Lecture - 16

Explaining Recommendations

Arpit Rana

1st Apr 2022

Recommendation Explanation

An explanation of a recommendation is any content additional to the recommendation itself that justifies the recommended item to the user.

> *We recommend you the movie "A Beautiful Mind" because it has features 'drama' and 'biography' that you liked before.*

This textual description justifies the movie "A Beautiful Mind" to the user by means of its features 'drama' and 'biography' which she liked before.

Goal of Explanations

Recommender systems provide explanations to -

- reveal how a recommender has reached its conclusions (transparency),
- help users to modify or correct the assumptions (scrutability),
- help users make better decisions (effectiveness),
- increase user trust in the system (trust),
- help users make decisions more quickly (efficiency),
- influencing user behaviour (persuasiveness), and
- improve user acceptance of recommendations (satisfaction).

Goal of Explanations

In addition to supporting end-users, explanations of recommendations may have a role in issues such as:

- detecting shilling attacks;
- detecting bias and discrimination; and
- contesting algorithmic decisions on personal data
  (as allowed for in government legislation such as GDPR in Europe (2016) and PDPR in India (2019))

Characterizing Explanations of Recommendations

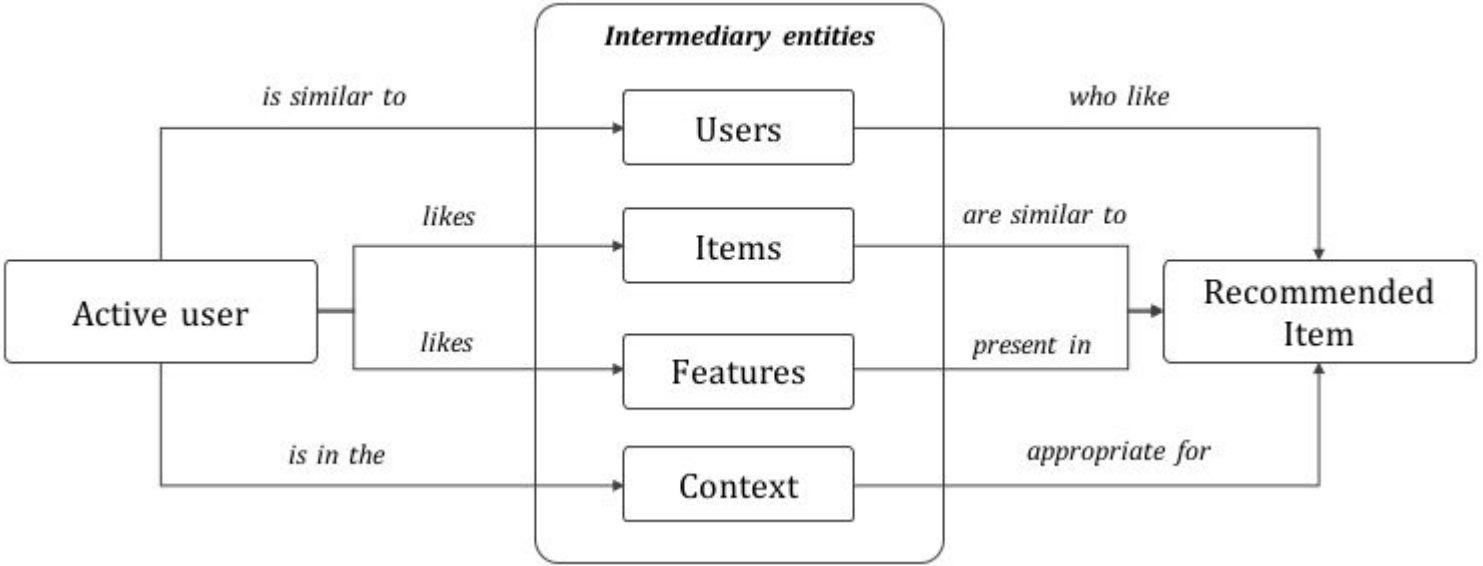
Explanations of recommendations can be characterized in a variety of different ways such as:

- the type of knowledge they use (e.g. user demographics, item descriptions, etc.)
- their fidelity to the recommender (i.e. white-box vs. black-box), and
- their role in producing recommendations.

