

Learning to Rank

Overview

Introduction of machine learning algorithms
for ranking problem

Outline

- Introduction
- Ranking Methods
 - Supervised Methods
 - Numeric Methods
- Practical Issues



I want top php engineers



I want top php engineers



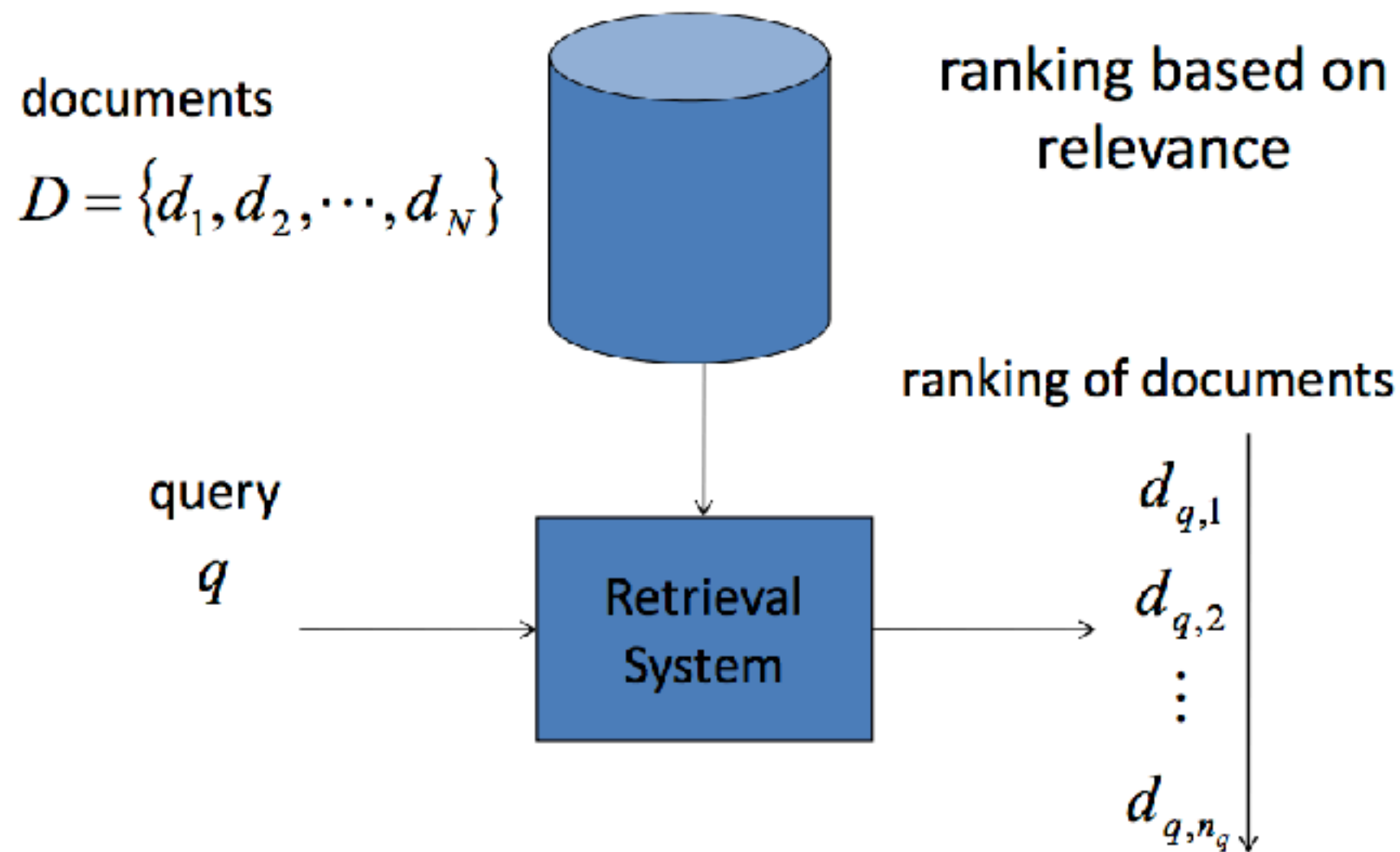
How to decide an order for these developers?



