Recommender Based On ViT using side information

장영수 지윤정
조기흠 백찬진

목차

#1 논문 제목

부가정보를 활용한 ViT기반 추천시스템

(Recommender Based On ViT using side information)

#2 연구의 필요성, 목적 및 의의

기존 ONCF의 장점

- Outer Product를 활용하여 MF의 Inner Product와 Element-wise Product를 사용함으로써 얻을 수 있는 장점을 모두 가짐
- MF, NCF에서 다루지 않던 다른 임베딩 차원 사이의 상관관계를 포함함
- Interaction Map의 Correlation을 CNN으로 모델링 함

연구의 필요성, 목적 및 의의

- 기존의 ONCF는 User와 Item의 Rating데이터만을 사용하기 때문에 Side Information은 사용하지 못하였으나,

  본 제안 방법론은 기존의 ONCF에 Side Information을 추가함으로 데이터를 더욱 풍부하게 만들어 줌
- 현재 Vision Task 분야에서 SOTA를 달성하고 있는 Vision Transformer (ViT)를 사용함으로

  정확도가 더욱 높은 모델을 기대해 볼 수 있음

#3 이론적 배경

#1. Side Information

Cold start 문제의 완화

- 데이터가 매우 적거나 존재하지 않는 사용자, 아이템에 관해 올바른 추천을 제공하기 어려운 경우 Side Information이 보완책이 될 수 있음

예측 결과에 도움

- Implicit Data인 Side information을 사용하여 결과를 예측하는데 보완할 수 있음

Recommender system with Side Information

- A Survey on Accuracy-oriented Neural Recommendation: From Collaborative Filtering to Information-rich Recommendation
- Factorization Machines, Deep FM
- Wide & Deep
- Latent Cross: Making Use of Context in Recurrent Recommender Systems

#2. ViT

보다 나은 예측 결과

- Transformer는 NLP를 넘어 Vision Task 분야에서도 SOTA를 달성하고 있기 때문에 해당 연구에서도 좋은 결과를 보일 것으로 예상함

연산량 감소

- 데이터의 양이 충분한 상태라면, Vit는 CNN 에 비해 연산량이 줄어든 상태로 더 높은 성능을 보여준다.

About ViT

- An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale
- Do Vision Transformers See Like Convolutional Neural Networks?
- Attention Is All You Need

