



# Word2Vec & ELMo

날짜	@2025년 2월 19일
분야	분석
주차	5주차

## Efficient Estimation of Word Representations in Vector Space

### 1. 연구 개요

이 논문은 2013년 Tomas Mikolov 등이 발표한 연구로, Word2Vec이라는 단어 임베딩 기법을 소개함

Word2Vec은 단어를 고차원 벡터 공간에 효율적으로 표현하는 방법으로, 단어 간 의미적 유사성을 잘 포착할 수 있도록 설계됨

### 2. Word2Vec의 주요 학습 방식

논문에서는 Word2Vec의 두 가지 주요 학습 방식인 CBOW(Continuous Bag of Words)와 Skip-gram을 제안

- **CBOW**

- 주변 단어(context words)를 기반으로 중심 단어(target word)를 예측하는 방식
- 학습 속도가 빠르고 소규모 데이터셋에서도 좋은 성능을 보임

- **Skip-gram**

- 중심 단어를 기반으로 주변 단어를 예측하는 방식
- 희귀한 단어에 대해서도 효과적인 벡터 표현을 학습할 수 있으며 대규모 데이터에서 강력한 성능을 발휘

### 3. 모델의 효율성 개선 기법

Word2Vec의 계산량을 줄이고 학습 속도를 높이기 위해 다음과 같은 기법을 도입

- **Negative Sampling**

- 소프트맥스 연산을 근사하여 계산량을 줄이는 방법

- **Hierarchical Softmax**

- 단어 출현 확률을 트리 구조로 모델링하여 연산 속도를 개선하는 방식

#### 4. 실험 결과 및 시사점

이 연구는 단어 벡터 간 연산을 통해 단어 의미를 해석할 수 있음을 실험적으로 증명

예를 들어, "king - man + woman = queen"과 같은 벡터 연산이 가능함을 보임

또한, 기존 단어 임베딩 기법보다 훨씬 빠르고 효율적인 학습이 가능하도록 설계되어 NLP 분야에서 전이 학습(Transfer Learning)의 기반을 마련함

#### 5. Word2Vec의 영향 및 활용

이 논문은 이후 GloVe, FastText, ELMo, BERT 등의 연구에 큰 영향을 미쳤으며, 현재도 기초적인 NLP 모델링에서 널리 활용되고 있음

## Deep Contextualized Word Representations

### 1. 연구 개요

이 논문은 2018년 Matthew E. Peters 등 연구진이 발표한 논문으로, ELMo (Embeddings from Language Models) 라는 새로운 단어 임베딩 기법을 소개한다. 기존의 Word2Vec이나 GloVe와 같은 정적(Static) 단어 임베딩과 달리, ELMo는 문맥(Context)에 따라 단어의 의미를 동적으로 조정하는 문맥화된(Contextualized) 단어 표현을 생성함

### 2. 기존 단어 임베딩의 한계

Word2Vec, GloVe와 같은 정적 임베딩은 각 단어가 **고정된 벡터 표현**을 가짐

- 예를 들어, "bank"라는 단어는 "river bank"(강둑)과 "bank account"(은행)에서 다른 의미로 사용되지만, 정적 임베딩은 이를 구분하지 못함

이러한 한계를 극복하기 위해, 문맥에 따라 단어 벡터를 동적으로 생성하는 모델이 필요

### 3. ELMo 모델 구조

ELMo는 사전 학습된 양방향 LSTM (BiLSTM) 기반의 언어 모델(Language Model, LM)을 활용하여 단어 표현을 학습한다.

BiLSTM 기반 언어 모델

- **순방향(Forward) LSTM**

- 앞에서부터 단어를 읽으며 다음 단어를 예측

- **역방향(Backward) LSTM**

- 뒤에서부터 단어를 읽으며 이전 단어를 예측

- 이렇게 학습된 두 방향의 정보를 결합하여 **문맥을 반영한 단어 표현을 생성**

다층 표현 (Deep Representations)

- ELMo는 단어 벡터를 단순히 한 층의 LSTM 결과로 표현하지 않고, **여러 층의 정보를 가중합하여 단어 벡터를 생성**
- 각 층의 정보가 다르게 학습되므로, 문맥에 따라 적절한 정보를 선택할 수 있음

#### 4. 주요 특징 및 장점

문맥에 따라 단어 벡터가 변함 → 같은 단어라도 문맥에 따라 다른 의미를 가질 수 있음

전이 학습(Transfer Learning)이 가능 → 사전 학습된 모델을 다양한 NLP 태스크 (SQuAD, SNLI, CoNLL-2003 등)에 적용 가능

기존의 단어 임베딩보다 성능 향상 → 감성 분석, 개체명 인식(NER), 질의응답(QA) 등 다양한 NLP 태스크에서 우수한 성능을 보임

#### 5. 실험 및 성능 평가

ELMo는 다양한 NLP 태스크에서 기존 단어 임베딩 기법을 크게 능가하는 성능을 보임

SQuAD (Question Answering), SNLI (Natural Language Inference), CoNLL-2003 (Named Entity Recognition) 등의 벤치마크에서 기존 모델 대비 성능 향상

- 벤치마크?
  - 성능평가 지표를 일컫는 말

특히, 문맥의 영향을 크게 받는 태스크(NER, QA 등)에서 더 큰 성능 향상

#### 6. 결론 및 영향

이 논문은 정적 단어 임베딩의 한계를 극복한 문맥화된 단어 표현을 제안하여, 이후의 NLP 연구(특히 BERT, GPT 등 Transformer 기반 모델)에 큰 영향을 미침

ELMo는 현재 BERT, GPT 등의 등장으로 직접적으로 사용되는 경우는 줄었지만, 딥러닝 기반 NLP 모델의 발전에 중요한 역할을 한 모델로 평가됨