成都理工大學中英合作办学
CDUT Sino-British Collaborative Education

OXFORD BROOKES UNIVERSITY

UNDERGRADUATE PROJECT PROGRESS REPORT

Project Title:	**Book Recommendation using variational autoencoders**
Surname:	**Wang**
First Name:	**Yutong (Ryan)**
Student Number:	**202018010315**
Supervisor Name:	**Joojo Walker**
Module Code:	**CHC 6096**
Module Name:	**Project**
Date Submitted:	**September 27, 2023**

Table of Contents

1 Introduction

1.1 Background

In the past few decades, with the rise of YouTube, Amazon, Netflix, and many other web services, recommender systems have taken an increasingly important place in our lives. From e-commerce (recommending items to buyers that interest them) to online advertising (recommending content to users that matches their preferences) [1]. The development of e-commerce has facilitated shopping by providing a new way to shop. There has been a dramatic increase in the use of e-commerce websites because of rapidly advancing technology. The contribution of e-commerce to total retail sales increased from 5.1% in 2007 to 16.0% in 2019. Books are also one of the commodity categories in shopping platforms, and the modern book industry is characterized by an extremely rich variety, which gives readers numerous choices when purchasing books on online platforms while also making it increasingly challenging to identify and select books that match their preferences and interests [2]. Therefore, an effective personalized book recommendation system (RS) is very important. At the same time, the act of reading is beneficial to both individuals and society, but studies have shown a decline in reading among young people. Recommender systems can also help stop this decline [3].

1.2 Aim

The goal of this project is to establish a user-centric recommendation platform utilizing Variational Autoencoders (VAEs) as the foundation. Optimize the recommendation process using VAE's latent representation capabilities to improve the platform's ability to provide tailored and diverse book recommendations that are closely aligned with users' personal preferences and reading habits. By implementing the VAE-based model and comprehensively evaluating its performance using evaluation metrics, a highly accurate recommendation system is finally obtained.

1.3 Objectives

- Collect a diverse dataset of books and perform thorough data cleansing and preprocessing to ensure uniformity and suitability for model training.
- Evaluate and compare the performance of various models based on VAE by implementing them, assessing metrics including recommendation accuracy, diversity, and coverage, and conducting a comprehensive analysis to identify the strengths and weaknesses of each model.
- Integrate the selected VAE model into the recommendation platform.

- Conduct performance testing to validate the efficiency and effectiveness of the system in generating personalized book recommendations tailored to individual user preferences.
- Optimize and fine-tune the model parameters and architecture, iteratively refining the selected VAE model to enhance recommendation accuracy and ensure an enriched overall user experience.
- Compile a comprehensive report detailing the project's methodology, implementation, and findings, summarizing key insights, challenges, and recommendations for future enhancements to the recommendation system.

1.4 Project Overview

1.4.1 Scope

This research aims to explore the efficacy of variational autoencoders (VAEs) in enhancing book recommendation systems with a focus on providing highly personalized and diverse book recommendations to users. By testing and evaluating the model based on VAE, this project seeks to integrate the model into a recommendation platform that can provide customized book recommendations according to readers' unique reading preferences and habits. This can contribute to improving the reading experience of users.

The significance of this research is that it develops a new book recommendation system utilizing cutting-edge deep learning techniques. Using VAE facilitates a closer connection between individuals and a wide range of literary domains. Furthermore, the findings and insights from this study are expected to contribute to broader research on personalized recommender systems.

1.4.2 Audience

The outcomes of this project will benefit several key stakeholders. These stakeholders include:

- Readers (users): Enhance their book exploration experience by exploring a wider range of literature to suit their individual preferences.
- Publishing professionals: Authors, editors, and publishers can use recommender systems to gain a deep understanding of reader preferences and trends. Adjust content creation and marketing strategies based on user preferences to increase book sales and readers' feeling of engagement.
- E-commerce platform: The platform can utilize the recommendation algorithms developed in this project to increase user engagement and conversion rates (the ratio of user purchases to views).

- Academics and Research: The findings of this project can contribute to existing research on personalized recommender systems and promote the understanding of the application of VAE.

2 Background Review

A recommender system (RS) predicts the likelihood that users will be interested in items that they may not yet know about. To make recommendations, RS typically require user data and user commentary on these elements. User's feedback on items can be obtained either explicitly or implicitly following a proposal is put forward. The system preserves the feedback in a database and utilizes it for future suggestions [3]. Deploying recommendations in e-commerce has many benefits for both sellers and consumers. The former's goal is to get more orders. On the other hand, consumers receive a list of products they are most likely to find useful. They also save time, energy, and money that they would have spent on discovering items they appreciate [3]. However, RS still faces challenges. The most well-known issue is rating scarcity, which can lead to cold starts for users or items.