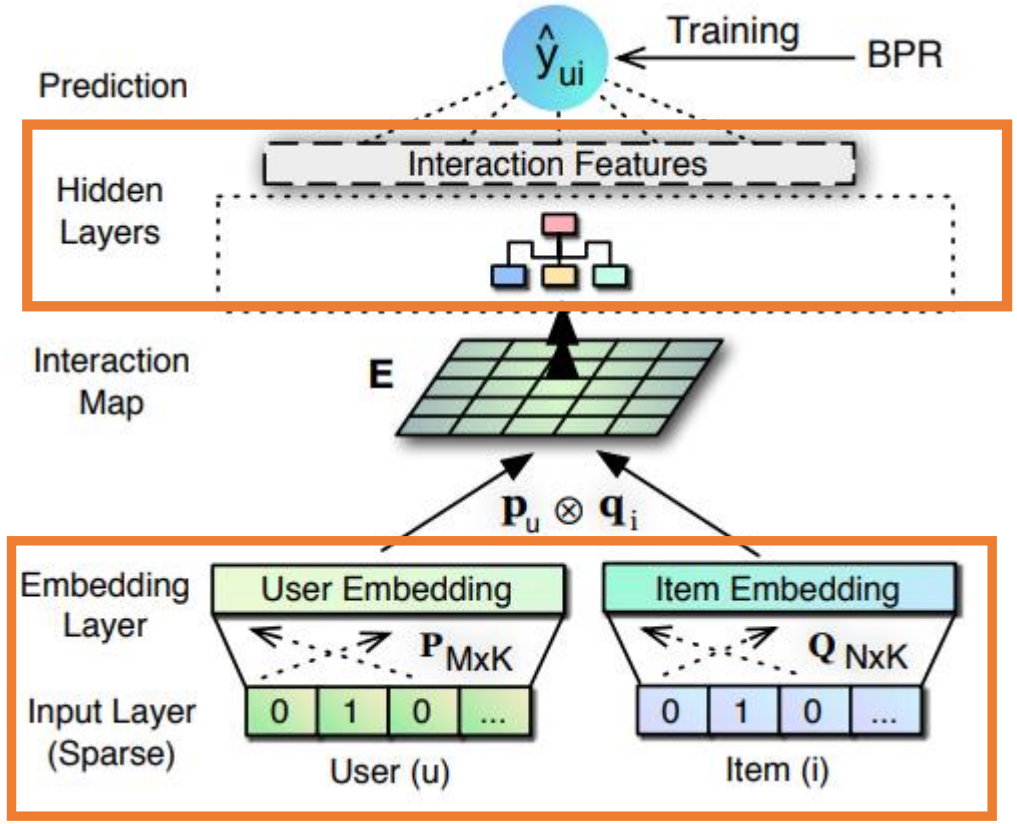
#4 연구 방법

Side Information과 ViT를 기반으로 한 여러가지 **방법론을 제시**한 이후

실험을 통해 각 모델과 기존 ONCF모델의 **성능을 비교**하고,

비교우위에 의해 제안한 방법론이 합리적이라는 것을 증명

#4 연구 방법, Method 1 (기존 ONCF변경)



