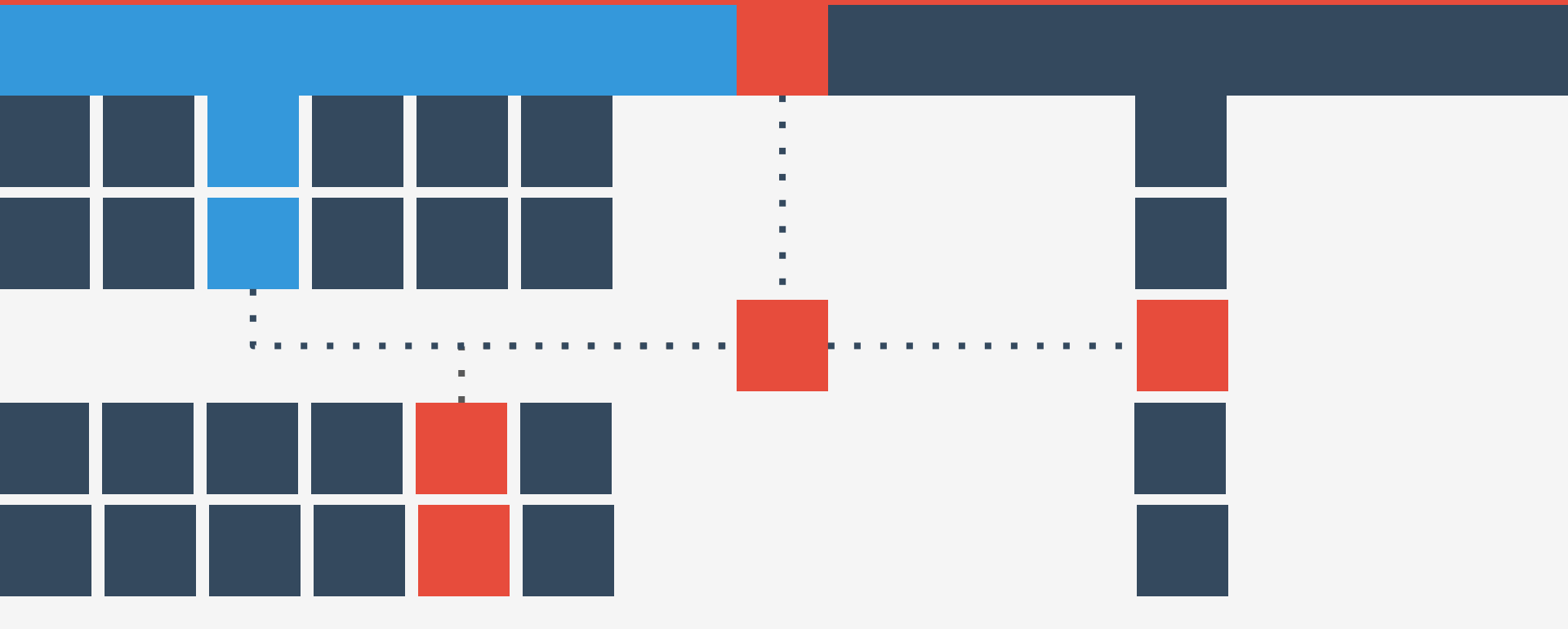
# RecSys 2017 Online Ranking Tutorial

Róbert Pálovics Domokos Kelen
Dániel Berecz András A. Benczúr

Informatics Laboratory of the Hungarian Academy of Sciences



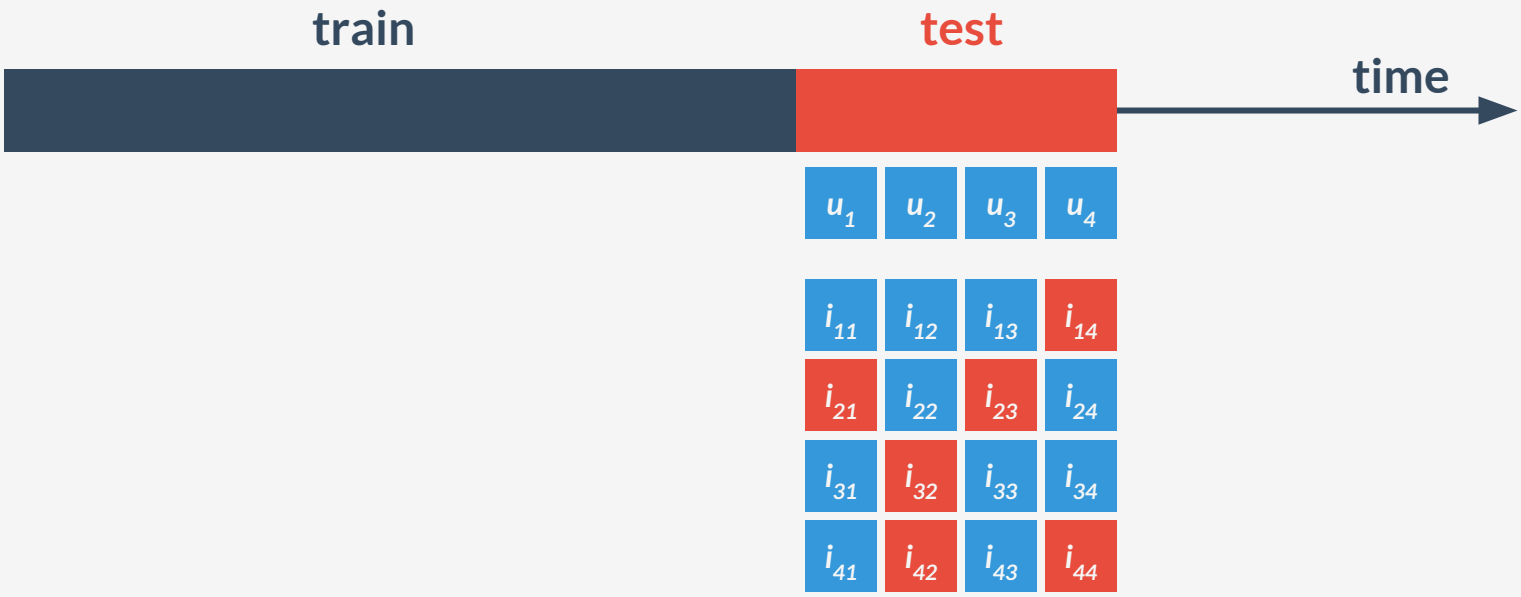