Collaborative filtering is a prevalent approach employed within recommender systems. It facilitates prediction of user preferences by identifying and utilizing similarities between users and items [3]. The variational autoencoder (VAE) extends to collaborative filtering with implicit feedback. VAE generalizes the linear latent factor model, allowing us to explore nonlinear probabilistic latent variable models supported by neural networks on large-scale recommendation datasets. The article presented a polynomial conditional likelihood model of neurodegeneration [4]. Recommender systems are often evaluated using ranking-based metrics. Previous research has shown that direct optimization of Top-N ranking loss is challenging, and previous work on this subject has resorted to employing relaxations and approximations [5], [6]. Here, it is evident that the multinomial likelihoods are well-equipped to simulate implicit feedback data and are a superior substitute to more traditional likelihood functions such as Gaussian and logistic in terms of their correspondence with ranking loss [5].

Kingma and Welling [7] showed how the reparameterization of variational lower bounds produces simple differentiable unbiased estimators of the lower bounds; such SGVB (Stochastic Gradient Variational Bayesian) estimators can be used for efficient approximate a posteriori inference in virtually any model with continuous latent variables and/or parameters and can be optimized directly using standard stochastic gradient ascent techniques. For the case of continuous latent variables in the dataset and at each data point, they propose the auto-encoding VB (AEVB) algorithm. In the AEVB algorithm, optimize the recognition model to make inference and learning particularly efficient by using an SGVB estimator that allows us to perform very efficient approximate a posteriori inference using simple ancestor sampling, which in turn allows us to efficiently learn the model parameters without the need for costly iterative inference schemes (e.g., MCMC) for each data point. The approximate a posteriori inference model, which has been learned, can be applied to a multitude of tasks, such as image recognition, noise reduction, data representation, and visualization [7]. When using neural networks for recognition models, we obtain variational autoencoders (VAEs).

In the paper "Adversarial Variational Self-Encoder for Top-N Recommender Systems", Zhang et al. [8] proposed a model of a recommender system that combines a Variational Auto-Encoder (VAE) and a Generative Adversarial Network (GAN). This model utilizes users' historical interaction data to predict items they may be interested in, especially for implicit feedback data. With this combination, the model not only improves recommendation accuracy, but also excels in handling data sparsity, a common and intractable problem in traditional recommender systems .The application of VAE allows the model to learn deep, probabilistic feature representations from the user's historical behaviors, and the introduction of GAN further optimizes the representation of these features. Experimental results show that this approach combining VAE and GAN outperforms traditional recommendation models on multiple publicly available datasets, especially in personalized recommendations and handling sparse user-item interaction data. This study not only demonstrates the potential of VAE in recommender systems, but also provides valuable insights on how to improve the performance of recommender systems through deep learning techniques.

Additionally, Krishnathasan [9] explored the use of Variable Auto-Encoder (VAE) in movie recommendation systems. This research presented an innovative approach using two VAEs running in parallel to learn and predict user preferences for movies and genres, respectively. This approach was especially effective when dealing with large sparse datasets, addressing some of the limitations in traditional collaborative filtering approaches. The article also emphasized the importance of leveraging neural networks and machine learning techniques, such as singular value decomposition (SVD) and matrix decomposition, to handle large sparse datasets. By compared with the AutoRec model, the study demonstrated the Concurrent Hybrid Variational Autoencoders approach in predicting user ratings of movies. This study demonstrates the potential application of VAE in movie recommendation systems and also provides a reference for book recommendation systems.

3 Technical Progress

3.1 Approach

Variational Auoto-Encoder [7]**:**

There are multiple advantages of applying VAE for collaborative filtering in recommender systems. VAE can map users and items into a continuous latent space, which helps to discover potential user interests and item characteristics. In addition, using VAE can fill in the gaps in the data by learning the representations in the latent space, which improves the system's ability to handle the missing data. VAE can alleviate the cold-start problem in recommender systems by learning the latent space and combining it with information about other relevant features to provide them with meaningful recommendations.

