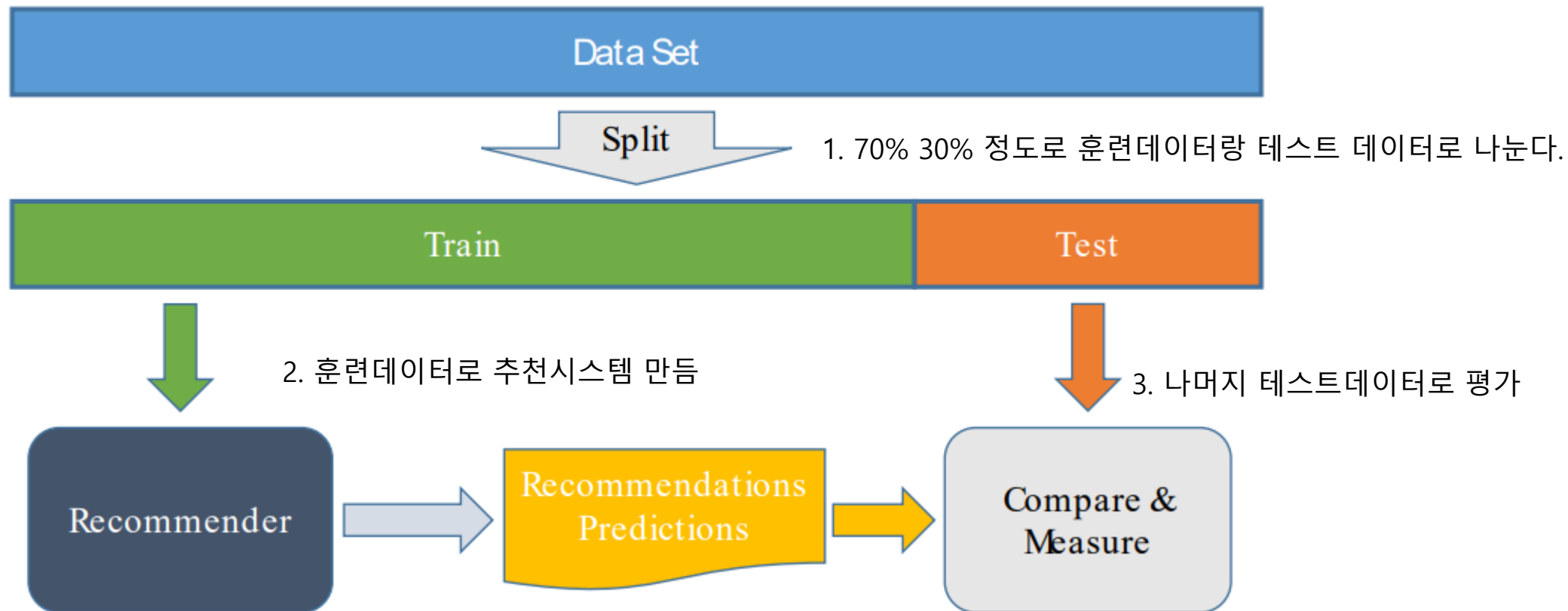


Data

- Ratings of movies
- Plays of songs
- Clicks of news articles

어떤 식으로 학습을 할까요?



