

# The Natural Language Decathlon: Multitask Learning as Question Answering – DecaNLP를 중심으로

이번 게시물에서는 DecaNLP와 MQAN을 통해 NLP에서 multitask learning의 가능성을 보여준 논문인 The Natural Language Decathlon: Multitask Learning as Question Answering을 리뷰한다.

원문 링크는 다음과 같다.

[The Natural Language Decathlon: Multitask Learning as Question Answering](#)

## Introduction

논문에서는 다양한 종류의 nlp task를 수행할 수 있는 generalized model을 찾기 위해 Natural Language Decathlon(decaNLP)를 소개한다. **decaNLP**는 **single model**이 동시에 **10개의 task**에 최적화할 수 있게끔 한다.

(해당 10개의 task는 question answering, machine translation, document summarization, semantic parsing, sentiment analysis, natural language inference, semantic role labeling, relation extraction, goal oriented dialogue, and pronoun resolution이다.)

저자들은 이러한 task들을 아래 자료에서 보여주는 것과 같이 question, context, answer를 가진 question answering의 형태로 변형하였다.

Examples		
Question	Context	Answer
What is a major importance of Southern California in relation to California and the US?	...Southern California is a major economic center for the state of California and the US....	major economic center
What is the translation from English to German?	Most of the planet is ocean water.	Der Großteil der Erde ist Meerwasser
What is the summary?	Harry Potter star Daniel Radcliffe gains access to a reported £320 million fortune...	Harry Potter star Daniel Radcliffe gets £320M fortune...
Hypothesis: Product and geography are what make cream skimming work. Entailment, neutral, or contradiction?	Premise: Conceptually cream skimming has two basic dimensions – product and geography.	Entailment
Is this sentence positive or negative?	A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film.	positive

Question	Context	Answer
What has something experienced?	Areas of the Baltic that have experienced eutrophication.	eutrophication
Who is the illustrator of Cycle of the Werewolf?	Cycle of the Werewolf is a short novel by Stephen King, featuring illustrations by comic book artist Bernie Wrightson.	Bernie Wrightson
What is the change in dialogue state?	Are there any Eritrean restaurants in town?	food: Eritrean
What is the translation from English to SQL?	The table has column names... Tell me what the notes are for South Australia	SELECT notes from table WHERE 'Current Slogan' = 'South Australia'
Who had given help? Susan or Joan?	Joan made sure to thank Susan for all the help she had given.	Susan

전통적으로, NLP example은 input  $x$ , output  $y$ , 그리고 modeling의 제약을 통해 model이 수행하는 underlying task  $t$ 로 구성되었다. Meta-learning 접근 방식은 input에 추가적으로 underlying task  $t$ 를 넣기도 하였다.

저자들은 이러한 전통적인 NLP example에서 벗어나, 본 연구에서는 더 이상 task  $t$ 에 대한 single representation을 사용하지 않고, 대신에 underlying task에 대한 description을 제공하는 natural language question을 이용한다고 밝힌다. 이러한 방법론은 model이 효과적으로 multitask를 수행하는데 도움을 주며, pre-trained model을 transfer learning과 meta learning에 더 적합하게 만든다고 주장한다.

즉, 정리하자면 natural language question은 다르지만, 그렇지만 task와는 관련된 description을 통해 model이 새로운 task에 대해 generalize 하게끔 한다고 주장하는 것이다.

이어서, decaNLP를 위해 설계된 multitask question answering network인 MQAN을 소개한다. decaNLP로 pre-train 된 MQAN은 machine translation, named entity recognition, domain adaptation for sentiment analysis and natural language inference task에 대한 transfer learning에서 향상된 성능을 보여주었고, text classification에 대한 zero-shot learning의 가능성을 보

여주었다고 말한다.

## Tasks and Metrics

decaNLP는 10개의 공개적으로 사용 가능한 dataset으로 구성되어 있으며, 각각의 example들은 위에서 언급한 것처럼 (question, context, answer)의 형태로 변환된다.

이후 논문에서는 각각의 task에 대해 설명하는데, 하나씩 살펴보도록 하겠다.

### Question Answering

Question Answering model의 경우 question과 바람직한 answer를 출력하기 위해 필요한 정보를 담고 있는 context를 입력으로 받는다. decaNLP에서는 SQuAD dataset을 이용하였다. normalized F1 score(nF1)를 통해 evaluation 한다.