(High dimensional) variable: X – generated from conditional distribution $P_{\theta*}(x|z)$

Unobserved continuous random variable: Z – generated from prior distribution $P_{\theta*}(z)$

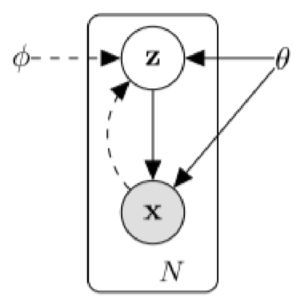

Figure 1 Directed Graphical Model [7]

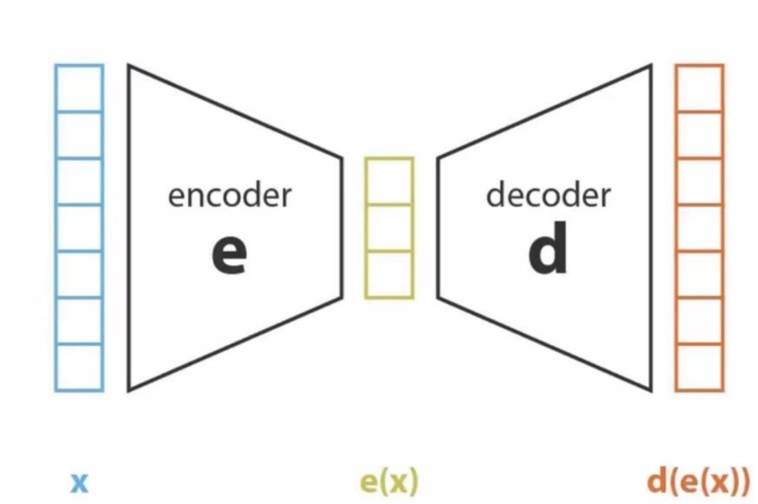

Figure 2 Abstract Figure of VAE

VAE is a network architecture belonging to the probabilistic graphical and variational Bayesian models.

Two neural networks of VAE:

- Encoder: Encoding and compression of data.
- Decoder: Data generation.

However, there is still a prerequisite before decoding: e(x) is a regularized latent space. In order to make latent space regularized, the way to do this is to add some randomness. That means input the same x, output different z every time. The specific mathematical representation of increased randomness is the application of Bayesian Rules.

$$P(z|x) = \frac{P(x|z)P(z)}{P(x)} = \frac{P(x|z)P(z)}{\int P(x|u)P(u)du}$$

Next, calculate loss function to ensure P(z) and P(z|x) are regularized.

$$q_\theta(z)(encoder) \approx p(z|x)$$

$$D(q_\theta(z)||p(z|x))(KL\ Divergence) = E_{z\sim q}\left[\log\frac{q_\theta(z)}{p(z|x)}\right] = E_{z\sim q}[\log q_\theta(z) - \log p(z|x)]$$

$$= E_{z\sim q}[\log q_\theta(z) - \log p(z,x)] + \log p(x)\ \ (evidence\ lower\ bound)$$

$$\log p(x) = E_{z\sim q}[\log p(z,x) - \log q_\theta(z)] + D(q_\theta(z)||p(z|x))$$

$$\log p(x) \geq E_{z\sim q}[\log p(z,x) - \log q_\theta(z)] \equiv \varsigma_q$$

The objective: minimizing KL divergence leads maximizing evidence lower bound (ELBO).

Recommender VAE (RecVAE) [10]**:**

The RecVAE model is an improved model based on the Mult-VAE model, which uses multinomial distributions as the likelihood function instead of Gaussian and Bernoulli distributions, which are commonly used in VAE models.

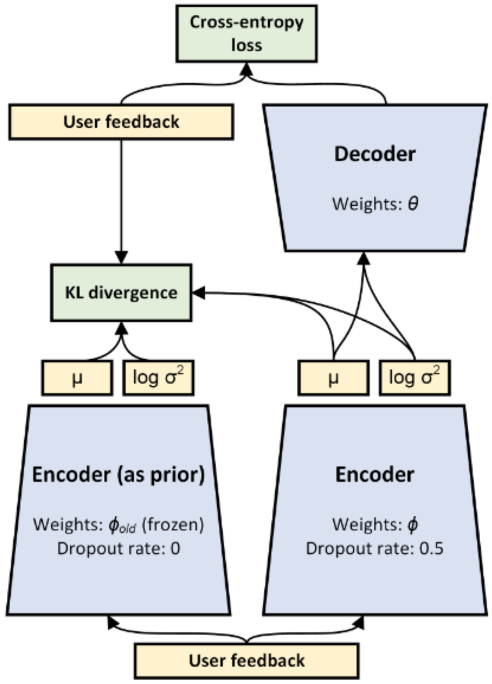

Figure 3 RecVAE Structure [10]

RecVAE is denoising variational autoencoder, it changes evidence lower bound (ELBO) to:

$$E_{q_{\phi}}(z_u|x'_u)E_q(x'|x)[\log p_\theta(x|z) - \beta'(x)KL(q_\varphi(z|x')||p(z|\varphi_{old}, x))]$$

Next, a method like Alternating Least Squares (ALS) (a common matrix factorization technique) is utilized to train the model, alternating before user and item embeddings. User embeddings are apportioned by the inference network, while each item embedding is trained separately.

