

아이템 공동 등장 엔트로피를 고려한 번들 표현을 통한 번들 추천 기법

이재리, 강유*
서울대학교

{jlunits2, ukang*}@snu.ac.kr

Representing Bundles with Co-Occurrence Entropy of Items for Bundle Recommendation

Jaeri Lee, U Kang*
Seoul National Univ.

요약

번들 추천에서 번들 내 아이템의 역할을 효과적으로 반영하여 정확한 추천을 제공하는 방법은 무엇인가? 효과적인 번들 추천은 관련이 있는 여러 아이템을 일괄 제공함으로써 사용자 만족도를 높이고 매출 향상에 기여한다. 그러나 기존 번들 추천 기법은 주로 번들을 단일 임베딩으로 표현하거나 단순히 구성 아이템들의 평균으로 처리함으로써, 번들 내 아이템 간의 중요도 차이를 포착하는데 한계가 있다. 본 연구에서는 번들 내 아이템의 공동 등장 엔트로피로 각 아이템의 중요도를 평가하고, 중요 아이템을 이용하여 번들을 표현하는 기법을 제안한다. 이는 각 번들에서 문맥적으로 중요한 의미를 가진 아이템을 선별해, 번들을 정확하게 표현한다. 실험 결과, 제안한 기법은 대표 번들 추천 모델인 CrossCBR 대비 Recall@10 에서 15.15%, nDCG@10 에서 12.55%의 성능 향상을 달성하였다.

I. 서론

번들 추천에서 번들 내 아이템의 역할을 효과적으로 반영하여 정확한 추천을 제공하는 방법은 무엇인가? 추천 시스템은 온라인 데이터의 증가에 맞춰 발전해왔으며, 이러한 변화에 따라 사용자들의 복잡한 선호를 더 잘 반영하기 위해 번들 추천이 등장했다. 번들 추천은 단일 아이템 추천을 넘어 사용자들에게 여러 아이템을 한 번에 추천하여 더 나은 편의성과 개인화된 경험을 제공한다[1].

그러나 기존 번들 추천 기법[2, 3]은 번들을 고정된 단일 임베딩 또는 번들 내 아이템의 단순 평균으로 표현하여, 번들 내 아이템 간의 차이를 고려하지 않는 문제가 있다. <그림 1>은 번들 내 각 아이템의 중요성이 다르며, 같은 아이템이라도 다른 번들 내에서는 중요성이 달라질 수 있음을 보인다. 번들 A 인 아보카도 타코에서는 아보카도와 토마토는 빠질 수 없는 재료로 중요 아이템이며, 양파는 부가적인 아이템이다. 반면 구운 양파 레시피인 번들 B 에서는 양파가 중요 재료가 된다. 이처럼 모든 아이템은 다른 중요성을 가지며, 어떠한 번들에 속하냐에 따라 아이템의 중요성이 달라진다.

따라서 본 연구에서는 번들 내 아이템의 공동 등장 엔트로피를 이용해 각 아이템의 중요도를 평가하고, 중요 아이템을 이용하여 번들을 표현하는 기법을 제안한다. 여러 실제 데이터셋에서 실험 결과, 제안한 기법은 대표 번들 추천 모델인 CrossCBR 대비 Recall@3 에서 최대 13.04%, nDCG3 에서 최대 10.29%의 성능 향상을 달성하였다.



번들 A: 아보카도 타코



번들 B: 구운 양파

<그림 1> 번들 내 아이템의 상이한 중요성. 각 번들의 중요 아이템은 컬러로 표시하였다.

II. 제안 기법

각 번들에서 중요한 의미를 갖는 아이템을 어떻게 찾을 수 있을까? 본 연구에서는 아이템의 중요성을 평가하기 위해 아이템의 공동 등장 엔트로피를 이용한다. 각 번들 조합에서의 아이템의 중요성을 파악하기 위해, 아이템의 특정 번들 내에서의 역할과 전체 번들에서 역할을 파악한다. 이를 위해 번들 내에서의 공동 등장 엔트로피와 시스템 전체에서의 공동 등장 엔트로피를 비교한다.

공동 등장 엔트로피는 한 아이템이 같은 그룹에 속한 다른 여러 아이템과 공통적으로 많이 소비되는지를 측정한다. 엔트로피가 높을수록 아이템은 다양한 아이템과 여러 조합으로 소비되며, 낮을수록 특정한 조합으로만 소비된다. 어떠한 아이템 i 의 그룹 \mathcal{X} 내에서의 공동 등장 엔트로피는 다음과 같이 정의된다:

$$H(i, \mathcal{X}) = - \sum_{j \in \mathcal{X}, j \neq i} p(j|i, \mathcal{X}) \log p(j|i, \mathcal{X}),$$
$$p(j|i, \mathcal{X}) = \frac{c(i, j)}{\sum_{j' \in \mathcal{X}, j' \neq i} c(i, j')},$$

여기서 $c(i, j)$ 는 아이템 i 와 아이템 j 가 같은 번들에서 등장하는 빈도이다.

