DMQA Open Seminar

# How to apply few-shot learning to Named Entity Recognition task?

2022.12.09


**D**ata **M**ining & **Q**uality **A**nalytics Lab.

이정민

# 발표자 소개


❖ **이정민(JungMin Lee)**

- 고려대학교 산업경영공학과 대학원 재학
- Data Mining & Quality Analytics Lab.(김성범 교수님)
- 석사 과정(2022.03~Present)


❖ **Research Interest**

- Time series Anomaly Detection
- Named Entity Recognition


❖ **Contact**

- jungmin9195@korea.ac.kr

# Contents

# Introduction

# Introduction

Named Entity Recognition

## ❖ 개체명 인식(Named Entity Recognition)이란?

- 미리 정의해 둔 사람, 회사, 장소, 시간, 단위 등에 해당하는 단어(개체명)를 문서에서 인식하여 추출 분류하는 기법으로 **token classification task**로 볼 수 있음
- 도메인 혹은 데이터셋 마다 사용되는 클래스(개체명)가 다를 수 있음

| | | |
|---|---|---|
| **Politics** | Hugo Chavez's political party, the United Socialist Party of Venezuela, drew 48% of the votes overall. | **politician**, **political party** |
| **Natural Science** | Mars has four known co-orbital asteroids, such as 5261 Eureka, all at the Lagrangian points. | **Astronomical object**, **Astronomical object**, **miscellaneous** |
| **Music** | House of Pain abruptly broke up in 1996 after the release of their third album, Truth Crushed to Earth Shall Rise Again. | **band**, **album** |
| **Literature** | Charies spent outdoors, but also read voraciously, including the picaresque noveis of Tobias Smollett. | **writer**, **literary genre**, **writer** |
| **Artificial Intelligence** | The gradient decent can take many interactions to compute a local minimum with a required accuracy. | **algorithm**, **miscellaneous**, **metrics** |



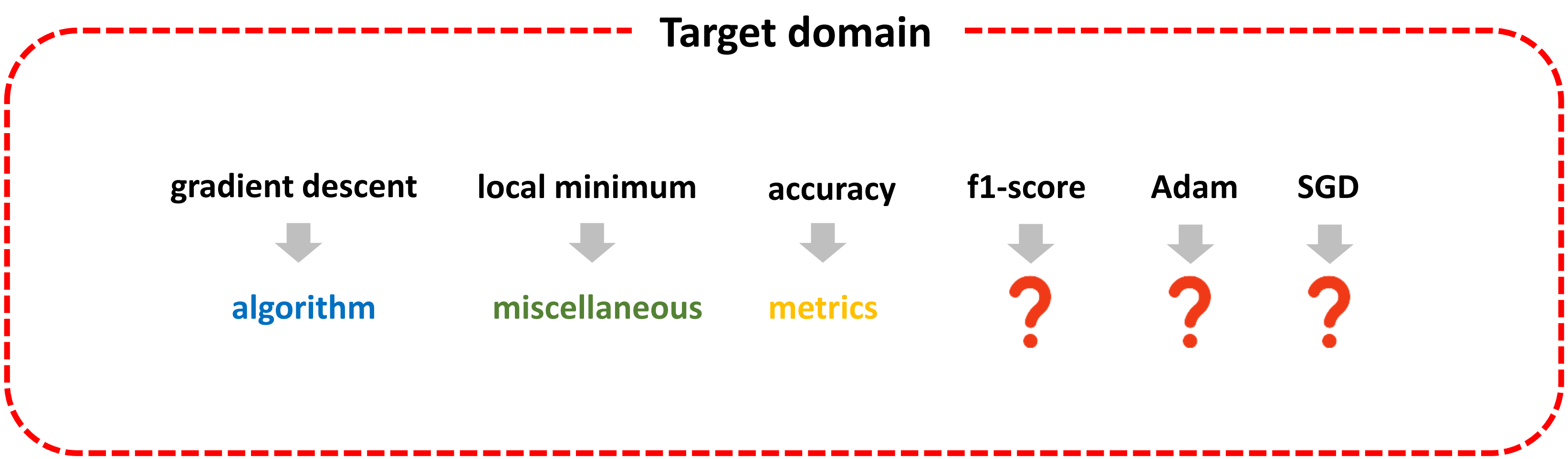
DMQA

# Introduction

Named Entity Recognition

## ❖ 문제 상황

- 타겟 도메인에서 다량의 레이블된 토큰 수집이 제한적
- 소량의 타겟 도메인 데이터로 fine-tuning 하는 방법은 부적절할 수 있음



# Introduction

Named Entity Recognition

❖ 문제 상황

- 타겟 도메인에서 다량의 레이블된 토큰 수집이 제한적
- 소량의 타겟 도메인 데이터로 fine-tuning 하는 방법은 부적절할 수 있음

Target domain

| gradient descent | local minimum | accuracy | f1_score | Adam | SGD |
| --- | --- | --- | --- | --- | --- |
| ⬇ | ⬇ | ⬇ | ⬇ | ⬇ | ⬇ |
| algorithm | miscellaneous | metrics | ? | ? | ? |

**“분류하는 방법 " 을 학습하는**
**few-shot learning의 필요성 대두**



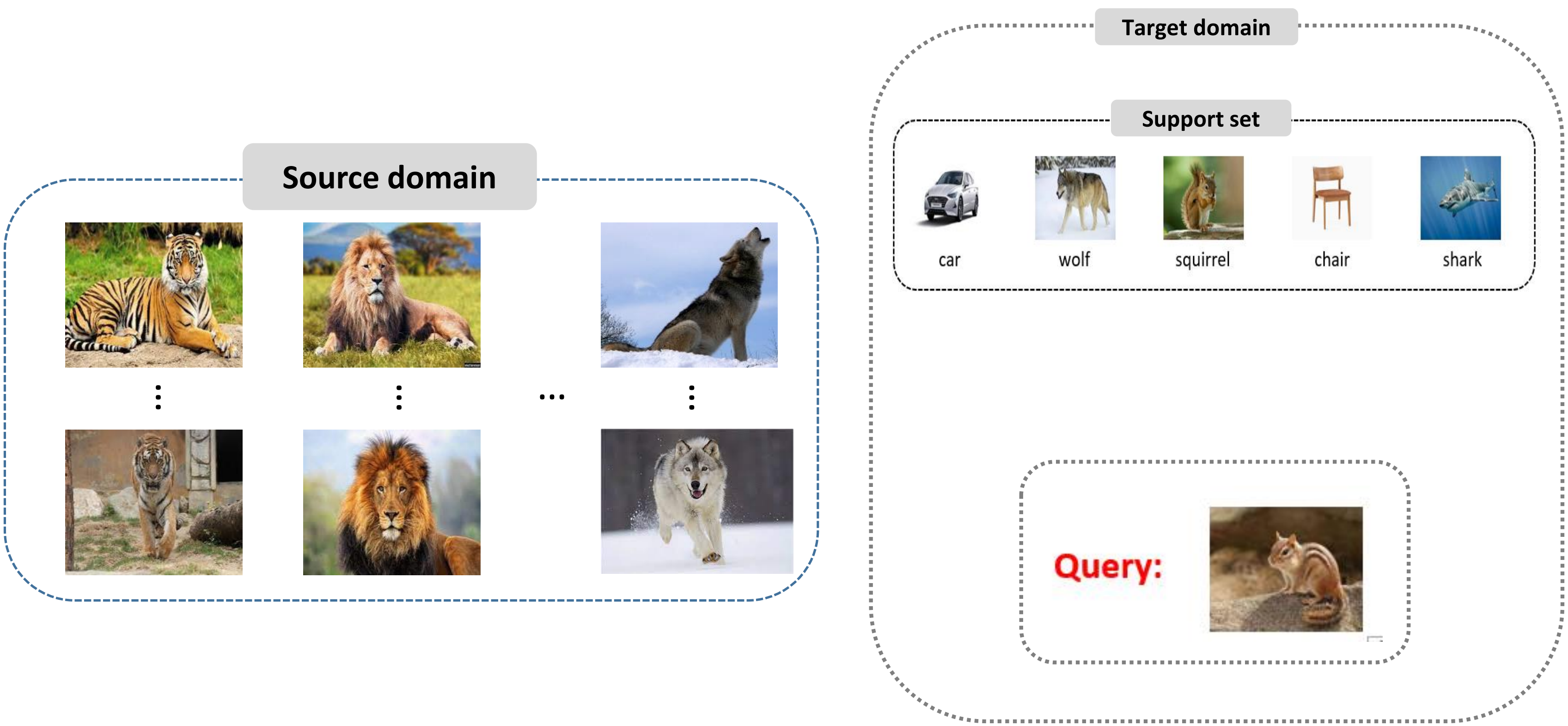
DMQA

# Introduction

## Few shot learning

### ❖ Few shot learning 이란?

- **"Few"** 한 양의 데이터로 모델을 학습하여 테스트 데이터에서 유의미한 성능을 내고자 하는 방법
- 지도 학습과 달리 **학습 데이터 셋에 없는 클래스**를 맞추는 문제

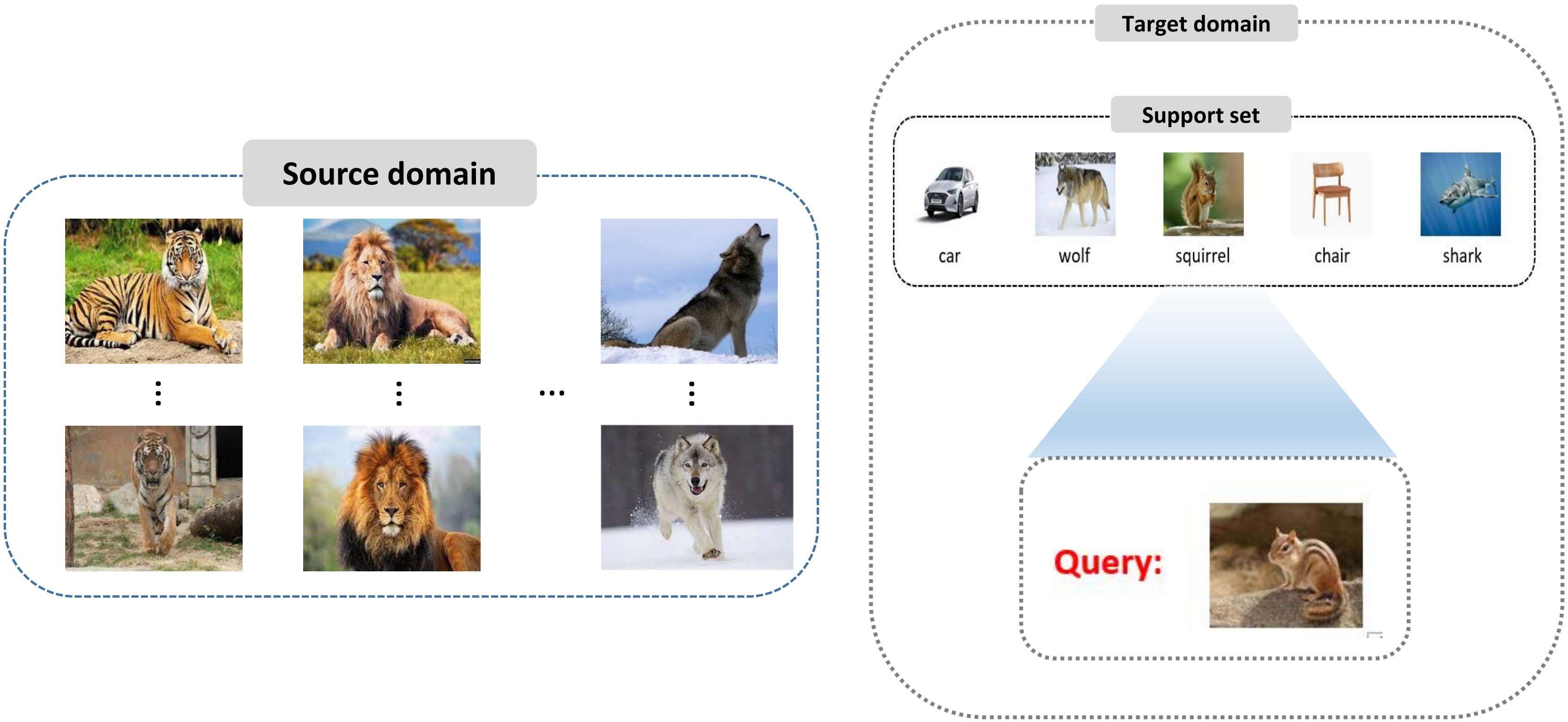


# Introduction

Few shot learning

## ❖ Few shot learning 이란?

- **"Few"** 한 양의 데이터로 모델을 학습하여 테스트 데이터에서 유의미한 성능을 내고자 하는 방법
- 학습 데이터에서 **구분하는 법**을 배우고 query sample이 support set 중 어떤 클래스와 같은 클래스인지를 맞춤