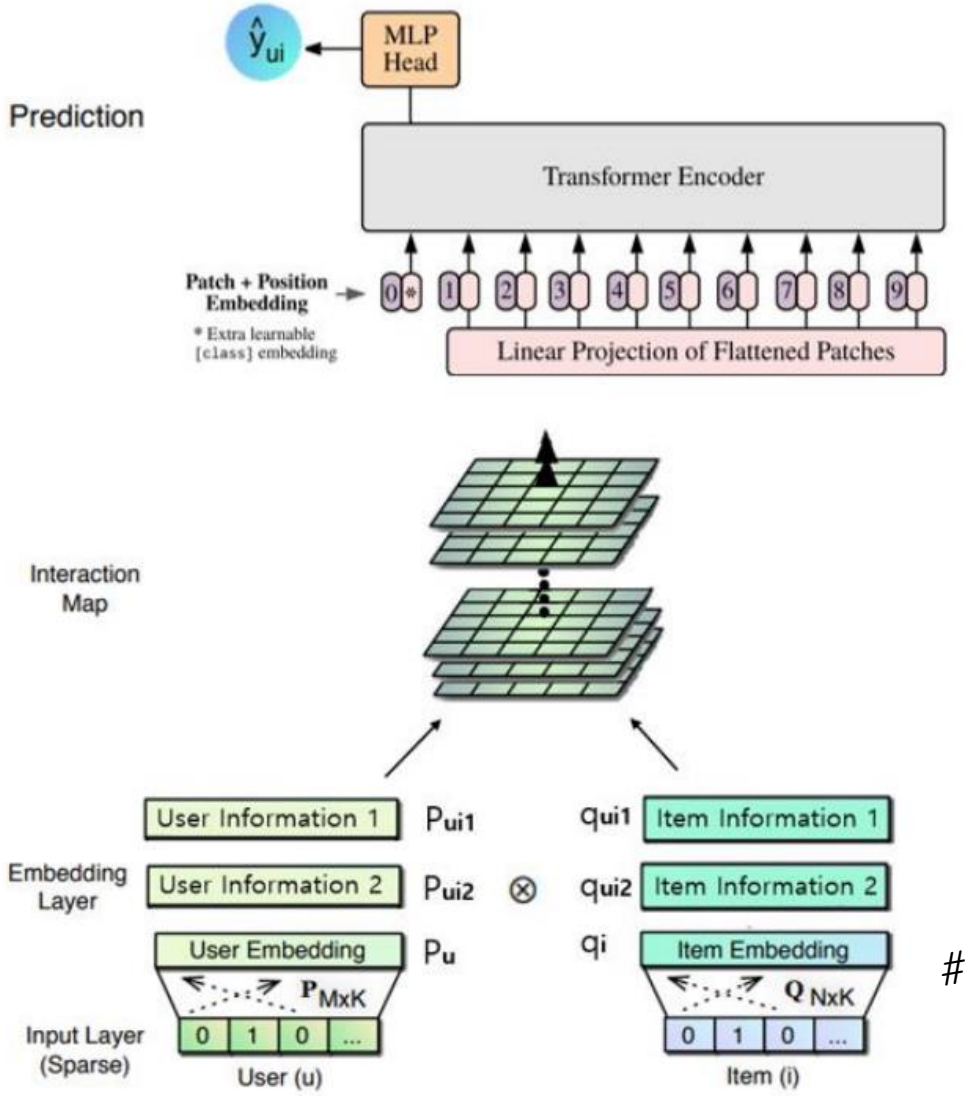
Hidden Layers

Vision Transformer (ViT)를 사용하여 CNN을 대체

Input Layer

- side information으로 데이터를 풍부하게 해줌

#4 연구 방법, Method 1



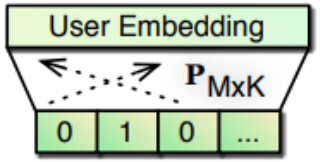
Interaction Map들을 여러 채널의 하나의 이미지로 간주하여
ViT 모델의 Input으로 사용

해당 Embedding Vector들을 서로 Outer-Product하여
여러 채널의 Interaction Map 생성

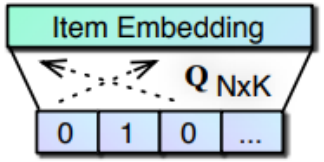
각각의 Side Information들을 User와 Item으로 나누어 각각
User Latent Embedding, Item Latent Embedding 과 같은 크기로 Embedding

#4 연구 방법, Method 1 (Input & Embedding)

