

추천 시스템을 위한 콜드-웜 아이템 다양화 방법

한정규, 천세진
동아대학교 컴퓨터 AI 공학부
e-mail : jkhan@dau.ac.kr, sjchun@dau.ac.kr

A Cold-Warm Item Diversification Method for Recommender System

Jungkyu Han, Sejin Chon
Division of Computer and AI Engineering, Dong-a University

요 약

사용자와의 상호작용 데이터(클릭 등)가 적어 협업필터로 정확도 높은 추천을 하기 힘든 콜드 아이템들에 대해, 상호작용과 상관없이 추출할 수 있는 아이템의 컨텐트 정보로부터 본래 협업 필터를 사용하여 학습하는 행위기반 특징을 “예측”하는 모델들이 최근에 연구되고 있다. 이들 연구에서는 콜드 아이템만을 추천 후보로 하여 모델이 사용자 취향에 부합하는 콜드 아이템을 얼마나 잘 찾아내는가를 평가하기에, 콜드 아이템과 상호작용 데이터가 충분히 존재하는 웜 아이템이 혼재된 추천 풀에서 아이템을 선택해야 하는 실제 추천 환경에서 사용자 취향에 맞는 콜드 아이템이 취향에 맞지 않는 웜 아이템을 제치고 사용자에게 잘 추천되는지를 알 수 없었다. 본 연구에서는 실제 추천 환경에서 행위기반 취향 정보를 예측하는 기존 모델의 콜드 아이템 추천 성능과 개선점을 분석하고, 사용자 취향에 부합하는 콜드 아이템과 웜 아이템을 조화롭게 추천하기 위해 추천 리스트의 아이템 콜드/웜 분포를 고려한 MMR 기반의 아이템 다양화 방법을 제안한다.

1. 서 론

협업필터 (Collaborative Filter)는 사용자의 클릭 등 사용자와 아이템의 상호작용 정보를 바탕으로 사용자의 취향과 아이템의 특징을 모델링하기 때문에 개인화된 추천에 필수적으로 사용되는 기술이나 상호작용 정보를 바탕으로 학습하기 때문에 예를 들어 시스템에 새로 등록된 아이템 등 상호작용 정보가 적은 콜드 아이템에 대해서는 그 성능을 완전히 발휘하기 힘들다. 이를 개선하기 위해 인공지능 경향을 활용하여 아이템의 제목, 카테고리, 태그 등, 컨텐트 정보로부터 협업필터가 학습하는 상호작용행위 기반의 특징 (이하, 행위기반 특징)를 예측하는 모델이 최근에 연구되어 왔다. [1][2][3]

이들 컨텐트 정보로부터 행위기반 특징을 예측하는 모델은 공통적으로 콜드 아이템만을 추천 후보 풀에 두고 모델이 사용자 취향에 부합하는 콜드 아이템을 얼마나 잘 찾아내는가를 평가하였다. 이는 모델의 순수한 콜드 아이템 추천 성능을 평가하는데는 유용하나, 콜드 아이템과 상호작용 데이터가 충분히 존재하는 웜 아이템이 혼재된 추천 풀에서 아이템을 선택해야 하는 실제 추천 환경에서 사용자 취향에 맞는 콜드 아이템이 그렇지 않은 웜 아이템을 제치고 취향에 부합하는 사용자에게 잘 추천되는지를 알 수 없다. 일반적으로 협업 필터는 상호작용 데이터가 충분한 아이템이 적은 아이템보다 아이템의 행위기반 특징을 강하게 학습할 수 있기 때문에 웜 아이템이 더 잘 추천되는 경향이 있다.

본 논문에서는 협업 필터로 행위기반 특징을 직접 학습한 웜 아이템과 컨텐트로부터 행위기반 특징을 예측한 콜드 아이템이 혼재한 상황에서 웜 아이템에 편향될 것으로 예측되는 추천 경향이 어떻게 변화하는지를 분석하고, 추천 정확도를 유지하면서 편향을 완화하기 위해 추천 리스트의 아이템 콜드/웜 분포를 고려한 MMR[3] 기반의 아이템 다양화 방법을 제안한다.

2. 관련 연구

2.1 콜드 아이템 추천

콜드 아이템의 정확한 추천을 위해 인공지능 경향을 활용하여 아이템의 제목, 카테고리, 태그 등, 컨텐트 정보로부터 협업필터가 학습하는 행위기반 특징을 예측하는 모델이 최근에 연구되어 콜드 아이템 추천 성능을 개선함을 보였다 [1][2]. 다만 이들은 모델의 순수한 콜드 아이템 추천 성능만 평가하였고 콜드 아이템과 상호작용 데이터가 충분히 존재하는 웜 아이템이 혼재된 추천 풀에서 아이템을 선택해야 하는 실제 추천 환경에서의 성능은 평가하지 않았다.

2.2 추천결과 다양화

추천 결과를 다양화하기 위해 MMR[3] 방식을 사용한 탐욕 알고리즘(Greedy Algorithm) 기반의 추천 다양화 기법이 제안되어 왔다. [4][5] 이들은 추천되는 아이템 카테고리의 다양화를 고려한 목적함수를 개발하고 이를 탐욕

알고리즘을 사용하여 최대화시키는 방법을 사용하였다.

