Targeted Meta-Learning for Critical Incident Detection in Weather Data

Mohammad Mahdi Kamani [1] **Sadegh Farhang** [1] **Mehrdad Mahdavi** [2] **James Z. Wang** [1]

Abstract

Due to imbalanced or heavy-tailed nature of weather- and climate-related datasets, the performance of standard deep learning models significantly deviates from their expected behavior on test data. Classical methods to address these issues are mostly data or application dependent, hence burdensome to tune. Meta-learning approaches, on the other hand, aim to learn hyperparameters in the learning process using different objective functions on training and validation data. However, these methods suffer from high computational complexity and are not scalable to large datasets. In this paper, we aim to apply a novel framework named as *targeted meta-learning* to rectify this issue, and show its efficacy in dealing with the aforementioned biases in datasets. This framework employs a small, well-crafted target dataset that resembles the desired nature of test data in order to guide the learning process in a coupled manner. We empirically show that this framework can overcome the bias issue, common to weather-related datasets, in a bow echo detection case study.

1. Introduction

Drastically improving their performance, machine learning and, more distinctively, deep learning models, are becoming the main propulsion of technology in a variety of domains. Notwithstanding their success, they are still suffering from different biases in the training data distribution. Biases, regardless of their nature, cause a mismatch between training and test data distributions, which leads to a poor generalization of the model on the test distribution. A palpable form of these biases appears when the size of different classes or groups are imbalanced. When class sizes are not balanced, imbalanced dataset problem arises (Buda et al., 2018; Huang et al., 2016; Ting, 2000), where large classes can dominate the training process, resulting in a model having low accuracy on small classes. A severe form of the problem, appearing in most real-world big datasets with immense number of classes, is long-tailed data distribution (Cui et al., 2019; Bengio, 2015; Ouyang et al., 2016), where the data distribution is skewed (Kendall et al., 1946). In this case, most of the data belongs to a few prevailing classes, while huge number of classes are represented by a few number of samples. Weather-related datasets, from real-time high temporal- or spatial-resolution satellite or radar image data, for detecting and predicting critical incidents mostly fall in this category. Important critical incidents are rare in such datasets, making detecting and forecasting these unusual events difficult for machine learning models.

A generic idea to address these biases is to adapt the training distribution to the test distribution, whether it is by resampling, or assigning weights based on the training loss to have a cost-sensitive weighting scheme (Elkan, 2001; Ting, 2000; Khan et al., 2018) such as boosting methods like AdaBoost (Freund & Schapire, 1997) or curriculum learning (Bengio et al., 2009; Jiang et al., 2018). However, relying merely on the training distribution has shown to be not practically efficient (Ren et al., 2018).

A new strand of research is to augment single-objective learning models with additional data-driven constraints in order to alleviate the effect of bias in training data (Cotter et al., 2018; Ren et al., 2018; Jiang et al., 2018; Andrychowicz et al., 2016). The main idea motivating this paradigm shift is that error rate on training data is not a satisfactory criterion by itself, and should be accompanied with a data-driven regularization. An appealing data-driven regularization idea is to create a target dataset that resembles desired properties of the test distribution, and impose it to the training process as an additional constraint. Meta-learning (Andrychowicz et al., 2016; Finn et al., 2017) introduces a coupled framework to interlace the hyperparameter tuning using a validation set with the training process in order to guide it.

In this paper, we adopt Targeted Meta-Learning framework, which employs a small target distribution and bilevel programming to model the multi-objective structure on both

[1]College of Information Sciences and Technology, The Pennsylvania State University, University Park, USA [2]Department of Computer Science & Engineering, The Pennsylvania State University, University Park, USA. Correspondence to: Mohammad Mahdi Kamani <mqk5591@psu.edu>.

Workshop at the 36^{th} International Conference on Machine Learning, Climate Change: How Can AI Help?, Long Beach, California, 2019.

training and target distributions. As a bilevel programming, targeted meta-learning has two levels, one dealing with the main training process while the other uses a well-tuned target dataset in order to optimize the weights of each desired class or group in the dataset. We will show that this framework can overcome the biases in weather-related learning problems by applying it to detect bow echoes in radar data.

2. Targeted Meta-Learning

Learning the learning process is a common practice nowadays in the machine learning field and in meta-learning approaches (Andrychowicz et al., 2016). However, most existing studies focus on tuning hyperparameters in the learning process. We will show that utilizing a well-crafted target dataset as a guidance for the main learning process can ameliorate the main learning problem. In targeted meta-learning, we impose a different objective function on the target distribution than the one for the training distribution.