N: user의 수 M: item의 수 k: 잠재요인 일때,



N * k



M * k

User Information 1

N * **k**

Item Information 1

M * **k**

User Information 2

N * **k**

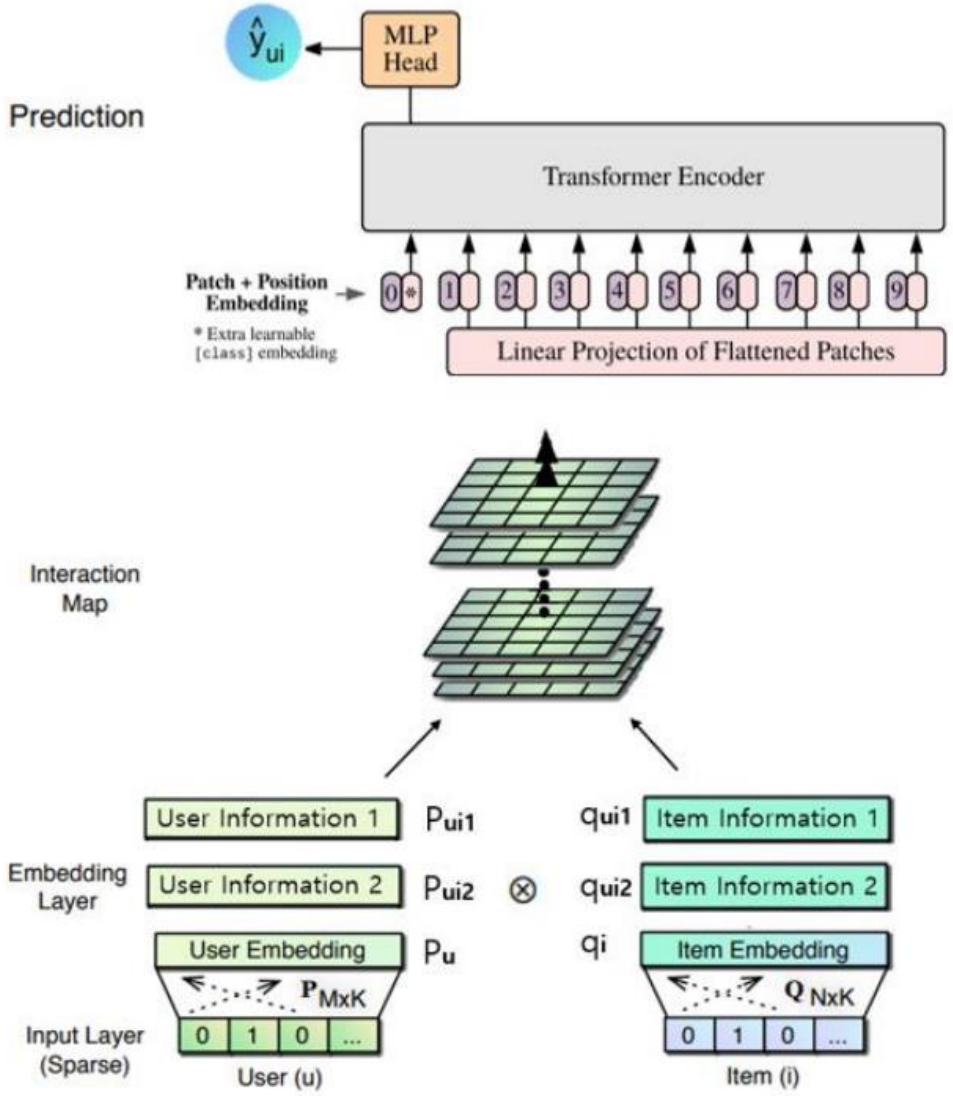
Item Information 2

M * **k**

User Side Information의 수: a

