TIML: Task-Informed Meta-Learning for crop type classification

Gabriel Tseng
NASA Harvest & Mila
gabriel.tseng@mila.quebec

Hannah Kerner
University of Maryland, College Park
hkerner@umd.edu

David Rolnick
McGill University & Mila
drolnick@cs.mcgill.ca

Abstract

Agricultural data is extremely spatially imbalanced. When trying to learn from the data-rich regions to improve model performance in data-sparse regions, metadata - such as the relative distances of the training data points to the data-sparse region of interest - can improve model performance. We introduce task-informed meta-learning (TIML), an augmentation to model-agnostic meta-learning which conditions the model with this additional task information. TIML achieves higher AUC ROC scores than all other models in a variety of agroecologies. In addition, TIML excels in small in-distribution dataset-size regimes, strongly outperforming other methods when trained on 20× less data.

1 Introduction

The global food system both drives the climate crisis - contributing to 26% of global greenhouse gas emissions [9] - and is vulnerable to it, with every 1% increase in temperature lowering yields by 2.28% [5]. Crop maps, which allow us to better understand where crops are being grown, are important in helping us mitigate to and minimize the worse effects of climate change on agriculture, such as by monitoring food security, more rapidly responding to food crises and increasing productive land without sacrificing carbon sinks such as forests.

Certain parts of the world collect plentiful agricultural data, but many regions are extremely data sparse. There have been numerous efforts to learn from data-rich areas and transfer this knowledge to data-sparse regions or crop types, including meta-learning [10, 13, 14], transfer learning [16] and multitask learning [4]. However, these approaches often fail to capture important meta-data and expert knowledge about the data-sparse tasks of focus, such as their location relative to the pre-training data, or the crop types being classified. Conditioning the model with this information - and allowing it to learn different distributions depending on the task - may allow it to more effectively learn, especially in low data regimes.

We build on previous work learning multiple distributions in meta-learning models depending on the task being considered [15, 12]. However, we consider the special case where we have task-specific metadata which stays static for a single task but which can condition the model.

In this paper, we present Task-Informed Meta-Learning, or TIML, an algorithm designed to augment model-agnostic meta-learning with task metadata. Appyling TIML to crop type classification, we demonstrate that TIML works well across a wide range of agroecologies and crops. In addition, we show that TIML performs especially well in the small in-distribution dataset-size regime, outperforming other methods trained on 20× more data.

Preprint. Under review.

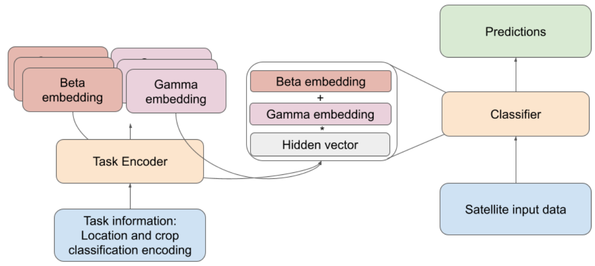

Figure 1: Task-Informed Meta-Learning, with feature wise linear modulation of parameters using task embeddings.

2 Data

We use the CropHarvest dataset [14] to train the model. This dataset consists of 88,145 datapoints, with the associated pixel time-series for each point. Of these datapoints, 28,564 (32.4%) contain land use labels; the remaining datapoints contain only "crop" or "non-crop" labels. Each datapoint is accompanied by a pixel timeseries from 4 remote sensing products: Sentinel-2 L1C, Sentinel 1, ERA5 and DEM, representing 1 year of data at monthly timesteps.

The CropHarvest dataset is additionally accompanied by 3 evaluation tasks, which test the ability of a model to learn from a small number of in-distribution datapoints in a variety of agroecologies:

Togo crop vs. non-crop: Classifying pixels containing crop from those which do not, in Togo. The training set consists of 1,319 datapoints, and the test set consists of 306 randomly sampled datapoints. We additionally measure the performance of the model when trained on a subset of training datapoints, using the following subset sizes: $\{20, 50, 126, 254, 382, 508, 636, 764, 892, 1020, 1148, 1319\}$.

