Domain Adaptation for Neural Machine Translation

Wang Yong, Longyue

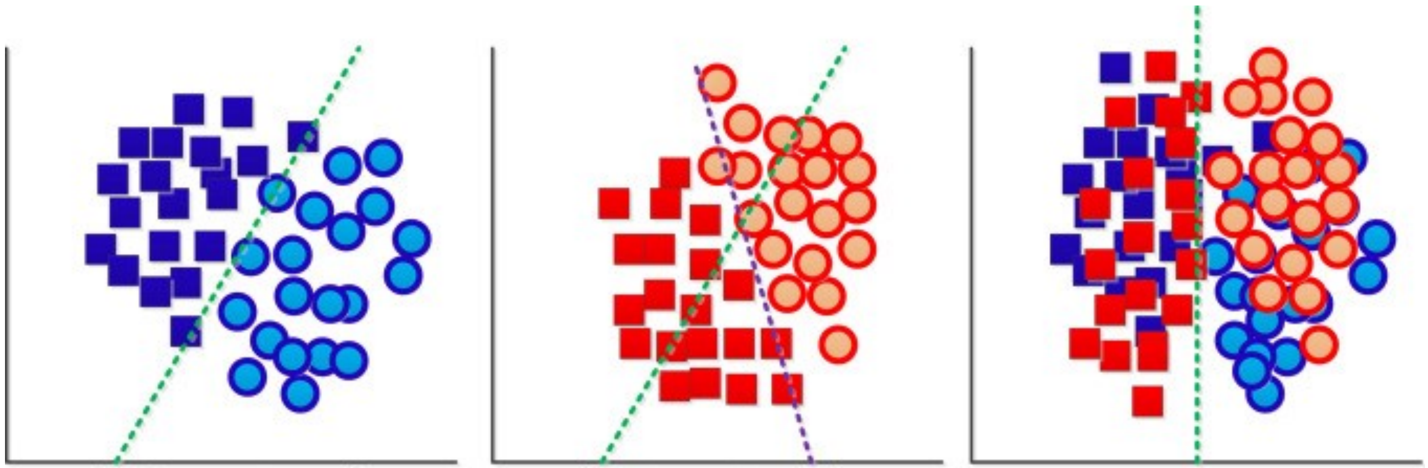
Outline

1. Background (Definition, Problems)
2. Related Work
3. Our Findings (三个维度)
4. Analysis and Interpretation of DA-NMT
5. Cons of Existing Methods
6. Our Proposed Model
7. 实验数据汇总

Background

Definition - ML

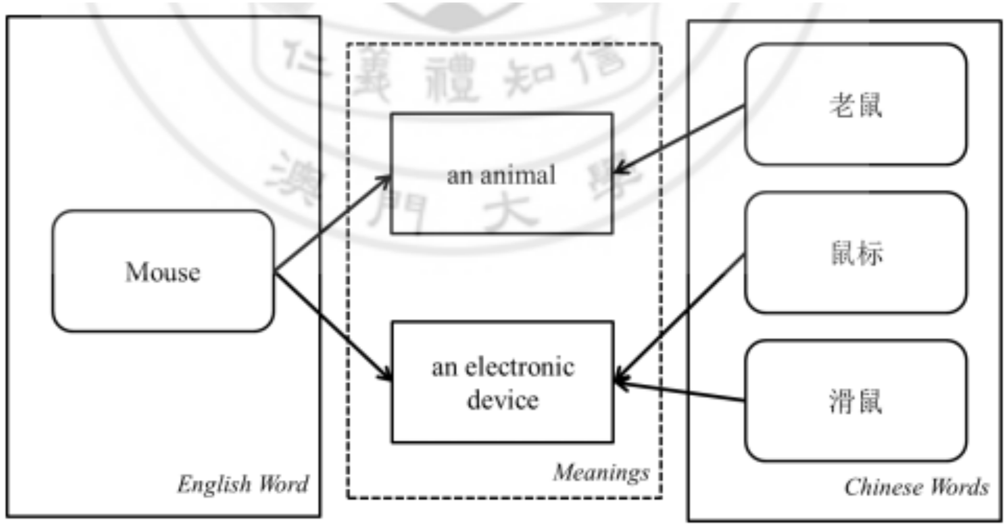
- [Domain] Data distribution.
- [Domain Adaptation] A field associated with machine learning and transfer learning. This scenario arises when we aim at learning from a source data distribution a well performing model on a different (but related) target data distribution.