### Machine Translation

Machine translation model은 source language의 input document를 input으로 받고, target language로 번역해준다.

decaNLP에서는 IWSLT dataset을 사용하였고, corpus-level BLUE score로 평가를 진행하였다.

### Summarization

Summarization model은 document를 input으로 받아 해당 document의 summary를 출력한다. decaNLP에서는 CNN/DM corpus를 summarization dataset으로 변환한 것의 non-anonymized version을 사용하였다. 해당 dataset의 example들은 평균적으로 decaNLP에서 가장 긴 document를 가지고 있었으며, 이는 model이 context에서 extracting 하는 것과 새롭고 추상적인 sequence of word를 생성하는 것의 균형을 유지하게끔 한다.

Evaluation의 경우, ROUGE-1과 ROUGE-2, ROUGE-L score를 모두 사용하여 3개의 score를 평균을 냄으로써 수행하였다.

## Natural Language Inference

Natural language inference model은 premise와 hypothesis의 2개의 input sentence를 받는다. 이후 두 문장 간의 관계를 살펴 premise(전제)가 hypothesis(가설)를 entail 하면 entailment, contradict 하면 contradiction, 둘 다 아니면 neutral의 label로 predict 한다.

(여기서의 entail은 "함의"의 뜻이며, premise가 사실일 때 hypothesis의 사실을 보장하는 것을 의미한다. contradict의 경우에는 premise와 hypothesis가 서로 모순된다는 의미이다.)

decaNLP에서는 Multi-Genre Natural Language Inference Corpus(MNLI)를 사용하였으며 exact match score(EM score)로 evaluation 하였다.

## Sentiment Analysis

Sentiment analysis model은 input text에 나타나 있는 감정을 분류하도록 학습된다. decaNLP에서는 영화 리뷰에 대해 positive, neutral, negative의 감정으로 labeling 된 Stanford Sentiment Treebank(SST) dataset을 이용하였으며, unparsed, binary version을 사용했다고 한다.

## Semantic Role Labeling

Semantic role labeling(SRL) model은 sentence와 predicate(술어, 논문에서는 일반적으로 동사가 주어진다고 한다.)가 주어진 이후, "누가", "누구에게", "무엇을" "했는지", "언제", "어디서"를 결정한다. decaNLP에서는 SRL dataset으로 SRL task를 question answering task처럼 처리하는 QA-SRL dataset을 사용하였다고 한다. SQuAD와 마찬가지로 normalized F1 score(nF1)를 통해 evaluation 한다.

## Relation Extraction

Relation extraction system은 unstructured text의 한 부분과 해당 text에서 추출될 kind of relation을 input으로 받는다. 이때, model이 관계가 없으므로 추출할 수 없다고 출력하는 능력이 중요하다.

decaNLP에서는 SRL task와 비슷하게 relation extraction task를 question answering task처럼 처리하는 QA-ZRE dataset을 사용한다. 이 dataset의 evaluation은 새로운 종류의 relation에서의 zero-shot performance를 측정하도록 설계되었는데, test time에 보이는 relation이 train time에서는 unseen 되도록 dataset을 나누어 이를 가능케 한다. 이러한 방식은 새로운 relation에 대해 model이 generalize 할 수 있게끔 한다.

QA-ZRE dataset은 corpus-level F1 score로 evaluate 하는데, 이는 특이한 성질을 가지고 있다.

다음은 corpus-level F1 score가 precision과 recall을 계산하는 방법이다.

- precision :  $\text{positive count} / \text{number of times the system returned a non-null answer}$
- recall :  $\text{true positive count} / \text{number of instances that have answer}$

이러한 계산식을 통해 산출된 precision과 recall을 가지고 f1 score를 산출하는 것이다.

## Goal-Oriented Dialogue

Dialogue state tracking은 goal-oriented dialogue system의 핵심 요소이다. 사용자의 발언, 이미 취한 행동, 대화 기록을 바탕으로 dialogue state tracker는 user가 dialogue system에 어떠한 사전 정의된 목표를 가지고 있는지와 system과 user가 차례대로 상호 작용을 할 때 어떤 종류의 요청을 하는지 추적한다. decaNLP에서는 English Wizard of Oz(WOZ) dataset을 사용하며, 해당 dataset은 customer의 목표에 대한 turn-based dialogue state EM(dsEM)을 통해 evaluate 된다.

