

Частина 1. Лабораторна робота №3, Рекомендаційні системи

1. ЗАГАЛЬНІ ВІДОМОСТІ

Даний документ містить інформацію щодо лабораторної роботи №3 зі статистичних алгоритмів навчання. Всього передбачено 12 варіантів. Ваш варіант відповідає Вашому номеру у списку з журналу (якщо Ваш номер більше 12, то відніміть від Вашого номера 12).

В кожному варіанті буде вказано посилання по якому можна скачати датасет для аналізу а також додаткову інформацію.

Завдання полягає у тому, щоб побудувати рекомендаційну систему.

1. Зчитайте дані, та перетворіть, за потреби у форму матриці (можливо з пропущеними даними).
2. Імплементуйте алгоритм ймовірнісної матричної факторизації в якому K -буде параметром.
3. Підберіть найкраще K .

2. ВИМОГИ ДО ВИКОНАННЯ, ОФОРМЛЕННЯ ТА ЗДАЧІ

Роботу можна виконувати в R або Python. Вибірку слід розбити на тестову та тренувальну.

Робота має бути оформлена у вигляді .pdf (можливо .doc/.docx) файлу який містить всю необхідну інформацію. Роботи потрібно здавати на парах або надсилати на email.

Кожна робота буде розглядатися на відповідність критеріям описаним вище, та на обґрунтованість прийнятих рішень. Кожен студент, повинен виконати свою роботу самостійно. Ідентичні, або майже ідентичні роботи прийматися до уваги не будуть.

3. ВАРІАНТИ

Варіант 1

Опис даних: <http://www2.informatik.uni-freiburg.de/cziegler/BX/>

Самі дані: <http://www2.informatik.uni-freiburg.de/cziegler/BX/BX-CSV-Dump.zip>

Варіант 2

Опис даних: <http://www.occamlab.com/petrick/data/>

Самі дані: <http://www.occamlab.com/petrick/data/ratings.dat>

Варіант 3

Опис даних: <http://files.grouplens.org/datasets/hetrec2011/hetrec2011-movielens-readme.txt>

Самі дані: <http://files.grouplens.org/datasets/hetrec2011/hetrec2011-movielens-2k-v2.zip>

Варіант 4

Опис даних: <http://files.grouplens.org/datasets/hetrec2011/hetrec2011-lastfm-readme.txt>

Самі дані: <http://files.grouplens.org/datasets/hetrec2011/hetrec2011-lastfm-2k.zip>

Додаткова інформація: Використовуйте кількість прослуховувань в якості метрики (рейтингу)

Варіант 5

Опис даних: <https://www.librec.net/datasets.html>, CiaoDVD

Самі дані: <https://www.librec.net/datasets/CiaoDVD.zip>

Варіант 6

Опис даних: <https://www.kaggle.com/tamber/steam-video-games/data>

Самі дані: <https://www.kaggle.com/tamber/steam-video-games/downloads/steam-video-games.zip/3>

Варіант 7

Опис даних: <http://www2.informatik.uni-freiburg.de/cziegler/BX/>

Самі дані: <http://www2.informatik.uni-freiburg.de/cziegler/BX/BX-CSV-Dump.zip>

Варіант 8

Опис даних: <http://www.occamslab.com/petricek/data/>

Самі дані: <http://www.occamslab.com/petricek/data/ratings.dat>

Варіант 9

Опис даних: <http://files.grouplens.org/datasets/hetrec2011/hetrec2011-movielens-readme.txt>

Самі дані: <http://files.grouplens.org/datasets/hetrec2011/hetrec2011-movielens-2k-v2.zip>

Варіант 10

Опис даних: <http://files.grouplens.org/datasets/hetrec2011/hetrec2011-lastfm-readme.txt>

Самі дані: <http://files.grouplens.org/datasets/hetrec2011/hetrec2011-lastfm-2k.zip>

Додаткова інформація: Використовуйте кількість прослуховувань в якості метрики (рейтингу)

Варіант 11

Опис даних: <https://www.librec.net/datasets.html>, CiaoDVD

Самі дані: <https://www.librec.net/datasets/CiaoDVD.zip>

Варіант 12

Опис даних: <https://www.kaggle.com/tamber/steam-video-games/data>

Самі дані: <https://www.kaggle.com/tamber/steam-video-games/downloads/steam-video-games.zip/3>