The way we define two objective functions and their parameters is problem-independent. However, for the sake of exposition, we tackle the problem of imbalanced datasets in classification models. In all these problems, we are solving a prediction problem on dataset $\mathcal{T}$ with n training samples, from input space $\mathcal{X}$ to label domain $\mathcal{Y}$, where each sample point is defined as $(\boldsymbol{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y}$. We use $g_i = \ell\left(\boldsymbol{\theta}; (\boldsymbol{x}_i, y_i)\right)$, where $\ell(\cdot;\cdot)$ is the training loss function and $\boldsymbol{\theta}$ is the parameters of the model, to denote the training loss on ith sample $(\boldsymbol{x}_i, y_i) \in \mathcal{T}$.

In order to address the aforementioned biases, we need to weight loss of samples from different classes or groups separately, hence, we define a weight vector, $\boldsymbol{w} \in \mathbb{R}_{+}^{c}$, where c is the number classes in the imbalanced data. Let $\mathbf{D} \in \{0,1\}^{n\times c}$ denote the assignments of n training samples to c classes. For a model parameter $\boldsymbol{\theta}$ and a fixed weight vector $\boldsymbol{w}$, we define the loss over training examples as

$$G(\boldsymbol{w}, \boldsymbol{\theta}; \mathcal{T}) = (\mathbf{D}\boldsymbol{w})^{\top} \boldsymbol{g}, \tag{1}$$

where $\boldsymbol{g} = [g_1, g_2, \ldots, g_n]^{\top}$ is the vector of losses over training samples. Equipped with Eq. (1) as the training goal, for a known weight vector $\boldsymbol{w}$ we can find the optimal parameters $\boldsymbol{\theta}$ by minimizing the objective. However, we use the samples in the target dataset to adaptively learn the optimal weight vector and guide the training process. To this end, we define the loss over a *small unbiased target dataset* $\mathcal{V}$ as:

$$F(\boldsymbol{w}, \boldsymbol{\theta}^*(\boldsymbol{w}); \mathcal{V}) = \frac{1}{|\mathcal{V}|} \sum_{(\boldsymbol{x}_i, y_i)\in\mathcal{V}} f(\boldsymbol{\theta}^*(\boldsymbol{w}); (\boldsymbol{x_i}, y_i)), \tag{2}$$

where $|\mathcal{V}|$ is the number of samples in $\mathcal{V}$, $\boldsymbol{\theta}^*(\boldsymbol{w})$ is the minimizer of the loss function in Eq. (1), and $f(\cdot;\cdot)$ is the target loss function which may or may not be same as training loss $\ell(\cdot;\cdot)$. We emphasize that target dataset could be a part of training dataset or it could be separated from it similar to the way that a validation dataset is generated.

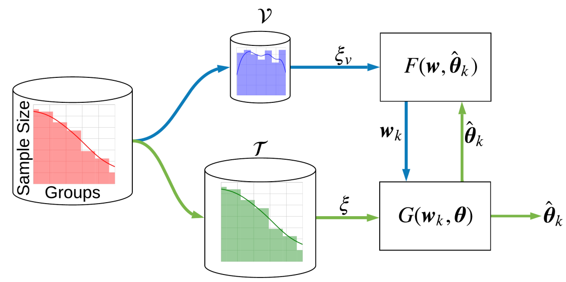

Figure 1. The proposed Targeted Meta-Learning framework.

It is worth noting that despite the similarity of this algorithm with a universal bilevel programming, the two levels are being optimized on different data distributions. This is the key difference that makes the targeted meta-learning framework capable of solving some challenging problems with a data-driven regularization using target distribution. The simple schematic of this framework is depicted in Fig. 1.

3. Case Study: Bow Echo Detection

3.1. Bow Echo

In real-world applications, the primary challenge is often to detect critical incidents in datasets. Notwithstanding their importance in the classification tasks, normally those critical incident samples are scarce in the dataset. Therefore, based on earlier discussions, a typical classifier would fail miserably in detecting these incidents. One of the conspicuous examples is severe weather detection using radar, satellite, and other sensor data. Severe weather conditions such as tornadoes, thunderstorms, and straight-line winds, are sporadic phenomena, but can be spotted in radar or satellite images with some specific patterns. One of these patterns, associated with severe weather conditions such as thunderstorms and straight-line winds, is called *bow echo*, because it has archer's bow shape in radar images as it is depicted in Fig. 2. The wind with a bow echo can be fierce and reach violent intensity. Detecting and predicting the formation of bow echoes, and thus, its related severe weather conditions, could help to prevent their detrimental consequences.