(a) Source Domain In Subject 1 (b) Target Domain In Subject 2 (c) Domain Adaptation

Definition - MT

- [Domain] Different domains may vary by topic or text style.
- [Domain Adaptation] A mismatch between the domain for which training data are available and the target domain of a machine translation system.



When an international treaty that relates to a contract and **which** the People's Republic of China has concluded on participated into has provisions of the said treaty shall be applied, but with the exception of clauses to which the People's Republic of China has declared reservation.

中华人民共和国缔结或者参加的与合同有关的国际条约同中华人民共和国法律有不同规定的,适用该国际条约的规定。但是,中华人民共和国声明保留的条款除外。

China's Li Duihong won the women's 25-meter sport pistol Olympic gold with a total of 687.9 points early this morning Beijing time. (Guangming Daily, 1996/07/02)

我国女子运动员李对红今天在女子运动手枪决赛中，以687.9环战胜所有对手，并创造新的奥运记录。（《光明日报》1996年7月2日）

Problems

Corpus	Words	Sentences	W/S
Law (Acquis)	18,128,173	715,372	25.3
Medical (EMEA)	14,301,472	1,104,752	12.9
IT	3,041,677	337,817	9.0
Koran (Tanzil)	9,848,539	480,421	20.5
Subtitles	114,371,754	13,873,398	8.2

Table 1: Corpora used to train domain-specific systems, taken from the OPUS repository. IT corpora are GNOME, KDE, PHP, Ubuntu, and OpenOffice.

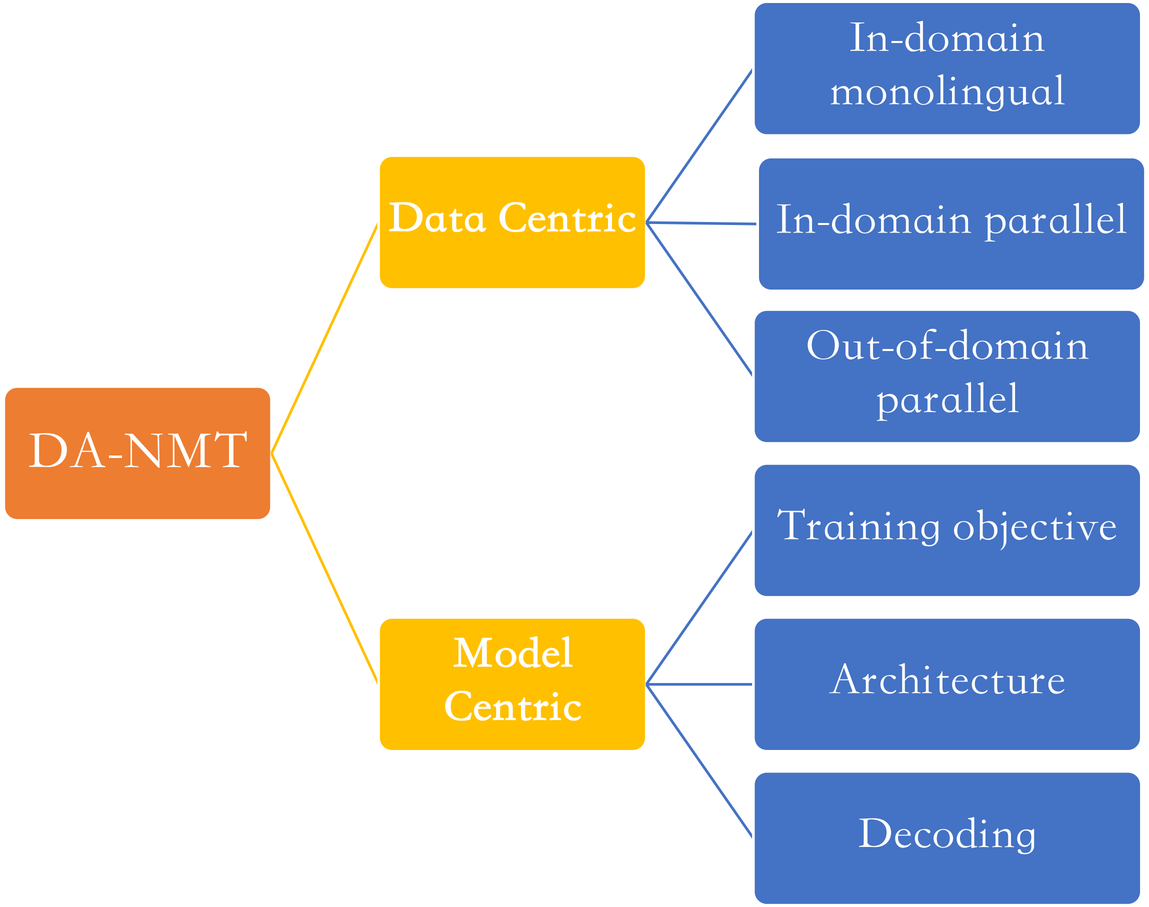
System ↓	Law		Medical		IT		Koran		Subtitles	
All Data	30.5	32.8	45.1	42.2	35.3	44.7	17.9	17.9	26.4	20.8
Law	31.1	34.4	12.1	18.2	3.5	6.9	1.3	2.2	2.8	6.0
Medical	3.9	10.2	39.4	43.5	2.0	8.5	0.6	2.0	1.4	5.8
IT	1.9	3.7	6.5	5.3	42.1	39.8	1.8	1.6	3.9	4.7
Koran	0.4	1.8	0.0	2.1	0.0	2.3	15.9	18.8	1.0	5.5
Subtitles	7.0	9.9	9.3	17.8	9.2	13.6	9.0	8.4	25.9	22.1