Ranking Everywhere

- Article ranking in community search page
- Article Infinite Scroll
 - Once the article opened, how to decide next 10 articles we recommend
- Applicant Quality Score
 - How to select top quality applicants to review first
- Codementor Request Matching
- ...

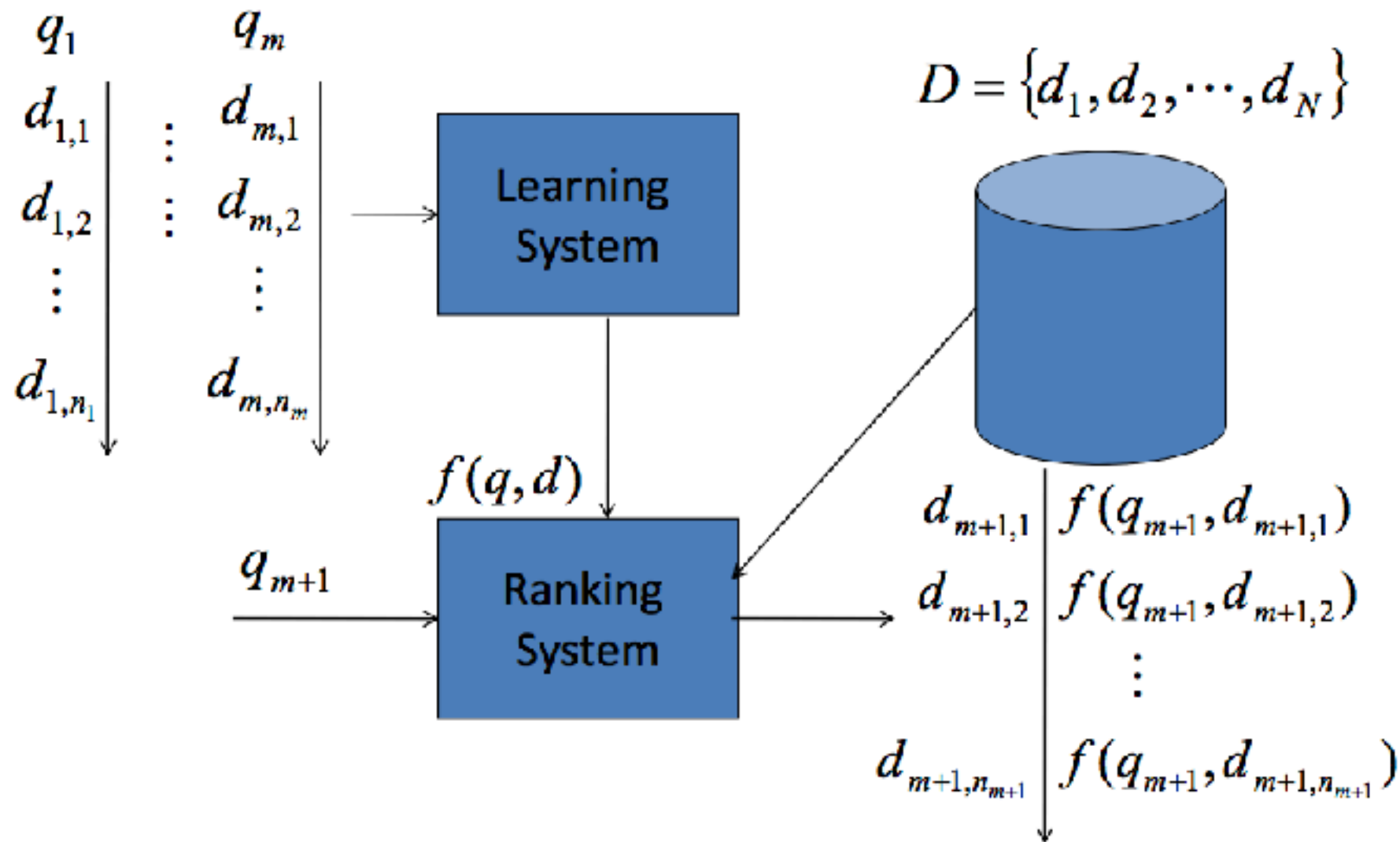
Native Ranking System



Native Ranking Methods

- The ranking expression is human experience
 - Need many time to tuning models
 - Hard to scale out if number of features increases
 - Overfitting problem
- Is it possible to use machine learning to solve ranking problem?

