**Slovak University of Technology in Bratislava**
**Faculty of Informatics and Information Technologies**

Ákos Lévárdy

Odporúčacie systémy založené na AI

Bachelor thesis

July 3, 2024

Degree course: Informatics
Field of study: 9.2.1 Informatics
Place: FIIT STU, Bratislava
Supervisor: PaedDr. Pavol Baťalík

# ANNOTATION
500 words

# ANOTÁCIA

500 slov

**DECLARATION OF OATH**

I hereby declare upon my honour that I wrote this thesis single-handed with usage of quoted literature and based on my knowledge and professional supervision of my supervisor.

**ACKNOWLEDGMENT**

First and foremost, I would like to thank my supervisor for their invaluable guidance and support throughout the duration of this project.

# Contents

# 1 Introduction

As Internet and Web technologies continue to evolve rapidly, the amount of information available online has expanded excessively across sections such as e-commerce, e-government, and e-learning. To help users navigate this vast sea of content, Recommender Systems have become fundamental. They are very effective tools for filtering out the most appropriate information any user would like to find. The primary focus of these recommendations is to predict if a specific user will be interested in the distinct items.

Recommendation types are divided to 3 different categories, which are Content-Based Filtering approaches (CB), Collaborative Filtering approaches (CF) and Hybrid approaches which are the combinations of the two. Both CB and CF approaches encounter a significant challenge known as the 'cold-start problem'. This problem arises when making recommendations to new users and/or items for which the available information is limited. As a result, the recommendations offered in such cases tend to be of poor quality and lack usefulness. [1]

Content-Based Filtering
- creates a profile for every user or item and characterizes them.

Collaborative Filtering
- analyzes relationships between users and interdependencies among products to identify new user-item associations
- recommend items or content to users by analyzing their interactions and similarities with other users

[2]

**2 Content-Based Filtering**

**3 Collaborative Filtering**

**4 Hybrid approach**

# 5 Semantics

# 6 Ontology

# 7 Matrix Factorization

Matrix factorization (MF) is a technique utilized in collaborative filtering to decompose a matrix of user-item ratings into lower-rank matrices capturing the latent factors under- lying the data [3].
People prefer to rate just a small percentage of items, therefore the user-item rating matrix, that tracks the ratings people assign to various items, is frequently sparse. In order to deal with this sparsity, matrix factorization (MF) algorithms split the matrix into two lower-rank matrices: one that shows the latent properties of the items and another that reflects the underlying user preferences. These latent representations can be used to predict future ratings or complete the matrix's missing ratings after factorization.

# 8 Search Engines

Search Engines have become crucial for navigating the vast amount of information available online. They make it possible for people to quickly look up solutions, learn new things, and browse the wide variety of resources available on the internet. Search engine optimization is now necessary to guarantee that search engines deliver relevant results, quick search times, and a top-notch user experience given the explosive growth of online information.
A search engine is essentially a software that finds the information the user needs using keywords or phrases.

It delivers results rapidly, even with millions of websites available online. The importance of speed in online searches is highlighted by how even minor delays in retrieval can negatively affect users' perception of result quality. [4]

# 9   Concept Drift

Concept Drift - Entropy-based

Information systems inevitably experience frequent data changes. This change in the statistical properties of the target variable, caused by unforeseeable variations in the underlying distribution of the data stream, is known as concept drift. [5]

# References

[1] Malak Al-Hassan, Bilal Abu-Salih, Esra'a Alshdaifat, Ahmad Aloqaily, and Ali Rodan. An improved fusion-based semantic similarity measure for effective collaborative filtering recommendations. 17(1), 2024. doi:10.1007/s44196-024-00429-4.

[2] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. 42(8):30 – 37, 2009. doi:10.1109/MC.2009.263.

[3] Srilatha Tokala, Murali Krishna Enduri, T. Jaya Lakshmi, and Hemlata Sharma. Community-based matrix factorization (cbmf) approach for enhancing quality of recommendations. 25(9), 2023. doi:10.3390/e25091360.

[4] Serge Stephane AMAN, Behou Gerard N'GUESSAN, Djama Djoman Alfred AGBO, and KONE Tiemoman. Search engine performance optimization: methods and techniques. 12, 2024. doi:10.12688/f1000research.140393.3.

[5] Yingying Sun, Jusheng Mi, and Chenxia Jin. Entropy-based concept drift detection in information systems. 290, 2024. doi:10.1016/j.knosys.2024.111596.

[6] Mehrbakhsh Nilashi, Othman Ibrahim, and Karamollah Bagherifard. A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. 92:507 – 520, 2018. doi:10.1016/j.eswa.2017.09.058.

[7] Hongbo Wang, Yizhe Wang, and Yu Liu. A sequential recommendation model for balancing long- and short-term benefits. 17(1), 2024. doi:10.1007/s44196-024-00460-5.

[8] Ben Halstead, Yun Sing Koh, Patricia Riddle, Russel Pears, Mykola Pechenizkiy, Albert Bifet, Gustavo Olivares, and Guy Coulson. Analyzing and repairing concept drift adaptation in data stream classification. 111(10):3489 – 3523, 2022. doi:10.1007/s10994-021-05993-w.

[9] Shilpa S. Laddha and Pradip M. Jawandhiya. Semantic search engine. 10(21):1–6, 2017. doi:10.17485/ijst/2017/v10i23/115568.

[10] Dirk Lewandowski. *Understanding Search Engines*. 2023.