NMT systems show more degraded performance out of domain.

Related Work

Related Work

There are mainly two approaches in DA-NMT as discussed in the recent **review paper** (Chenhui Chu et al., 2018):



Related Work

(1) **Data centric**

- Use **in-domain monolingual** corpora: **Language Modeling.**
  - Zhang and Zong, 2016b; Cheng et al., 2016; Currey et al., 2017;
  - Domhan and Hieber, 2017.

- Enlarge **in-domain parallel** data: Synthetic parallel corpora generation, **Translation Model.**
  - Sennrich et al., 2016b; Zhang and Zong, 2016b; Park et al., 2017.

- Use large **out-of-domain parallel** corpora: Multi-domain, Data selection.
  - Chu et al., 2017; Sajjad et al., 2017; Britz et al., 2017;
  - Wang et al., 2017a; van der Wees et al., 2017.

Related Work

(2) **Model centric**

- **Training objective** centric: Instance/cost weighting, Fine-tune, Mixed fine-tune, Regularization.

➢ Luong and Manning, 2015; Sennrich et al., 2016b; Servan et al., 2016; Freitag and Al-Onaizan, 2016; Wang et al., 2017b;

➢ Chen et al., 2017a; Varga, 2017; Dakwale and Monz, 2017; Chu et al., 2017; Miceli Barone et al., 2017.

- **Architecture** centric: Deep fusion, Domain discriminator, Domain control.

➢ Kobus et al., 2016; Gulcüehre et al., 2015; Britz et al., 2017.

- **Decoding** centric: Shallow fusion, Ensemble, Neural lattice search.

➢ Gulcüehre et al., 2015; Dakwale and Monz, 2017; Khayrallah et al., 2017.

Related Work

More recently, researchers mainly focus on two parts:

- **Develop new models**

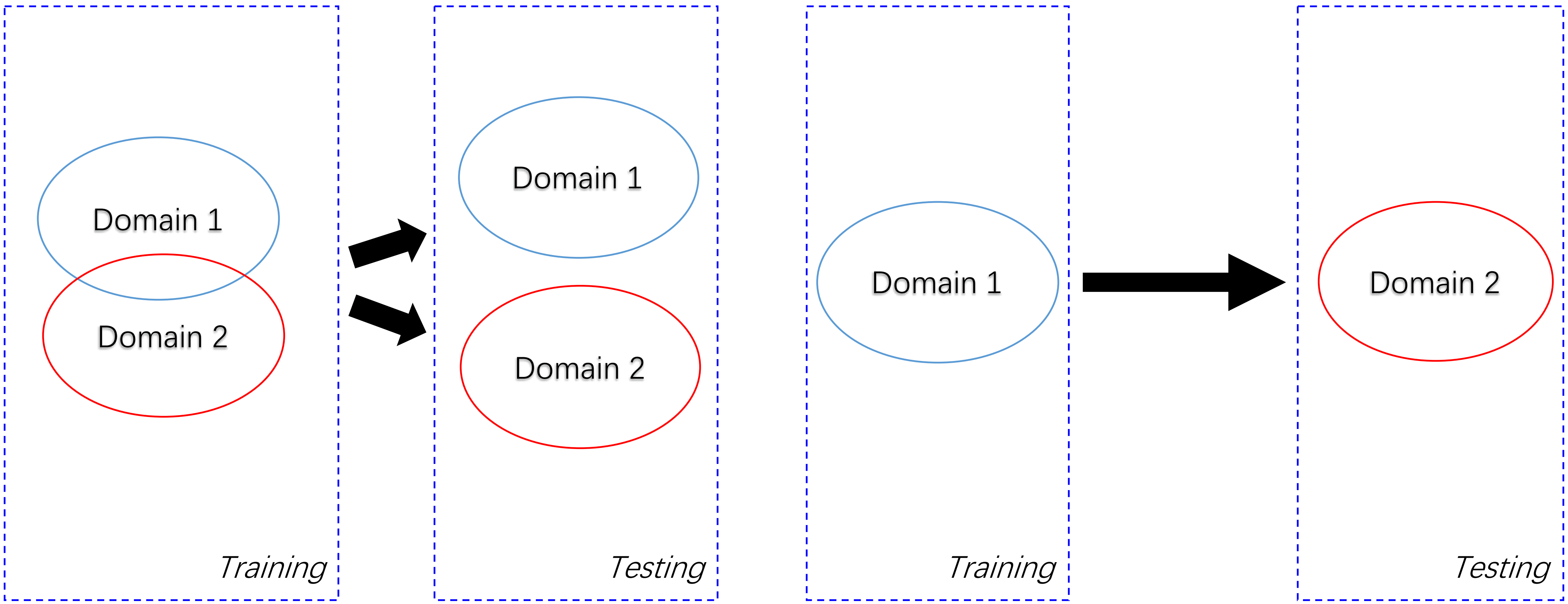
➢ Jiali Zeng et al., 2018, Asa Cooper Stickland et al., 2019.

- **Model interpretation and analysis**

➢ Brian Thompson et al., 2018.

Our Findings

维度一：DA-NMT 任务场景



Domain-Mix

Domain-Shift

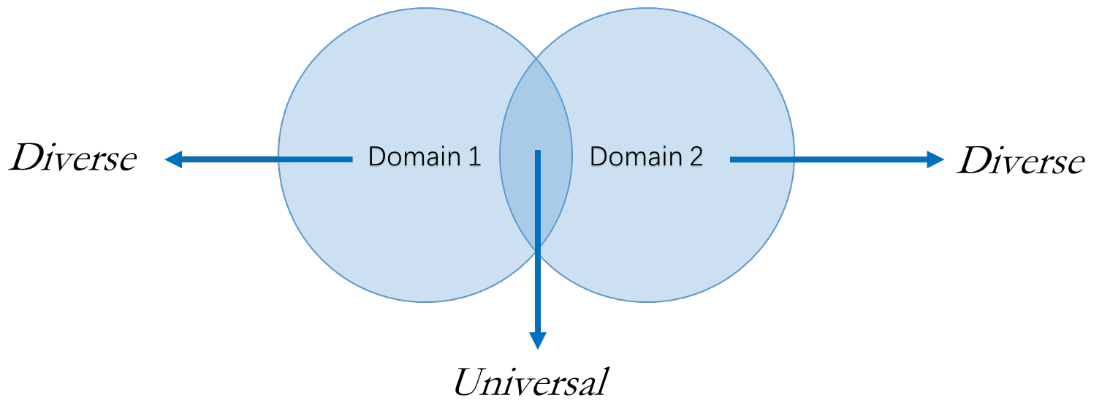
维度二：DA-NMT 核心问题

How to deal with both **universal** and **diverse knowledge**:

- **Universal**: domain-shared representations
- **Diverse**: domain-specific representations

Fine-tune举例：

- **Universal**: mix data and train a general model;
- **Diverse**: tune domain-aware parameters on in-domain data.



