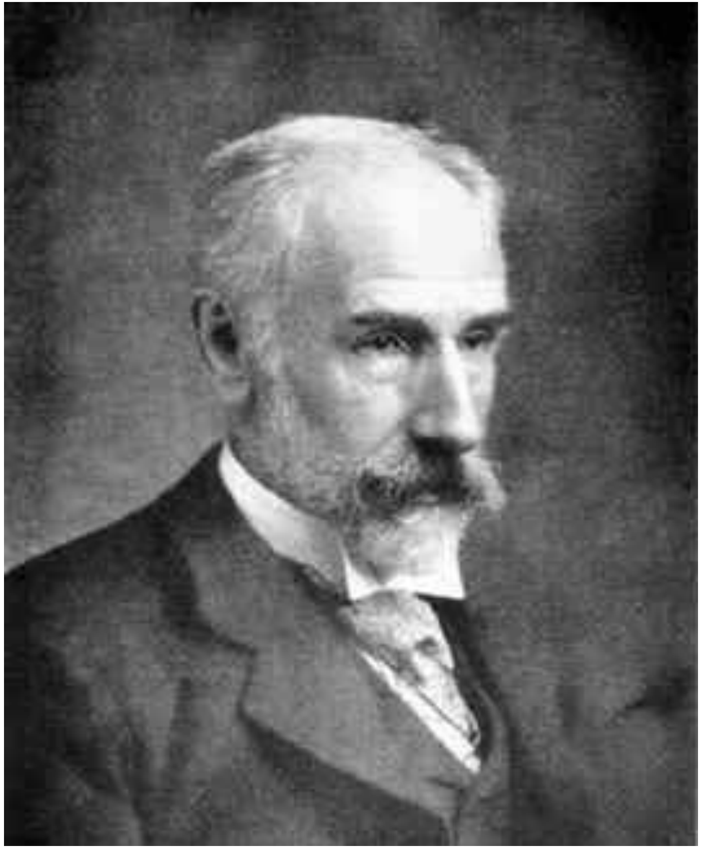
zalando

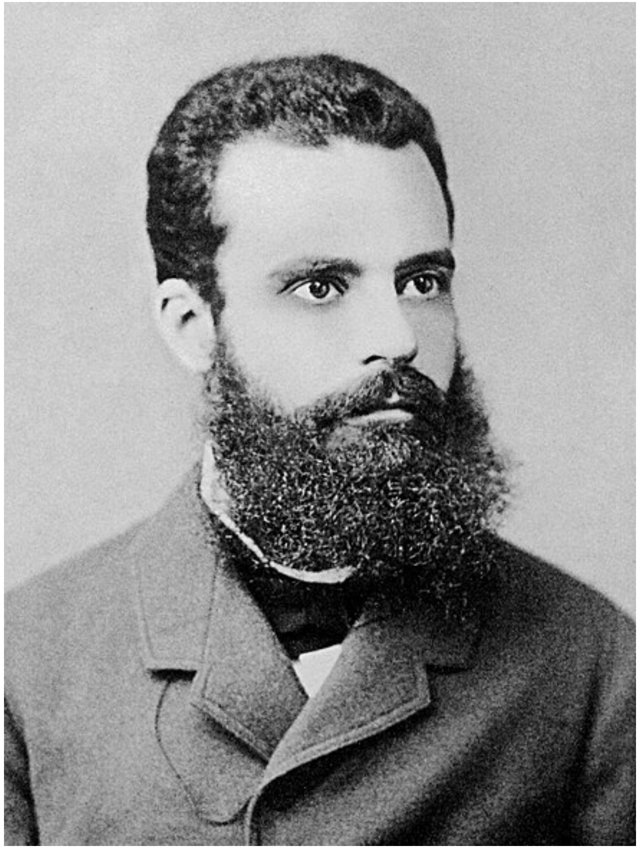
HOW TO KILL TWO BIRDS WITH ONE STONE: LEARNING TO RANK WITH MULTIPLE OBJECTIVES

HAYSTACK EU 2019

ALEXEY KURENNOY

28-10-2019





MULTI-OBJECTIVE OPTIMISATION IN EVERYDAY LIFE

- Product quality vs price
- Hotel location vs facilities
- Job satisfaction vs compensation
- ...

