Ranking in Recommenders

Authors: Felipe Perez, Ella Chen, Serena McDonnell, and Werner Chao

Nov 2019

aggregate intellect

What is ranking and why is it necessary?

In giving recommendation the user is presented with options.

Better recommendation should be at the top.

Evaluate the order (ranking) so the conditions above are met.

Helps keep the user engage.

Each list can be personalized.

How does it work? (pointwise)

1. A query is given: User u, search for info q.

$$u = (u_1, \ldots, u_n) \qquad q = (q_1, \ldots, q_l)$$

1. A score is computed for each item.

$$s_i = \text{score}(u, q, i)$$

1. Scores are used to rank.

$$\text{argsort}$$

How does it work? (pairwise)

1. A query is given: User u, search for info q.

$$u = (u_1, \ldots, u_n) \qquad q = (q_1, \ldots, q_l)$$

1. A score is computed for each pair of items.

$$s_{ij} = \text{score}(u, q, i, j)$$

1. Sort according to the rule

$$P(i < j) = s_{ij}$$

Evaluating Ranking: Mean Average Precision

Query: Round fruits

Most relevant

Less relevant

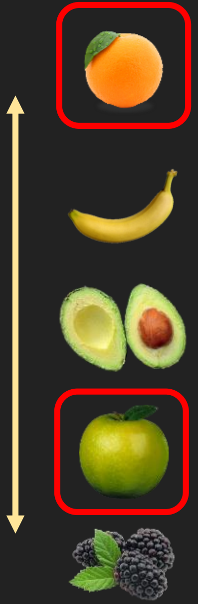
Precision at k

$$P(k) = \frac{|\{\text{relevant documents}\} \cap \{\text{top } k\}|}{k}$$

Evaluating Ranking: Mean Average Precision

Query: Round fruits

Most relevant



Less relevant

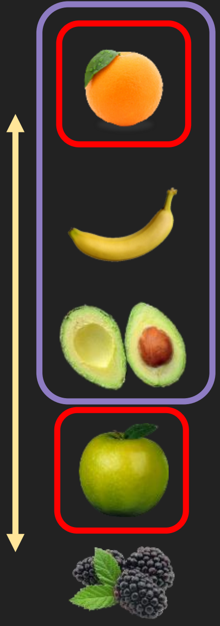
Precision at k

$$P(k) = \frac{|\{\text{relevant documents}\} \cap \{\text{top } k\}|}{k}$$

Evaluating Ranking: Mean Average Precision

Query: Round fruits

Most relevant



Less relevant

Precision at k

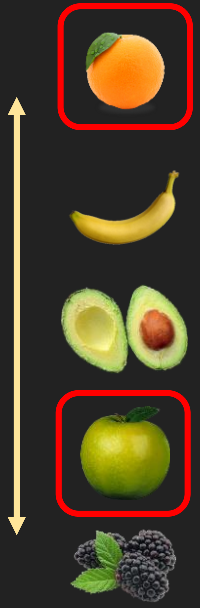
$$P(k) = \frac{|\{\text{relevant documents}\} \cap \{\text{top } k\}|}{k}$$

$$\boxed{P(3) = 1/3}$$

Evaluating Ranking: Mean Average Precision

Query: Round fruits

Most relevant



Less relevant

Precision at k

$$P(k) = \frac{|\{\text{relevant documents}\} \cap \{\text{top } k\}|}{k}$$

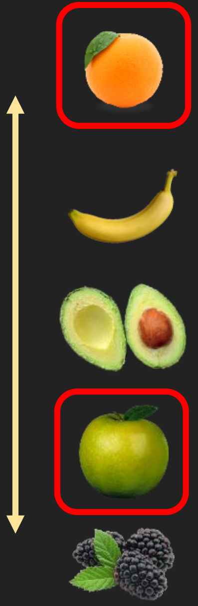
$$P(1) = 1$$
$$P(2) = 1/2$$
$$P(3) = 1/3$$
$$P(4) = 2/4 = 1/2$$
$$P(5) = 2/5$$

Evaluating Ranking: Mean Average Precision

Query: Round fruits

Most relevant

Less relevant

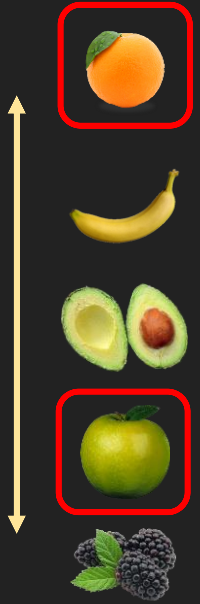


$$\mathrm{rel}(k) = \begin{cases} 1 & \text{if item } k \text{ is relevant} \\ 0 & \text{otherwise} \end{cases}$$

Evaluating Ranking: Mean Average Precision

Query: Round fruits

Most relevant



Less relevant

$$\mathrm{rel}(k) = \begin{cases} 1 & \text{if item } k \text{ is relevant} \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \mathrm{rel}(1) &= 1 \\ \mathrm{rel}(2) &= 0 \\ \mathrm{rel}(3) &= 0 \\ \mathrm{rel}(4) &= 1 \\ \mathrm{rel}(5) &= 0 \end{aligned}$$