# Objectives

*Examples: hashtag recommendation on Twitter, news recommendation, music recommendation*

- **Implicit** recommendation
- **Top-*k*** recommendation
- **Time-aware models** and **time-aware evaluation**

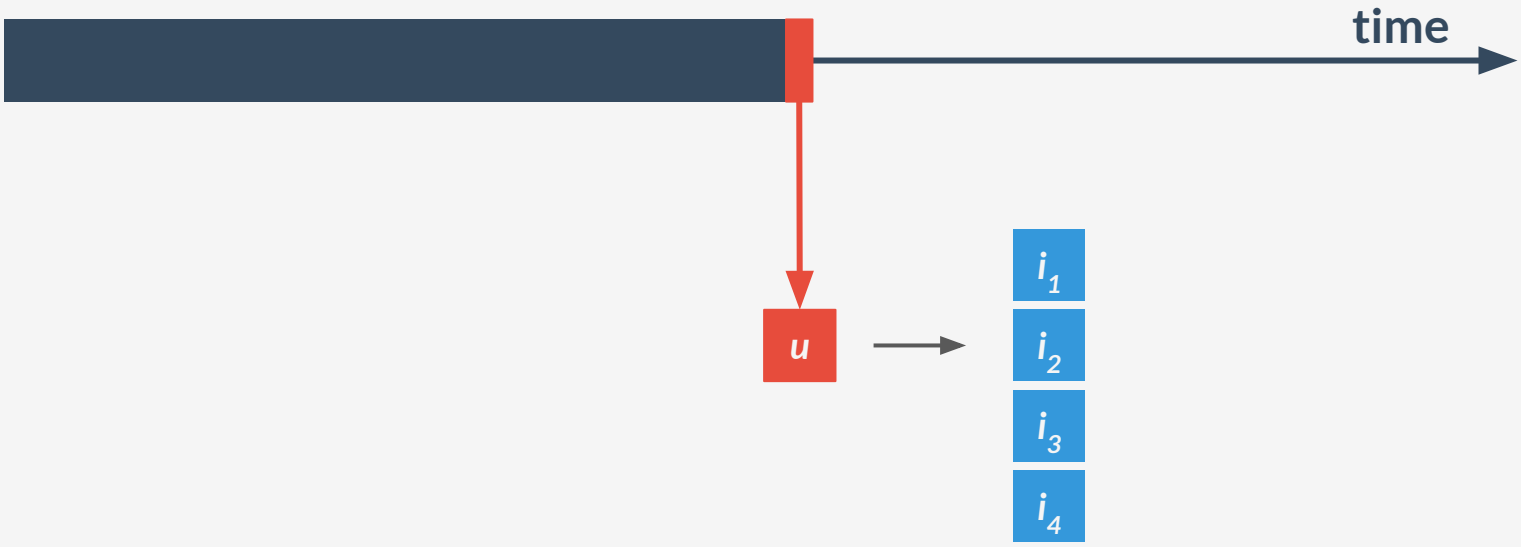
# Batch [time-aware] Top-K Recommendation Task

