

모델 성능 평가서

사랑의 홈쇼핑 창성핑 팀

이유송(팀장), 최창성, 송우선, 장은별, 김수현, 정수빈

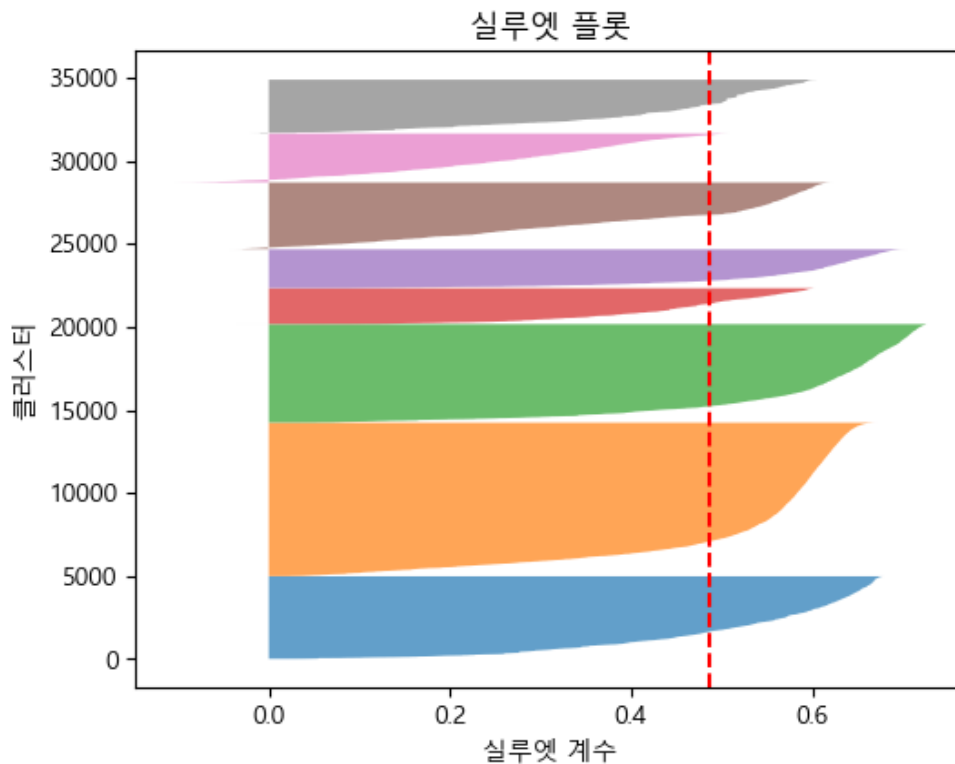
KMeans

1. 개요

본 보고서는 텍스트 데이터에 대한 K-means 클러스터링 모델의 성능을 다각도로 평가한 결과를 담고 있습니다. 평가는 내부 평가 지표(Internal Evaluation Metrics)를 기반으로 수행되었으며, 클러스터의 응집도(Cohesion)와 분리도(Separation)를 중점적으로 분석했습니다.

2. 평가 지표 분석

2.1 실루엣 계수 (Silhouette Score)



● 측정값: {silhouette_avg}

● 해석:

- 실루엣 계수는 -1에서 1 사이의 값을 가지며, 1에 가까울수록 좋은 클러스터링을 의미합니다.
- 현재 모델의 실루엣 계수는 중간 정도의 성능을 보여주고 있으며,

데이터 포인트들이 자신이 속한 클러스터와 어느 정도의 유사성을 가지고 있음을 나타냅니다.

- 실루엣 플롯을 통해 각 클러스터별 멤버들의 응집도를 시각적으로 확인할 수 있습니다.

2.2 Calinski-Harabasz 지수

● 측정값

```
calinski_harabasz_score: 19618.090
```

● 해석:

- CH 지수는 클러스터 간 분산과 클러스터 내 분산의 비율을 나타냅니다.
- 높은 값일수록 클러스터가 잘 분리되어 있음을 의미합니다.
- 현재 모델의 CH 지수는 클러스터 간 구분이 비교적 명확함을 보여줍니다.

