Delivery Hero

Shedding Light on Positional Bias: Strategies for Mitigation

Burak ISIKLI

Delivery Hero

Haystack EU 2023

Burak Isikli

- PhD Candidate (Computer Science)
- 16 years of experience in various fields
- Staff ML Engineer @ Delivery Hero, Search Personalization Ranking

Who are we?

- Food ordering & delivery
- Restaurants & Quick commerce
- 70+ countries
- People we serve: up to 2.2 billion
- Restaurant partners & local vendors: 1.5 million

Delivery Hero

foodpanda

Glovo

PedidosYa

Yemeksepeti

foodora

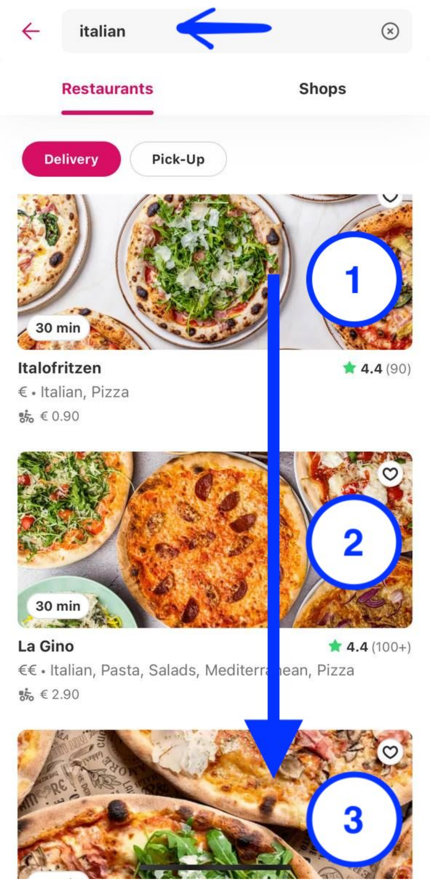
talabat

Why Positional Bias

- When was the last time you scrolled to the second page?
- What if you would be hungry :)

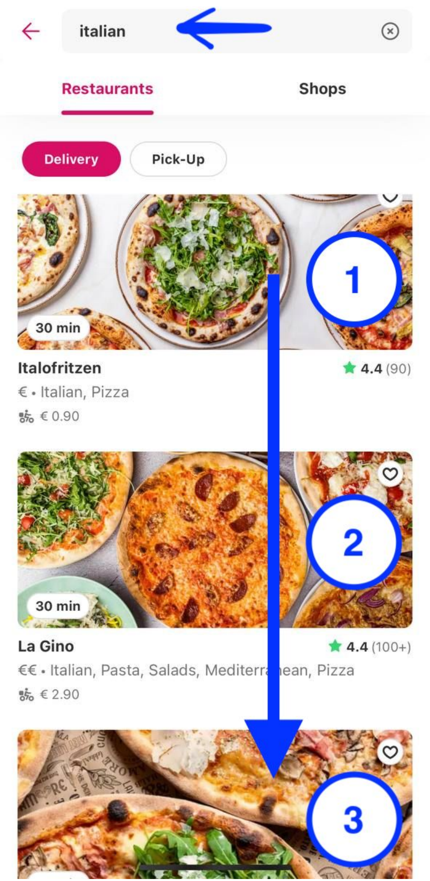
Why Positional Bias

- Goal:
  - Presentation Bias
  - To make fair results for all of the vendors
  - E.g On Google Search, users click on the first position 10x more than the tenth position



Why Positional Bias

- Reasons:
  - Trust the algorithm and select from the top result
  - Truly relevant results on the top result
  - Training the model on biased historical data perpetuates the bias (Interaction features)



italian

Restaurants

Shops

Delivery

Pick-Up

Italofritzen ★ 4.4 (90)
€ • Italian, Pizza
€ 0.90

La Gino ★ 4.4 (100+)
€€ • Italian, Pasta, Salads, Mediterranean, Pizza
€ 2.90

German Tank Problem

https://en.wikipedia.org/wiki/German_tank_problem

- During World War II, production of German tanks such as the Panther was accurately estimated by Allied intelligence using statistical methods

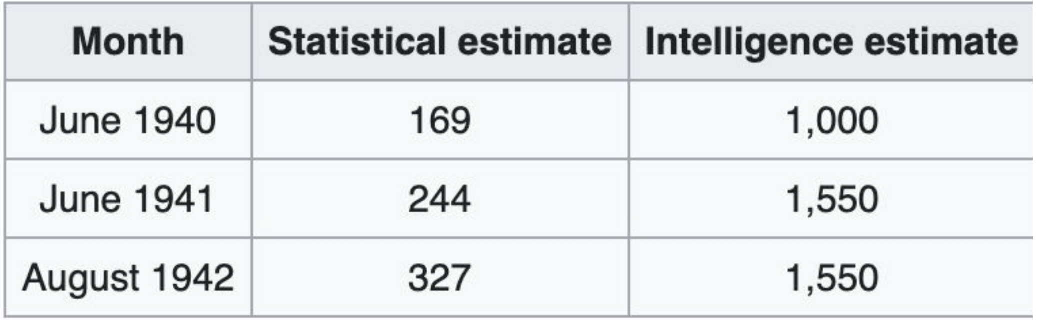
German Tank Problem

https://en.wikipedia.org/wiki/German_tank_problem

- Serial numbers: 19, 40, 42 and 60
- Let
  - N equal the total number of tanks predicted to have been produced,
  - m equal the highest serial number observed and
  - k equal the number of tanks captured.
- $N \sim= m + m/k - 1$
- $N = 60 + (60/4) - 1 = 74$

