

# paper-rec

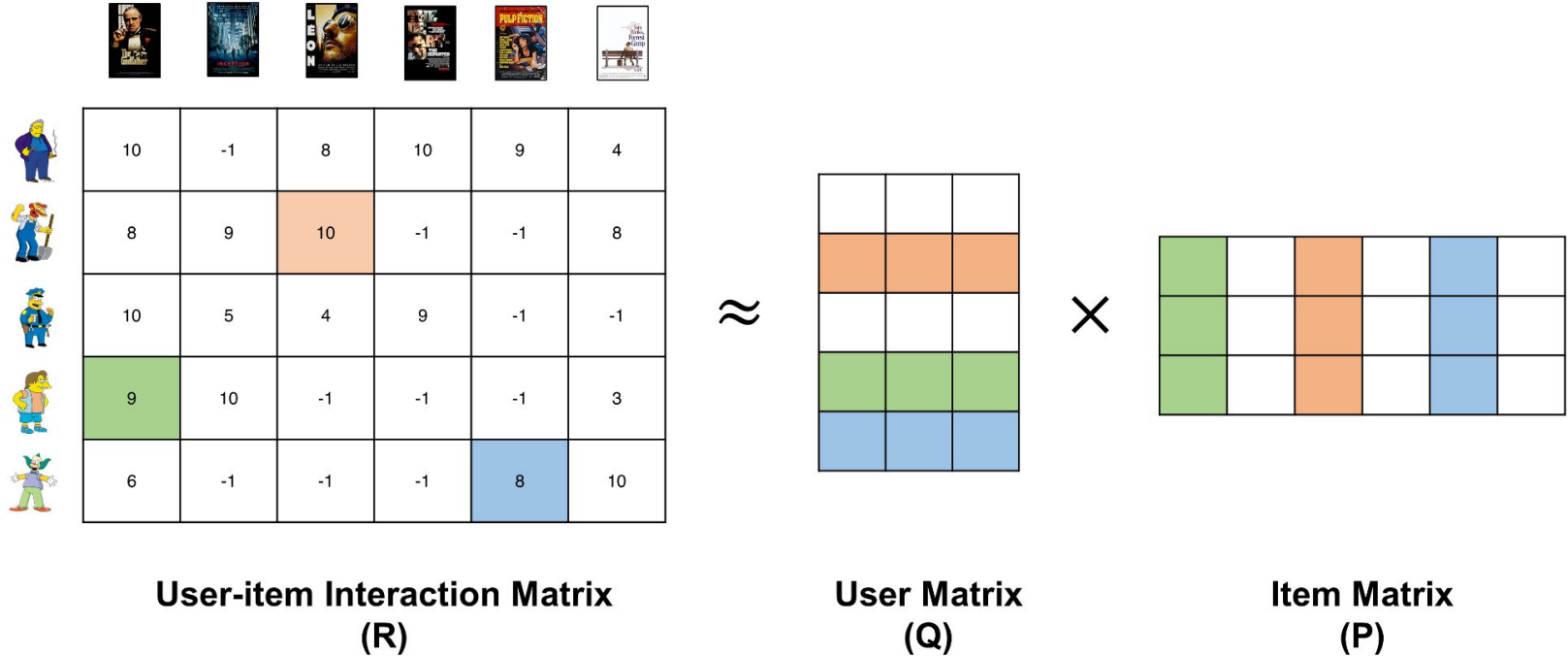
notes and demo

Dr. Ernesto Diaz-Aviles  
ernesto@libreai.com  
2022-10-26

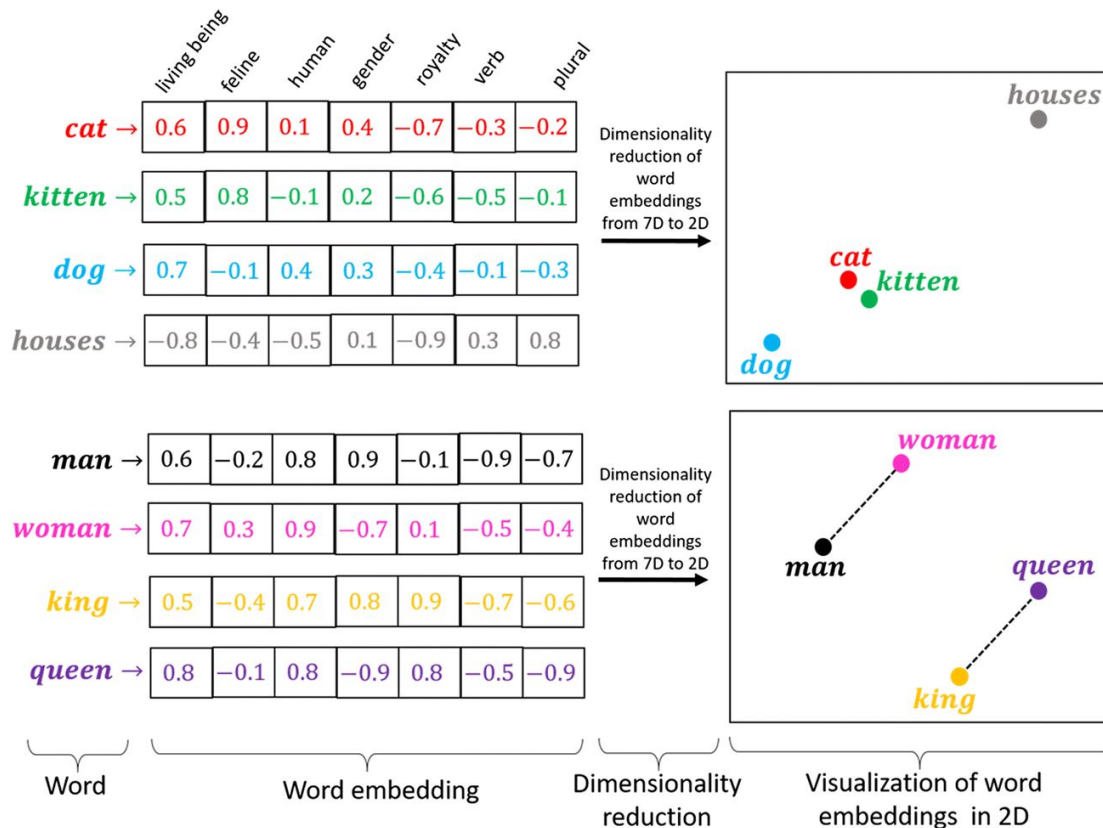
## `paper-rec`

- In **recommender systems**, the idea is to model **users** and **items** using **feature vectors** (i.e., **embeddings**) that represent their characteristics, and then compute scores based on operations on those feature vectors (e.g., dot product), the computed score is used to create a ranked list of personalized items for a given user
- In the case of **items**, such (latent) features represent properties of the objects, e.g., for movies they may correspond to metadata such as genre, cast, director, or to content features extracted from the video and audio feed
- For **users** the embedding represents their **preferences based on their history**, e.g., what users have bought, read, or listened to before helps to capture their preferences over time

