



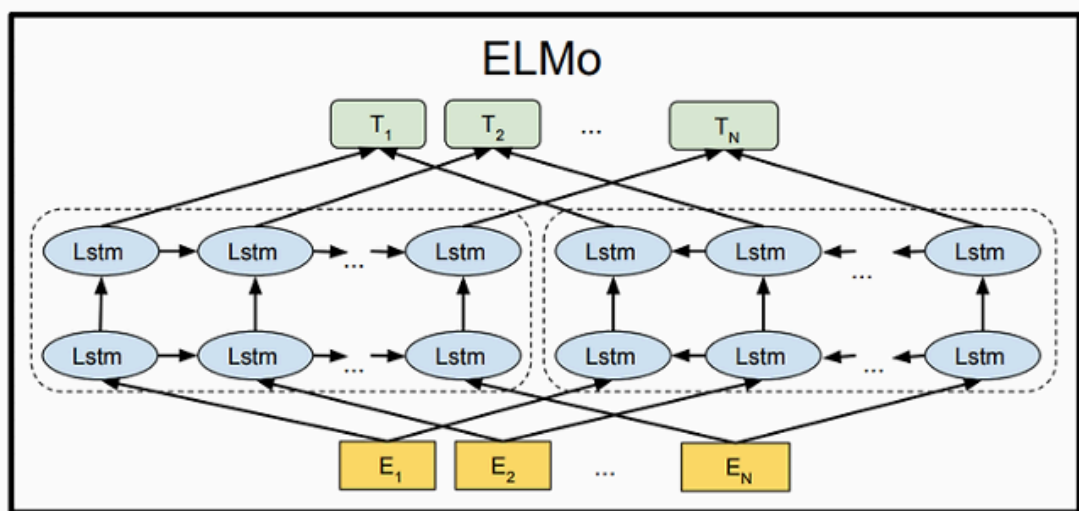
# ELMo

## ▼ Introduction

전통적인 단어 임베딩(Word2Vec, GloVe 등)은 단어의 고정된 표현을 제공하지만, 문맥에 따라 단어의 의미가 달라지는 다의성을 효과적으로 반영하지 못하는 한계가 있다. 이를 해결하기 위해 ELMo(Embeddings from Language Models) 라는 새로운 단어 표현 기법을 제안한다.

ELMo는 양방향 LSTM 기반의 언어 모델(biLM)의 내부 상태를 활용하여, 문맥에 따라 변하는 동적 단어 표현을 학습한다. 이는 다양한 NLP 작업에서 성능을 크게 향상시키며, 기존의 정적인 단어 벡터보다 뛰어난 성능을 발휘한다.

## ▼ Proposed method



## 1 ELMo의 특징

- 기존 단어 임베딩과 달리 문장 전체를 고려하여 동적으로 단어 표현을 생성
- 단어의 구문적(syntactic) 및 의미적(semantic) 특성을 심층적으로 모델링 가능
- 양방향 LSTM 기반 언어 모델을 사전 학습하여, 다양한 NLP 작업에서 전이 학습 가능

- 여러 NLP 태스크에서 SOTA 기록을 갱신

## 2 ELMo의 학습 방식

### 1. 양방향 언어 모델(biLM) 학습

- 문장의 앞과 뒤에서 오는 단어를 동시에 고려하는 양방향 LSTM을 사용
- 각 단어의 문맥을 반영한 벡터를 생성

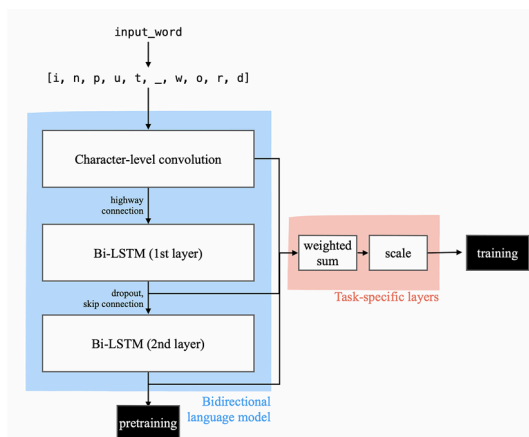
### 2. 다층 구조 활용

- LSTM의 각 층이 다른 언어적 정보를 학습
  - 예: 낮은 층 → 문법 정보, 높은 층 → 의미 정보
- 특정 NLP 태스크에 맞게 다양한 층을 조합하여 최적의 표현 학습

### 3. NLP 모델에 쉽게 통합 가능

- 기존의 NLP 모델 입력에 ELMo를 추가하는 방식으로 쉽게 적용 가능
- 단순한 추가만으로도 성능이 크게 향상됨

## 3 ELMo의 아키텍처



Bidirectional LSTM(Left to Right, Right to Left)

+

Language Model

(Embedding From Language Model)

=> Bidirectional Language Model

- ELMo는 크게 두 가지 주요 컴포넌트로 구성됨:

### 1. 양방향 언어 모델 (biLM)

- Forward LSTM: 앞에서 뒤로 진행하며 단어 예측
- Backward LSTM: 뒤에서 앞으로 진행하며 단어 예측
- 각각의 층은 문맥을 반영한 표현 생성

### 2. ELMo 표현 생성 모듈

- biLM의 여러 층을 조합하여 최적의 단어 표현을 학습
  - 특정 태스크에 따라 각 층을 다른 가중치로 조합하여 사용
  - 최종적으로 각 단어에 대해 문맥에 따른 동적인 벡터를 제공
- 학습 과정
    1. biLM 사전 학습 (Pretraining the biLM)
      - 대규모 코퍼스를 사용하여 학습
      - 언어 모델 (LM) 기반 학습: 다음 단어 예측 (Forward) + 이전 단어 예측 (Backward)
      - LSTM의 내부 상태(hidden states)를 활용하여 단어 표현 학습
    2. 특정 NLP 태스크에 적용
      - 사전 학습된 biLM을 그대로 사용하거나, 도메인 특화된 데이터로 fine-tuning
      - biLM의 각 층을 조합하여 최적의 단어 표현을 생성
      - 결과적으로 각 단어의 표현이 문맥에 따라 동적으로 변화

#### ▼ Results & Performance

- 질문 답변(Question Answering), 문장 추론(Textual Entailment), 감정 분석(Sentiment Analysis) 등 6가지 NLP 태스크에서 성능을 크게 향상시킴

태스크	기존 SOTA	Baseline	ELMo 적용 후	향상률
SQuAD (질문 답변)	84.4	81.1	<b>85.8</b>	+4.7%
SNLI (문장 추론)	88.6	88.0	<b>88.7</b>	+0.7%
SRL (의미역 분석)	81.7	81.4	<b>84.6</b>	+3.2%
Coref (대명사 해소)	67.2	67.2	<b>70.4</b>	+3.2%
NER (개체명 인식)	91.93	90.15	<b>92.22</b>	+2.06%
SST-5 (감정 분석)	53.7	51.4	<b>54.7</b>	+3.3%

➡ ELMo는 문맥을 반영하는 동적인 단어 표현 방식으로, NLP 분야에서 기존 정적 단어 임베딩의 한계를 극복하는 혁신적인 기법이다.