German Tank Problem

https://en.wikipedia.org/wiki/German_tank_problem

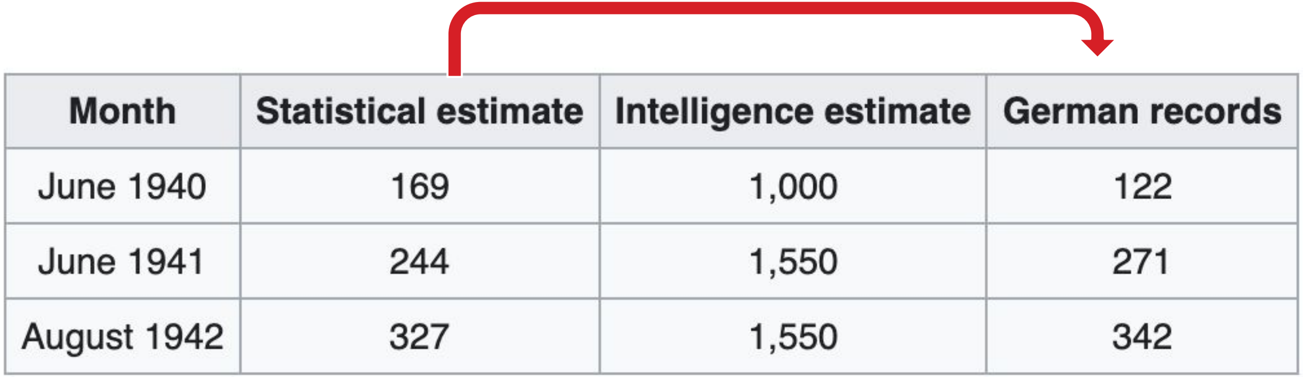


Month	Statistical estimate	Intelligence estimate
June 1940	169	1,000
June 1941	244	1,550
August 1942	327	1,550

German Records?

German Tank Problem

https://en.wikipedia.org/wiki/German_tank_problem



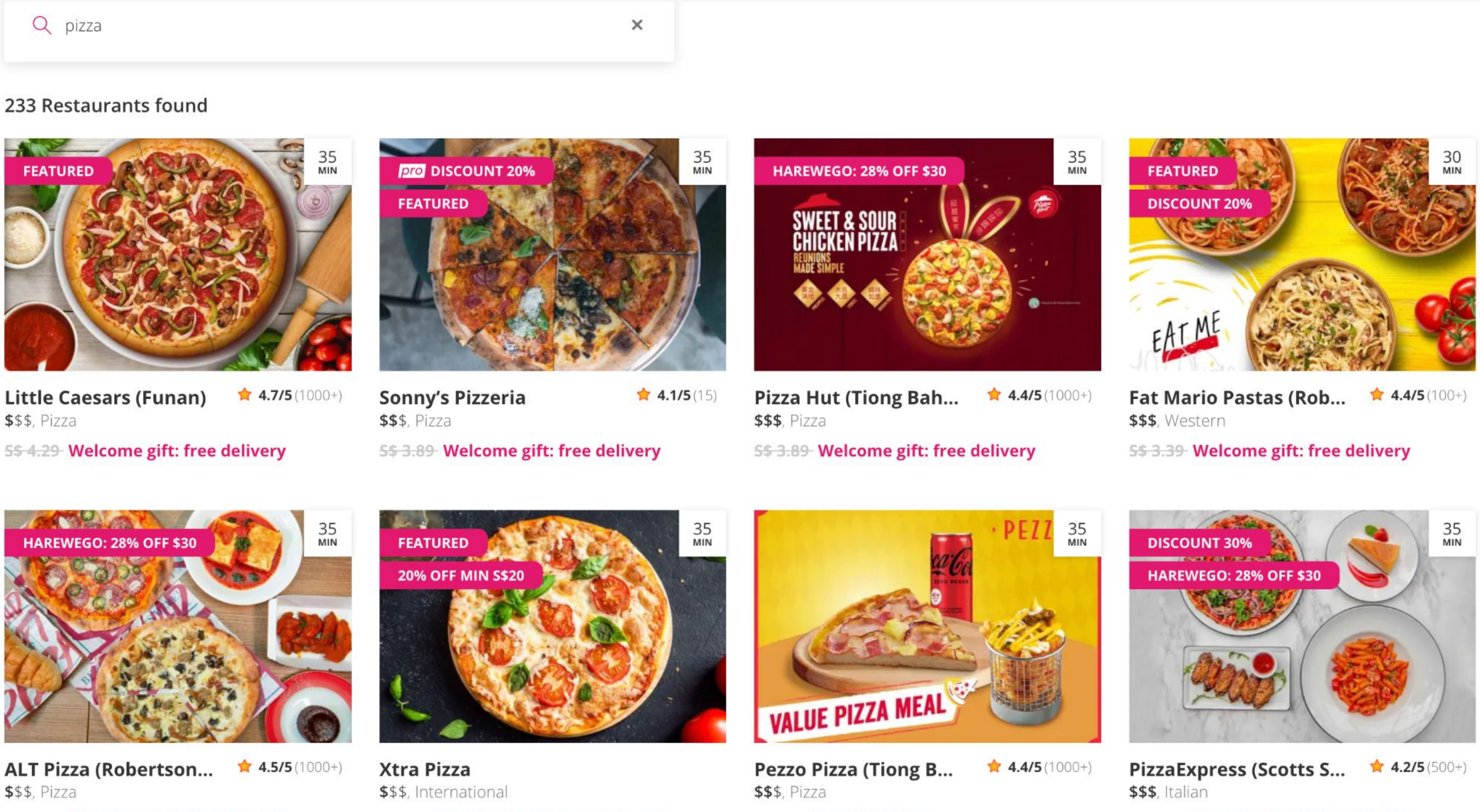
Month	Statistical estimate	Intelligence estimate	German records
June 1940	169	1,000	122
June 1941	244	1,550	271
August 1942	327	1,550	342

What does this have to do with us?

- No impression data
- We don’t know how far user scrolls

How?

