

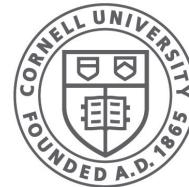
Exploring Recommendations Under User-controlled Data Filtering

Hongyi Wen, Longqi Yang, Michael Sobolev, Deborah Estrin

CORNELL
TECH
HOME OF THE
JACOBS
INSTITUTE



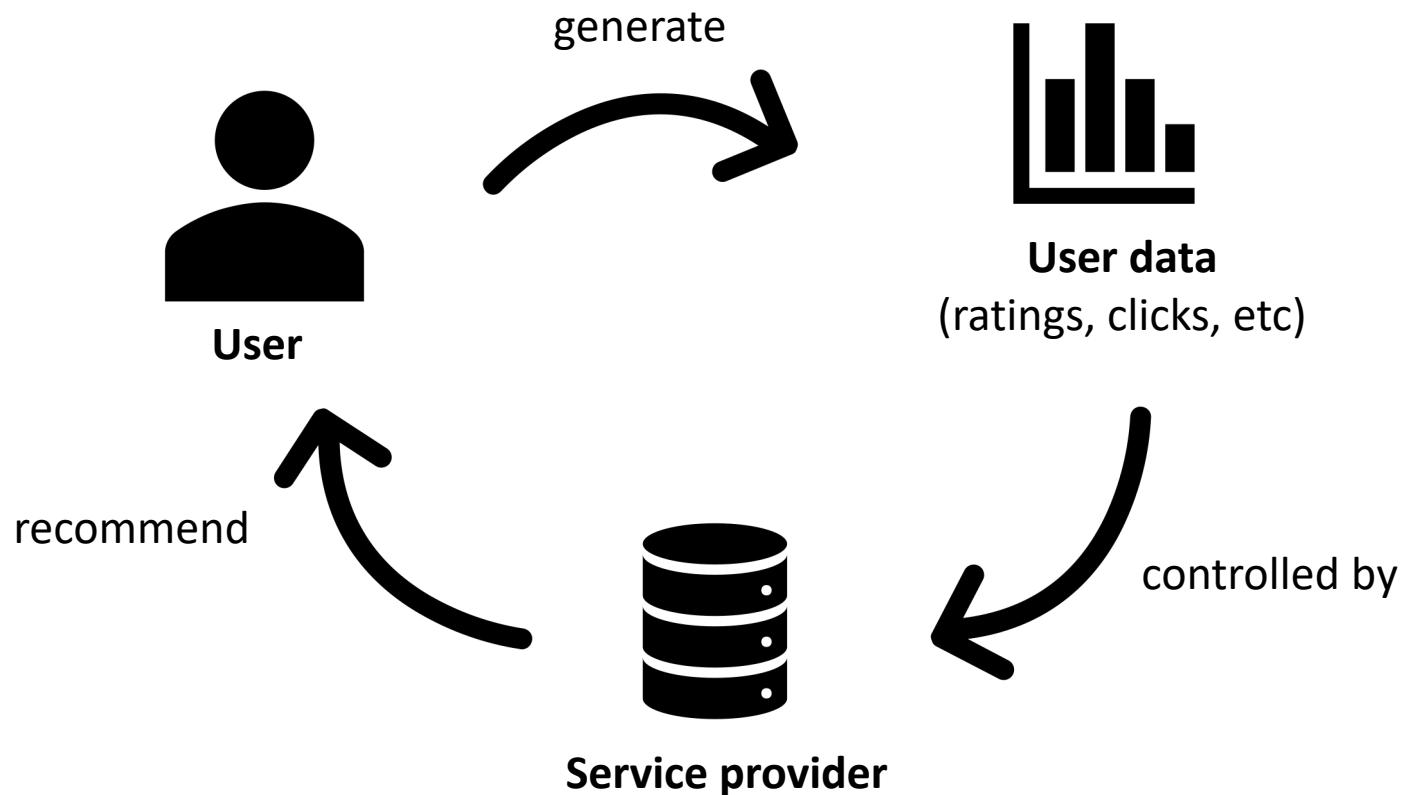
CORNELL
TECH



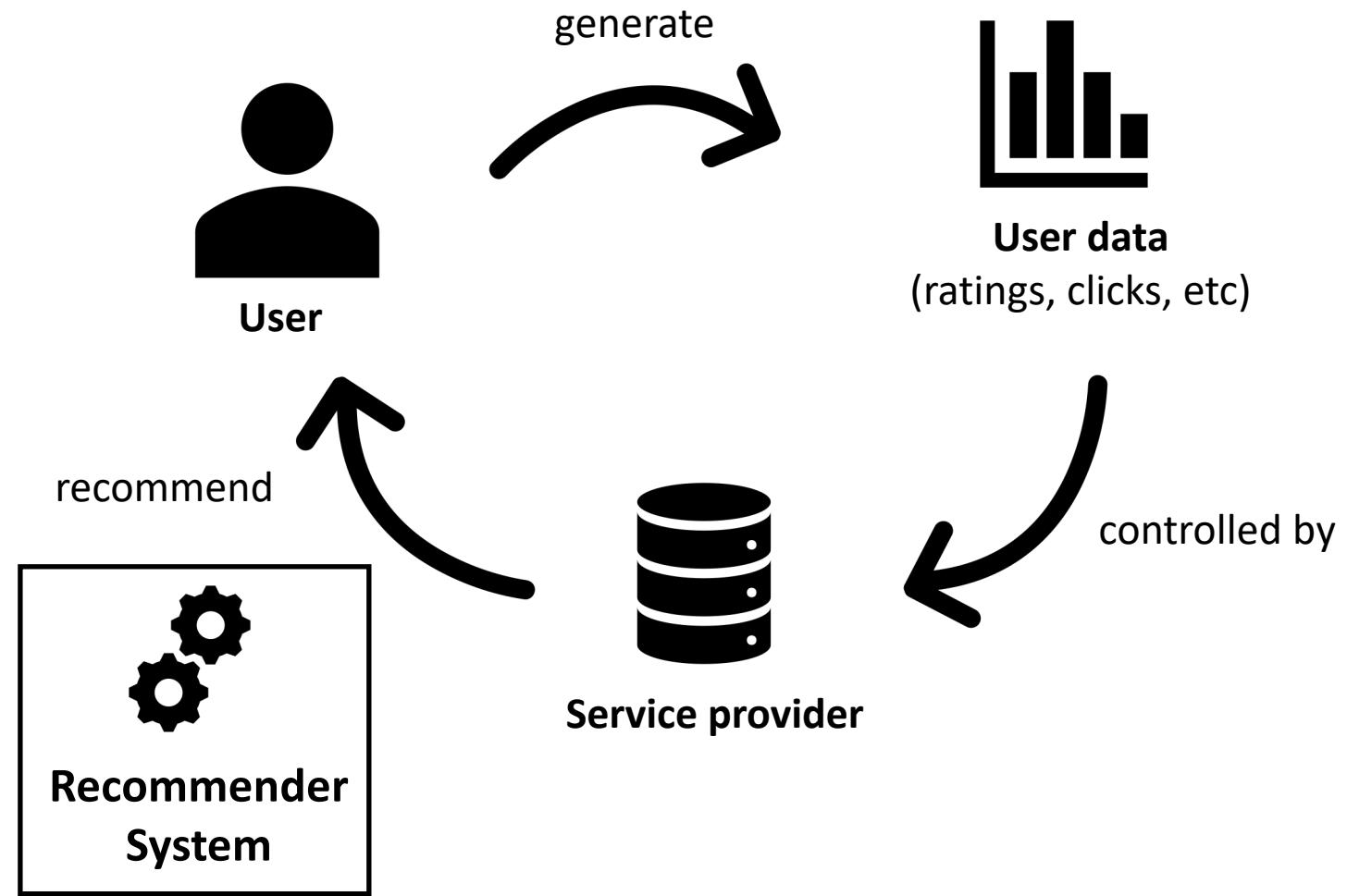
Cornell CIS
COMPUTING AND INFORMATION SCIENCE