- Learn from the past batch training set
- Recommend for each user in the test a top-*k* list
- Evaluate based on the test set



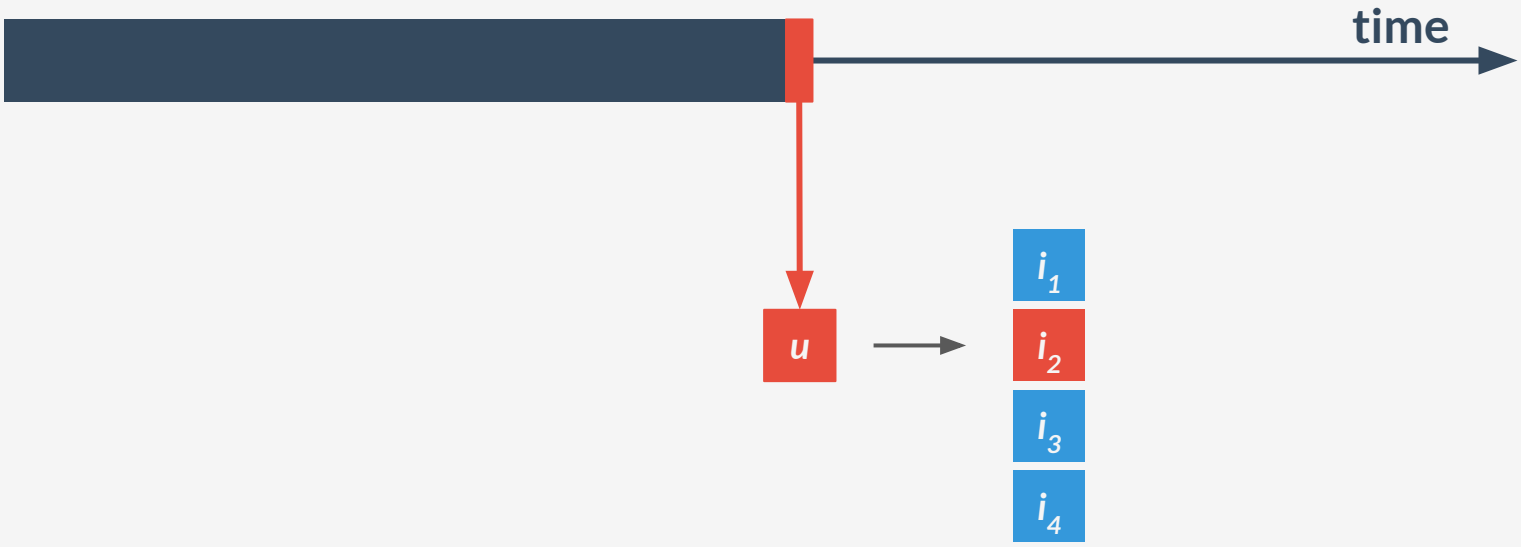
# Online [time-aware] Top-K Recommendation Task