참 고 문 헌

- [1] Hyunsik J., Jeong-eun L., Jeongin Y., and U K., "Cold-start Bundle Recommendation via Popularity-based Colaescence and Curriculum Heating", WWW, 2024
- [2] Yunshan M., Yingzhi H., An Z., Xiang W., and Tas-Sng C., "CrossCBR: Cross-view Constrastive Learning for Bundle Recommendation", KDD, 2022
- [3] Jianxin C., Chen G., Xiangnan H., Yong L., and Depeng J., "Bundle Recommendation with Graph Convolutional Networks", SIGIR, 2020
- [4] Zhankui H., Handong Z., Tong Y., Sungchul K., Fan D., and Julian M., "Bundle MCR: Towards Conversational Bundle Recommendation", RecSys, 2022

어떠한 아이템 i 의 번들 $b = \{i_1, \dots, i_{|b|}\}$ 내 공동 등장 엔트로피 $H(i, b)$ 가 낮다는 것은, 아이템이 해당 번들 내 특정 아이템과 연관이 높아 아이템이 번들 내에서 중요한 아이템임을 나타낸다. 반면 아이템 i 의 전체 아이템 \mathcal{I} 에서의 공동 등장 엔트로피 $H(i, \mathcal{I})$ 가 높다는 것은, 아이템이 여러 번들에 잦은 빈도로 포함되어 특정 번들에서 중요도를 가질 확률이 낮다는 것이다. 따라서 번들 b 내 공동 등장 엔트로피가 $H(i, b)$ 가 전체 아이템 \mathcal{I} 에서의 공동 등장 엔트로피 $H(i, \mathcal{I})$ 보다 낮은 경우, 번들이 해당 번들 b 에서 중요한 의미를 가진다. 따라서 어떠한 아이템 i 의 번들 b 에서의 중요도는 다음과 같이 정의된다:

$$s(i, b) = \frac{H(i, \mathcal{I})}{H(i, b)}$$

각 아이템의 중요도를 번들 표현에 반영하기 위해, 각 번들의 아이템 임베딩을 중요도에 비례하게 합산하는 번들 임베딩을 제안한다. 어떠한 번들 b 의 임베딩 \mathbf{e}_b 는 다음과 같이 정의된다:

$$\mathbf{e}_b = \sum_{i \in b} s(i, b) \mathbf{e}_i,$$

여기서 $\mathbf{e}_i \in \mathbb{R}^d$ 는 아이템 i 의 학습 가능한 임베딩이다. 이러한 번들 표현을 일반 아이템 추천 모델과 결합하여 최종적인 번들 추천을 한다.

IV. 실험 결과

제안한 기법의 성능을 검증하기 위해, 제안한 기법을 대표 번들 추천 시스템인 CrossCBR[2]과 실제 데이터셋에서 비교한다. 검증 데이터셋으로는 추천 시스템 평가에서 가장 널리 쓰이는 데이터셋인 MovieLens 데이터셋을 활용한다. MovieLens는 영화 리뷰 사이트이며, 데이터셋은 사용자가 남긴 영화의 기록이다. 기존 번들 논문[4]과 같이 같은 시점에 리뷰 된 영화를 번들로 취급한다. Leave-one-out 기법을 사용하여, 가장 마지막 상호작용을 검증세트, 나머지 상호작용을 훈련세트로 사용한다. 상위 10개의 추천 결과에서, 추천 순서에 상관없는 Recall@10와 순서를 반영한 nDCG@10로 정확도를 평가한다. 실험 결과, CrossCBR은 0.0462의 Recall@10와 0.0247의 nDCG@10을 보이며, 제안한 기법은 0.0532의 Recall@10와 0.0278의 nDCG@10을 보인다. 제안한 기법은 최대 15.15%의 향상을 보이며 우수한 번들 추천 성능을 달성한다.

III. 결론

본 논문에서는 정확한 번들 추천을 위해 번들 내 아이템의 공동 등장 엔트로피를 이용해 각 아이템의 중요도를 평가하고, 중요 아이템을 이용하여 번들을 표현하는 기법을 제안한다. 제안한 번들 표현 기법은 번들 내 아이템의 의미를 정확하게 파악하여, 대표 번들 추천 모델인 CrossCBR 대비 Recall@10에서 15.15%, nDCG@10에서 12.55%의 성능 향상을 달성하였다.

ACKNOWLEDGMENT

본 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(NO.RS-2021-II211343, 인공지능대학원지원(서울대학교))