维度三：DA-NMT 解决手段

To balance the **trade-off**, existing work investigates approaches from:

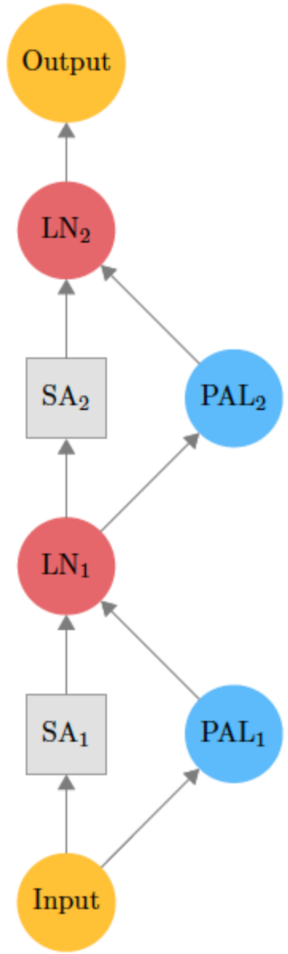
- **Parameters:**
  - ➢ shared & domain-specific parameters (Cooper Stickland et al., 2019).
- **Representations:**
  - ➢ split the annotations of encoder into common and domain-specific parts (Jiali Zeng et al., 2018).
- **Optimization:**
  - ➢ fine-tune (Chenhui Chu et al., 2017).

Parameters

BERT and PALs: Projected Attention Layers for Efficient Adaptation in Multi-Task Learning

Asa Cooper Stickland, Iain Murray

Arxiv 2019



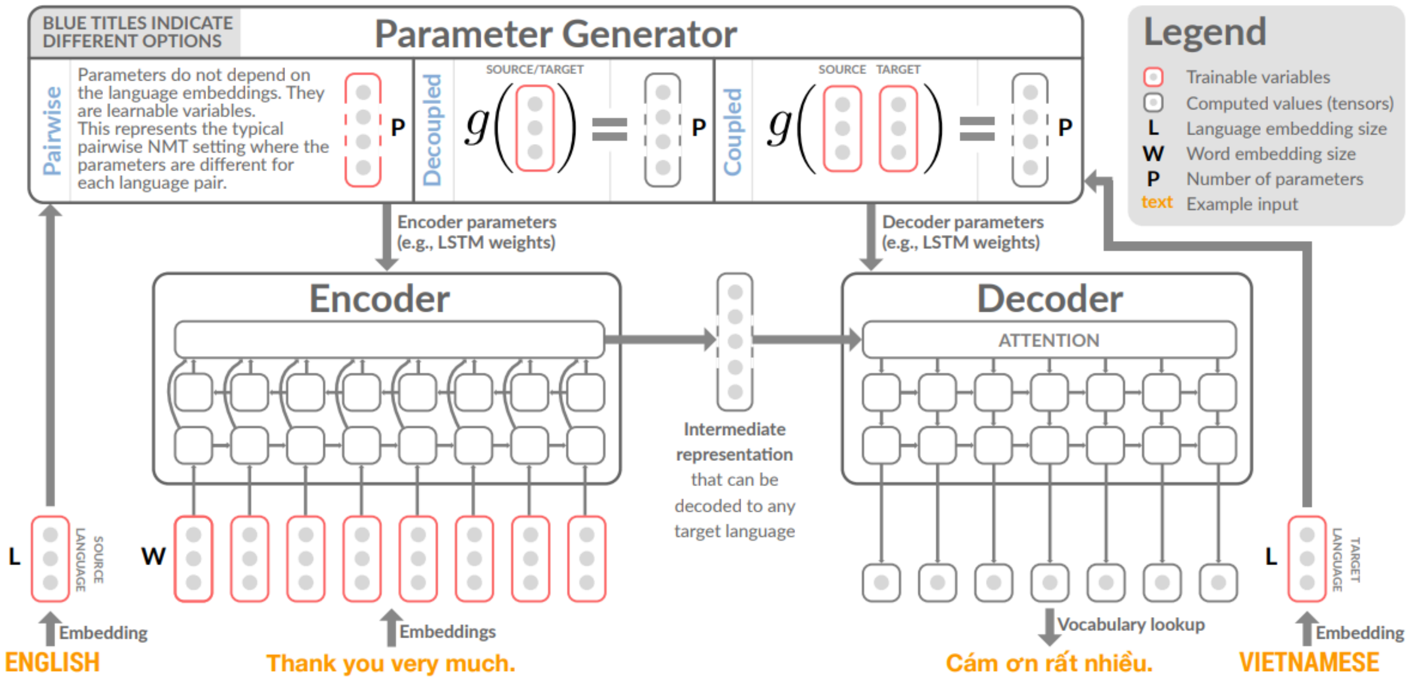
PAL: Projected Attention Layers
LN: Layer Normalization
SA: Self-attention