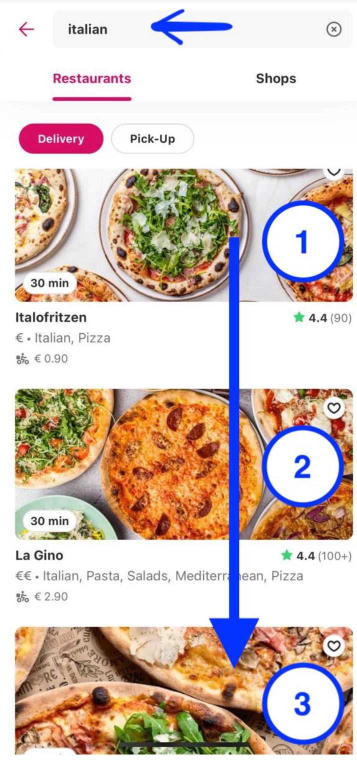
https://en.wikipedia.org/wiki/German_tank_problem



pizza

233 Restaurants found

FEATURED 35 MIN Little Caesars (Funan) ★ 4.7/5 (1000+) $$$, Pizza ~~S$ 4.29~~ Welcome gift: free delivery	pro DISCOUNT 20% FEATURED 35 MIN Sonny's Pizzeria ★ 4.1/5 (15) $$$, Pizza ~~S$ 3.89~~ Welcome gift: free delivery	HAREWEGO: 28% OFF $30 35 MIN Pizza Hut (Tiong Bah... ★ 4.4/5 (1000+) $$$, Pizza ~~S$ 3.89~~ Welcome gift: free delivery	FEATURED DISCOUNT 20% 30 MIN Fat Mario Pastas (Rob... ★ 4.4/5 (100+) $$$, Western ~~S$ 3.39~~ Welcome gift: free delivery
HAREWEGO: 28% OFF $30 35 MIN ALT Pizza (Robertson... ★ 4.5/5 (1000+) $$$, Pizza	FEATURED 20% OFF MIN S$20 35 MIN Xtra Pizza $$$, International	35 MIN Pezzo Pizza (Tiong B... ★ 4.4/5 (1000+) $$$, Pizza	DISCOUNT 30% HAREWEGO: 28% OFF $30 35 MIN PizzaExpress (Scotts S... ★ 4.2/5 (500+) $$$, Italian



italian

Restaurants Shops

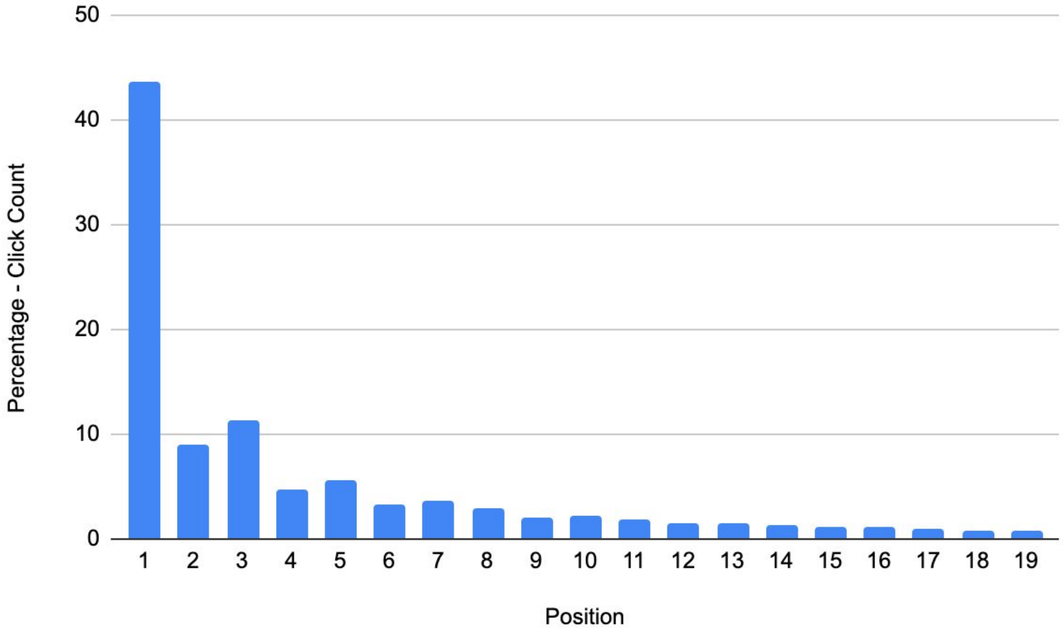
Delivery Pick-Up

30 min
Italofritzen ★ 4.4 (90)
€ • Italian, Pizza
€ 0.90

30 min
La Gino ★ 4.4 (100+)
€€ • Italian, Pasta, Salads, Mediterranean, Pizza
€ 2.90

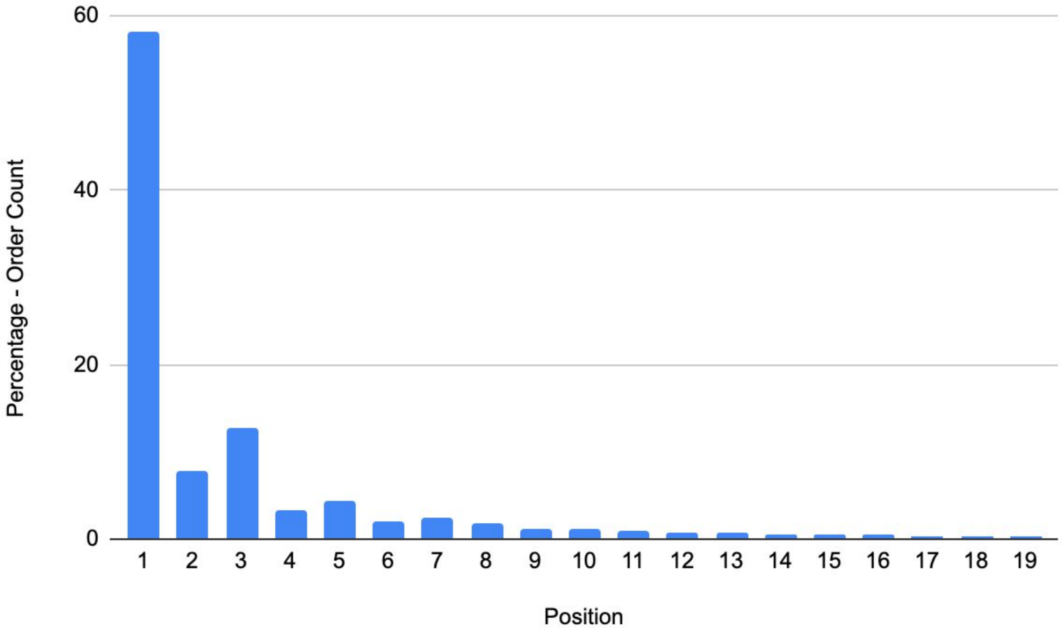
How

How?

