

# Word2Vec의 네거티브 샘플링(negative sampling)과 이진분류(binary classification) 효과 설명

Word2Vec는 단어를 벡터로 임베딩하는 대표적인 방법 중 하나로, CBOW(Continuous Bag of Words)와 Skip-gram 두 가지 모델이 있습니다. 특히 Skip-gram 모델은 중심 단어로부터 주변 단어를 예측하는 방식입니다. 하지만 전체 단어 집합(vocabulary)이 매우 클 경우, 모든 단어에 대해 softmax를 계산하는 것은 연산량이 너무 많아 비효율적입니다.

이를 해결하기 위해 **네거티브 샘플링(negative sampling)** 기법이 도입되었습니다.

## 네거티브 샘플링이란?

- **핵심 아이디어:** 전체 단어 집합이 아닌, 일부(소수)의 '네거티브(음성) 샘플'만을 뽑아 이진 분류 문제로 단순화합니다.
- **방법:**
  - 실제로 주변에 등장한 단어(positive sample)는 1(참)로,
  - 임의로 뽑은 주변에 등장하지 않은 단어(negative sample)는 0(거짓)으로 레이블링합니다.
  - 모델은 중심 단어와 주변 단어 쌍이 실제로 문맥상 함께 등장하는지(1), 아닌지(0)를 구분하는 **이진 분류(binary classification)** 문제를 학습합니다.

## 이진분류를 통한 효과

- **계산 효율성:** 전체 단어 집합에 대해 softmax를 계산하지 않고, 소수의 네거티브 샘플에 대해서만 sigmoid(이진 분류)를 수행하므로 연산량이 크게 줄어듭니다.
- **학습 속도 향상:** 연산량 감소로 인해 대용량 코퍼스에서도 빠르게 학습할 수 있습니다.
- **임베딩 품질 유지:** 적절한 네거티브 샘플 수를 선택하면, 전체 softmax를 사용한 경우와 비슷한 품질의 임베딩을 얻을 수 있습니다.

## 수식적 설명

- 중심 단어  $w_c$ 와 주변 단어  $w_o$ 가 실제로 함께 등장하면  $D = 1$ , 임의로 뽑은 단어면  $D = 0$ .
- 확률은 sigmoid 함수로 계산:
  - $P(D = 1 | w_c, w_o) = \sigma(\mathbf{v}_{w_c} \cdot \mathbf{v}_{w_o})$
  - $P(D = 0 | w_c, w_k) = \sigma(-\mathbf{v}_{w_c} \cdot \mathbf{v}_{w_k})$

- 전체 손실 함수는 실제 쌍(positive)와 네거티브 샘플(negative)에 대해 각각 이진 분류 손실을 합산합니다.

## 요약

- **네거티브 샘플링**은 Word2Vec에서 전체 단어 집합 대신 일부 샘플만을 사용해 이진 분류 문제로 변환함으로써, 계산 효율성과 학습 속도를 크게 높여주는 핵심 기법입니다.
- 이진 분류를 통해 중심 단어와 주변 단어의 관계를 효과적으로 학습할 수 있습니다.