

# **A survey of Domain Adaptation for Neural Machine Translation**

**COLING, 2018**

# Introduction

- NMT는 대규모의 데이터를 이용가능한 환경에서 좋은 성능을 자랑함
- Domain adaptation은 그 중에서도 매우 중요한 분야
  - 양질의 parallel corpora는 영어와 일부 유럽언어를 제외하고는 부족
  - Domain specific data는 더욱 부족 : Domain specific translation 성능이 낮은 이유
  - Domain specific MT에 일반적인 목적을 가진 MT를 적용하는데는 한계가 있음
- 본 논문에서는 NMT의 SOTA domain adaptation에 적용되는 포괄적인 연구를 소개

# SMT domain adaptation

- Data Centric
  - 모델을 특화시키는 것이 아닌 데이터를 특화시키는 방법
  - Data selection, generation과 관련된 기법
- Model Centric
  - 모델, 학습방식을 특화시키는 방법
  - In-domain model과 out-of-domain model을 합쳐서 사용
  - Model level, Instance level에서 수정

# NMT domain adaptation

- Data Centric
  - Using Monolingual Corpora
  - Synthetic Parallel corpora Generation
  - Using out-of-Domain Parallel corpora
- Model Centric
  - Training Objective Centric
  - Architecture Centric
  - Decoding Centric

# NMT domain adaptation

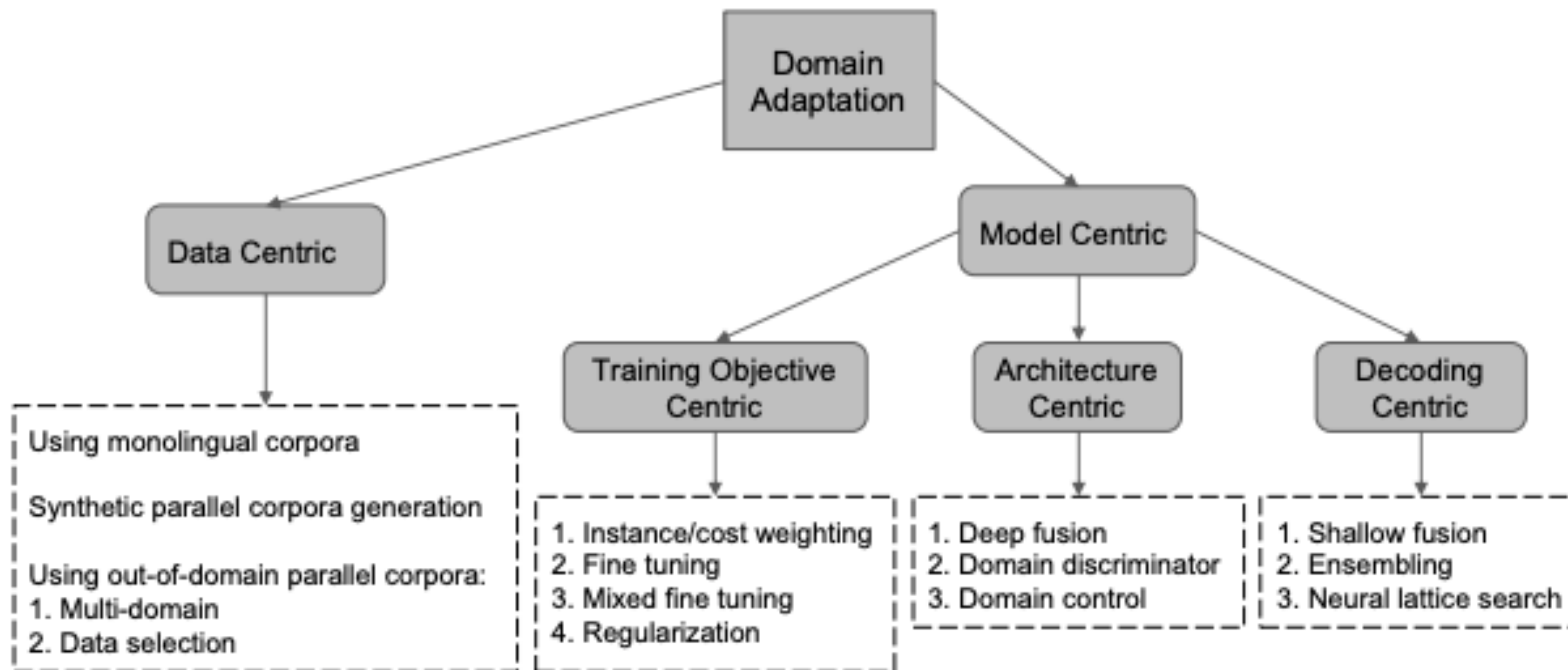


Figure 1: Overview of domain adaptation for NMT.