Based on their Type of Knowledge

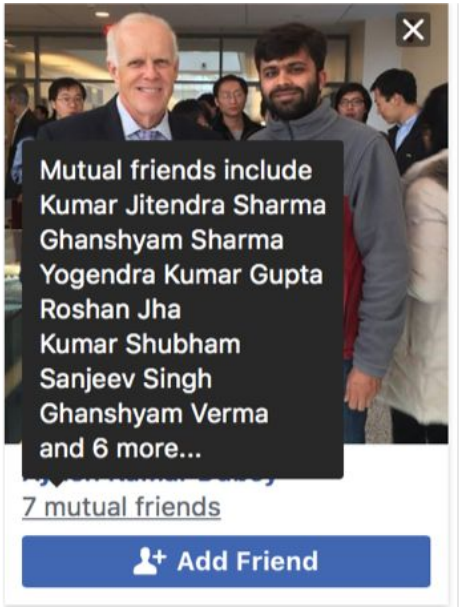
Explanations of recommendations often relate the recommended item to the user through intermediary entities,

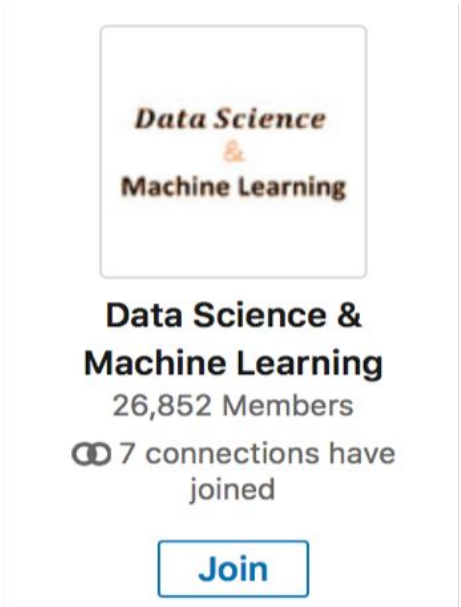
e.g. other users, other items, item features, or context