3. 기존모델 성능분석 및 제안 기법

이 장에서는 기존 모델을 사용한 성능 분석 환경과 분석 내용에 대해 설명하고, 추천 정확도를 유지하면서 워밍 아이템의 편향을 완화하기 위해 MMR[3] 기반의 다양화 알고리즘을 응용하는 방식을 설명한다.

3.1. 성능 분석

(그림 1) 성능분석 환경

상호작용데이터가 존재하지 않는 아이템을 콜드 아이템, 상호작용 데이터가 하나 이상 존재하는 아이템을 워밍 아이템으로 구분한 다음, 그림 1과 같이 기존의 협업 필터 방법을 사용하여 행위기반 특징을 학습한 워밍 아이템의 임베딩 벡터(Embedding Vector)와, 아이템 콘텐츠 정보로부터 행위기반 특징을 예측하는 모델을 사용하여 예측한 콜드 아이템의 임베딩 벡터를, 추천 타겟 사용자의 임베딩 벡터와 내적을 통하여 계산한 사용자 취향 점수가 높은 순으로 정렬하고 그중 상위 k 개의 아이템을 대상으로 분석한다.

분석을 위해 상위 k 개의 아이템 중, 워밍 아이템과 콜드 아이템의 비율 (워밍 아이템에 편향되지 않고, 콜드 아이템이 적절히 추천되는가?), 추천한 k 개의 아이템 중 사용자가 실제로 상호작용한 정답 아이템이 몇 개인가 (전체 시스템의 추천 정확도), 워밍 아이템, 콜드 아이템에 대해서 정답을 맞춘 비율 (각각의 아이템 그룹에 대한 추천 정확도)를 측정한다. 이를 바탕으로 워밍 아이템과 콜드 아이템이 혼재된 환경에서 기존모델을 사용하였을 경우의 효과에 대해서 분석한다.

3.2. 제안 기법

제안 기법은 워밍 아이템, 콜드 아이템, 타겟 사용자의 임베딩 벡터가 이미 계산되어 있는 상태에서 식 1의 값을 최대화하는 k 개의 아이템을 선택하는 후처리 (Post Processing) 방식이다.

$$obj_k = \lambda \times REL_k + (1 - \lambda) \times Div_k \quad (\text{식 1})$$

식 1에서 REL_k 는 이미 선택한 k 개의 아이템의 취향 점수의 합으로, 워밍 아이템은 $w_{warm} \times (u \cdot i_w)$, 콜드 아이템은 $w_{cold} \times (u \cdot i_c)$ 으로 계산된다. u , i_w , i_c 는 사용

자, 워밍 아이템, 콜드 아이템 각각의 임베딩 벡터이고, w_{warm} , w_{cold} 는 각각의 가중치이다. 이 가중치는 데이터 분석에 기반하여 시스템 운영자가 결정하거나 데이터로부터 학습한다.

DIV_k 는 k 개 아이템 중에 콜드 아이템과 워밍 아이템의 비율을 측정하는 지표이며, ILD(Intra-List-Diversity)[5], 비율 균형[4] 등의 지표를 사용할 수 있다. λ 는 REL_k 와 DIV_k 의 혼합 비율을 결정하는 시스템 운영자가 설정하는 가중치로 λ 값이 클수록 시스템의 추천 정확도가 높아지고, 작을수록 콜드 아이템과 워밍 아이템의 추천 개수가 시스템에서 정의한 균형을 이룬다.

식 (1)을 최대화 하기위해 MMR[3]에서 설명한 탐욕 알고리즘을 사용한다. 그림 1의 결과 생성된 아이템 리스트 전체 L 에서 최상위 $w_{warm} \times (u \cdot i_w)$ (혹은 Warm 대신 Cold) 값을 가지는 아이템을 하나 선택하고 L 에서 남아 있는 아이템 중에서 식 (1)의 추천하였을 때 값을 최대로 증가시키는 아이템을 하나하나 선택해가면서 최종적으로 k 개 아이템을 선택한다.

4. 결론

본 논문에서는 실제 추천 환경에서 행위기반 취향 정보를 예측하는 기존모델의 콜드 아이템 추천 성능과 개선점을 분석하고, 사용자 취향에 부합하는 콜드 아이템과 워밍 아이템을 조화롭게 추천하기 위해 추천 리스트의 아이템 콜드/워밍 분포를 고려한 아이템 다양화 방법을 제안하였다.

참고문헌

- [1] M. Volkovs, G. Yu, and T. Poutanen. "Dropoutnet: Addressing cold start in recommender systems," Proceedings of Advances in Neural Information Processing Systems, 2017.
- [2] Z. Zhu, S. Sefati, P. Saadatpanah, J. Caverlee, "Recommendation for New Users and New Items via Randomized Training and Mixture-of-Experts Transformation," Proceedings of the 43rd International ACM SIGIR Conference, pp. 1121-1130, 2020.
- [3] J. Carbonell and J. Goldstein. "The use of MMR, diversity-based reranking for reordering documents and producing summaries," Proceedings of SIGIR '98, 1998
- [4] V. Dang and W. Bruce Croft. "Diversity by proportionality: an election-based approach to search result diversification." In Proceedings SIGIR '12, pp. 65-74, 2012.
- [5] C.-N. Ziegler, S.M. McNee, J.A. Konstan, and G. Lausen. "Improving recommendation lists through topic diversification," In Proceedings of WWW '05. pp.22-32. 2005.