# Data Centric

## Using Monolingual Corpora

- in-Domain Parallel corpora는 데이터가 매우 부족
- in-Domain monolingual corpora는 그에 비해 구하기 쉬움
- Method
  - in-Domain monolingual data를 LM에 학습시키고, NMT model과 합침
    - NMT의 decoder에 사용
    - NMT의 encoder에 강화하는데 사용
    - LM과 NMT를 fuse 시킴

# Data Centric

## Synthetic Parallel corpora Generation

- NMT는 그 자체로 LM을 학습하는 능력을 가짐
- Target monolingual data 또한 NMT의 decoder를 강화하는데 사용될 수 있다.
- Target data를 translate해서 새로운 Synthetic Parallel corpora를 생성하는 방법
  - Back translating : target -> synthetic source
- Target monolingual data뿐만 아니라 Source side monolingual data를 사용해도 매우 효과적인 방법



Figure 3: Synthetic data generation for NMT (Sennrich et al., 2016b).

# Data Centric

## Using out-of-domain Parallel corpora

- Mixed domain MT system(Multi-domain)
  - In-domain translation 성능은 향상
  - out-of-domain translation 성능은 유지
  - In-domain 데이터와 out-of-domain data를 함께 사용
- Multiple domain의 corpora는 두개의 작은 수정을 거침
  - <2domain> tag를 source sentence에 append
  - In-domain corpus를 더 작게 조정
  - NMT decoder가 특정 도메인에 필요한 문장을 생성하게 해줌

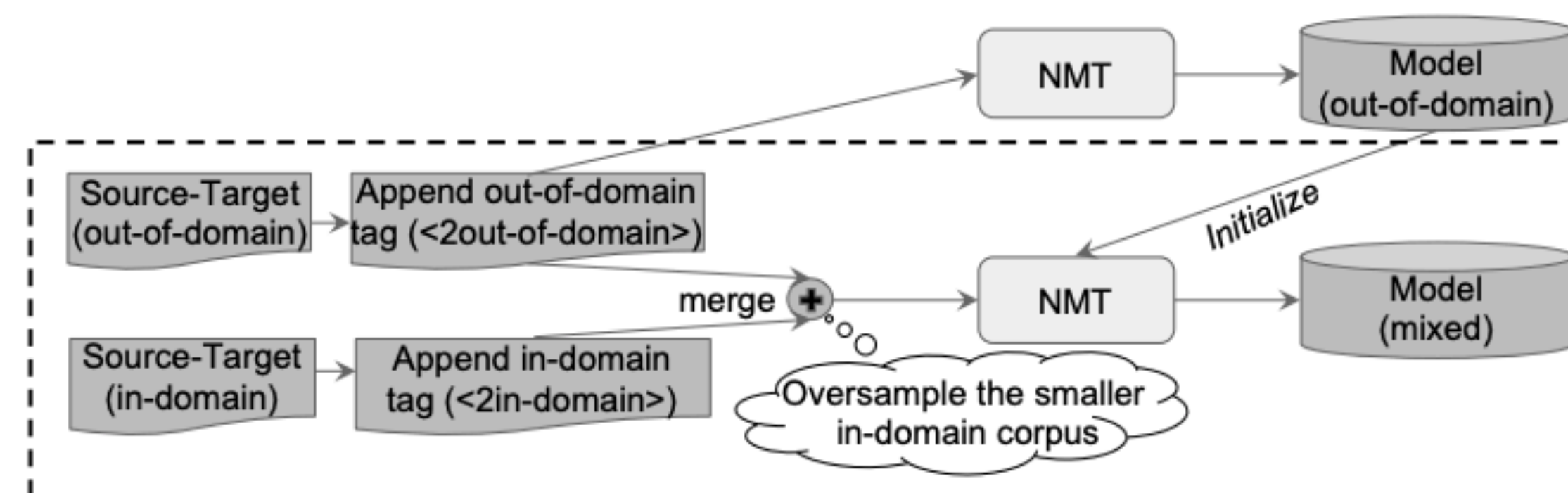


Figure 6: Mixed fine tuning with domain tags for domain adaptation (Chu et al., 2017). The section in the dotted rectangle denotes the *multi-domain* method .