## Semantic Parsing

Semantic parsing은 SQL query 생성과 연관이 있다. Semantic parsing model은 WikiSQL dataset을 기반으로 동작하며, 자연어로 구성된 question을 structured SQL query로 변환하여 user로 하여금 자연어로 database와 상호작용할 수 있게 한다. WikiSQL dataset은 logical form exact match(lfEM)으로 evaluate 된다.

## Pronoun Resolution

Pronoun Resolution은 Winograd schemas를 기반으로 한다. Winograd Schema는 reading comprehension task이다. 이 reading comprehension task란, model이 pronoun(대명사)가 있는 문장을 읽고 해당 pronoun(대명사)의 referent(참조)를 선택 목록 중에서 선택해야 하는 task인 것이다.

decaNLP에서는 Winograd Schema Challenge의 example들을 answer가 context로부터 주어지는 single word가 되도록 수정한

modified Winograd Schema Challenge (MWSC)를 사용하였고, 이 MWSC는 context, question, answer 간 불일치 혹은 표현의 특이함으로 인해 score가 상승하거나 하락하지 않도록 한다. 또한 MWSC는 EM score로 evaluate 된다.

## The Decathlon Score

지금까지 살펴본 10개의 task에 대해, model은 각 task 별 metric의 additive combination으로 평가 되고, 모든 metric은 0과 100 사이에 있다. 즉, decaScore는 0과 1000 사이의 값이 된다. decaScore additive combination을 사용하기 때문에 각각의 다른 metric에 대해 가중치를 설정할 때 발생하는 문제를 피할 수 있다는 장점이 있다.

## Multitask Question Answering Network(MQAN)

저자들은 decaNLP와 함께 제안하는 model의 이름을 모든 task가 question answering의 형태로 수정 되고 함께 학습되기 때문에

multitask question answering network(MQAN)이라고 명명한다.

각각의 example은 context, question, answer로 구성된다. 위에서 살펴봤던 예시를 다시 한번 살펴 보자.

Examples		
Question	Context	Answer
What is a major importance of Southern California in relation to California and the US?	...Southern California is a major economic center for the state of California and the US...	major economic center
What is the translation from English to German?	Most of the planet is ocean water.	Der Großteil der Erde ist Meerwasser
What is the summary?	Harry Potter star Daniel Radcliffe gains access to a reported £320 million fortune...	Harry Potter star Daniel Radcliffe gets £320M fortune...
Hypothesis: Product and geography are what make cream skimming work. Entailment, neutral, or contradiction?	Premise: Conceptually cream skimming has two basic dimensions – product and geography.	Entailment
Is this sentence positive or negative?	A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film.	positive

Question	Context	Answer
What has something experienced?	Areas of the Baltic that have experienced eutrophication.	eutrophication
Who is the illustrator of Cycle of the Werewolf?	Cycle of the Werewolf is a short novel by Stephen King, featuring illustrations by comic book artist Bernie Wrightson.	Bernie Wrightson
What is the change in dialogue state?	Are there any Eritrean restaurants in town?	food: Eritrean
What is the translation from English to SQL?	The table has column names... Tell me what the notes are for South Australia	SELECT notes from table WHERE 'Current Slogan' = 'South Australia'
Who had given help?	Joan made sure to thank Susan for all the help she had given.	Susan

논문에서 나오지는 않았지만, 예시에서 주어지는 각 task별 example을 살펴보며 해당 task가 어떠한 방식으로 question answering의 형태로 수정되었는지 살펴보자

## Question Answering

Question Answering model의 경우 원래 그대로이다. 여기서는 다음과 같은 형태를 가진다.

- **Question** : What is a major importance of Southern California in relation to California and the US?
- **Context** : ...*Southern California is a major economic center for the state of California and the US....*
- **Answer** : major enconomic center

Question에 대한 Answer에 대한 정보가 Context에 담겨있음을 확인할 수 있다.

## Machine Translation

Machine translation의 경우, Question으로 영어-독일어 번역이 무엇인지에 대한 sentence를 입력하고, Context는 source language인 영어로 된 sentence, Answer로 target language인 독일어 sentence를 넣음을 확인할 수 있다.

