

ContentWise Impressions

An Industrial Dataset with Impressions Included

Authors:

Fernando B. Pérez Maurera¹²³

Maurizio Ferrari Dacrema¹

Lorenzo Saule¹²

Mario Scriminaci²

Paolo Cremonesi¹

¹ Politecnico di Milano ² ContentWise ³ Speaker

CONTENTWISE



POLITECNICO
MILANO 1863

Outline

- State of the art
- Features of the dataset
- Final Remarks
- Acknowledgements

Classical recommendation scenario

Top TV Shows for Fernando



Top Movies for Fernando



Classical recommendation scenario

Top TV Shows for Fernando



Relevant

Less Relevant

Top Movies for Fernando



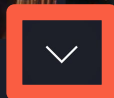
Classical recommendation scenario

Top TV Shows for Fernando



Relevant

Top Movies for Fernando



Less Relevant

Classical recommendation scenario

Top TV Shows for Fernando



Top Movies for Fernando



Interactions = Interests of users



Classical recommendation scenario

Top TV Shows for Fernando



Interactions = Interests of users

Top Movies for Fernando



Classical recommendation scenario

Top TV Shows for Fernando

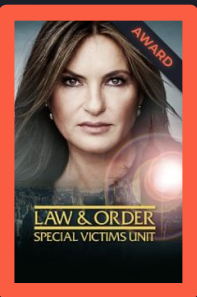


Interactions = Interests of users

Top Movies for Fernando



State of the Art Datasets



Only Interactions



ContentWise Impressions

Top TV Shows for Fernando



Top Movies for Fernando



Interactions + Impressions



Interest based on Item Position

Top TV Shows for Fernando



(1, 1)

Top Movies for Fernando



Less exploration ~ Lower interest

Interest based on Item Position

Top Movies for Fernando



Most viewed movies



(3, 9)

Features of the Dataset

- Data collected from OTT service
- Content related to television & cinema
- Open Source
 - Dataset and code available on GitHub ([Link](#))
 - Flexible Creative Commons license ([License](#))

Features of the Dataset

# Users	42,153
# Items	145,074
# Interactions	10,457,810
# Impressions	23,650,070

Date Range	3 Months
# Attributes Interactions	10
# Attributes Impressions	4

Impressions

Interaction density	90.55% (first 3 rows)
Row Position (min, 25%, 50%, 75%, max) [mean \pm std]	(0, 0, 1, 2, 17) [1.43 \pm 2.06]
Recommendation Length (min, 25%, 50%, 75%, max) [mean \pm std]	(1, 10, 10, 12, 481) [10.71 \pm 3.83]

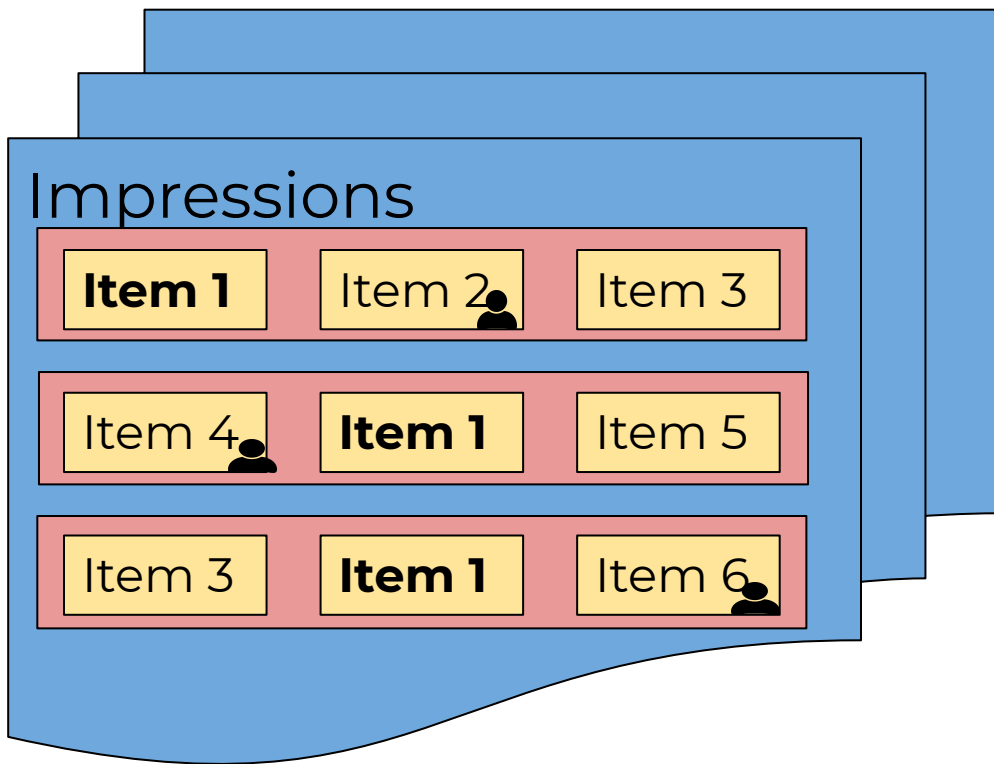
Interactions

Interaction Type	Count	Percentage
<i>View</i>	6,122,105	58.54%
<i>Access</i>	4,105,530	39.26%
<i>Purchase</i>	221,066	2.11%
<i>Rating</i>	9,109	0.09%
Total	10,457,810	100%

