**Neural Collaborative Filtering – 16기 김건우**

* **Abstract**

최근 들어, 딥러닝 기술은 computer vision이나 NLP 등에는 많이 활용되었지만, 추천시스템에 적용된 연구는 상대적으로 적었다. 그래서 이 논문에서 추천 시스템 중collaborative filtering의 문제점을 neural network로 다루는 것을 보여준다.

Matrix Factorization은 user-item latent features들의 내적을 통해 학습을 하는 방법인데, 이 논문은 내적을 neural architecture로 대체하며 학습하는 NCF, Neural Network-based Collaborative Filtering을 제시한다. 선형을 가정한 MF의 한계를 지적하며 비선형으로 더 general한 모델인 NCF를 설명한다.

* **Introduction**

추천 시스템에서는 item기반인 collaborative filtering 방법론 중 MF, Matrix Factorization,이 가장 인기가 많았다. 이는 user와 item을 나타내는 latent vector들의 내적을 통해 모델링 된다. 최근에는 DNN을 추천 시스템에 접목하는 것도 많이 제안 되었는데, 대부분 비정형 데이터의 features를 추출하는 용도로만 사용되고 linear한 MF의 기저인 내적을 활용하는 것은 여전히 사용되었다. 하지만 이 논문에서는 implicit data를 활용해 non-linear한 DNN을 활용하며 collaborative filtering을 접근하는 방법을 보여줄 것이다.

1. NCF (Network-based Collaborative) 방법론 설명
2. MF는 NCF의 특별한 경우
3. 실험을 통해 NCF의 효율성 입증

* **Preliminaries**
  + **Learning from Implicit data**

M명의 사용자와 N개의 items가 있다고 할 때, MxN행렬을 다음과 같이 표기해보자.

A picture containing text

Description automatically generated

여기서 ‘1’의 값은 특정 사용자와 item 간에 상호작용이 있다는 것이지만 사용자가 그 item을 좋아한다는 것은 아니다. 마찬가지로 ‘0’의 값도 사용자와 item 간에 관계가 없다는 것이고 사용자가 item을 싫어한다는 것도 아니다. 그래서 이 문제를 machine learning 관점으로 풀 때, y가 ‘1’이 되는 확률을 예측하는 문제로 접근한다.

를 추정할 때, 대체로 두가지 방법으로 추정할 수 있다.

1. Pointwise Loss
2. Pairwise Loss

NCF는 DNN을 f로 활용을 하고 위에 두가지 학습을 다 할 수 있다.

* + **Matrix Factorization**

Text

Description automatically generated with medium confidence

MF는 user와 item의 latent vector의 내적으로 추정한다.

이 논문은 MF의 표현에 대한 한계가 있음을 다음 그림을 예시로 들며 설명한다.

Chart, diagram

Description automatically generated with medium confidence

(a)를 기반으로 는 i행과 j행의 코사인 유사도를 나타내고, (b)는 user간의 관계를 기하학적으로 나타낸다.

인 상황에서, 새로운 user 4가 들어올 때, 문제가 생긴다고 볼 수 있다.

User 4는 다음과 같은 유사도를 따르는데, 이를 (b)와 같이 이 관계를 표기하려면 불가능하다.

이 문제를 해결할 수 있는 유일한 방법은 K차원을 매우 높이는 것이지만, 이는 모델의 일반화를 해치므로 과적합을 유발할 수 있다. 그래서 DNN을 활용한 방법론이 제시된다.

* **Neural Collaborative Filtering**
  + **General Framework**

Diagram

Description automatically generated

Input: 다음 그림은 논문에서 제시하는 NCF모델의 개요이다. 밑에 input layer로는 user와 item을 나타내는 binary one-hot-encoding 벡터이다.

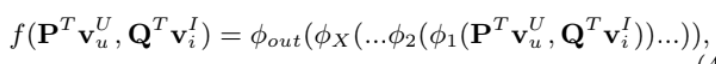
Embedding: input층 위에 embedding layer는, sparse representation을 dense하게 나타내는 fully connected layer이다.

