Design of AI Systems (DAT410)
Project proposal for assignment 8

Luca Modica
Hugo Manuel Alves Henriques e Silva

February 27, 2024

1 Goals and outcomes

The main goal of this project is to build a sophisticated recommender system: It will first cluster users and items for a better understanding of patterns and preferences in the data; this followed by a suitable neural network to generate personalized recommendations. One of our focus will be to use AI tools capable of textual data processing: the purpose is to offer even more precise recommendations based on user reviews and items descriptions, alongside a more interpretable decision-making process.

2 Motivation

Enhancing the quality of recommendations of a system like described above, alongside a better understanding of the motivations behind such recommendations, are among our main motivations behind this project. Moreover, this work aligns with current interest in developing and delve deeper into insights about data (of a specific domain knowledge), AI and its decision-making mechanisms.

3 Relevant modules of the course

- **AI problem solving**: the development of a recommender system involving text processing and clustering of data is considered as a well-known AI problem.
- **Recommendation systems**: this module represents the core of our project, since we want to build and enhance a system of this type.
- **AI tools**: as mentioned, our work will aim to involve different type of tools, from a clustering algorithms to a neural network model.

- **Diagnostic system**: aligned with one of the main topic of this model, one of our key considerations will be the interpretability of the system's recommendation.
- **Natural language processing**: relevant textual information (such as text reviews) will be processed and analyzed, in order to get more data insights and improve the quality of the system.

4 Preliminary technical ideas

Our idea is to divide the project into the following steps.

1. **Data acquisition**: making sure to secure a dataset for the purposes described above. 2 suitable examples can be the "Amazon product data" dataset [1][4] and the Microsoft News Recommendation Dataset (or MIND) [6].
2. **Data analysis and clustering**: Exploratory Data Analysis (EDA) and use a clustering algorithm to group users and items. This for a first understanding of the data to be fed into the neural network model.
3. **Neural network model**: development of the actual hybrid neural network system. There will be 2 main components.
    - **Text analysis component**: the idea is to start by having a baseline with a TF-IDF vectorization of the texts. Then, a suitable fine-tunable model to process textual data would be used to extract the semantic embedding (such as SBERT [5]); we can follow a similar approach to Budi Juarto et al., and their related work with Neural Collaborative with Sentence BERT for News Recommender System [3].
    - **Collaborative filtering component**: a Neural Collaborative Filtering (NCF) to generate the latent features for both users ans items, aiming to an approach that can be similar to the one described by He et al. [2].
4. **Training and testing**: train and testing of the model (mainly model fine-tuning and evaluations, with the most common metrics for this specific type of system).
5. **Interpretability**: discussion about the interpretability of the model recommendations, from the previous clustering to possibly more sophisticated methods (post-hoc analysis frameworks and the sentimental analysis from the sentence embedding).

5 Project feasibility and timeline

- **Week 1 (Feb 26 - Mar 4)**: dataset selection, Exploratory Data Analysis and data preprocessing.
- **Week 2 (Mar 5 - Mar 11)**: data clustering, neural network design and model training.
- **Week 3 (Mar 12 - Mar 14)**: model evaluation, interpretability analysis, final report preparation.

References

[1] Ruining He and Julian McAuley. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proceedings of the 25th International Conference on World Wide Web*, 2016.

[2] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering, 2017.

[3] Budi Juarto and Abba Girsang. Neural collaborative with sentence bert for news recommender system. *JOIV : International Journal on Informatics Visualization*, 5:448, 12 2021.

[4] Julian McAuley, Christopher Targett, Jinfeng Shi, and Anton van den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2015.

[5] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks, 2019.

[6] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, and Ming Zhou. Mind: A large-scale dataset for news recommendation. *arXiv preprint arXiv:2006.04155*, 2020.