

Motivation

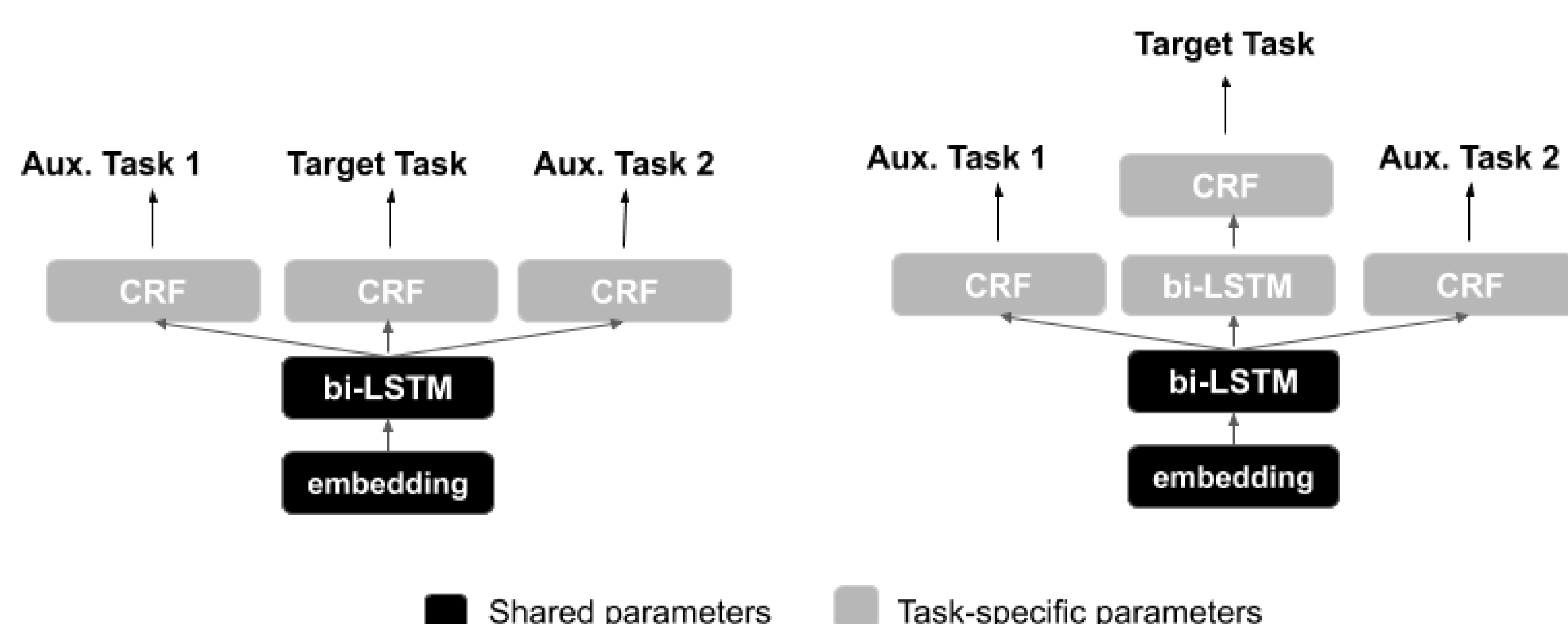
- **Problem:** Slot filling datasets used in conversational agents are **expensive** to obtain being task specific.
- **Idea:** Leverage a more **general, cheaper, and semantically related** task in a **multi-task learning (MTL)** setting.
- **Question:** Can **non-conversational resources** help the slot filling performance in **low resource settings**?

Multi-Task Learning: Target Task + Aux. Tasks

Sentence	what	is	the	most	expensive	flight	from	boston	to	dallas
ATIS Slot (Target Task)	O	O	O	B-COST_REL	I-COST_REL	O	O	B-FROM_LOC	O	B-TO_LOC
NER (Aux. Task)	O	O	O	O	O	O	O	B-GPE	O	B-GPE
SemTag (Aux. Task)	B-QUE	B-ENS	B-DEF	B-TOP	B-IST	B-CON	B-REL	B-GPE	O	B-GPE

- Named Entity (NER) often occurs as **slot values**. Semantic Tagging (Abzianidze and Bos, 2017) **complements** NER as its labels subsume NER labels.
- Previous work (Mesnil et al., 2013, 2015; Zhang and Wang, 2016; Gong et al., 2019) incorporate NER through the output of NER systems or ground-truth NER label as features. We learn these features from **disjoint datasets** through MTL.

Experiments and Results



MTL-Fully Shared Network (left) and Hierarchical-MTL (right) (Sogaard and Goldberg, 2016; Sanh et al., 2019)

Dataset	Task	#train	#dev	#test	#label
ATIS	Slot Filling	447	500	893	79
MIT Restaurant	Slot Filling	612	1532	3385	8
MIT Movie	Slot Filling	782	1955	2443	12
OntoNotes 5.0	NER	34970	5896	2327	18
PMB	SemTag	67965	682	650	73

