Recommender Systems || Lecture 11

Summer 2022

Dr. Karthik Mohan

UW, Seattle

August 3, 2022

Logistics

1. Assignment 3 due next Wednesday

Logistics

1. Assignment 3 due next Wednesday
2. Project checkpoint due next Monday
3. Please pick a time slot for your team to discuss project propsoal if you haven't yet!

Logistics

1. Assignment 3 due next Wednesday
2. Project checkpoint due next Monday
3. Please pick a time slot for your team to discuss project propsoal if you haven't yet!
4. Anything else?!

Today

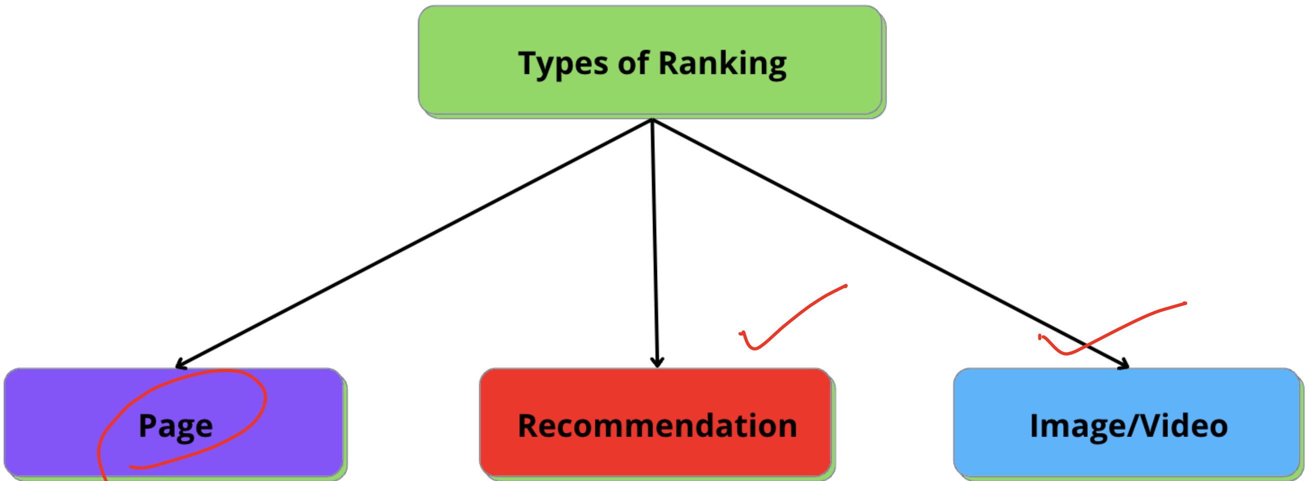
Today

1. Ranking loss function and metrics

Today

1. Ranking loss function and metrics
2. Page Ranking and Ranking Algorithms

Ranking Problems



Traditional Information Retrieval (IR)

IR → Traditional (Query → List of documents)

