

오토인코더 기반의 그룹 추천을 위한 사용자 클러스터링

임재형, 박진태, 박대영
인하대학교

Jhlim2993@gmail.com, qkrwlsxo1954@gmail.com, dpark@inha.ac.kr

User Clustering for Group Recommendation using Autoencoder

Jaehyung Lim, Jintae Park, Daeyoung Park
Inha University

요 약

다양한 플랫폼에서 사용자 맞춤형으로 아이템을 추천해주는 추천시스템에 대한 연구가 활발하다. 하지만 사용자의 수와 아이템의 수가 늘어날수록 사용자 맞춤형 추천을 위한 연산의 수가 급격하게 증가하므로 그룹 추천시스템에 대한 관심도 높아지고 있다. 비슷한 취향의 사용자를 그룹으로 묶어야 추천 품질이 높아지므로 본 연구에서는 해당 추천의 품질을 높이기 위해 오토인코더를 사용해서 그룹 형성 과정에 대한 방법을 제안한다. MovieLens 데이터에 여러 그룹 추천 연구의 추천 알고리즘을 사용하여 오토인코더를 이용한 사용자 클러스터링의 그룹 추천 효과를 검증한다.

I. 서 론

각 사용자들에게 맞춤형으로 알맞은 서비스를 추천해주는 개인화 추천시스템 연구가 활발하게 진행되고 있다. 하지만 플랫폼을 사용하는 사용자의 수와 아이템(서비스 품목)의 종류가 늘어남에 따라 사용자에게 아이템을 추천하기 위해 필요한 계산 복잡도가 매우 높은 단점이 있다. 취향이 비슷한 사용자에게 비슷한 아이템을 추천해주는 그룹 추천시스템은 낮은 복잡도에도 추천 품질이 높으므로 개인화 추천의 대안으로 주목받고 있다.

기존에는 그룹 추천을 위해 추천의 성능을 높이는 추천 알고리즘 위주로 연구되었으며, 그룹을 형성하는 과정에서는 랜덤하게 사용자를 그룹으로 만들거나, 휴리스틱으로 결정되었다. 예를 들어서 동시에 특정한 활동을 했던 사람들을 같은 그룹으로 간주하거나 [1], 정해진 개수 이상의 아이템을 동시에 사용한 사용자들을 그룹으로 간주하였다 [2].

그룹 추천을 위해 사용자 임베딩을 클러스터링하여 그룹을 만드는 연구[3]는 행렬 분해 방법[4]으로 얻은 사용자 행렬의 벡터를 사용자 임베딩으로 활용하여 클러스터링하는 방법이다. 행렬 분해 방법에서 사용자 행렬을 얻는 방식은 사용자-아이템 평점 행렬을 두 행렬의 곱으로 나타내는 방식이다. 개인화 추천에서는 이러한 행렬 완성(matrix completion)을 보완하기 위해 입력과 출력의 데이터 값을 같도록 신경망의 파라미터를 학습하는 오토인코더(autoencoder)를 활용하였다 [5]. 본 논문에서는 사용자 임베딩의 선형적 특성을 보완하기 위해 오토인코더 신경망의 가운데 은닉층의 데이터(Latent vector)를 사용자 임베딩으로 사용하여 벡터 공간에서 사용자를 클러스터링하여 그룹 추천하고자 한다. 그리고 이 과정에서 얻은 그룹 정보를 바탕으로 그룹 추천 알고리즘을 적용하여 사용자들에게 만족스러운 추천했는지에 대한 성능을 확인해보고자 한다.

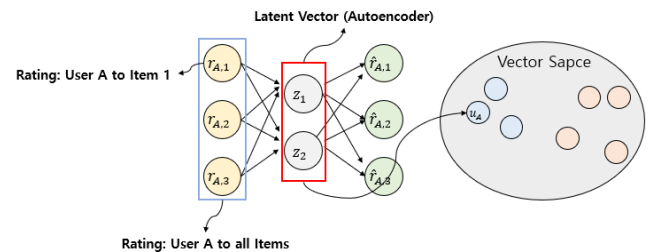


그림 1. 오토인코더를 이용한 사용자 클러스터링

II. 제안하는 사용자 클러스터링 방법

오토인코더를 이용하는 사용자 클러스터링 방법을 설명하고 실험 설계를 제시한 후에 제안하는 방법이 우수한 추천 품질을 나타내는 실험 결과를 보여준다.

2.1 오토인코더를 활용한 클러스터링

기존에는 사용자-아이템 평점 행렬의 결측치를 예측하기 위해 행렬 분해 방법으로서 사용자-아이템 평점 행렬을 사용자 행렬과 아이템 행렬의 곱으로 분해하게 된다 [4]. 사용자 행렬의 각 행은 사용자의 특징을 나타낸다고 생각할 수 있으므로, K-Means 알고리즘 사용자 임베딩인 각 행벡터에 적용하여 클러스터링 하였다 [3].

한편 개인화 추천을 위한 행렬 완성을 위해 오토인코더를 이용해서 평점 행렬의 결측치를 채우는 연구가 있었다 [5]. 오토인코더 신경망의 입력 데이터는 한 명의 사용자의 모든 아이템에 대한 평점이 된다. 따라서 본 연구에서는 그림 1 과 같이 신경망 가운데 은닉층의 데이터 (Latent vector)를 사용자의 평점 정보를 벡터 공간으로 투영한 사용자 임베딩 벡터들을 K-Means 알고리즘을 사용하여 클러스터링한다.