Learning to Rank System



Supervised Methods

Methods

Pointwise Approach

1



5



2



2



3



4

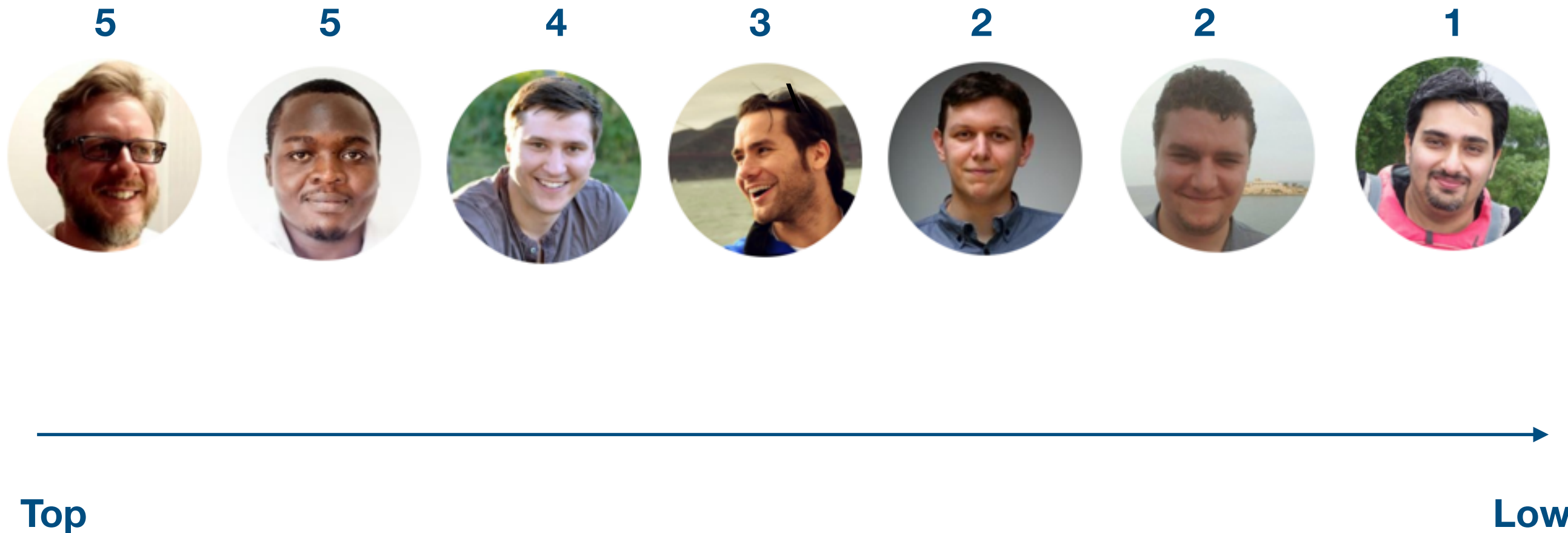


5



Methods

Pointwise Approach



Methods

Pointwise Approach

- Model Complexity: $O(|q| * n_q)$
 - $|q|$: number of queries
 - n_q : number of matched documents of query q
- Algorithms
 - SVM, McRank, SLR ...