Based on their Type of Knowledge

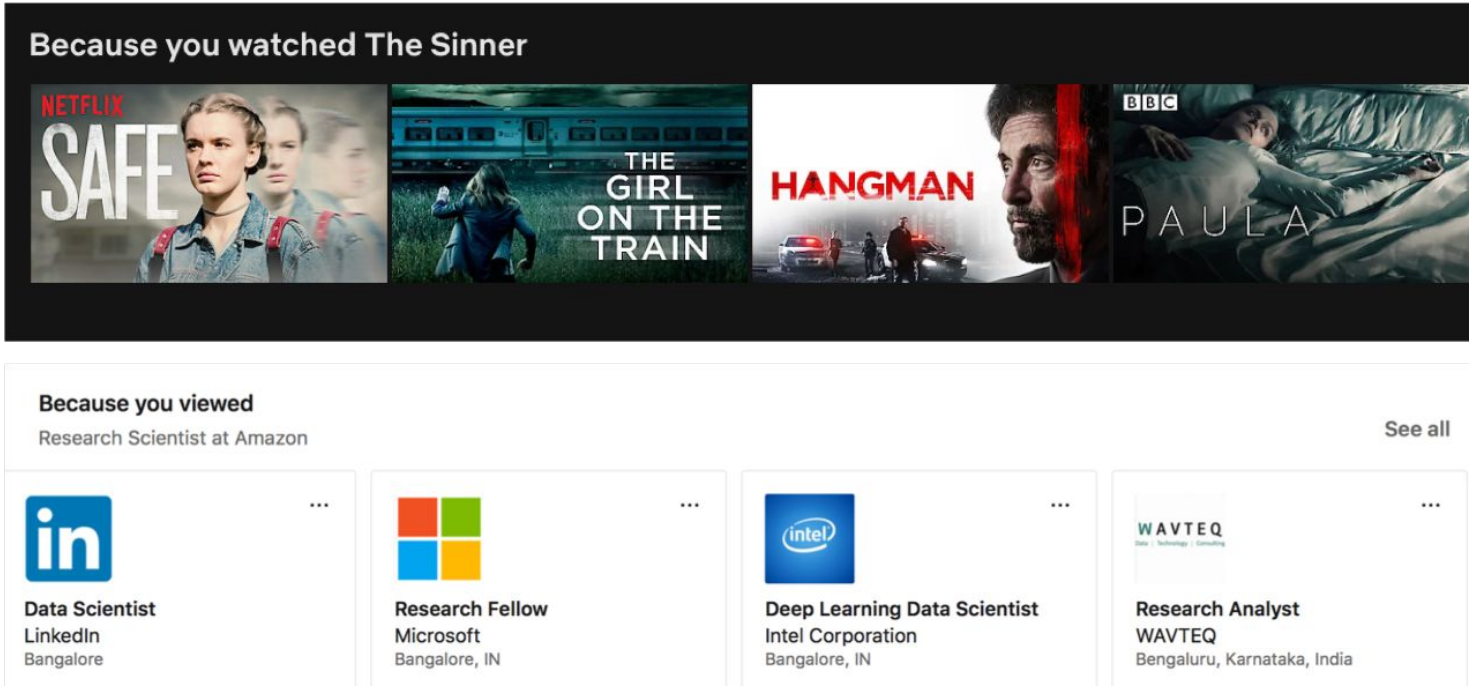
User-based Explanations





Based on their Type of Knowledge

Item-based Explanations



Based on their Type of Knowledge

Feature-based Explanations:

- *Attribute-value pairs*
  A movie based on the user's most preferred actor, genres and director
- *Item content*
  News, books, articles or blogs based on the keywords extracted from their textual content
- *User-generated tags*
  Items using tags that users assign to the items
- *User-reviews*
  Items using information extracted from user reviews
- *Linked data*
  Items using linked open data on DBpedia

Based on their Type of Knowledge

Contextual Explanations:

- Time, location, weather, or companions, can influence how a person perceives a product or service.
- Such observable contextual factors are used to explain the recommended items to the user.

Based on their Fidelity

In Artificial Intelligence in general, explanations are sometimes categorized as

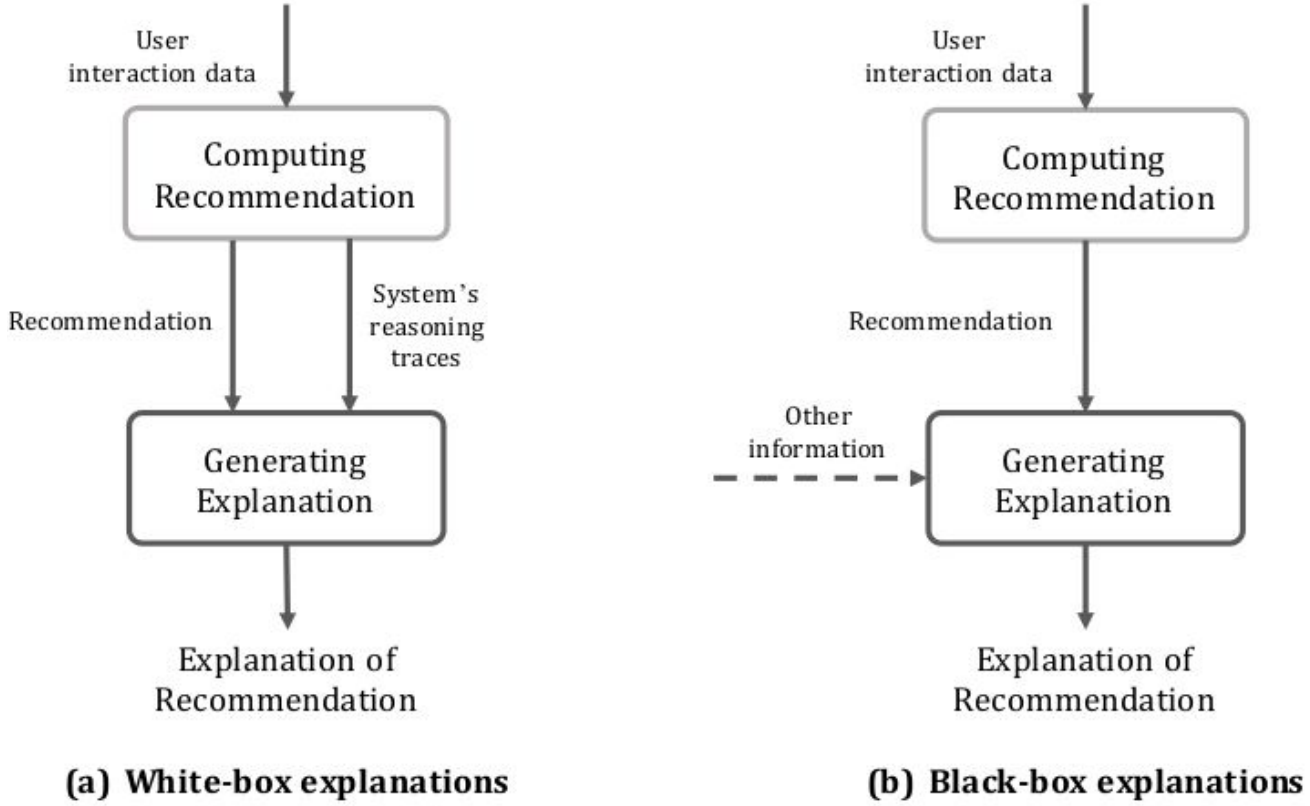
- White-box (a.k.a. *model-based*)