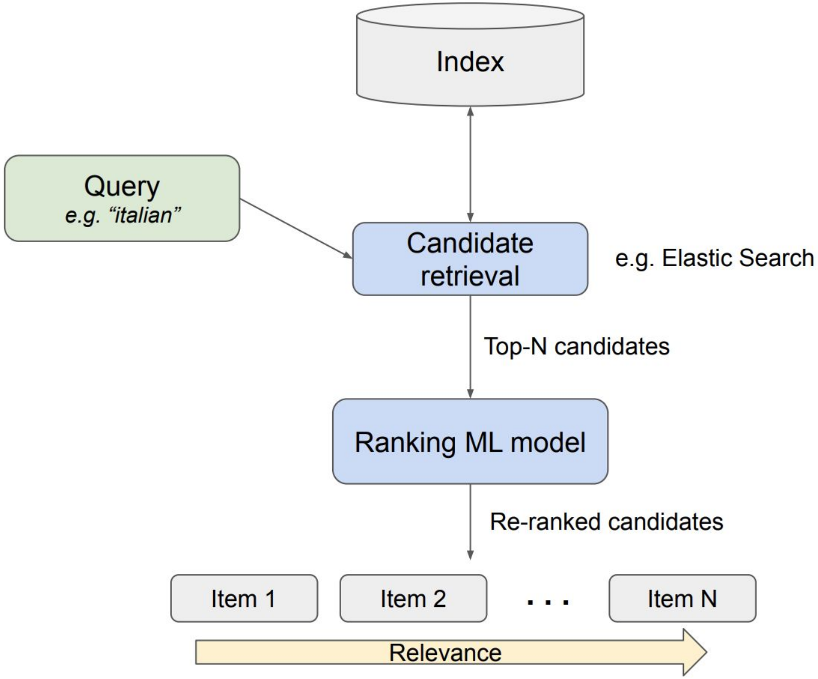


50
40
30
20
10
0
Percentage - Click Count
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
Position

How?


60
40
20
0
Percentage - Order Count
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
Position

Architecture



Features

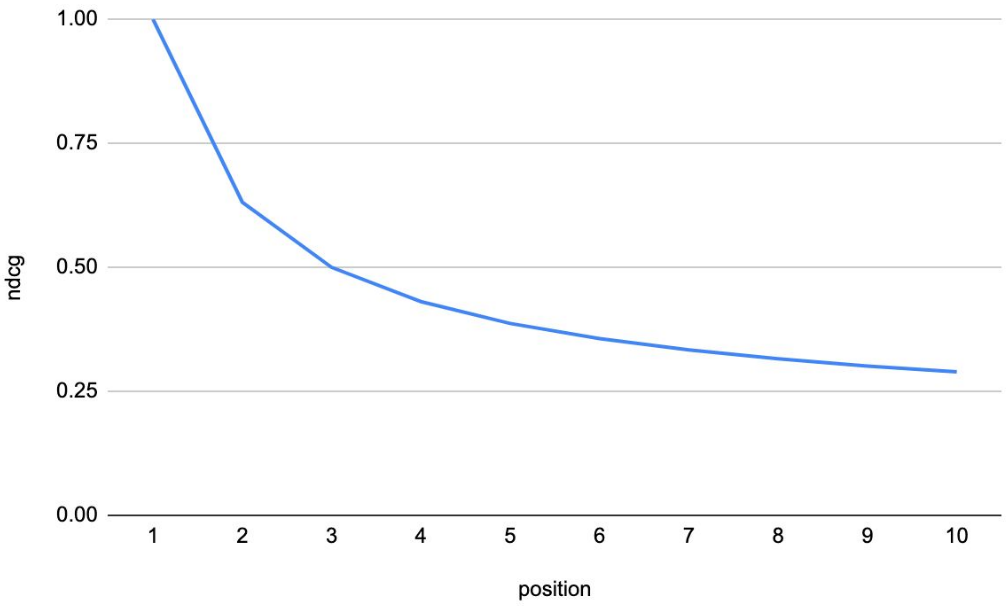
- Textual Similarities
- **Behavioural/Interaction Features**
- Price
- Context

Relevance

- Binary
  - Ordered/Not Ordered (1/0)