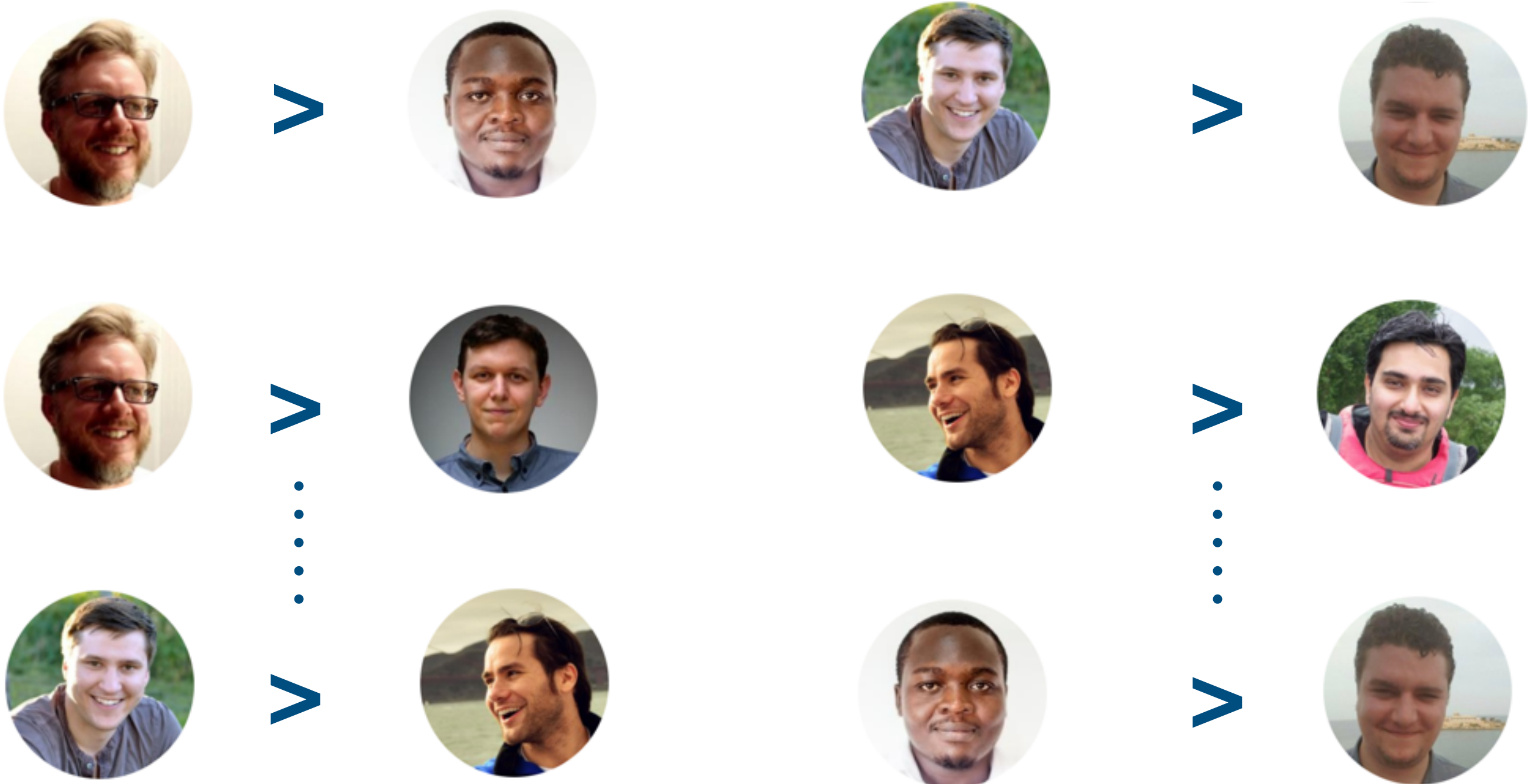
Methods

Pointwise Approach

- Pros
 - Easy and intuitive to implement
 - Fast modeling
- Cons
 - Cannot consider ordering
 - Labeling is hard