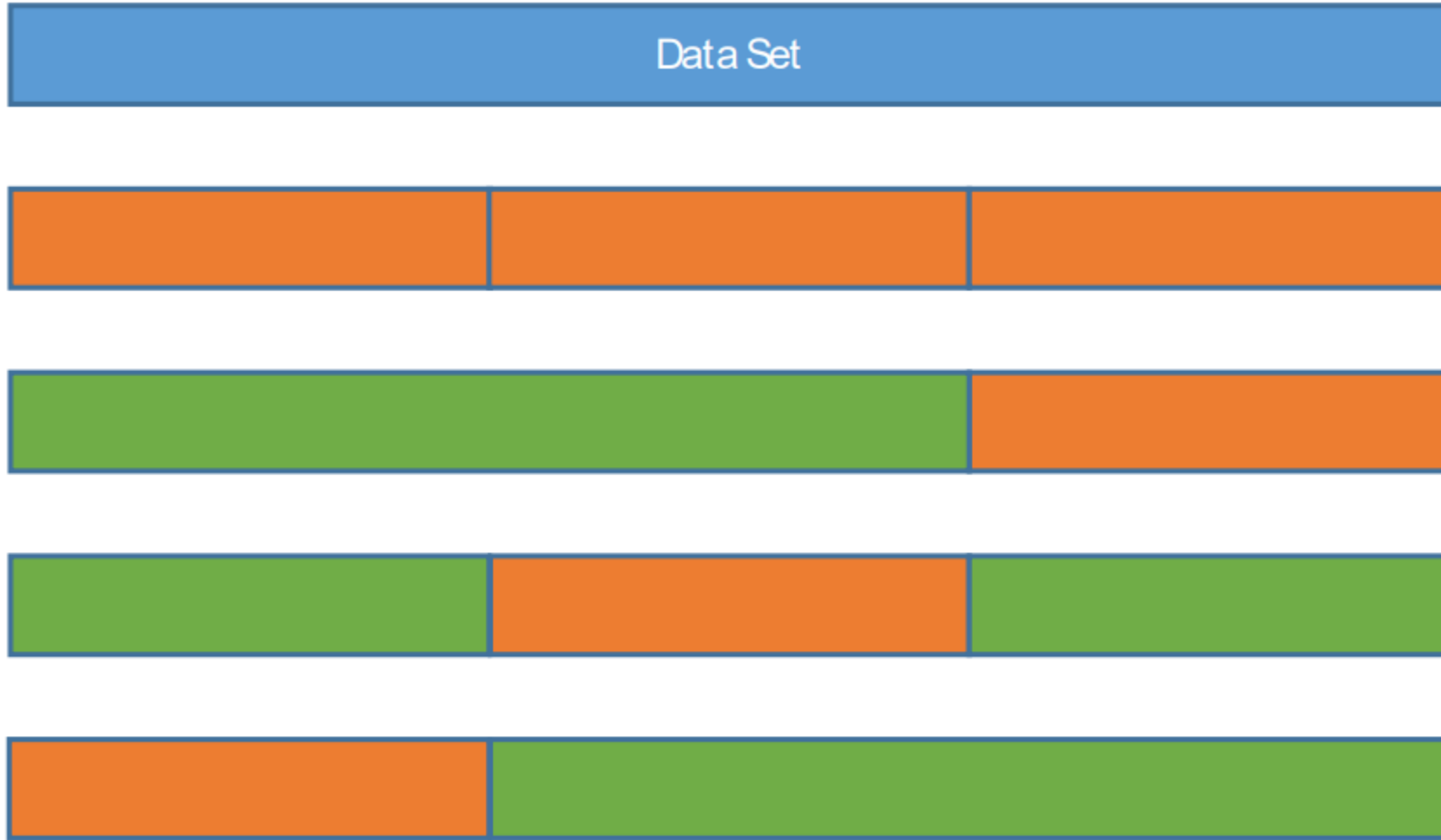
Cross - Validation

만약에 우연히라도 데이터가 너무 편향적으로 나누어진다면??



여러 번 나눈다!

Cross validation?



예를 들면 세 번에 걸쳐서 학습

단점

- 정확성이 추천시스템에서 중요한가?
즉, 사용자가 기존에 구매했던 물품을 다시 추천해주는게 의미가 있나?

Prediction Accuracy Metrics

- MAE
- MSE
- RMSE

MAE

[Prediction: P , ground_truth: R]

$$\begin{aligned}\text{MAE} &= E[|P - R|] \\ &= \frac{\sum ratings |P - R|}{\# ratings}\end{aligned}$$

단점

100등 - 95등 차이와 6등 - 1등 차이를 같다고 계산한다.
실제로 6등 - 1등 차이가 더 크다!

MSE

[Prediction: P , ground_truth: R]

$$\begin{aligned}\text{MSE} &= E[(P - R)^2] \\ &= \frac{\sum_{\text{ratings}} (P - R)^2}{\# \text{ ratings}}\end{aligned}$$

단점

스케일을 직관적으로 이해하기가 어렵다

RMSE

[Prediction: P , ground_truth: R]

$$\begin{aligned}\text{RMSE} &= \sqrt{E[(P - R)^2]} \\ &= \sqrt{\frac{\sum_{\text{ratings}} (P - R)^2}{\# \text{ ratings}}}\end{aligned}$$

잠깐 고민해볼 것

Rating에 중심을 뒤탈까? Vs. 유저당 rating에 중심을 뒤탈까?

잠깐 고민해볼 것

Rating에 중심을 뒤탈까? Vs. 유저당 rating에 중심을 뒤탈까?

왜 중요한데?

A 유저가 평가를 3000개

B 유저가 평가를 10개

이 상황에서 Rating에 중심을 두는게 맞을까?

잠깐 고민해볼 것

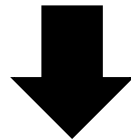
Rating에 중심을 뒤탈까? Vs. 유저당 rating에 중심을 뒤탈까?

왜 중요한데?