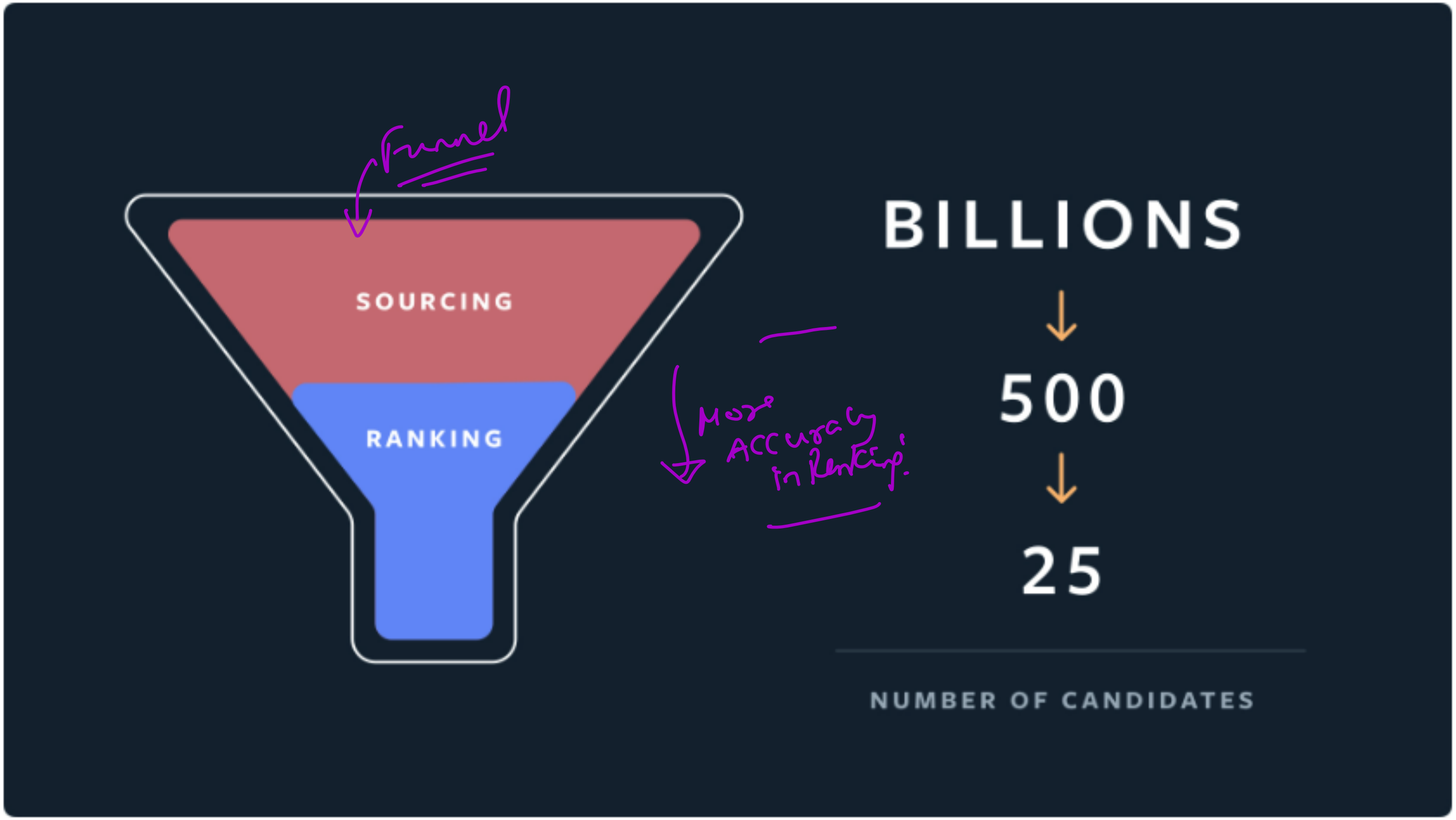
Latency

Retrieval Info.

