Gradual Fine-Tuning for Low-Resource Domain Adaptation

Haoran Xu, Seth Ebner, Mahsa Yarmohammadi, Aaron Steven White[†], Benjamin Van Durme, Kenton Murray
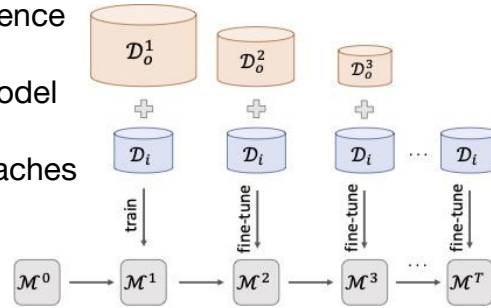
Johns Hopkins University, [†]University of Rochester

Introduction

Domain adaptation: Directly learn a model for a task in a particular domain with too few instances of in-domain data

Fine-tuning:
1) Pretrain a model on a large amount of out-of-domain but task-relevant data,
2) Further train the pretrained model on in-domain data until convergence

Gradual FT: Iteratively train a model to convergence on data whose distribution progressively approaches that of the in-domain data.

Method

```
 1: function GRADUAL-FT($\mathcal{D}_i, \mathcal{D}_o^0, \mathcal{M}^0, \mathcal{S}$)
 2:     $t \leftarrow 0$
 3:     for $amount$ in $\mathcal{S}$ do
 4:         $t \leftarrow t + 1$
 5:         $\mathcal{D}_o^t \leftarrow$ SAMPLE($\mathcal{D}_o^{t-1}, amount$)
 6:         $\mathcal{D}_{train}^t \leftarrow \mathcal{D}_i \cup \mathcal{D}_o^t$
 7:         $\mathcal{M}^t \leftarrow$ TRAIN($\mathcal{M}^{t-1}, \mathcal{D}_{train}^t$)
 8:     end for
 9:     return $\mathcal{M}^t$
10: end function
```

- Train a model on a sequence of datasets, each of which would be increasingly difficult to learn on its own due to its small size.

- At each stage, we increase the similarity between the current domain and the target domain, which enables the model to potentially better fit the distribution of the target domain.

Datasets

Dialogue State Tracking
MultiWOZ v2.0:
    N-1 → {Rest., Hotel}

Event Extraction
ACE 2005:
    English → Arabic

	Rest.	Hotel	Attract.	Taxi	Train
# Slots	7	10	3	4	6
# Turns	3011	3472	577	1667	1771
# Dialogues					
Train	523	513	127	326	282
Dev	50	56	11	57	30
Test	61	65	12	52	33

	English	Arabic
# Event types	33	30
# Role types	22	21
# Events/Arguments		
Train	4202/4859	1743/2506
Dev	450/605	117/174
Test	403/576	198/287

Results

Rest.

Hotel
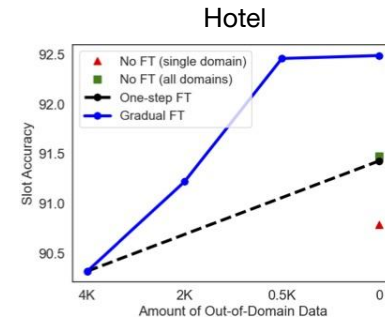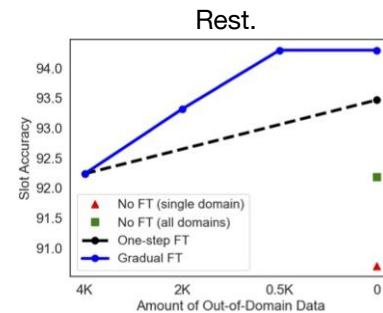
	TrigID	TrigC	ArgID	ArgC
No FT (Ar)	64.77	57.03	47.76	42.83
No FT (mixed)	64.12	59.48	46.57	43.21
One-stage FT	63.61	59.88	46.79	43.44
Gradual FT	**66.29**	**62.87**	**48.11**	**44.21**

Conclusions

Gradually fine-tuning in a multi-stage process can:
    - yield substantial gains over one-stage FT
    - be applied without modifying the model or objective