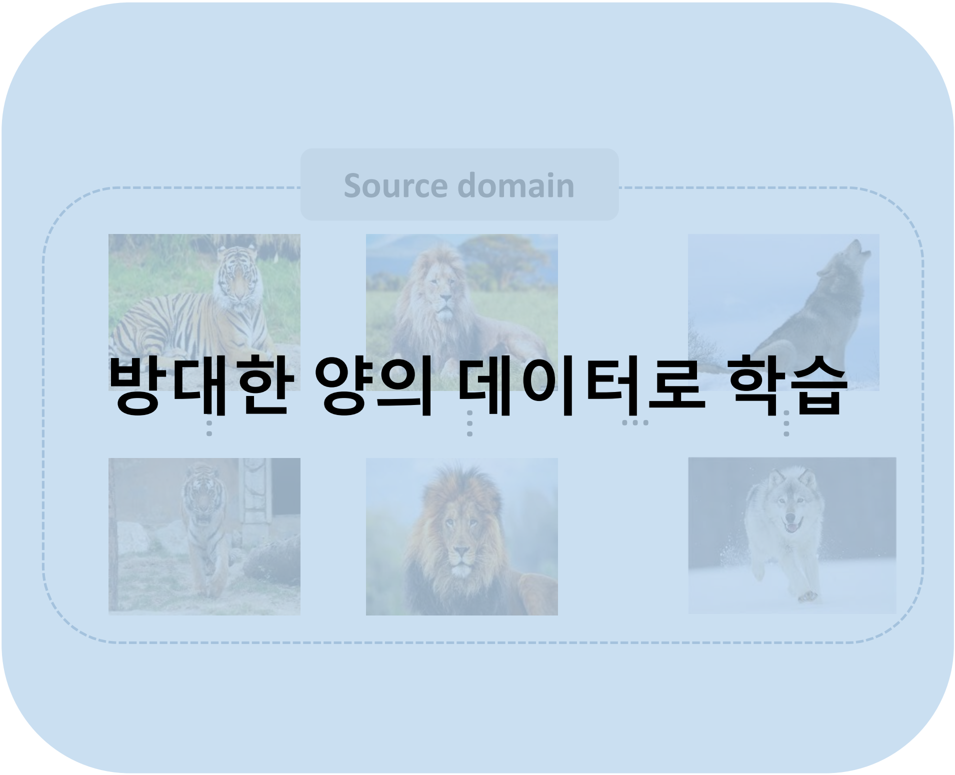


# Introduction

Few shot learning

❖ **Few shot learning 이란?**

- **"Few"** 한 양의 데이터로 모델을 학습하여 테스트 데이터에서 유의미한 성능을 내고자 하는 방법
- 학습 데이터에서 **구분하는 법**을 배우고 query sample이 support set 중 어떤 클래스와 같은 클래스인지를 맞춤



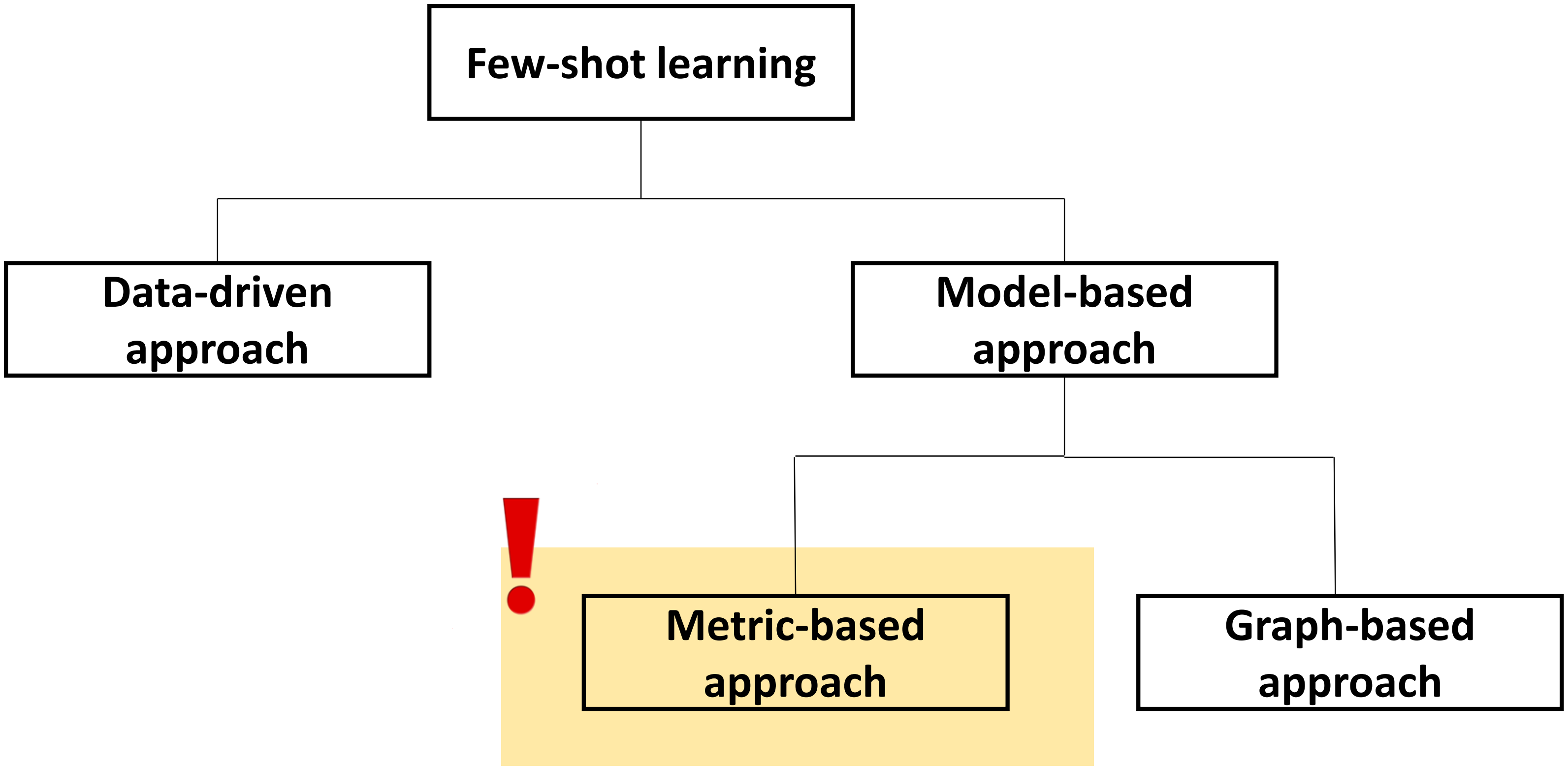




DMQA

# Introduction

Few shot learning








DMQA

# Introduction

Few shot learning

## ❖ Few shot learning 학습 방법

- Similarity 계산
  - ✓ Euclidean distance
  - ✓ Cosine similarity

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2}$$

**< Euclidean distance>**

$$similarity = cos(\theta) = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