Evaluating Ranking: Mean Average Precision

Query: Round fruits

Most relevant

Less relevant

$$\mathrm{AveP}(q) = \frac{\sum_{k=1}^{n} P(k) \times \mathrm{rel}(k)}{\text{number of relevant documents}}$$

$$= \frac{1 \cdot 1 + 1/2 \cdot 0 + 1/3 \cdot 0 + 1/2 \cdot 1 + 2/5 \cdot 0}{2}$$

$$= \frac{1.5}{2} = 0.75$$

Evaluating Ranking: Mean Average Precision

Query: Round fruits

Most relevant

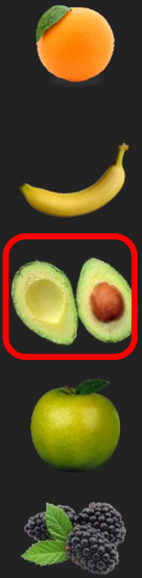
Less relevant

$$\mathrm{MAP} = \frac{\sum_q \mathrm{AveP}(q)}{\text{number of queries}}$$

Evaluating Ranking: Mean Reciprocal Rank

Query: Avocado

Most relevant



Less relevant

$$\text{Reciprocal rank} = \frac{1}{\text{rank of first relevant}}$$

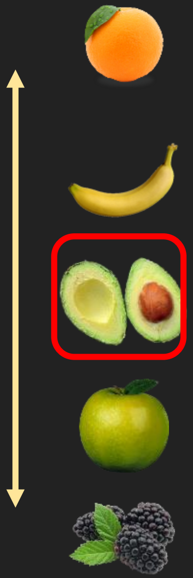
$$= \frac{1}{3}$$

Evaluating Ranking: Mean Reciprocal Rank

Query: Avocado

Most relevant

Less relevant



$$\mathrm{MRR} = \sum_{\text{queries } q} \frac{\text{reciprocal rank for } q}{\text{number of queries}}$$

Hands on exercise

1. Implement MAP
2. Implement MRR

Feature Generation

Why creating features?

Features allow certain models to “understand” the problem by creating a better representation of the elements involved.

Feature Generation

Why creating features?

Features allow certain models to “understand” the problem by creating a better representation of the elements involved.

Feature Generation

Why creating features?

Features allow certain models to “understand” the problem by creating a better representation of the elements involved.

Feature Generation

How to create features?

- Understand the problem.
- Understand the data distribution.
- Understand the features distribution.
- Understand how the model uses the features.

How to recommend clocks:

A clock store

- Customers enter and spent time in the store.
- You observe the customer actions.
- The customer may ask questions.
- You always collect data.

Local features

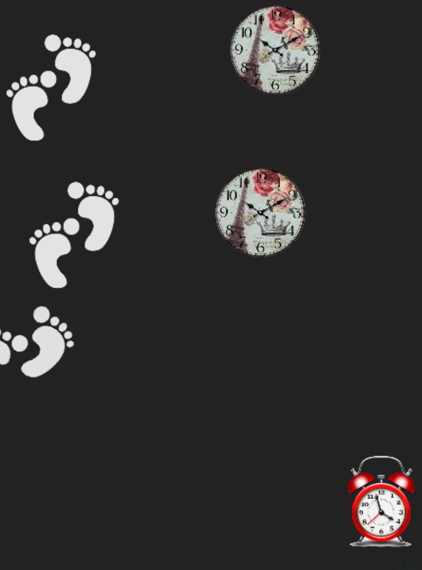
Associated to the current interaction

- How much time has the user spent in the store?
- How many clocks have they looked at?
- How many red clocks?
- Have they picked more than one?
- How many sections of the store has the customer visited?
- etc.

Local features

Associated to the current interaction

- How much time has the user spent in the store?
- How many clocks have they looked at?
- How many red clocks?
- Have they picked more than one?
- How many sections of the store has the customer visited?
- etc.



User features

Associated to the person who came in

- High
- Gender
- Economical status
- House size
- Job
- Etc

User features

Associated to the person who came in

- High
- Gender
- Economical status
- House size
- Job
- Purchase history
- Etc

Ethical question:

Is it okay to use this kind of information?

Item features

Associated to the items in store

- Color
- Size
- Price
- Roundness
- Digital vs Analog
- Etc

Global features

Associated to the items-item or user-item global

- Total number of red clocks
- Total number of round clocks
- Total number clock A has been bought with clock B
- Total number of times user U has bought round clocks
- Number of users that have bought clock A
- Etc

Hands on exercise

1. Study the trivago Dataset
2. Create local features
3. Create global features
4. Create item features
5. Evaluate
6. Discuss computational bottlenecks

Feature Generation

~900k training sessions

Time

timestamp	step	action_type	reference	city	current_filters
1541108169	1	filter selection	Camping Site	Bacalar, Mexico	Camping Site
1541108268	2	interaction item image	3799074	Bacalar, Mexico	NaN
...					
1541108331	12	clickout item	8939464	Bacalar, Mexico	Camping Site
1541108336	13	clickout item	8939464	Bacalar, Mexico	Camping Site
1541108349	14	clickout item	9144656	Bacalar, Mexico	Camping Site

Device
mobile

Session_id:
0273bcf96dddf

User_id:

3IZW0KL458DJ

platform:
MX