The two other evaluation tasks consist of classifying a specific crop in a pixel. Thus, "rest" below refers to all other crop and non-crop classes. For both these tasks, entire polygons (as opposed to single pixels) were collected, allowing evaluation across them:

Kenya maize vs. rest: The training set consists of 1,345 imbalanced samples (266 positive and 1,079 negative samples). The test set consisted of 45 polygons.

Brazil coffee vs. rest: The training set consists of 203 imbalanced samples (21 positive and 182 negative samples). The test set consisted of 62 polygons.

3 Model Agnostic Meta-Learning

Model-agnostic meta-learning (MAML) [2] is an algorithm for learning a set of model weights θ which are close to the optimums of a variety of different tasks. The optimum for a specific task can then be reached with little data and/or few gradient steps. These initial weights θ are learnt by training on many other tasks, such that these tasks behave as training examples.

As with the CropHarvest benchmarks, we define tasks spatially using bounding boxes for countries drawn by [6]. Tasks consist of classifiying pixels that contain crop or classifying a specific crop-type.

3.1 Task-Informed Meta-Learning

We build on the original model-agnostic meta-learning [2] algorithm, considering the case where there is additional task-specific information which would inform the model, such as the spatial relationships between tasks. Information such as the spatial coordinates of a task (which we represented as the central coordinates of the task-country) remains static for all datapoints in the task, so is not useful to differentiate positive and negative instances. However, it may be useful to condition the model prior to inner loop training.

Algorithm 1 Task-Informed Meta-Learning

Require: $p(\mathcal{T})$: Distribution over tasks
Require: α, β: step size hyperparameters
randomly initialize meta model θ_m, task encoder θ_e
while not done **do**
  Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$ with task information t_i
  for all $\mathcal{T}_i, t_i$ **do**
    Generate task embeddings $\mu_i = f(t_i; \theta_e)$
    Evaluate $\nabla_{\theta_m} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_m}, \mu_i)$ with respect to K examples
    Compute adapted meta parameters with gradient descent: $\theta'_{m_i} \leftarrow \theta_m - \alpha \nabla_{\theta_m} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_m}, \mu_i)$
  Update $\theta_m \leftarrow \theta_m - \beta \nabla_{\theta_m} \Sigma_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_{m_i}}, \mu_i)$
  Update $\theta_e \leftarrow \theta_m - \beta \nabla_{\theta_e} \Sigma_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_{m_i}}, \mu_i)$

We therefore introduce Task-Informed Meta-Learning (TIML) (Algorithm 1), which modulates (using feature-wise linear modulation (FiLM)[8]) the hidden vectors in the meta-model based on embeddings calculated using task information. These embeddings are updated in the outer loop of the MAML training procedure (Figure 1).

For the CropHarvest dataset, task information is encoded in a 13-dimensional vector. 3 dimensions are used to encode spatial information, with latitude and longitude transformed to $[cos(\text{lat}) \times cos(\text{lon}), cos(\text{lat}) \times sin(\text{lon}), sin(\text{lat})]$. This ensures transformed values at the extreme longitudes are close to each other. The remaining 10 dimensions are used to communicate the type of task the model is being asked to learn. Specifically, a one-hot encoding of crop type categories was constructed from the FAO's indicative crop classification [1], with an additional class for non-crop. For crop vs. non-crop tasks, all crop type categories were assigned value $\frac{1}{n}$, for n crop type categories.

4 Experiments

MAML (and by extension, TIML) can be applied to any neural network architecture. We evaluate TIML by training it on the CropHarvest dataset, and fine-tuning it on the evaluation tasks, as happens for the benchmarks released alongside the dataset. We use the same classifier base as the benchmarks: a 1-layer LSTM model followed by a linear classifier.