Neural IR → challenge in Keeping Inference Latency Low

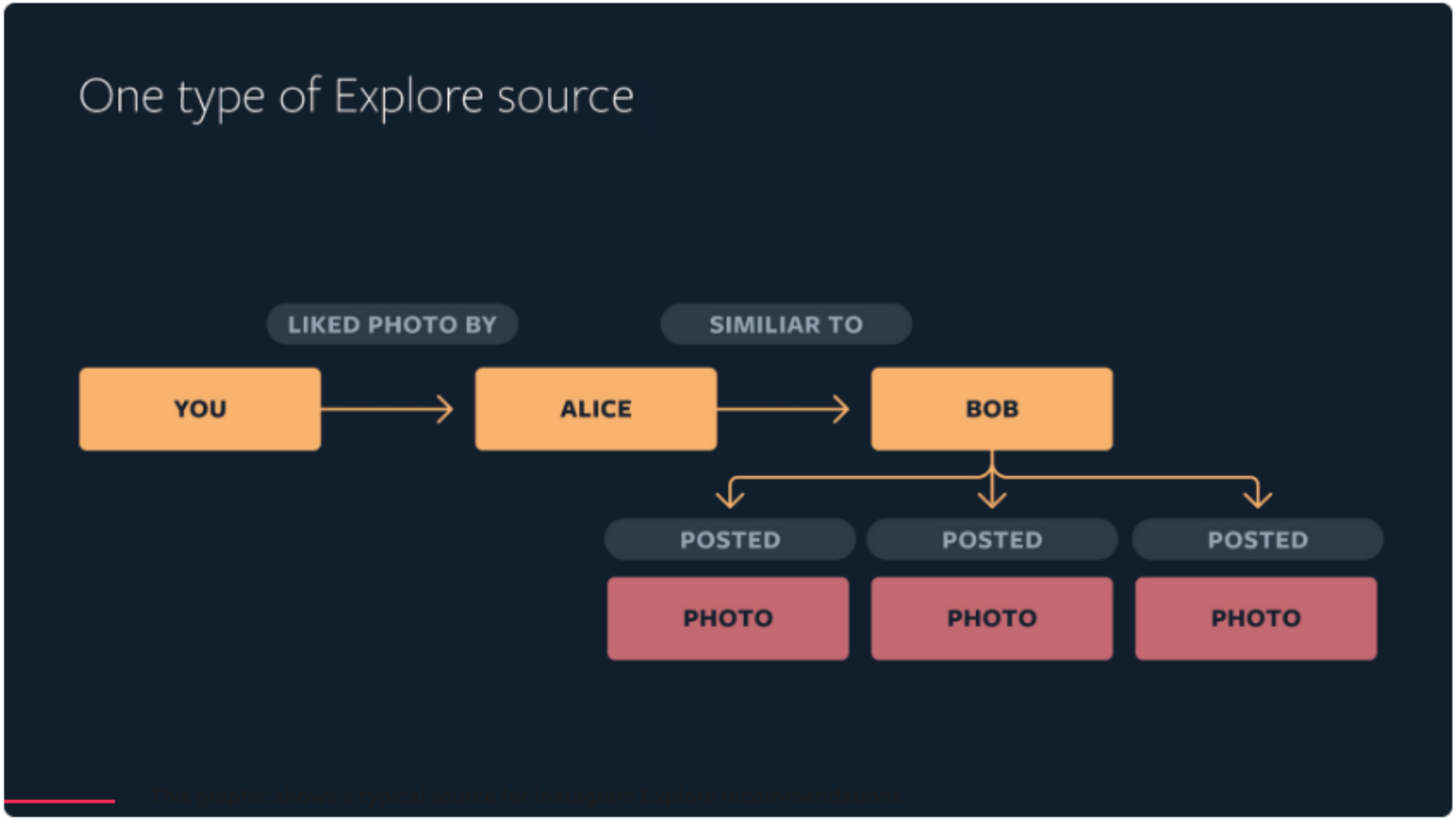
Latency/Memory

Image/Video Ranking (Instagram)



Instagram Recommendations

Image/Video Ranking (Instagram Candidate Generation)



Instagram Recommendations

Image/Video Ranking (Amazon Video)

Behavior-based Popularity Ranking on Amazon Video

With the growth in the number of video streaming services, providers have to strive hard to make relevant content available and keep customers engaged. A good experience would help customers discover new and popular videos to stream with ease. Customer streaming behavior tends to be a strong indicator of whether they found a video engaging. Aggregate customer behavior serves as a useful predictor of popularity. We discuss the use of past streaming behavior to learn patterns and predict a video's popularity using tree ensembles.

CCS Concepts: • **Information systems** → **Recommender systems**; **Content ranking**.