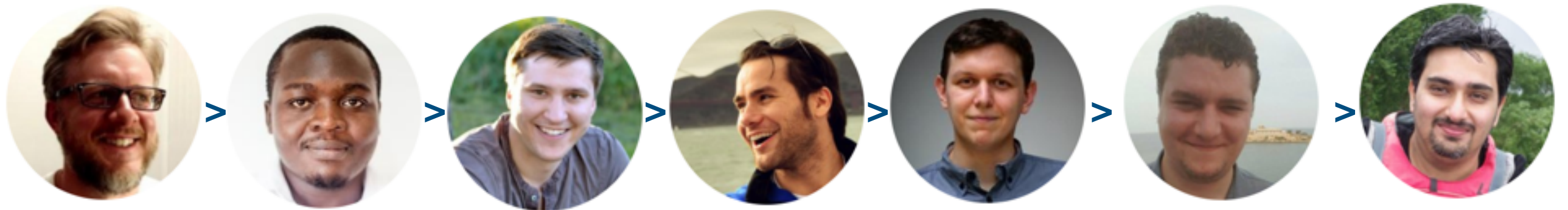
Methods

Pairwise Approach



Methods

Pairwise Approach



Top

Low

Methods

Pairwise Approach

- Model Complexity: $O(|q| * n^2)$
- Algorithms
 - RankNet, IR-SVM, RankBoost

Methods

Pairwise Approach

- Pros
 - Consider relative order
 - Most popular concept in real world
- Cons
 - Need to generate $O(n^2)$ tags, high cost
 - Not consider the relevance between query and document

