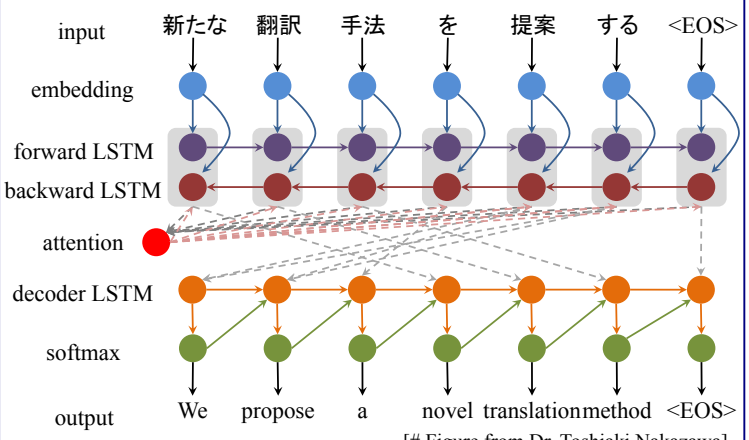
An Empirical Comparison of Domain Adaptation Methods for Neural Machine Translation

Chenhui Chu,[1] Raj Dabre,[2] Sadao Kurohashi[2]
chu@ids.osaka-u.ac.jp, raj@nlp.ist.i.kyoto-u.ac.jp, kuro@i.kyoto-u.ac.jp
[1]Osaka University, [2]Kyoto University

1. Attention-based Neural Machine Translation [Bahdanau+ 2015]



[# Figure from Dr. Toshiaki Nakazawa]

It requires large-scale data for learning large amounts of parameters!

2. The Low Resource Domain Problem

BLEU-4 scores for translation performance

System	NTCIR-CE	ASPEC-CJ	IWSLT-CE	WIKI-CJ
SMT	29.54	36.39	14.31	36.83
NMT	37.11	42.92	7.87	18.29

- Resource rich: NTCIR-CE patent (train: 1M), ASPEC-CJ scientific (train: 672k)
- Resource Poor: IWSLT-CE spoken (train: 209k), WIKI-CJ *automatically extracted* Wikipedia (train: 136k)

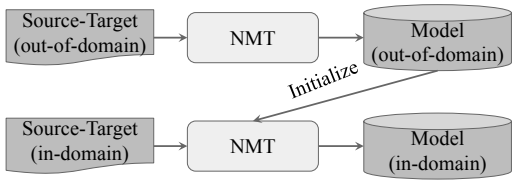
3. Overview of This Study

We conducted an empirical study on different domain adaptation methods for NMT

Methods / Features	Fine tuning [Luong+ 2015] etc.	Multi domain [Kobus+ 2016]	Proposed: mixed fine tuning
Performance	Good	Good	Best
Speed	Fast	Slowest	Slower
Overfitting	Yes	No	No

4. Domain Adaptation Methods for Comparison

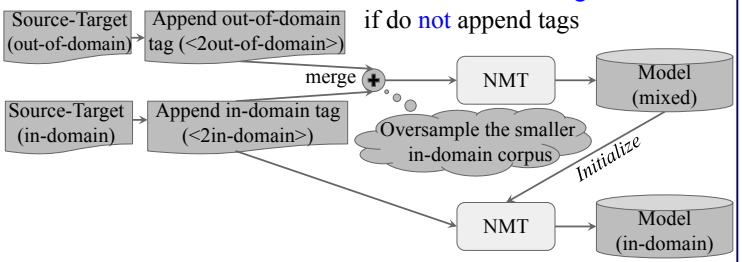
4.1. Fine tuning [Luong+ 2015; Sennrich+ 2016; Servan+ 2016; Freitag+ 2016]



Fine tuning tends to overfit due to the small size of the in-domain data

4.2. Multi domain [Kobus+ 2016]

Multi domain w/o tags: if do not append tags



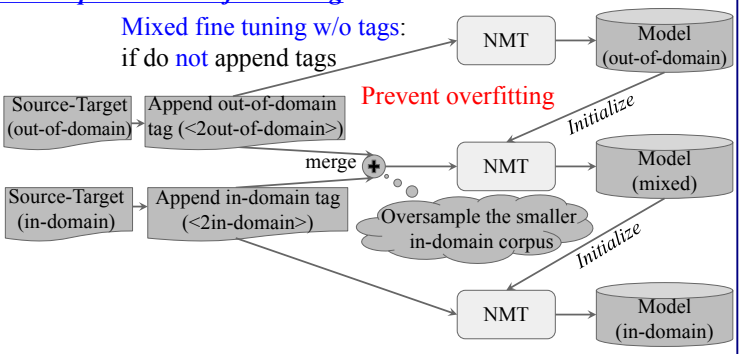
Multi domain + Fine tuning

[Kobus+ 2016] studies multi domain translation, but it is not for domain adaptation in particular

4.3. Proposed: mixed fine tuning

Mixed fine tuning w/o tags: if do not append tags

Prevent overfitting



Mixed fine tuning + Fine tuning

5. Experimental Results & Future Work

- High Quality In-domain Corpus Setting Results (BLEU-4)

System	NTCIR-CE	IWSLT-CE
IWSLT-CE SMT	-	14.31
IWSLT-CE NMT	-	7.87
NTCIR-CE SMT	29.54	4.33
NTCIR-CE NMT	37.11	2.60
Fine tuning	17.37	16.41
Multi domain	36.40	16.34
Multi domain w/o tags	37.32	14.97
Multi domain + Fine tuning	14.47	15.82
Mixed fine tuning	37.01	18.01
Mixed fine tuning w/o tags	39.67	17.43
Mixed fine tuning + Fine tuning	32.03	17.11

- Low Quality In-domain Corpus Setting Results (BLEU-4)

System	ASPEC-CJ	WIKI-CJ
WIKI-CJ SMT	-	36.83
WIKI-CJ NMT	-	18.29
ASPEC-CJ SMT	36.39	17.43
ASPEC-CJ NMT	42.92	20.01
Fine tuning	22.10	37.66
Multi domain	42.52	35.79
Multi domain w/o tags	40.78	33.74
Multi domain + Fine tuning	22.78	34.61
Mixed fine tuning	42.56	37.57
Mixed fine tuning w/o tags	41.86	37.23
Mixed fine tuning + Fine tuning	31.63	37.77

Red numbers indicate the best system and all systems that were not significantly different from the best system

Future work: leverage in-domain monolingual corpora