

Project – RecSys Challenge 2019

Recommender Systems

VU 194.035, 2019S

Dimitris Sacharidis



FAKULTÄT
FÜR INFORMATIK

Faculty of Informatics



acknowledgements

- some content from Peter Knees

RecSys Challenge

- in cooperation with the ACM RecSys Conference
- real-world (industry) recommendation task with real data
- competition with leaderboard (held-out data)
- prizes, invitation to write a report

RecSys Challenge 2019

- Industry Organizer:
- Travel **meta-search engine**
- Search for and compare hotels
- Prices and availability of 2M+ hotels from 400+ booking sites
- 55 localized platforms in 190+ countries
 - + enabling advertisers of hotels to access a broad audience of travelers...



<https://recsys.trivago.cloud/>
<http://www.recsyschallenge.com/2019/>
<https://recsys.acm.org/recsys19/challenge/>



Anreise

Mo, 16.09.19



Abreise

Fr, 20.09.19



Zimmer

Doppelzimmer

Suchen

Preis / Nacht

500€+

Unterkunft

Alle Typen



Gästebewertung

Beliebig



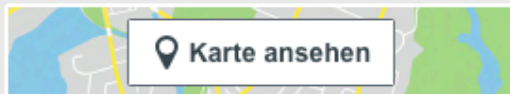
Hotelstandort

Zentrum



Mehr Filter

Wählen

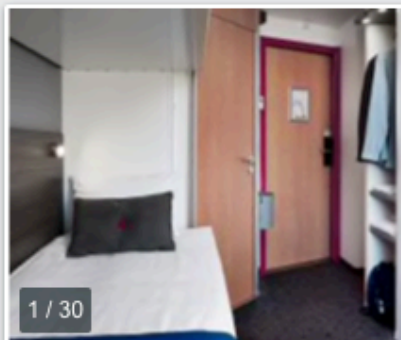


Sortiert nach

Unseren Empfehlungen



So beeinflussen Zahlungen an uns unser Ranking



Hotel CABINN City

★★ Hotel



Kopenhagen, 0.5 km bis Copenhagen Central Station



7.4

(13129 Bewertungen)



Hervorragende Lage · Sehr gute Sauberkeit

Expedia
107€Hotels.com
107€Amoma.com
143€Mehr Angebote
ab 107€

Stayforlong

~~172€~~

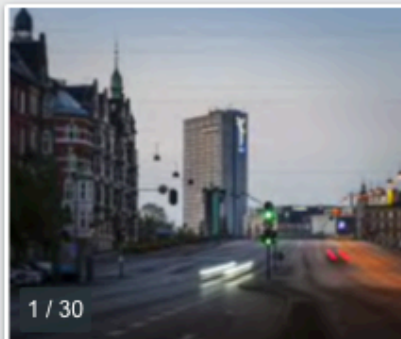
Booking.com

107€

4 Nächte für 428€

-37%

zum Angebot



Radisson Blu Scandinavia Hotel Copenh...

★★★★ Hotel



Kopenhagen, 1.5 km bis Copenhagen Central Station



7.7

Gut (5981 Bewertungen)



Sehr gute Sauberkeit · Sehr guter Service

Expedia
240€Hotels.com
240€Flixbus
241€Mehr Angebote
ab 237€

ZenHotels.com

~~265€~~

Booking.com

240€

4 Nächte für 960€

zum Angebot



RecSys Challenge 2019

- **session-aware** recommendation task
- sessions are sequences of various type of **actions**
- users can have multiple sessions
- goal is to predict the **clickout** action:
 - the hotel offer the user will click on (and be redirected for booking)

Dataset – Session Actions (excerpt)

- **user_id**: identifier of the user
- **session_id**: identifier of each session
- **step**: step in the sequence of actions within the session
- **action_type**: identifier of the action that has been taken by the user.
 - **clickout item**: user makes a click-out on the item.
 - **interaction item rating**: user interacts with a rating or review.
 - **interaction item info**: user interacts with item information.
 - **change of sort order**: user changes the sort order.
 - **search for item**: user searches for an accommodation.
- **reference**: reference value of the action
- **impressions**: items displayed at the time of a click-out
- **prices**: prices of items displayed