**< Cosine similarity >**



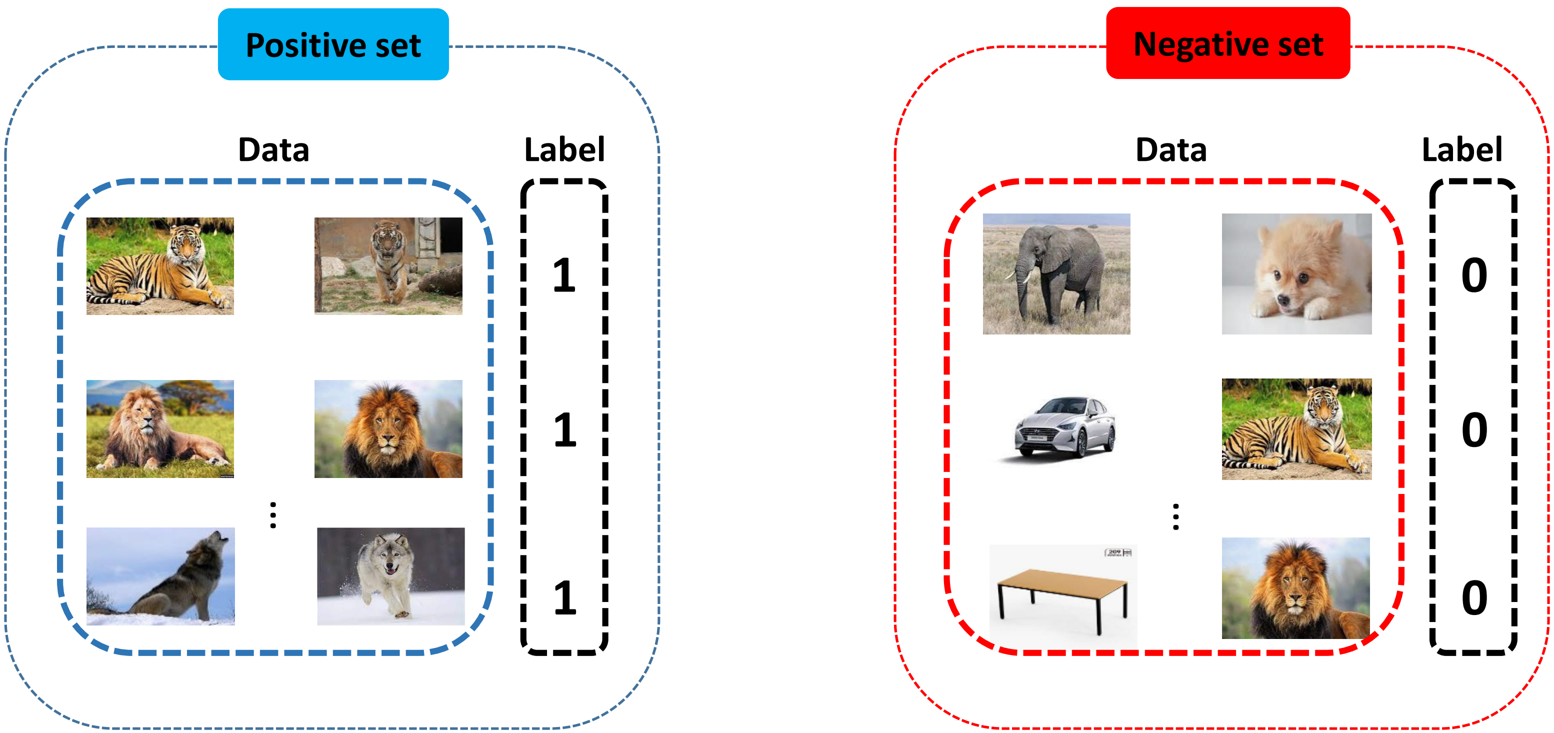
DMQA

# Introduction

Few shot learning

## ❖ Few shot learning 학습 방법

- 다량의 source domain 데이터로 metric 기반 학습
- 데이터셋 구성: positive set + negative set






DMQA

# Introduction

Few shot learning

## ❖ Few shot learning 학습 방법

- **Positive pair**는 서로 **가까워지도록**, **negative pair**는 서로 **멀어지도록** 학습

**< Positive pair>**

**< Negative pair >**



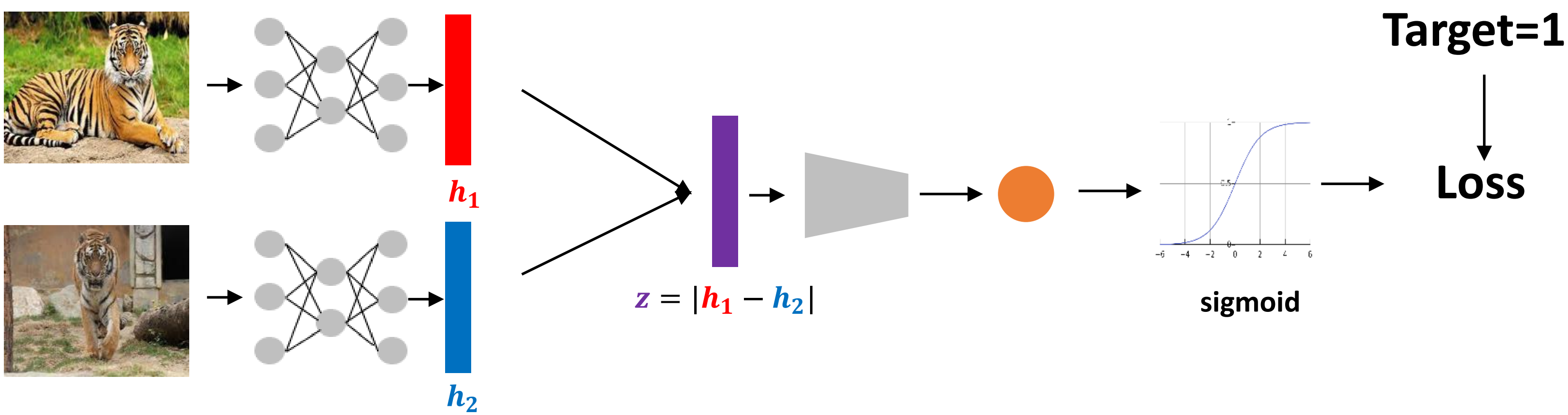
DMQA

# Introduction

Few shot learning

❖ **Few shot learning 학습 방법**

- **Positive pair**는 서로 **가까워지도록**, **negative pair**는 서로 **멀어지도록** 학습
- **Positive pair**, negative pair 에 대해서 번갈아 가며 학습




**< Positive pair >**



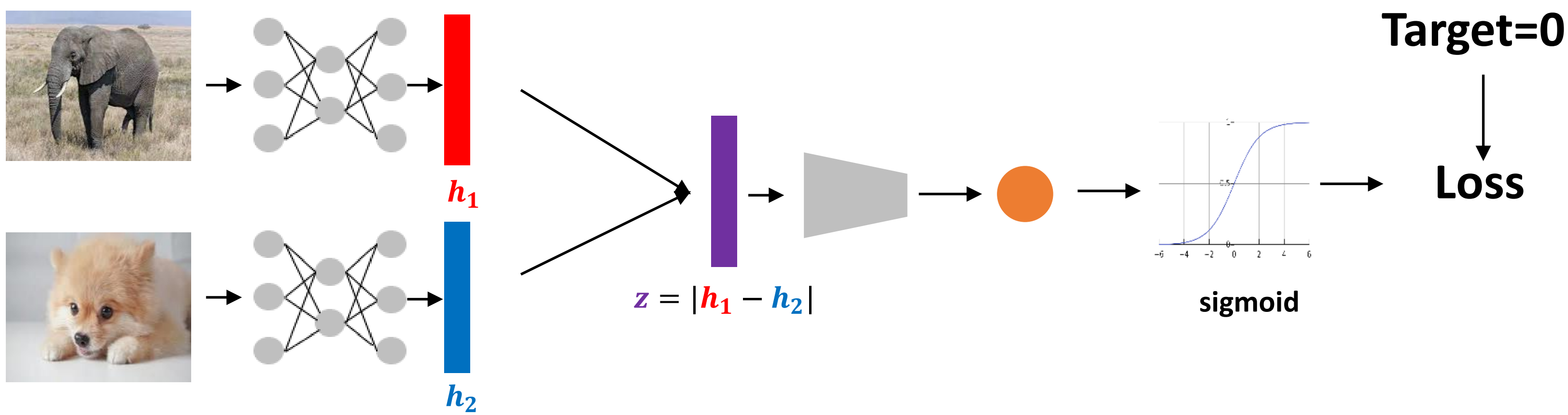
DMQA

# Introduction

Few shot learning

## ❖ Few shot learning 학습 방법

- **Positive pair**는 서로 **가까워지도록**, **negative pair**는 서로 **멀어지도록** 학습
- Positive pair, **negative pair** 에 대해서 번갈아 가며 학습




**< Negative pair >**



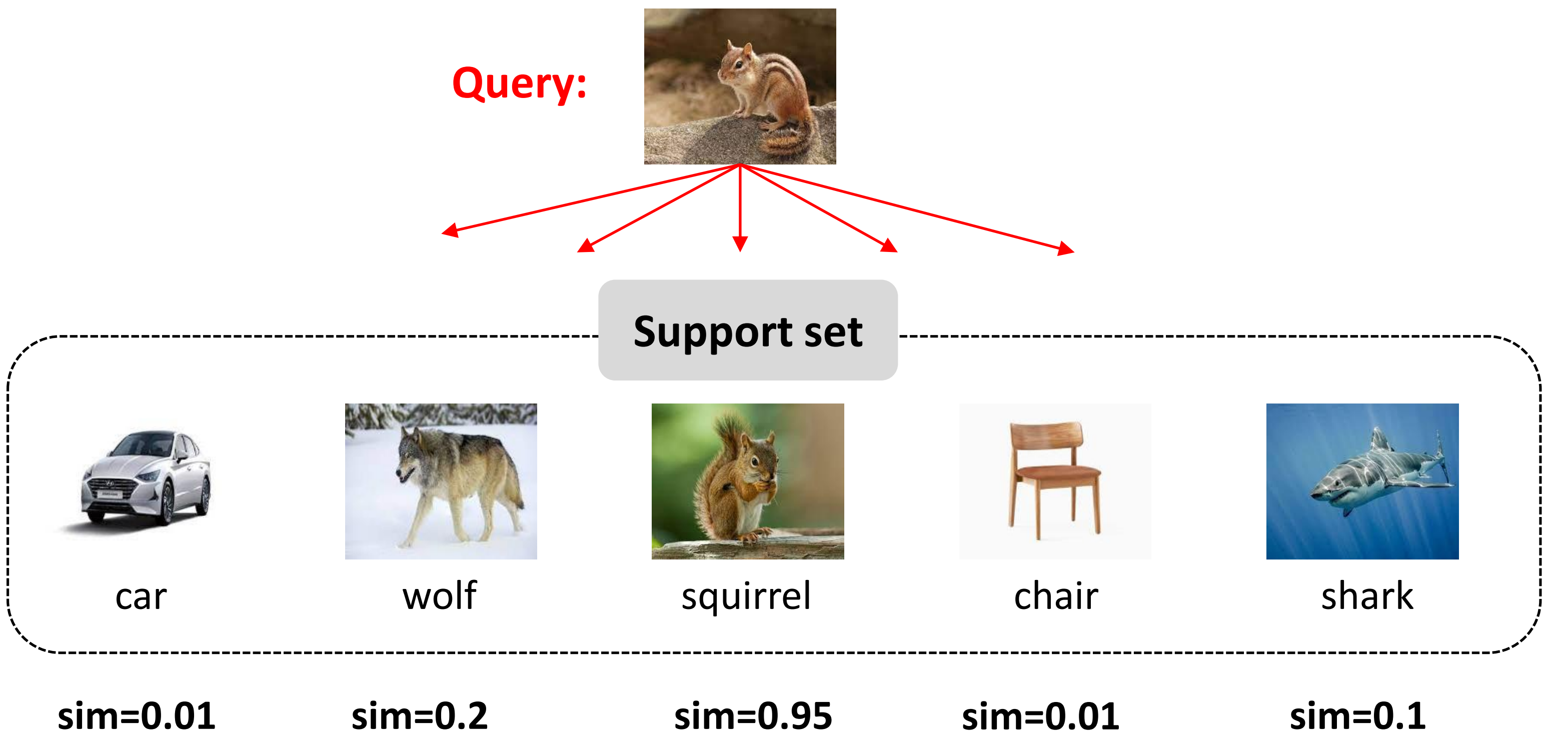
DMQA

# Introduction

Few shot learning

## ❖ Few shot learning 학습 방법

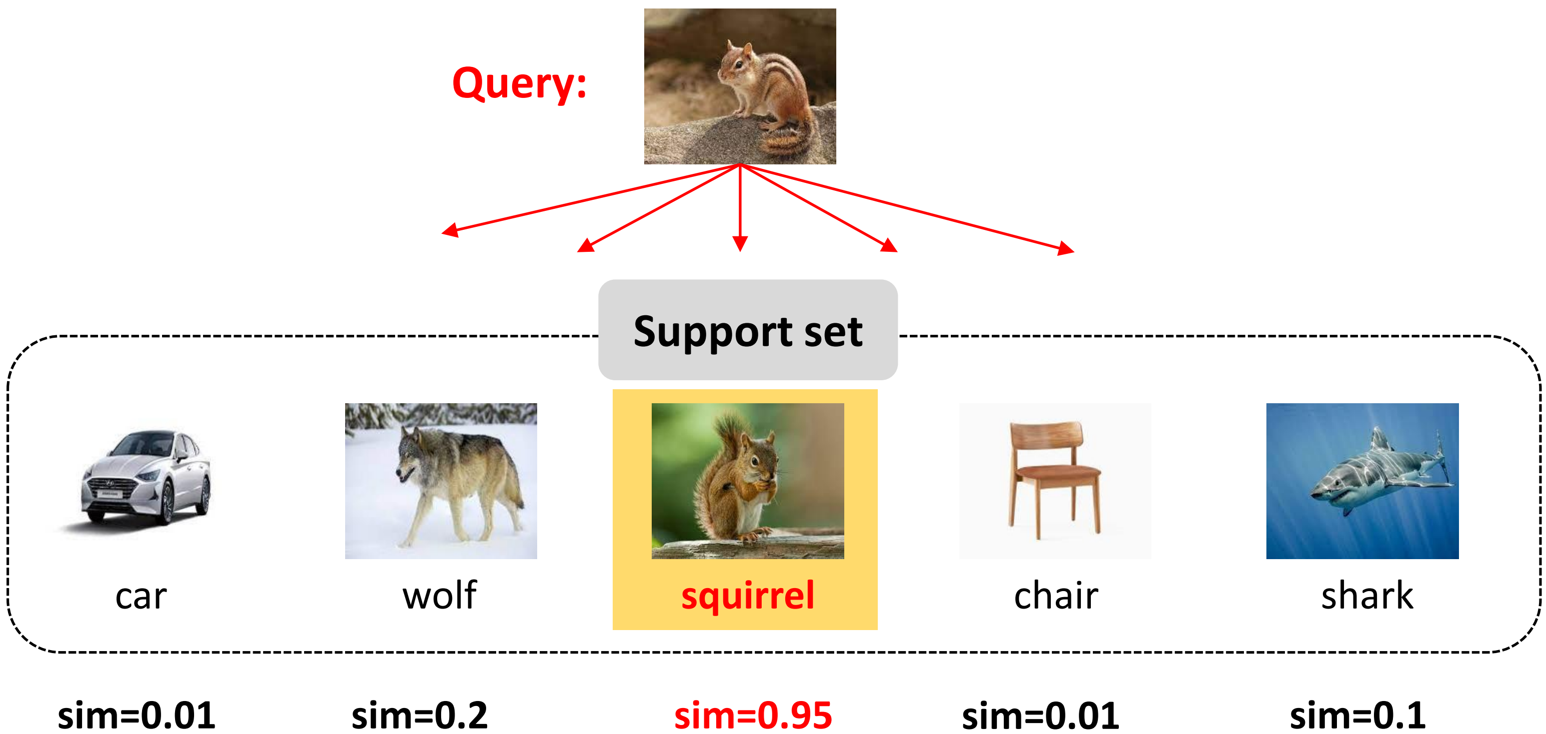
- Query 이미지와 support set의 이미지의 representation 간 차이를 통해 유사성 계산



Few shot learning

## ❖ Few shot learning 학습 방법

- Query 이미지와 support set의 이미지의 representation 간 차이를 통해 유사성 계산



# Few shot learning to Named Entity Recognition

# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

## ❖ 연구 배경

- 대부분의 언어에서 큰 규모의 labeled data 부족
- 큰 규모의 labeled data가 있어도 개체명이 거의 존재하지 않음

### Few-shot classification in Named Entity Recognition Task

Alexander Fritzler*†‡
Russia
afritzler449@gmail.com

Varvara Logacheva*
Russia
varvara.logacheva@gmail.com

Maksim Kretov*
Russia
kretov.mk@mipt.ru

**ABSTRACT**

For many natural language processing (NLP) tasks the amount of annotated data is limited. This urges a need to apply semi-supervised learning techniques, such as transfer learning or meta-learning. In this work we tackle Named Entity Recognition (NER) task using Prototypical Network — a metric learning technique. It learns intermediate representations of words which cluster well into named entity classes. This property of the model allows classifying words with extremely limited number of training examples, and can potentially be used as a zero-shot learning method. By coupling this technique with transfer learning we achieve well-performing classifiers trained on only 20 instances of a target class.