  These are built from traces of the system's reasoning, e.g. content-based methods

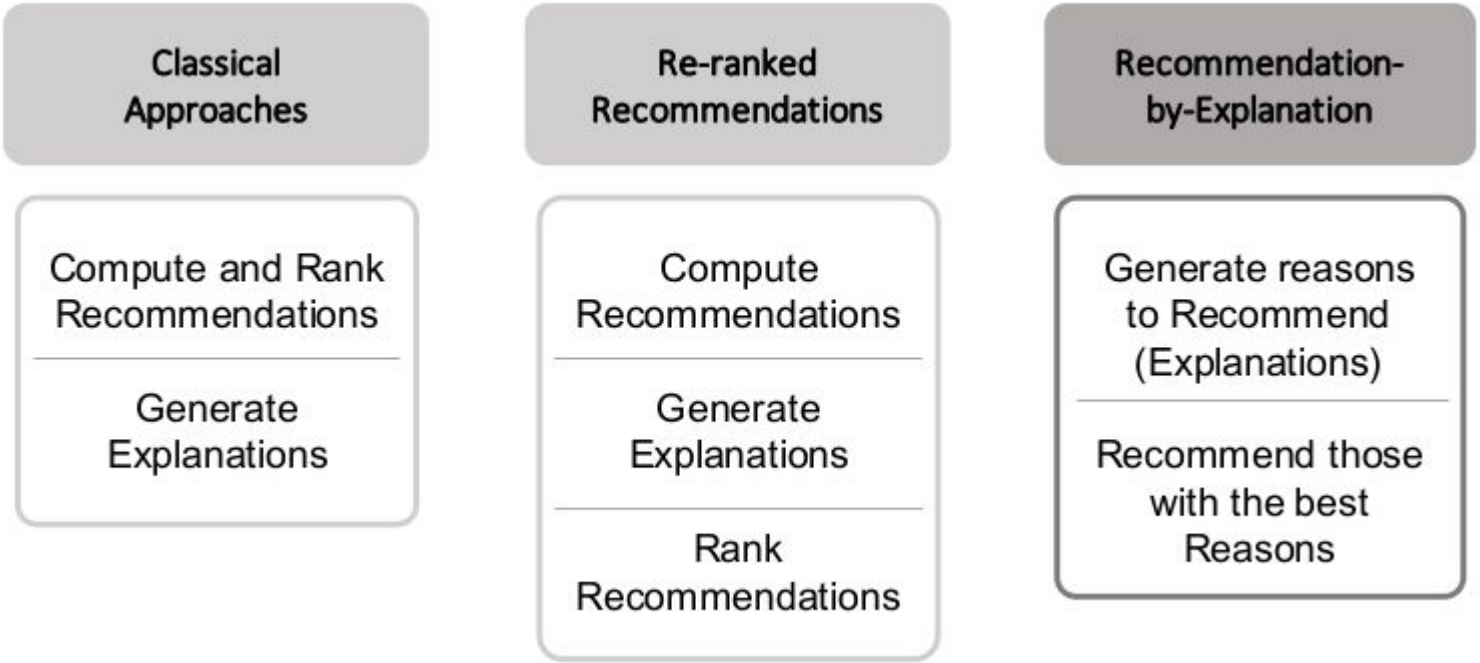
- Black-box (a.k.a. *model-agnostic*)

  These explanations make no use of knowledge of how the system produced its decision, e.g. matrix factorization for recommendation

Based on their Fidelity



Based on their Role in Producing Recommendations



Evaluating Explanations of Recommendations

Explanations are user-centric.

- Offline experiments are of very limited use; for example, we can measure the size of a system's explanation (e.g. how many items or features they contain).
- To evaluate the subjective perception of the users and their impact on user behaviour really requires either user trials or online evaluation with a deployed system.

Next lecture -

Explaining Recommendations