Dataset – Session Actions – Example

user_id	session_id	timestamp	step	action_type	reference	platform	city	device	current_filters	impressions	prices
93F7WGHBP03A	569f5ea70df51	1541543231	1	search for destination	Barcelona, Spain	US	Barcelona, Spain	desktop			
93F7WGHBP03A	569f5ea70df51	1541543269	2	filter selection	Focus on Distance	US	Barcelona, Spain	desktop	Focus on Distance		
93F7WGHBP03A	569f5ea70df51	1541543269	3	search for poi	Port de Barcelona	US	Barcelona, Spain	desktop	Focus on Distance		
93F7WGHBP03A	569f5ea70df51	1541543371	4	interaction item deals	40255	US	Barcelona, Spain	desktop			
93F7WGHBP03A	569f5ea70df51	1541543425	5	clickout item	40255	US	Barcelona, Spain	desktop		6744 40181 40630 84610 2282416 1258693 974937 147509 128238 7998246 40255 3058538 1637385 40285 147502 921707 40849 6757 12770 893733 685091 147522 40708 860451 6819	162 91 218 190 176 365 272 159 139 240 136 5099 164 116 90 192 191 213 109 178 131 128 168 101 331
93F7WGHBP03A	569f5ea70df51	1541543741	6	search for item	81770	US	Barcelona, Spain	desktop			
93F7WGHBP03A	569f5ea70df51	1541543770	7	interaction item info	81770	US	Barcelona, Spain	desktop			
93F7WGHBP03A	569f5ea70df51	1541543813	8	clickout item	81770	US	Barcelona, Spain	desktop		6832 40396 6621784 40197 6743 147488 40635 6177052 6742 1319782 40763 945255 83855 39937 1870125 1354432 6812 82400 40181 6834 81770 5056102 40797 923935 40284	347 245 199 65 359 233 227 270 294 625 208 174 121 217 226 616 293 166 91 198 274 272 123 130 131

Dataset – Session Actions

- Train Set
 - ~16M session actions
 - belonging to 910.683 sessions
 - by 730.803 unique users
- Test Set
 - ~3.8M session actions
 - belonging to 291K sessions
 - by 251K users (only 33K also in training set)
 - 528.779 (~14%) actions are clickouts
 - **253.573 clickouts need to be predicted**

Dataset – Item Metadata

- item_id: identifier of the accommodation
- properties: list of filters applicable for the given item

```
7818446,Free WiFi (Public Areas)|Hotel Bar|4 Star|WiFi (Public  
Areas)|Good Rating|Free WiFi (Rooms)|Free WiFi  
(Combined)|Beach|Swimming Pool (Indoor)|Swimming Pool (Outdoor)|WiFi  
(Rooms)|Satisfactory Rating|Restaurant|Gym|Car Park  
6468,Hotel Bar|Laundry Service|Luxury Hotel|Television|WiFi (Rooms)|Gay  
Friendly|From 2 Stars|Hotel|Hairdryer|Satellite TV|Singles|Computer  
with Internet|Free WiFi (Rooms)|Lift|3 Star|Reception (24/7)|Openable  
Windows|Conference Rooms|Family Friendly|Non-Smoking  
Rooms|Shower|Design Hotel|Childcare|Free WiFi (Combined)|WiFi (Public  
Areas)|Good Rating|Central Heating|From 3 Stars|Safe  
(Hotel)|Telephone|Free WiFi (Public Areas)|Satisfactory  
Rating|Business Hotel|Romantic|Cot
```

Evaluation

- For each clickout to be predicted, return a **ranked list** of the items in impressions
- One among them is the actual clickout item
- single evaluation criterion: Mean Reciprocal Rank (MRR)
- for the i -th prediction task, let rank_i be the position in the returned list of the actual clickout item
- MRR is the average (mean) over all tasks of the inverse (reciprocal) of rank_i :

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

Project

- implement recommenders for the RecSys Challenge 2019
- optional but advised: submit your solutions to the challenge
- each week we discuss your **progress** and **ideas**
 - **participation** is part of the project's grade
- at the end of project, hand in **project report**, and **code**
- we will provide a **smaller subset** of the dataset for your testing