimization algorithm so that the model can be good at learning with a few samples.

2.2 ProtoMAML

ProtoMAML (Triantafillou et al., 2020) is a meta-learning method that combines the simple inductive bias of prototypical networks, which is effective for very-few-shot-learning, with the flexible adaptation mechanism of MAML (Finn et al., 2017). Prototypical Networks (Snell et al., 2017) use an embedding network f_θ to encode all samples in a support set S_c and compute *prototypes* μ_c per class $c \in C$ as

$$\mu_c := \frac{1}{|S_c|} \sum_{(x_i, y_i) \in S_c} f_\theta(x_i)$$

Using the definition of prototypes μ_c above, the weights w_c and bias b_c for a class c are computed as follows

$$w_c := 2\mu_c \quad b_c := -\mu_c^T \mu_c$$

ProtoMAML is an adaptation of MAML where the final linear layer is parameterized as per the equation above at the start of each episode using the support set. ProtoMAMLn is an adaptation of ProtoMAML, inspired by Wang et al. (2019), where L_2 Normalization is applied to the prototypes themselves. Further, to reduce the computational complexity of the meta-optimization step of MAML, which relies on second derivatives, a first-order approximation is used. Referred to as fo-ProtoMAMLn, van der Heijden et al. (2021) show that these adaptations stabilize training and lead to faster convergence to better local optima.

3 Datasets

First, we provide an overview of the datasets used in the paper replication (Amazon Sentiment Analysis) and in the proposed extension (MM-COVID).

Amazon Sentiment Analysis The Amazon Sentiment Analysis (Prettenhofer and Stein, 2010) dataset is a collection of product reviews in three different categories (books, dvds, and music) in English, French, German and Japanese. For each language, all product categories have been concatenated and the reviews distinguished between positive (rating > 3) and negative (rating < 3). The total number of samples for each language amounts to 12 000. To augment this dataset, 22 000 Chinese product reviews (distinguished between positive and negative) from the website JD.com have been added.

MM-COVID The MM-COVID (Li et al., 2020) dataset contains fake news content, social engagements, and spatial-temporal information in English, Spanish, Portuguese, Hindi, French, and Italian, but we only focus on the news content section. For each language, the reliable labels (fake and real) are obtained from fact-checking websites, and the source content of the news are retrieved from the original articles. In order to balance the dataset labels for each language, additional verified news are sourced from reliable websites selected by the authors. Table 1 displays a summary of the dataset after pre-processing.

Label	en	es	pt	fr	hi	it
Fake	1847	564	293	147	260	81
Real	4749	1830	637	246	1205	937

Table 1: Statistics of COVID-MM after pre-processing

4 Methods & Experimental Setup

Paper Replication The authors of the paper we decided to replicate conducted experiments in both cross-lingual adaptation and multi-lingual joint-training frameworks, and to prove effectiveness of the meta-learning approach in the few-shot learning framework, different meta-learning methods have been compared to a standard non-episodic supervised learning approach. For the experiments two different datasets have been used: MLDoc (Schwenk and Li, 2018), consisting of news stories in 8 languages (English, Spanish, French, Italian, Russian, Japanese and Chinese); and the above cited Amazon Sentiment Analysis. Moreover, for each cross- and multi-lingual framework

Setting	Method	Replication					Original Paper				
		de	fr	ja	zh	Δ	de	fr	ja	zh	Δ
High-Resource	Zero-Shot	87.4	87.3	83.7	79.8	84.6	91.2	90.7	87.0	84.6	88.4
	Non-Episodic	89.1	86.5	84.1	80.3	85.0	91.6	91.0	85.5	87.9	89.0
	foMAML	89.6	90.4	84.9	86.0	87.7	91.4	92.5	88.0	90.4	90.6
	foProtoMAMLn	90.2	90.7	86.8	87.2	88.7	92.0	93.1	88.6	89.8	90.9

Table 2: Average accuracy of 5 different seeds on the unseen target languages for Amazon. Δ corresponds to the average accuracy across test languages.