- **Question** : What is the translation from English to German?
- **Context** : Most of the planet is ocean water
- **Answer** : Der Großteil der Erde ist Meerwasser

## Summarization

Summarization은 Question에는 summary가 무엇인지에 대한 질문을 넣고, Context에는 document, Answer에 해당 document의 summary를 넣어주는 것을 확인할 수 있다.

- **Question** : What is the summary?
- **Context** : Harry Potter star Daniel Radcliffe gains access to a reported £320 million fortune...
- **Answer** : Harry Potter star Daniel Radcliffe gets £320M fortune...

## Natural Language Inference

Natural language inference의 경우, Question에는 hypothesis(가설)를 넣은 뒤, Entailment, neutral, or contradiction?을 넣어주었다. Context에는 premise(전제)를 넣어주고, Answer에는 entailment, contradiction, neutral이 있음을 확인할 수 있다.

- **Question** : Hypothesis: Product and geography are what make cream skimming work. Entailment, neutral, or contradiction?
- **Context** : Premise: Conceptually cream skimming has two basic dimensions — product and geography.
- **Answer** : Entailment

## Sentiment Analysis

Sentiment analysis는 Question으로 Is this sentence positive or negative?, Context는 sentiment analysis가 될 sentence, Answer는 sentiment로 구성된다.

- **Question** : Is this sentence positive or negative?
- **Context** : A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film.
- **Answer** : positive

## Semantic Role Labeling

Semantic role labeling(SRL)에서는 Question으로 What has something experienced? 가 있으며, Context는 주어지는 sentence, Answers는 Question에서 물어본 "무엇을"에 해당하는 단어로 구성된다.

- **Question** : What has something experienced?
- **Context** : Areas of the Baltic that have experienced eutrophication.
- **Answer** : eutrophication

## Relation Extraction

Relation extraction의 경우, Question으로 input text에서 추출될 kind of relation인 Who is the illustrator of Cycle of the Werewolf?라는 질문으로 구성되었고, Context는 input unstructured text, Answer은 해당 text에서 추출된 kind of relation이다.

- **Question** : Who is the illustrator of Cycle of the Werewolf?
- **Context** : Cycle of the Werewolf is a short novel by Stephen King, featuring illustrations by comic book artist Bernie Wrightson.
- **Answer** : Bernie Wrightson



## Goal-Oriented Dialogue

Dialogue state tracking의 경우, Question으로는 dialogue state에서의 변화를 묻는 What is the change in dialogue state? 와 Context로 user의 dialogue, Answer로 user의 dialogue system에 대한 사전 정의된 목표를 나타낸다.

- **Question** : What is the change in dialogue state?
- **Context** : Are there any Eritrean restaurants in town?
- **Answer** : food: Eritrean

## Semantic Parsing

Semantic parsing에서는 machine translation과 비슷하다. Question으로 What is the translation from English to SQL? 과 Context는 자연어로 구성된 question, Answer는 structured SQL query로 구성됨을 확인할 수 있다.

- **Question** : What is the translation from English to SQL?
- **Context** : The table has column names... Tell me what the notes are for South Australia
- **Answer** : SELECT notes from table WHERE 'Current Slogan' = 'South Australia'

## Pronoun Resolution

Pronoun Resolution은 Question으로 해당 pronoun(대명사)의 referent(참조)에 대한 질문인 Who had given help? Susan or Joan? 과 Context는 pronoun(대명사)가 있는 문장, Answer는 spronoun(대명사)의 referent(참조)로 구성됨을 확인할 수 있다.

- **Question** : Who had given help? Susan or Joan?
- **Context** : Joan made sure to thank Susan for all the help she had given.
- **Answer** : Susan

저자들은 최근의 많은 QA model들은 일반적으로 answer를 context로부터 복사할 수 있다고 가정하지만, 이러한 가정은 general question answering을 할 수 없게끔 한다고 지적한다. 그러면서 question이 종종 answer space를 제한하는 주요 정보를 포함하는 경우를 가지고 있음을 주목하며 input뿐만 아니라 question의 representation도 풍부하게 하게끔 하는 coattention을 확장하는 것을 제안한다.

training동안, MQAN model은 다음과 같은 3개의 sequence를 input으로 받는다.

