

부산 관광지 활성화를 위한 음식점 추천 시스템 개발

김재영*, 배준수*, 조민서*, 장진규*,
*동아대학교 컴퓨터공학과
e-mail : tjfansghk@naver.com

Restaurant recommendation system using BPR model

Jaeyoung Kim*, Junsu Bae*, Minseo Jo*, Jingyu Jang*,
* * Department of Computer Engineering, Dong-A University

요약

부산 관광지 활성화를 위한 BPR 모델은 관광객의 과거 행동 데이터를 바탕으로 개인화된 추천을 제공하여, 다양한 음식점을 효과적으로 소개함으로써 관광객들이 보다 만족스러운 경험을 할 수 있도록 도와줍니다. 이를 통해 부산의 관광 활성화와 지역 경제에 긍정적인 영향을 미칠 수 있습니다.

1. 서론

부산은 인기 있는 관광지로, 관광객들의 경험을 향상시키기 위해 음식점 추천 서비스가 중요한 역할을 할 수 있습니다. 이를 위해서 BPR 모델을 기반으로 한 음식점 추천 시스템을 개발하여, 사용자 개인의 취향을 반영하고 효율적인 맞춤형 추천을 제공하는 것을 목표로 합니다.

2. 관련연구

기존 추천 시스템은 주로 콘텐츠 기반 필터링(Content-based Filtering)과 협업 필터링(Collaborative Filtering) 두 가지 접근 방식을 활용합니다. 본 연구에서는 협업 필터링 기법 중 하나인 행렬 분해(Matrix Factorization, MF)를 적용하여 초기 추천 시스템을 구현하였습니다.

행렬 분해는 협업 필터링의 대표적인 방법으로, 사용자와 아이템 간의 상호작용을 행렬 형태로 표현하고, 이를 잠재 요인(Latent Factors)으로 분해하여 학습하는 방식입니다. 이 접근 방식의 주요 목표는 사용자가 과거에 평가한 아이템 정보를 기반으로, 새로운 아이템에 대한 사용자 선호도를 예측하는 것입니다.

행렬 분해 아이디어: (그림 1 참조)

- 행렬 표현:** 사용자의 평가 정보를 담고 있는 Rating Matrix R 를 사용자 행렬 U 와 아이템 행렬 I 로 분해합니다. 여기서 R 은 사용자가 아이템에 대해 부여한 평가 점수로 구성된 행렬입니다.
- 행렬 분해:** Rating Matrix R 를 두 개의 저차원 행렬, 사용자 행렬 U 와 아이템 행렬 I 로 분해합니다. 각 행렬의 차원은 사용자의 잠재 요인과 아이템의 잠재 요인을 나타내며, 이 과정에서 사용자와 아이템 간의 잠재적 관계를 학습합니다.

- 예측 생성:** 분해된 행렬 U 와 I 를 다시 내적하여 예측된 평점 행렬 R^{\wedge} 을 생성합니다. 이 예측된 평점 행렬은 원래의 Rating Matrix R 에서 비어있는 부분에 대한 예측값을 제공합니다.

이러한 행렬 분해 기법은 비어있는 평가 점수에 대한 예측을 통해, 사용자에게 적합한 아이템을 추천하는데 있어 효과적인 방법입니다.

$$\begin{array}{c} \text{Item} \\ \begin{array}{c|c|c|c|c} & W & X & Y & Z \\ \hline A & & 4.5 & 2.0 & \\ B & 4.0 & & 3.5 & \\ C & & 5.0 & & 2.0 \\ D & & 3.5 & 4.0 & 1.0 \end{array} \\ \text{User} \end{array} = \begin{array}{c} \begin{array}{c|c|c} A & 1.2 & 0.8 \\ B & 1.4 & 0.9 \\ C & 1.5 & 1.0 \\ D & 1.2 & 0.8 \end{array} \times \begin{array}{c|c|c|c} W & X & Y & Z \\ \hline 1.5 & 1.2 & 1.0 & 0.8 \\ 1.7 & 0.6 & 1.1 & 0.4 \end{array} \\ \text{Rating Matrix} \qquad \text{User Matrix} \qquad \text{Item Matrix} \end{array}$$

(그림 1)

표 1. MF 모델의 한계점

| 시스템 명 | 기술적인 한계점 |
|-------|---|
| MF 모델 | 사용자는 종종 극단적인 평가를 남기거나, 모든 아이템에 대해 평점을 제공하지 않아서 모델이 사용자의 실제 선호도를 반영하지 못할 수 있습니다. |
| | Explicit 데이터를 이용하는 MF 모델은 사용자의 주관적인 평가로 구성되어 있으며, 이로 인해 평가의 일관성 문제가 발생할 수 있습니다. |

3. BPR 모델 개요

BPR 모델은 협업 필터링 기법의 일종으로, 사용자의 선호를 모델링하는 데 있어 중요한 역할을 합니다. BPR은 순위 기반 손실 함수(rank-based loss function)를 사용하여 사용자가 선호하는 아이템과 비선호 아이템 간의 상대적인 차이를 학습하여 기존의 평점 기반 접근법과 비교하여 더 나은 개인화된 추천 성능을 제공합니다.