**KEYWORDS**

Named Entity Recognition, Prototypical networks, Few-shot learning, Semi-supervised learning, Transfer learning

**ACM Reference Format:**

Alexander Fritzler, Varvara Logacheva, and Maksim Kretov. 2019. Few-shot classification in Named Entity Recognition Task. In *The 34th ACM/SIGAPP Symposium on Applied Computing (SAC '19), April 8–12, 2019, Limassol, Cyprus.* ACM, New York, NY, USA, Article 4, 8 pages. https://doi.org/10.1145/3297280.3297378

for low-resourced languages. And even if we have a large labelled corpus, it will inevitably have rare entities that occur not enough times to train a neural network to accurately identify them in text.

This urges the need for developing methods of **few-shot** NER — successful identification of entities for which we have extremely small number of labelled examples. One solution would be semi-supervised learning methods, i.e. methods that can yield well-performing models by combining the information from a small set of labelled data and large amounts of unlabelled data which are available for virtually any language. Word embeddings which are trained in the unsupervised manner and are used in the majority of NLP tasks as the input to a neural network, can be considered as incorporation of unlabelled data. However, they only provide general (and not always suitable) information about word meaning, whereas we argue that unsupervised data can be used to extract more task-specific information on the structure of the data.

A prominent approach to the task of learning from few examples is metric learning [2]. This term denotes techniques that learn a metric to measure fitness of an object to some class. Metric learning methods, such as matching networks [20] and prototypical networks [18], showed good results in few-shot learning for image classification. These methods can also be considered as semi-



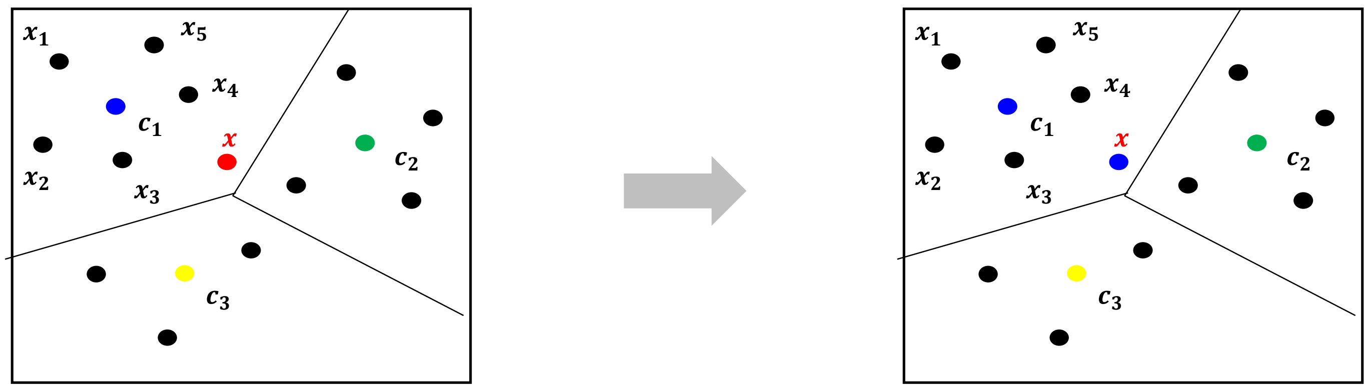
DMQA

# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

❖ **Prototypical network**

- 레이블된 데이터가 희소한 상황의 분류 문제에서 많이 사용
- K-means 클러스터링과 유사

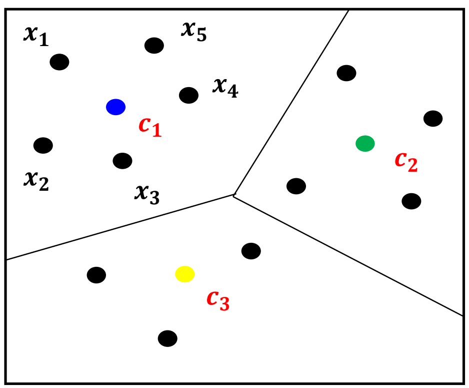


# Few shot learning to Named Entity Recognition

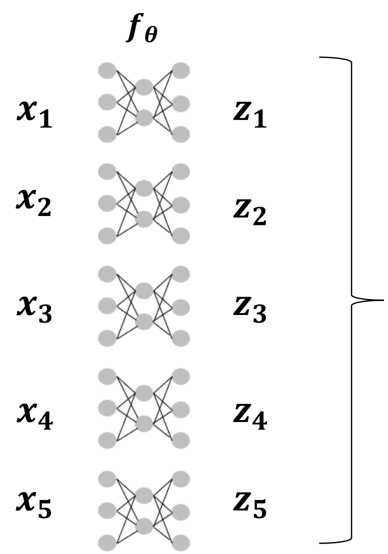
Few shot classification in Named Entity Recognition Task

## ❖ Prototypical network

- Support set 에서 각 class별로 prototype 형성




$$c_k = \frac{1}{||S_k||}\sum_i \overbrace{f_\theta(x_i)}^{z_i}$$




$$mean(z_1, \ldots, z_5) = c_1$$



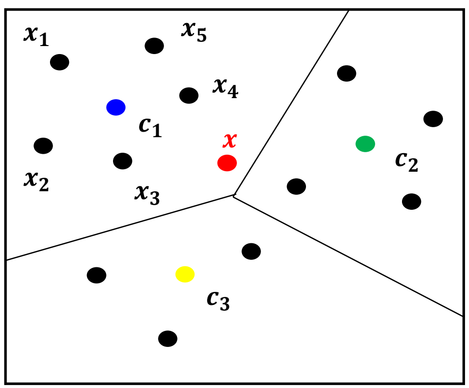
DMQA

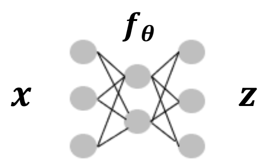
# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

## ❖ Prototypical network

- Support set 에서 각 class별로 prototype 형성
- Query set에서 x의 embedded vector(z)와 class별 prototype(c)과의 거리 계산
- 가장 가까운 class로 예측 후 cross entropy loss를 통해 학습




$$l_i = -d(z_i, c_k) \xrightarrow{softmax} L(y,\hat{y}) = -\sum_{i=1}^{N'} y_i \log \hat{y}_i$$

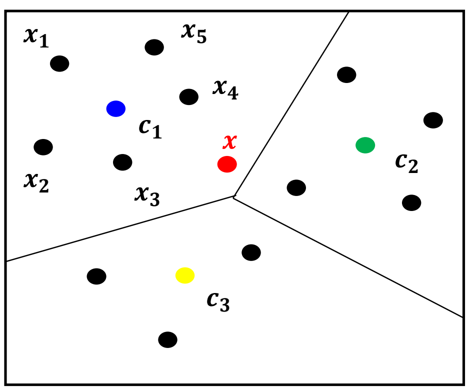
$$c_k = \frac{1}{||S_k||}\sum_i \overset{z_i}{f_\theta(x_i)}$$

# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

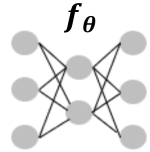
## ❖ Prototypical network

- Support set 에서 각 class별로 prototype 형성
- Query set에서 x의 embedded vector(z)와 class별 prototype(c)과의 거리 계산
- 가장 가까운 class로 예측 후 cross entropy loss를 통해 학습




$$c_k = \frac{1}{||S_k||}\sum_i \overbrace{f_\theta(x_i)}^{z_i}$$

$x$




$z$ → $l_i = -d(z_i, c_k)$ $\xrightarrow{softmax}$ $L(y,\hat{y}) = -\sum_{i=1}^{N'} y_i \log \hat{y}_i$