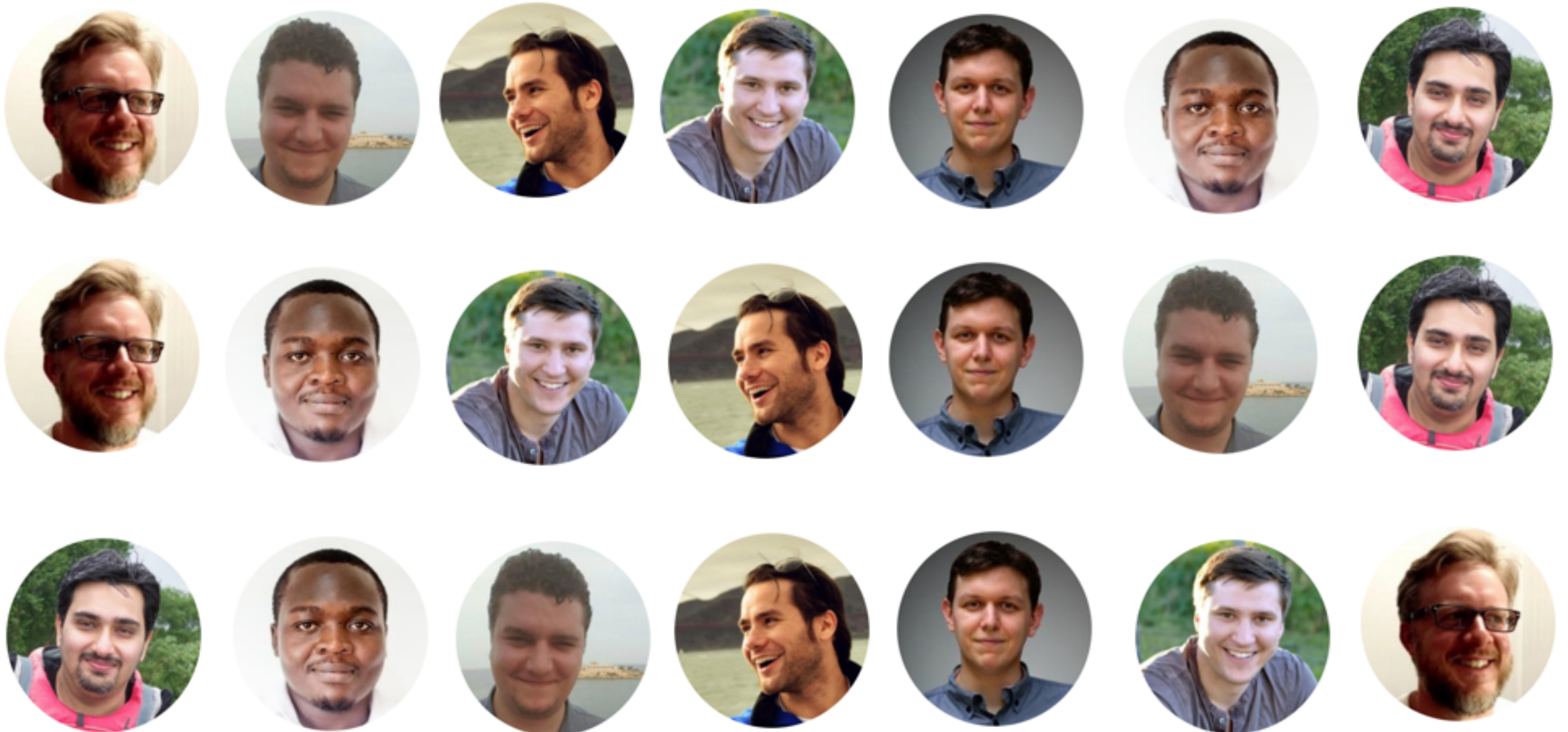
Numerical Methods

Methods

Listwise Approach

Top

Low

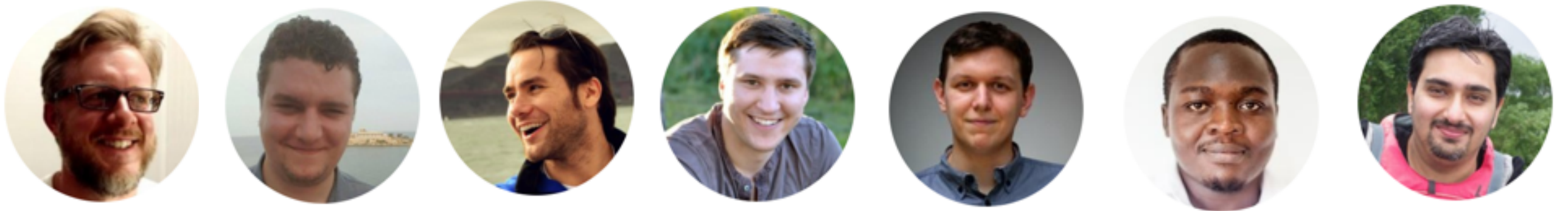


Methods

Listwise Approach

Top

Low



Which one is better ordering?



Methods

Listwise Approach

Best Order



**Measure the difference with
best orders**



One of the orders

Methods

Listwise Approach

- Model Complexity: $O(|q| * n!)$
 - Many researches focus on reducing model complexity by numerical methods
- Algorithms
 - AdaRank, ListNet, SVM-MAP ...

Methods

Listwise Approach

- Pros
 - Consider all situations
 - Based on the origin problem definition
- Cons
 - $O(n!)$ permutations (hard to implement and time bound)
 - Hard to optimize (objective functions not always continuous)

Practical Issues

- What is the objective metrics we pursue?
- How to generate labels?
- Cold-start problem

Objective Metrics

- Based on application
- Search
 - Most page views
 - Click Through Rate (CTR)
- CMX Job match
 - Contract Conversion Rate (CCR)
 - Interview Conversion Rate (ICR)