# Data Centric

## Data selection

- Source sentence에 internal embedding을 진행
- Sentence embedding similarity를 사용해서 문장들을 select
- Out-of-domain data에서 in-domain data에 가까운 문장들을 선택해서 학습에 이용
- Dynamic data selection method : Training epoch사이에 선택된 training data의 subset을 줄임

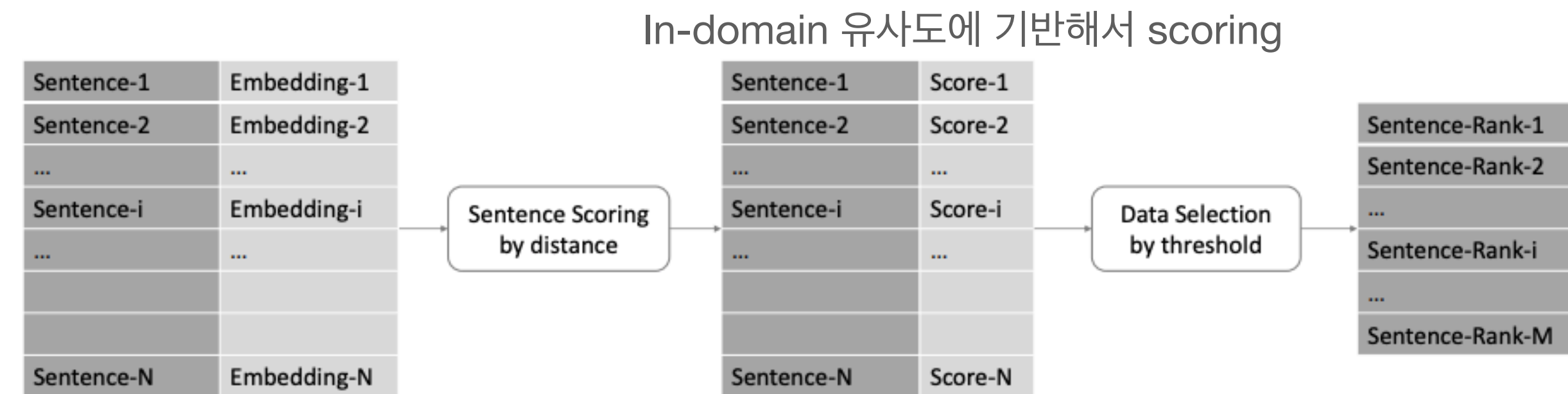


Figure 4: Data selection for NMT (Wang et al., 2017a).

# Model Centric

## Training Objective Centric

- Instance/Cost Weighting(Wang et al, 2017)
  - Objective function의 weight를 설정하고, cross-entropy로 in-domain LM과 out-of-domain LM을 학습
  - Instance weighting 대신, domain classifier와 함께 NMT의 cost function을 수정
  - Domain classifier를 development data로 학습
  - Domain classifier의 output probability는 domain weight로 전이
- Fine Tuning(Varga et al, 2017)
  - Out-of-domain corpus 자원이 풍부한 상황에서 NMT를 수렴할 때 까지 훈련, 이후에 in-domain corpus 자원으로 fine tuning
  - Comparable corpora에서 parallel corpus를 추출해서 사용 가능

# Model Centric

## Training Objective Centric

- Mixed Fine Tuning : multi-domain + fine tuning (Chu et al, 2017)
  - NMT model을 out-of-domain data로 수렴할 때까지 학습
  - Mixed domain data : in-domain data와 out-of domain data(oversampling the in-domain data)를 혼합
  - Mixed domain data로 fine-tuning
- Regularization(Barone et al, 2017)
  - Fine tuning동안 overfitting problem이 존재
  - Regularization techniques들을 실험 (Dropout, L2-regularization)
  - Tune out : 다양한 Dropout을 적용해 regularization하는 방법