Offline Evaluation Metrics

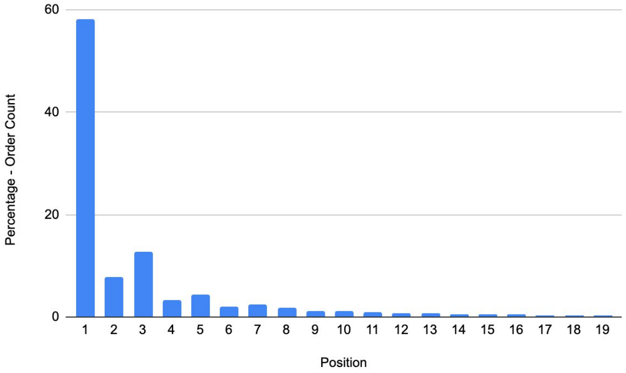
- nDCG
  - nDCG@50
  - nDCG@2-10

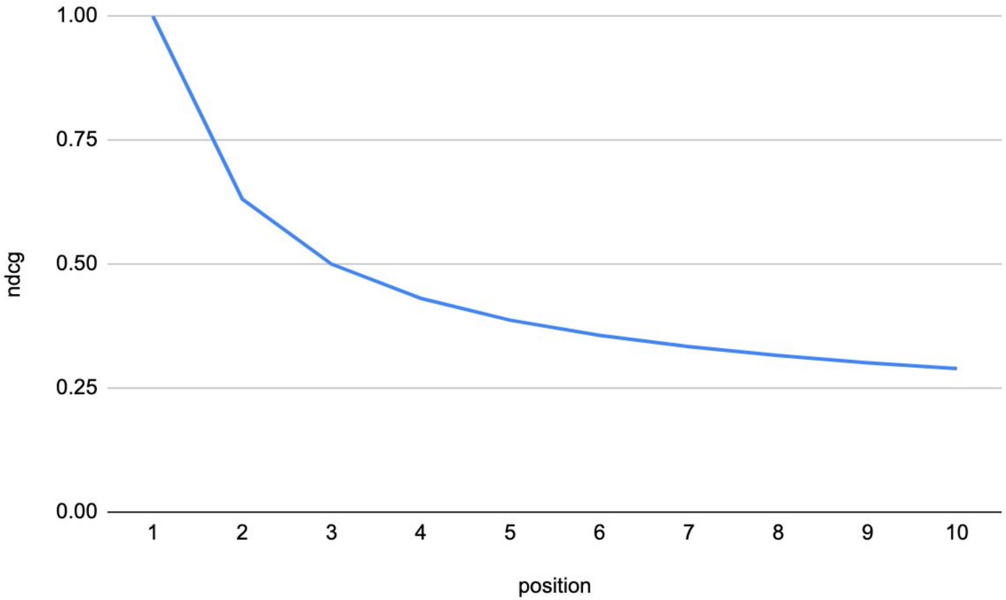


How

Offline Evaluation Metrics

- nDCG
  - nDCG@50
  - nDCG@2-10
- Recap





How ● ● ● ● ●

Offline Evaluation Metrics

- nDCG
  - nDCG@50
  - nDCG@2-10
- Diversity of vendors
  - How many distinct items have been shown at first k slots (k=10, 20)
- Novelty
  - How well the restaurants are shown in top-k result?

1. Positional-aware Learning

- Position as a feature
- Training: Enabled
- Inference
    - Fixed position: 0, 1, 5, 50?

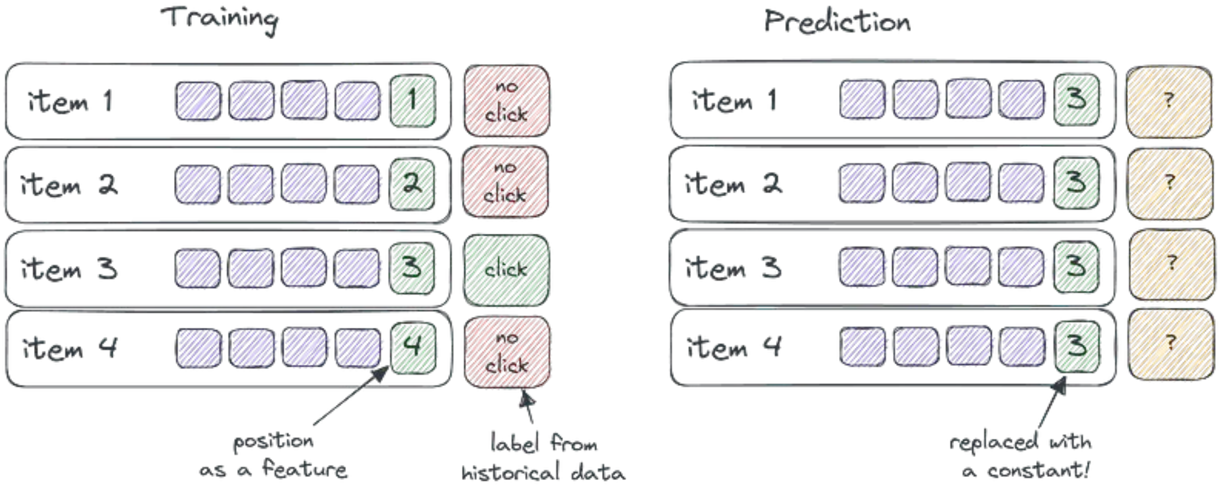


Image: https://www.kdnuggets.com/wp-content/uploads/grebennikov_dealing_position_bias_recommendations_search_13.png

1. Positional-aware Learning

- Position as a feature
- Training: Enabled
- Dropout
    - Positions of some sessions can be discarded (10%, 20%, etc.)

Session_id	Features	Positions	Use this session
1	…..	1,2,3,4,5	✓
2	…..	1,2,3,4,5	✓
3	…..		X (20%)
4	…..	1,2,3,4,5	✓
5	…..	1,2,3,4,5	✓

Image: https://www.kdnuggets.com/wp-content/uploads/grebennikov_dealing_position_bias_recommendations_search_13.png