# Embeddings Using Collaborative Filtering

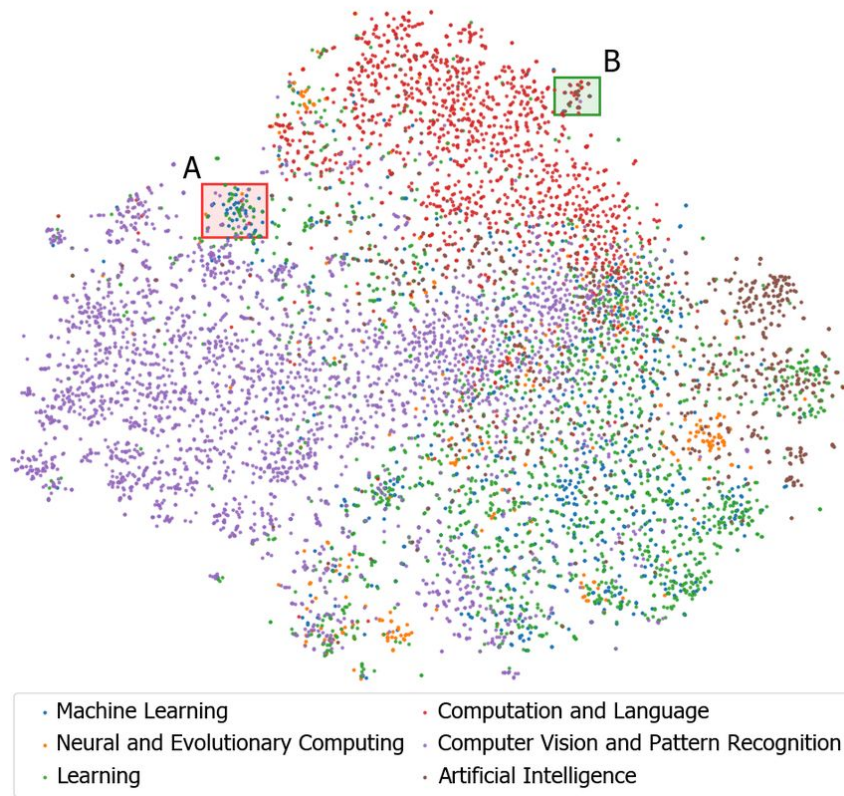


# Content-based Embeddings



# Content-based Embeddings

Embeddings for arXiv papers (6 ML categories)



## `paper-rec`

- In **recommender systems**, the idea is to model **users** and **items** using **feature vectors** (i.e., **embeddings**) that represent their characteristics, and then compute scores based on operations on those feature vectors (e.g., dot product), the computed score is used to create a ranked list of personalized items for a given user
- In the case of **items**, such (latent) features represent properties of the objects, e.g., for movies they may correspond to metadata such as genre, cast, director, or to content features extracted from the video and audio feed
- For **users** the embedding represents their **preferences based on their history**, e.g., what users have bought, read, or listened to before helps to capture their preferences over time