ZALANDO AT A GLANCE

~ 5.4 billion EUR
revenue 2018

> 300 million visits per month

> 400,000
product choices

> 27 million
active customers

> 15,500
employees in Europe

> 80%
of visits via mobile devices

~ 2,000
brands

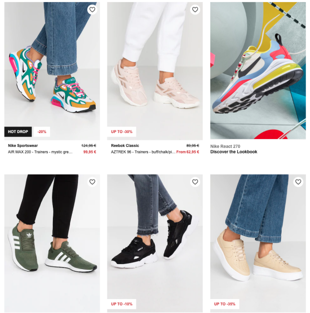
17
countries

zalando

OUTLINE

- Introduction
- Learning to Rank recap
- Multi-objective optimisation
- LambdaMART recap
- Multi-objective Learning to Rank with LambdaMART
- Experiments
- Conclusion and further steps

LTR RECAP



HISTORICAL DATA

MODEL TRAINING

OBJECTIVE FUNCTION

MODEL

RANKING

zalando

EXAMPLE: NDCG OBJECTIVE

Interaction	r_{ij}
purchase	3
click	1
no interaction	0

$$i = 1, \ldots, N,$$
$$j = 1, \ldots, n_i.$$

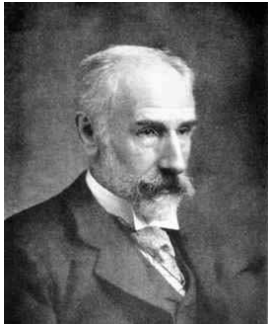
$$DCG_i = \sum_{j=1}^{n_i} \frac{r_{ij}}{\log_2(j+1)}$$

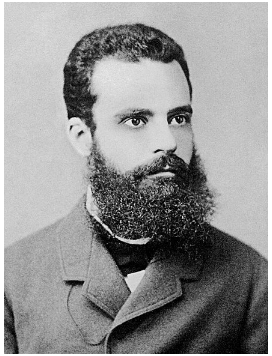
$$NDCG = \sum_{i=1}^{N} \frac{DCG_i}{\max DCG_i}$$

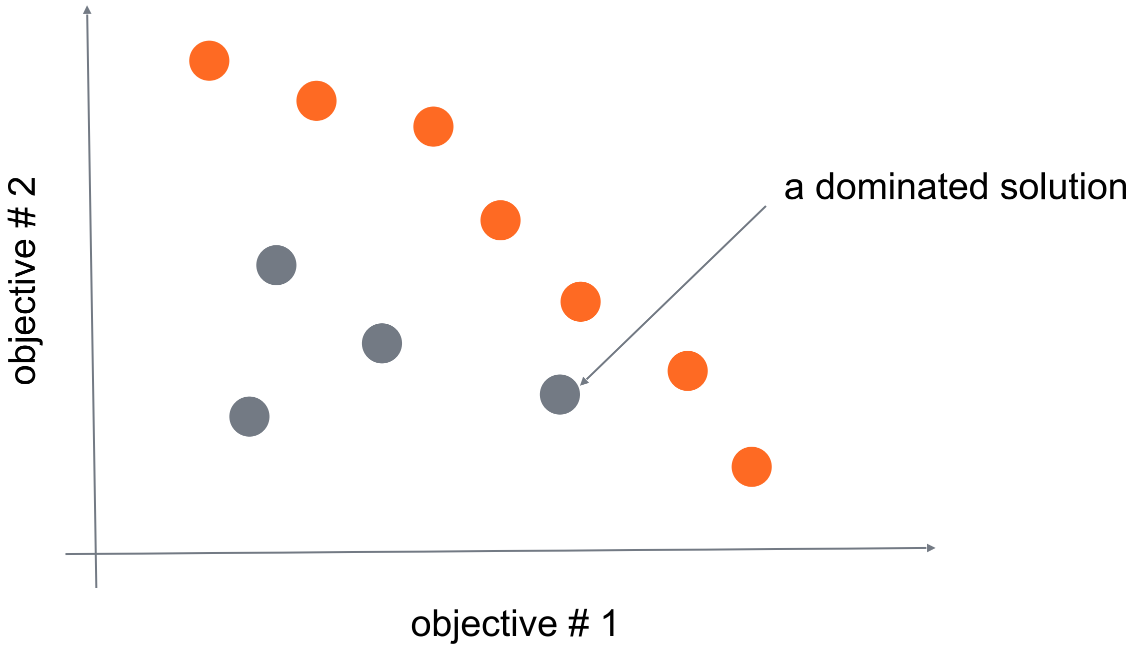
WHY A SINGLE OBJECTIVE MAY BE NOT ENOUGH?