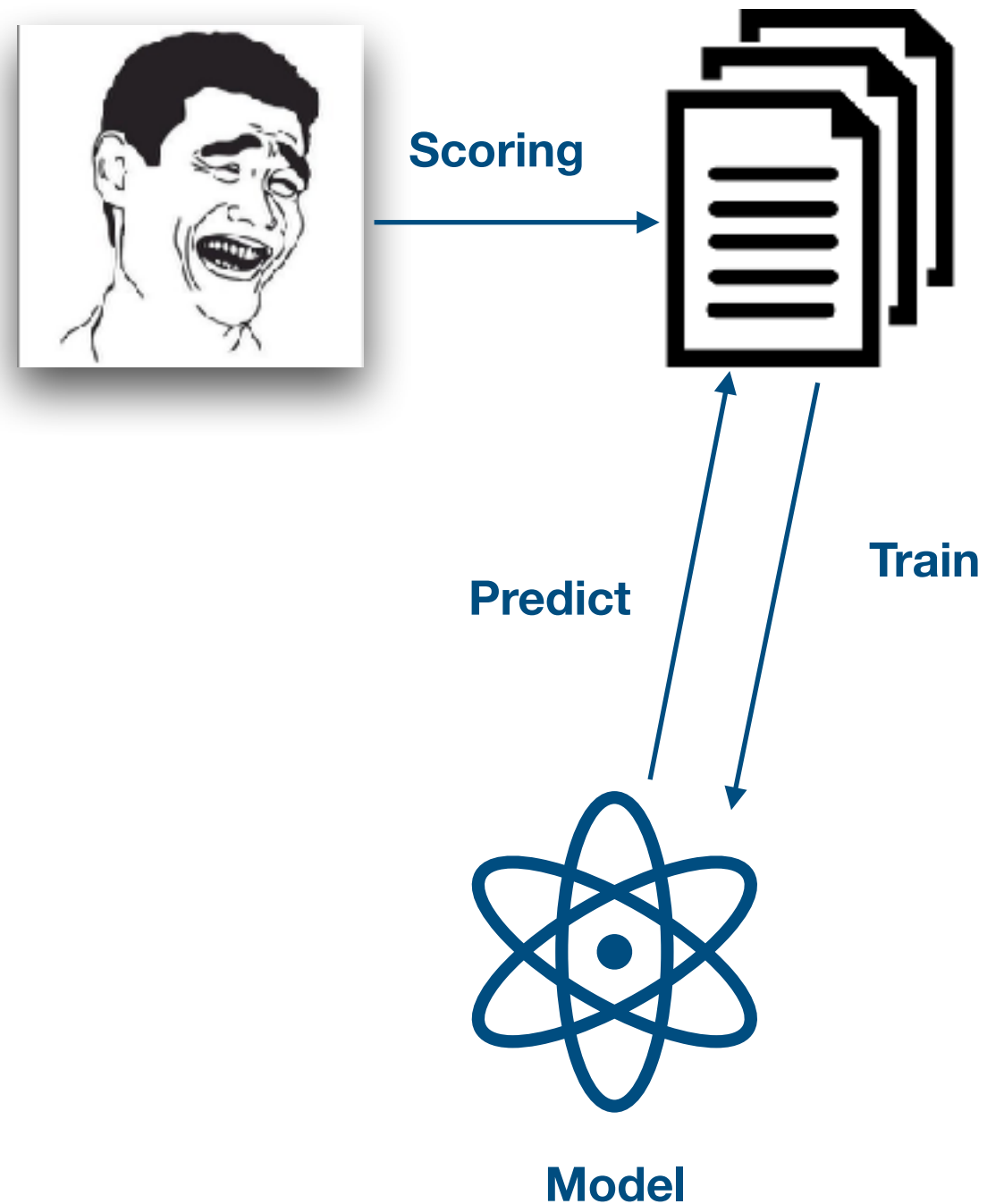
Labeling Problem

- How to label data if we have nothing
 - Small dataset - human labeling is ok
 - Large dataset
 - Semi-supervised learning
 - Matrix factorization