The term bow echo was coined by Fujita (1978), to describe strong outflow winds associated with storms that spread out in straight lines over the ground. Przybylinski (1995) categorize bow echoes in two categories, namely, bow echo patterns associated with derechoes or straight-line winds, and bow echo patterns associated with vortices, including tornadoes. Klimowski et al. (2004) classify different types of bow echoes and their evolution from meteorologists' point of view. Our previous studies (Kamani et al., 2016; 2018)

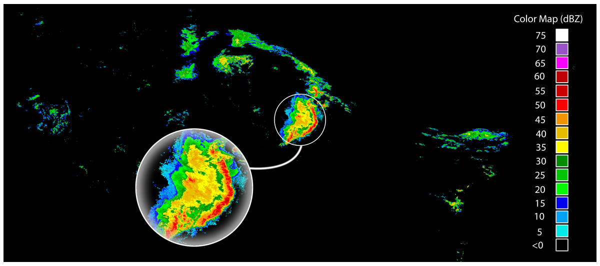

Figure 2. Radar image of the Continental United States with a bow echo, May 24, 2008, 10:20 GMT. We magnify the part that bow echo happened (*i.e.* red regions).

devoted to automate the detection process of these bow echo patterns using computer vision techniques followed by a classification stage. However, it needs some preprocessing stages, which could be time consuming.

3.2. Radar Images

Our dataset consists of images from NEXRAD level III radar of National Weather Service (technical name WSR-88D), which can measure precipitation and wind movement in the atmosphere. These images are gathered from 160 active high-resolution radar sites in the Continental United States. We use base reflectivity images from NEXRAD level III radar, which represent the amount of returned power to the radar from transmitted signal after hitting precipitation in the atmosphere. The images have 4-bit color map with $6,000 \times 2,600$ pixels of spatial resolution, which are stored every five minutes.[1] That is, in a whole year there are more than $105K$ images with mentioned quality. The colormap associated to these radar images (shown in Fig. 2) has the range from 0 dBZ to 75 dBZ for reflectivity. The range of the reflectivity from 0 dBZ to -30 dBZ, alongside with "No Data" regions (due to spots with beam blockage) is represented by a black color. Bow echoes can be spotted in heavy precipitation red regions on radar images (*i.e.*, with reflectivity higher than 50 dBZ).

3.3. Experimental Results

We use NEXRAD level III radar data in order to create our dataset of radar images for a whole year of 2008 gathered from 160 high resolution radars across the Continental United States. We will test the model on a balanced set of bow echo and non-bow echo samples from the year 2009. The year 2008 is chosen for training because of high number of severe weather activities in that year. Despite the huge number of images each year, number of images with a bow echo sample on it is very limited. For instance, in the year 2008 we only have $1,821$ images from 81 different instances that are labeled as bow echo samples. Therefore, this dataset, similar to other severe weather detection and prediction datasets, is greatly imbalanced with class size ratio of 0.017. The data distribution is immensely skewed toward normal data points, as it is the case for most critical incident detection applications. Thus, we apply targeted meta-learning framework on this dataset to overcome the imbalance problem.

[1] http://mesonet.agron.iastate.edu/docs/nexrad_composites/

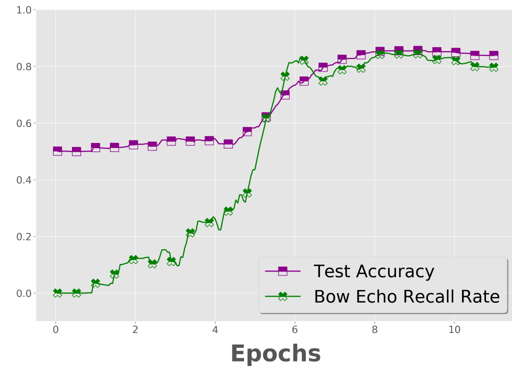

Figure 3. Accuracy and recall rate on balanced test dataset after 11 epochs of training. The training dataset contains the complete radar images from year 2008 with class size ratio of 0.017, while the test set is a balanced dataset of bow echo and non bow echo samples from year 2009. Test accuracy and recall rate on bow echo samples reach to 0.8605 and 0.855, respectively.