1. Positional-aware Learning

- nDCG

nDCG@50		
Value	Dropout	Δ current (%)
0	0.1	-0.416
1	0.1	-0.76

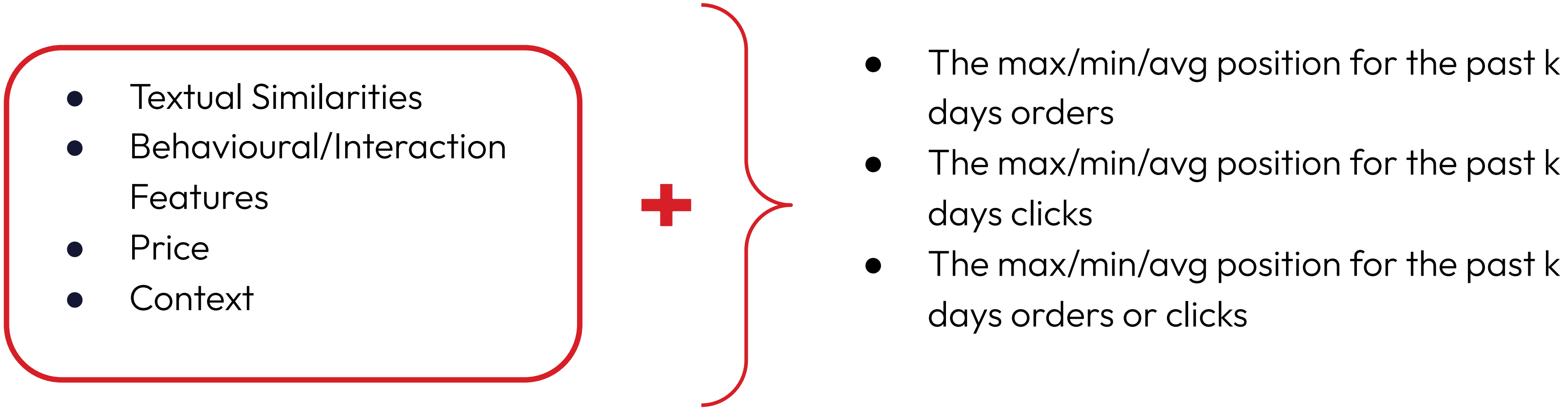
nDCG@2-10		
Value	Dropout	Δ current (%)
0	0.1	6.293
1	0.1	5.964

1. Positional-aware Learning

- Diversity
  - Number of distinct vendors in top-k positions

diversity@10		
Value	Dropout	Δ current (%)
0	0.1	4.919

2. Proximity Features



- Textual Similarities
- Behavioural/Interaction Features
- Price
- Context

+

- The max/min/avg position for the past k days orders
- The max/min/avg position for the past k days clicks
- The max/min/avg position for the past k days orders or clicks

2. Proximity Features

- Best possible feature set: [order_avg_position_30d + click_avg_position_30d + all_avg_position_30d + order_max_position_30d]

nDCG@50
Δ current (%)
0.62

nDCG@2-10
Δ current (%)
8.31

2. Proximity Features

- Diversity

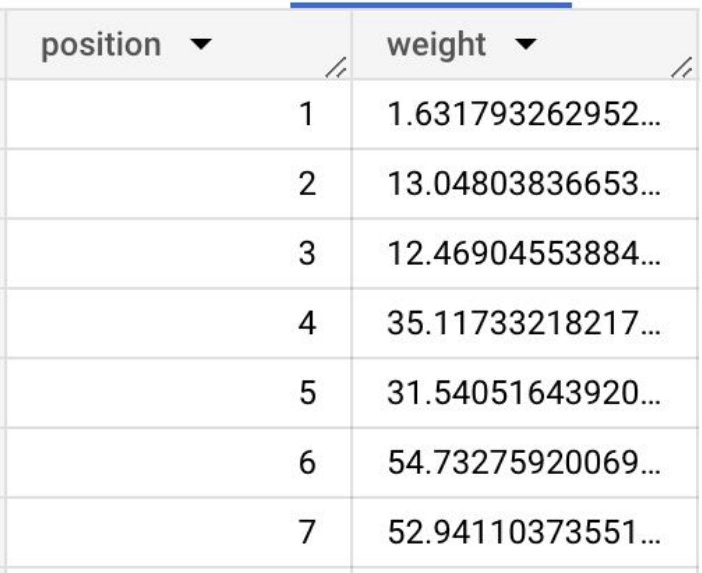
diversity@10
Δ current (%)
- 0.23

3. Punished Interaction Features

- During the click count calculation, formula is defined to punish if the position is higher
- Similar to Inverse Propensity Weighting

Click_count = Σ (click_conversion*weight)

Weight can be calculated based on the number of orders



position	weight
1	1.631793262952...
2	13.04803836653...
3	12.46904553884...
4	35.11733218217...
5	31.54051643920...
6	54.73275920069...
7	52.94110373551...

3. Punished Interaction Features

- nDCG

nDCG@50
Δ current (%)
0.52

nDCG@2-10
Δ current (%)
8.18

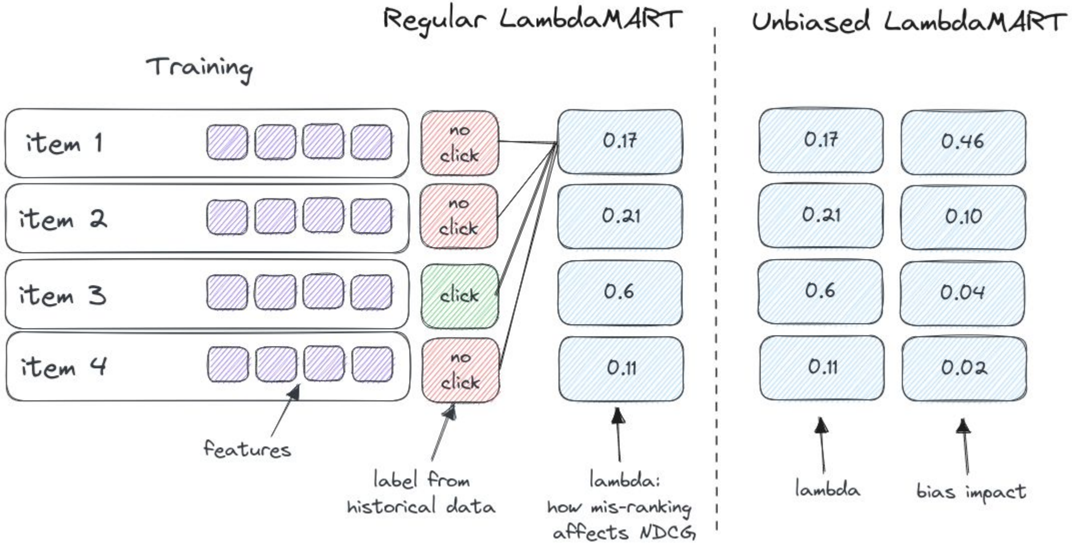
3. Punished Interaction Features

- Diversity

diversity@10
Δ current (%)
3.18

4. Unbiased Lambdamart

- Unbiased LambdaMART: An Unbiased Pairwise Learning-to-Rank Algorithm, https://dl.acm.org/doi/10.1145/3308558.3313447



