ADL Final
Domain Adaptive QA

R11944074 汪宣甫
R12922104 蔡昀燁
R12922051 陳韋傑
R12922062 鄭雅勻
R12944062 李泓賢

Abstract

This work focuses on exploring domain adaptation in Question Answering (QA):

- We use two different backbone models (Vanilla & Context Network) and two types of embedding paired with Context Network (BERT-base & SimCSE)

- We experiment three different training strategies: In-domain, Out-domain and Two-stage finetuning (Pre-training on out-of-domain knowledge and fine-tuning on in-domain datasets)

- We conclude that pretraining on out-domain datasets and finetuning on in-domain dataset can enhance performance hugely

Introduction

While Language models excel in performance during finetuning and testing within a specific dataset, their generality diminishes when applied to other similar datasets. [1]

Q: Can we find a strategy that solves the task for domain adaptive QA well?

Fine-tuned on	Evaluated on: SQuAD	TriviaQA	NQ	QuAC	NewsQA
SQuAD	**75.6**	46.7	48.7	20.2	41.1
TriviaQA	49.8	**58.7**	42.1	20.4	10.5
NQ	53.5	46.3	**73.5**	21.6	24.7
QuAC	39.4	33.1	33.8	**33.3**	13.8
NewsQA	52.1	38.4	41.7	20.4	**60.1**

Method - Training Strategy

IN: Dt

OUT: D1, D2, D3...

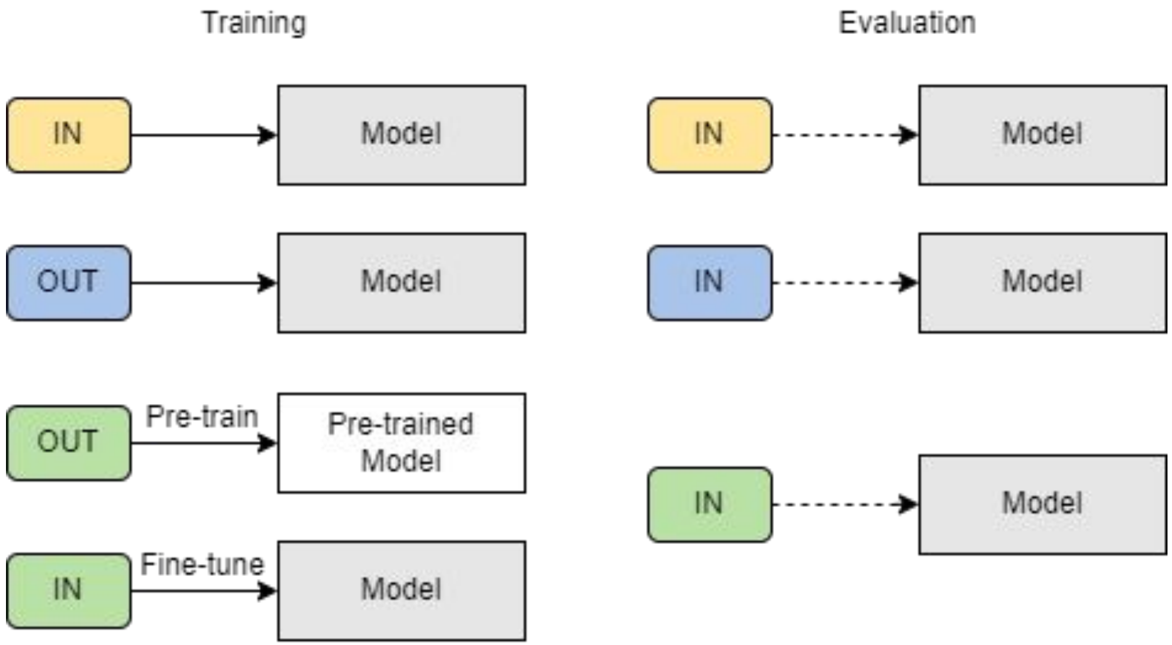
In-Domain (ID)

Out-Domain (OD)

Two Stage:

OD-Pretraining,

ID-Finetuning



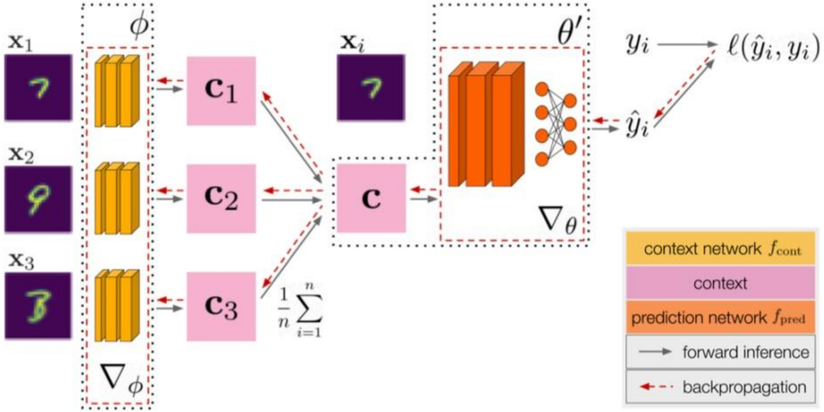
Method - Model

We have three different model settings: origin, cls and sim.