$c_1$ → 0.6

$c_2$ → 0.2

$c_3$ → 0.2

$x \to c_1$



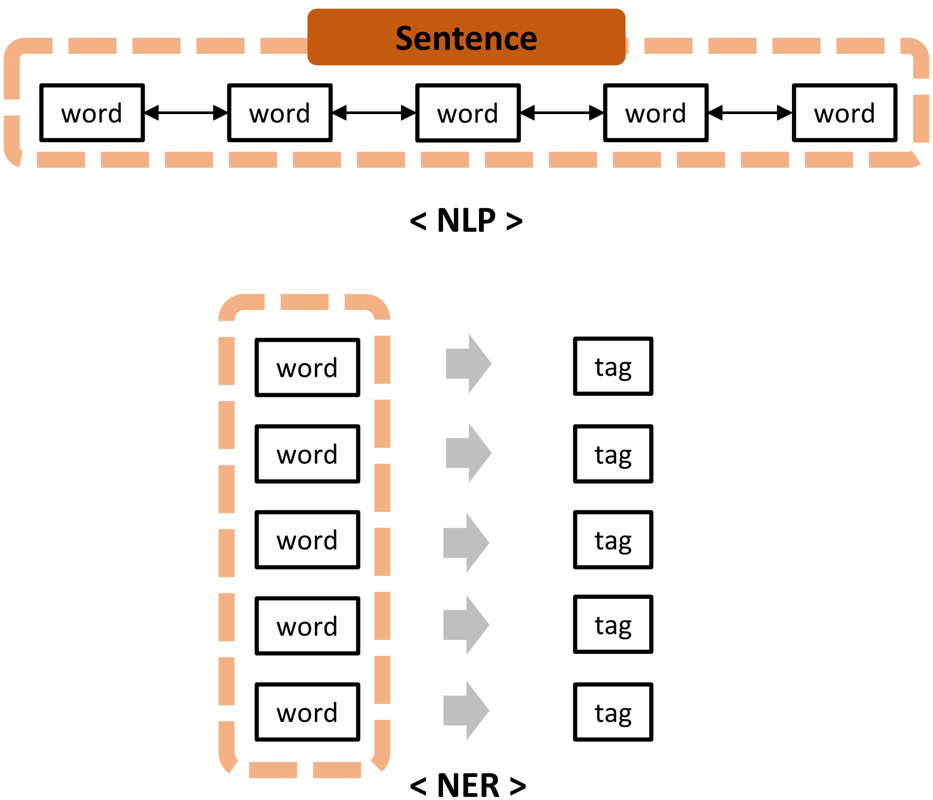
DMQA

# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

## ❖ Adaptation to NER - Sequential

- 일반적으로 NER task에서는 연속성을 반영하지 않고 **각 단어마다 classification 수행**

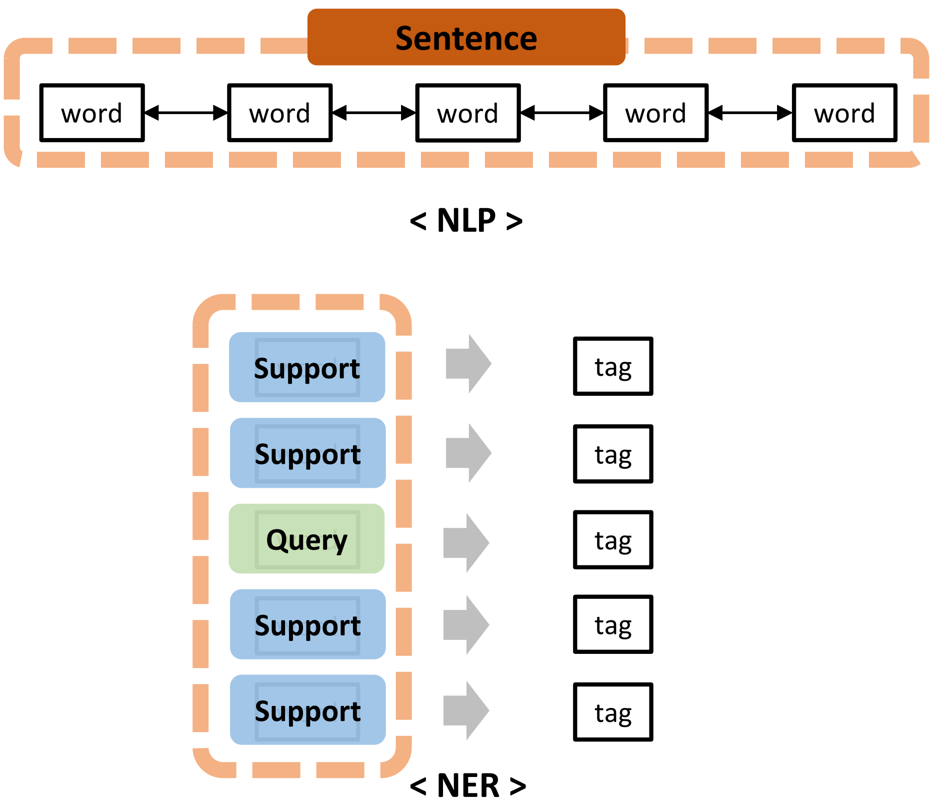


# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

## ❖ Adaptation to NER - Sequential

- 일반적으로 NER task에서는 연속성을 반영하지 않고 **각 단어마다 classification 수행**
- 한 문장에서 일부는 support set, 나머지는 query set으로 분리될 수 있음





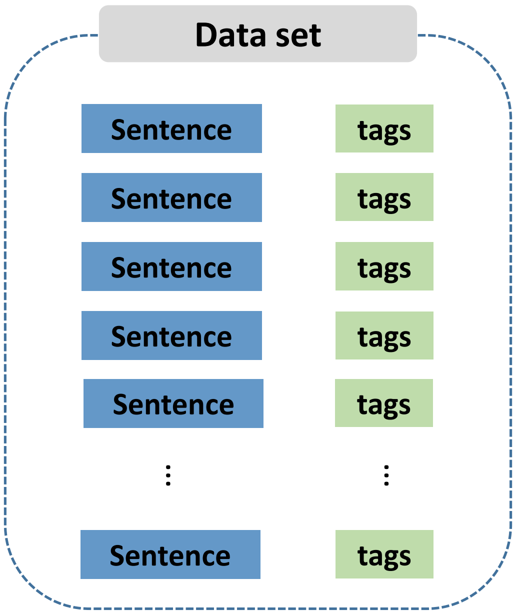

DMQA

# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

## ❖ Adaptation to NER - Sequential

- 일반적으로 NER task에서는 연속성을 반영하지 않고 **각 단어마다 classification 수행**
- 한 문장에서 일부는 support set, 나머지는 query set으로 분리될 수 있음
- Support set과 query set을 **문장 단위로 형성**





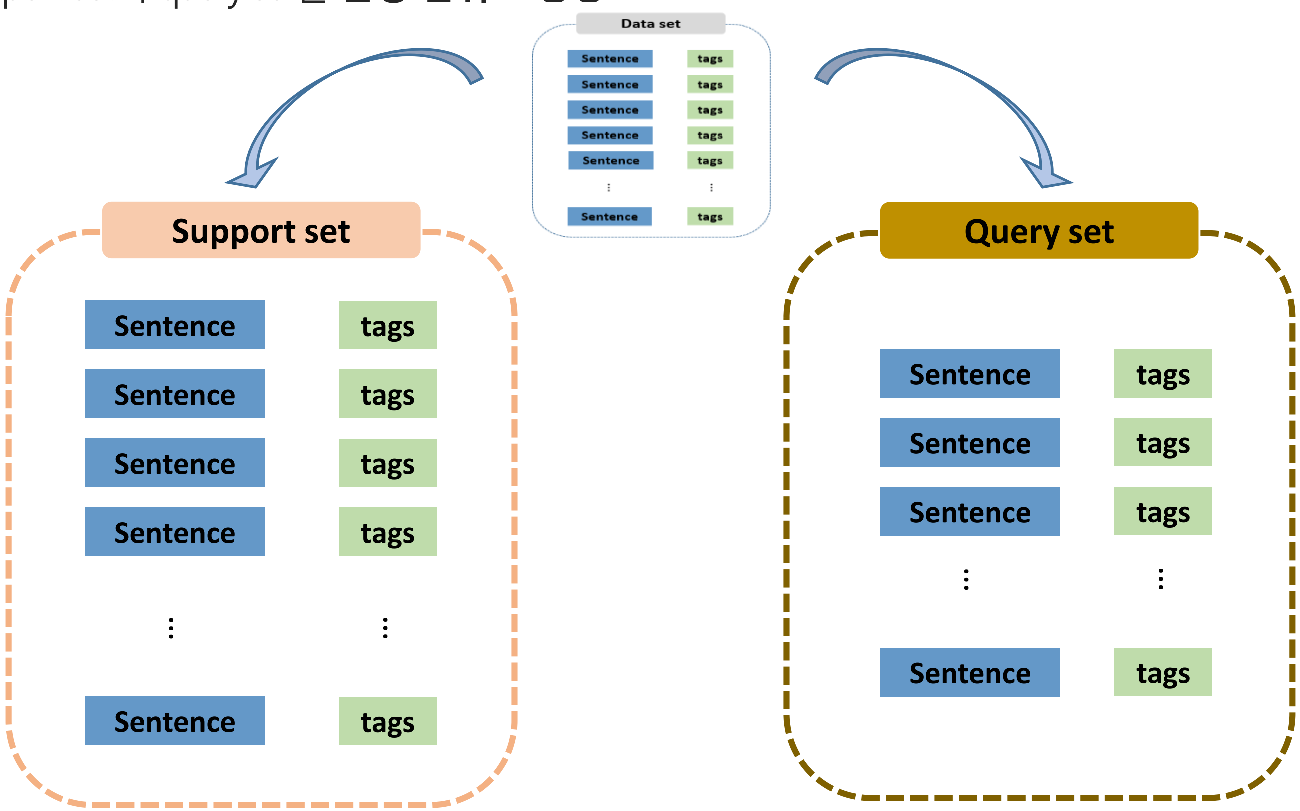

DMQA

# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

## ❖ Adaptation to NER - Sequential

- 일반적으로 NER task에서는 연속성을 반영하지 않고 **각 단어마다 classification 수행**
- 한 문장에서 일부는 support set, 나머지는 query set으로 분리될 수 있음
- Support set과 query set을 **문장 단위로 형성**

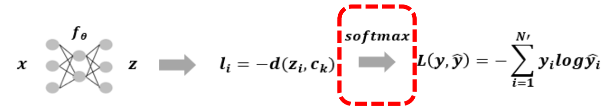





DMQA

# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task




## ❖ Adaptation to NER – "no entity" class

- 개체명과 상관 없는 단어 존재 → "O" class (no entity)

**no entity**

do excess the maybe should

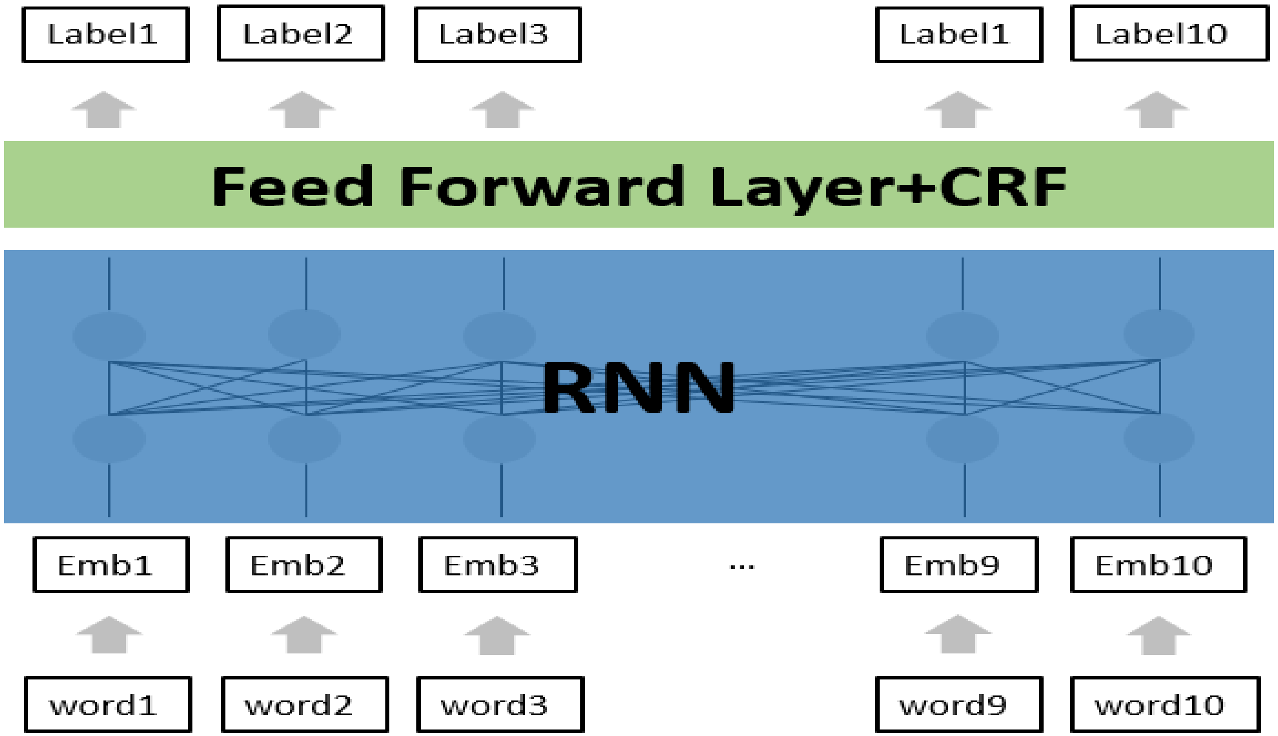
$$\boldsymbol{softmax(l_1, l_2, \dots, l_k)} \quad \Rightarrow \quad \boldsymbol{softmax(l_1, l_2, \dots, l_{k-1}, b_o)}$$

# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

## ❖ 방법론

- Baseline models : RNN, Prototypical network
- CRF : Sequence labeling을 위해 potential functions을 이용하는 softmax regression

