

1. 실험 단어(3단어)

- “사랑”, “무관심”, “증오” 순으로 단어 리스트 나열함.

2. 사용한 임베딩 모델과 링크

- 모델명: snunlp/KR-SBERT-V40K-klueNLI-augSTS
링크: (<https://huggingface.co/snunlp/KR-SBERT-V40K-klueNLI-augSTS>)

3. 단어 임베딩 결과 비교표

```
유사도 행렬 (Cosine Similarity):  
[[1.0000002  0.34904817 0.4880125 ]  
 [0.34904817 1.0000001  0.32961398]  
 [0.4880125  0.32961398 1.          ]]
```

4. 유사도 차이에서 흥미로웠던 결과

- 사랑과 무관심은 통상적으로 반의어로 여겨지기도 하여, 유사도가 음수에 가까울 것이라 예상했다. 하지만 실제 결과는 **35%** 정도의 유사도를 보여 의외였다. 반면, 증오와 무관심 간의 유사도는 더 낮게 나타났다. 이는 일상 언어에서 두 단어가 함께 쓰이는 빈도나 의미적 맥락 차이에서 기인할 수 있다고 생각한다. 자연어처리의 결과가 언어 사용의 실제와 얼마나 연관되는지 관찰해보고자 한다.

5. 벡터 연산 실험: 의미적 연산이 가능한가?

- 의미적 벡터 연산(예: 사랑 - 증오 + 무관심)이 직관적 결과를 주는 경우도 있지만, 이는 제한된 범위에서만 유효하다.
- 현시점에서 단순 벡터 산술만으로 모든 언어적 뉘앙스를 포착하는 데에는 한계가 있다.
- 최근 연구들은 비선형 변환 네트워크, 그래프 기반 의미 연결망, 기하학적 모델링 등을 도입해 이 한계를 보완하고 있다.
- 이러한 연구 흐름의 핵심에는 **self-attention mechanism**이 있으며, 대표적 모델로는 **GPT-4** 등이 있다.

6. 임베딩 시각화(시각화는 실패했지만, 수학적 설명을 하려함)

- 이번 실험에서는 PCA 및 t-SNE를 통한 시각화는 구현하지 못했지만, 이 두 기법의 수학적 차이를 정리해보았다.

항목	PCA	t-SNE
목적	분산 최대화 방향 찾기	유사한 이웃 구조 보존
수학 기법	고유값 분해 (eigendecomposition) 기반	확률적 거리 분포 보존 (t-분포 기반)
사용 행렬	공분산 행렬 (항상 PSD)	조건부 확률 행렬 (PSD 아님)
보존 정보	전역적 거리 정보	국소적 거리 정보 (시각화에 특화됨)
계산 비용	상대적으로 낮음	상대적으로 높음

보완 개념 정리

- 코사인 유사도 행렬은 대칭 행렬이지만, 반드시 양의 준정부호(Positive Semi-Definite, PSD)는 아니다.
- 따라서 PCA에 사용할 경우에는, 별도의 정규화 전처리 또는 공분산 행렬 기반의 분석으로 변환 후 사용하는 것이 일반적이라고 생각합니다.

마무리 소감

자연어의 감정 표현을 수치적으로 모델링하면서, 단어 간의 관계가 통계적으로도 의미를 가질 수 있다는 점이 흥미로웠다. 그러나, 임베딩 벡터의 해석은 수학적 엄밀함과 언어적 직관의 균형이 필요하다는 것을 느꼈다. 향후에는 시각화도 성공시키고, 보다 넓은 단어군으로 실험을 확장하고자 한다.