Efficient Multi-Task Auxiliary Learning: Selecting Auxiliary Data by Feature Similarity

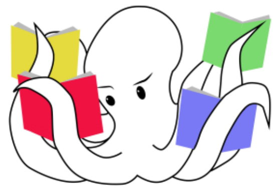
Po-Nien Kung, Yi-Cheng Chen, Sheng-Siang Yin, Tse-Hsuan Yang, Yun-Nung (Vivian) Chen

國立臺灣大學 National Taiwan University

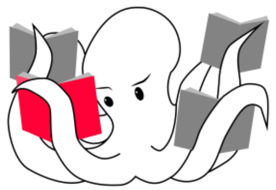
Summary: This paper proposes a **feature similarity**-based approach to **select beneficial auxiliary data** to fasten multi-task auxiliary learning.

1. Background

Multi-Task Learning V.S. **Auxiliary Learning**



All tasks are important!



Using auxiliary tasks improves the primary task.

More tasks (data), more computing.

Treating RTE as the primary task:

MT-DNN setting → 400x computing cost

Muppet setting → 2000x computing cost

Q: Should we use all auxiliary data?

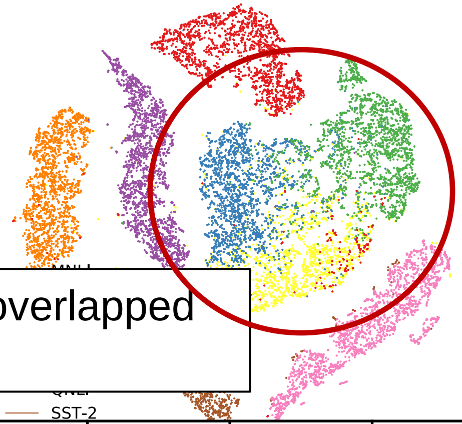
Using all auxiliary data is **time-consuming**.

Some auxiliary data might be **useless** or even **harmful**!

Q: How to select the most beneficial data?

A: Feature Similarity!

more similar feature
↓
more beneficial



RTE, **MRPC**, and **STS-B** more overlapped
→ more benefit from MTL!

