

[Review]

본 논문은 추천시스템 기술의 바이블이라고 할 수 있는 Matrix Factorization에 대하여 다룬 논문이다. Matrix Factorization은 Neighborhood Method와 Latent Factor Method로 나뉘어지는 Collaborative Filtering Approach 중에서 Latent Factor Method를 구현하는 가장 좋은 방법 중 하나이다.

우선 Collaborative Filtering Approach는 도메인 지식이 없어도 아이템과 유저에 대한 추천이 가능하다는 장점을 가진다. 그러나 새로운 사용자와 아이템을 다루기에 부적합해 Cold Start Problem이라는 단점을 가진다.

Matrix Factorization은 유저와 아이템의 협업으로 생성된 평점 데이터의 패턴 속에서 Latent Factor를 추론해 사용자와 아이템의 특성을 찾아내는 방식으로 구성되어 있다. 이 방식은 높은 정확도, 확장성, 유연성이라는 장점을 가진다고 본 논문의 저자는 이야기한다.

기본적으로 Matrix Factorization은 SVD 방식과 매우 유사하다. 왜냐하면 평점 데이터를 임의의 K개의 차원을 가지는 Latent Factor로 분해하기 때문이다. 그런데 SVD 방식을 직접적으로 추천시스템에 직접적으로 적용하는 것은 무리가 있다. SVD의 경우 결측치가 존재하면 안되는데 추천 시스템의 평점 데이터에는 수 많은 결측치가 존재하기 때문이고 이 결측치를 임의로 채운다면 데이터의 왜곡 가능성이 매우 높아진다. 따라서 오직 관측된 평점만을 직접적으로 모델링하는 Matrix Factorization 방법을 제시하게 되었다.

Matrix Factorization은 크게 아이템의 Latent Factor와 유저의 Latent Factor를 동시에 추론하는 SGD 방식과 하나의 Latent Factor를 먼저 추론한 후에 그 다음의 Latent Factor를 추론하는 ALS 방식으로 나뉘어진다. SGD 방식이 간단한 방법이지만 시스템의 병렬화가 가능하고 암시적 데이터에 집중되어 있다면 ALS가 더 높은 효과를 보인다고 한다.

그런데 실제에 추천시스템 모델을 구축하기 위해서는 사용자의 행동 정보를 반영할 필요가 있으며 현실에서 제품에 대한 인식과 인기 등은 시간에 따라 변화한다. 따라서 추천 시스템은 시간에 따라 변하는 사용자와 아이템 간의 상호작용의 동적인 성질을 반영하여 Temporal Effect에 대하여 설명할 수 있어야 한다.

본 논문에서 제시한 Matrix Factorization은 추천 알고리즘 대회인 Netflix Prize에서 1등을 했으며 현재 까지도 추천 시스템에서 많이 활용되는 바이블과 같은 기술이다.

[본문 URL]

<https://velog.io/@tobigs-recsys/MATRIX-FACTORIZATION-TECHNIQUES-FOR-RECOMMENDER-SYSTEM>