

Review : Being Accurate is Not Enough: How Accuracy Metrics have hurt Recommender Systems

Sean M. McNee, John Riedl, Joseph A. Konstan

May 20, 2022

Abstract

-recommender system은 거대한 정보공간 속에서 user가 흥미로워하고 연관있는 아이템을 찾는데 도움을 주는 엄청난 잠재력을 보여주었다.
-지금까지의 대부분 연구는 추천시스템의 정확성을 향상시키는 것에 초점을 맞췄다.
-standard metrics에 따른 높은 정확도가 user들에게 대부분 유용한 것은 아니다.
-저자들은 이 것이 잘못되었다고 판단하여, 새로운 방법을 제시하고자한다.
-recommender systems을 평가하기 위해 new user-centric direction을 제안한다.

Keywords

Recommender systems, personalization, collaborative filtering, metrics

ACM Classification Keywords

H3.3. Information Search and Retrieval: Information Filtering, Search Process, Retrieval Models

Introduction

-예시 : 여행 Recommender system을 사용하는데, 이미 간 곳(선호도 순으로 Ranking을 매겨서 추천하는 것이 뛰어나다 할지라도)만 추천한다면, 이 recommender system을 계속 사용할 것인가?
-현재 standard methodology는 이미 방문한 장소 대신 새로운 장소를 추천하면 불이익을 받게 되는 형태이다.
-MAE[Herlocker 1999]와 같은 현재의 accuracy metrics는 user's rating of an item에 대해 알고리즘의 예측을 비교하여 성능을 측정한다.
-가장 흔하게 사용되는 방법론은 leave-n-out approach[Breese 1988]이다.(전체 dataset의 일정 비율을 recommender한테 받아서, test data로 사용)
-User가 방문한 장소가 아닌 방문할 새로운 장소를 추천해주는 것이 중요하다.

- Current accuracy metrics가 아닌 새로운 측면을 기준으로 recommender system을 검토할 것이다.
- aspects** : the similarity of recommendation lists, recommendation serendipity, the importance of user needs and expectations in a recommender.

Similarity

- 'The item-item collaborating filtering algorithm'은 사용자를 'similarity hole'에 빠지게한다.(example : amazon.com의 Robert Heiemlein의 책만 추천/ star track 영화를 평가하면 더 많은 star track 영화만 추천)
- 이러한 문제는 recommender system에 new user들의 data가 적을 때 더 자주 발생한다.
- 이러한 문제로 인해 user가 recommender system을 영원히 이용 안하게 되는 때가 생긴다.
- 따라서 recommender list는 user가 이미 본 목록의 내용을 추천하는 것이 아니라 유용성으로 판단되어야한다.
- Intra-List Similarity Metric[Ziegler 2005]는 topic의 다양성(list의 item 다양성을 늘리거나 줄일수 있도록 변경)을 제안하였다. accuracy는 더 낮을지라도 user들이 이 변경된 recommender list를 더 좋아하였다.
- user의 의도에 따라 list에 나타나는 item 구성이 정확도 높은 recommender list보다 user들의 만족도가 더 높았다.

Serendipity

- recommender의 serendipity는 뜻밖의 우연한 추천을 받는 경험을 의미한다.
- leave-n-out metric은 높은 정확도를 나타내는 machine learning algorithm이다[Billsus 1998, Breese 1998].
- leave-n-out metric은 user가 소비할 다음 항목이 추천 될 확률이 높은 item을 분류를 통해 user 프로필 다음에 배치한다.
- 하지만 user는 자신이 생각하지 못한 item에 대한 추천을 선호하는 경우가 있다.
- 예를 들어 온라인 음악사이트에서 User-User Collaborative Filtering algorithm를 이용하여, 가장 일반적인 비틀즈의 'White Album'을 추천했다. 이는 정확도는 높을지라도, 모든 user가 'White Album'을 소유하거나 소유하지 않기를 결정했기때문에 가치가 있는 recommender가 아니었다.
- user들은 현재 작업을 기반으로 다양한 recommender algorithm의 list를 선호하였다[McNee 2002, Torres 2004].
- 이것은 recommender algorithm을 분류하는 새로운 방법이 필요함을 나타낸다.
- serendipity metric은 user의 feedback없이 생성되기 어렵지만, recommender algorithm의 다양한 측면을 판단하면, 좀더 상세한 추천이 가능하다.

User Experiences and Expectations

- 신규 user와 기존의 user의 recommender의 needs는 다르다.
- 신규 user는 recommender가 제공하는 list를 활용하기 전에 recommender와의 신뢰 관계를 확립해야

하기 때문에 높은 비율의 항목을 생성하는 알고리즘으로부터 이익을 얻을 수 있다.

-신규 user에게 사용되는 알고리즘 선택이 user의 경험과 생성된 recommender의 정확성에 크게 영향을 미친다는 것을 보여준다[Rashid 2001].

-user의 만족도에는 그들의 언어와 문화적 배경의 차이도 영향을 끼친다[Toress 2004].

-user의 모국어로 추천한 recommender가 다른 언어로 추천된 recommender 보다 더 선호도가 높았다.

Moving Forward

-recommender를 사용하는 user들도 생각해야한다. 그들은 좋은점수를 받은 모델과 상관없이 자신들에게 의미 있는 recommender를 원한다.

-1.user가 recommender list를 보는 것처럼 추천의 품질을 판단해야한다. recommender list에 작용하는 다양한 metrics를 만들어야한다.

-2. 추천 알고리즘간의 차이를 이해하고 평가가능성을 넘어서는 방식으로 이를 측정해야한다.(user들은 추천알고리즘의 차이를 파악가능하다.)

-3. new user부터 기존 user들 까지 각 사용자들의 목적에 맞는 요구사항을 충족시켜줘야 한다.