Additional Key Words and Phrases: video streaming; popularity; recommendations

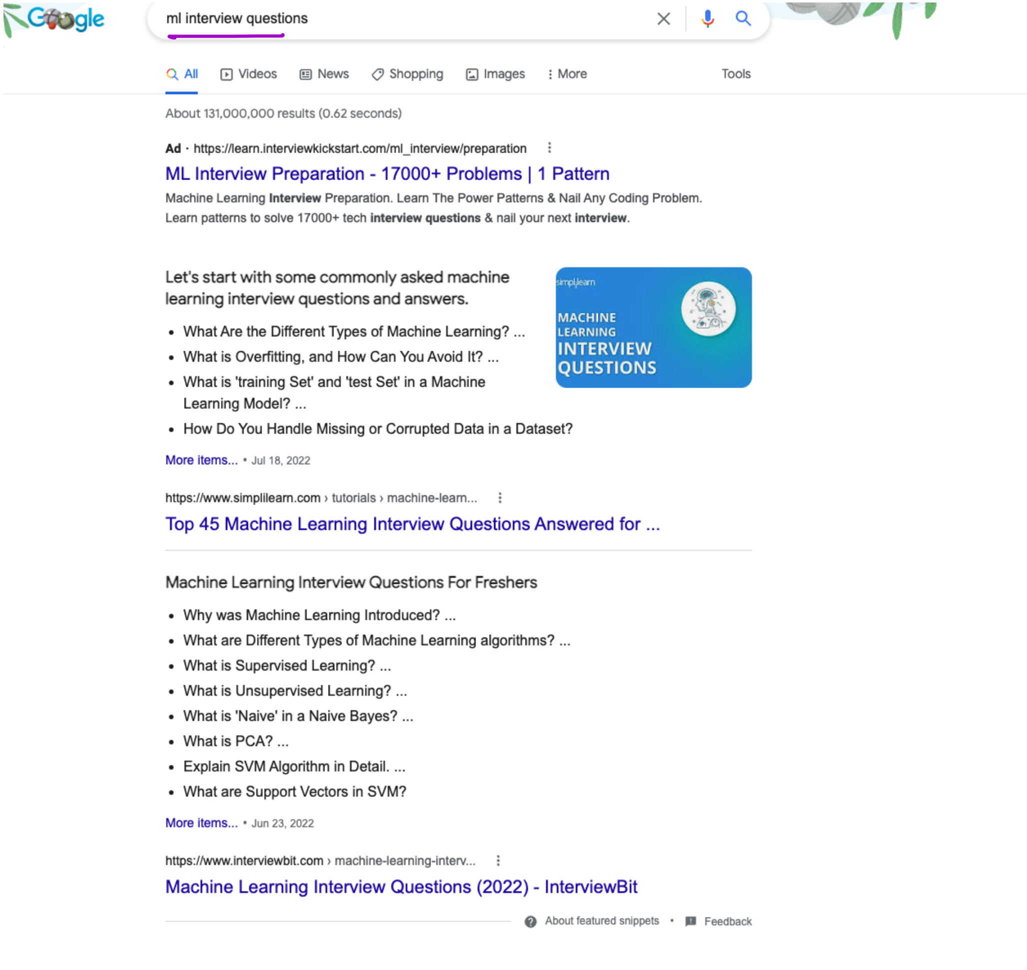
ACM Reference Format:

. 2020. Behavior-based Popularity Ranking on Amazon Video. In *Fourteenth ACM Conference on Recommender Systems (RecSys '20), September 22–26, 2020, Virtual Event, Brazil.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3383313.3411555

