MetaMix: Improved Meta-Learning with Interpolation-based Consistency Regularization

Yangbin Chen¹, Yun Ma², Tom Ko³, Jianping Wang¹, Qing Li²

1. City University of Hong Kong 2. The Hong Kong Polytechnic University 3. Southern University of Science and Technology

BACKGROUND

- Few-Shot Learning (FSL) problem is a machine learning problem that learns with limited labelled data of the target tasks by incorporating external source data with a different distribution.
- Few-Shot Classification is a few-shot learning task defined as N-way, K-shot, where N is the number of classes in the target task and K is the number of labelled examples per class.
- Model-Agnostic Meta-Learning (MAML) and its variants aim to train a model, which can adapt quickly to any new tasks using only a few examples.

MOTIVATION

- Conventional meta-learning algorithms face meta-overfitting problems, where the learned decision boundary stays too close to the limited labelled examples in few-shot classification tasks.
- The Empirical Risk Minimization (ERM) allows large neural networks to memorize (rather than generalize from) the training data.
- We aim to propose a regularization technique to solve the meta-overfitting problem.

METHODOLOGY

![Diagram of MetaMix algorithm]

Algorithm 1 MetaMix with MAML

Require: \( p(T) \) : distribution over tasks
Require: \( S_t \) : support set; \( Q_t \) : query set
Require: \( \alpha, \beta \) : learning rate
Require: \( \alpha \) : Beta distribution parameter
Require: \( \text{mix}_A(a,b) = \alpha a + (1 - \alpha)b, \lambda \sim \text{B}(\alpha, \alpha) \)

1. Randomly initialize model parameters \( \theta \)
2. while not done do
3. Sample a batch of episodes \( T_t \sim p(T) \)
4. for all \( T_t \) do
5. Sample a support set \( S_t = \{(x_j, y_j)\}_{j=1}^T \)
6. Evaluate \( \nabla_{\theta} \mathcal{L}_{S_t}(f_{\theta}) \) using \( S_t \) and \( \mathcal{L}_{S_t}(f_{\theta}) \)
7. Compute adapted parameters with gradient descent:
8. \( \theta_t^\prime = \theta - \alpha \cdot \nabla_{\theta} \mathcal{L}_{S_t}(f_{\theta}) \)
9. Sample a query set \( Q_t = \{(x_z, y_z)\}_{z=1}^Z \)
10. Randomly select pairs of examples \( \{(x_m, y_m)\}_{m=1}^M, \{(x_n, y_n)\}_{n=1}^N \) from \( Q_t \)
11. Get new query set \( Q_t = \{(x_z, y_z)\}_{z=1}^Z \)
12. end for
13. Update \( \theta \leftarrow \theta - \beta \cdot \nabla_{\theta} \sum_{Q_t} \mathcal{L}_{Q_t}(f_{\theta}) \)
14. end while

EXPERIMENT

- Performance comparison of MetaMix and baseline approaches on 5-way classification tasks over three datasets

<table>
<thead>
<tr>
<th>Models</th>
<th>mini-ImageNet 1-shot</th>
<th>CUB 5-shot</th>
<th>FC100 5-shot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching Network</td>
<td>50.47 ± 0.80</td>
<td>64.83 ± 0.67</td>
<td>57.70 ± 0.87</td>
</tr>
<tr>
<td>Prototypical Network</td>
<td>49.33 ± 0.82</td>
<td>65.71 ± 0.76</td>
<td>51.34 ± 0.86</td>
</tr>
<tr>
<td>Relation Network</td>
<td>50.48 ± 0.80</td>
<td>65.39 ± 0.72</td>
<td>59.47 ± 0.96</td>
</tr>
<tr>
<td>MAML</td>
<td>48.18 ± 0.78</td>
<td>63.05 ± 0.71</td>
<td>54.32 ± 0.91</td>
</tr>
<tr>
<td>MetaMix+MAML</td>
<td>50.51 ± 0.86</td>
<td>65.73 ± 0.72</td>
<td>57.70 ± 0.92</td>
</tr>
<tr>
<td>FOMAML</td>
<td>45.22 ± 0.77</td>
<td>60.97 ± 0.70</td>
<td>53.12 ± 0.93</td>
</tr>
<tr>
<td>MetaMix+FOMAML</td>
<td>47.78 ± 0.77</td>
<td>63.55 ± 0.70</td>
<td>54.81 ± 0.97</td>
</tr>
<tr>
<td>MetaSGD</td>
<td>49.93 ± 1.73</td>
<td>64.01 ± 0.90</td>
<td>56.19 ± 0.92</td>
</tr>
<tr>
<td>MetaMix+MetaSGD</td>
<td>50.60 ± 1.80</td>
<td>64.47 ± 0.88</td>
<td>57.64 ± 0.88</td>
</tr>
<tr>
<td>MTL</td>
<td>61.37 ± 0.82</td>
<td>78.37 ± 0.60</td>
<td>71.90 ± 0.86</td>
</tr>
<tr>
<td>MetaMix+MTL</td>
<td>62.74 ± 0.82</td>
<td>79.11 ± 0.58</td>
<td>73.04 ± 0.86</td>
</tr>
</tbody>
</table>
**CONTRIBUTIONS**

- We propose MetaMix as a regularization technique, which can be integrated with many meta-learning algorithms, including MAML and its variants, and improve their performance.
- MetaMix with MAML-based algorithms perform more robust with the reduction of training data, compared with original MAML-based algorithms.
- MetaMix with MTL achieves state-of-the-art performance.