Setting	Method	Replication					Original Paper				
		de	fr	ja	zh	Δ	de	fr	ja	zh	Δ
Included	Non-Episodic	84.4	85.1	82.8	82.9	83.8	91.0	91.0	87.3	89.4	89.8
	foMAML	87.0	88.6	84.6	85.3	86.4	90.1	90.7	87.2	89.5	89.4
	foProtoMAMLn	89.4	90.0	86.7	87.2	88.3	90.7	91.5	88.0	90.4	90.2

Table 3: Average accuracy of 5 different seeds on the target languages in the joint-training setting for Amazon. Δ corresponds to the average accuracy across test languages.

both limited-resource and high-resource settings are considered, while for multi-lingual an additional distinction is made between English Included or Excluded from the meta-training. Considering that for each of this settings, the performances of foProtoMAMLn are compared against 5 baselines, the number of models to run would amount to 292.

Due to time and computational resources limitations, we decided to replicate only a subset of the experiments proposed in the paper. We thus focused on implementing cross-lingual document classification (High-Resource setting) and multilingual document classification (Included setting) tasks only on the Amazon Sentiment Analysis dataset, and we selected three baselines (Zero-Shot, Non-Episodic, foMAML), among the five proposed (Zero-Shot, Non-Episodic, ProtoNet, foMAML, Reptile), to compare the foProtoMAMLn results. This allowed us to reduce the number of models to run to 32.

As base-learner for all models we use XLM-RoBERTa model from HuggingFace, and for each meta-update method use the optimal hyperparameters specified by van der Heijden et al. (2021). Regarding the task language division into l_{aux}, l_{dev} and l_{tgt}, we set source language (l_{src}) to English, and adapted the auxiliary languages based on the testing language. The validation language (l_{dev}) is set to French in all cases except when we meta-test on French, in which we used l_{dev} = Chinese. The meta parameter update for each language is performed on a support set of 16 samples.

Extension For the proposed extension of COVID-19 fake-news detection, we use the same framework developed for the replication part, focusing on the case of cross-lingual classification. We compare the performances of foProtoMAMLn on the MM-COVID languages against dEFEND\C, the baseline without meta-learning proposed by Li et al. (2020). dEFEND\C is a variant of dEFEND (Shu et al., 2019), which uses sentence attention LSTM model to learn the news content representation and XLM-RoBERTa to get the initial text representation.

Also here, as base-learner we used XLM-RoBERTa model from HuggingFace, and after performing cross-validation on {*inner-loop learning rate, outer-loop learning rate*} the optimal set of parameters is {1e-4, 1e-5}. Concerning the task language division into l_{aux}, l_{dev} and l_{tgt}, we set source language (l_{src}) to English, and adapted the auxiliary languages based on the testing language. The validation language (l_{dev}) is set to French in all cases except when we meta-test on French, in which we used l_{dev} = Italian. The meta parameter update for each language is performed on a support set of 8 samples.

5 Results

Paper Replication Table 2 shows the accuracy scores obtained on Amazon Sentiment Analysis

Method	Metric	pt	fr	hi	it	Δ
dEFEND\C	Acc	75.0	84.0	79.0	85.0	77.2
	F_1	75.0	84.0	78.0	85.0	76.6
foProtoMAMLn	Acc	92.7	92.9	82.9	90.6	89.8
	F_1	94.8	94.3	88.5	94.6	93.1

Table 4: Average accuracy and F_1 score of 5 different seeds on the target languages in the cross-lingual setting for MM-COVID. Note that Li et al. (2020) trained dEFEND\C in a multilingual joint-training setting (with additional language dataset).

dataset in the cross-lingual (high-resource) setting, while Table 3 shows the accuracy scores obtained on Amazon Sentiment Analysis dataset in the multilingual (included) setting. We can see that, in both cases, we reach lower performances than the ones stated in the original paper but we observe a similar trend, with foProtoMAMLn still being the best performing model on average. The reasons for this discrepancy could rely on a potentially different data preprocessing, since the authors of the paper don't give information about it, or some minor changes in the models implementations.