--model type	Backbone-Model Type	Embedding Type
origin	Vanilla	BERT-based
cls	Context Network	BERT-based
sim	Context Network	SimCSE

Method - Context Network

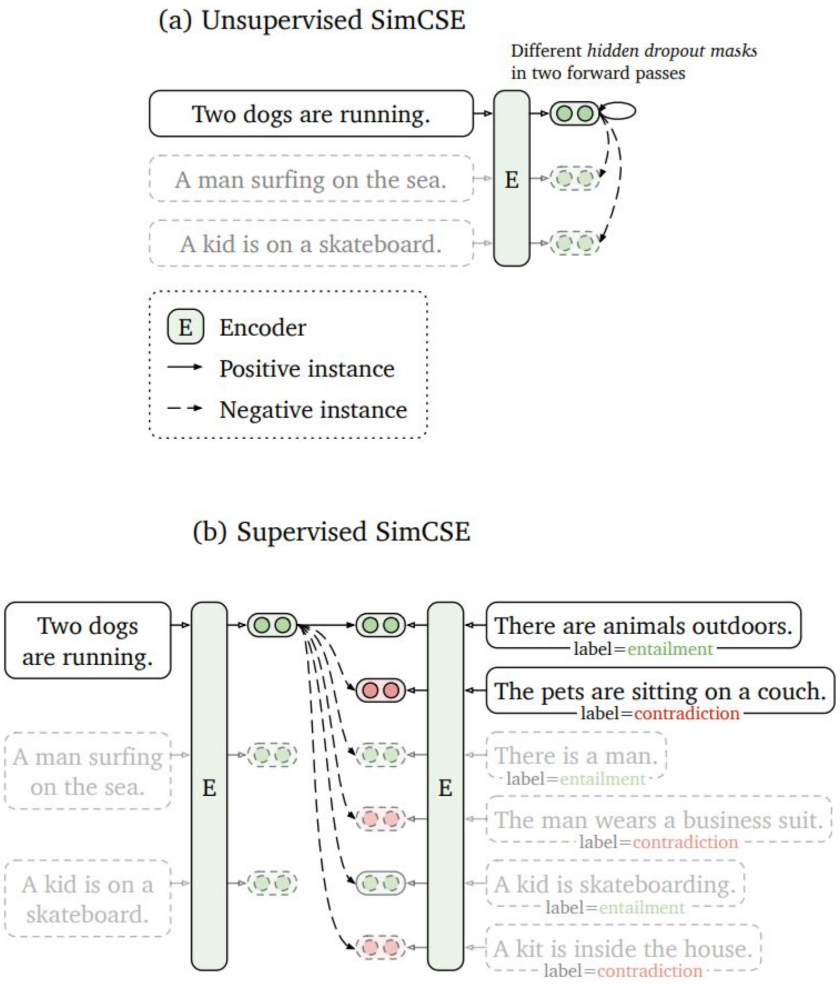
By incorporating context, the network can improve its robustness and generalization capabilities of the overall model. [3]



Method - SimCSE

SimCSE [2] directly targets sentence-level representations using the technique of contrastive learning.

It is an approach specifically designed for learning fixed-size embedding for the entire sentence.



Dataset

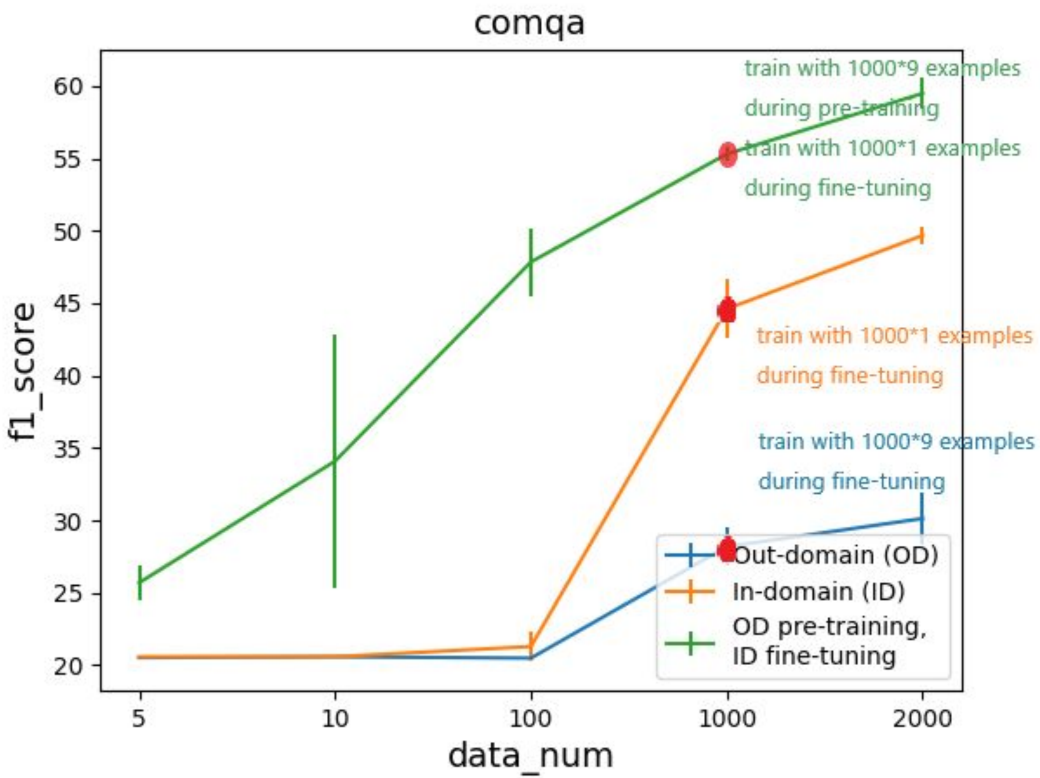
- We use 10 question answering datasets.
- Note that some questions do not have answers.

