

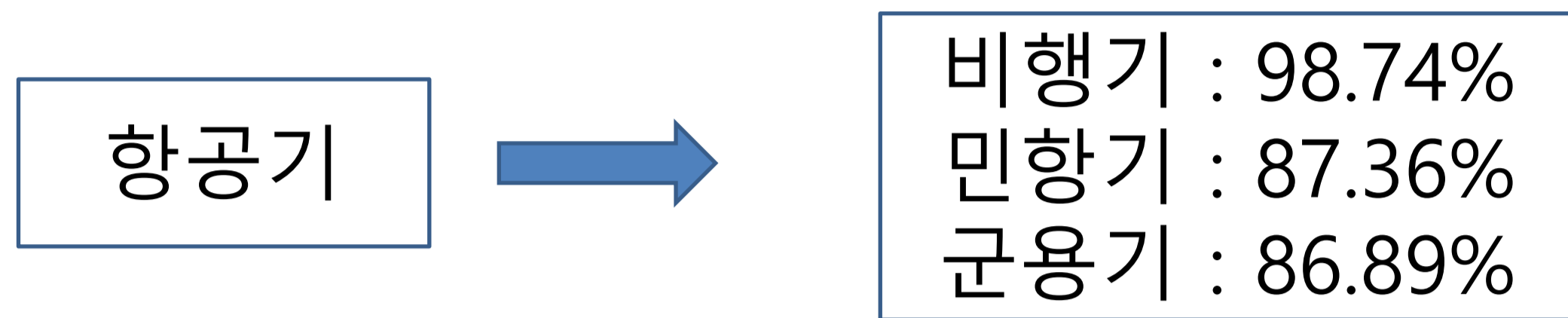
사전학습 언어모델의 Perplexity에 기반한 Zero-shot 어휘 의미 모델

최형준^o, 나승훈
전북대학교

modera2017@gmail.com, nash@jbnu.ac.kr

I. 유의어 추천

유의어 추천은 입력된 단어와 유사한 단어를 제시하는 작업을 의미한다.



유의어 추천을 위해 각 단어간 유사도를 계산하여 제시할 단어를 선택해야 하지만, 데이터셋 내에 존재하지 않는 단어에 대해서는 유사도를 계산하기 어렵다. 이러한 문제를 해결하기 위해 Masked LM(Language Model)을 사용하여 각 단어에 대한 PPL(Perplexity)를 계산, 이를 통해 단어간 유사도를 계산하였다.

II. PPL 기반 단어 유사도

PPL은 언어모델의 출력이 얼마나 실제 문장을 잘 예측하는지에 대한 척도로, 다음과 같이 계산할 수 있다.

$$PPL(s) = \exp\left(-\frac{1}{l} \sum_{x \in s} p(x)\right)$$

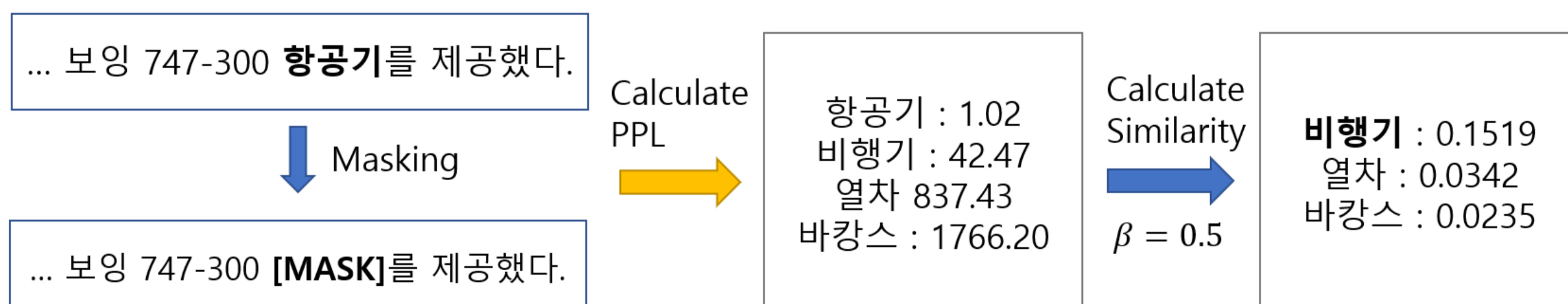
Masked LM은 입력 문장에서 일부 토큰을 임의로 [MASK] 토큰 등으로 마스킹한 후, 마스킹 되기 전의 원래 단어를 예측하도록 사전학습을 진행한다. 이를 통해 마스킹된 N-gram 토큰 $w = [x_k, x_{k+1}, \dots, x_{k+n-1}]$ 에 대해 Conditional PPL을 다음과 같이 계산할 수 있다.

$$CondPPL(s) = \exp\left(-\frac{1}{N} \sum_{x \in w} p(x)\right)$$

이때, 원본 N-gram 토큰을 다른 N-gram 토큰으로 치환하는 것을 통해 원본 단어와 다른 단어 간의 PPL 차이를 얻을 수 있는데, 원본 문장과 어울리는, 원본 N-gram과 유사한 단어일수록 PPL이 낮게 측정되고, 그렇지 않은 단어일수록 PPL이 높게 측정된다. 따라서 단어간 유사도는 다음과 같이 계산할 수 있다.

$$sim(w, w'|s) = CondPPL(w|s)^{-\beta} \cdot CondPPL(w'|s)$$

$$sim(w, w') = \frac{1}{N} \sum_{w \in s} sim(w, w'|s)$$



III. 실험 구성

실험을 위해 한국어 위키백과를 사용하여 데이터셋을 구성했다. 각 단어 사이의 유의어 관계는 다음과 같이 수작업으로 구성했고, 유의어 집합은 총 113개, 전체 단어 수는 326개이고, 이중 한국어 위키백과 데이터셋에 포함되지 않은 단어는 131개이다.

| | | | |
|-------|----|-------|----|
| 대영제국 | 영국 | 잉글랜드 | |
| 대한민국 | 한국 | 남한 | |
| 공상 과학 | SF | SF 영화 | |
| 선거 | 투표 | 총선 | 대선 |

한국어 위키백과 문장 중 하이퍼링크가 걸린 부분을 각 단어에 대한 멘션으로 사용했고, 공백문자를 제거한 후 유의어 집합 내의 단어들과 문자열 비교 후, 일치하는 단어를 추출하여 데이터셋을 구성했다. 이때 각 단어마다 최대 4개의 문장만 데이터셋에 포함되게 구성했다.

IV. 실험

| Beta | MRR |
|------|--------|
| 0.5 | 41.31% |
| 1.0 | 41.30% |
| 2.0 | 40.29% |

위의 성능은 데이터셋 내에 존재하는 모든 단어간 유사도를 계산한 후 각 단어에 대한 유의어 집합 내의 자기 자신이 아닌 다른 단어에 대한 MRR을 Beta 값을 조정해가며 측정해본 것이다. MRR은 유의어 집합 내의 각 단어를 유사도 기준으로 내림차순으로 정렬한 후, 각 유의어의 위치에 대한 역수의 평균을 매긴 것이다.

| Method | MRR |
|--------|--------|
| Top-1 | 38.71% |
| MRR | 34.05% |
| Recall | 52.77% |

위의 성능은 Beta가 0.5일 때 각 문장에 대해 PPL 기반으로 유의어를 추출한 결과에 대한 성능이다. Top-1은 가장 유사도가 높은 단어가 유의어 집합에 포함이 되어있는지에 대한 비중이고, Recall@5는 유의어 집합 내의 각 단어가 상위 5개 단어 안에 속해 있는 비중이다.