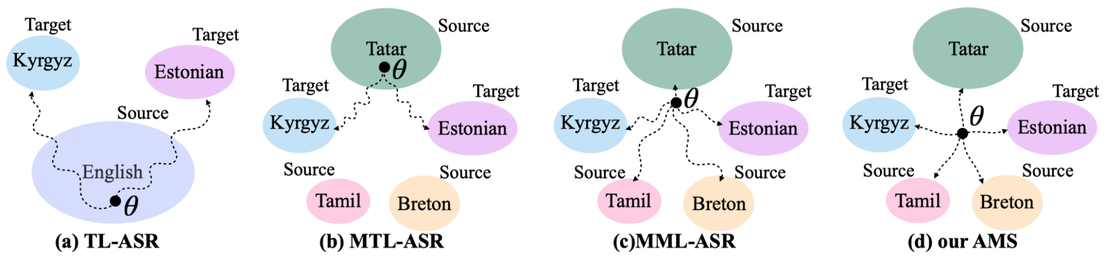
国际人工智能会议
AAAI 2021 论文北京预讲会

Adversarial Meta Sampling for Multilingual Low-Resource Speech Recognition

Yubei Xiao[1], Ke Gong[2], Pan Zhou[3], Guolin Zheng[1], Xiaodan Liang[1,2], Liang Lin[1,2*]
[1] Sun Yat-sen University, [2] Dark Matter AI Research, [3] SalesForce

Motivations



- Low-resource automatic speech recognition (ASR) is challenging, as the low-resource target language data 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.

Target	Kyrgyz	Estonian	Spanish		Dutch		Kabyle	
Source	*Diversity11*		*Indo9*	*Indo12*	*Indo9*	*Indo12*	*Indo9*	*Indo12*
Monolingual training (Hori et al. 2017)	76.25	86.04	80.30		68.71		85.41	
TL-ASR (Kunze et al. 2017)	68.28	82.04	79.39		56.58		89.12	
MTL-ASR (multi-head) (Dalmia et al. 2018)	67.56	81.50	75.82	73.00	57.80	56.41	82.10	81.23
MTL-ASR (Watanabe, Hori, and Hershey 2017)	64.90	83.70	73.85	71.40	62.55	58.46	84.25	81.88
our AMS (MTL-ASR)	**59.55**	**79.33**	**71.02**	**68.97**	**58.22**	**54.96**	**83.90**	**79.26**
MML-ASR (Hsu, Chen, and yi Lee 2020)	58.29	79.66	66.75	65.24	53.33	52.56	79.45	75.96
our AMS (MML-ASR)	**50.72**	**72.26**	**65.21**	**64.40**	**51.18**	**49.13**	**78.21**	**73.69**

