Recommender systems

Alaa BAKHTI - OCTO Technology

alaa.bakhti@epita.fr

28 / 11 / 2020

Who Am I?

EPITA 2018

Machine learning engineer at OCTO Technology

Interested in:

- Data Science
- Software craftsmanship
- AI on production

What about you?

- Your profile.
- Any knowledge about RecSys?
- Your expectations for the course.

The course

- 5 sessions - 21 hours
- Each session = 2h course + 2h30 practical work
- Grading
    - Weekly reflections
    - Practical work
    - Project
    - Exam
- Bonus
    - Participation

Syllabus

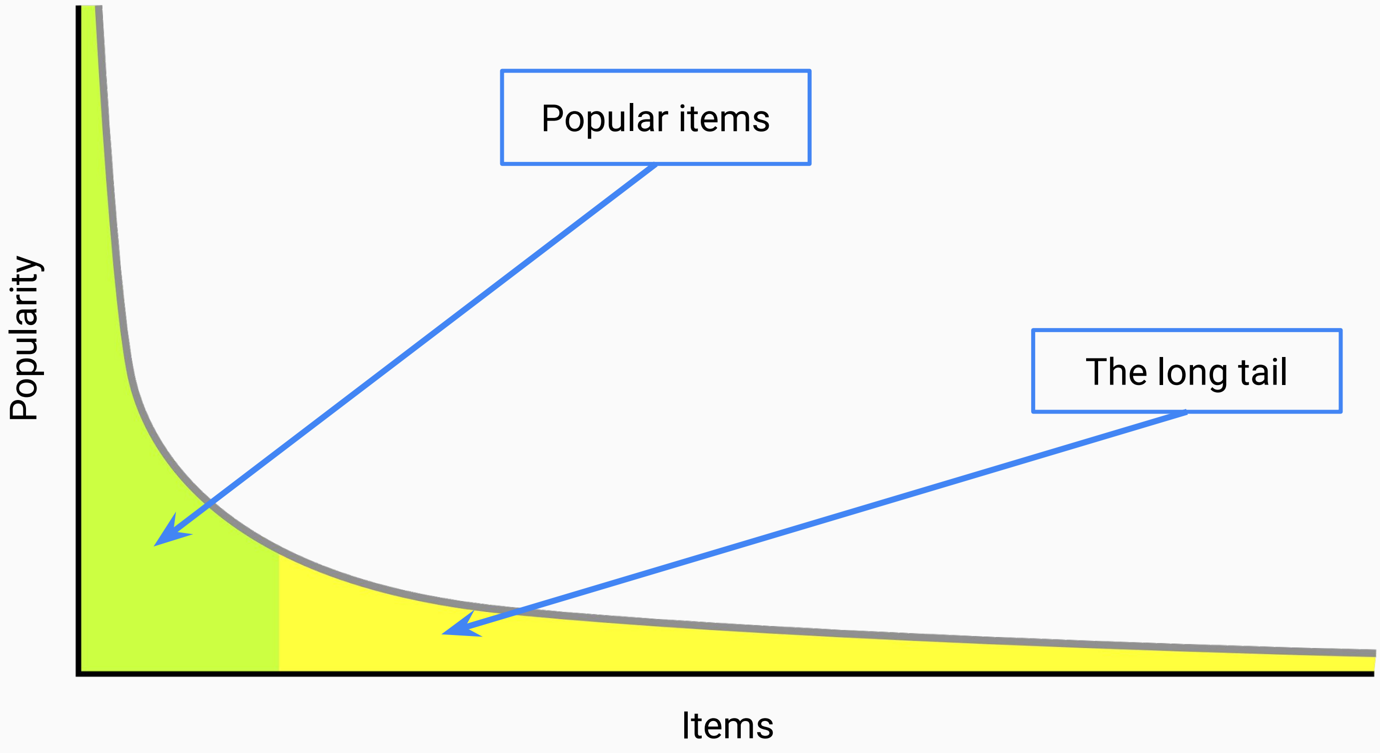
- *Session 1:* Introduction to recommender systems
- *Session 2:* Content-based filtering
- *Session 3:* Model-based collaborative filtering
- *Session 4:* Memory-based collaborative filtering
- *Session 5:* Evaluation of recommender systems and Application of deep learning in this domain.

Do you know?

TF IDF
Embedding space
word2vec
Cosine similarity
SVD
NMF
EDA
Loss function
L2 regularization
AB testing?

Introduction

The long tail problem



Picture by Hay Kranen / PD

Information overload

The paradox of choice

Source: Mark Rowland - Your marketing rules

The paradox of choice by Barry Schwartz (book review) - Youtube video

Recommender systems

- Help users find compelling content in a large corpora.
- Reduce information overload by estimating relevance.
- Personalise the user experience.