Item Side Information의 수: b 일때, a*b개의 채널

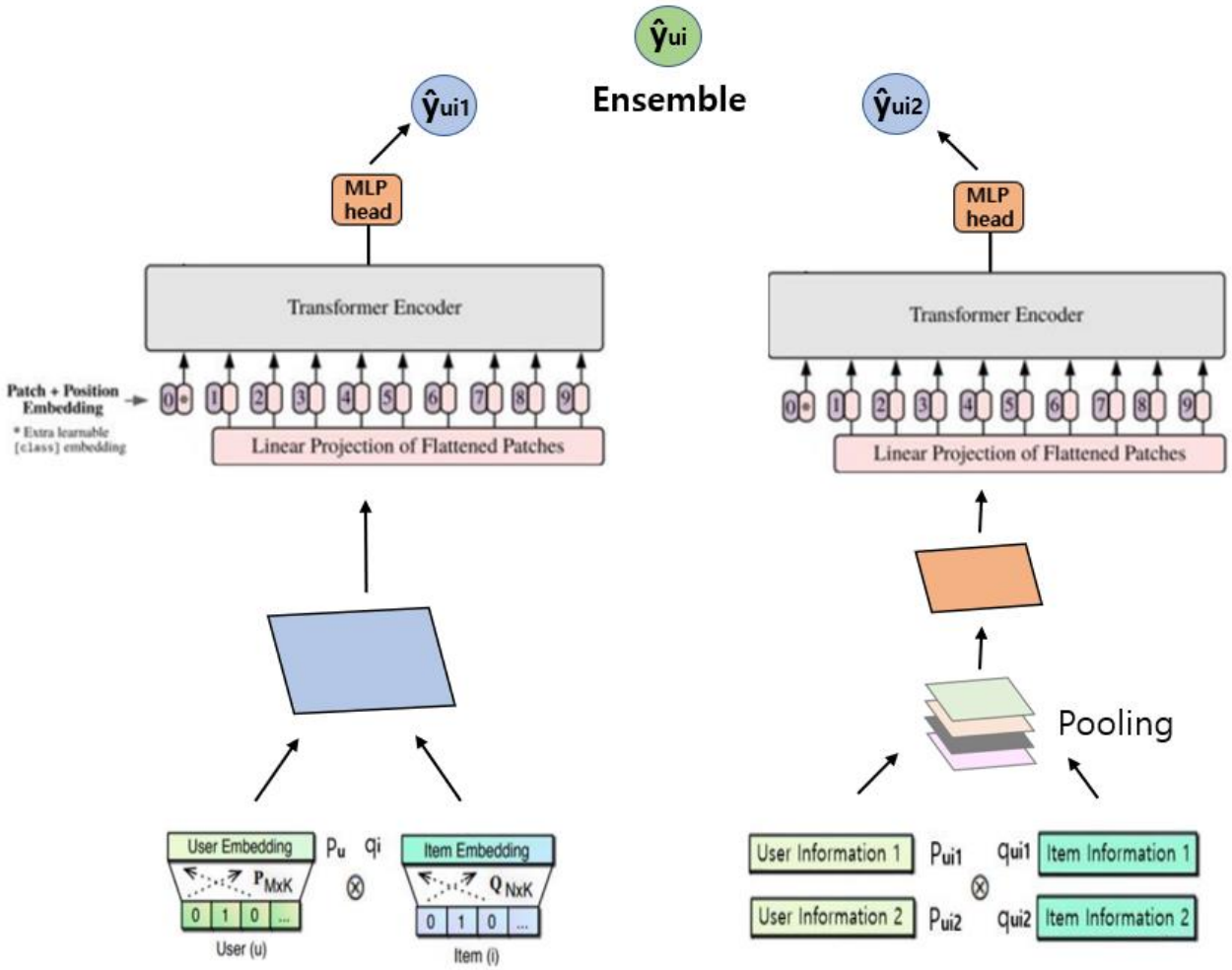
#5 예상 결과, Method 1



CNN보다 성능이 좋은 모델인 ViT를 사용함으로 **더 좋은 결과**를 기대할 수 있음

Side Information을 활용한 User와 Item 간의 관계를 채널로 활용하여 다양한 관계를 고려할 수 있을 것으로 기대