- $l$ 개의 token으로 이루어진 Context  $c$
- $m$ 개의 token으로 이루어진 Question
- $n$ 개의 token으로 이루어진 Answer

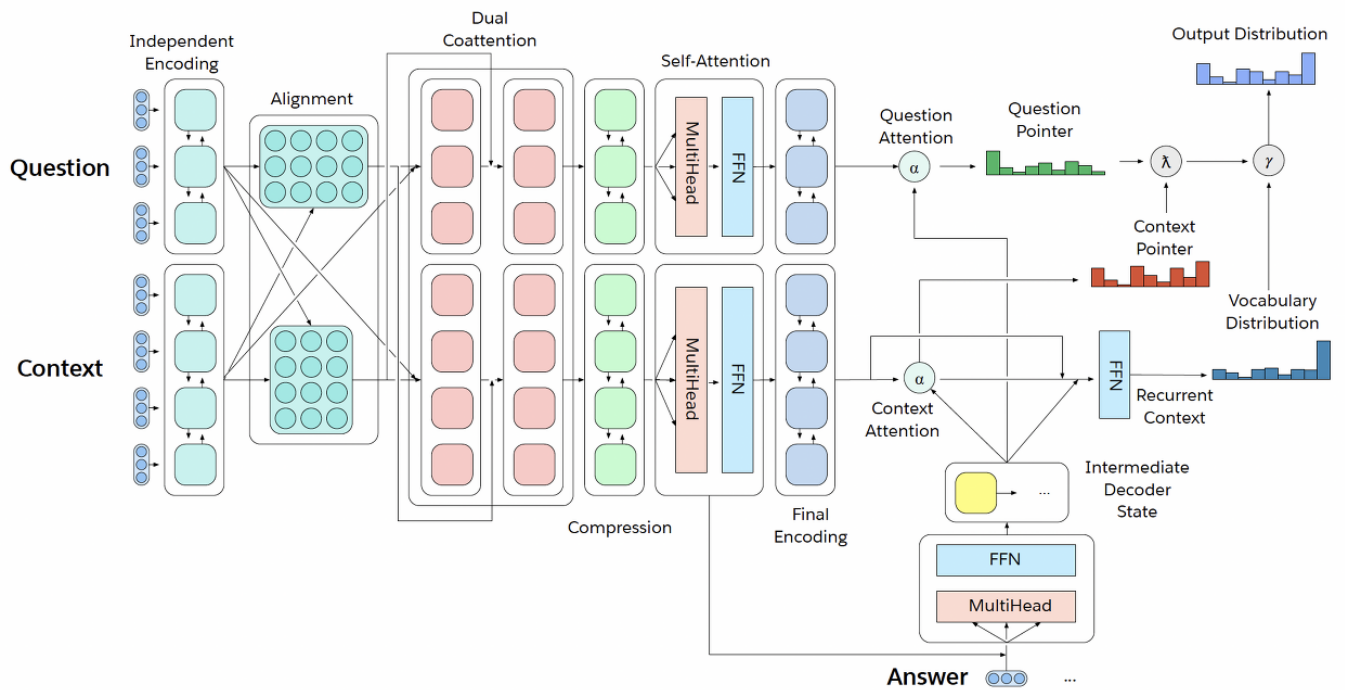
각각의 요소들은  $i$ 번째 행이 sequence의  $i$ 번째 token에 대한  $d_{emb}$  차원의 embedding matrix인 행렬로 표현된다.

수식은 다음과 같다.

$$C \in \mathbb{R}^{l \times d_{emb}} \quad Q \in \mathbb{R}^{m \times d_{emb}} \quad A \in \mathbb{R}^{n \times d_{emb}}$$

Encoder는 이  $C$ ,  $Q$ ,  $A$ 의 input representation들을 input으로 받은 뒤 final representation  $C_{fin} \in \mathbb{R}^{l \times d}$ 와  $Q_{fin} \in \mathbb{R}^{m \times d}$ 을 생성하기 위해 deep stack of recurrent, coattentive, and selfattentive layer를 사용한다.

전체 model의 구조는 다음과 같다.



(model에 관한 세부 내용은 추후 reference로 언급된 논문들을 읽고 난 후 상세히 기술할 예정이다.)

## Experiment and Analysis

본 연구에서 저자들은 MQAN model에 여러 요소들을 추가해가며 ablation study를 진행하였다.

해당 ablation study의 결과는 다음과 같다.