4. Unbiased Lambdamart

Currently;

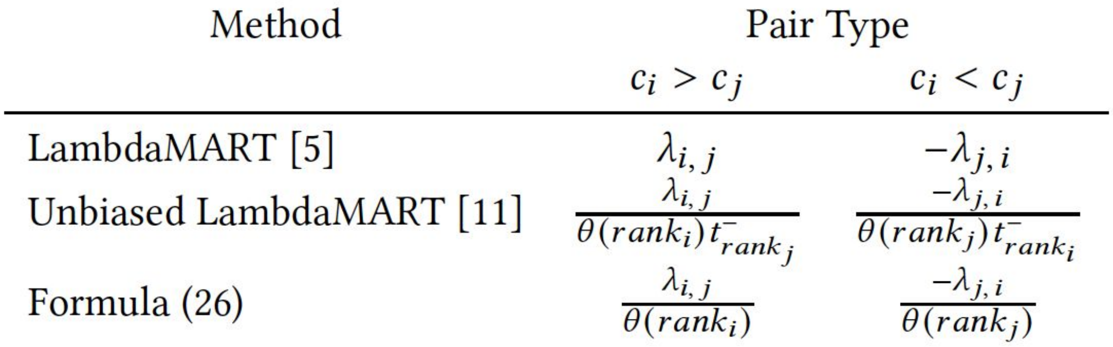
- Ranking metric is optimised on nDCG (conversion aka orders)
- Objective: LambdaRank
- Method: LambdaMART
    - https://github.com/microsoft/LightGBM/blob/master/src/objective/rank_objective.hpp

In positional bias;

- Unbiased LambdaMART: An Unbiased Pairwise Learning-to-Rank Algorithm
- Same nDCG metric with different cutoff (k=10, k=20)
- Objective: Custom
    - https://github.com/burakisiklidh/unbiased-lambdamart
- 2 Additional parameters
    - Positional bins
    - Punishment Ratio
- Method: LambdaMART

4. Unbiased Lambdamart

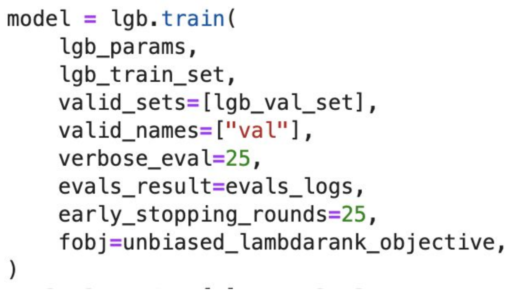
- A General Framework for Pairwise Unbiased Learning to Rank https://dl.acm.org/doi/abs/10.1145/3539813.3545119
- https://github.com/zalandoresearch/pairwise-debiasing



Method	Pair Type	
	$c_i > c_j$	$c_i < c_j$
LambdaMART [5]	$\lambda_{i,j}$	$-\lambda_{j,i}$
Unbiased LambdaMART [11]	$\frac{\lambda_{i,j}}{\theta(rank_i)t^-_{rank_j}}$	$\frac{-\lambda_{j,i}}{\theta(rank_j)t^-_{rank_i}}$
Formula (26)	$\frac{\lambda_{i,j}}{\theta(rank_i)}$	$\frac{-\lambda_{j,i}}{\theta(rank_j)}$

4. Unbiased Lambdamart

- Lightgbm supports custom objective (fobj)

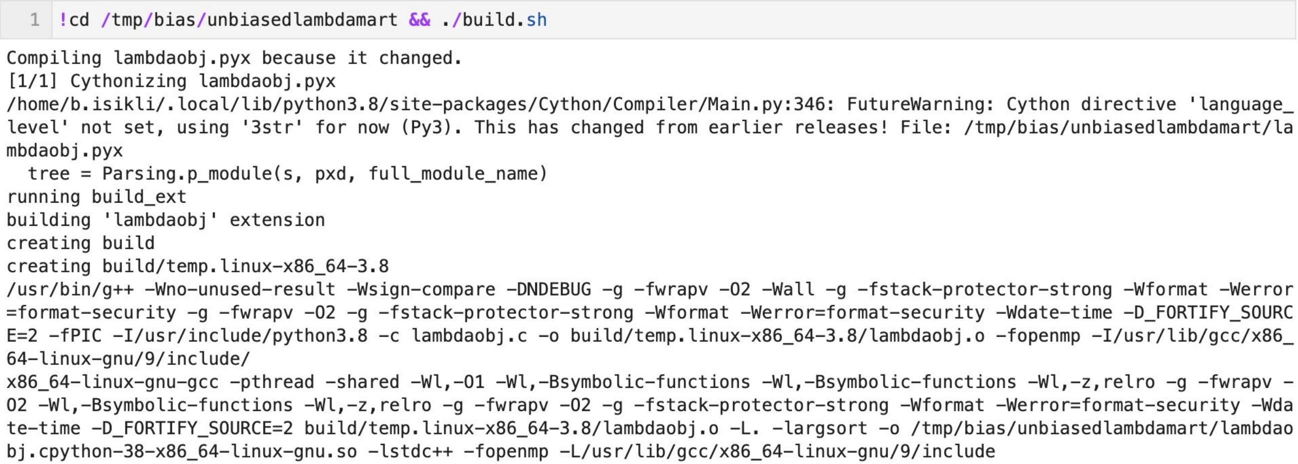


```
model = lgb.train(
    lgb_params,
    lgb_train_set,
    valid_sets=[lgb_val_set],
    valid_names=["val"],
    verbose_eval=25,
    evals_result=evals_logs,
    early_stopping_rounds=25,
    fobj=unbiased_lambdarank_objective,
)
```