그림 2. 그룹 추천 실험 결과

2.2 실험 설계 및 실험 결과

실험에서 데이터셋은 추천시스템 연구에서 사용되는 MovieLens-100k (ML-100K)와 MovieLens-1m (ML-1M)를 사용하였다. ML-100K는 사용자 수가 943명, 아이템의 수가 1682개이고, 사용자가 부여한 평점의 개수는 100,000개이다. ML-1M은 사용자 수가 6040명, 아이템의 수가 3952개이고, 사용자가 부여한 평점의 개수는 1,000,000개이다.

ML	# Users	# Items	# Ratings	Sparsity
100K	943	1682	100,000	0.9370
1M	6040	3952	1,000,000	0.9581

표 1. MovieLens 데이터셋

각 데이터셋 별로 입력으로 주는 아이템의 수가 다르기 때문에 ML-100K 실험은 은닉층의 셀의 개수를 1100개로, ML-1M 실험은 셀의 개수를 4000개로 하였고, 이 값은 사용자 임베딩 벡터의 차원이 된다.

클러스터링 이후의 그룹 추천을 시행하기 위해 그룹 추천 비교 알고리즘으로 주요하게 사용되는 2가지 Group Aggregation Method인 Borda Count (BC)와 Copeland Rule (CR)을 사용하였다 [6]. 추천 성능은 추천된 아이템이 실제 사용자가 선호하는지에 대한 지표인 NDCG와 Recall을 사용한다 [6]. 또한 NDCG 및 Recall은 모두 20개를 추천한 경우로 실험을 진행하였다.

2.3 실험 결과

그림 2는 본 연구에서 제안하는 사용자 클러스터링의 실험 결과를 나타낸다. 붉은색 선은 오토인코더를 이용한 클러스터링, 파란색 선은 [3]에서 제안한 행렬 분해를 이용한 클러스터링 방식을 이용한 결과이다. 실선은 NDCG를 이용한 결과이고, 점선은 Recall의 결과이다. 그룹의 개수는 64개부터 5개의 간격으로 254개까지 진행하였고, 각 데이터셋 별, 그룹 추천 알고리즘 별로 그래프로 구성하였다. 그림 2에서 ML-모든 데이터셋에 BC와 CR 그룹 추천 알고리즘을 적용했을 때, 64~254개의 그룹 개수에 대해서 오토인코더를 이용한 클러스터링이 행렬 분해를 이용한 클러스터링보다 NDCG와 Recall의 값이 더 높은 것을 관찰할 수 있다. 따라서, 오토인코더의 은닉층의 latent 벡터가 행렬 분해로 얻는 것보다 사용자의 특징을 추출하는데 있어서 더 낫다는 것을 확인할 수 있다.

III. 결론

본 논문에서는 기존의 그룹 추천시스템 연구에서 집중하지 않은 그룹 형성에 초점을 맞춘 연구를 제안한다. 기존 행렬 분해 모델에서 사용자 그룹화를 제안한 것처럼, 본 논문은 오토인코더에 사용자의 평점

정보를 입력으로 넣어줄 때 은닉층의 latent 벡터를 사용자 임베딩으로 사용하여 클러스터링을 진행한다. 이 그룹의 결과를 바탕으로 각 그룹의 사용자들에게 BC와 CR 알고리즘을 적용하여 그룹 추천을 해보고 실제 추천의 성능을 검증했다. 실험 결과 실험에 사용한 모든 데이터셋과 그룹 추천 알고리즘에 대하여 비교 모델인 행렬 분해보다 NDCG와 Recall 성능지표에서 우수한 결과를 얻었다. 이는 선형 연산으로 사용자 임베딩을 추출한 방법에 비하여 오토인코더는 비선형 연산으로 사용자 임베딩을 얻기 때문에 벡터 공간에서의 사용자에 대한 표현력이 더 좋다고 볼 수 있는 것이다.

IV. 후속 연구

향후에는 오토인코더의 은닉층의 개수를 늘리거나, 다른 신경망을 적용하여 사용자 임베딩을 얻어 이를 클러스터링 할 수 있을 것이다. 또한 K-Means 이외의 클러스터링 알고리즘에 적용하여 더욱 성능을 높일 수 있을 것이다.

ACKNOWLEDGMENT

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. RS-2022-00155915, 인공지능융합혁신인재양성(인하대학교)). 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2022R1H1A2092541).

참고 문헌

- [1] L. Yu, Y. Leng, D. Zhang, and S. He. "Collaborative group embedding and decision aggregation based on attentive influence of individual members: A group recommendation perspective." accepted in Decision Support Systems, 2022.
- [2] G. Fang, L. Su, D. Jiang, and L. Wu. "Group recommendation systems based on external social-trust networks," Wireless Commun. Mobile Comput., vol. 2018, May 2018.
- [3] J. Shi, B. Wu, and X. Lin. "A latent group model for group recommendation," in IEEE Int. Conf. Mobile Services, pp. 233-238, 2015.
- [4] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems." Computer, vol. 42, no. 8, pp. 30-37, Aug. 2009.
- [5] S. Sedhain, A. K. Menon, S. Sanner, and L. Xie. "AutoRec: Autoencoders meet collaborative filtering," in Proc. Int. Conf. World Wide Web, pp. 111-112, May 2015.
- [6] S. Dara, C. R. Chowdary, and C. Kumar, "A survey on group recommender systems," J. Intelligent Information Systems, vol. 54, no. 2, pp. 271-295, Jan. 2020.