Final Remarks

- Impressions are vital for Recommender Systems
 - Better & realistic recommendation models
 - User intentions
- First open-sourced dataset with impressions
- Integrate impressions to recommendations
 - Re-ranking
 - Increase coverage

Final Remarks



- Is **Item 1** really relevant to the user?
- Is **Item 8** relevant to the user? (Never shown)



Acknowledgements

SIGIR Student Travel Grants



CONTENTWISE

Headquarters

Via Schiaffino, 11
20158 Milano
T. +39 02 4951 7001

USA East

211 Congress Street
Boston, MA 02110
T. +1 617 936 0212

USA West

12655 W Jefferson Blvd
Los Angeles, CA 90066
T. +1 323 524 0524

Singapore

5 Temasek Boulevard
Singapore, 038985

contentwise.com



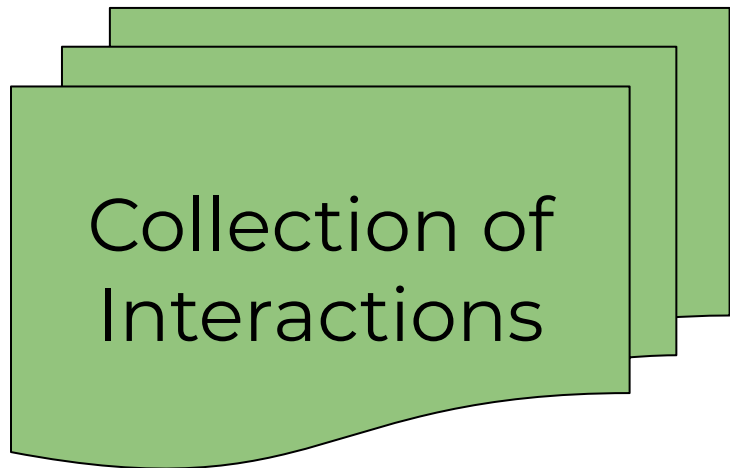
Extra Slides

Item Types

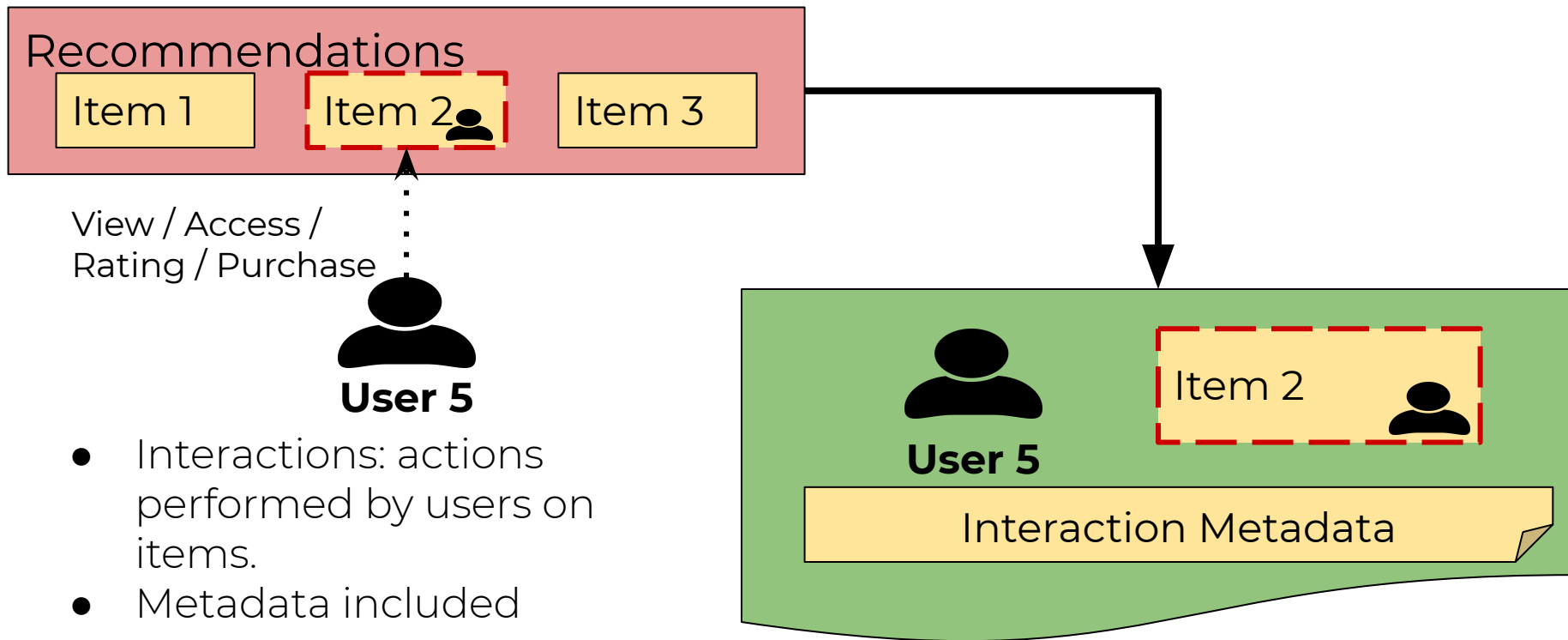
Item Type	Count	Percentage
<i>Episodes of TV series</i>	123,831	85,36%
<i>Movies</i>	13,733	9.47%
<i>Movies and clips in series</i>	5,722	3,94%
<i>TV movies and shows</i>	1,788	1.23%
Total	145,074	100%