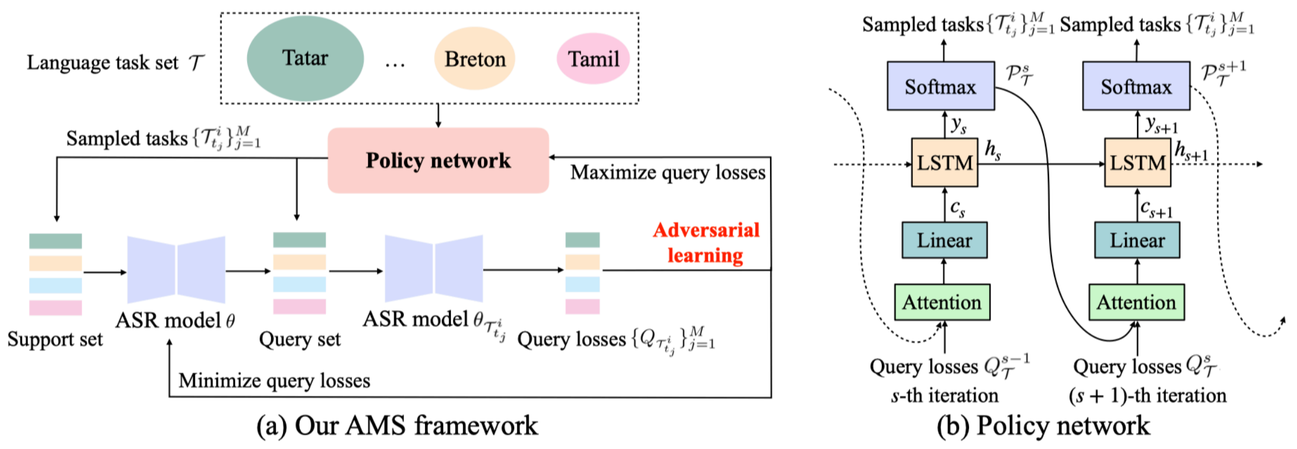
➢ *Ablation Study*

Method	Kyrgyz	Estonian
MML-ASR (Reptile) (Nichol, Achiam, and Schulman 2018)	66.51	83.17
MML-ASR (FOMAML) (Hsu, Chen, and yi Lee 2020)	59.23	78.64
MML-ASR (MAML) (Uniform) (Hsu, Chen, and yi Lee 2020)	58.29	79.66
PPQ-MAML (Dou, Yu, and Anastasopoulos 2019)	58.95	77.26
PPQL-MAML (Sun et al. 2018a)	54.87	74.97
PPEAQL-MAML	55.14	75.41
PPAQL-MAML	53.15	73.33
our AMS-MAML w/o attention	54.16	74.29
our AMS-Reptile	59.30	78.49
our AMS-FOMAML	53.04	74.77
our AMS-MAML	**50.72**	**72.26**
our AMS-MAML (80% target)	59.02	75.87
our AMS-MAML (50% target)	70.27	81.97
our AMS-MAML (20% target)	87.11	91.72

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: $\min_\theta \ \mathbb{E}_{\mathcal{T}_i\sim\mathcal{T}}\mathcal{L}_{D_{\text{query}}}(\theta-\alpha\nabla_\theta\mathcal{L}_{D_{\text{support}}}(\theta))$



➢ *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:

$$\min_\theta \ \mathbb{E}_{\pi\sim f_\phi}\mathbb{E}_{\mathcal{T}_i\sim\pi(\mathcal{T})}\mathcal{L}_{D_{\text{query}}}(\theta-\alpha\nabla_\theta\mathcal{L}_{D_{\text{support}}}(\theta))$$

- Our policy network attempts to increase the query loss of MML-ASR model through adversarial learning for sampling the proper language task set for training. Formally, the objective loss of our policy network is defined as:

$$\phi^*=\arg\max_\phi \mathcal{J}(\phi), \text{where}\mathcal{J}(\phi)=$$
$$\mathbb{E}_\pi\sim f_\phi\mathbb{E}_{\mathcal{T}_i\sim\pi(\mathcal{T})}\mathcal{L}_{D_{\text{query}}}(\theta-\alpha\nabla_\theta\mathcal{L}_{D_{\text{support}}}(\theta))$$

- The policy network injects attention mechanism into LSTM to utilize the long-term information in LSTM and the current query losses at each training iteration shown in (b).
- The policy network can be jointly trained with MML-ASR model in an end-to-end way when applying REINFORCE algorithm (Williams 1992) to solve the issue of non-differentiable sampling operation and optimize the policy network via the following gradient,

$$\nabla_\phi\mathcal{J}(\phi)=\nabla_\phi\mathbb{E}_{\pi\sim f_\phi}\mathbb{E}_{\mathcal{T}_i\sim\pi(\mathcal{T})}\mathcal{L}_{D_{\text{query}}}(\theta-\alpha\nabla_\theta\mathcal{L}_{D_{\text{support}}}(\theta))$$
$$\approx\nabla_\phi\sum_{i=1}^M\mathcal{P}_{\mathcal{T}_i}\mathcal{L}_{D_{\text{query}}}(\theta-\alpha\nabla_\theta\mathcal{L}_{D_{\text{support}}}(\theta)).$$

Codes & Contact

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

主办方：中国中文信息学会青年工作委员会
承办方：智源社区