	train_num	eval_num
comqa	7416	1069
duorc_p	69524	15591
duorc_s	60721	12961
hotpotqa	90447	7405
newsqa	92549	5166
nq	110865	3369
quac	83568	7354
squad	130319	11873
triviaqa	110647	14229
wikihop	43748	5128

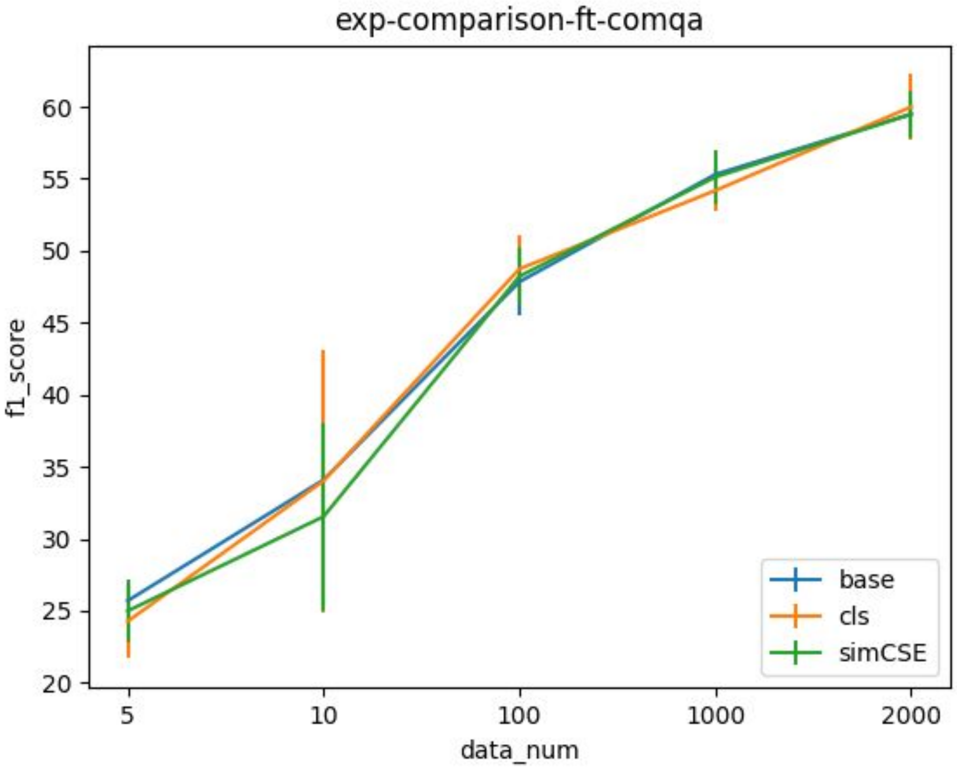
Experiment settings

- `data_num` and domain types decides training samples number of each corpus. If dataset size < training samples number, then we take **whole** dataset for training.
- Three different domain types (in-domain, out-domain, ODID).
- What is in-domain? out-domain? ODID?

Explain of data_num



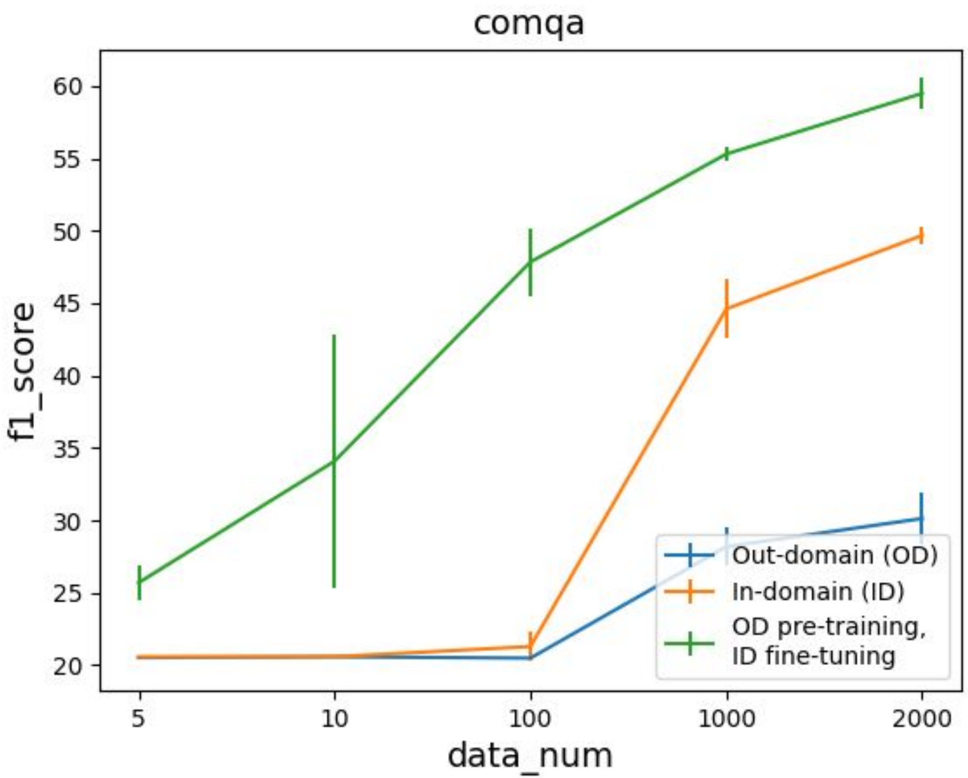
How different model perform **ODID** like?