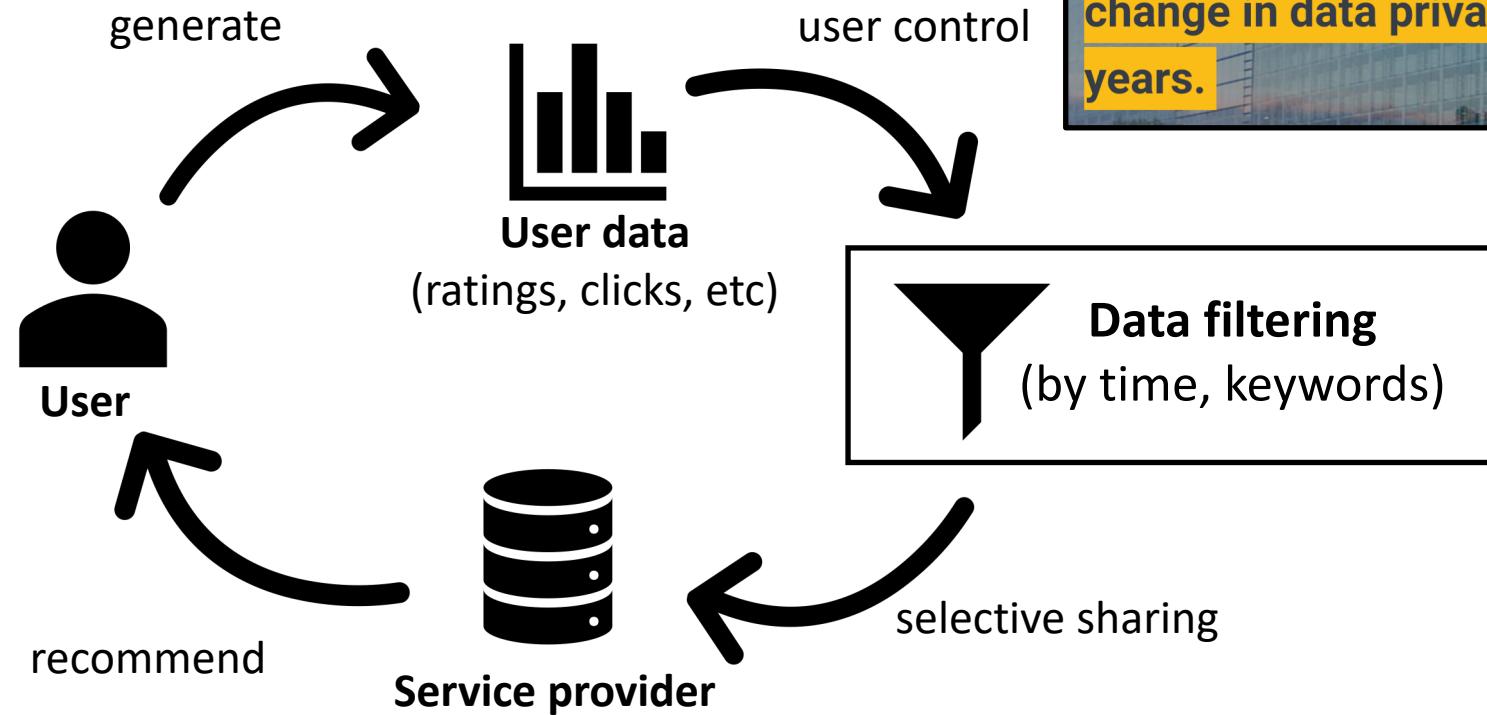
Classic recommendation framework



Classic recommendation framework

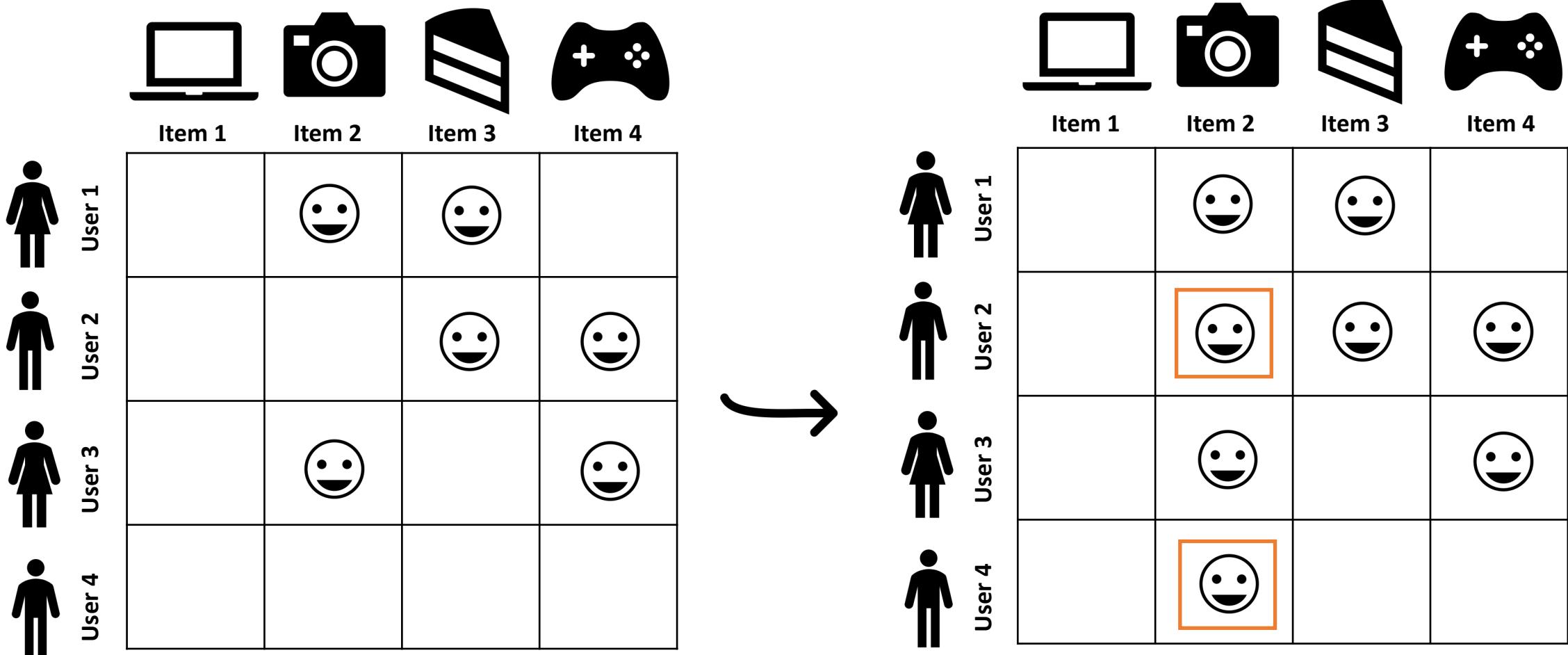


User control on personal data

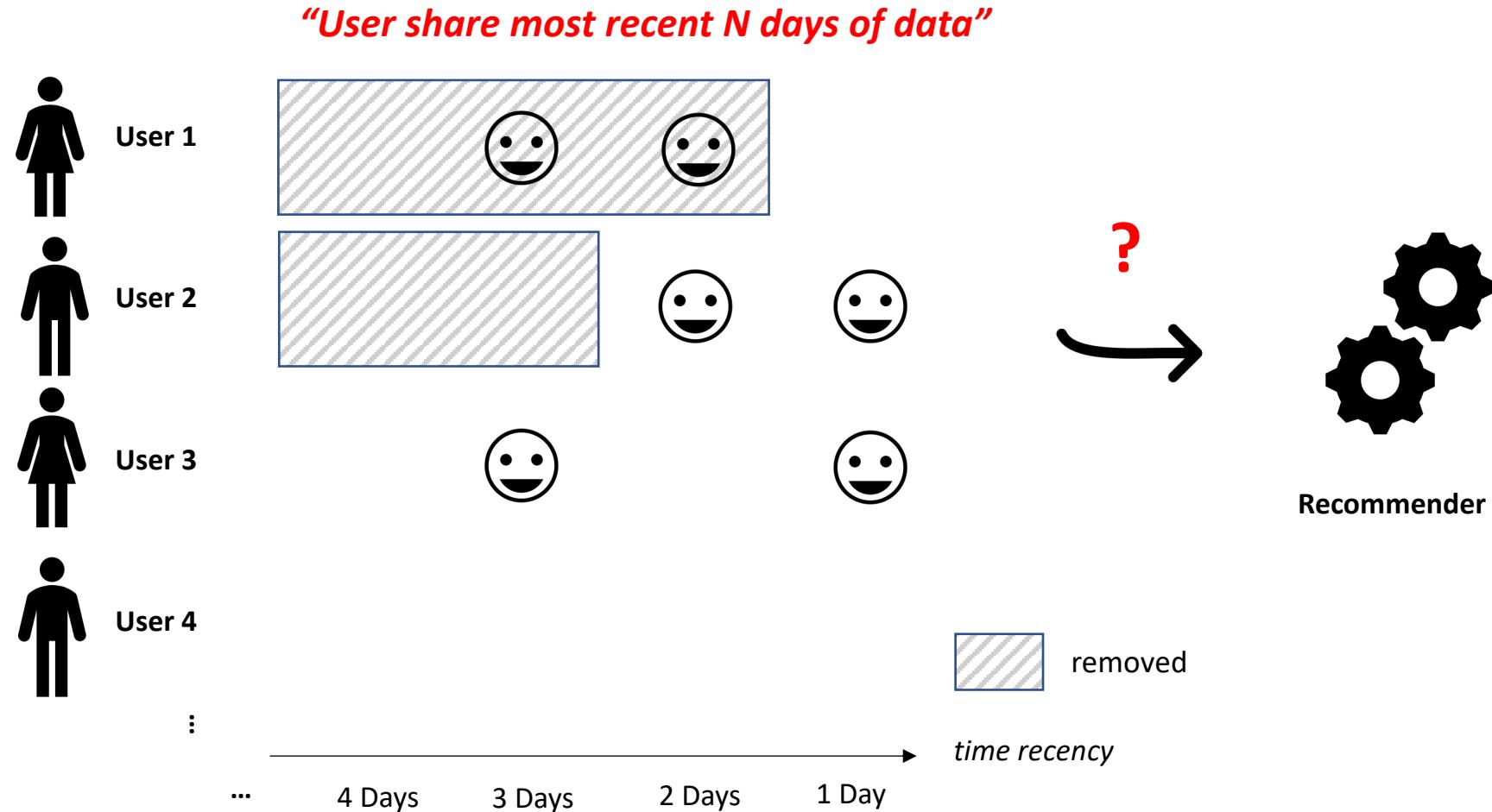


The EU General Data Protection Regulation (GDPR) is the most important change in data privacy regulation in 20 years.

Previous work: data modification



This work



Experiments

MovieLens 20M dataset [Harper et.al.]