For this dataset, we apply targeted meta-learning on ResNet20 model (He et al., 2016), with image size of 52×180. The target distribution is a balanced dataset of both bow echo and non-bow echo samples from year 2008 with 273 samples that have 137 bow echo samples. The balanced test set contains $3,524$ images from year 2009, which has $1,762$ bow echo samples. In this experiment, we set the training batch size to 50, and target batch size to 10, with learning rates of 0.001 and 0.2 for main and target learning, respectively. The result of this training after 11 epochs in Fig. 3 reveals that targeted meta-learning has an exceptional capacity in addressing biases in these kinds of problems.

4. Conclusions

To alleviate the deficiency of conventional deep learning models on imbalanced real-world weather prediction datasets, we advocate the use of a small unbiased target dataset in a bilevel fashion as a data-driven regularizer for the main training with biased datasets. Our targeted meta-learning utilizes this target dataset to learn the weight of each designated class or category in the training process using a bilevel program. We empirically show the efficacy of this framework in dealing with imbalanced data problem in bow echo dataset.

Acknowledgments. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation under grant number ACI-1548562. We thank Stephen Wistar of AccuWeather, Inc. for valuable discussions during our earlier research on bow echoes.

References

Andrychowicz, M., Denil, M., Gomez, S., Hoffman, M. W., Pfau, D., Schaul, T., Shillingford, B., and De Freitas, N. Learning to learn by gradient descent by gradient descent. In *Advances in Neural Information Processing Systems*, pp. 3981–3989, 2016.

Bengio, S. Sharing representations for long tail computer vision problems. In *ACM International Conference on Multimodal Interaction*, pp. 1. ACM, 2015.

Bengio, Y., Louradour, J., Collobert, R., and Weston, J. Curriculum learning. In *International Conference on Machine Learning*, pp. 41–48. ACM, 2009.

Buda, M., Maki, A., and Mazurowski, M. A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249–259, 2018.

Cotter, A., Gupta, M., Jiang, H., Srebro, N., Sridharan, K., Wang, S., Woodworth, B., and You, S. Training well-generalizing classifiers for fairness metrics and other data-dependent constraints. *arXiv preprint arXiv:1807.00028*, 2018.

Cui, Y., Jia, M., Lin, T.-Y., Song, Y., and Belongie, S. Class-balanced loss based on effective number of samples. *arXiv preprint arXiv:1901.05555*, 2019.

Elkan, C. The foundations of cost-sensitive learning. In *International Joint Conference on Artificial Intelligence*, volume 17, pp. 973–978. Lawrence Erlbaum Associates Ltd, 2001.

Finn, C., Abbeel, P., and Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning-Volume 70*, pp. 1126–1135. JMLR.org, 2017.

Freund, Y. and Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

Fujita, T. T. *Manual of downburst identification for project NIMROD*. Satellite and Mesometeorology Research Project, Department of the Geophysical Sciences, University of Chicago, 1978.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Huang, C., Li, Y., Change Loy, C., and Tang, X. Learning deep representation for imbalanced classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5375–5384, 2016.

Jiang, L., Zhou, Z., Leung, T., Li, L.-J., and Fei-Fei, L. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *International Conference on Machine Learning*, pp. 2309–2318, 2018.

Kamani, M. M., Farhat, F., Wistar, S., and Wang, J. Z. Shape matching using skeleton context for automated bow echo detection. In *IEEE International Conference on Big Data*, pp. 901–908, 2016.

Kamani, M. M., Farhat, F., Wistar, S., and Wang, J. Z. Skeleton matching with applications in severe weather detection. *Applied Soft Computing*, 70:1154–1166, 2018.

Kendall, M. G. et al. *Advanced Theory of Statistics*. Charles Griffin: London, 1946.

Khan, S. H., Hayat, M., Bennamoun, M., Sohel, F. A., and Togneri, R. Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8):3573–3587, 2018.

Klimowski, B. A., Hjelmfelt, M. R., and Bunkers, M. J. Radar observations of the early evolution of bow echoes. *Weather and Forecasting*, 19(4):727–734, 2004.

Ouyang, W., Wang, X., Zhang, C., and Yang, X. Factors in finetuning deep model for object detection with long-tail distribution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 864–873, 2016.

Przybylinski, R. W. The bow echo: Observations, numerical simulations, and severe weather detection methods. *Weather and Forecasting*, 10(2):203–218, 1995.

Ren, M., Zeng, W., Yang, B., and Urtasun, R. Learning to reweight examples for robust deep learning. *arXiv preprint arXiv:1803.09050*, 2018.

Ting, K. M. A comparative study of cost-sensitive boosting algorithms. In *International Conference on Machine Learning*, pp. 983–990, 2000.