

Evaluating Recommender Systems: Choosing the best one for your business

<https://medium.com/recombee-blog/evaluating-recommender-systems-choosing-the-best-one-for-your-business-c688ab781a35>

**** 학계에서의 오프라인 평가: 낮은 예측 오류(RMSE/MAE) searching, 높은 Recall, 높은 항목 coverage ****

RMSE

1. 다양한 상품에 대한 과거로부터의 평점들을 모은다.
2. split validation을 이용하여 train set, test set를 나눈다.
3. 예측을 실시한다. 실제 값(4)이랑 예측 값(3.5)을 비교한다, 오류 값 - 0.5
4. 오류 값들의 평균을 계산한 것이 RMSE

Implicit rating(unary ratings)

- 페이지 방문 기록, 구매기록 페이스북 좋아요 등등
- 평점에 대한 오류 값을 예측할 수 없다.
- 다른 방법으로 예측을 시행한다. 구매 데이터베이스에 있는 80%를 추천시스템에 적용하고 나머지 20%는 추천시스템에 의해서 예측을 한다.
- 구매를 예측한 건수에 따라서 정확도(Recall)가 나오게 된다.(연관성이 있는 상품들끼리는 좋은 예측을 한다.)

Catalog coverage

- reasonable한 recall을 위해선 일반적으로 가장 유명한 제품을 추천한다.
- 새로운 item을 추천해주기 위해서 catalog coverage를 계산해준다.

**** 비즈니스세계에서의 온라인 평가 : 높은 CLV(Customer Lifetime Values) searching,
CTR(클릭율), CR(구매전환율), ROI(투자자본수익률), QA(품질보증)****

오프라인 평가에서는 일반적으로 split-validation을 사용하지만, 온라인 평가에서는 A/B 시험 또는 다변량 시험이 가장 두드러진 접근방식이다.

1. A와 B로 그룹을 나누고 user id cookie hashing한다.
2. 실험을 하기 위해서 추천 test를 하기 좋은 website나application의 좋은 장소를 결정한다.
3. Small traffic이면 결과가 충분히 수집될 수 있는 곳에 장소를 결정하고 반대인 경우 결과가 과도하게 수집되지 않게 결정한다. 그렇지 않으면 이상한 추천을 할 가능성이 높다.

주요 측정변수로는 CTR(Click-Through Rate) , CR(Conversion Rate)가 존재한다.

CTR = 3의 클릭/20번의 추천 15%

Conversion Rate = 사용자를 만족시킨 권장사항 수

CTR과 CR은 추천을 확인하기에 좋은 estimate이다. 그러나 도메인에 따라서 조심해야 한다. 뉴스 기사를 볼 때는 높은 클릭 수를 보이지 않을 때가 있다. 그러나 클릭 수를 높이기 위해서 자극적인 제목의 기사를 올리게 되면 장기적으로 보았을 때 나쁜 이미지를 유지할 수 있다.

Empiric evaluation(경험적 평가)

- 사람의 직감으로 어떤 추천이 좋고 나쁜지를 판단한다. 하지만 이것을 수치화 하지 못한다.

ROI(Return of Investment) 투자대비수익률

- 추천아이템을 성공적으로 추천해주었을 때 현재 대비 얼마나 수익률이 올랐는지 확인

CLV(Customer Lifetime Value)

- 이용자들이 좋아하는 서비스를 제공하여 comfortable하게 느끼도록 한다.