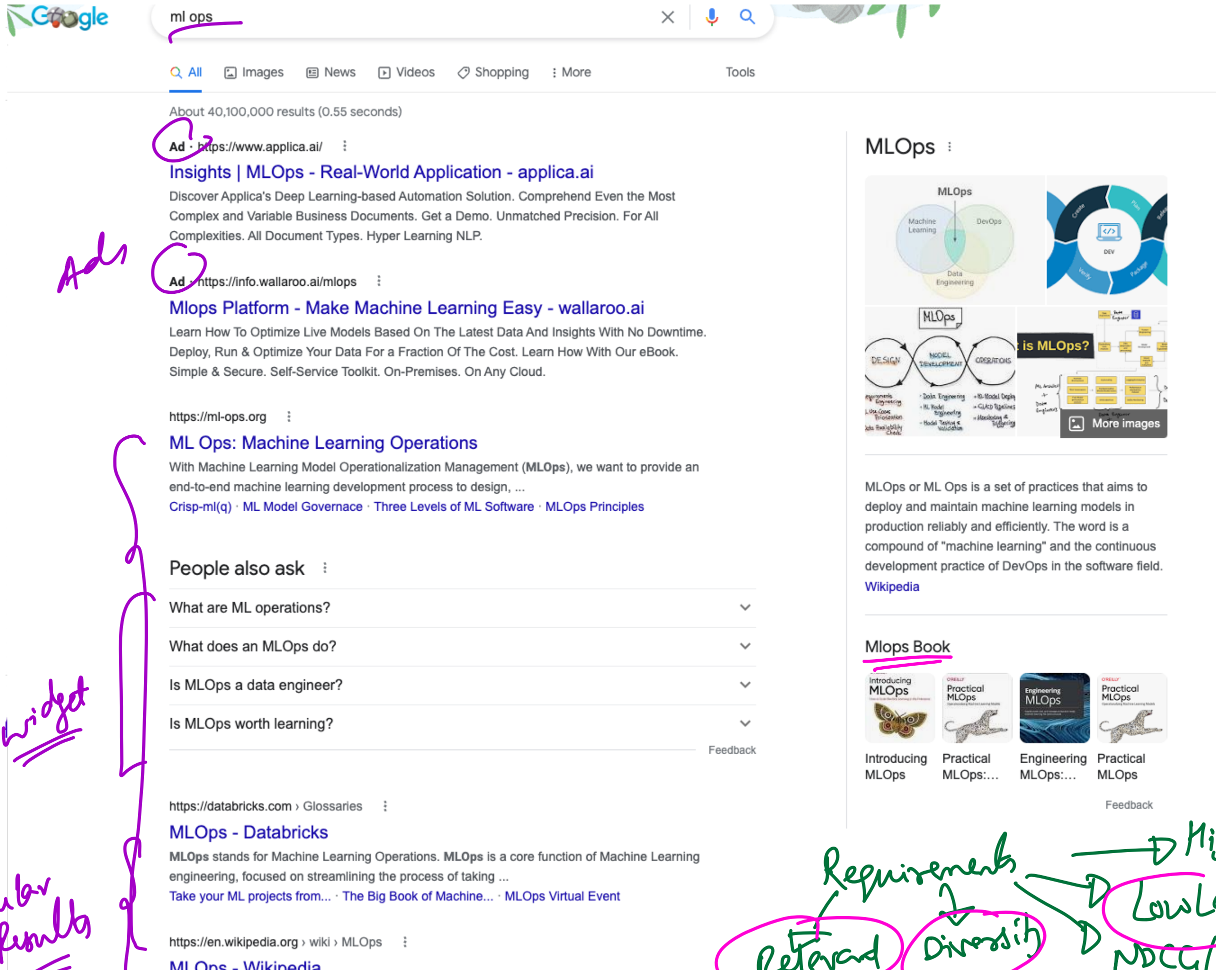
1 INTRODUCTION

On streaming platforms customers use two main approaches to find videos—search and discovery. Customers can enter a query to search for a video or browse the catalog to discover content. Streaming platforms use customer preferences,

Page/Search Ranking



Page/Search Ranking



Ads

widget

Regular Results

Requirements

Relevant

Diversity

→ High clicks

→ Low Latency

→ NDCG/MAP metrics

Types of Ranking

Types of Ranking (search)

Ranking Creation → Rank a List of Documents

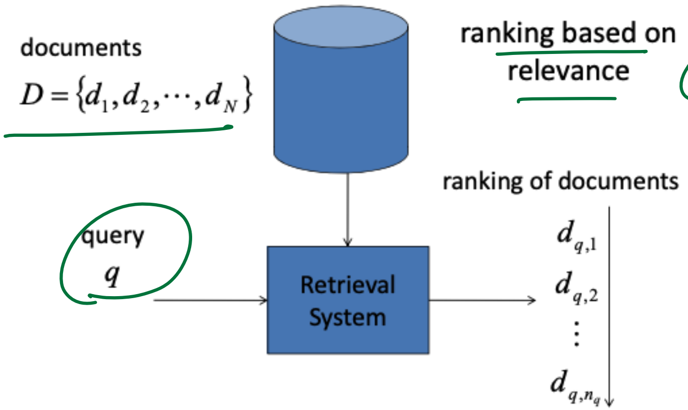
Ranking Aggregation

Aggregation of Ranking Lists → Example:- News Aggregator

Learning to Rank Survey

Search Ranking - Document Retrieval

Generation
↓
Filter
↓
Rank → ReRank



(diversity / freshness / recency / Filter Spam / Filter Inapp. Content)

Learning to Rank Survey

Recommendations Ranking - Collaborative Filtering

	Item1	Item2	Item3	...	ItemN
User1	5	4			
User2	1		2		2
...		?	?	?	
UserM	4	3			

Rank Items for each user

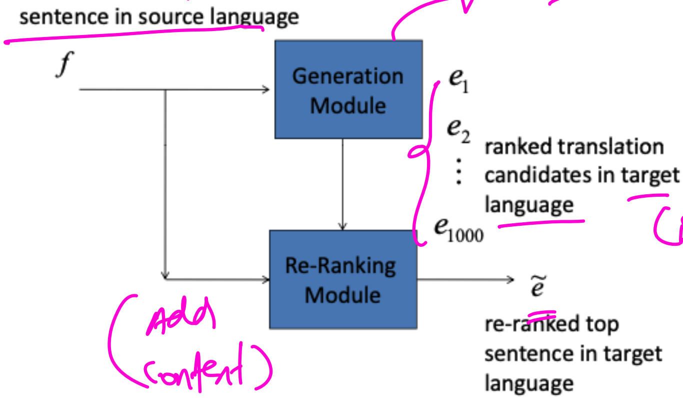
These into query like (q) in search!