A 유저가 평가를 3000개

B 유저가 평가를 10개

이 상황에서 Rating에 중심을 두는게 맞을까?



상황에 따라 다르다

Comparing Different Algorithms

- What to do when computing MAE in different cases:
 - Remember, must be same data set/scale
 - If coverage is different (different set of user/item pairs for which predictions are available, two choices):
 - Check against common subset
 - Supplement algorithm with default for full coverage

Decision – Support Metrics

좋은 건 좋다고 싫은 건 싫다고 말하는 게 좋아

예를 들면 (영화 평점 5점 만점)

좋은 영화를 4.5점 줘서 나는 본다 -> 꿀잼!

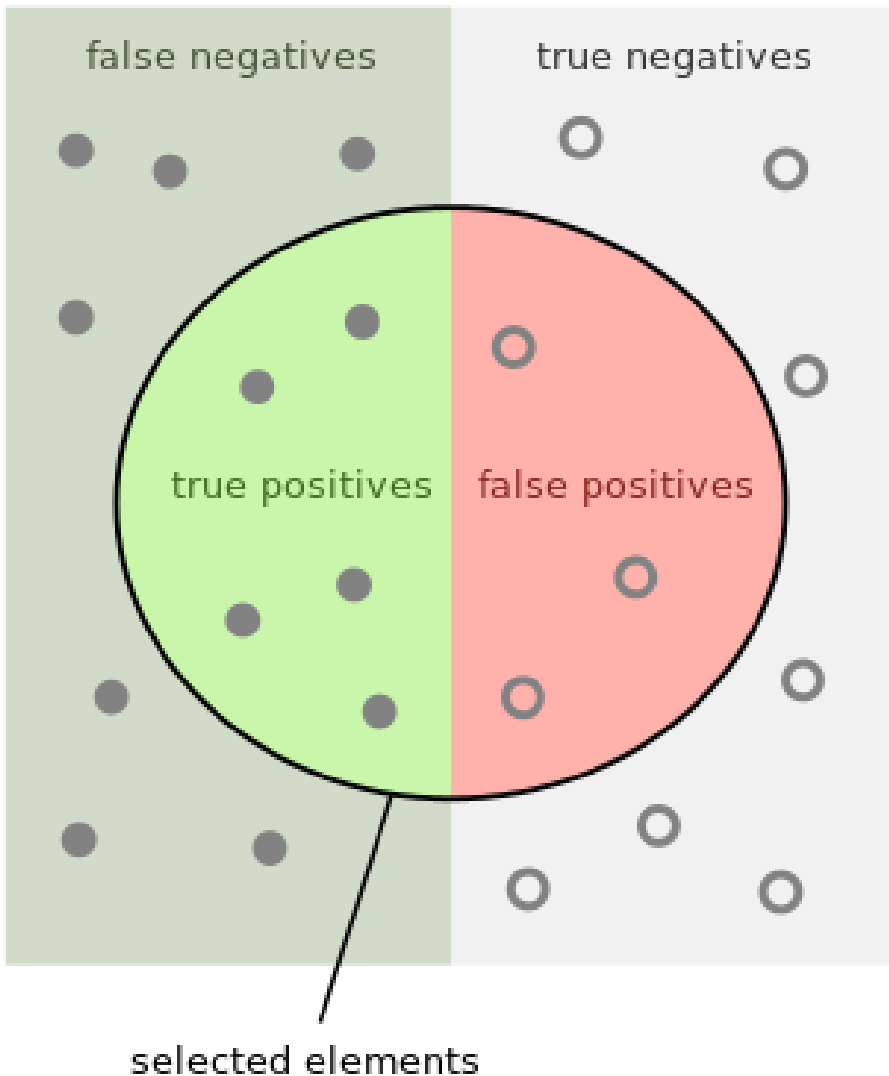
노잼 영화를 1점 줘서 보지 않는다 -> 시간은 금!

하지만

좋은 영화를 2점을 줘서 나는 보지 않는다 -> 좋은 영화를 볼 기회를 놓침

노잼 영화를 4점을 줘서 나는 본다 -> 시간 낭비

relevant elements



Precision and Recall

정밀도와 재현율

- Precision = $\frac{N_{rs}}{N_s}$

Precision = $\frac{\text{green semi-circle}}{\text{green semi-circle} + \text{red semi-circle}}$

- Recall = $\frac{N_{rs}}{N_r}$

Recall = $\frac{\text{green semi-circle}}{\text{green semi-circle} + \text{green rectangle}}$

- $F_1 = \frac{2PR}{P + R}$

Precision and Recall

문제점

1. ground truth 가 필요하다.
 - 추천시스템을 만드는데 ground-truth는 있을 수 가 없다.
(ground truth가 있으면 추천시스템을 만들 필요가 없다)
2. 상위 추천데이터가 아니라 전체 데이터를 다뤄야한다.