- We use timestamped implicit data
- Process the events in the data in temporal order
- After each event (*u*,*i*,*t*)
  - recommend a new top list of items
  - then update the recommender model



# Online [time-aware] Evaluation

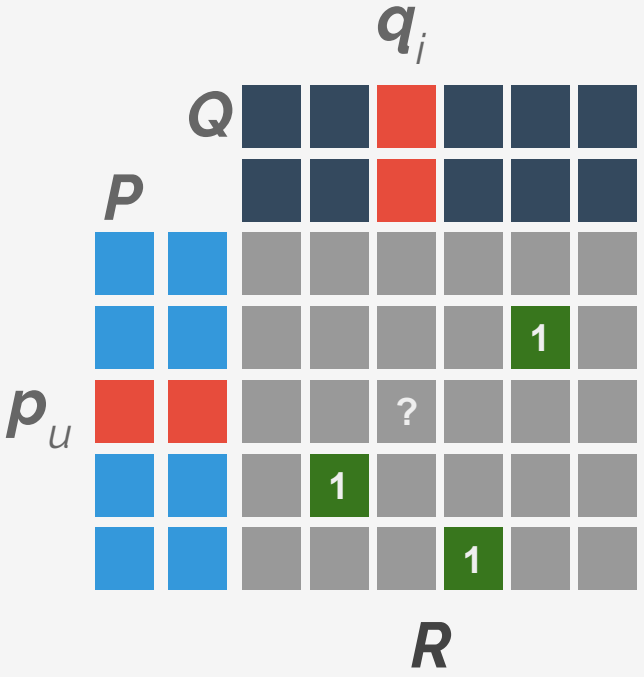
- Evaluate the given single tuple (*u*,*i*,*t*) in question against the recommended top list
- DCG = 1 / ($\log_2$( rank( ***i*** ) + 1)
- Compute timely averages, e.g. daily average DCG



# Matrix Factorization

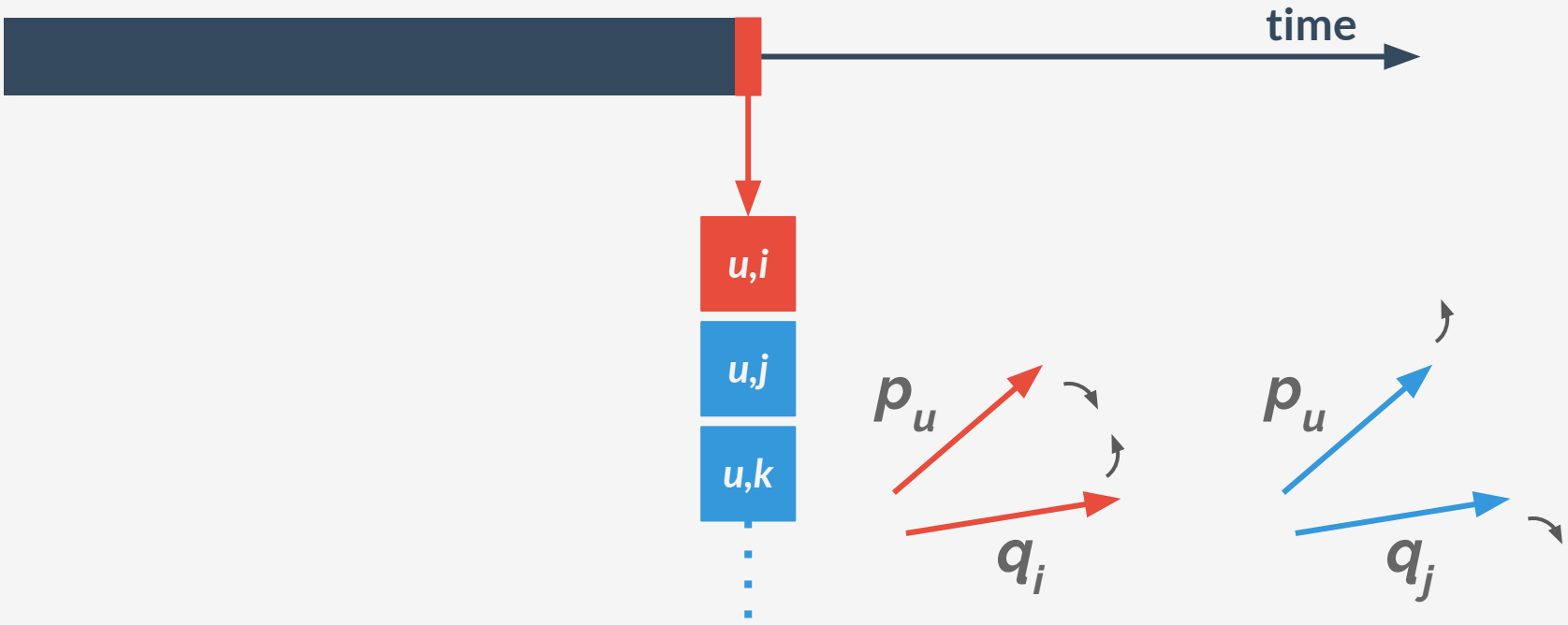
- Data
  - sparse matrix $\boldsymbol{R}$
  - $r(u, i) = 1$, if $u$ interacted with $i$
- Model
  - $\boldsymbol{P}$ and $\boldsymbol{Q}$ matrices for the users and the items respectively
  - "probability" of an interaction