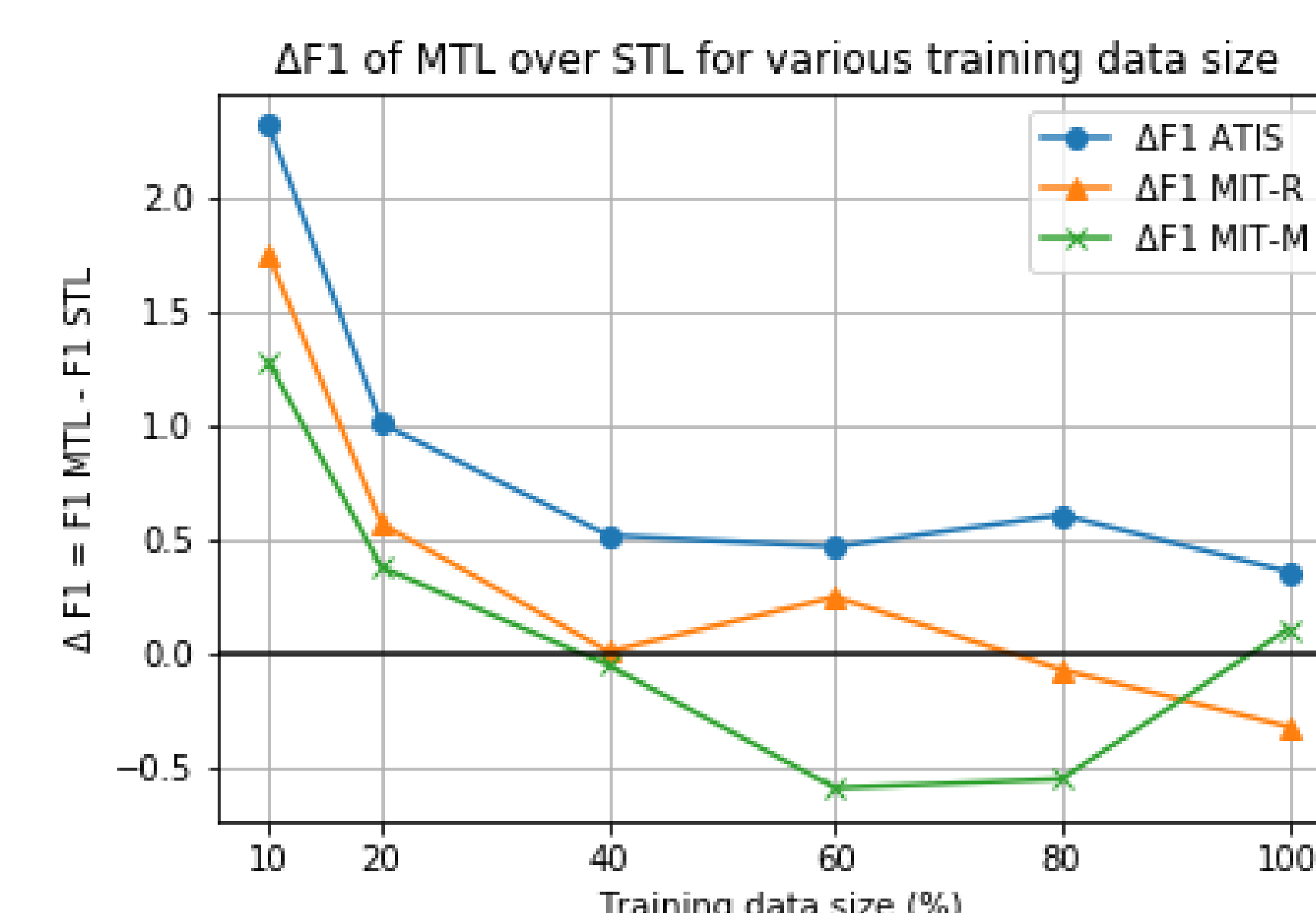
Statistics about the datasets, reporting the number of sentences in train/dev/test set, and the number of labels. The #train size for slot filling is 10% of the original training data size

Model	Aux. Task	Target Task		
		ATIS	MIT-R	MIT-M
STL	-	87.91 _{0.56}	67.37 _{0.26}	80.71 _{0.63}
STL+ FB	-	87.79 _{0.67}	67.27 _{0.64}	80.56 _{0.54}
MTL-FSN	NER	89.56 _{0.16}	68.82 _{0.18}	80.77 _{0.13}
	SemTag	89.19 _{0.26}	68.21 _{0.71}	80.57 _{0.32}
	NER, SemTag	89.10 _{0.41}	68.21 _{0.43}	79.69 _{0.33}
H-MTL	NER	89.17 _{0.33}	69.22 _{1.00}	81.79 _{0.26}
	SemTag	88.96 _{0.41}	69.09 _{0.24}	81.59 _{0.17}
	NER, SemTag	88.78 _{0.37}	68.96 _{0.50}	81.15 _{0.25}

- Leveraging NER and SemTag through MTL gives improvements over baselines for all target tasks.
- Supervising tasks with coarse-grained labels and fine-grained labels on different layers can be beneficial

Target Task	Concept	Model	
		STL	MTL
ATIS	LOC	94.74 _{0.37}	95.82 _{0.34}
	ORG	92.52 _{0.89}	93.37 _{0.29}
MIT-R	LOC	75.29 _{0.46}	76.02 _{0.39}
MIT-M	PER	85.04 _{0.24}	84.58 _{0.56}

- MTL improves the performance of slots related to coarse-grained concepts.



- MTL is more useful in very low resource scenarios

Conclusion & Future Work

- Using **NER and Semantic Tagging as auxiliary tasks** in a multi-task learning setting can improve low resource slot filling.
- **Non-conversational resources** have the potential to help low resource slot filling through **transfer learning**.
- Ongoing & Future Work: Data Selection for MTL, Data Augmentation