Training/Validation
(1 year history)



Experiments

MovieLens 20M dataset [Harper et.al.]

Training/Validation
(1 year history)

User percentage (P)
 $P \in \{0.25, 0.5, 0.75, 1.0\}$

Time span (N)
 $N \in \{1, 7, 14, 30, 60, 90, 180\}$

P of users

Share most recent
 N days of data

time-sensitive filtering

Experiments



MovieLens 20M dataset [Harper et.al.]

Training/Validation
(1 year history)

User percentage (P)
 $P \in \{0.25, 0.5, 0.75, 1.0\}$

Time span (N)
 $N \in \{1, 7, 14, 30, 60, 90, 180\}$

P of users

$1 - P$ of users

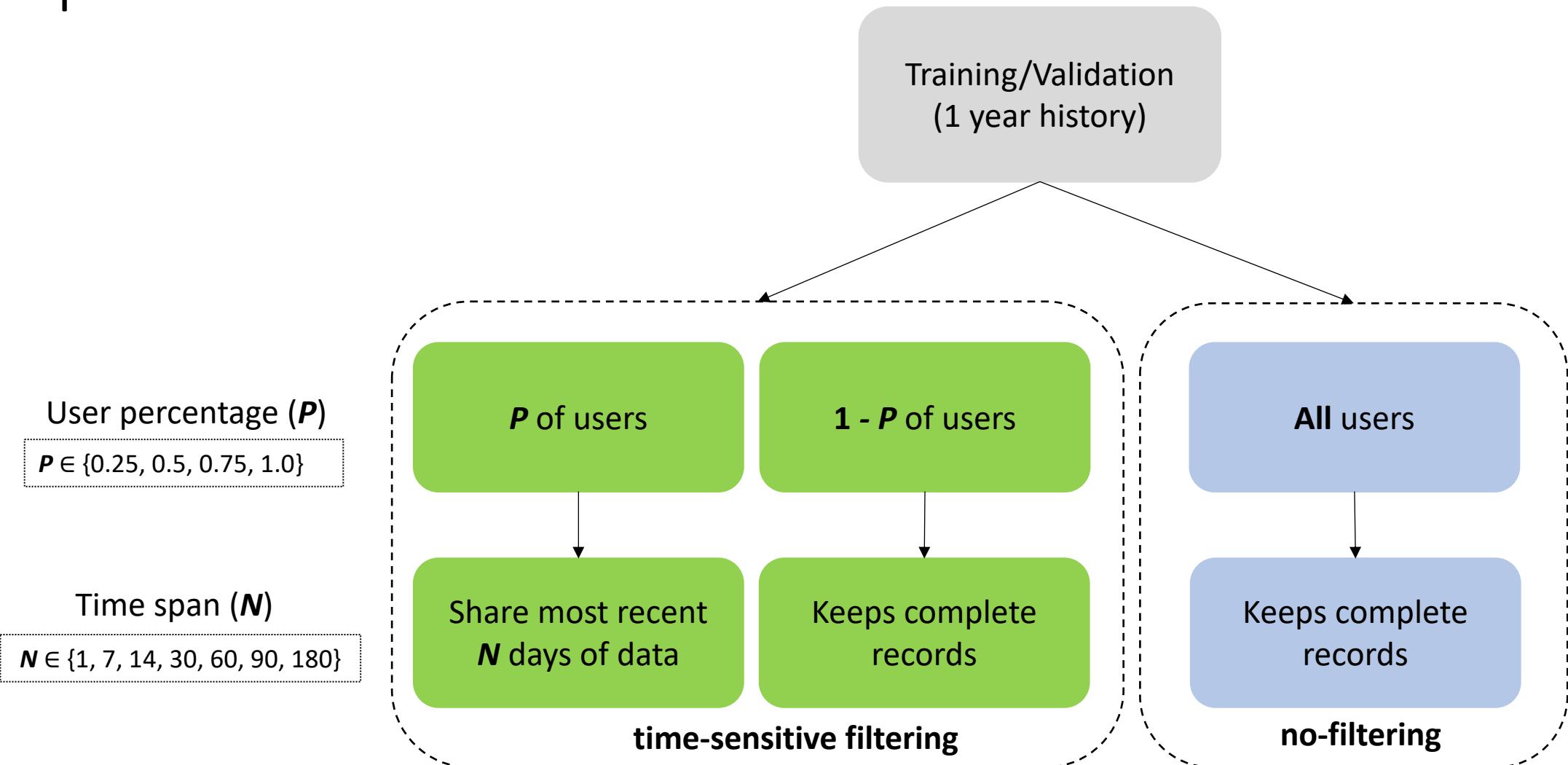
Share most recent
 N days of data

Keeps complete
records

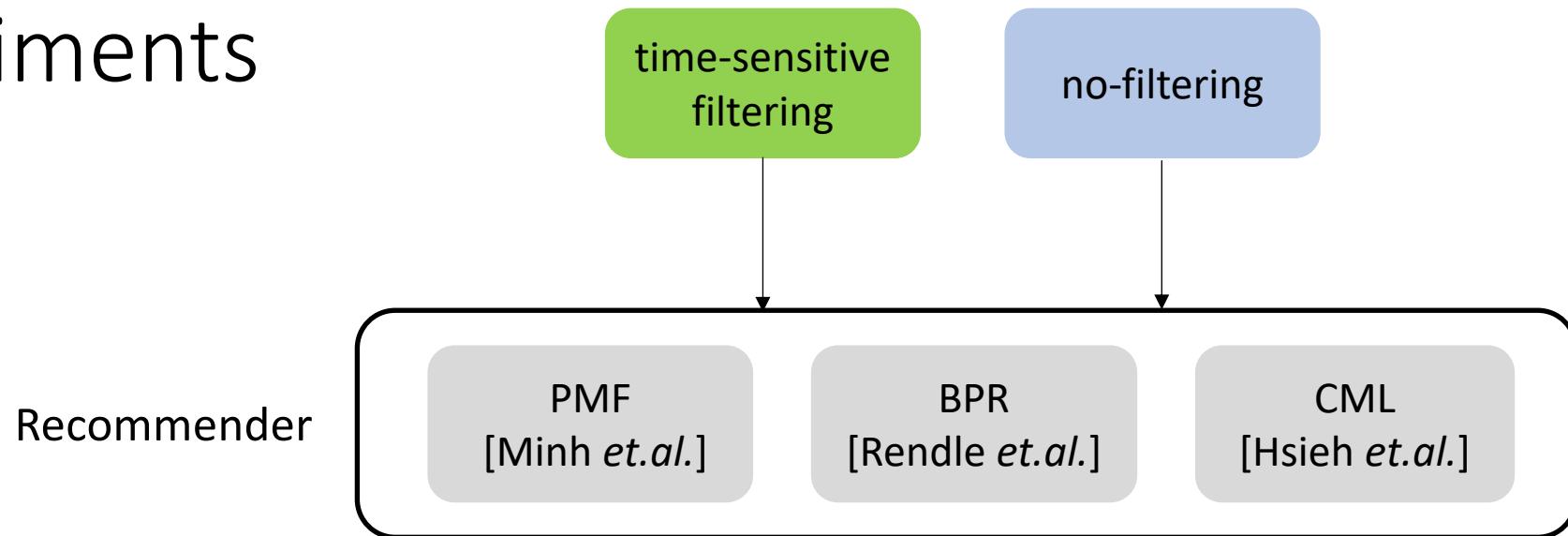
time-sensitive filtering

Experiments

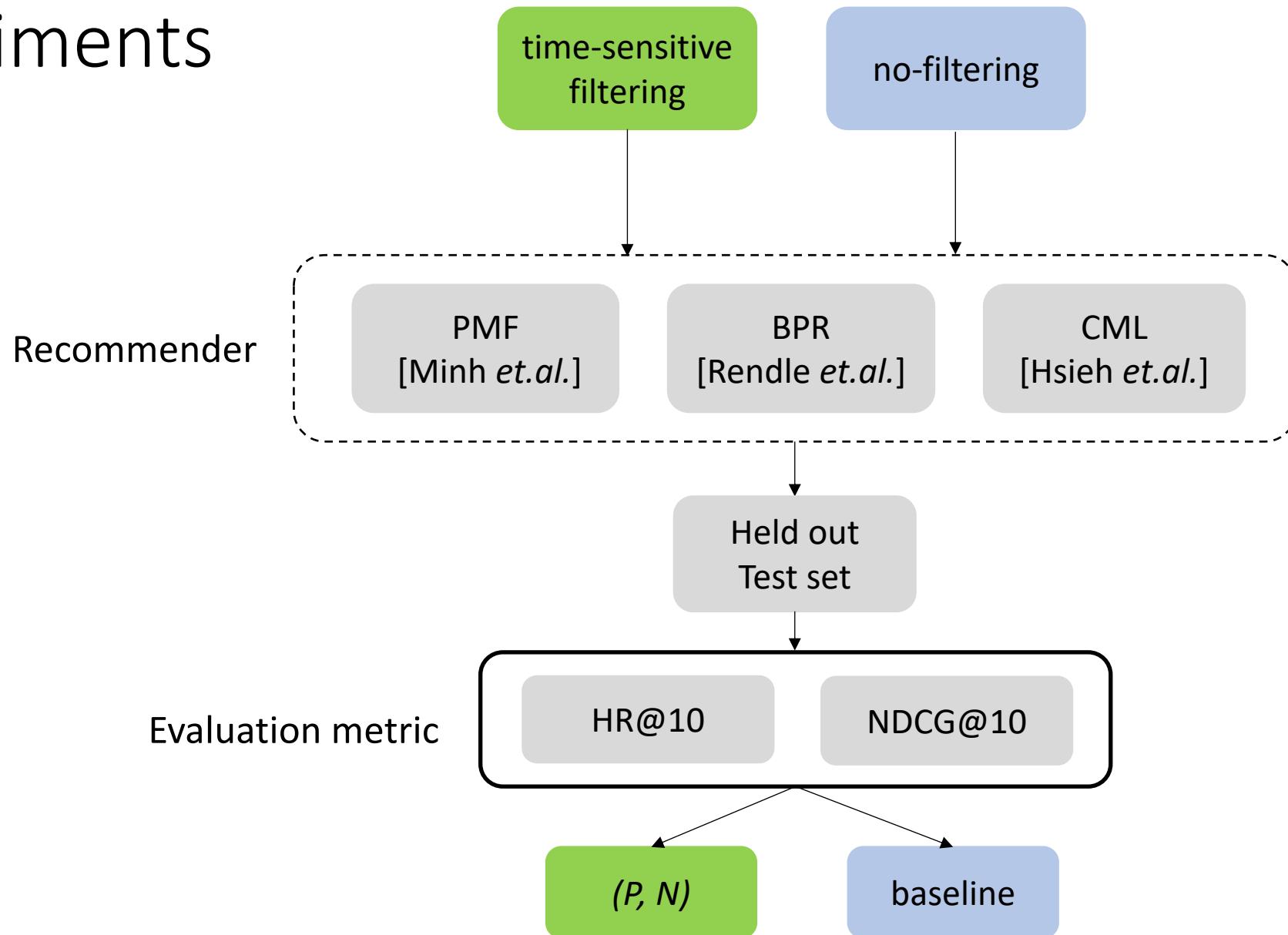
MovieLens 20M dataset [Harper et.al.]