- Overcome GIL
    - Cython (Python with C data types)

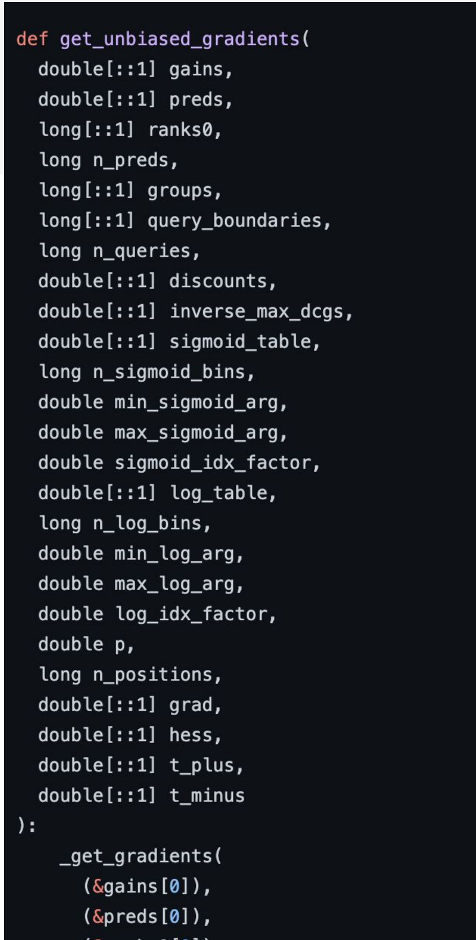
4. Unbiased Lambdamart

- C Programming Language
- Makefile
- Building



```
!cd /tmp/bias/unbiasedlambdamart && ./build.sh
```

```
Compiling lambdaobj.pyx because it changed.
[1/1] Cythonizing lambdaobj.pyx
/home/b.isikli/.local/lib/python3.8/site-packages/Cython/Compiler/Main.py:346: FutureWarning: Cython directive 'language_
level' not set, using '3str' for now (Py3). This has changed from earlier releases! File: /tmp/bias/unbiasedlambdamart/la
mbdaobj.pyx
  tree = Parsing.p_module(s, pxd, full_module_name)
running build_ext
building 'lambdaobj' extension
creating build
creating build/temp.linux-x86_64-3.8
/usr/bin/g++ -Wno-unused-result -Wsign-compare -DNDEBUG -g -fwrapv -O2 -Wall -g -fstack-protector-strong -Wformat -Werror
=format-security -g -fwrapv -O2 -g -fstack-protector-strong -Wformat -Werror=format-security -Wdate-time -D_FORTIFY_SOURC
E=2 -fPIC -I/usr/include/python3.8 -c lambdaobj.c -o build/temp.linux-x86_64-3.8/lambdaobj.o -fopenmp -I/usr/lib/gcc/x86_
64-linux-gnu/9/include/
x86_64-linux-gnu-gcc -pthread -shared -Wl,-O1 -Wl,-Bsymbolic-functions -Wl,-Bsymbolic-functions -Wl,-z,relro -g -fwrapv -
O2 -Wl,-Bsymbolic-functions -Wl,-z,relro -g -fwrapv -O2 -g -fstack-protector-strong -Wformat -Werror=format-security -Wda
te-time -D_FORTIFY_SOURCE=2 build/temp.linux-x86_64-3.8/lambdaobj.o -L. -largsort -o /tmp/bias/unbiasedlambdamart/lambdao
bj.cpython-38-x86_64-linux-gnu.so -lstdc++ -fopenmp -L/usr/lib/gcc/x86_64-linux-gnu/9/include
```



```
def get_unbiased_gradients(
  double[::1] gains,
  double[::1] preds,
  long[::1] ranks0,
  long n_preds,
  long[::1] groups,
  long[::1] query_boundaries,
  long n_queries,
  double[::1] discounts,
  double[::1] inverse_max_dcgs,
  double[::1] sigmoid_table,
  long n_sigmoid_bins,
  double min_sigmoid_arg,
  double max_sigmoid_arg,
  double sigmoid_idx_factor,
  double[::1] log_table,
  long n_log_bins,
  double min_log_arg,
  double max_log_arg,
  double log_idx_factor,
  double p,
  long n_positions,
  double[::1] grad,
  double[::1] hess,
  double[::1] t_plus,
  double[::1] t_minus
):
    _get_gradients(
      (&gains[0]),
      (&preds[0]),
```

4. Unbiased Lambdamart

- Unbiased LambdaMART can jointly estimate the biases at click positions and the biases at unclick positions, and learn an unbiased ranker
- https://github.com/microsoft/LightGBM/blob/master/src/objective/rank_objective.hpp
  - Init
  - GetGradients
  - GetGradientsForOneQuery
- Should return
  - The value of the first order derivative (gradient)
  - The value of the second order derivative (Hessian) of the loss
  - https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMRanker.html
- No need to change anything on inference side

4. Unbiased Lambdamart

- Xgboost 2.0 Experimental Feature: Unbiased Lambdamart
- https://github.com/dmlc/xgboost/releases/tag/v2.0.0rc1

4. Unbiased Lambdamart

- REG_P_VALUES = [0.0, 1.0, 2.0]
- position_bias_power = 0.5
- nDCG

nDCG@50
Δ current (%)
0.14

nDCG@2-10
Δ current (%)
7.09

4. Unbiased Lambdamart

- Diversity

diversity@10
Δ current (%)
2.73

5. Conclusion

	Δ nDCG@50	Δ nDCG@2-10	Δ Diversity@10
PAL	-0.416	6.293	**4.919**
Proximity	**0.62**	**8.31**	-0.23
Punished Interactions	0.52	8.18	3.18
Unbiased Lambdamart	0.14	7.09	2.73

5. Conclusion

- If you're using clicks/purchases as implicit labels, then their position also matters!
- We can use custom loss to unbias the model
- A lot of researches are done on neural ranking, a few for gbdt

Delivery Hero

Feedbacks/Questions?