

Deep contextualized word representations

<https://arxiv.org/abs/1802.05365>

0. Introduction

- 기존 단어 임베딩(Word2Vec, GloVe 등)은 **고정 벡터**로 단어를 표현하여 문맥 변화를 반영하지 못함
- 자연어에서는 동일한 단어라도 문맥에 따라 의미가 달라짐 (예: "bank" → 강변 vs. 은행)
- **문맥(Context)을 반영하는 단어 표현**을 제안하여 NLP 성능을 향상시키고자 함
- 제안 모델 : **ELMo (Embeddings from Language Models)**

1. Overview

- ELMo는 문맥에 따라 변화하는 단어 표현(Contextualized Word Embeddings)을 생성
- 특징 :
 - 단어 표현이 문장의 문맥에 따라 동적으로 변화
 - 다층 Bi-directional LSTM 기반
 - 다양한 NLP 태스크에서 성능 향상 가능

2. Challenges

- 고정 임베딩의 문맥 불일치 문제
- 문맥 정보를 효과적으로 통합하는 방법 설계의 어려움
- 모델 복잡도 및 계산 비용 증가
- 대규모 언어 모델 학습 시 리소스 요구량

3. Method

- **Bi-directional Language Model (BiLM)** 학습 : 양방향 LSTM 사용
 - 전방(forward) LSTM : 왼쪽에서 오른쪽으로 단어 예측
 - 후방(backward) LSTM : 오른쪽에서 왼쪽으로 단어 예측
- ELMo 표현 : 다층 BiLM의 각 층 출력을 가중합하여 단어 표현 생성
- 학습 : 대규모 말뭉치에 대해 언어 모델 방식으로 사전 학습
- 기존 태스크에 ELMo 임베딩을 추가하면 성능 향상 가능

4. Experiments

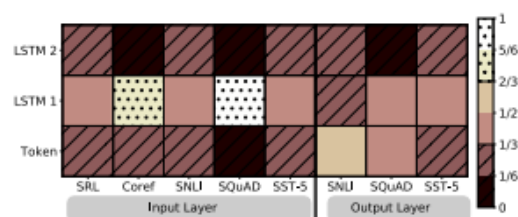
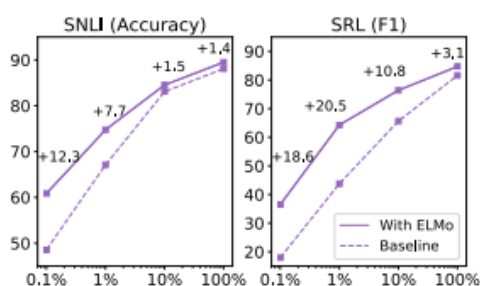
- 데이터셋 : 1 Billion Word Benchmark, OntoNotes 5.0, SQuAD 등
- 평가 태스크 : 개체명 인식(NER), 문장 관계 추론(SNLI), 질문 답변(QA), 의미역 분석(SRL)
- 비교 baseline : GloVe, Word2Vec 등
- 다양한 층 수 및 가중치 설정 실험

5. Results

TASK	PREVIOUS SOTA		OUR BASELINE	ELMo + BASELINE	INCREASE (ABSOLUTE/ RELATIVE)
SQuAD	Liu et al. (2017)	84.4	81.1	85.8	4.7 / 24.9%
SNLI	Chen et al. (2017)	88.6	88.0	88.7 ± 0.17	0.7 / 5.8%
SRL	He et al. (2017)	81.7	81.4	84.6	3.2 / 17.2%
Coref	Lee et al. (2017)	67.2	67.2	70.4	3.2 / 9.8%
NER	Peters et al. (2017)	91.93 ± 0.19	90.15	92.22 ± 0.10	2.06 / 21%
SST-5	McCann et al. (2017)	53.7	51.4	54.7 ± 0.5	3.3 / 6.8%

Model	F ₁
WordNet 1st Sense Baseline	65.9
Raganato et al. (2017a)	69.9
Iacobacci et al. (2016)	70.1
CoVe, First Layer	59.4
CoVe, Second Layer	64.7
biLM, First layer	67.4
biLM, Second layer	69.0

Model	Acc.
Collobert et al. (2011)	97.3
Ma and Hovy (2016)	97.6
Ling et al. (2015)	97.8
CoVe, First Layer	93.3
CoVe, Second Layer	92.8
biLM, First Layer	97.3
biLM, Second Layer	96.8



- ELMo 임베딩 추가 시 모든 태스크에서 성능 개선
- 문맥 의존성이 중요한 태스크에서 특히 효과적
- 기존 고정 임베딩 대비 평균 4~7% 성능 향상
- ELMo의 문맥 표현력이 모델 성능 향상의 핵심 요인임 확인

6. Insight

- ELMo는 **문맥 기반 단어 표현**의 효시로 NLP 모델의 성능을 크게 향상시킴
- 이후 BERT, GPT 등 **Transformer 기반의 Contextualized Embedding** 연구의 기반이 됨
- 문맥 의존성 반영이 NLP 전반의 패러다임 변화를 이끌었음
- 핵심: 단어 표현을 문맥과 결합하는 방식이 NLP 성능에 중요한 영향을 미침