Labeling Problem Semi-supervised

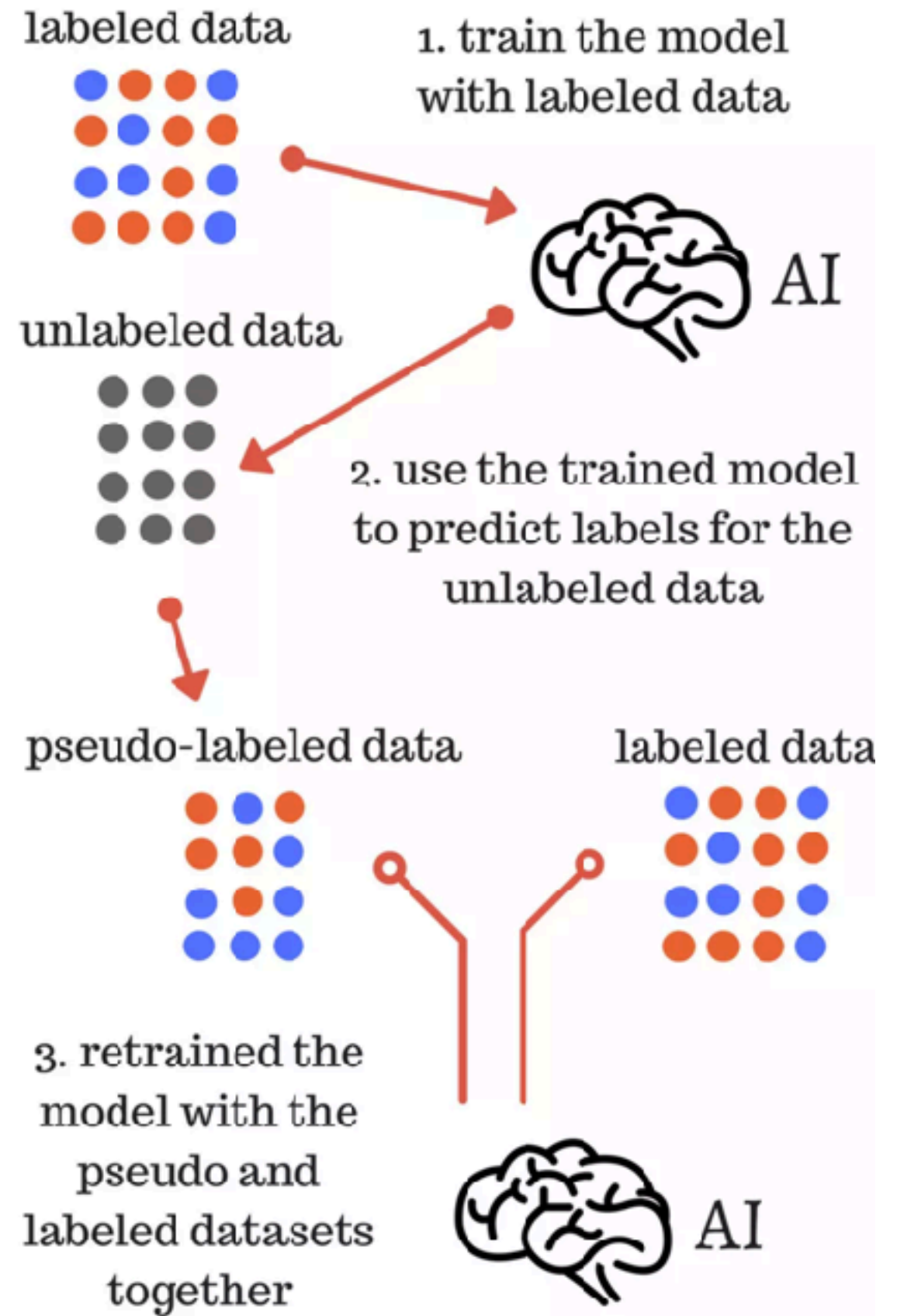
Iterative training with
relevance Labels

- Rule-based models
- Human ranking expressions



Labeling Problem Semi-supervised

Pseudo Labeling



Labeling Problem - Matrix Factorization

- When data size is large, we can label part of pairs

	P1	P2	P3	P4	P5
Php	1	2	5	?	5
React	4	?	1	3	1
Javascript	3	5	?	1	?
Lavarel	?	1	4	?	5

Cold-start Problem

- Occurs in using some online metrics as features
 - Click count
 - View count
- How to handle new items?
 - Find similar items based on static features

Reference

- Ranking

- <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.222.6261&rep=rep1&type=pdf>
- <http://www.shuang0420.com/2016/10/25/Search%20Engines%E7%AC%94%E8%AE%B0%20-%20Learning%20to%20Rank/>
- <http://net.pku.edu.cn/~wbia/2011/slides/LEARNING%20TO%20RANK%20TUTORIAL%20-%20tyliu.pdf>
- https://people.cs.umass.edu/~jpjiang/cs646/16_learning_to_rank.pdf

- Semi-supervised learning techniques

- <https://www.analyticsvidhya.com/blog/2017/09/pseudo-labelling-semi-supervised-learning-technique/>
- <https://datawhatnow.com/pseudo-labeling-semi-supervised-learning/>