#4 연구 방법, Method 2

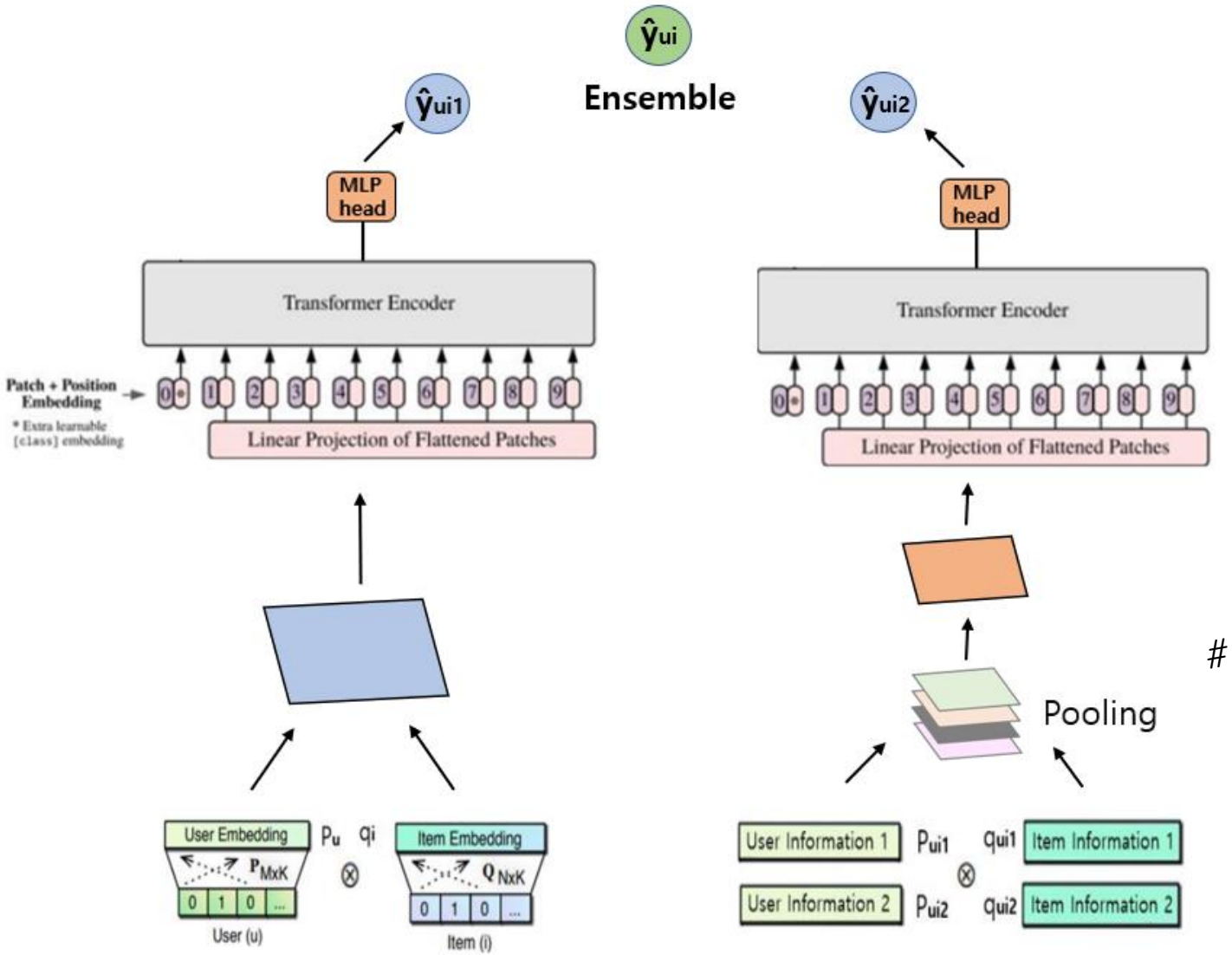


평점 데이터를 통해 도출된 데이터와 Side Information을 분리하여 각각 ViT에 넣은 후 앙상블함

다양한 Side Information에 대한 정보를 채널 간 Average Pooling을 통하여 Side Information을 하나로 고려하는 Interaction Map이 생성