- Different types of user feedback (implicit vs explicit)
- Different sources of feedback (user vs annotator)
- E-Commerce: engagement vs post-order experience
- E-Commerce: engagement vs fashionability

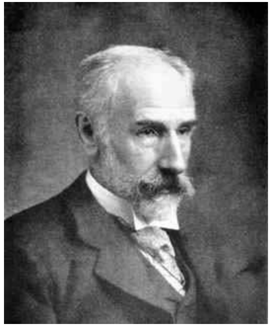
MULTI-OBJECTIVE OPTIMISATION

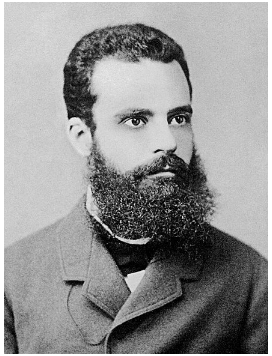


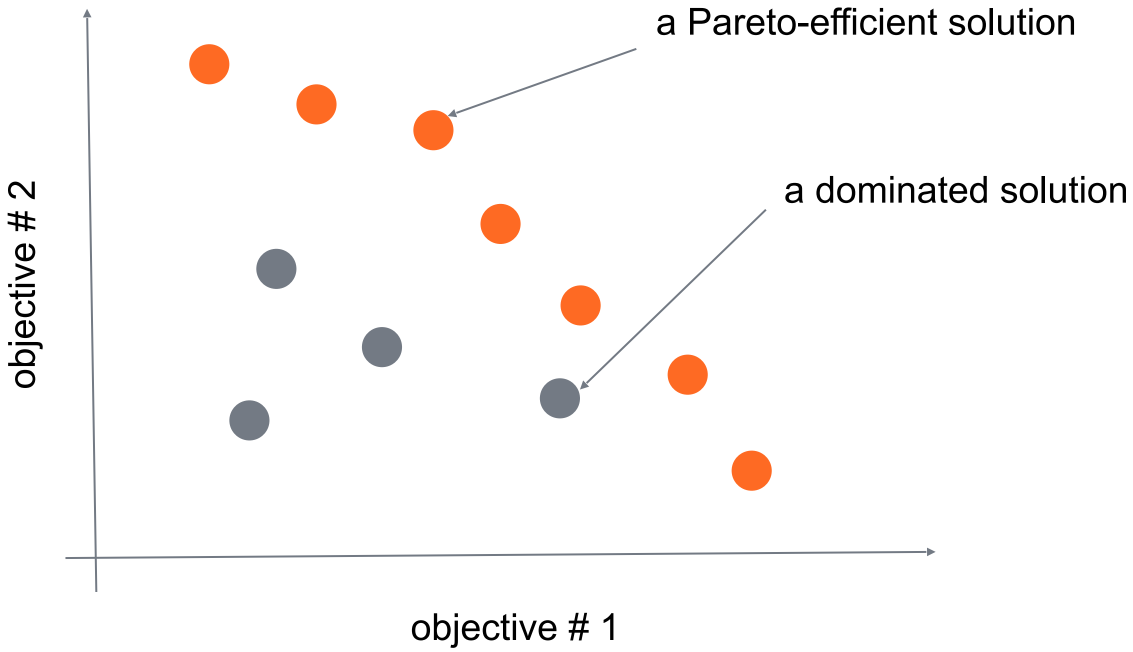




MULTI-OBJECTIVE OPTIMISATION

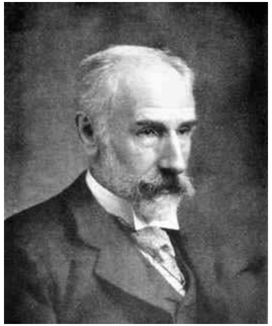


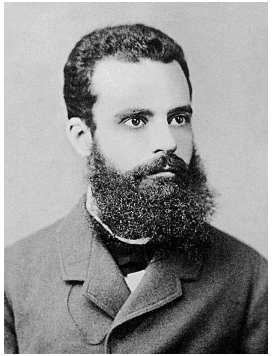


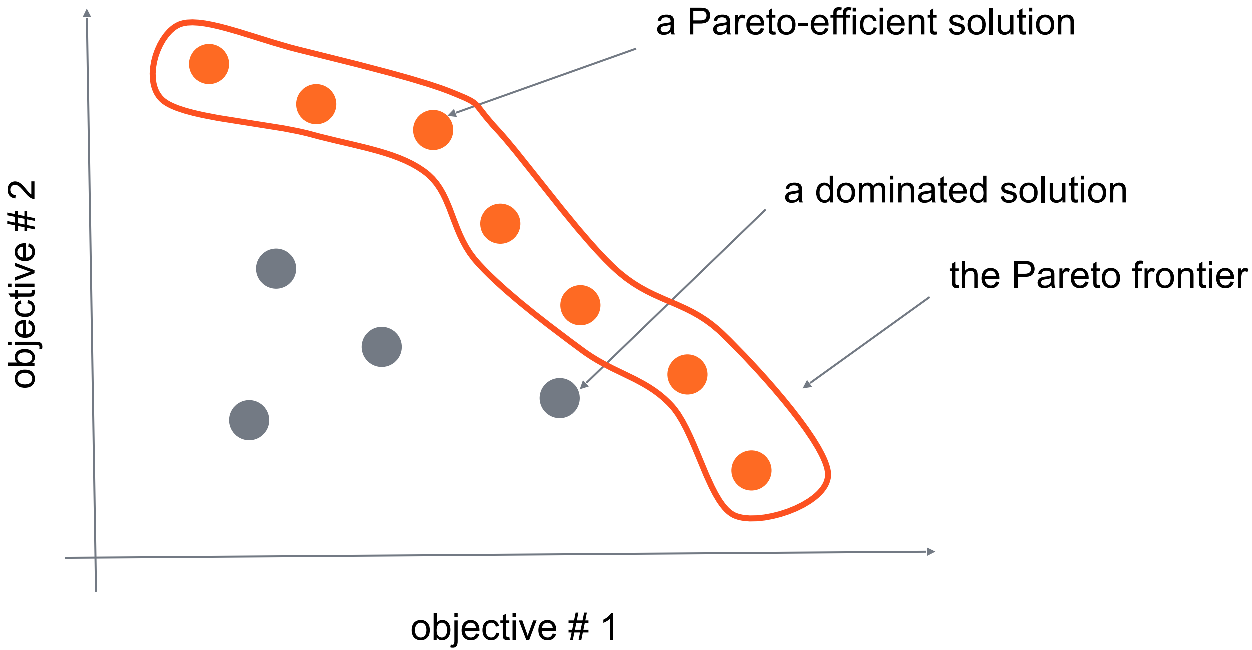


zalando

MULTI-OBJECTIVE OPTIMISATION







HOW TO SEARCH FOR PARETO SOLUTIONS?

$$r', \ r'', \ F_{r'}, \ F_{r''}$$

HOW TO SEARCH FOR PARETO SOLUTIONS?

$$r', \ r'', \ F_{r'}, \ F_{r''}$$

- Post-augmentation of the sorting rule

HOW TO SEARCH FOR PARETO SOLUTIONS?

$$r', \; r'', \; F_{r'}, \; F_{r''}$$

- Post-augmentation of the sorting rule
- Redefining the relevance

zalando

HOW TO SEARCH FOR PARETO SOLUTIONS?

$$r', \; r'', \; F_{r'}, \; F_{r''}$$

- Post-augmentation of the sorting rule
- Redefining the relevance

Interaction	r_{ij}
purchase, fashionable	7
purchase, non-fashionable	3
click	1
no interaction	0

$$i = 1, \ldots, N,$$
$$j = 1, \ldots, n_i.$$

HOW TO SEARCH FOR PARETO SOLUTIONS?