# Model Centric

## Architecture Centric

- Deep Fusion
  - In-domain monolingual data를 이용해 adaptation
  - In-domain LM을 훈련시켜 NMT의 decoder로 사용
  - 하나의 모델로 사용
- Domain Discriminator
  - Domain Discriminator와 NMT를 연결해서 사용

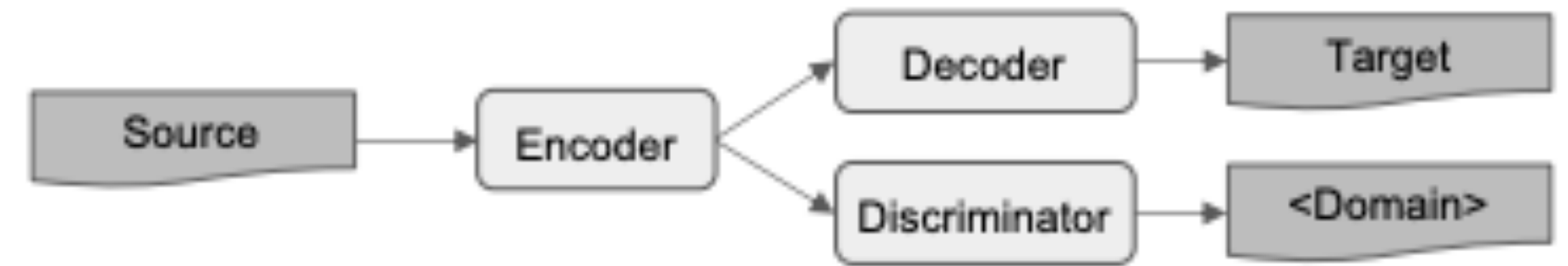


Figure 7: Domain discriminator (Britz et al., 2017).

# Model Centric

## Architecture Centric

- Domain Control (Kobus et al, 2016)
  - Domain tokens을 사용해서 domain control을 진행
  - NMT의 embedding layer에 Word level features를 추가함
  - Domain tag를 각 단어에 추가했고, Domain tag를 예측하는데 TF-IDF에 기반한 방법을 제안

# Model Centric

## Decoding Centric

- Shallow fusion
  - LM이 큰 monolingual corpora로 학습되어 있을때 이전에 훈련된 NMT model과 결합해서 사용하는 방법
  - Deep fusion의 경우 decoder로 LM이 종속, 하나의 모델
  - Shallow fusion의 경우 별개의 model로서 각각의 output(다음에 올 단어의 probability)를 weighted sum으로 rescoring
- Ensembling (Freitag, 2016)
  - Out-of-domain model과 fine-tuned in-domain model의 ensemble
  - Motivation : In-domain data finetuning한 후에도 out-of-domain translation 성능의 유지를 위해 사용



Figure 8: LM shallow fusion (Gülçehre et al., 2015).

# Real-World Scenarios

- Domain adaptation 기법은 특정 시나리오에 따라 채택될 수 있음
  - Out-of-domain data에서 in-domain이 포함되어 있는 경우 : sentence selection
  - Monolingual data만 이용가능한 경우 : LM + NMT fusion
  - 두가지 다 있는 경우, 여러 방법의 combination이 가능
  - Chu et al(2018) : Mixed finetuning + synthetic parallel data
- 최고의 방법을 직접 찾아보기를 권장함

# Future Directions

- Domain adaptation for State-of-the-art NMT Architectures
  - 기존의 연구가 대부분 RNN기반으로 이루어졌기 때문에 최신 모델(transformer)에 domain adaptation techniques을 적용하는 방향을 제시
- Domain Specific Dictionary Incorporation
  - 사전과 같은 외부의 Knowledge를 사용하는 것
- Multilingual and Multi-Domain Adaptation
  - 항상 out-of-domain parallel corpus를 사용하는 것이 가능하진 않음
  - 다른 언어로 부터 언어쌍을 가져오는 방법을 사용
  - Crosslingual transfer learning : 여러 언어들 사이의 NMT model parameter를 전이(parameter sharing)
  - Out-of-domain data가 같은 언어쌍에 존재하더라도, multi-lingual, multi-domain data 각각 사용하는 것이 번역 성능을 향상시킴
- Adversarial Domain Adaptation and Domain Generation
  - GAN을 이용한 방법을 domain adaptation에서 적용
  - NLP에 GAN기법을 적용한 연구들이 존재(Chen,Britz, 2017)