Coursework Assignment:

Designing and Developing
a Personalised Recommender System

Overview

Lecturer/Marker

Suncica Hadzidedic
suncica.hadzidedic@durham.ac.uk
Room E231

Hand-out to students: 20 November 2020

Type: summative assessment
Level: 3
Components marked: report, application (with a video demo)
Total marks: 100
Weight of module mark: 100%
Expected workload: 15 days, 3-4 h/day = 60 h

Submission instructions

Submission deadline	**8 February 2021; 14:00**
Format	• DUO – **Turnitin:** report/**documentation**, as one PDF file. • **DUO** – compressed (.zip or .rar) file to include: video, source code**,** resource files (including datasets[1]), readme .txt

I. Requirements

You are required to submit a single piece of coursework for the Recommender System module which is worth 100% of your final module mark.

The task is for you to design and develop a **hybrid recommender system** using the **Yelp dataset**. Yelp.com provides user reviews (ratings, text feedback, etc.) of different businesses and services in a specific location. You are to:

I.   Select a business/service category (e.g. bars, health, accountants, landscapers, etc.) as an application domain for which you will design and develop an RS – you can select one, or several related services. Note: the number of services you include will not influence your mark.

---

[1] I have access to the overall Yelp dataset. Submit **only** the sampled and pre-processed dataset that you have used for your recommender system.

II. Define the time-span for your data – you do not need to use the whole Yelp dataset, you can use a sample, determined by your time-span and business category/type.

III. Take into account Covid-19 data in generating recommendations (also available in Yelp data).

Your coursework should meet the following 4 categories of requirements:

1. System design and development

▪ ***Programming language and testing environment***:
  - *Python 3.8.3*
  - available packages are listed in the accompanying file "ListofModules.txt" (ask before using additional packages/modules);
  - include comments in the code, explaining which part performs which part of the process;
  - readme.txt should state which is the main file (file to be run first).
  - Your RS solution will be tested on: laptop (1.7 GHz, 8 GB RAM); Windows 10 OS; using Anaconda 3 Prompt.

▪ ***Dataset***:
  - Address data preparation and feature selection/weighting
    o you are free to choose the methods, but you have to justify why these are appropriate for your data and application domain.

▪ ***Hybrid RS***
  - Mandatory: **two** personalised **recommenders,** one of which is: collaborative filtering.
    o Optional[2]: additional recommenders.
  - Hybrid **scheme** (i.e.: weighted, mixed, etc.): you are free to choose the scheme, however justify the selection.
  - Recommender 1 and 2 algorithms: you are free to choose a learning/prediction method/algorithm, but you have to provide a justification.
    o Optional: for an additional challenge try a *deep learning method.*

▪ ***Methods/techniques***:
  - You are to research, select and justify (in the <u>report</u>) all the methods (i.e., hybrid scheme, feature weighting, algorithms used, etc.) used for your hybrid RS, aligning the applicability and relevance of the selected methods to the purpose, domain of application and data used for this recommender system.

▪ ***User interface:***
  - This should be a **command line** based interface.
    o Note: Do not develop graphical, web-based interfaces, as these will not be marked.
  - Mandatory to account for:
    o *Input*: How does the system recognise the active user? Which user data is gathered – explicit/implicit? Are users aware which data is collected, how and for which purposes? Updating of user profiles?
    o *Output*: How are recommendations (and prediction scores) presented to the user? Consider target user needs, in terms of e.g.: number of recommendations presented; style of presentation; context – environment, device used; user

---

[2] Including the optional components will **not** result in additional marks.

characteristics – age, disabilities, etc. Explainability/transparency - within the output presentation, include *explanations* for why the specific set of recommendations is presented.

- Make sure to reference any external sources you have used for the code, data, algorithm logic, etc.

2. Evaluation

- Evaluate the performance of your RS by carrying out an **offline experiment**.
- Apply <u>one evaluation metric</u> from each of the four categories listed below. You are to justify why each of the selected metrics is appropriate for the purpose of the RS and its domain of application.
    - i.   Accuracy of rating predictions
    - ii. Accuracy of usage predictions
    - iii. Novelty, diversity, coverage
    - iv. Explainability
- **Ethics**
    - You are to identify 3 ethical issues that your recommender system might exhibit including those that result from: data collection (experiments, user awareness, etc.), data storage, algorithms used, output presentation, business/provider interests, etc.
    - Briefly discuss approaches that can be used to address/solve these 3 ethical issues.
- Compare your hybrid RS against a **baseline RS**, e.g. hybrid RS against an only collaborative-filtering RS.

3. Report
The report should be up to 3 pages (including references) and use IEEE conference paper formatting[3]. It should include all of the following sections:

- Introduction
    - Domain of application
    - Related work review
    - Purpose/Aim
- Methods
    - Data description
    - Data preparation and feature selection
    - Hybrid scheme
    - Recommendation techniques/algorithms
    - Evaluation methods
- Implementation
    - Input interface
    - Recommendation algorithm[4]
    - Output interface[5]
- Evaluation results
    - Comparison against *baseline implementation*

---

[3] https://www.ieee.org/conferences/publishing/templates.html
[4] N*ote*: illustrate and describe the process
[5] N*ote*: Present the recommendations (and prediction/matching or other scores shown to the use)

- Comparison against *hybrid recommenders in related studies*
- Ethical issues
▪ Conclusion
- Limitations
- Further developments
▪ References

4. Video

The video should showcase your hybrid RS. In up to **2** minutes, you are to demonstrate:

- Input interface: how the user inputs data to the system.
- Output interface: how recommendations are presented to the user and any other interaction (e.g. feedback) a user is allowed at this stage.
- Back-end: explain how recommendations are generated – illustrate the hybrid RS phases and process.

II. Marking Criteria

The marks are distributed to three marking criteria, as presented in the table below. Make sure you covered all the requirements within each of the categories.

Design and development methods	40
Use of, description and justification of methods for: data preparation, hybrid scheme, model learning and prediction methods, user interface (input, output)	
System evaluation	**35**
Experiments, evaluation metric selection and justification for each of the 4 evaluation categories, identification and addressing ethical issues, comparison of results against a baseline and literature	
Presentation	**25**
Report, video	
TOTAL	**/100**

III. Learning Outcomes

Subject-specific Knowledge demonstrated via:
- an understanding of the different types of recommender systems, their purpose and domains of application
- an understanding of recommender system users: usage behaviour, demographics, preferences, contextual information
- an in-depth knowledge of recommender system algorithms, specifically hybrid techniques.
- an understanding of recommender system evaluation methods.

Subject-specific and Key Skills demonstrated via:
- an ability to undertake self-study and independent research
- an ability to critically analyse and evaluate state of the art practices
- an ability to apply RS methods and techniques
- an ability to implement a recommender system for a specific domain
- an ability to evaluate RS performance, including any ethical issues.