Parameters

Contextual Parameter Generation for Universal Neural Machine Translation

EA Platanios, M Sachan, G Neubig

EMNLP 2018

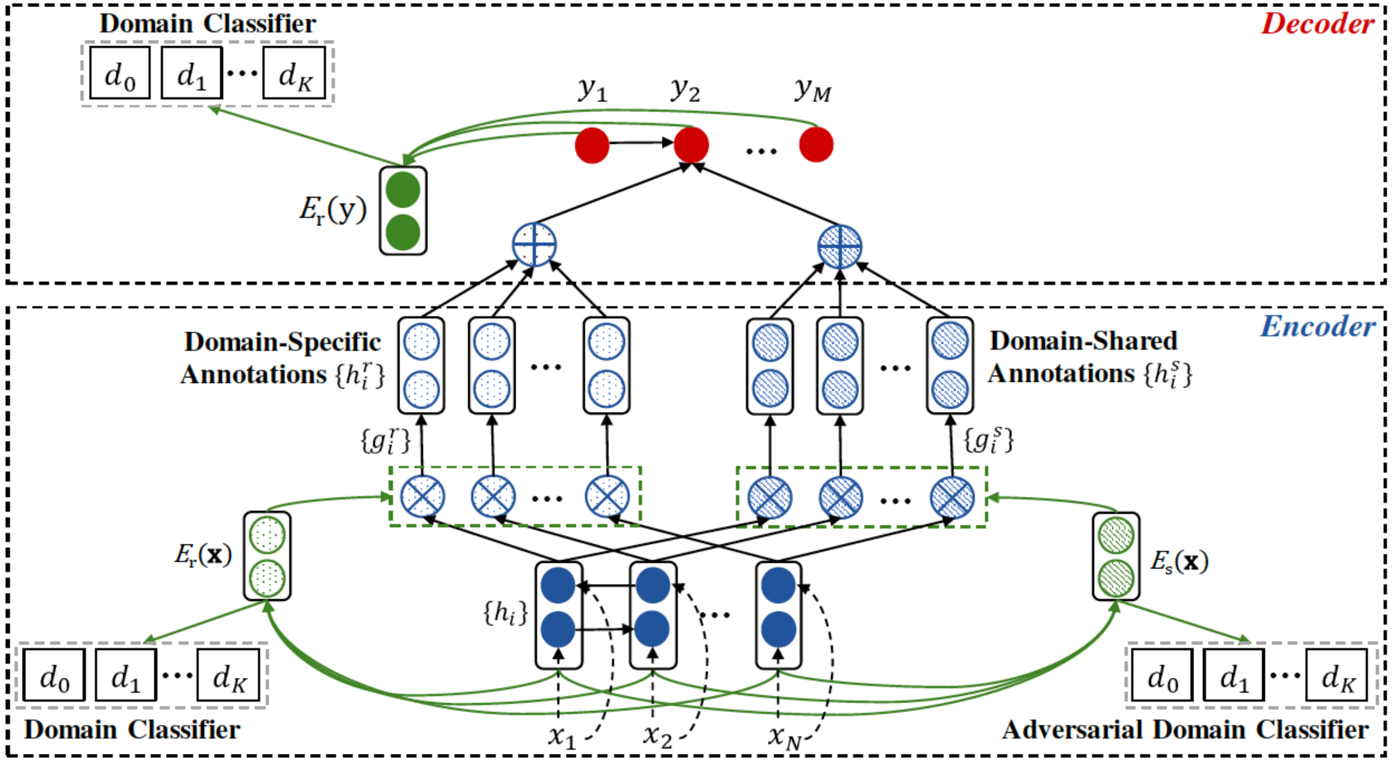


Representations

Multi-Domain Neural Machine Translation with Word-Level Domain Context Discrimination