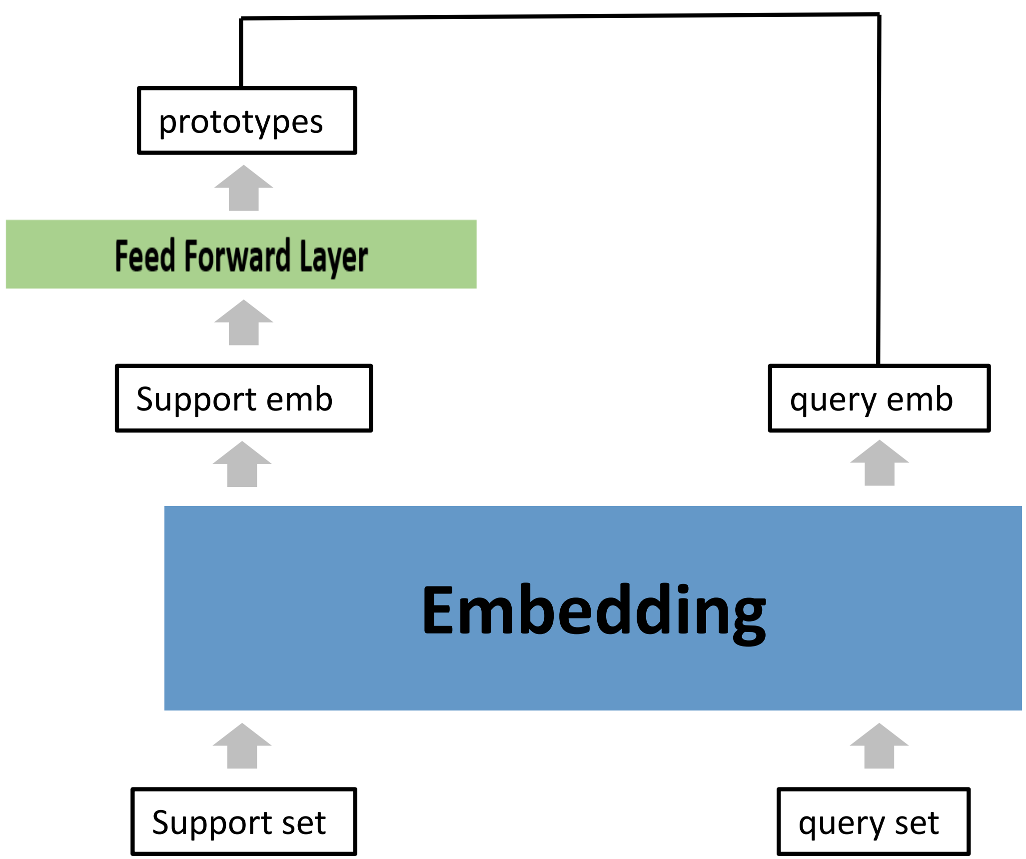


# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

## ❖ 방법론

- Baseline models : RNN, **Prototypical network**
- CRF : Sequence labeling을 위해 potential functions을 이용하는 softmax regression





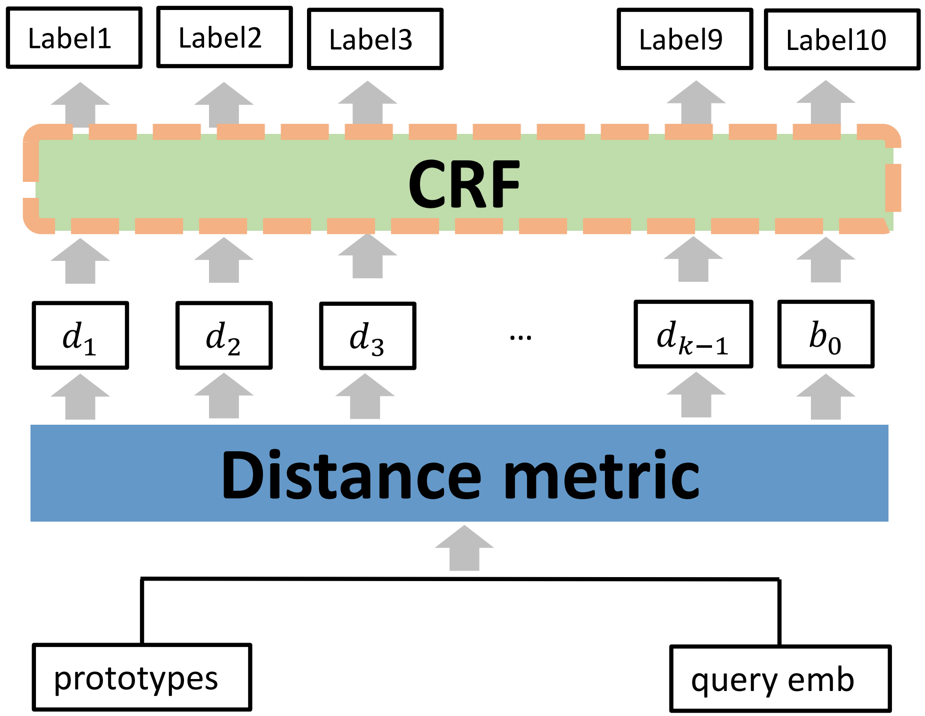

DMQA

# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

## ❖ 방법론

- Baseline models : RNN, **Prototypical network**
- CRF : Sequence labeling을 위해 potential functions을 이용하는 softmax regression

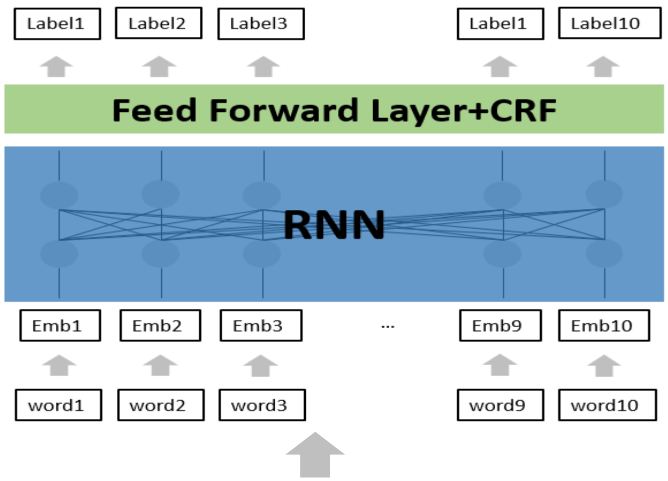


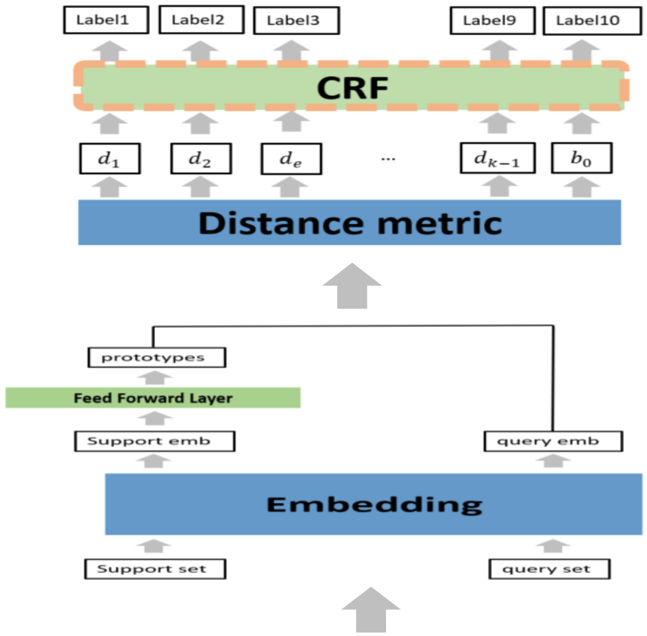
# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

## ❖ 제안 방법론

- RNN, Prototypical network + **Transfer learning**




**Pretrain with out of domain**

**Pretrain with out of domain**



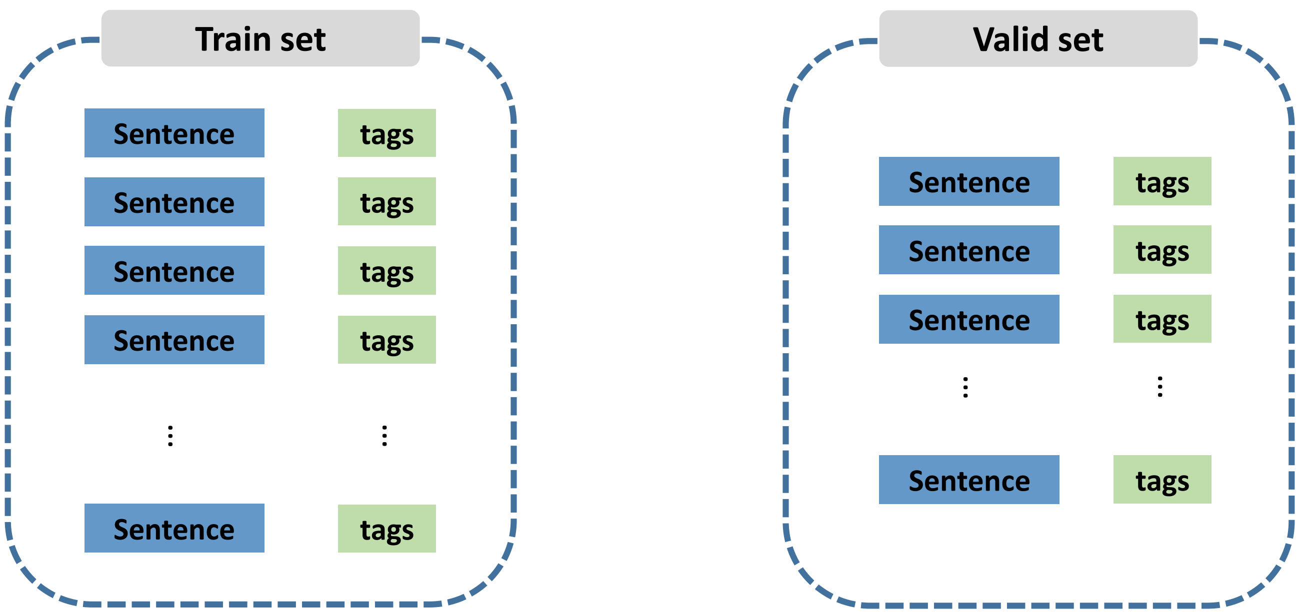
DMQA

# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

## ❖ 실험

- **Ontonotes** 데이터셋
  - ✓ 총 14개 클래스
  - ✓ 학습 문장 : 약 150개, 검증 문장: 약 19개
  - ✓ 학습 tag : 약 3만개 이상 common instance의 tag + 약 100개의 rare instance의 tag



# Few shot learning to Named Entity Recognition

Few shot classification in Named Entity Recognition Task

## ❖ 실험 결과

- 평가 척도: F1-score

| Class name | Base | BaseProto | WarmProtoZero | Protonet | WarmProto | WarmBase | WarmProto-CRF |
|---|---|---|---|---|---|---|---|
| **Validation Classes** | | | | | | | |
| GPE | 69.75 ± 9.04 | 69.8 ± 4.16 | 60.1 ± 5.56 | 78.4 ± 1.19 | **83.62 ± 3.89** | 75.8 ± 6.2 | 80.05 ± 5.4 |
| DATE | 54.42 ± 3.64 | 50.75 ± 5.38 | 11.23 ± 4.57 | 56.55 ± 4.2 | 61.68 ± 3.38 | 56.32 ± 2.32 | **65.42 ± 2.82** |
| ORG | 42.7 ± 5.54 | 39.1 ± 7.5 | 17.18 ± 3.77 | 56.35 ± 2.86 | 63.75 ± 2.43 | 63.45 ± 1.79 | **69.2 ± 1.2** |
| **Test Classes** | | | | | | | |
| EVENT | 32.33 ± 4.38 | 24.15 ± 4.38 | 4.85 ± 1.88 | 33.95 ± 5.68 | 33.85 ± 5.91 | 35.15 ± 4.04 | **45.2 ± 4.4** |
| LOC | 31.75 ± 9.68 | 24.0 ± 5.56 | 16.62 ± 7.18 | 42.88 ± 2.03 | 49.1 ± 2.4 | 40.67 ± 4.85 | **52.0 ± 4.34** |
| FAC | 36.7 ± 8.15 | 29.83 ± 5.58 | 6.93 ± 0.62 | 41.05 ± 2.74 | 49.88 ± 3.39 | 45.4 ± 3.01 | **56.85 ± 1.52** |
| CARDINAL | 54.82 ± 1.87 | 53.7 ± 4.81 | 8.12 ± 7.92 | 64.05 ± 1.61 | 66.12 ± 0.43 | 62.98 ± 3.5 | **70.43 ± 3.43** |
| QUANTITY | 64.3 ± 5.06 | 61.72 ± 4.9 | 12.88 ± 4.13 | 65.05 ± 8.64 | 67.07 ± 5.11 | 69.65 ± 5.8 | **76.35 ± 3.09** |
| NORP | 73.5 ± 2.3 | 72.1 ± 6.0 | 39.92 ± 10.5 | 83.02 ± 1.42 | **84.52 ± 2.79** | 79.53 ± 1.32 | 82.4 ± 1.15 |
| ORDINAL | 68.97 ± 6.16 | 71.65 ± 3.31 | 1.93 ± 3.25 | **76.08 ± 3.55** | 73.05 ± 7.14 | 69.77 ± 4.97 | 75.52 ± 5.11 |
| WORK_OF_ART | 30.48 ± 1.42 | 27.5 ± 2.93 | 3.4 ± 2.37 | 28.0 ± 3.33 | 23.48 ± 5.02 | 30.2 ± 1.27 | **32.25 ± 3.11** |
| PERSON | 70.05 ± 6.7 | 74.1 ± 5.32 | 38.88 ± 7.64 | 80.53 ± 2.15 | 80.42 ± 2.13 | 78.03 ± 3.98 | **82.32 ± 2.51** |
| LANGUAGE | 72.4 ± 5.53 | 70.78 ± 2.62 | 4.25 ± 0.42 | 68.75 ± 6.36 | 48.77 ± 17.42 | 65.92 ± 3.52 | **75.62 ± 7.22** |
| LAW | 58.08 ± 4.9 | 53.12 ± 4.54 | 2.4 ± 1.15 | 48.38 ± 8.0 | 50.15 ± 7.56 | **60.13 ± 6.08** | 57.72 ± 7.06 |
| MONEY | 70.12 ± 5.19 | 66.05 ± 1.66 | 12.48 ± 11.92 | 68.4 ± 6.3 | 73.68 ± 4.72 | 68.4 ± 5.08 | **79.35 ± 3.6** |
| PERCENT | 76.88 ± 2.93 | 75.55 ± 4.17 | 1.82 ± 1.81 | 80.18 ± 4.81 | 85.3 ± 3.68 | 79.2 ± 3.76 | **88.32 ± 2.76** |
| PRODUCT | 43.6 ± 7.21 | 44.35 ± 3.48 | 3.75 ± 0.58 | 39.92 ± 7.22 | 35.1 ± 9.35 | 43.4 ± 8.43 | **49.32 ± 2.92** |
| TIME | 35.93 ± 6.35 | 35.8 ± 2.61 | 8.02 ± 3.05 | 50.15 ± 5.12 | 56.6 ± 2.28 | 45.62 ± 5.64 | **59.8 ± 0.76** |