Side Information의 정보를 기존 Interaction Map을 통한 결과에 부가적인 정보로 활용

#5 예상 결과, Method 2

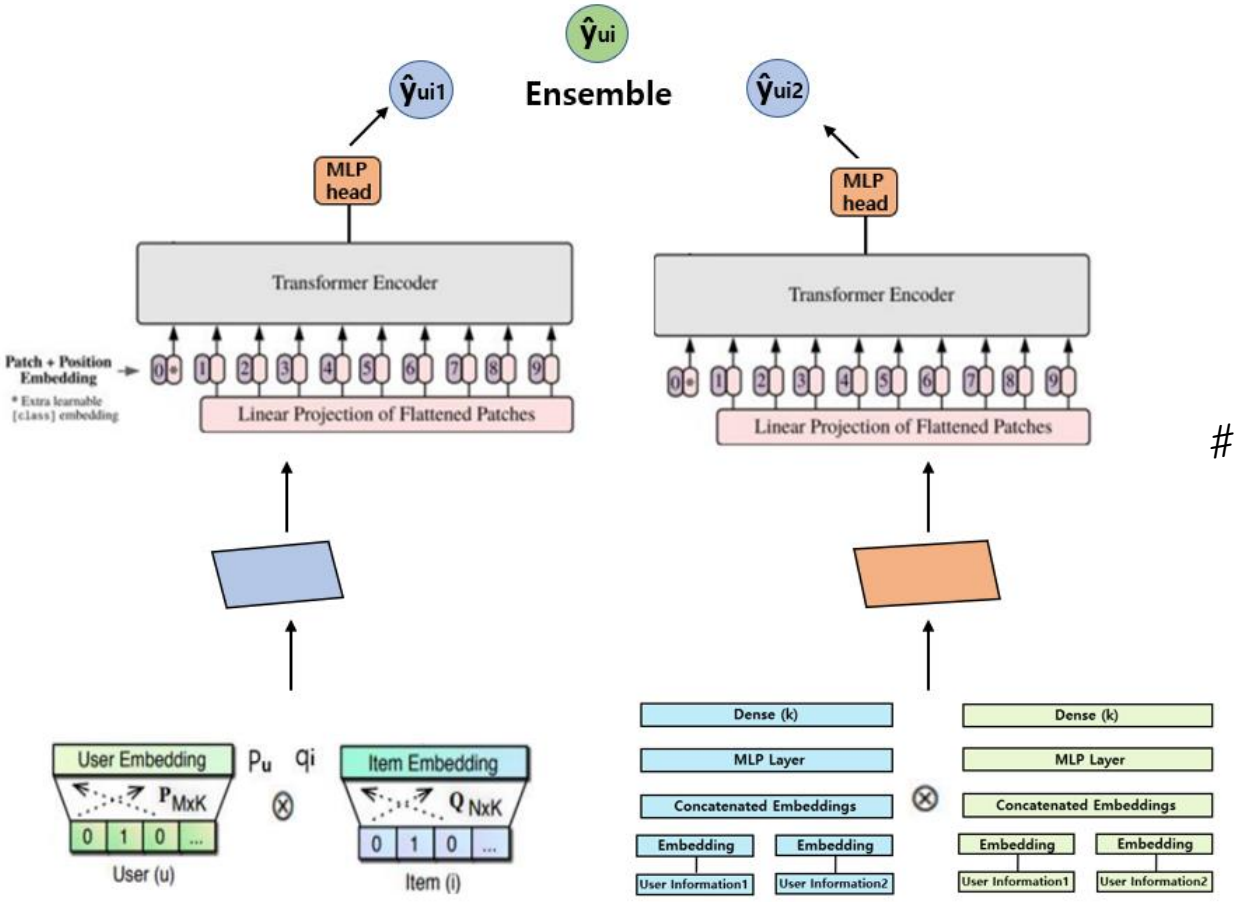


기존 ONCF Interaction Map에 부가적인 정보로 활용하여 결합하는 방식으로 진행

기존의 정보에 Side Information을 참고하는 형태로 예측에 활용한다면, 기존 예측에 도움이 될 것으로 기대

Side Information에 대한 정보를 Average Pooling을 통하여 하나의 Image Map으로 표현하기 때문에, 다양한 정보를 고려한 학습이 가능

