Adversarial Meta Sampling for Multilingual Low-Resource Speech Recognition

Yubei Xiao1, Ke Gong2, Pan Zhou3, Guolin Zheng1, Xiaodan Liang1,2, Liang Lin1,2*
1Sun Yat-sen University, 2Dark Matter AI Research, 3SalesForce

Motivations

• Low-resource automatic speech recognition (ASR) is challenging, as the low-resource target language cannot well train an ASR model.
• For different source languages, the quantity and difficulty vary greatly because of their different data scales and diverse phonological systems, which leads to task-quantity and task-difficulty imbalance issues. So sampling approaches become especially important in ASR.
• Existing low-resource ASR approaches such as multilingual transfer learning ASR (MTL-ASR) and multilingual meta-learning ASR (MML-ASR) often ignore the task imbalance issues which could result in unsatisfactory performance.

Experiments

Dataset
We validate our method on two public multilingual speech recognition datasets: Mozilla Common Voice Corpus (Mozilla.org 2019) and the IARPA BABEL dataset (Gales et al. 2014) and we also conduct experiments on speech classification dataset in AutoSpeech 2020 competition (InterSpeech 2020) and multilingual speech translation corpus CoVoST (Wang et al. 2020) to demonstrate the applicability of AMS to other low-resource speech tasks.

Results
The experimental results demonstrate that our AMS significantly improves the performance over the existing approaches on low-resource ASR, especially under the realistic task-imbalance scenarios and shows its great generalization capacity in other low-resource speech tasks.

Methodology

Preliminaries: Multilingual Meta-learning ASR
• We train a multilingual meta-learning ASR (MML-ASR) model (Hsu, Chen, and Yi Lee 2020) on all languages to pursue the few-shot learning ability to handle the low-resource recognition problems. MML-ASR can be formulated as:

\[
\min \ E_{T \sim \mathcal{D}_{\text{Train}}(T)} \mathcal{L}_{\text{Sup}}(\theta) - \alpha \nabla \mathcal{L}_D(\mathcal{D}_{\text{Sup}}(\theta))
\]

Adversarial Meta Sampling
• We propose a novel and effective adversarial meta sampling (AMS) approach that adaptively determines the sampling probability for each language task set in the meta-training process to balance both task quantity and difficulty in different language domains.
• The query losses of tasks from each language domain can well measure both task-quantity imbalance and task-difficulty imbalance. So, we design a policy network to increase the query loss of MML-ASR model through adversarial learning for sampling from proper language domain shown in (a).
• At each meta-training iteration, our policy network predict the most befitting task sampling probability for each language domain to form training task set for meta-training of MML-ASR model. So the meta-objective of MML-ASR model can be reformulated as:

\[
\min \ E_{T \sim \mathcal{D}_{\text{Train}}(T)} \mathcal{L}_{\text{Sup}}(\theta) - \alpha \nabla \mathcal{L}_D(\mathcal{D}_{\text{Sup}}(\theta))
\]

Ablation Study

Method | Keyword | Evaluation
--- | --- | ---
MML-ASR (Replica) (Nichol, Azuma, and Schulman 2018) | 68.61 | 33.17
MML-ASR (PSMAML) (Hsu, Chen, and Yi Lee 2020) | 59.38 | 37.64
MML-ASR-MAML (Hsu, Chen, and Yi Lee 2020) | 56.29 | 36.46
PPQ-MAML (Dua, Yu, and Anagnostopoulos 2019) | 58.95 | 72.30
PPQA-MAML (Sun et al. 2019a) | 54.87 | 74.97
PPQAQ-MAML | 55.14 | 75.41
PPQ-MAML | 53.15 | 73.33
our MAML-MAML (w/o attention) | 54.16 | 74.29
our AMS-MAML | 59.30 | 76.49
our AMS-POMAML | 53.04 | 73.73
our AMS-MAML | 58.72 | 72.34
our AMS-MAML (80% target) | 59.02 | 73.87
our AMS-MAML (50% target) | 70.27 | 81.97
our AMS-MAML (30% target) | 87.11 | 91.72

Codes & Contact
• Codes: https://github.com/iamxiaoyubei/AMS
• Contact: xiaoyb5@mail2.sysu.edu.cn