**Table 1: Results of experiments in terms of chunk-based $F_1$-score. Numbers in bold mean the best score for a particular class, underlined numbers are the second best results. Numbers are averaged across 4 runs with standard deviations calculated.**

# Few shot learning to Named Entity Recognition

CONTAINER: Few shot Named Entity Recognition via Contrastive Learning

## ❖ 연구 배경

- 기존 방법론은 few-shot 상황에서 unseen target domain에 대한 일반화 성능이 떨어짐
  - ✓ Source domain에서 클래스별 semantic feature와 중간 representation만 학습

➡ 이 문제를 해결하기 위해, **contrastive learning**을 사용한 새로운 방법론을 제안

**CONTAINER: Few-Shot Named Entity Recognition via Contrastive Learning**

**Sarkar Snigdha Sarathi Das, Arzoo Katiyar, Rebecca J. Passonneau, Rui Zhang**
Pennsylvania State University
{sfd5525, arzoo, rjp49, rmz5227}@psu.edu

**Abstract**

Named Entity Recognition (NER) in Few-Shot setting is imperative for entity tagging in low resource domains. Existing approaches only learn *class-specific* semantic features and intermediate representations from source domains. This affects generalizability to unseen target domains, resulting in suboptimal performances. To this end, we present CONTAINER, a novel contrastive learning technique that optimizes the inter-token distribution distance for Few-Shot NER. Instead

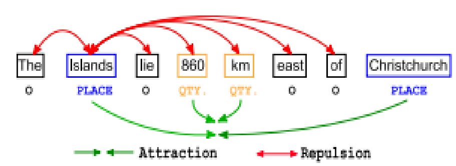



Figure 1: Contrastive learning dynamics of a token (*Islands*) with all other tokens in an example sentence from GUM (Zeldes, 2017). CONTAINER decreases the embedding distance between tokens of the same category (PLACE) while increasing the distance between different categories ( QTY. and O).



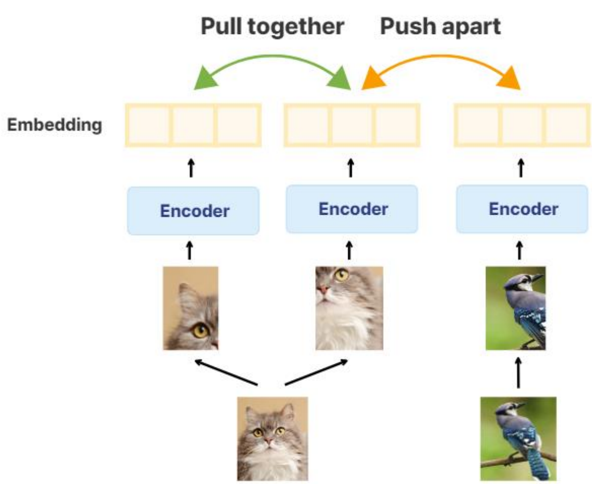
DMQA

# Few shot learning to Named Entity Recognition

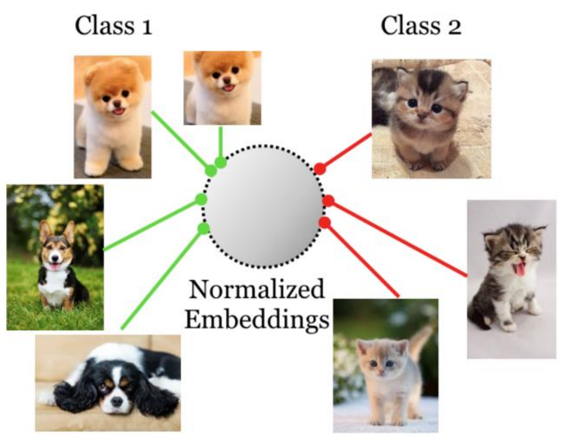
CONTAINER: Few shot Named Entity Recognition via Contrastive Learning

## ❖ Contrastive learning

- 입력 샘플 간의 비교를 통해 학습을 하는 방법
  - ✓ Positive pair 간 임베딩 거리는 가깝게, negative pair 간 임베딩 거리는 멀게 학습
- 주로 self-supervised learning에서 사용되지만 supervised learning의 맥락에서도 수행됨




< Self-supervised contrastive learning >




< Supervised contrastive learning >



DMQA

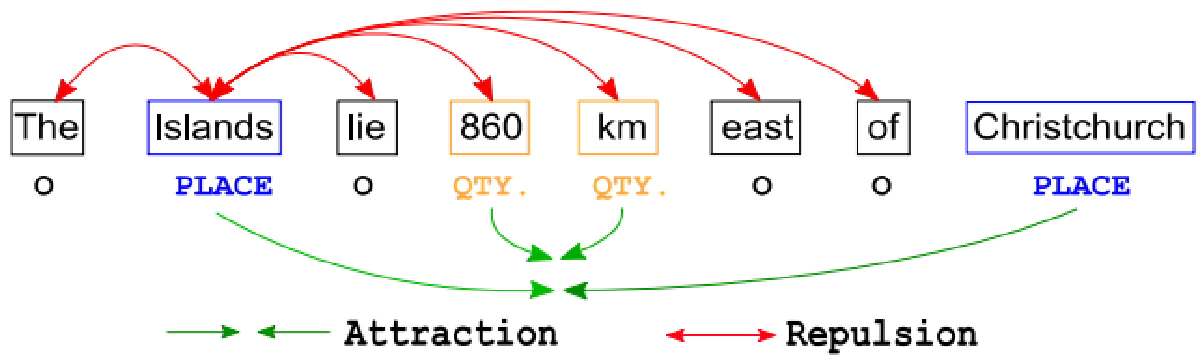
# Few shot learning to Named Entity Recognition

CONTAINER: Few shot Named Entity Recognition via Contrastive Learning

## ❖ Contrastive learning

- **문장 내 토큰 간 contrastive learning:**

  같은 개체명에 속하는 토큰 간 임베딩 거리는 가깝게, 다른 개체명에 속하는 토큰 간 임베딩 거리는 멀게 학습



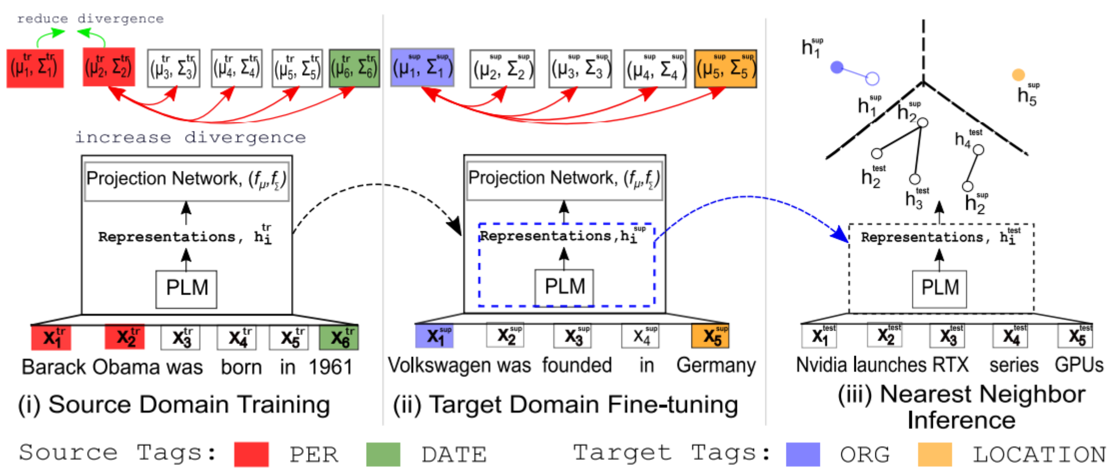

**< 문장 내 토큰 간 contrastive learning >**

# Few shot learning to Named Entity Recognition

CONTAINER: Few shot Named Entity Recognition via Contrastive Learning

## ❖ 제안 방법론

- **가우시안 임베딩에 대한 contrastive learning에 기반한 CONTAINER**
  1) Source domain에서 모델 학습
  2) Target domain에서 모델 representation을 fine-tuning
  3) Nearest neighbor classifier 사용하여 inference



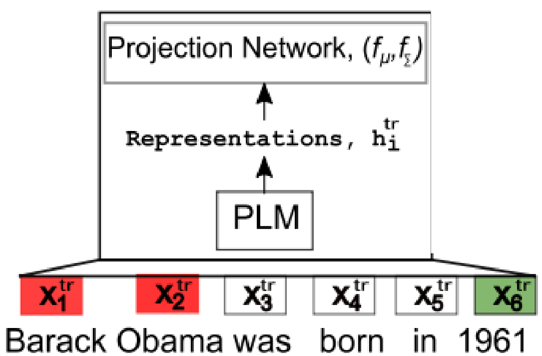

**< Overall framework >**

# Few shot learning to Named Entity Recognition

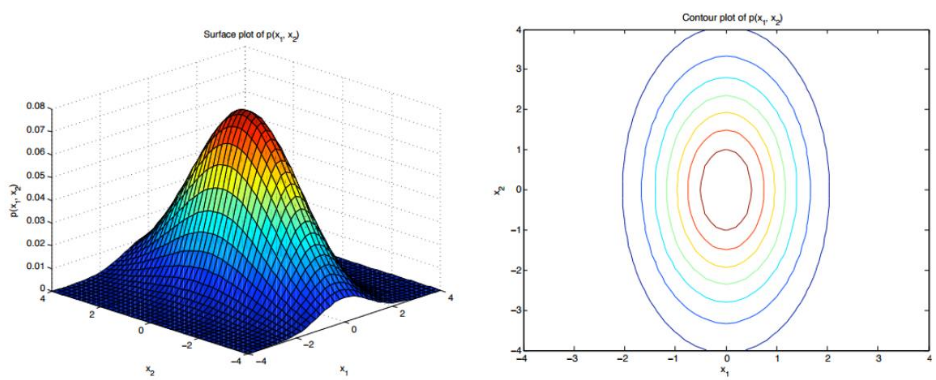
CONTAINER: Few shot Named Entity Recognition via Contrastive Learning

