



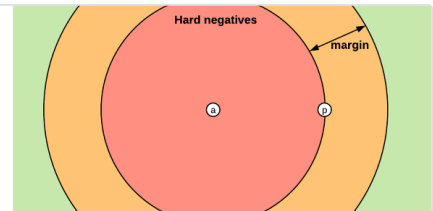
Learning To Rank(Point, Pair, List + Wise)

랭킹 하는 법을 학습한다 ! (ML + IR)

[개념정리] Ranking Loss Functions & Metric Learning

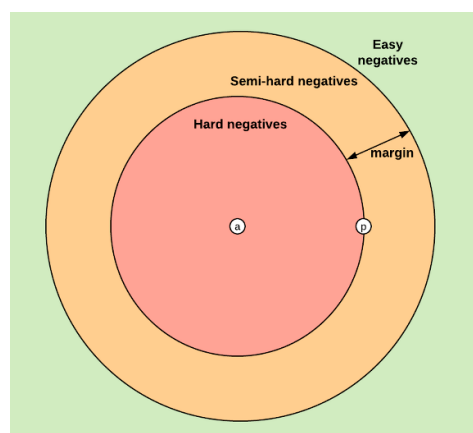
direct하게 label, value, set, input으로 부터 주어진 value를 예상하도록 학습하는 것이 목표인 Cross-Entropy Loss나 Mean Square Error Loss와 같은 다른 loss function들과 달리 Ranking losses의 목표는 inputs 사이의 상대적인 거리를 예측하는 것이다. 이 작

📄 <https://hyeonni.tistory.com/277>




Ranking Losses의 목표는 입력 값들 사이에 상대적인 거리를 예측하는 것이다.

여기서는 Triplet을 다루는데 샘플을 기준으로 Positive Sample과 Negative Sample 사이의 거리를 최적화 한다.



Learning to Rank와 nDCG

최근 검색 모델에도 관심을 가지게 되면서, 랭킹모델과 IR에 대해 다시 공부중이다. 기존에 관심있었던 커머스 분야의 추천 랭킹 시스템에서는, '유저가 실제로 살 것 같은 하나의 아이템'을 잘 추천해주는 것이 중요했기 때문에, MRR이나 nDCG 같은 순위 기준의 랭킹 평가

 <https://yamalab.tistory.com/119>

$$CG_p = \sum_{i=1}^p rel_i$$

가장 널리 알려진 랭킹 평가 방법인 nDCG에 대해 소개한다.

1. Point-Wise Approaches

- 하나의 아이템 단위의 접근. 질의에 대해 가장 연관성이 높은 (확률이 높은) 아이템을 예측하는 것. 예측 결과를 정렬 해서 보여주는 것이 최종 결과의 형태. 하나의 predictor가 Ir을 아이템 단위로 적용 후 결과 정렬
- Pointwise Approaches는 Loss Function에서 한번에 한 문서를 본다. 이 접근법 들은 필수적으로 한 문서를 가져서 classifier/regresser를 학습해서 현재 Query에 관련성이 있는 것을 예측한다. 최종 랭킹은 달성된다. 간단하게 결과 리스트를 정렬해서 문서 점수들을 기반 해서. Pointwise Approaches를 위해, 각각 문서를 위한 점수는 독립적이다. Query 내의 다른 문서 결과와

2. Pair-wise Approaches

- point의 label 혹은 score에 의존하는 Point-Wise에 비교하여 ground-truth에 가까운 order를 찾는 과정. 순서가 뒤바뀌는 것을 cost function에 활용하여, 잘못된 순서를 바로 잡는다
- Pairwise approaches는 Loss Function에서 한번에 한 쌍의 문서를 본다. 주어진 한 쌍의 문서에서 그 쌍의 최적의 순서를 찾고 결과를 ground truth와 비교하려고 시도한다. 랭킹을 매기는 도구의 목적은 랭킹이 역전 되는 횟수를 최소화 하는 것이다. 랭킹의 역전은 Ground Truth와 비교해서 순서가 잘못되는 경우다

Pairwise Approaches는 실제에서 Pointwise Approaches 보다 더 잘 동작한다. 그 이유는 상대적인 순서를 예측 하는 방식이 관련성에 대한 스코어 나 클래스 라벨을 예측하는 것에 비해 랭킹의 특성과 가깝기 때문이다. 몇 가지 가장 유명한 Learning to Rank 알고리즘은 RankNet, LambdaRank, LambdaMart 는 Pairwise Approaches다

3. List-Wise Approaches


- 모든 반환된 결과를 Ground Truth와 한방에 비교하는 방법. 타입(시나리오) 별로 나누어 랭크를 학습하는 것.
- Listwise Approaches는 직접적으로 전체 문서 리스트를 보고 최적의 순서를 찾으려고 시도한다. 두 가지 메인 sub-technique 들이 있다.
 1. 검색 방법에 대한 직접적인 최적화 방식. 예를 들어 NDCG, SoftRank, AdaRank
 2. 달성 하고자 하는 랭킹 종류의 unique한 특성들에 대한 이해를 기반으로 정의된 loss function 을 최소화하는 방법. ListNet, ListMLE
- Listwise Approaches는 Pointwise 나 Pairwise 방식에 비해 꽤나 복잡해 질 수 있다

4. nDCG(normalizing Discounted Cumulative Gain)

특정 아이템에 skew 되어져 있는 rank. 각각의 순위에 가중치를 준다. 하나의 랭킹으로 여러 명을 평가 하기 위해 밀도를 추가하여, 2차원 점수를 평가한다는 의미

tensorflow/ranking

You can't perform that action at this time. You signed in with another tab or window. You signed out in another tab or window. Reload to refresh your session. Reload to refresh your session.

 https://github.com/tensorflow/ranking/tree/master/tensorflow_ranking