Applications and business value

Many domains where the recommender systems can be used

Where the RecSys is used?

- E-commerce websites
- Search engines
- Social networks
- Movie or music streaming sites
- mobile app stores
- etc

Netflix: movie recommendation

75%

of the watched content is from some sort of recommendation

Source: Netflix recommendations: Beyond the 5 stars - X. Amatriain and J. Basilico - Netflix Inc - 2012.

Netflix: movie recommendation

$1B

per year is the estimated business value of recommendation

Source: The Netflix recommender system: Algorithms, business value, and innovation - C. A. Gomez-Uribe and N. Hunt - Netflix Inc - 2015.

Amazon: product recommendation

35%

of Amazon sales originate from cross-sales (recommendation)

[Source: How retailers can keep up with consumers - Mckinsey & Company - 2013.]

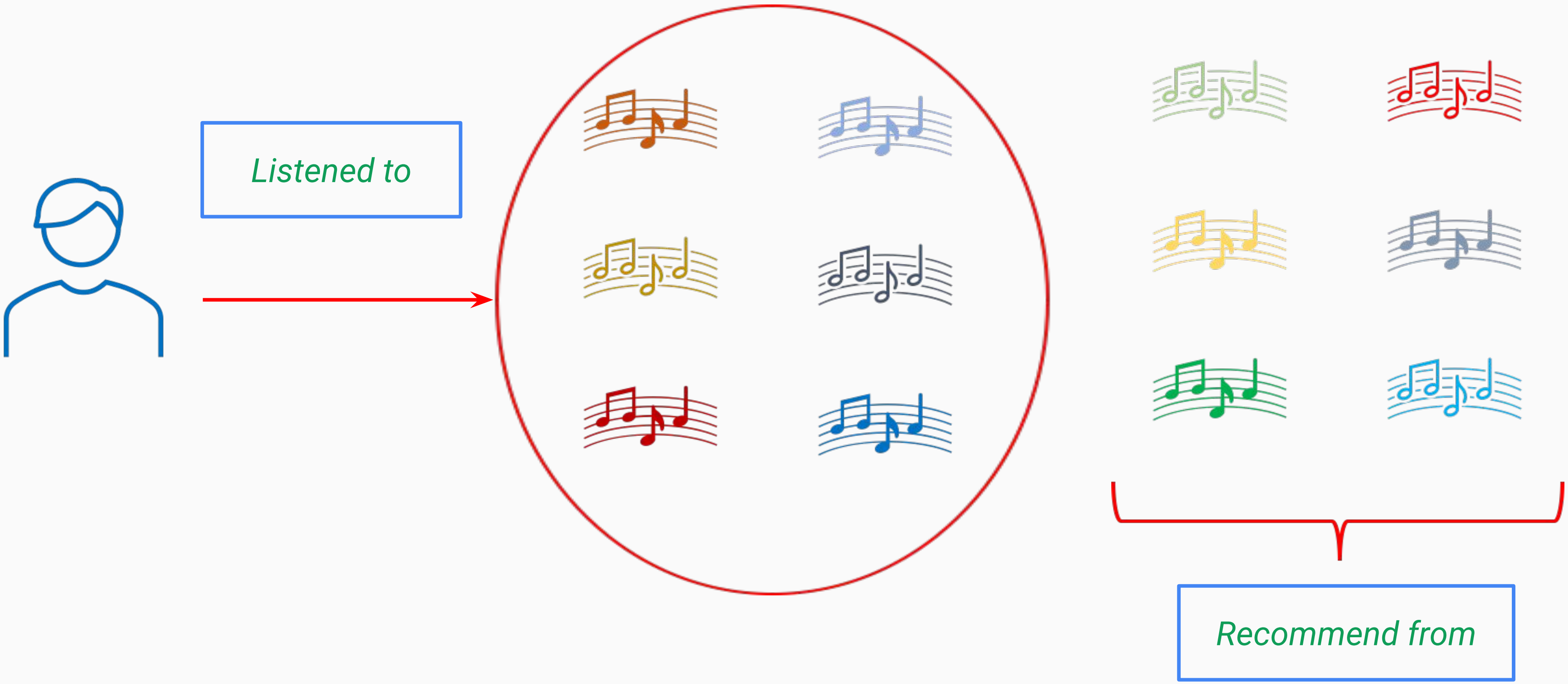
Youtube: video recommendation

60%

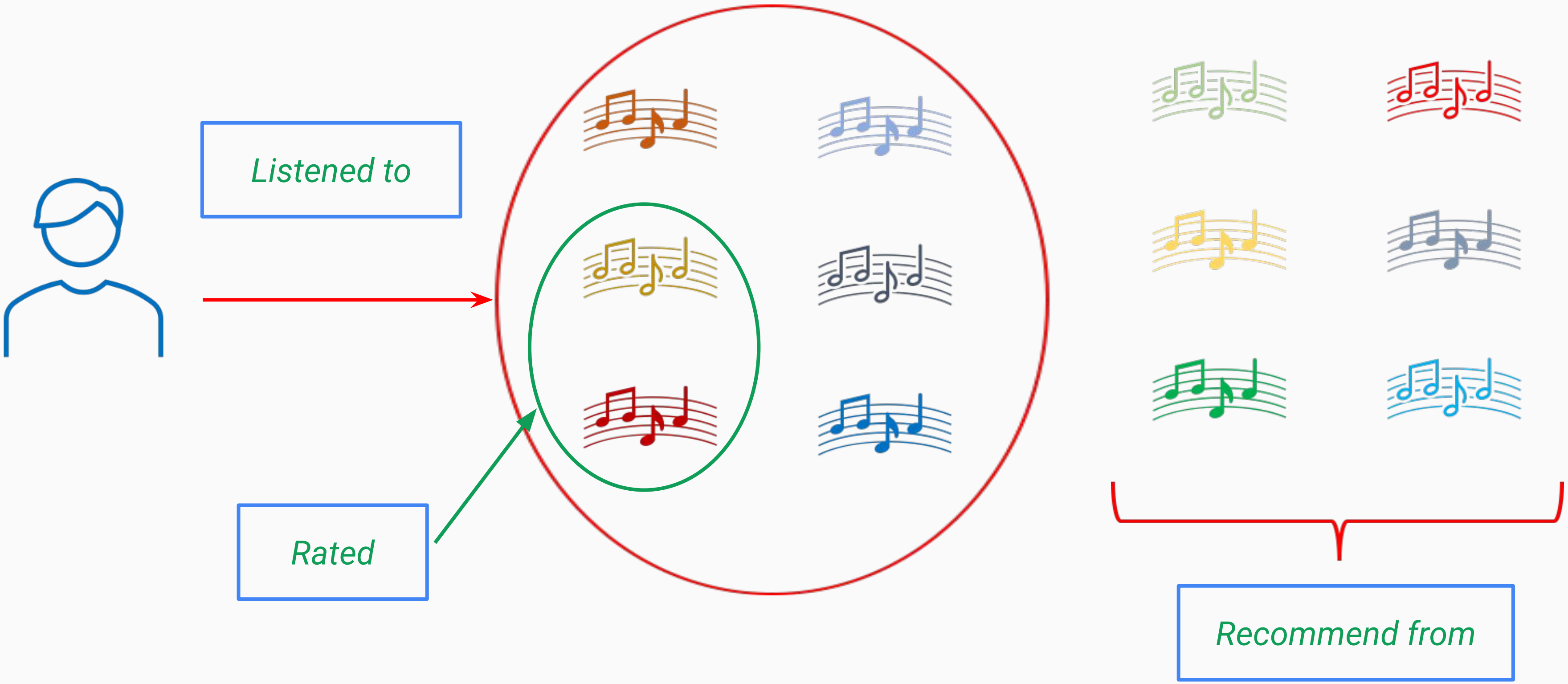
of the clicks on the home screen
are on the recommendations

[Source: The YouTube Video Recommendation System - J. Davidson & al - Google Inc - 2010.]

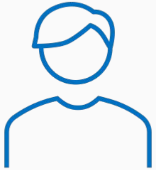
Problem formulation



How to determine items that the user may be interested in?



Rating matrix



					
4/5	2/5	?	?	?	?

How to determine the user rating of items he didn’t ***explicitly*** rate?

User interactions feedback

Explicit

- Data provided by users intentionally.
- Example : Press the like button on a YouTube video.
- Problem : it requires effort from the user => doesn't scale.

Implicit

- Data generated based on the user interaction with items (easier to collect).
- Example : purchased an item => high rating.
- Problem : poorly learns low ratings (what the user doesn't like).

Recommender systems use the combination of explicit and implicit user feedbacks.

NETFLIX

How to determine the user ratings for items he interacted with?

How to extrapolate the user ratings for items he didn’t interact with?

Weekly reflection

- Implicit feedback in Recommender systems
- Recommender system architecture design

Practical work

Subject

- Exploratory Data Analysis (EDA) on the Movies dataset
- Final dataset will be used in the next sessions

Grading criteria

- Respect of the submission instructions (git, repository structure, documentation, quoting ressources, etc)
- Logical sequence of exploration (illustrate problems and then resolve them)
- Comment the identified problem and the solution you propose
- Focus your exploration on the user-movie recommendation use case
- Saved the processeed dataset with the newly generated features at the end of the notebook
- Notebook presentation (Titles, spelling mistakes, etc)