## ❖ 제안 방법론 – 가우시안 임베딩 추출

- 토큰 임베딩이 가우시안 분포를 따른다고 가정
  - ✓ PLM 인코더(BERT)를 사용해 중간 representation 추출
  - ✓ 중간 representation에서 projection network를 통해 가우시안 임베딩의 평균과 대각 공분산을 얻음




**< 가우시안 임베딩 추출 >**




**< Gaussian with diagonal covariance matrix >**

# Few shot learning to Named Entity Recognition

CONTAINER: Few shot Named Entity Recognition via Contrastive Learning

## ❖ 제안 방법론 – Training in source domain

- Contrastive loss를 구하기 위해서는 임베딩 간 거리를 계산해야 함
- 가우시안 임베딩 간 거리를 계산하기 위해 KL-divergence를 사용
- KL-divergence: 분포 간 거리를 계산할 수 있는 지표

**두 분포가 가까울 때:**
**KL divergence는 작은 값을 가짐**

**두 분포가 멀 때:**
**KL divergence는 큰 값을 가짐**

**< KL divergence 예시 >**

# Few shot learning to Named Entity Recognition

CONTAINER: Few shot Named Entity Recognition via Contrastive Learning

## ❖ 제안 방법론 – Training in source domain

- $D_{KL}[\mathcal{N}_q||\mathcal{N}_p]$: 가우시안 임베딩 $\mathcal{N}(\mu_p, \Sigma_p)$와 $\mathcal{N}(\mu_q, \Sigma_q)$ 간 KL-divergence
- KL-divergence 값은 symmetric하지 않기 때문에 (1)과 같이 양방향을 모두 구함
- 한 배치 내에서 contrastive loss (2)를 최소화하도록 학습

$$d(p,q) = \frac{1}{2}\left(D_{KL}[N_q||N_p] + D_{KL}[N_p||N_q]\right) \quad \textbf{(1)}$$

**Positive pair 간 임베딩은 가깝게**

$$l(p) = -log\frac{\boxed{\sum_{(x_q,y_q)\in\chi_p} \exp(-d(p,q)/|\chi_p|)}}{\boxed{\sum_{(x_q,y_q)\in\chi,p\neq q} \exp(-d(p,q))}} \quad \textbf{(2)}$$

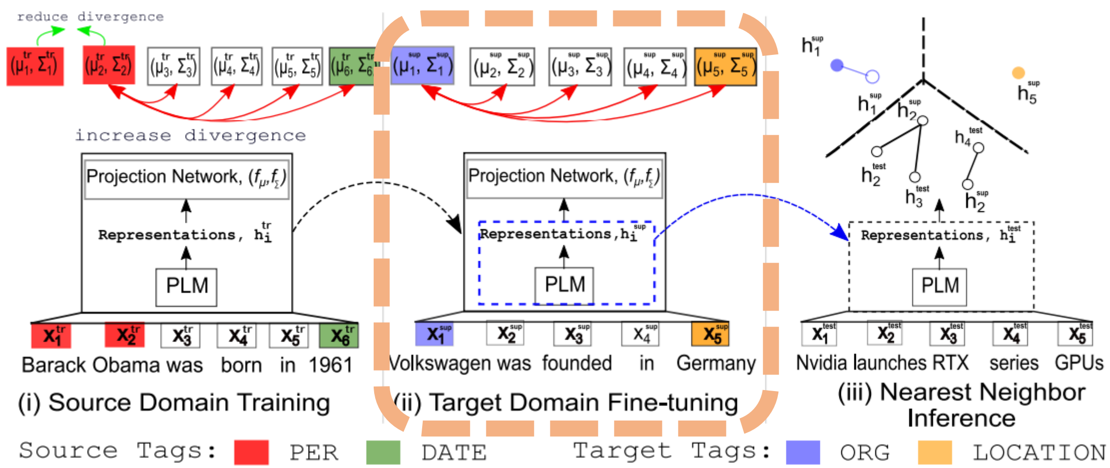
**Negative pair 간 임베딩은 멀게**

# Few shot learning to Named Entity Recognition

CONTAINER: Few shot Named Entity Recognition via Contrastive Learning

❖ **제안 방법론 – Finetuning to target domain using support set**

- Source domain에서 학습 후 유사한 절차에 따라 target domain의 support sample들로 fine-tuning 수행
- Source domain과 target domain의 클래스 레이블이 겹치지 않을 수 있음






DMQA

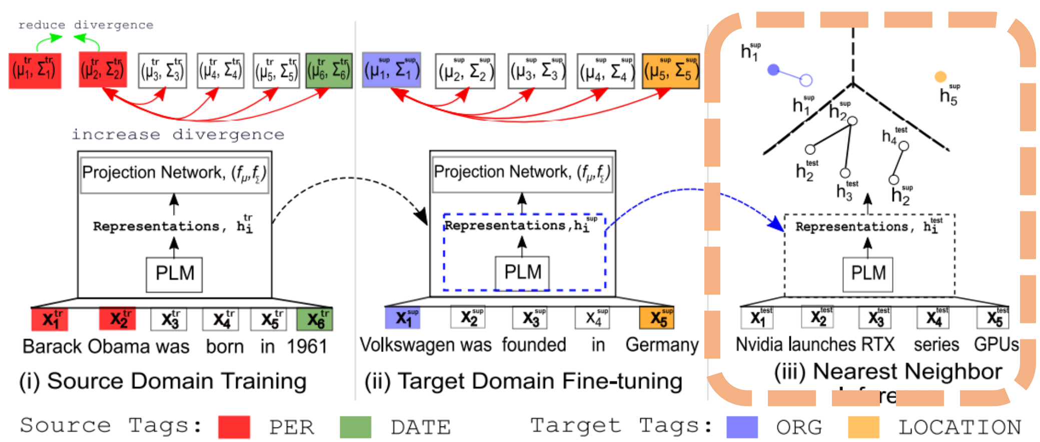
# Few shot learning to Named Entity Recognition

CONTAINER: Few shot Named Entity Recognition via Contrastive Learning

## ❖ 제안 방법론 – Instance level nearest neighbor inference

- Inference 단계에서는 projection head를 사용하지 않고 중간 representation만 사용
- Test data의 representation과 가장 가까이 위치한 support set의 representation을 찾아 해당 클래스로 예측
- Viterbi decoding 사용하여 sequence labeling

$$y_k^{test} = argmin \, ||h_i^{test} - h_k^{sup}||_2^2$$






DMQA

# Few shot learning to Named Entity Recognition

CONTAINER: Few shot Named Entity Recognition via Contrastive Learning

## ❖ 실험

- 데이터셋

| Dataset | Domain | # Class | # Sent |
|---|---|---|---|
| OntoNotes | General | 18 | 76K |
| I2B2'14 | Medical | 23 | 140K |
| CoNLL'03 | News | 4 | 20K |
| WNUT'17 | Social | 6 | 5K |
| GUM | Mixed | 11 | 3.5K |
| FEW-NERD | Wikipedia | 66 | 188K |



DMQA

# Few shot learning to Named Entity Recognition

CONTAINER: Few shot Named Entity Recognition via Contrastive Learning

## ❖ 실험 결과

- 평가 척도: F1-score

| Model | 1-shot | | | | 5-shot | | | |
|---|---|---|---|---|---|---|---|---|
| | Group A | Group B | Group C | Avg. | Group A | Group B | Group C | Avg. |
| Proto | 19.3 ± 3.9 | 22.7 ± 8.9 | 18.9 ± 7.9 | 20.3 | 30.5 ± 3.5 | 38.7 ± 5.6 | 41.1 ± 3.3 | 36.7 |
| NNShot | 28.5 ± 9.2 | 27.3 ± 12.3 | 21.4 ± 9.7 | 25.7 | 44.0 ± 2.1 | 51.6 ± 5.9 | 47.6 ± 2.8 | 47.7 |
| StructShot | 30.5 ± 12.3 | 28.8 ± 11.2 | 20.8 ± 9.9 | 26.7 | 47.5 ± 4.0 | 53.0 ± 7.9 | 48.7 ± 2.7 | 49.8 |
| **CONTaiNER** | **32.2 ± 5.3** | **30.9 ± 11.6** | **32.9 ± 12.7** | **32.0** | **51.2 ± 5.9** | **55.9 ± 6.2** | **61.5 ± 2.7** | **56.2** |
| **+ Viterbi** | **32.4 ± 5.1** | **30.9 ± 11.6** | **33.0 ± 12.8** | **32.1** | **51.2 ± 6.0** | **56.0 ± 6.2** | **61.5 ± 2.7** | **56.2** |

Table 2: F1 scores in Tag Set Extension on OntoNotes. Group A, B, C are three disjoint sets of entity types. Results vary slightly compared to Yang and Katiyar (2020) since they used different support set samples (publicly unavailable) than ours.

| Model | 1-shot | | | | | 5-shot | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | I2B2 | CoNLL | WNUT | GUM | Avg. | I2B2 | CoNLL | WNUT | GUM | Avg. |
| Proto | 13.4 ± 3.0 | 49.9 ± 8.6 | 17.4 ± 4.9 | 17.8 ± 3.5 | 24.6 | 17.9 ± 1.8 | 61.3 ± 9.1 | 22.8 ± 4.5 | 19.5 ± 3.4 | 30.4 |
| NNShot | 15.3 ± 1.6 | 61.2 ± 10.4 | 22.7 ± 7.4 | 10.5 ± 2.9 | 27.4 | 22.0 ± 1.5 | 74.1 ± 2.3 | 27.3 ± 5.4 | 15.9 ± 1.8 | 34.8 |
| StructShot | 21.4 ± 3.8 | **62.4 ± 10.5** | 24.2 ± 8.0 | 7.8 ± 2.1 | 29.0 | 30.3 ± 2.1 | 74.8 ± 2.4 | 30.4 ± 6.5 | 13.3 ± 1.3 | 37.2 |
| **CONTaiNER** | 16.4 ± 1.7 | 57.8 ± 10.7 | 24.2 ± 2.9 | 17.9 ± 1.8 | 29.1 | 24.1 ± 1.9 | 72.8 ± 2.0 | 27.7 ± 2.2 | 24.4 ± 2.2 | 37.3 |
| **+ Viterbi** | **21.5 ± 1.7** | 61.2 ± 10.7 | **27.5 ± 1.9** | **18.5 ± 4.9** | **32.2** | **36.7 ± 2.1** | **75.8 ± 2.7** | **32.5 ± 3.8** | **25.2 ± 2.7** | **42.6** |

Table 3: F1 scores in Domain Extension with OntoNotes as the source domain. Results vary slightly compared to Yang and Katiyar (2020) since they used different support set samples (publicly unavailable) than ours.



DMQA

# Conclusion

# Conclusion

❖ **Summary**

- NER task의 타겟 도메인의 레이블이 적은 상황에서 few-shot learning 필요성 대두
- Few-shot learning
  - ✓ "Few" 한 양의 데이터로 모델을 학습하여 테스트 데이터에서 유의미한 성능을 내고자 하는 방법
- NER task에 few-shot learning을 접목시킨 방법론 소개
  - ✓ **Few shot classification in Named Entity Recognition Task**
    - ➢ 이미지에 쓰이던 prototypical network를 NER task에 적용함으로써 성능 개선
  - ✓ **CONTAINER: Few shot Named Entity Recognition via Contrastive Learning**
    - ➢ Few-shot NER에 contrastive learning을 접목시킨 framework 제안함으로써 성능 개선



DMQA

# References

# References

[1] Das, S. S. S., Katiyar, A., Passonneau, R. J., & Zhang, R. (2021). Container: Few-shot named entity recognition via contrastive learning. *arXiv preprint arXiv:2109.07589*.

[2] Fritzler, A., Logacheva, V., & Kretov, M. (2019, April). Few-shot classification in named entity recognition task. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing* (pp. 993-1000).

[3] Huang, J., Li, C., Subudhi, K., Jose, D., Balakrishnan, S., Chen, W., ... & Han, J. (2020). Few-shot named entity recognition: A comprehensive study. *arXiv preprint arXiv:2012.14978*.

[4] Wang, Y., Yao, Q., Kwok, J. T., & Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, *53*(3), 1-34.



DMQA

# 감사합니다