ContentWise Impressions?

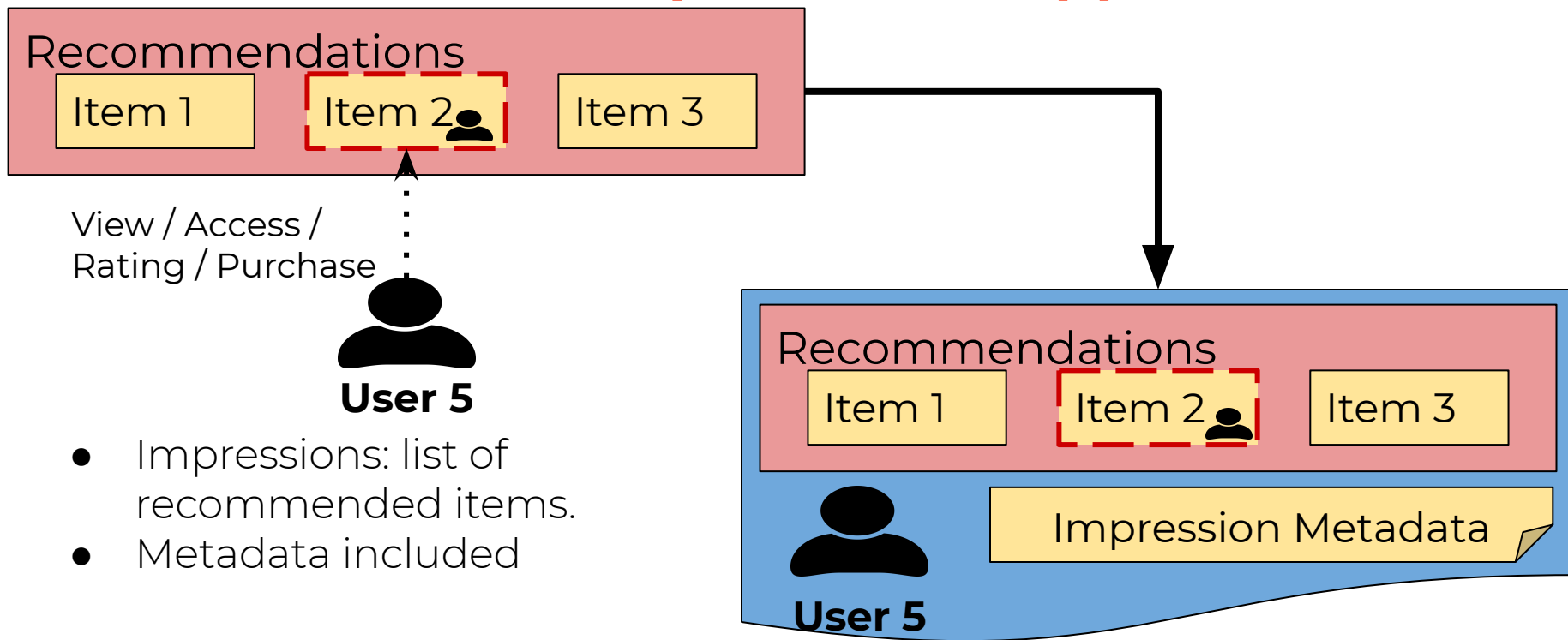
- Collections of **interactions** & **impressions**



Collection of interactions



Collection of impressions (1)



Collection of impressions (2)