Table 2: Validation metrics for decaNLP baselines: sequence-to-sequence (S2S) with self-attentive transformer layers (w/SAAtt), the addition of coattention (+CAtt) over a split context and question, and a question pointer (+QPtr). The last model is equivalent to MQAN. Multitask models use a round-robin batch-level sampling strategy to jointly train on the full decaNLP. The last column includes an additional anti-curriculum (+ACurr) phase that trains on SQuAD alone before switching to the fully joint strategy. Entries marked with '-' would correspond to decaScores for aggregates of separately trained models; this is not well-defined without a mechanism for choosing between models.

Dataset	Single-task Training				Multitask Training				
	S2S	w/SAAtt	+CAtt	+QPtr	S2S	w/SAAtt	+CAtt	+QPtr	+ACurr
SQuAD	48.2	68.2	74.6	75.5	47.5	66.8	71.8	70.8	74.3
IWSLT	25.0	23.3	26.0	25.5	14.2	13.6	9.0	16.1	13.7
CNN/DM	19.0	20.0	25.1	24.0	25.7	14.0	15.7	23.9	24.6
MNLI	67.5	68.5	34.7	72.8	60.9	69.0	70.4	70.5	69.2
SST	86.4	86.8	86.2	88.1	85.9	84.7	86.5	86.2	86.4
QA-SRL	63.5	67.8	74.8	75.2	68.7	75.1	76.1	75.8	77.6
QA-ZRE	20.0	19.9	16.6	15.6	28.5	31.7	28.5	28.0	34.7
WOZ	85.3	86.0	86.5	84.4	84.0	82.8	75.1	80.6	84.1
WikiSQL	60.0	72.4	72.3	72.6	45.8	64.8	62.9	62.0	58.7
MWSC	43.9	46.3	40.4	52.4	52.4	43.9	37.8	48.8	48.4
decaScore	-	-	-	-	473.6	546.4	533.8	562.7	<b>571.7</b>

Single-task setting부터 살펴보자.

먼저, 첫 번째 baseline으로 pointer-generator sequence-to-sequence (S2S) model을 설정하였다. S2S model의 경우 single sequence를 입력으로 받기 때문에 해당 model에 한해서 context와 question을 concat 하여 진행하였다고 한다.

S2S model는 SQuAD에서 좋은 성능을 내지 못함을 확인할 수 있었으며, 기존 선행 연구의 baseline보다는 높은 성능을 거두었지만 본 연구의 타 baseline에 비해서는 낮은 성능을 내는 것을 확인할 수 있었다.

S2S model에 self attentive encoder, decoder를 추가한 model(w/ SAAtt)은 context와 question의 information을 통합하는 model의 capacity를 늘려주는 효과가 있다. 이 model은 S2S model에 비해 SQuAd, QA-SRL, WikiSQL에서 성능 향상이 있음을 확인할 수 있었다. 특히, WikiSQL에서는 기존 SOTA model과 비슷한 성능을 냄을 확인할 수 있었다.

context와 question을 두 개의 input sequence로 나눈 뒤 coattention mechanism으로 S2S model을 보강한 (+CAtt) model에서는 SQuAD와 QA-SRL에서의 성능 향상이 있었지만, 다른 task들에서는 그다지 좋은 성능을 보이지 못하였고 MNLI와 MWSC에서는 성능이 대폭 하락하였다.

저자들은 이를 회복하기 위해 이전의 baseline에서 question pointer를 추가한 model (+QPtr)을 테스트해보았다. 이 방법은 MNLI, MWSC, SQuAD에서 성능을 향상했다.

Multi-task setting에서도 비슷한 양상의 결과가 도출되었다. 다만, QA-ZRE에서 single task setting에서의 성능보다 더 좋은 결과가 도출되었고, 저자들은 이러한 현상에 대해 **mutitask learning**이 **zero-shot learning**을 위한 더 나은 **generalization**으로 이끈다는 가설을 뒷받침해주는 증거라고 주장한다.

또한 저자들은 decaNLP에서 pre-training 수행 이후 다른 domain과 다른 task에 대해 pretrained MQAN model과 random initialization MQAN의 성능을 비교하였다.