Learning to Rank Survey

Re-ranking in Machine Translation

(German)

→ NN model



(English)

(Add content)

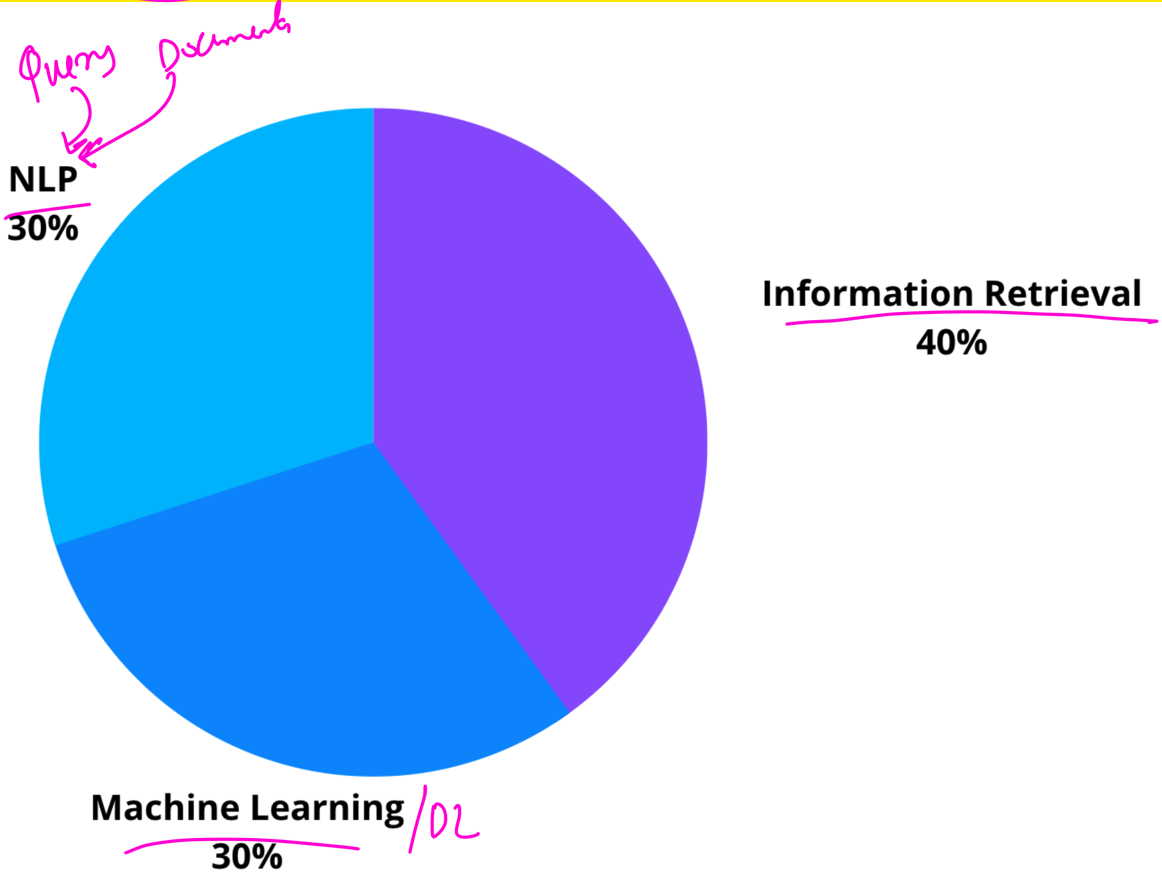
→ (Style of Language, personalization, etc)

MT is based on Seq2Seq models (SOTA)