Neural CF layers: 각 embedding들은 multi-layer neural architecture에 들어가는데, 이 때문에 NCF라 불린다. 각 층은 user-item간의 관계의 latent structure를 찾는다. 마지막 층인 X는 모델의 성능을 결정하는데, y의 추정치를 내뱉는다. target값과 추정치 간의 최소/최대화 하는 방향으로 학습을 한다. (pointwise, pairwise loss function)

NCF를 식으로 표현하면 다음과 같고,



f함수는 multi-layer neural network이므로 다음과 같이 표현할 수 있다.



NCF가 예측하는 값은 확률이기에 0~1사이 값이고, target variable 값도 0또는 1밖에 없는 binary이기에, 베르누이 분포를 이용하면서 likelihood function을 정의해야 한다.

Text

Description automatically generated with medium confidence

따라서 loss function은 언급한 likelihood function을 최소화 하는 파라미터를 찾는 것이다.

Text, letter

Description automatically generated

* + **Generalized Matrix Factorization (GMF)**

여기서 MF가 NCF의 특별한 케이스인 것을 보여준다. Input층이 embedding 층으로 가서 dense data로 변환된 것은, latent vector로 볼 수 있다.

Text

Description automatically generated with medium confidence

여기서 활성함수 ‘a’를 identity function으로 보고, ‘h’를 1의 uniform벡터로 보면 MF모델이랑 같음을 알 수 있다.

실제로는 활성함수 ‘a’를 sigmoid function으로 보고, ‘h’를 log loss로부터 학습되는 값으로 보면 GMF가 된다.

* + **Multi-Layer Perceptron (MLP)**

단순히 user와 item의 latent vector를 합쳐서 모델링하면 collaborative filtering에 충분하지 못한다. 그래서 고정된 element-wise product를 사용하는 GMF와 달리 hidden layers를 추가한 MLP를 활용하면 latent vector (user,item)간의 상호작용을 더 유연하고 비선형적으로 학습할 수 있다.

Text, letter

Description automatically generated

실험적으로 봤을 때, 여기서 사용되는 활성함수는 RELU가 사용된다.

* + **Fusion of GMF and MLP**

Linear kernel를 적용하는 GMF와 Non-linear kernel를 적용하는 MLP를 다뤘다. 여기서는 어떻게 GMF와 MLP가 합쳐지는지 설명할 것 이다.

Diagram

Description automatically generated

GMF와 MLP는 같은 embedding층을 공유하므로, 각 함수의 결과값을 합치면 된다. 같은 embedding층을 공유하는 GMF의 1층과 MLP의 마지막층을 결합한 것은 이런 수식으로 나타낼 수 있다.

A picture containing diagram

Description automatically generated

Fused model에 더욱더 유연함을 주기 위해서 (무조건 같은 크기의 embedding을 사용할 필요 없다), 서로 다른 크기의 embedding에서 학습을 하고 마지막 서로 마지막 층에서 합치면, 이런 수식으로도 표현이 된다.

A picture containing text

Description automatically generated

이는 non-convexity 목적 함수이기 때문에, gradient기반 최적화는 local optimal에 빠지는 문제점이 있다. 그래서 NeuMF를 시작할 때, GMF와 MLP의 pretrained 모델을 활용하는 것을 제안한다. Initialization을 랜덤하게 수렴할 때까지 GMF와 MLP를 각각 훈련시켜 놓고, 그 모델의 parameters들을 NeuMF의 parameters에 넣어서 훈련을 시작한다.

* **Conclusion and future work**

Neural network 기반인 collaborative filtering에 유용한 NCF를 제안했다.

1. GMF: MF의 일반형
2. MLP: DNN을 활용한 구조
3. NeuMF: GMF와 MLP의 앙상블

나아가서는, NCF 모델의 pairwise learners도 연구할 것이고, 다른 부가 정보(리뷰, 지식, 시그날)등 을 활용해서 NCF모델에 적용할 것이다. 지금까지 추천 시스템은 개인에게 초점이 맞춰졌지만, 앞으로는 그룹단위로도 연구를 해볼 것이다.