How about in-domain and out-domain?

- Result 1:

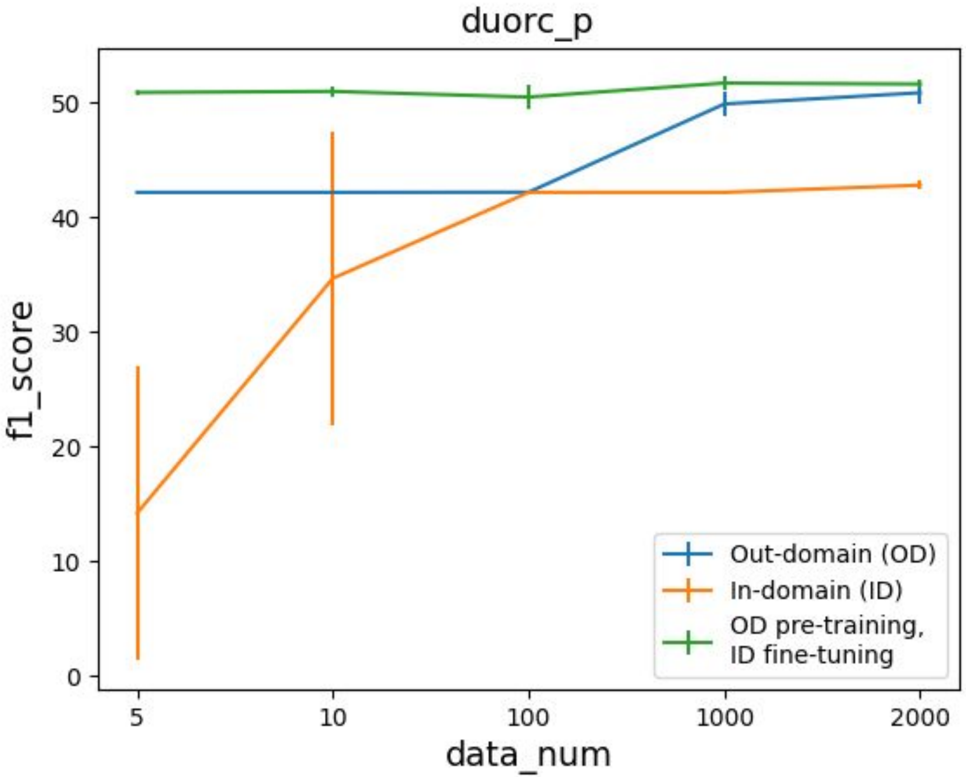
(in-domain > out-domain)



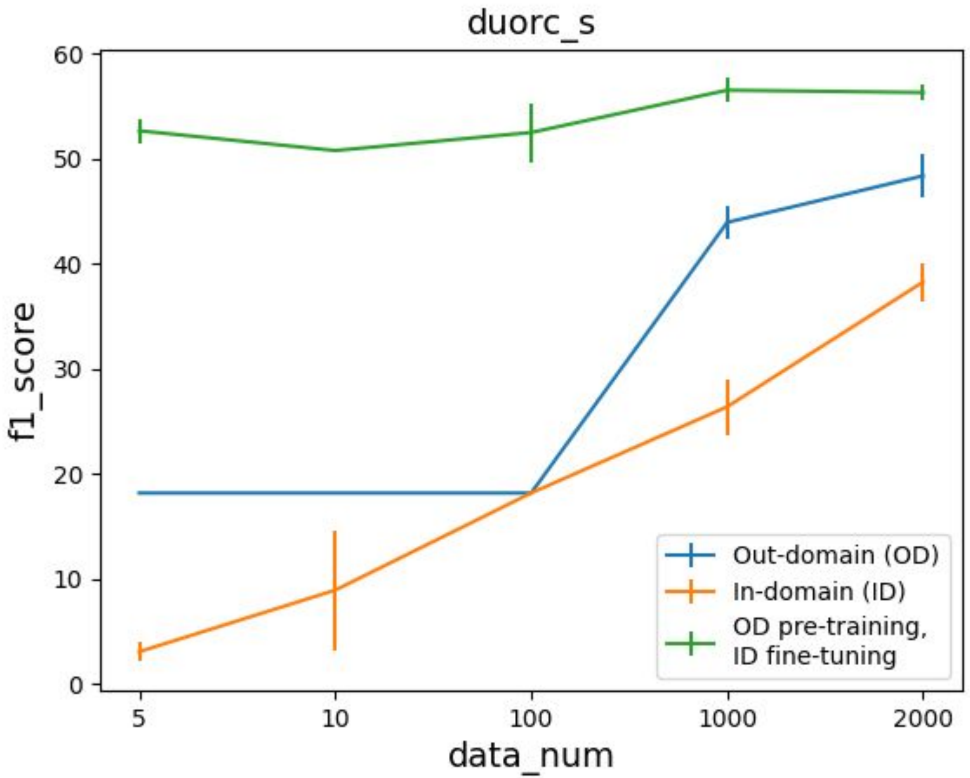
How about in-domain and out-domain?