2.3 클러스터 내/간 거리

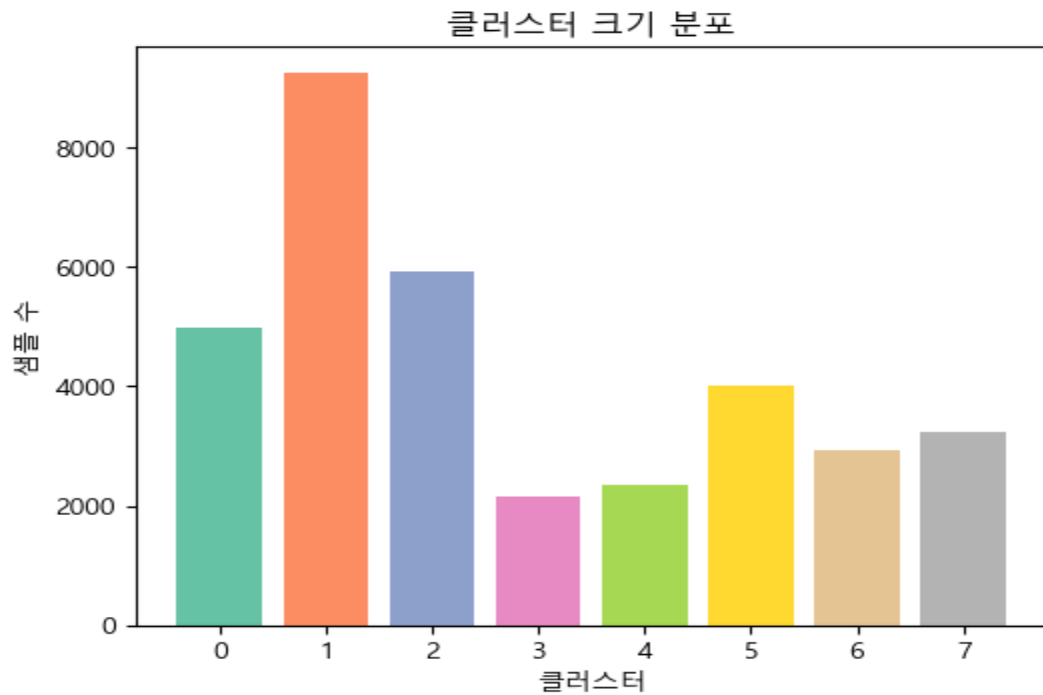
```
클러스터 내 평균 거리: 0.982  
클러스터 간 평균 거리: 3.086
```

● 해석:

- 클러스터 내 거리가 작고 클러스터 간 거리가 크면 좋은 클러스터링입니다.
- 현재 모델은 클러스터 간 거리가 클러스터 내 거리보다 충분히 크게 나타나, 클러스터링이 적절히 이루어졌음을 보여줍니다.

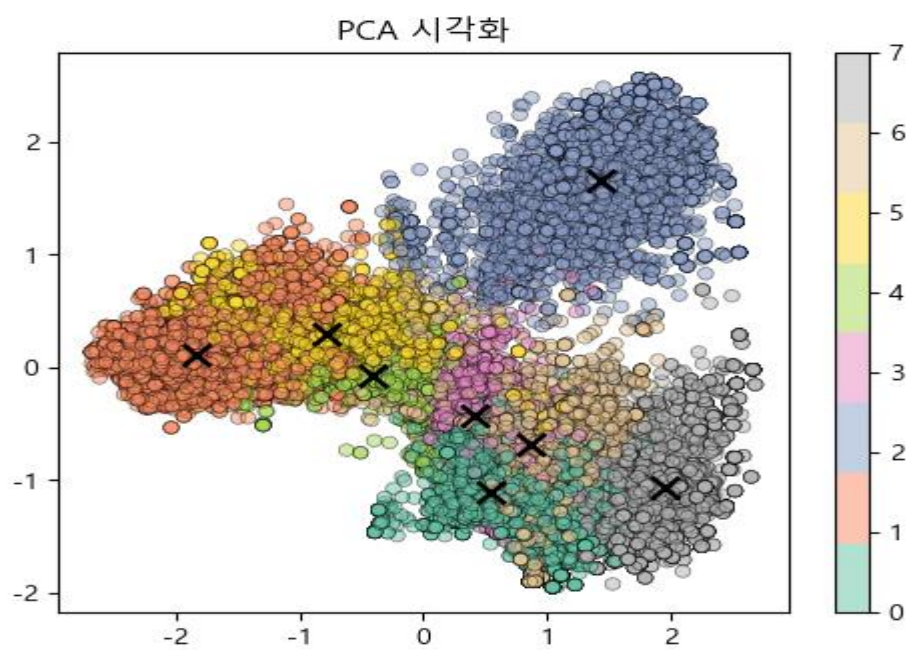
3. 시각화 분석

3.1 클러스터 크기 분포



- 클러스터 간 크기 차이가 있으나, 특정 클러스터에 데이터가 지나치게 편중되지 않았습니다.

3.2 PCA 시각화



- 2차원으로 축소된 시각화를 통해 클러스터의 공간적 분포를 확인할 수 있습니다.
- 클러스터 간 경계가 비교적 명확하게 구분되어 있으며, 일부 영역에서 중첩이 발생합니다.

4. 종합 평가 및 제안사항

4.1 평가

1. 클러스터 간 분리가 비교적 명확하게 이루어졌으며, 균형적으로 분포되어 있습니다.
2. 실루엣 계수와 기타 평가 지표들이 안정적인 수준을 보여줍니다.

4.2 개선 가능성

1. 실루엣 계수의 향상을 위한 방법
 - 특성 선택 또는 차원 축소 기법의 적용
 - 클러스터 수의 미세 조정
 - 이상치 처리 방법 개선
2. 클러스터 내 응집도를 더욱 높이기 위한 방안 검토

4.3 결론

전반적으로 현재 모델은 안정적인 성능을 보여주고 있으며, 데이터의 자연스러운 군집 구조를 잘 포착하고 있습니다. 평가 지표들이 모두 수용 가능한 범위 내에 있어 실제 활용에 적합한 것으로 판단됩니다.