In addition, we learn the embeddings using a task encoder. This encoder consists of linear blocks, where a block consisted of a linear layer with a GeLU activation [3] and Dropout [11] with value 0.2. The task information was encoded into a vector of size 128 by two linear blocks. An independent linear block then generated an embedding for each hidden vector to be modulated.

4.1 Baselines

We compare the TIML architecture to 5 baselines: (1) **MAML**: A model-agnostic meta-learning classifier without the task information, (2) **MAML + task info**: a MAML classifier, as before, but with the task information appended to the raw inputs. This evaluates the effect of adding the task information to the model independently of the TIML architecture, (3) **Pretrained**: A classifier pretrained to classify all data as crop or non-crop, (4) **Random**: A randomly initialized classifier and (5) **Random Forest**: A random forest, trained using scikit-learn [7] with the default hyperparameters.

For all deep learning-based models, we fine-tuned the models on the test tasks for 250 gradient steps. Each batch contained 10 positive and 10 negative examples.

4.2 Results & Discussion

We run each experiment 10 different times, to account for randomness and (for the subsampled Togo task) for differences in performance based on the subsample selected. Results for the tasks using all datapoints are shown in Table 1. The results on the Togo test task, with subsampling, are plotted in Figure 2. We report the AUC ROC score and the F1 score calculated using a threshold of 0.5.

Table 1: Results for the evaluation tasks. All results are averaged from 10 runs; results are reported with the accompanying standard error. We report the area under the receiver operating characteristic curve (AUC ROC) and the F1 score using a threshold of 0.5 to classify a prediction as the positive or negative class. The best metric for each task is in **bold**.

	Model	Kenya	Brazil	Togo
AUC ROC	Random Forest	0.598 ± 0.009	0.918 ± 0.003	0.889 ± 0.001
	Random	0.344 ± 0.006	0.901 ± 0.006	0.864 ± 0.003
	Pre-trained	0.712 ± 0.001	0.828 ± 0.002	0.895 ± 0.000
	MAML	0.675 ± 0.002	0.821 ± 0.004	0.886 ± 0.004
	MAML+info	0.729 ± 0.002	0.943 ± 0.003	0.888 ± 0.000
	TIML	$\mathbf{0.769 \pm 0.002}$	$\mathbf{0.945 \pm 0.001}$	$\mathbf{0.904 \pm 0.000}$
F1 score	Random Forest	0.583 ± 0.002	0.002 ± 0.001	0.750 ± 0.003
	Random	0.781 ± 0.001	$\mathbf{0.783 \pm 0.008}$	0.728 ± 0.004
	Pre-trained	0.821 ± 0.001	0.696 ± 0.002	0.666 ± 0.003
	MAML	0.840 ± 0.001	0.548 ± 0.002	0.656 ± 0.002
	MAML+info	0.835 ± 0.001	0.496 ± 0.001	0.649 ± 0.001
	TIML	$\mathbf{0.839 \pm 0.000}$	0.514 ± 0.001	$\mathbf{0.774 \pm 0.002}$

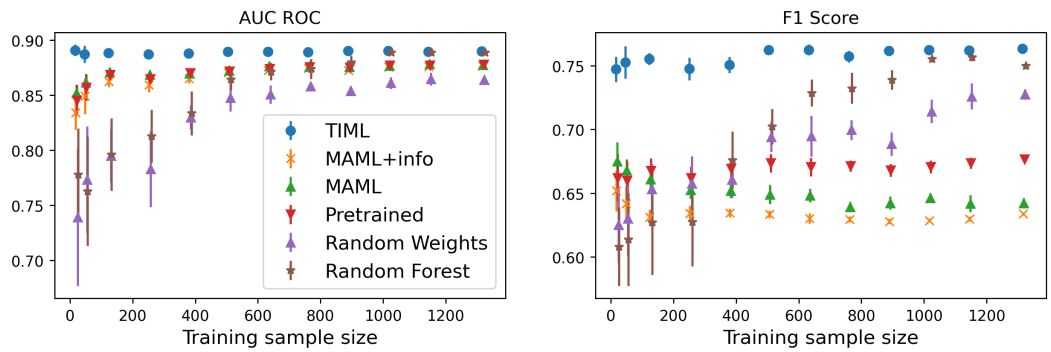

Figure 2: Results for the Togo test task as a function of dataset size. All results are averaged from 10 runs; results are reported with error bars representing the standard error. We report the area under the receiver operating characteristic curve (AUC ROC) and the F1 score. We highlight the improved performance of TIML for small subset sizes.