$$r', \; r'', \; F_{r'}, \; F_{r''}$$

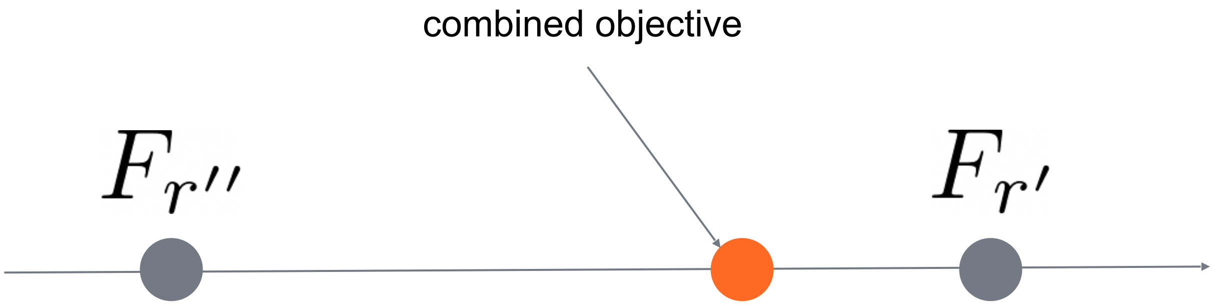
- Post-augmentation of the sorting rule
- Redefining the relevance
- Scalarization

$$\alpha F_{r'} + (1 - \alpha) F_{r''} \to \max$$
$$\alpha \in (0, 1)$$

zalando

SCALARIZATION

$$\alpha F_{r'} + (1 - \alpha) F_{r''} \rightarrow \max$$
$$\alpha \in (0,\, 1)$$

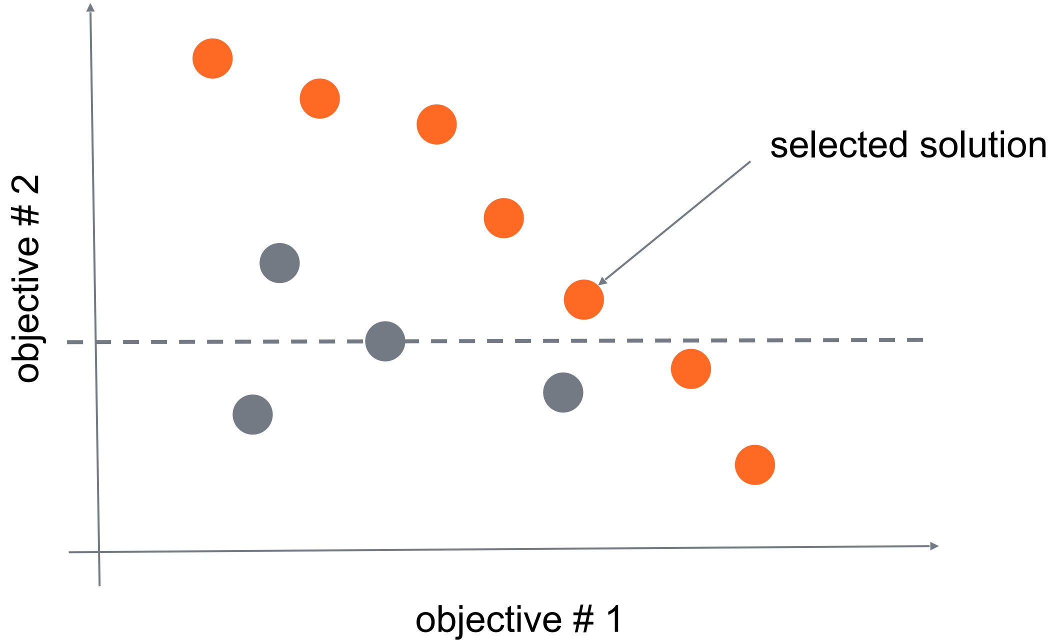


SCALARIZATION

$$\alpha F_{r'} + (1 - \alpha) F_{r''} \to \max$$
$$\alpha \in (0,\, 1)$$

- Produces only Pareto optimal solutions
- Allows finding any Pareto optimal solution (under certain assumptions)
- Straightforward to apply when there are more than two objectives

HOW TO CHOOSE AMONG PARETO SOLUTIONS?