Precision and Recall

해결법

1.
 - Fake precision/recall by limiting to rated items
 - Common – results in interesting biases
 - Human-rating experiments that compute precision/recall over some random subset
2.
 - Precision@n is the percentage of the top-n items that are “good”: $P@n = \frac{N_{r@n}}{n}$
 - Some have proposed computing this as an average over a set of experiments with 1 “hit” and a large number of presumed misses
 - Recall@n is effectively the same

Rank-Aware Top-N Metrics

- Mean Reciprocal Rank (MRR)
- Mean Average Precision (MAP)
- Discounted Cumulative Gain (DCG)

MRR도 사용하지만 nDCG랑 MAP를 주로 사용한다

Mean Reciprocal Rank (MRR)

$$MRR(O, U) = \frac{1}{|U|} \sum_{u \in U} \frac{1}{k_u}$$

Query	Proposed Results	Correct response	Rank	Reciprocal rank
cat	catten, cati, cats	cats	3	1/3
tori	torii, tori , toruses	tori	2	1/2
virus	viruses , virii, viri	viruses	1	1

$$MRR = \frac{\left(\frac{1}{3} + \frac{1}{2} + 1\right)}{3} = \frac{11}{18} = 0.61$$

Mean Reciprocal Rank (MRR)

$$MRR(O, U) = \frac{1}{|U|} \sum_{u \in U} \frac{1}{k_u}$$

Query	Proposed Results	Correct response	Rank	Reciprocal rank
cat	catten, cati, cats	cats	3	1/3
tori	torii, tori , toruses	tori	2	1/2
virus	viruses , virii, viri	viruses	1	1

$$MRR = \frac{\left(\frac{1}{3} + \frac{1}{2} + 1\right)}{3} = \frac{11}{18} = 0.61$$

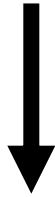
장점

- 간단
- 사용자가 원하는 것을 정확히 타겟팅

단점

- 일반적인 추천에는 적합하지 않다.

Average Precision



Rank	Correct?	Precision
1	True	1.0
2	True	1.0
3	False	0.67
4	False	0.5
5	False	0.4
6	True	0.5
7	True	0.57
8	False	0.5
9	False	0.44
10	True	0.5

Mean Average Precision (MAP)

$$MAP(O, U) = \frac{1}{|U|} \sum_{u \in U} AP(O(u))$$

Rank	Correct?	Precision
1	True	1.0
2	True	1.0
3	False	0.67
4	False	0.5
5	False	0.4
6	True	0.5
7	True	0.57
8	False	0.5
9	False	0.44
10	True	0.5

$$MAP(O, U) = \frac{1}{10} (1 + 1 + 0.67 + 0.5 \dots)$$

Spearman correlation

$$\frac{\sum_i (k_o(i) - \overline{k_o})(k_{o_u}(i) - \overline{k_{o_u}})}{\sqrt{\sum_i (k_o(i) - \overline{k_o})^2} \sqrt{\sum_i (k_{o_u}(i) - \overline{k_{o_u}})^2}}$$

Spearman correlation

$$\frac{\sum_i (k_o(i) - \bar{k}_o)(k_{o_u}(i) - \bar{k}_{o_u})}{\sqrt{\sum_i (k_o(i) - \bar{k}_o)^2} \sqrt{\sum_i (k_{o_u}(i) - \bar{k}_{o_u})^2}}$$

문제점

모든 차이를 같다고 본다
(1이랑 3이랑 바꾸는건 치명적
11이랑 13이랑 바꾸는건 별로 안 치명적)



랭크가 낮을수록 비중을 뒤야함

Discounted Cumulative Gain (DCG)

$$DCG(O, u) = \sum_i \frac{r_{ui}}{disc(i)}$$

$$disc(i) = \begin{cases} 1, & i \leq 2 \\ \log_2 i, & i > 2 \end{cases}$$

$$nDCG(O, u) = \frac{DCG(O, u)}{DCG(O_u, u)}$$

Discounted Cumulative Gain (DCG)

A 4점
B 3점
C 0점
D 5점

$$DCG = \frac{4}{1} + \frac{3}{1} + \frac{0}{1.58} + \frac{5}{2} = 9.5$$

$$DCG_{perfect} = \frac{5}{1} + \frac{4}{1} + \frac{3}{1.58} + 0 = 10.9$$

$$nDCG = \frac{9.5}{10.9} = 0.872$$