또한, BPR 모델은 주로 클릭, 구매, 조회 시간 등과 같이 데이터 대부분이 Implicit 형태로 명시적으로 관심을 표현하지 않아도 되기 때문에, Explicit 데이터보다 훨씬 수집하기가 수월합니다.

4. 구조

BPR 모델은 행렬 분해(Matrix Factorization) 기법을 사용하여 사용자와 아이템의 잠재 요인을 학습합니다. 모델의 주요 구성 요소는 다음과 같습니다:

- 사용자 임베딩(User Embedding): 각 사용자를 고차원 잠재 공간에 매핑하는 임베딩 벡터를 학습하여 사용자의 선호를 모델링합니다.
- 아이템 임베딩(Item Embedding): 각 아이템을 고차원 잠재 공간에 매핑하는 임베딩 벡터를 학습하여 아이템의 특징을 모델링합니다.

BPR 모델은 다음과 같은 수학적 표현으로 정의됩니다:

$$\hat{r}_{ui} = \mathbf{p}_u^\top \mathbf{q}_i$$

$$\hat{r}_{uj} = \mathbf{p}_u^\top \mathbf{q}_j$$

\hat{r}_{ui} = 사용자 u가 아이템 i를 선호할 확률

\hat{r}_{uj} = 사용자 u가 아이템 j를 선호할 확률

\mathbf{p}_u = 사용자 u의 임베딩 벡터

$\mathbf{q}_i, \mathbf{q}_j$ = 각각 아이템 i와 j의 임베딩 벡터

5. 학습과정

BPR 모델의 학습과정은 순위 기반 손실 함수를 최소화 하는 것을 목표로 합니다. 다음과 같은 손실 함수를 사용하여 모델을 학습합니다

$$L = - \sum_{(u,i,j) \in \mathcal{D}} \log \sigma(\hat{r}_{ui} - \hat{r}_{uj})$$

\mathcal{D} = 훈련 데이터 셋

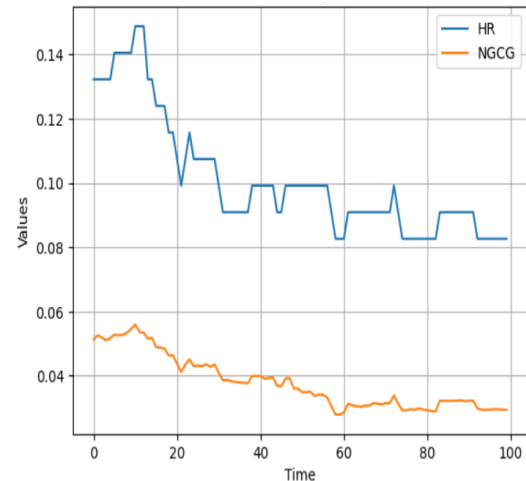
σ = sigmoid 함수

여기서 손실 함수는 사용자가 선호하는 아이템에 대한 점수가 비선호 아이템에 대한 점수보다 높아지도록 학습됩니다.

6. 구현 및 평가

본 논문에서는 BPR 모델을 PyTorch 프레임워크를 사용하여 구현하였으며, 학습 데이터와 테스트 데이터를 준비하기 위해 부정적 샘플링 기법을 적용하였습니다. 구현 과정은 데이터 로드, 모델 훈련, 예측 및 평가 단계로 나누어져 있으며, 각 단계에서의 세부 사항은 다음과 같습니다:

- 데이터 로드: 학습 데이터와 테스트 데이터를 로드하고, 부정적 샘플링을 통해 훈련 데이터를 생성합니다.
- 모델 훈련:SGD(Stochastic Gradient Descent)를 사용하여 모델을 학습합니다.
- 예측 및 평가: 학습된 모델을 기반으로 추천을 수행하고, HR 및 NDCG 지표를 통해 추천 성능을 평가합니다. (그림 2 참고)



(그림 2)

7. 결론

부산은 다양한 음식점과 관광 명소로 잘 알려져 있지만, 관광객들은 많은 선택지 속에서 최적의 장소를 찾는 데 어려움을 겪습니다. 본 논문에서 제안한 BPR 기반 추천 시스템은 관광객의 개인적인 취향과 과거 행동 데이터를 분석하여, 각 관광객에게 맞춤형 음식점 추천을 제공합니다.

이러한 개인화된 추천 시스템은 관광객들의 부산에서의 식사 경험을 더욱 만족스럽게 만들어주며, 이는 재방문율을 높이고 새로운 관광객을 유치하는 데 중요한 역할을 할 것입니다. 또한, 제안된 BPR 기반 추천 시스템은 부산의 다양한 음식 문화와 관광 자원을 더욱 효과적으로 홍보하는 도구로 활용될 수 있을 것으로 기대됩니다.

참고문헌

[1] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner and Lars Schmidt-Thieme, "BPR: Bayesian Personalized Ranking from Implicit Feedback",

Machine Learning Lab, University of Hildesheim
Marienburger Platz 22, 31141 Hildesheim, Germany

[2] Hao Wang, Huawei Shen, Wentao Ouyang, Xueqi Cheng, “Exploiting POI-Specific Geographical Influence for Point-of-Interest Recommendation”,
1 Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China 2University of Chinese Academy of Sciences, Beijing, China

[3]부산관광공사,2023 년 2,3,4 분기 부산관광기업 모니터링 조사,2023 부산 방문 관광객 실태조사