HOW TO SEARCH FOR PARETO SOLUTIONS?

$$r', \ r'', \ F_{r'}, \ F_{r''}$$

- Post-augmentation of the sorting rule
- Redefining the relevance
- Scalarization
- Constrained optimisation
  (Momma et. al. Multi-objective Relevance Ranking. SIGIR 2019)

$$F_{r'} \to \max \ \ s.t. \ F_{r''} \geq c$$

zalando

LAMBDAMART RECAP

- A boosting algorithm for optimising ranking metrics (such as NDCG)
- Iteration:
  - compute the lambda-gradient

$$\Delta s_{ij}(r)$$

  - construct a tree that approximates the lambda-gradient and add it to the ensemble

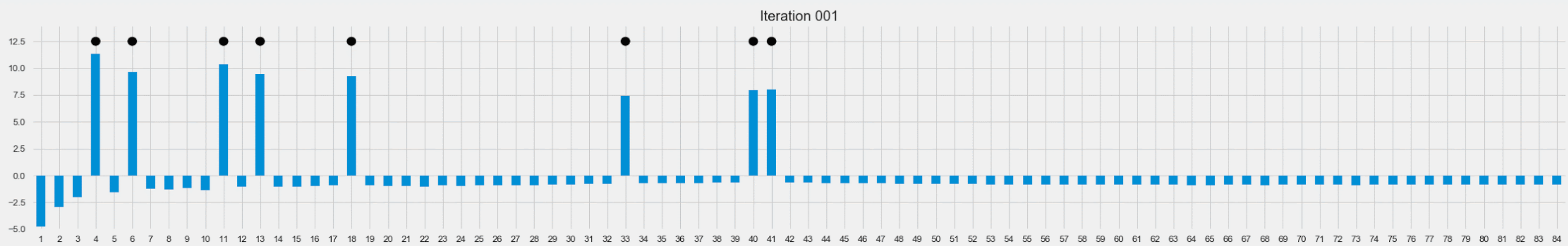
zalando

LAMBDAMART RECAP

- A boosting algorithm for optimising ranking metrics (such as NDCG)
- Iteration:
  - compute the lambda-gradient

$$\Delta s_{ij}(r)$$

  - construct a tree that approximates the lambda-gradient and add it to the ensemble



MULTI-OBJECTIVE OPTIMISATION WITH LAMBDAMART

$$\Delta s_{ij}(r)$$

$$r',\ r''$$

$$\Delta s_{ij} = \alpha \Delta s_{ij}(r') + (1-\alpha)\Delta s_{ij}(r'')$$

IMPLEMENTATION

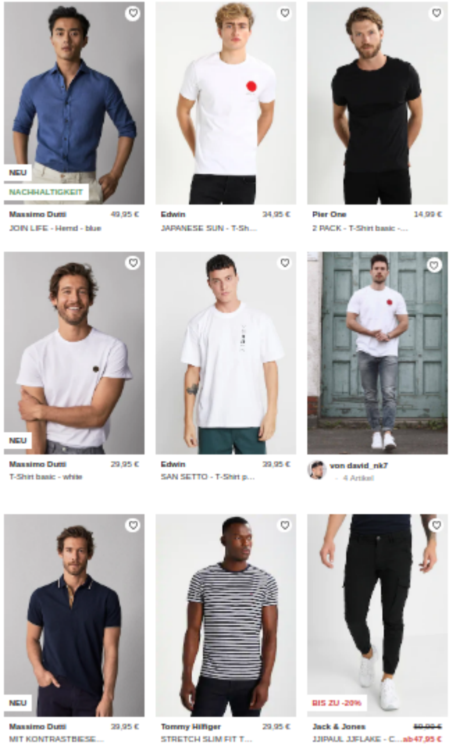
```
lightgbm.train(..., fobj=...)
```

- Used Cython
- Parallelisation by means of OpenMP
  ~5x speed-up (Dual-Core Intel Core i5)
- The speed is on par with the original implementation

```
cimport cython
cimport openmp
from cython.parallel import prange

...

cdef int cnt
cnt = <int>(end - start)

...

cdef int i
cdef int j
for high in range(cnt):
    label_high = labels[start + high]
    score_high = preds[start + high]
    for low in range(cnt):
        label_low = labels[start + low]
        score_low = preds[start + low]
        if label_high > label_low:
            sigmoid = get_sigmoid(score_high - score_low, ...)
            abs_delta_ndcg = get_abs_delta_ndcg(...)

            grad[start + high] += -abs_delta_ndcg * sigmoid
            grad[start + low] -= p_lambda
```