Jiali Zeng et al.

EMNLP 2018



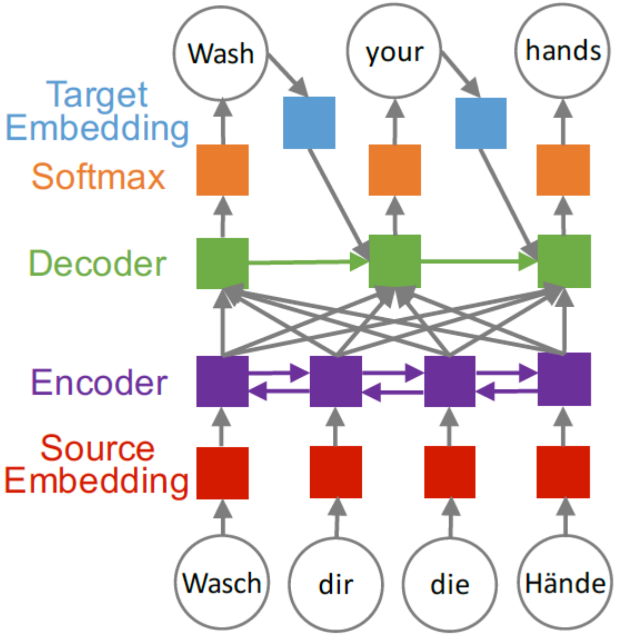
Analysis and Interpretation of DA-NMT

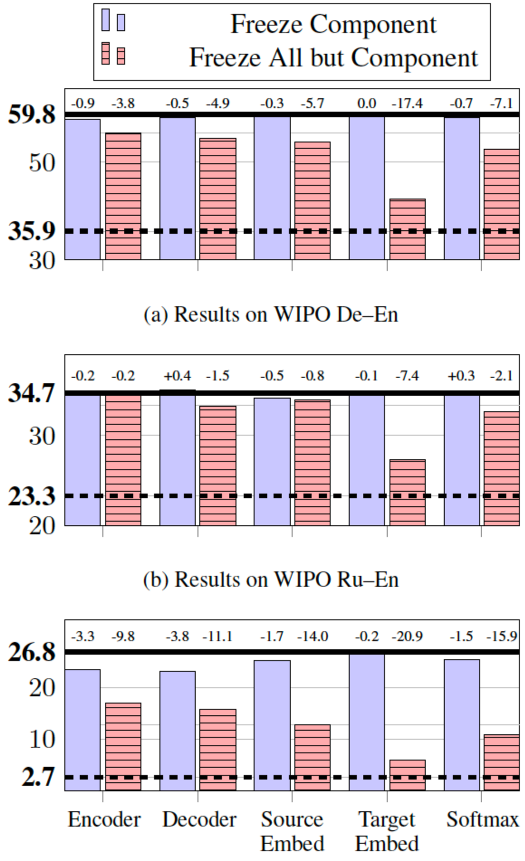
Analysis and Interpretation of DA-NMT

Freezing Subnetworks to Analyze Domain Adaptation in Neural Machine Translation

Brian Thompson et al.

Arxiv 2019





(a) Results on WIPO De–En

(b) Results on WIPO Ru–En

(c) Results on WIPO Ko–En

Analysis and Interpretation of DA-NMT

Effective Domain Mixing for Neural Machine Translation

Reid Pryzant et al.

WMT 2017

Lanuage	**Domain 1**	**Domain 2**	$\hat{d}_A$
Japanese	ASPEC	SubCrawl	1.89
Chinese	News	TED	1.73
French	Europarl	OpenSubs	1.23