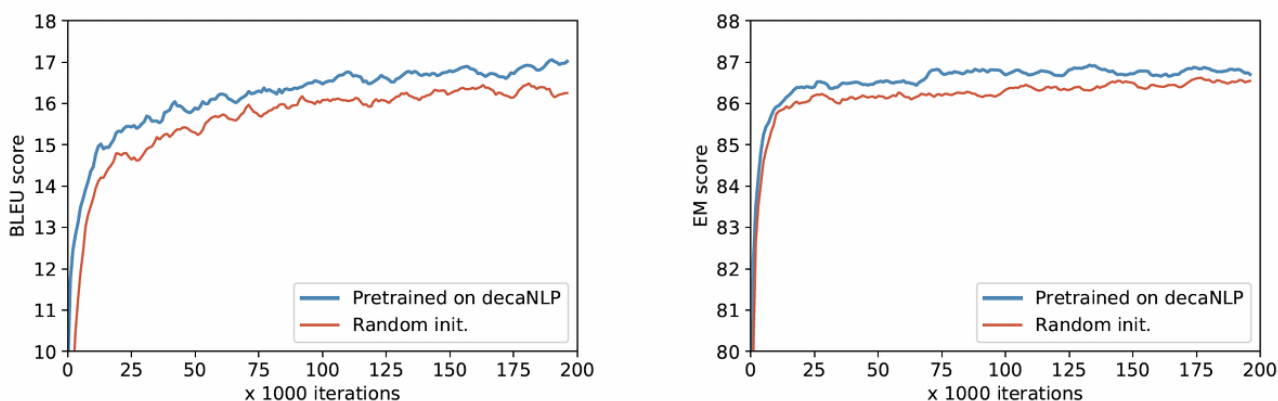


Figure 4: MQAN pretrained on decaNLP outperforms random initialization when adapting to new domains and learning new tasks. Left: training on a new language pair – English to Czech, right: training on a new task – Named Entity Recognition (NER).

왼쪽은 영어-체코어 번역 task이며, 이는 새로운 domain에서의 task 수행이다(machine translation). 오른쪽은 NER(개체명인식) task이며, 새로운 task의 수행이다. 자료에서 볼 수 있듯이, decaNLP에서 pre-training을 수행한 model이 더 좋은 성능을 내는 것을 확인할 수 있다.

또한 MMLI가 decaNLP에 포함되어있기에 Stanford Natural Language Inference Corpus(SNLI) task에서도 adapt 하는 것이 가능하다. 실제로 decaNLP에서 pretraining 된 MQAN은 SNLI task에서 기존 SOTA model의 weight random initialization의 성능보다 2 point 향상된 성능을 낸다고 한다. 또한, 저자들은 주목할만한 점으로 **SNLI task에 대한 fine-tuning 없이도 62%의 성능을 유지한다는 것을 강조한다.**

이와 비슷하게 decaNLP는 SST를 포함하고 있기에 다른 이진 분류 task에서도 잘 작동할 수 있다. 실제로 Amazon and Yelp reviews dataset에서 **fine-tuning을 거치지 않은 decaNLP로 pre-training 된 MQAN model은 각각 82.1%(Amazon), 80.8%(Yelp)의 성능을 보여주었다.**

저자들은 이러한 결과에 대해 decaNLP로 train 된 model이 여러 task에 대해 out of domain context and question을 동시에 generalization 할 수 있으며, text classification을 위해 unseen classes에 대해서도 adapt 할 수 있는 잠재성이 있음을 보여준다고 말한다. 추가적으로, 이러한 input space와 output space에서의 zero-shot domain adaptation은 decaNLP의 task 범위가 single-task training으로 달성할 수 있는 정도 이상으로 generalization을 장려한다는 것을 시사한다고 주장한다.

## Conclusion

본 연구에서는 10개의 task에 대한 성능을 측정하는 새로운 benchmark인 decaNLP를 제안하였으며, 해당 10개의 task는 question answering으로 통합된다. 또한 task에 대한 자연어의 description으로 question을 활용하기 위해 multi-pointer-generator decoder를 사용하는, general question answering model인 MQAN을 제안하였다.

task-specific module이 없음에도 불구하고, 본 연구에서는 MQAN을 decaNLP의 task에 대해 공동으로 train 시켰고, 이러한 anti-curriculum learning이 더 많은 개선점을 가져오는 것을 보였다. 또한 decaNLP에 대한 training 이후 transfer learning과 zero-shot capability에 대한 연구도 진행하였다. **Pretrained weight를 사용할 때 MQAN의 성능이 새로운 task에서도 향상함을 보였다. 이어서 새로운**

domain에서의 text classification에 대한 zero-shot domain adaptation capability를 보였다.