Extension Table 4 displays the accuracy scores obtained for the fake-news detection problem. We can see how foProtoMAMLn is able to substantially outperform the classical multilingual baselines. The meta-learning approach gains 12 percentage points in accuracy on average, with a good performance increase on low-resource languages such as Hindi. Figure 1 shows the comparison of the ROC curve of foProtoMAMLn for all test languages. With respect to the baseline proposed by Li et al. (2020) (dEFEND\C), it is important to underline that we keep the target language unseen until meta-testing, while the baseline we are referring to has been trained on a small amount of data from the target language. We are thus comparing our cross-lingual implementation with a multilingual joint-training one. Nonetheless, the improvements obtained suggest that meta-learning is actually effective in improving the model generalization capabilities to unseen tasks.

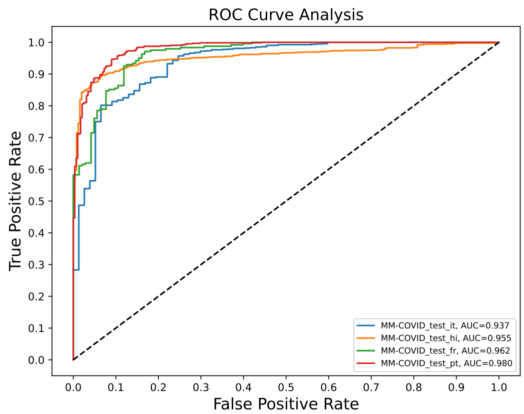

Figure 1: ROC curve in meta-testing for each unseen target language in the cross-lingual setting for MM-COVID. The predictions used to produce the curve are the ones obtained from the last of the 5 different seeds used to compute the test metrics.

6 Conclusion

We replicated the paper by van der Heijden et al. (2021) proposing a meta-learning framework for few-shot cross- and multilingual joint-learning for document classification tasks in different domains. In both frameworks, we see that foProtoMAMLn is the best performing model on average, although we haven't been able to reach the same results as the ones stated in the paper. As regards the extension, we showed the ability of the meta-learning approach to outperform classical multilingual models for fake-news detection.

7 Authors Contribution

Giacomo Bugli: literature review, development of code-base for meta-learning, Amazon dataset preprocessing, cross-lingual experiments on Amazon dataset, report writing
Luigi Noto: literature review, development of code-base for meta-learning, multilingual joint-training experiments on Amazon dataset, MM-COVID dataset preprocessing, cross-lingual experiments on MM-COVID
Anirudh Nistala: literature review, development of code-base for meta-learning, model performance analysis, report writing

References

Y. Bengio, S. Bengio, and J. Cloutier. Learning a synaptic learning rule. In *IJCNN-91-Seattle International Joint Conference on Neural Networks*, volume ii, pages 969 vol.2–, 1991. doi: 10.1109/IJCNN.1991.155621.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. *CoRR*, abs/1703.03400, 2017. URL http://arxiv.org/abs/1703.03400.

Yichuan Li, Bohan Jiang, Kai Shu, and Huan Liu. MM-COVID: A multilingual and multimodal data repository for combating COVID-19 disinformation. *CoRR*, abs/2011.04088, 2020. URL https://arxiv.org/abs/2011.04088.

Peter Prettenhofer and Benno Stein. Cross-language text classification using structural correspondence learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1118–1127, Uppsala, Sweden, July 2010. Association for Computational Linguistics. URL https://aclanthology.org/P10-1114.

Holger Schwenk and Xian Li. A corpus for multilingual document classification in eight languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA). URL https://aclanthology.org/L18-1560.

Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. Defend: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining*, KDD '19, page 395–405, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450362016. doi: 10.1145/3292500.3330935. URL https://doi.org/10.1145/3292500.3330935.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/cb8da6767461f2812ae4290eac7cbc42-Paper.pdf.

Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Utku Evci, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, and Hugo Larochelle. Meta-dataset: A dataset of datasets for learning to learn from few examples. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=rkgAGAVKPr.

Niels van der Heijden, Helen Yannakoudakis, Pushkar Mishra, and Ekaterina Shutova. Multilingual and cross-lingual document classification: A meta-learning approach. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1966–1976, Online, April 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.eacl-main.168. URL https://aclanthology.org/2021.eacl-main.168.

Yan Wang, Wei-Lun Chao, Kilian Q. Weinberger, and Laurens van der Maaten. Simpleshot: Revisiting nearest-neighbor classification for few-shot learning. *CoRR*, abs/1911.04623, 2019. URL http://arxiv.org/abs/1911.04623.