

# Recommendation system

GROUP 6  
CAO KY HIEU  
NGUYEN NGOC THIEN

June 2019

## 1 trivago RecSys Challenge 2019 Dataset

- Session actions (train.csv and test.csv)

- **user\_id**: identifier of the user
- **timestamp**: UNIX timestamp for the time of the interaction.
- **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 and gets forwarded to a partner website. The reference value for this action is the item\_id. Other items that were displayed to the user and their associated prices are listed under the ‘impressions’ and ‘prices’ column for this action.
  - \* **interaction item rating**: user interacts with a rating or review of an item. The reference value for this action is the item id.
  - \* **interaction item info**: user interacts with item information. The reference value for this action is the item id.
  - \* **interaction item image**: user interacts with an image of an item. The reference value for this action is the item id.
  - \* **interaction item deals**: user clicks on the view more deals button. The reference value for this action is the item id.
  - \* **change of sort order**: user changes the sort order. The reference value for this action is the sort order description.
  - \* **filter selection**: user selects a filter. The reference value for this action is the filter description.
  - \* **search for item**: user searches for an accommodation. The reference value for this action is the item id.
  - \* **search for destination**: user searches for a destination. The reference value for this action is the name of the destination.

- \* **search for poi**: user searches for a point of interest (POI). The reference value for this action is the name of the POI.
  - **reference**: reference value of the action as described for the different action types
  - **platform**: country platform that was used for the search, e.g. trivago.de (DE) or trivago.com (US)
  - **city**: name of the current city of the search context
  - **device**: device that was used for the search
  - **current\_filters**: list of pipe-separated filters that were active at the given timestamp
  - **impressions**: list of pipe-separated items that were displayed to the user at the time of a click-out (see `action_type = clickout_item`)
  - **prices**: list of pipe-separated prices of the items that were displayed to the user at the time of a click-out (see `action_type = clickout_item`)
- **Item metadata (item\_metadata.csv)**
  - **item\_id**: identifier of the accommodation as used in the reference values for item related action types, e.g. `clickout_item` and item interactions, and impression list
  - **properties**: pipe-separated list of filters that are applicable for the given item

## 2 Problem definition

- The training set contains user actions up to a specified time (split date). It can be used to build models of user interactions and specifies the type of action that has been performed (filter usage, search refinements, item interactions, item searches, item click-outs) as well as information about impressed items and prices at the time of a click-out.
- The recommendations should be provided for a test set that contains information about sessions after the split date but is missing the information about the accommodations that have been clicked in the last part of the sessions. The required output is a list of maximum 25 items for each click-out ordered by preferences for the specific user. The higher the actually clicked item appears on the list the higher the score.

## 3 References

- **Neural Factorization Machines for Sparse Predictive Analytics**
- **Attentional Factorization Machines: Learning the Weight of Feature Interactions via Attention Networks**

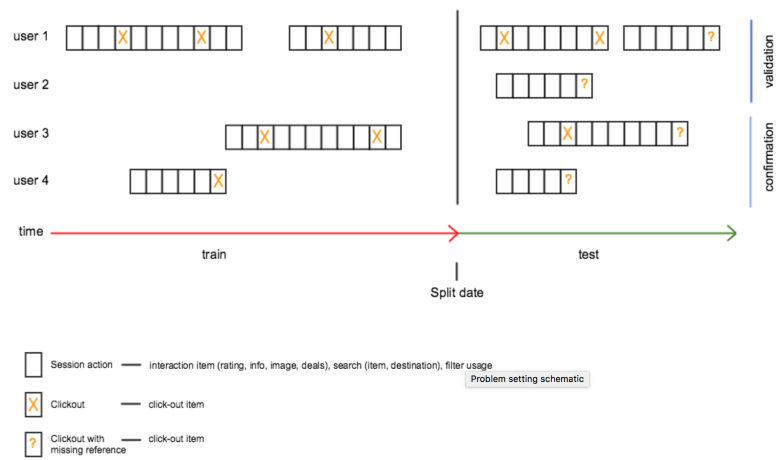


Figure 1: Caption