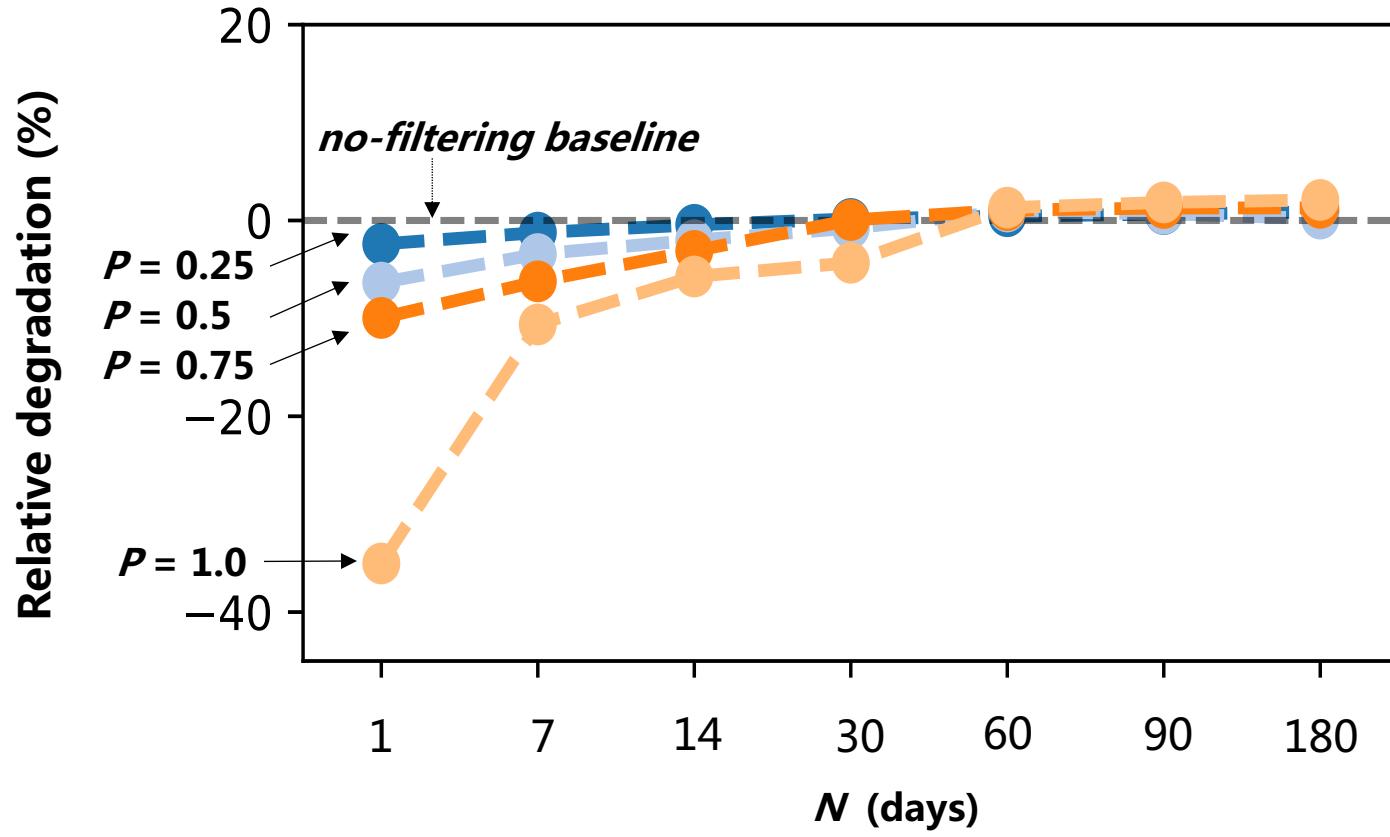
Experiments



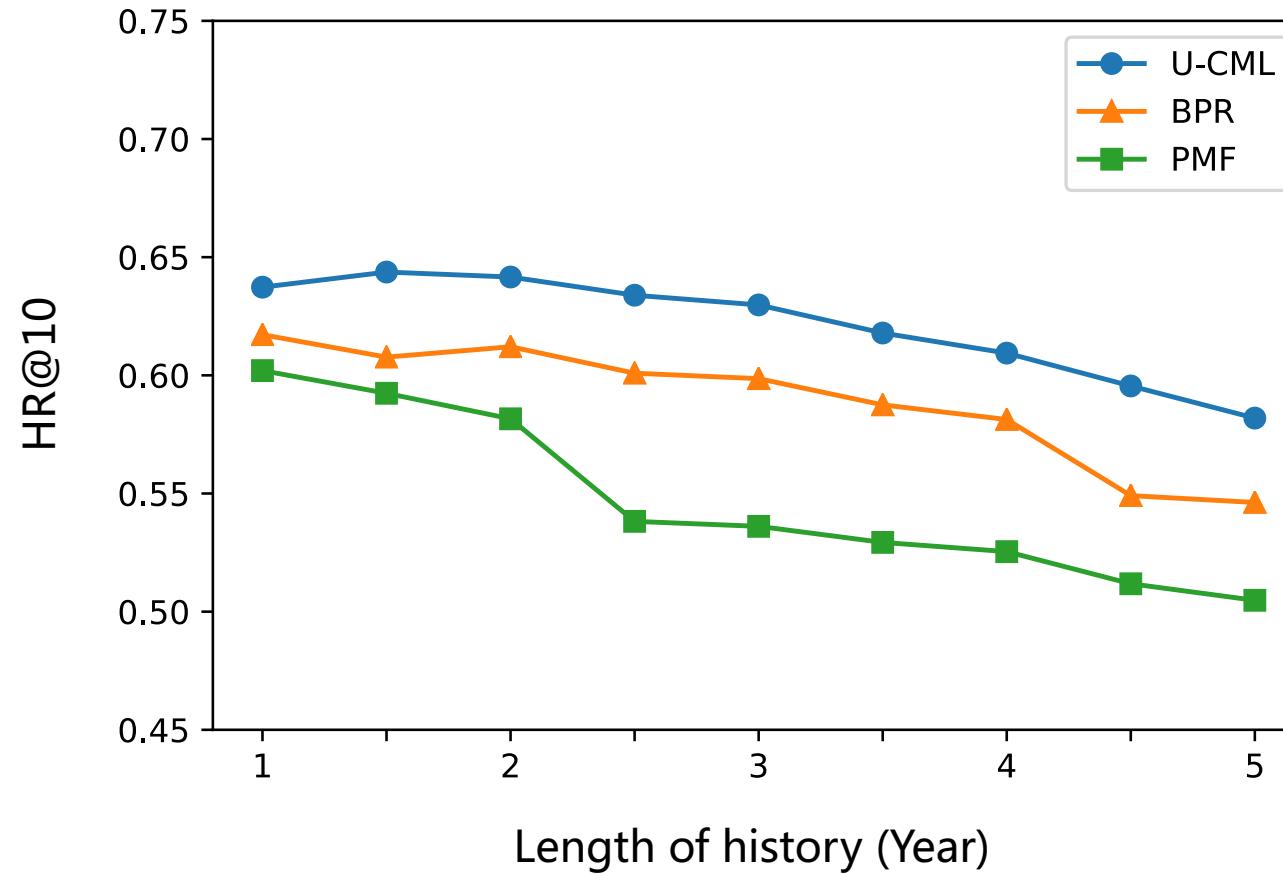
Experiments



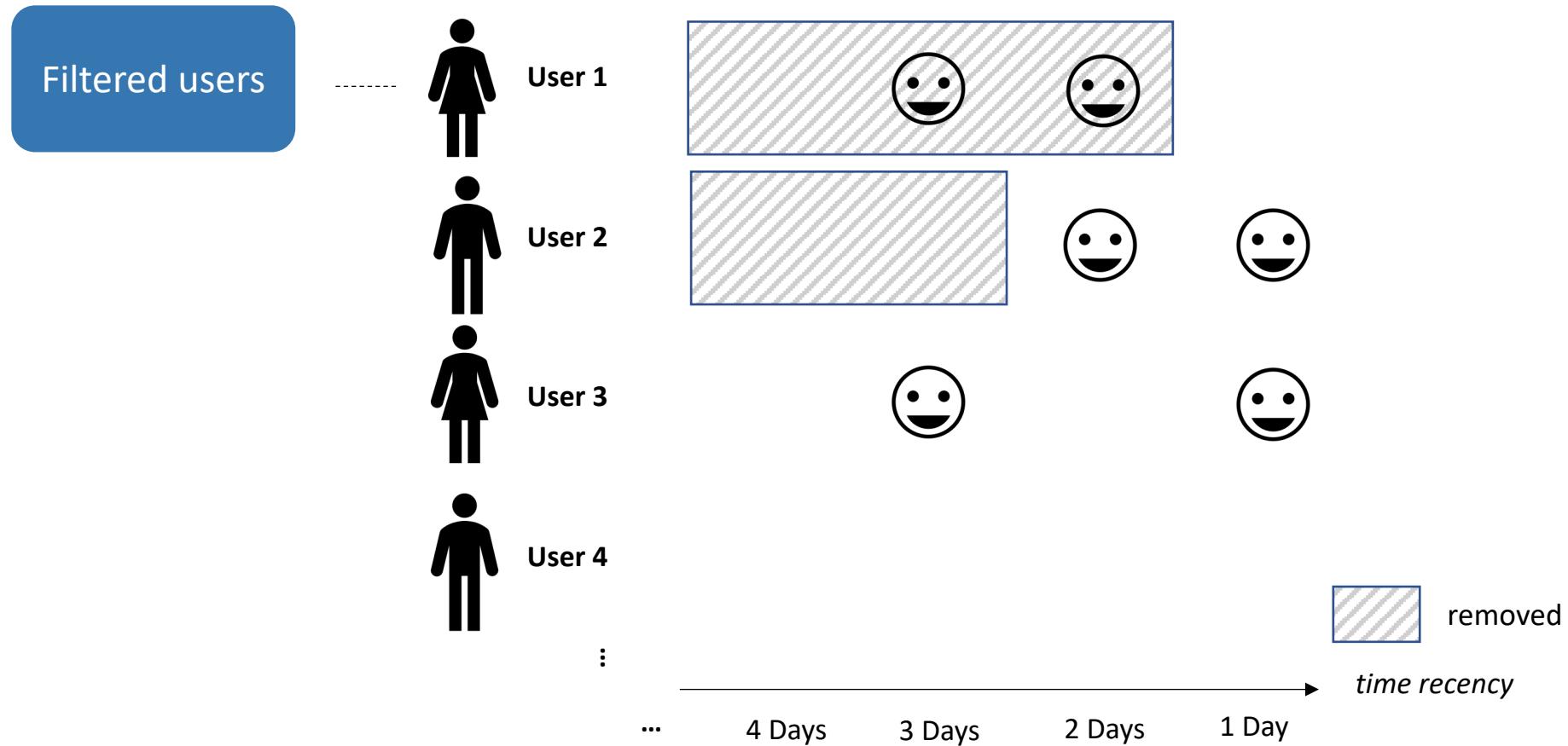
Population-level performance



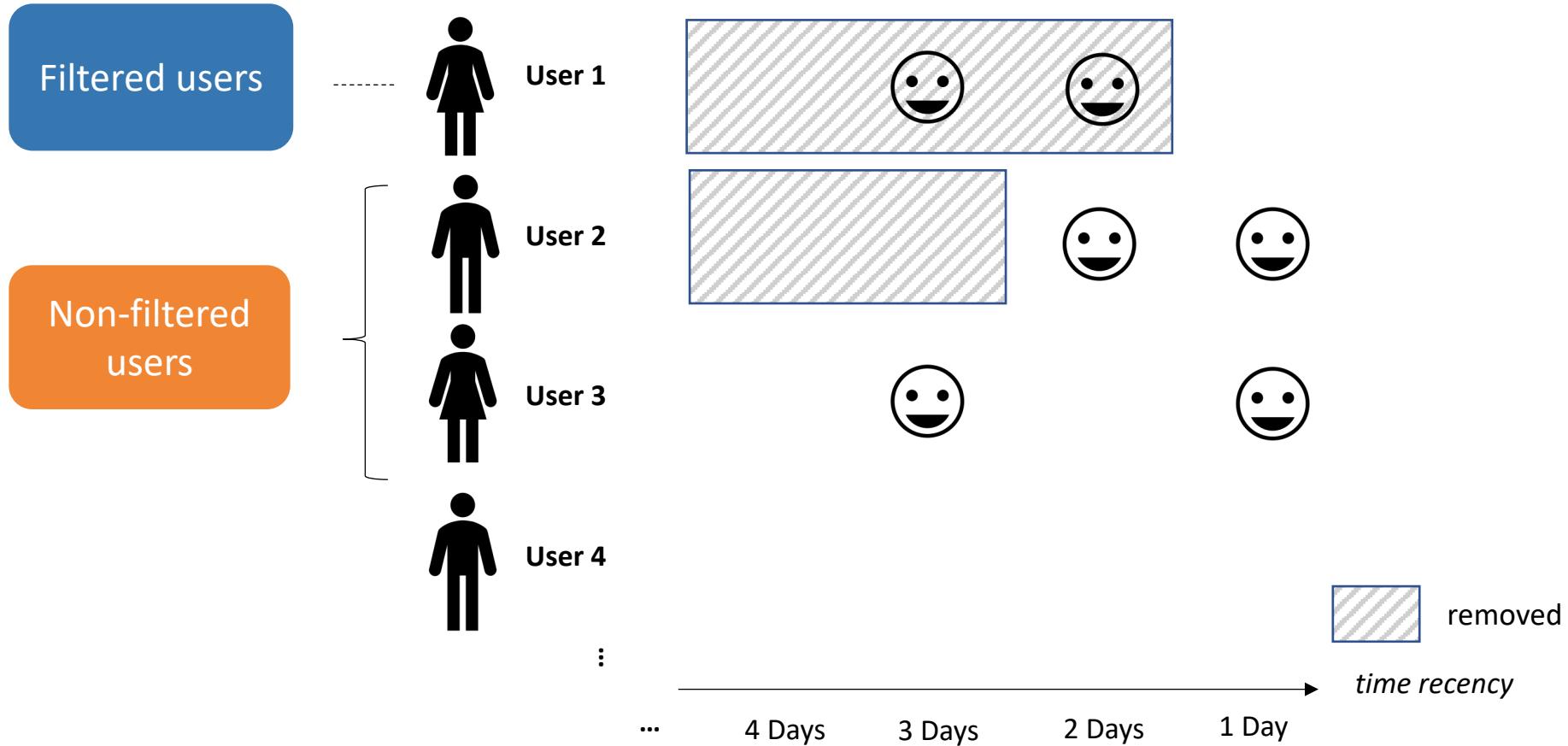
Effect of length of history



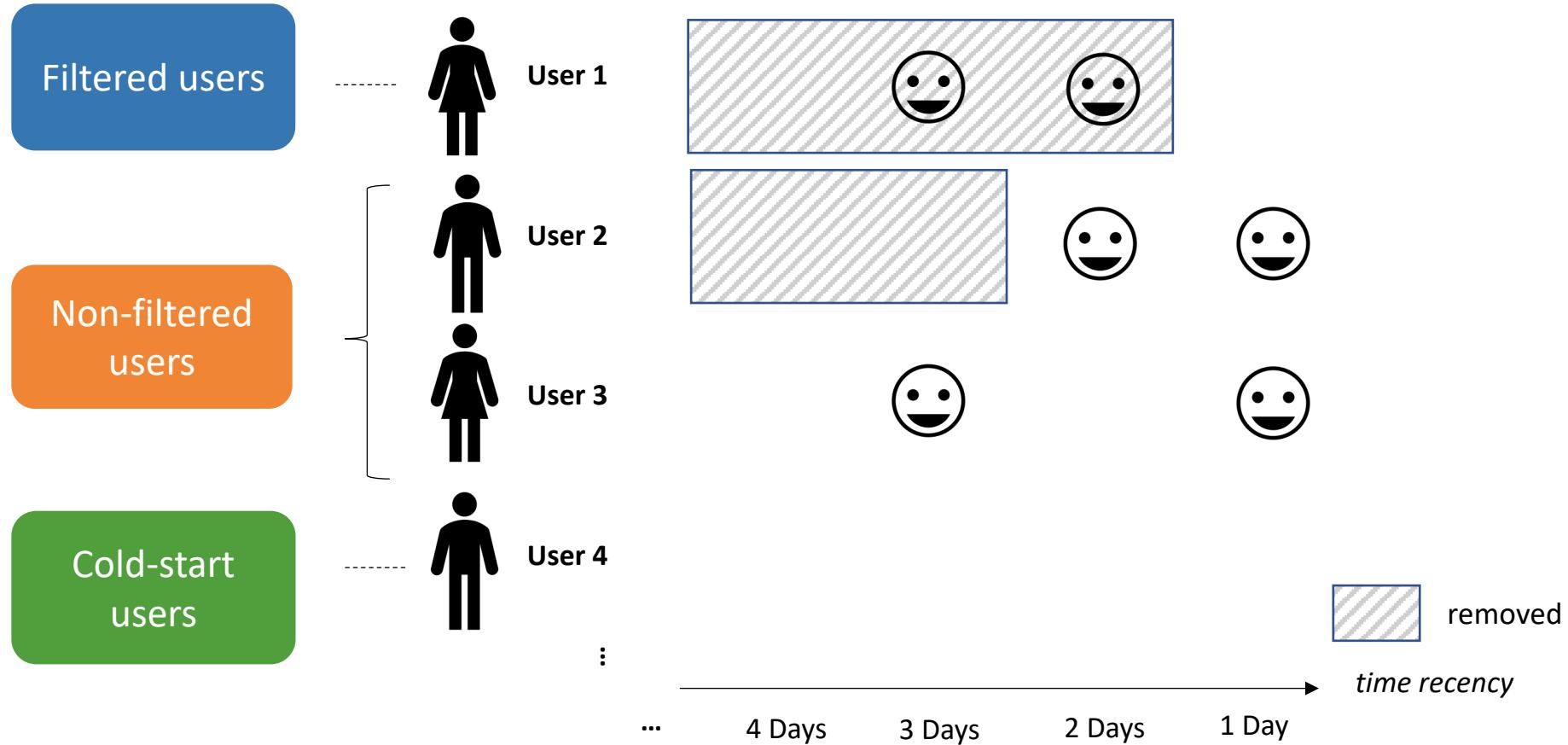
Disaggregated performance



Disaggregated performance



Disaggregated performance



Disaggregated performance

Filtered users

negatively affected when they share only a few days of data

Non-filtered
users

not affected by others choose to filter

Cold-start
users

the impact is algorithm-dependent

Disaggregated performance

Filtered users

negatively affected when they share only a few days of data

