YetiRank: Everybody Lies

Andrey Gulin and Igor Kuralenok
Yandex (company.yandex.ru)
{gulin,solar}@yandex-team.ru

YR: Agenda

- Idea and MatrixNet boosting
- Pair weighting
- Transfer-learning?
- Results

YR: Idea

- Boosting based on decision trees
- We want to optimize on pairs
- Pair could be in right or wrong order
- Pair weight depends on:
  - Importance
  - Confidence in order

YR: Target Function

Target function:

$$\sum_{(i,j)} w_{ij} \log \frac{e^{x_i}}{e^{x_i} + e^{x_j}}, \quad (1)$$

Divergence optimization in terms of $d(x_i - x_j)$:

$$\underset{dx_i, dx_j}{\operatorname{argmin}} \sum_{(i,j)} w_{ij} \left(dx_i - dx_j - \frac{e^{x_j}}{e^{x_i} + e^{x_j}} \right)^2 \quad (2)$$

$$\underset{x}{\operatorname{argmin}} ||Ax - b||^2$$

YR: MatrixNet boosting

We use oblivious decision trees as weak learner.

$$R^n = \{C_i\}_{i=1}^{2^h}$$

Substitute quadrant values instead of x:

$$x = Cy$$

$$B = AC$$

$$y = \left(B^T B\right)^{-1} B^T b$$

Я

YR: Pair weight

$$w_{ij} = d_{ij} c(r_i, r_j)$$

where d_{ij} answers the question

"is this pair important for ranking?"

and $c(r_i, r_j)$

"are we sure in the ordering?"

YR: Pair importance

- Introduce variance in scores for each query

$$x'_i = x_i \frac{\xi}{1-\xi}$$

- Sample ordering n times

$$d_{ij} = \frac{1}{n} \sum_{t=1}^{n} \frac{1}{index_t((i,j))}$$

- What to do with inverted pairs?

YR: Order confidence

Motivation:

- Experts can be wrong
- Mistakes are not uniformly distributed among relevances
- This distribution looks similar for different experts

Model:

- Introduce transition matrix $p(J|r_i)$
- Measure confidence as

$$c(r_i, r_j) = \sum_u \sum_v sign(u - v) p(J_u|r_i) p(J_v|r_j)$$

- No access to experts → emulate with multi-classifier

YR: Transfer learning?

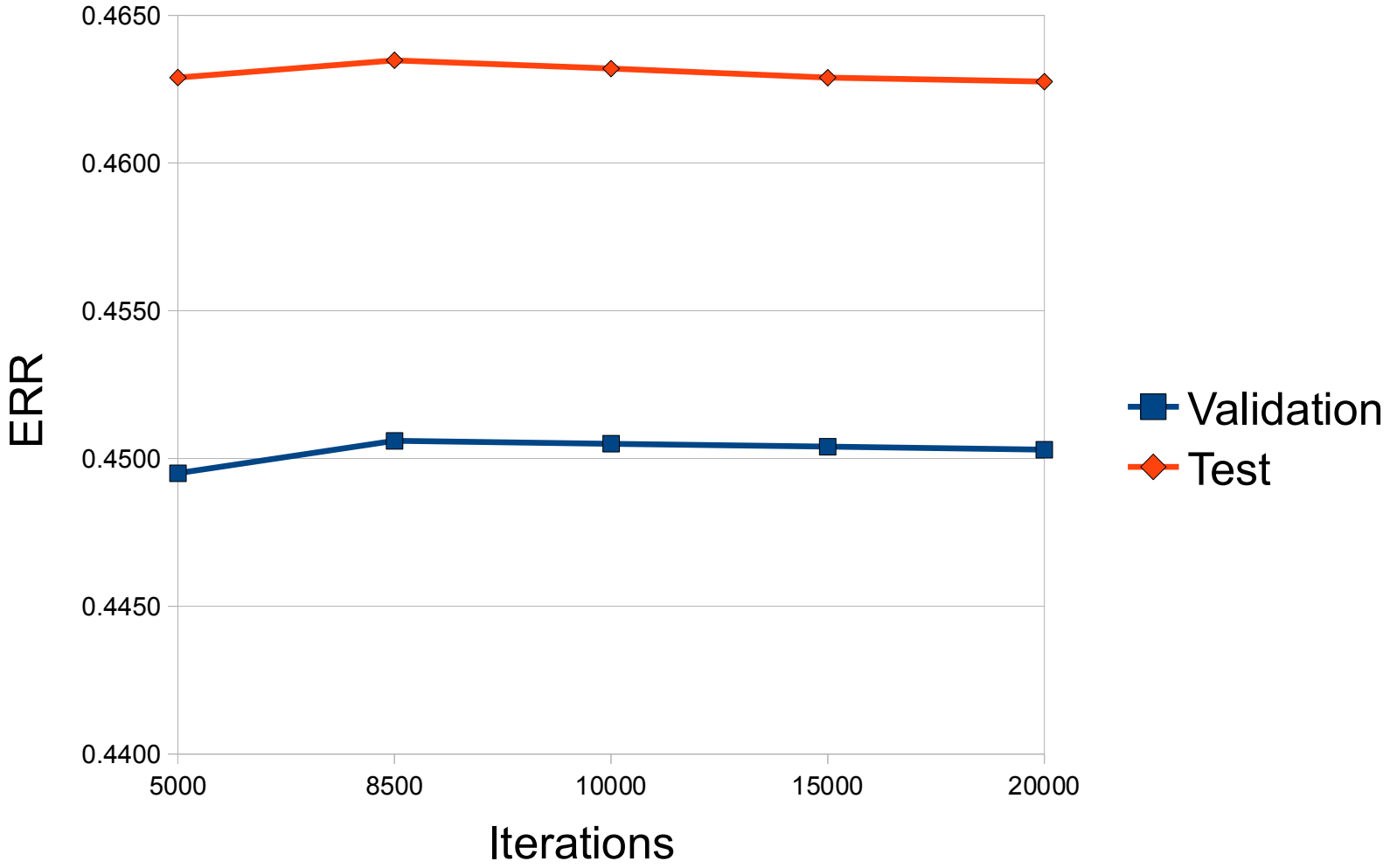
Set2 TM emulation

	bad	fair	good	excellent	perfect
bad	0.48	0.40	0.10	0.02	0
fair	0.08	0.68	0.21	0.03	0
good	0.01	0.44	0.47	0.08	0
excellent	0	0.07	0.54	0.38	0.01
perfect	0	0.01	0.09	0.43	0.48

Set1 TM emulation

	bad	fair	good	excellent	perfect
bad	0.35	0.59	0.06	0	0
fair	0.07	0.69	0.24	0	0
good	0.02	0.48	0.48	0.02	0
excellent	0.01	0.28	0.6	0.1	0.01
perfect	0	0.07	0.31	0.37	0.24

YR: Results


0.4650
0.4600
0.4550
0.4500
0.4450
0.4400
ERR
5000
8500
10000
15000
20000
Iterations
Validation
Test

YR: Conclusions

- Experts can be wrong and modeling this helps ranking
- Marks are the same != judgments are the same
- Judgments “transfer” may be a key point to transfer-learning in ranking