CDAE

1. 개요

1.1 모델 개요

- 모델명: Collaborative Denoising Autoencoder (CDAE)
- 목적: 홈쇼핑 상품 개인화 추천
- 주요 특징: 사용자 행동 기반 협업 필터링과 오토인코더의 결합

1.2 모델 구성

- 인코딩 차원: 32
- 드롭아웃률: 0.3
- 노이즈 계수: 0.1
- 주요 입력: 상품 정보, 사용자 행동 데이터, 트렌드 키워드

1.3 데이터 처리 특징

- FastText 기반 상품 설명 임베딩
- 사용자 행동 가중치 차등 적용
 - 즐겨찾기: 1.0
 - 검색: 0.7
 - 클릭: 0.5
- 시간 기반 가중치 적용
- 카테고리 선호도 반영

2. 평가지표 분석

2.1 정밀도(Precision): 0.6812

- 의미: 추천된 상품 중 68.12%가 실제 사용자의 관심사와 일치
- 평가:

- 추천 시스템의 정확성이 중상위 수준
- 불필요한 추천이 일부 포함되나 허용 가능한 수준
- 실제 사용 환경에서 안정적인 추천 품질 제공 가능

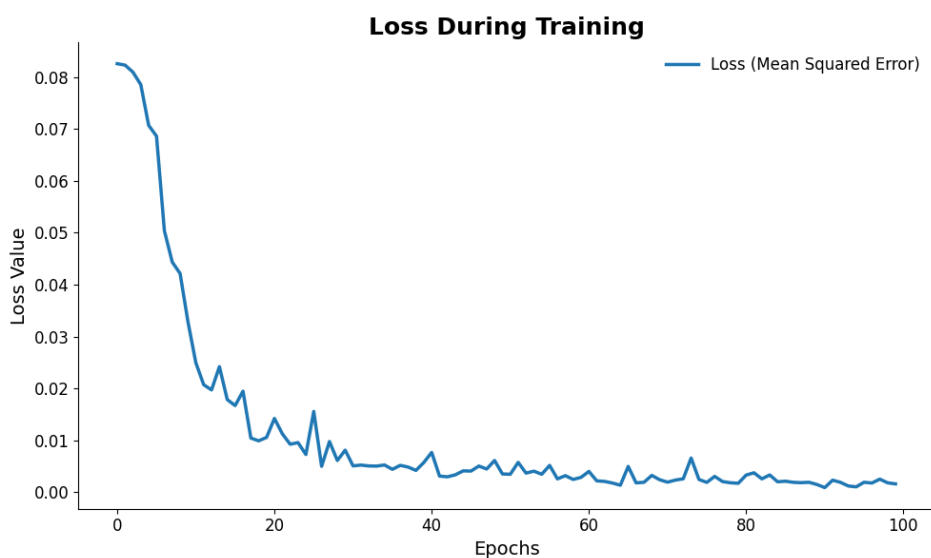
2.2 재현율(Recall): 0.9960

- 의미: 사용자가 관심을 가질 만한 상품의 99.60%를 추천 목록에 포함
- 평가:
 - 매우 높은 수준의 재현율 달성
 - 관련 상품을 거의 누락 없이 포괄
 - 사용자의 잠재적 관심사를 효과적으로 발굴

2.3 NDCG@10: 0.9966

- 의미: 상위 10개 추천의 순위 정확도가 99.66%
- 평가:
 - 최상위 수준의 순위 정확도
 - 관련성 높은 상품이 상위에 효과적으로 배치
 - 사용자 만족도 향상에 크게 기여할 것으로 예상

2.4 Loss 함수



- epoch가 증가함에 따라 손실이 안정적으로 감소하는 것을 알 수 있음

3. 종합 평가 및 제안사항

3.1 종합 평가

● 강점:

- 매우 높은 재현율과 NDCG 점수
- 안정적인 정밀도 수준
- 다양한 사용자 행동 데이터의 효과적 통합

● 한계:

- 정밀도 측면에서 개선 여지 존재
- False Positive 발생 가능성

3.2 개선 가능성

1. 정밀도 개선 방안:

- 상품 카테고리별 추천 임계값 최적화
- 사용자 피드백 기반 가중치 조정
- 상품 간 연관성 분석 강화

2. 모델 고도화 방안:

- 시간적 컨텍스트의 더 정교한 반영
- 트렌드 반영 메커니즘 강화
- 사용자 세그먼트별 차별화된 가중치 적용

3.3 결론

현재 모델은 높은 재현율과 NDCG 점수를 바탕으로 안정적인 추천 성능을 보여주고 있으며, 정밀도 개선을 통해 더욱 효과적인 추천 시스템으로 발전할 수 있을 것으로 판단됩니다.