- Result 2:

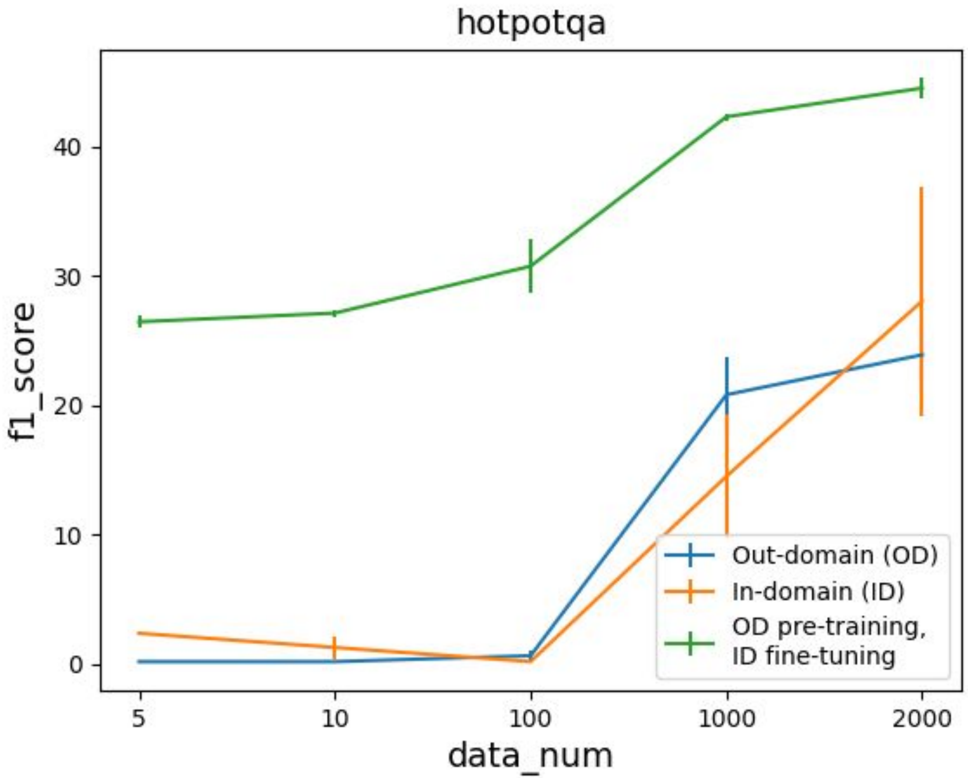
(out-domain > in-domain)