Learning to Rank Survey

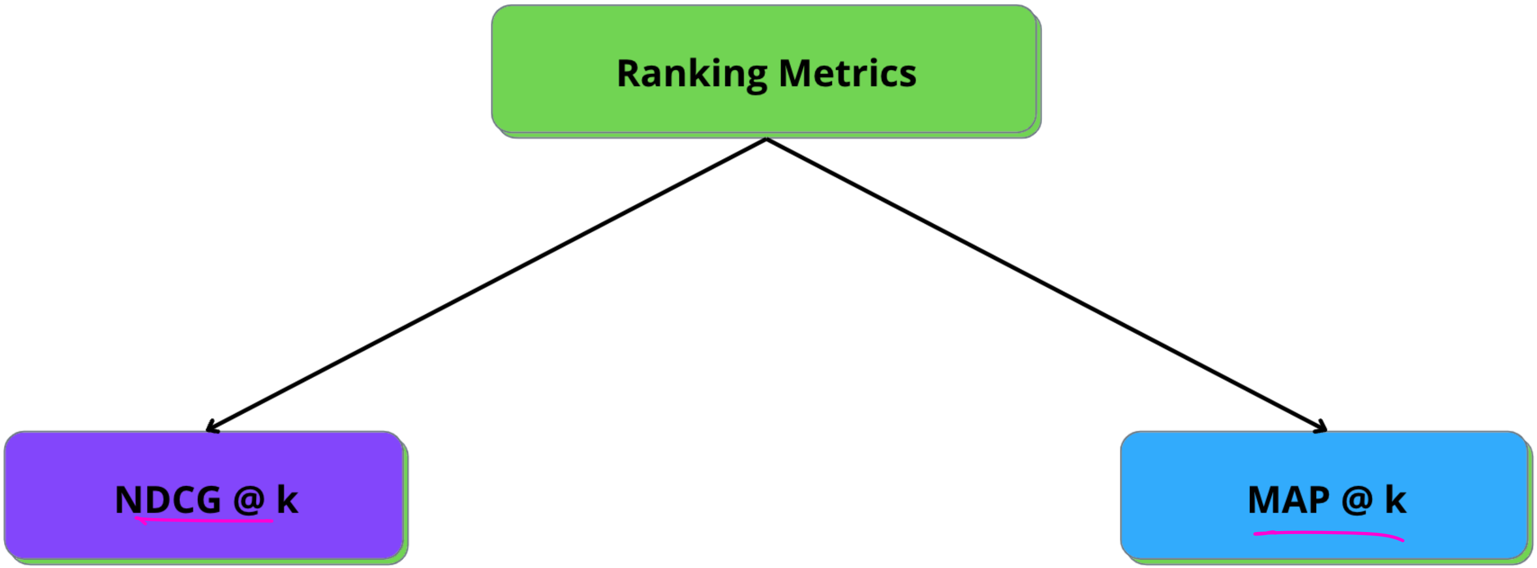
(Traditional IR) Concepts

Learning to Rank



Ranking Metrics

Top Interview Questions in search Ranking
→ Ranking loss
→ Ranking metrics



Learning to Rank Survey

Ranking Metric - DCG & NDCG

normalizing

Discounted Cumulative Gain

(Position) sum it up (Relevance of doc for query)

(Input) (Output)

Queries Documents

q_1 $d_{1,1}\ d_{1,2}, \dots d_{1,10}$

q_2

q_3 → First doc for query 3: $d_{3,1}\ d_{3,2} \dots d$

$\vdots$

q_{100}

DCG@K

Table 2.4: Example of NDCG

Perfect ranking	Formula	Explanation
(3, 3, 2, 2, 1, 1, 1)		grades: 3,2,1
(7, 7, 3, 3, 1, 1, 1)	$2^{y_{i,j}} - 1$	gains
(1, 0.63, 0.5, ···)	$1/\log(\pi_i(j)+1)$	position discounts
(7, 11.41, 12.91, ···)	$\sum_{j:\pi_i(j)\le k} \frac{2^{y_{i,j}}-1}{\log(\pi_i(j)+1)}$	DCG scores
(1/7, 1/11.41, 1/12.91,···)	$DCG_{max}^{-1}(k)$	normalizing factors
(1,1,1,···)	$NDCG(k)$	NDCG scores

Imperfect ranking	Formula	Explanation
(2, 3, 2, 3, 1, 1, 1)		grades: 3,2,1
(3, 7, 3, 7, 1, 1, 1)	$2^{y_{i,j}} - 1$	gains
(1, 0.63, 0.5, ···)	$1/\log(\pi_i(j)+1)$	position discounts
(3, 7.41, 8.91, ···)	$\sum_{j:\pi_i(j)\le k} \frac{2^{y_{i,j}}-1}{\log(\pi_i(j)+1)}$	DCG scores
(1/7, 1/11.41, 1/12.91,···)	$DCG_{max}^{-1}(k)$	normalizing factors
(0.43, 0.65, 0.69, ···)	$NDCG(k)$	NDCG scores

