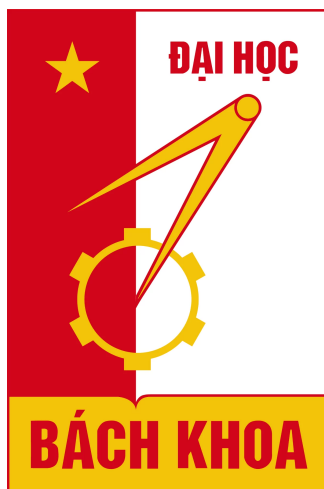


HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION
TECHNOLOGY



PROPOSAL FOR COURSE PROJECT

INTRODUCTION TO ARTIFICIAL INTELLIGENT
(IT3160E)

Book Recommendation System Using Collaborative Filtering

Team members

Vu Duc An 20215174
Le Ha Ngan 20215230
Dang Viet Hoang 20215206
Trinh Diem Quynh 20210737

Email

an.vd215174@sis.hust.edu.vn
ngan.lh215230@sis.hust.edu.vn
hoang.dv215206@sis.hust.edu.vn
quynh.td210737@sis.hust.edu.vn

Class

136462 - IT3160E

Lecturer

Nguyen Nhat Quang
quang.nguyennhat@hust.edu.vn
quangnn@soict.hust.edu.vn

November 1, 2022

Time period: 1st Nov - 15th Dec

I Describe the practical problem:

- **Goal:** Build an intelligent recommender systems that can learn to give book recommendations using information collected from users.
- **Important characteristic:** The dataset must have some characteristics extracted from books:
 - Information related to books (author, title, genre, etc).
 - Users (users id, location, age).
 - Ratings (which user has given how much rating to which book).
- **Application scenarios:**
 - Shopping websites / ebook websites to recommend similar books for readers or help people find good books in a genre.
 - Group books that people like near each others in bookstores.

II Explain intuitive idea/direction:

- * **Basic idea:** Find other users whose past ratings or interactions are similar to that of the current user and use their item preferences to predict what the current user will like.

Therefore, a **Collaborative-based filtering system** should be used!

- **Step 1: Data Preparation:**
 - Import the necessary libraries and load the datasets.
- **Step 2: Data Preprocessing, Data Cleaning:**
 - Remove redundant or irrelevant data such as publishing date, format, cover image, number of pages, etc.
 - Avoid repetitions by creating an ID for each book and clearing duplicate ratings.
- **Step 3: Exploratory Data Analysis:**
 - Identify the interesting aspects of the books datasets: distribution of ratings, number of ratings per user, different languages, the average rating, and multiple editions of each book.
- **Step 4: Build Recommender:**
 - Choose user-based (user-user) collaborative filtering.
 - Define similarity between users (in terms of ratings on same set of books) using memory-based approach with matrix.
 - Normalize the missing ratings that haven't been provided in the matrix.
 - Compute the similarity using cosine similarity function.
 - Use KNN to make rating predictions for each book.
 - Recommend a number of books with the highest average rating to the current user.
- **Step 5: Evaluate Recommender System:**
 - Assess the performance using a combination of multiple evaluation metrics (MAE, Coverage, Diversity, etc)

III Explain the input and output of system:

- **Input:** Readers data, books data and book reviews
- **Dataset to be used:** <http://www2.informatik.uni-freiburg.de/~ctiegle/BX/>
- **Output:** A recommended list of similar books

IV Execution plan:

- **Start date:** 1/11/2022
- **End date:** 15/12/2022