#4 연구 방법, Method 3

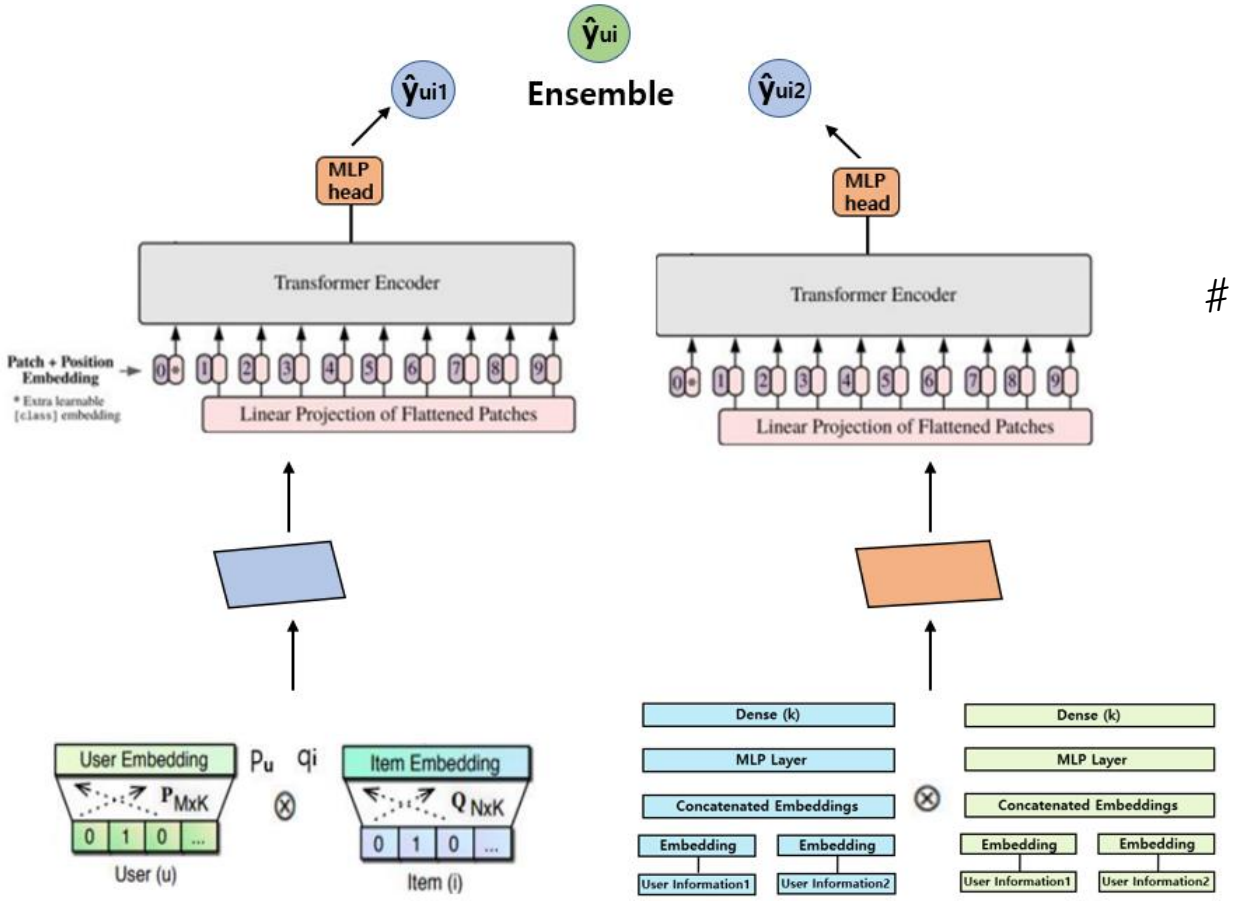


Method2와 같은 방식으로 앙상블 진행

User와 Item 각각을 MLP Layer에 통과시켜 원하는 Embedding size인 Dense(k)로 출력된 결과를 이용하여 Outer-Product 진행

User와 Item의 SIde Information을 각각 하나로 통합하기 위하여 임베딩을 거친 뒤에 Concat

#5 예상 결과, Method 3



Side Information들을 User, Item 각각 하나의 Embedding Vector로 표현하기 때문에 기존 ONCF의 이점을 방해하지 않으면서 Side Information을 추가할 수 있을 것으로 기대함

#6 참고 문헌

- Neural Collaborative Filtering (NCF)
- Outer Product-based Neural Collaborative Filtering (ONCF)
- Wide & Deep Learning for Recommender Systems
- Attention Is All You Need (Transformer)
- An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (Vision Transformer)
- A Deep Learning Based Recommender System Using Visual Information

Question?