Grade to Gain

Ground Truth data

$y_{i,j}$ → Grade of document j for query i

i – Query

j – document

$$DCG@4 = \sum_{j=1}^{4} \frac{Gain(j)}{Discount(j)}$$

Max value for NDCG(K) = 1

Min value for NDCG(K) = 0

$$NDCG(K) = \frac{DCG(K)}{DCG_{max}(K)}$$

Grades → Higher the grade – More relevant the document

Relevance of a Document

↳ human annotated

3 grades (not relevant, relevant, highly relevant)

Gains ← Function of Grade (exponentiation)

ICE #1

DCG computation

For a query, "show me top 10 thriller movies of this century" - a search engine returns 10 results: $d_1, d_2, \ldots, d_{10}$, where d_j represents document/page j. Let the grade of the documents be given as follows: $\{3, 2, 1, 5, 2, 4, 5, 6, 7, 4\}$. The $DCG@3$ for the results of the search is closest to:

1. 11
2. 21
3. 31
4. 41

$$DCG@K = \frac{Gain(1)}{\log(1+1)} + \frac{Gain(2)}{\log(2+1)} + \cdots + \frac{Gain(K)}{\log(K+1)}$$

$$\frac{2^3-1}{\log(2)} + \frac{2^2-1}{\log(3)} + \frac{2^1-1}{\log(4)} \approx \underline{31!}$$

$$Gain(K) = 2^{grade(K)} - 1$$

ICE #2

NDCG computation

For a query, "show me top 10 thriller movies of this century" - a search engine returns 10 results: $d_1, d_2, \ldots, d_{10}$, where d_j represents document/page j. Let the grade of the documents be given as follows: $\{3, 2, 1, 5, 2, 4, 5, 6, 7, 4\}$. The $NDCG@1$ for the results of the search is closest to:

1. 0.05
2. 0.1
3. 0.15
4. 0.2

$$NDCG@K = \frac{DCG@K}{DCG_{max}@K} = \frac{2^3 - 1}{2^7 - 1} = \frac{7}{127} = 0.05$$

$$DCG@1 \propto 2^3 - 1 \qquad DCG_{max}@1 \propto 2^7 - 1 =$$

Ranking Metric - Mean Average Precision (MAP)

(1, 0, 1, 1, 0, 0, 0)

Precision@1 = 1/1   P@4 = 3/4

Precision@2 = 1/2

Precision@3 = 2/3

Similar to Precision@K!!

MAP = AP for all queries averaged/mean

Table 2.5: Example of MAP

Perfect ranking	Formula	Explanation
(1, 0, 1, 1, 0, 0, 0)		Ranks:1,0
(1, -, 0.67, 0.75, –, –, –)	$P(j)$	Precision at position j with label 1
0.81	AP	Average Precision

Grades are 0 or 1

AP@3 = ICE#3

MAP ← Improved precision metric that incorporates ranking!!

$AP@4 = \frac{1 + 0.67 + 0.75}{4} = 0.81!$

ICE #3

AP (Average Precision) computation

For a query, "show me top 10 thriller movies of this century" - a search engine returns 10 results: $d_1, d_2, \ldots, d_{10}$, where d_j represents document/page j. Let the normalized grade of the documents be given as follows: $\{1, 0, 1, 1, 0, 0, 1, 0, 1, 0\}$. The $AP@3$ for the results of the search is closest to:

1. 0.35
2. 0.45
3. 0.55
4. 0.65

Different Example

Precision $= 3/7 = 0.42$

P@3 P@6 P@1 P@2

$1/3$ $2/6$ 1 $2/2 = 1$

1 0 0 1 0 0 1 1 2 1 1 0 0 0 0 1

$AP@7 = \frac{1/3 + 2/6 + 3/7}{3} = 0.36$ $AP = \frac{1 + 1 + 3/7}{3} = 0.809$

Next Class

NDCG and approx. NDCG

1. Ranking Loss functions
2. Neural models for Search Ranking and Re-ranking