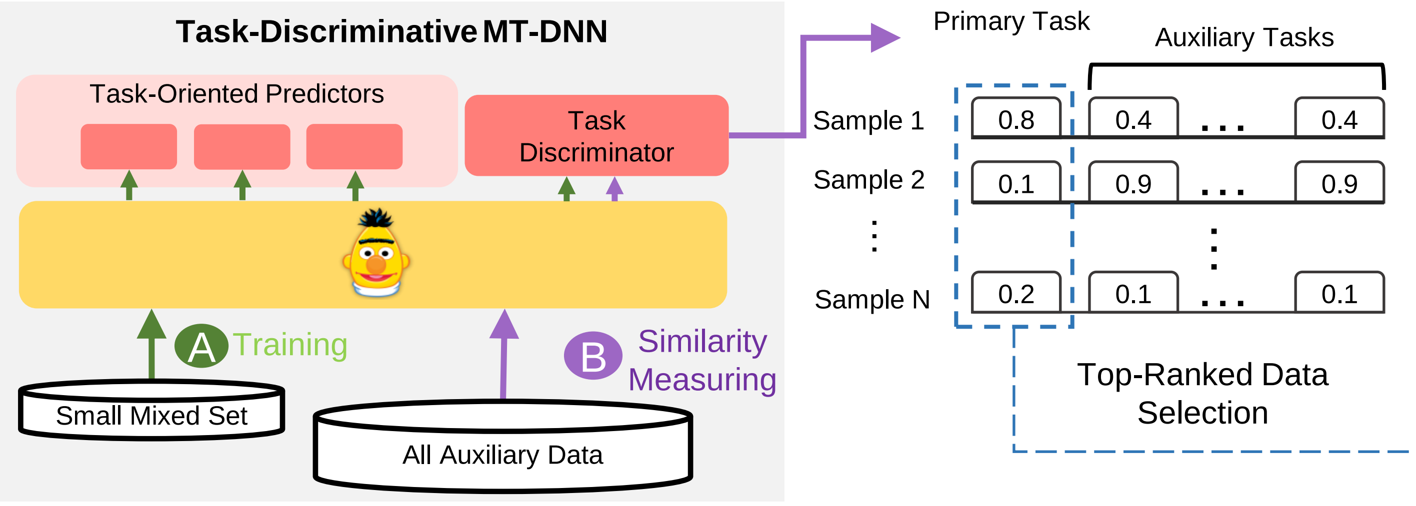
MNLI	RTE	MRPC	STS-B	QQP	QNLI	SST-2	CoLA

2. Two-Stage Approach

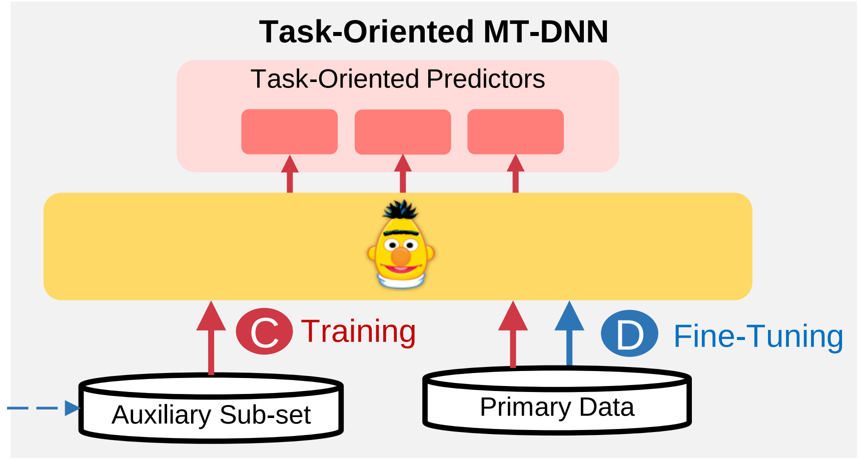
Stage 1: Train a proxy MT-DNN along with a **task discriminator** with small data and **predict** the similarity.

Stage 2: Use the auxiliary subset with highest similarity scores in the MT-DNN framework

Stage 1: Similarity Ranking

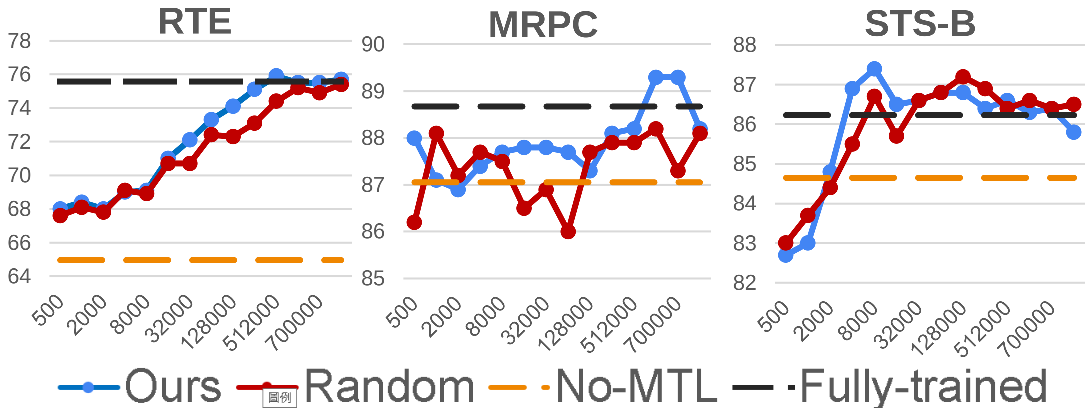


Stage 2: Multi-Task Auxiliary Learning & Fine-tuning



3. Experiments

Data: three tasks from GLUE (benefit from MTL)



Our method can use **less data** to achieve **better results**, and is much **faster** than training with full data!

Runtime(s) STS-B	Similarity Sampling		Auxiliary MTL		Total	Speed x
	Training a small proxy model	Predict similarity	MTL	Fine-tuning		
Fully-trained	--		15801	190	15991	--
Ours	95	775	260	190	1320	12x

Contributions

- Address the **efficiency issue** in multi-task auxiliary learning
- Propose **data sampling** to shrink auxiliary data size → computing cost reduction
- First use **feature similarity** to determine data usefulness