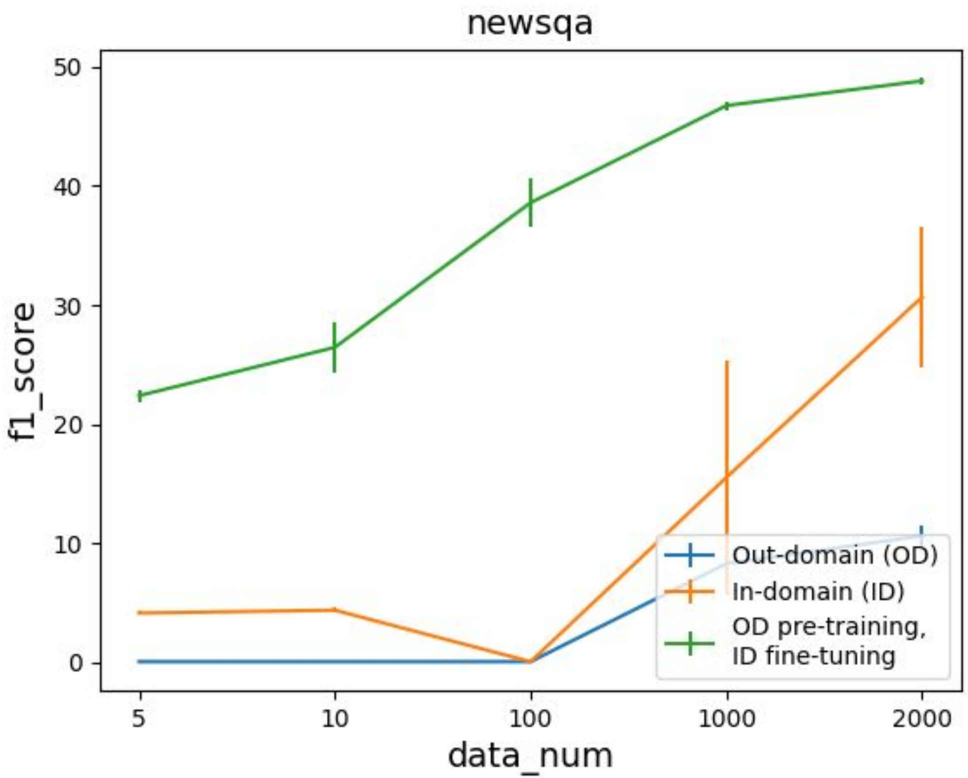
Experiments - duorc_s



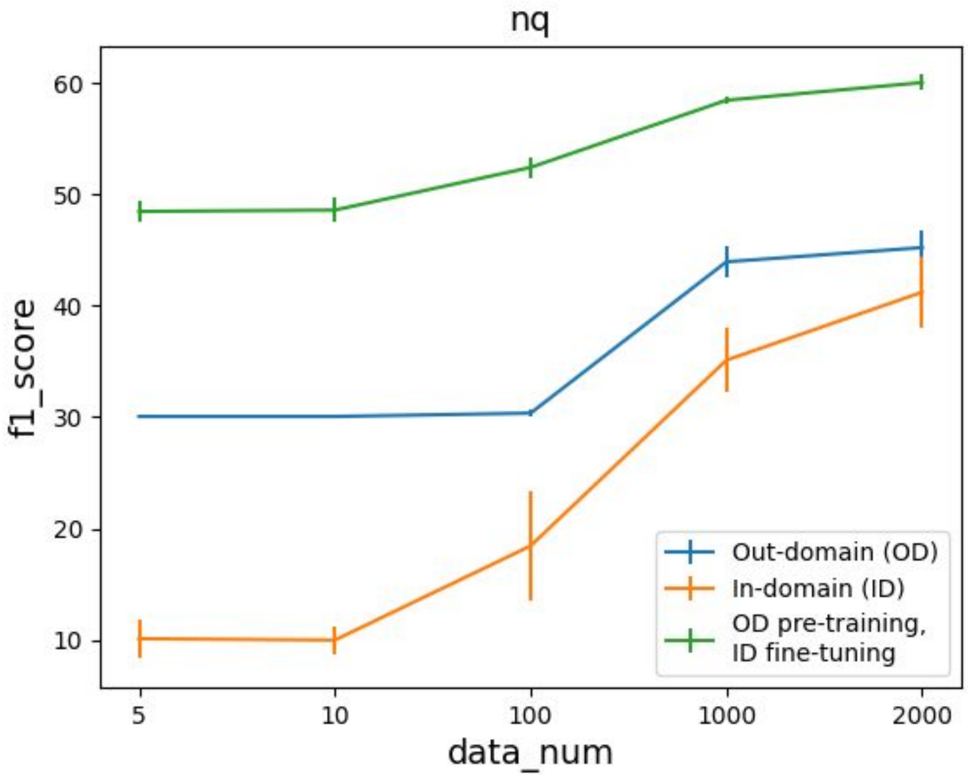
Experiments - hotpotqa



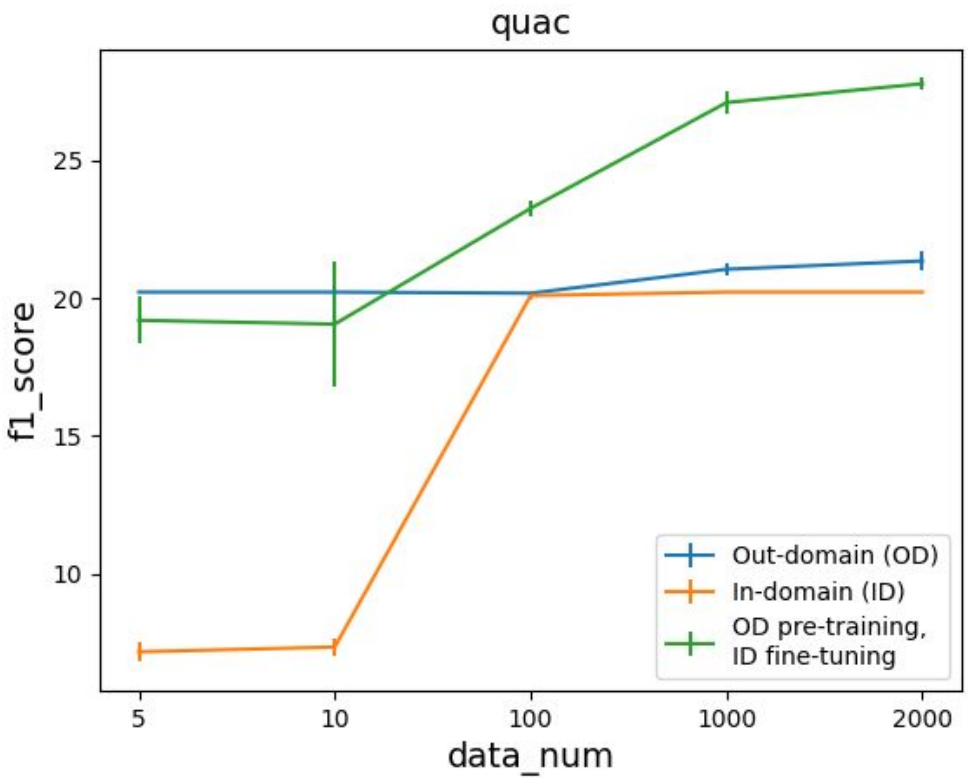
Experiments - newsqa



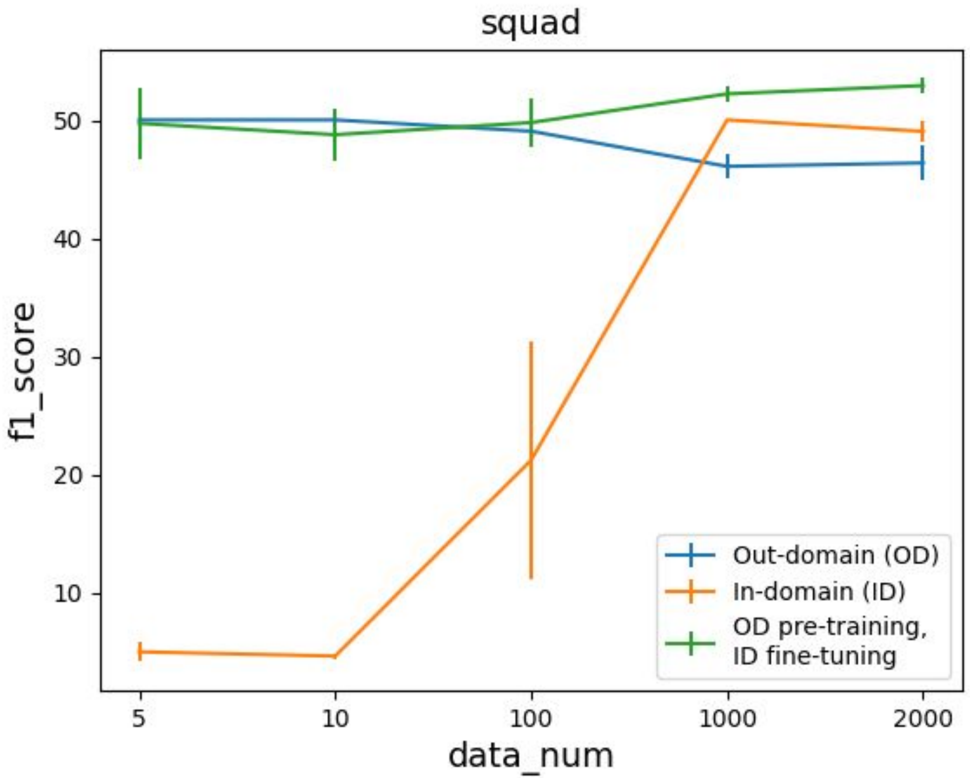
Experiments - nq



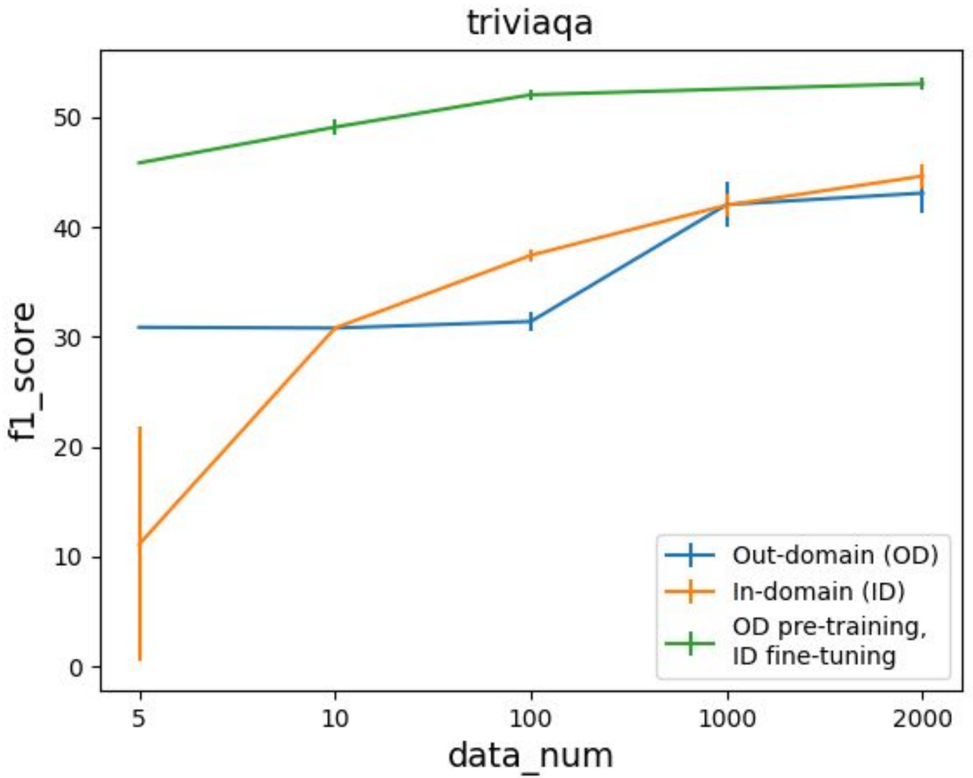
Experiments - quac



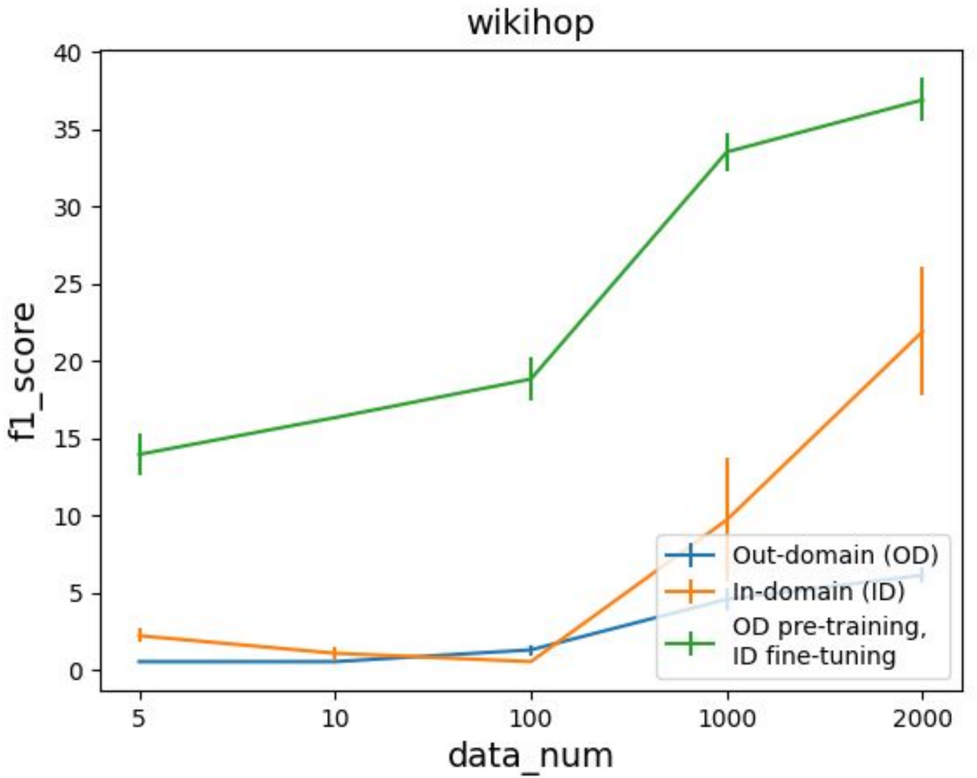
Experiments - squad



Experiments - triviaqa

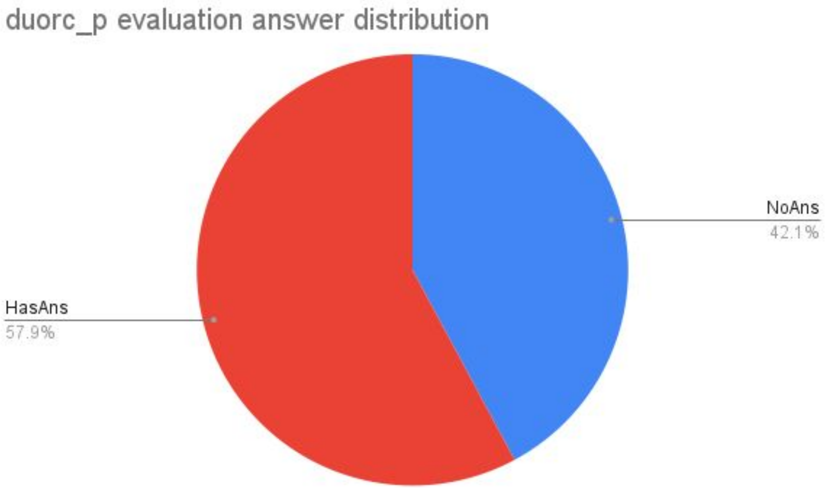


Experiments - wikihop



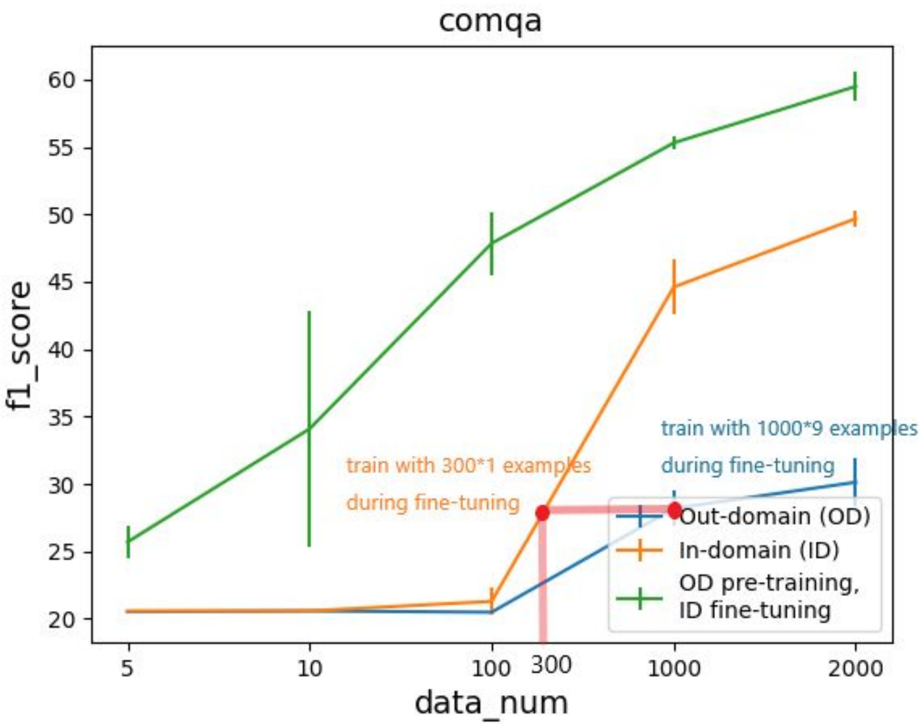
Discussion

- Why there are 2 results?
    - In-domain > Out-domain? Reasonable!
    - Out-domain > In-domain? Why?
- We find that out_domain experiment tends to predict “no answer” in the few-shot setting.



Discussion

- Out-domain useless?
- A preliminary step for ODID.



Conclusion

- **Pretraining on out-domain datasets** and **finetuning on in-domain dataset** can enhance performance hugely.
    - That is what our ODID do!
- Training on out of domain with few shots of data will make model output nothing.
- If we have enough out-domain data, it is useful when our in-domain is insufficient.

References

[1] Priyanka Sen, Amir Saffari. What do Models Learn from Question Answering Datasets? In EMNLP 2020.

[2] Tianyu Gao, Xingcheng Yao, Danqi Chen. SimCSE: Simple Contrastive Learning of Sentence Embeddings. In EMNLP 2021.

[3] Marvin Zhang, Henrik Marklund, Nikita Dhawan, Abhishek Gupta, Sergey Levine, Chelsea Finn. Adaptive Risk Minimization: Learning to Adapt to Domain Shift. In NeurIPS 2021.

Thanks for Listening!