Non-filtered
users

not affected by others choose to filter

Cold-start
users

the impact is algorithm-dependent

Disaggregated performance

Filtered users

negatively affected when they share only a few days of data

Non-filtered
users

not affected by others choose to filter

Cold-start
users

the impact is algorithm-dependent

Takeaways and future work

- User-controlled data filtering is not always bad for recommendation.
- Explore models that use temporal, auxiliary and contextual information
- Apply this approach to real-world scenarios where user data sharing patterns are more complicated

Takeaways and future work

- User-controlled data filtering is not always bad for recommendation.
- Explore models that use temporal, auxiliary and contextual information
- Apply this approach to real-world scenarios where user data sharing patterns are more complicated

Takeaways and future work

- User-controlled data filtering is not always bad for recommendation.
- Explore models that use temporal, auxiliary and contextual information
- Apply this real-world scenarios where user data sharing patterns are more complicated

Thank you!

Code: <https://github.com/whongyi/datafilter-recsys>

Contact: hw557@cornell.edu

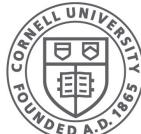
Twitter: @hongyi_wen



Small Data Lab



**CORNELL
TECH**



Cornell CIS
COMPUTING AND INFORMATION SCIENCE

Funders:



Oath:
A Verizon company