    $r(u, i) = \boldsymbol{p}_u \boldsymbol{q}_i$
- Learning
  - objective: MSE
  - optimization: iALS or SGD



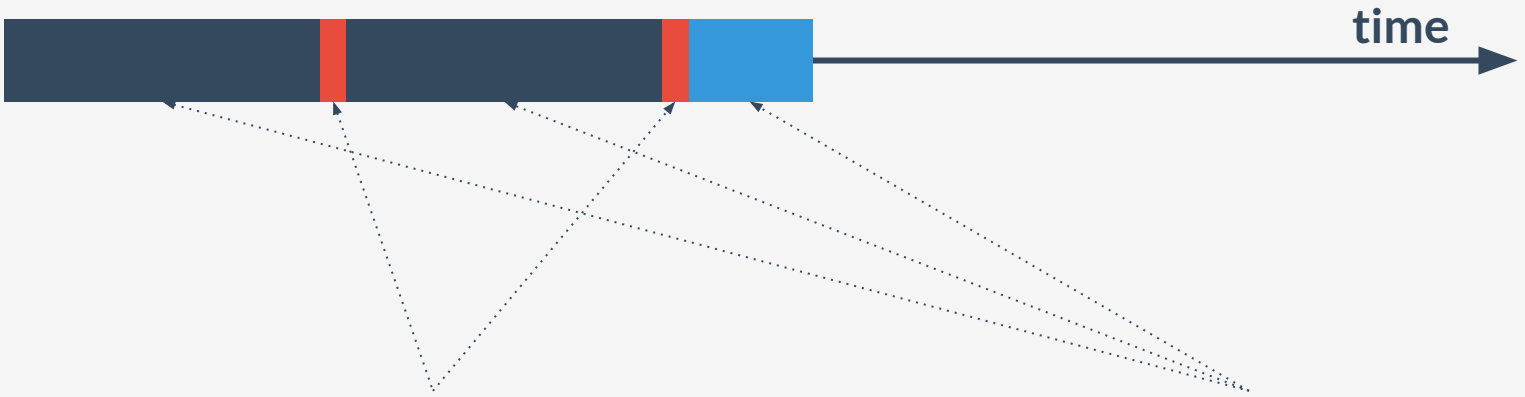
# Online Matrix Factorization on Implicit Data

- Single iteration over the training data
- We process the events in temporal order
- Optimize for MSE with SGD
- Generate random negative samples for the given user



# Batch then Online Matrix Factorization

- Periodically re-learn the batch model
- Between two batch model building, continue the learning of the previous batch model via online matrix factorization



# Recommender frameworks

## Alpenglow

free and open source C++ based framework with Python API for conjoint batch and online learning

## Flink

open-source stream processing framework with batch and streaming API
MF algorithms are available only as pull requests

## SparkML


provides batch iALS baseline

# Outline

- Installation
- Batch time-aware top-k evaluation of batch models

**Alpenglow - Flink - Spark**

- Online time-aware top-k evaluation of both batch and online models

**Alpenglow - Flink**