Adding task-information to the model using the TIML architecture improves AUC ROC score in all test-regions. We highlight the improved performance of the model given small in-distribution training set sizes (Figure 2) - the TIML model achieves comparable results when trained on 20 points as when it is trained on the whole dataset, outperforming all benchmarks.

The F1 score is variable relative to the AUC ROC score, suggesting the TIML model is not well-calibrated for the Brazil task. However, this a challenge for all models (e.g. the random forest in Brazil, or the pretrained model for the Togo tasks). This suggests that better calibrating the models, or picking better thresholds given very small (or non-existent) validation sets may significantly improve the utility of crop maps.

5 Conclusion

Accurate crop-type maps are important tools in minimizing the worst effects of climate change, but some regions lack the data to develop these crop maps, particularly when targeting specific crops. We introduce task-informed meta-learning (TIML), a method for conditioning the model with prior information about a specific task and apply it to the CropHarvest dataset, improving performance when the in-distribution training dataset is small. In addition, this method achieves better AUC ROC scores than all benchmark models across a range of agroecologies and dataset sizes.

All code used to train this model will be made available upon publication.

References

[1] Programme, concepts and definitions. In *World Programme for the Census of Agriculture*, volume 1. FAO, 2020.

[2] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135. PMLR, 2017.

[3] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (GELUs). *arXiv preprint arXiv:1606.08415*, 2016.

[4] Hannah Kerner, Gabriel Tseng, Inbal Becker-Reshef, Catherine Nakalembe, Brian Barker, Blake Munshell, Madhava Paliyam, and Mehdi Hosseini. Rapid response crop maps in data sparse regions. In *ACM SIGKDD Conference on Data Mining and Knowledge Discovery Workshops*, 2020.

[5] Dicle Ozdemir. The impact of climate change on agricultural productivity in asian countries: a heterogeneous panel data approach. *Environ Sci Pollut Res Int*, 2021.

[6] Tom Patterson and Nathaniel Vaughn Kelso. Natural earth. `https://www.naturalearthdata.com/`.

[7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[8] Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[9] Hannah Ritchie. Food production is responsible for one-quarter of the world's greenhouse gas emissions. *Our World in Data*, 2019.

[10] Marc Rußwurm, Sherrie Wang, Marco Korner, and David Lobell. Meta-learning for few-shot land cover classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 200–201, 2020.

[11] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[12] Eleni Triantafillou, Hugo Larochelle, Richard Zemel, and Vincent Dumoulin. Learning a universal template for few-shot dataset generalization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 10424–10433, 2021.

[13] Gabriel Tseng, Hannah Kerner, Catherine Nakalembe, and Inbal Becker-Reshef. Learning to predict crop type from heterogeneous sparse labels using meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1111–1120, 2021.

[14] Gabriel Tseng, Ivan Zvonkov, Catherine Nakalembe, and Hannah Kerner. CropHarvest: a global satellite dataset for crop type classification. Under submission, `https://openreview.net/forum?id=JtjzUXPEaCu`.

[15] Risto Vuorio, Shao-Hua Sun, Hexiang Hu, and Joseph J. Lim. Multimodal model-agnostic meta-learning via task-aware modulation. In *Neural Information Processing Systems*, 2019.

[16] Anna X. Wang, Caelin Tran, Nikhil Desai, David Lobell, and Stefano Ermon. Deep transfer learning for crop yield prediction with remote sensing data. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, COMPASS '18, 2018.