# `paper-rec`

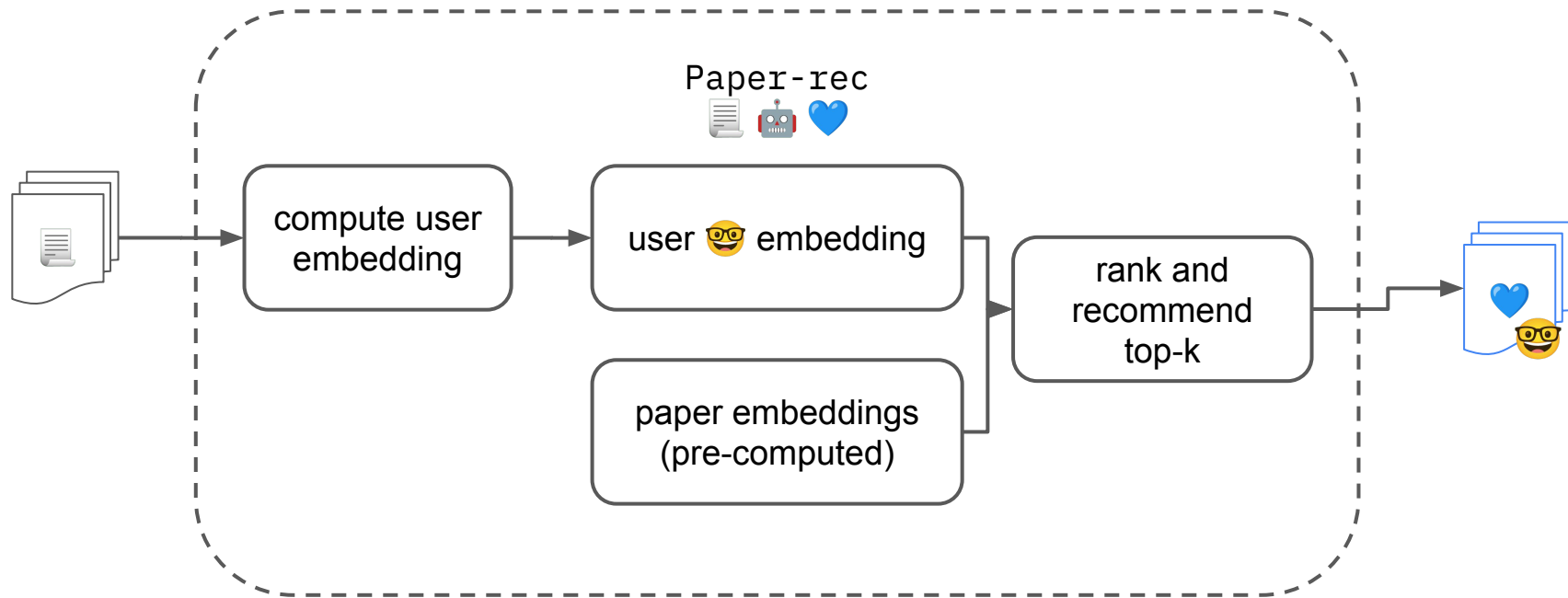
- **What paper in ML/AI should I read next?** It is difficult to choose from all the great research publications published daily. The goal of **paper-rec** is to offer a personalized selection of papers from the latest scientific contributions available
- Data: arXiv RSS feeds of latest contributions in particular:
  - Artificial Intelligence
  - Computation and Language
  - Computer Vision and Pattern Recognition
  - Information Retrieval
  - Machine Learning (cs)
  - Machine Learning (stat)
- Use [sentence-transformers](#) as feature extractor

## `paper-rec` : representing items and users

- Represent each **paper** (item) using an embedding computed using a **SentenceTransformer** from the sentences in the title and abstract of the paper
- For the **users**: represent their preferences as an embedding computed on extracts of text from articles they like, i.e., apply a **SentenceTransformer** on sentences extracted from text that users like



# `paper-rec`



# Conclusion and Outlook

- `paper-rec` recommends articles relevant to user interest
- Limitations:
  - Paper-rec follows a content-based approach, however, combining collaborative filtering based on the observed interactions, e.g., when user clicks on a recommendation, to fine tune the embeddings could yield a better performance. In the demo is not possible to keep track if the users click or not on a recommendation and to use this information
  - Integrating a recsys library as part of the Hugging Face Hub Libraries outlined here <https://huggingface.co/docs/hub/libraries> would be an interesting exercise and something we would like to explore with more time. Currently, the integration is to support downstream capabilities, but not the tasks associated with the libraries integration in the Hub

## Conclusion and Outlook ctd.

- Interesting to explore: experiments on the embeddings extracted by the sentence transformer vs BERT using a recsys as downstream task, i.e., measure recommendation quality
- **recsys**  $\neq$  NLP, CV, Audio tasks

# Output

- ``paper-rec`` recommender library using huggingface\_hub library for downstream support: <https://github.com/bluebalam/paper-rec>
- Dataset 🙌: <https://huggingface.co/bluebalam/paper-rec>
- Demo 😊: <https://huggingface.co/spaces/bluebalam/paper-rec>