zalando

OFFLINE EXPERIMENTS: SETUP

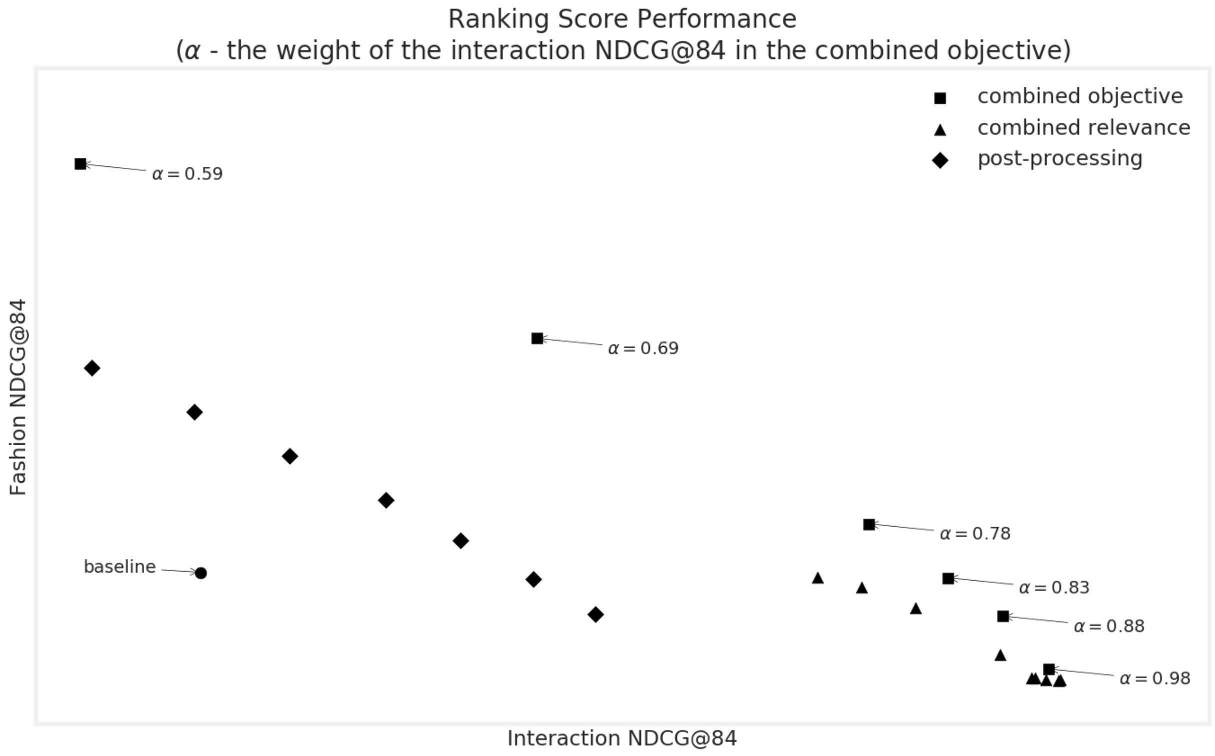


- Result pages from Zalando catalog (~50K pages)
- Two objectives:
  - Interaction NDCG

$$r_{ij} = \begin{cases} 3 & if\ purchased_{ij}, \\ 1 & if\ clicked_{ij}, \\ 0 & otherwise. \end{cases} \qquad \begin{array}{l} i = 1, \dots, N, \\ j = 1, \dots, n_i. \end{array}$$

  - Fashionability NDCG (relevance = fashion score)
- Goal: up-sort popular *but fashionable* articles

OFFLINE EXPERIMENTATION: RESULTS



zalando

OFFLINE EXPERIMENTATION: RESULTS

- Visibility of popular but non-fashionable articles went UP
- Visibility of fashionable but unpopular articles went UP
- Visibility of fairly popular and relatively fashionable articles went DOWN

CONCLUSIONS

- A Pareto-efficient solution in ranking can lead to a “mixed” effect
- Evaluation in the “multi-objective” space may not be reliable

FUTURE STEPS

- Experimenting with the “combined” relevance
- Experimenting with dataset augmentation

P.S.

No birds were killed during the preparation of this talk.