ALGORITHM 1: Proposed training procedure

Data: $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|U|}\}$
Result: $\boldsymbol{\phi}, \boldsymbol{\theta}$
for $n := 1, \ldots, N$ **do**
  for $m := 1, \ldots, M_{\text{enc}}$ **do**
    Sample batch $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_b\} \sim \mathcal{D}$;
    Update $\boldsymbol{\phi}$ based on $\widetilde{\mathcal{L}}$;
  end
  $\boldsymbol{\phi}_{old} := \boldsymbol{\phi}$;
  for $m := 1, \ldots, M_{\text{dec}}$ **do**
    Sample batch $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_b\} \sim \mathcal{D}$;
    Update $\boldsymbol{\theta}$ based on $\widetilde{\mathcal{L}}_{\text{dec}}$;
  end
end

Figure 4 Training Method

Dataset [1]**:**

In this project, different book datasets will be used. We will use the user_id column, the book_id (ISBN) column and the rating column from the dataset. Currently we are using the Book-Crossing dataset: https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset/data.

3.2 Technology

Software	Operating System	Windows 11
	Framework	Pytorch 1.12.1 Cudatoolkit 11.3.1
	Language	Python 3.9.18
	Libraries	Matplotlib 3.8.1 Numpy 1.26.0 Pandas 2.1.3
Hardware	Central Processing Unit (CPU)	Intel(R) Core™ i7-10750H CPU @ 2.60GHz
	Graphics Processing Unit (GPU)	NVIDIA GeForce GTX 1660 Ti

3.3 Testing and Evaluation Plan

- Data Testing
  - Prior to model training, ensure that the data conforms to pre-determined assumptions. (The dataset should contain three columns: user_id, book_id and rating)
  - Verify the completeness and consistency of the dataset to ensure that there is no missing or incorrect data.
  - Determine the specific values of the parameters, such as rating threshold, minimum items per user and minimum users per item, used in the model through data visualization.
- Model Testing & Evaluation
  - Perform basic testing of model architecture and logic to catch potential bugs and issues. (Pre-training tests)
  - After model training, a comprehensive performance evaluation is performed, including but not limited to the evaluation of metrics such as NDCG@100, recall@20, recall@50. (Post-training tests)

- Pipeline Testing
  - Validate the entire recommender system pipeline, including data preprocessing, model training, prediction and feedback mechanisms, to ensure that each component of the system works together
  - Ensure that the model is repeatable and consistent, i.e., that the model's performance remains stable when run in different environments or conditions (Different CPU and GPU, different versions of Python and dependent libraries ,and different dataset).
- Cross-validation
  - Test model performance on different data subsets through cross-validation to ensure model adaptability to different data. The aim is to test the robustness of the model.

3.4 Design and Implementation

- Data preprocessing
  - Not all datasets are appropriate. The beginning of the project downloaded different datasets online. Such as Amazon, Good-Reading and Book-Crossing. However, the RecVAE model requires data that are user_id, book_id and rating. Only Book-Crossing meets the requirements. Therefore, we apply Book-Crossing to Data preprocessing.
  - Use data visualization to identify key parameters (rating threshold, minimum items per user, minimum users per item, number of held out users).
- Implementation of the RecVAE mode
  - The encoder and decoder of the model are optimized separately using the Adam optimizer. The training process consisted of selectable alternating training methods and the performance of the model was evaluated after each training cycle using NDCG and recall as evaluation metrics. The final model performance is determined by its performance on a test set, which enables effective recommendation of books that may be of interest to the user.

4 Project Management

4.1 Activities

To complete each objective, some details will be given in the table.

Objectives	**Activities**

Learning the Variational Autoencoders (VAE) model	Research articles and codes about the model like VAE, Mult-VAE, and RecVAE **(Completed)**
Dataset	Download suitable datasets about books and data processing **(Completed)**
Book Recommendation	Using the book dataset, implement it in the RecVAE model and get the initial results **(Completed)**
Comparative Study	Evaluate the performance of different models and list their strengths and weaknesses **(Uncompleted)**
Implementing the model	Based on RecVAE, develop an efficient model and use the dataset to perform multiple tests to demonstrate its performance **(Uncompleted)**
Report	● Weekly log ● Record test results and improvement measures ● Final report

4.2 Schedule

Using Gantt charts to show the activities and their deadlines.

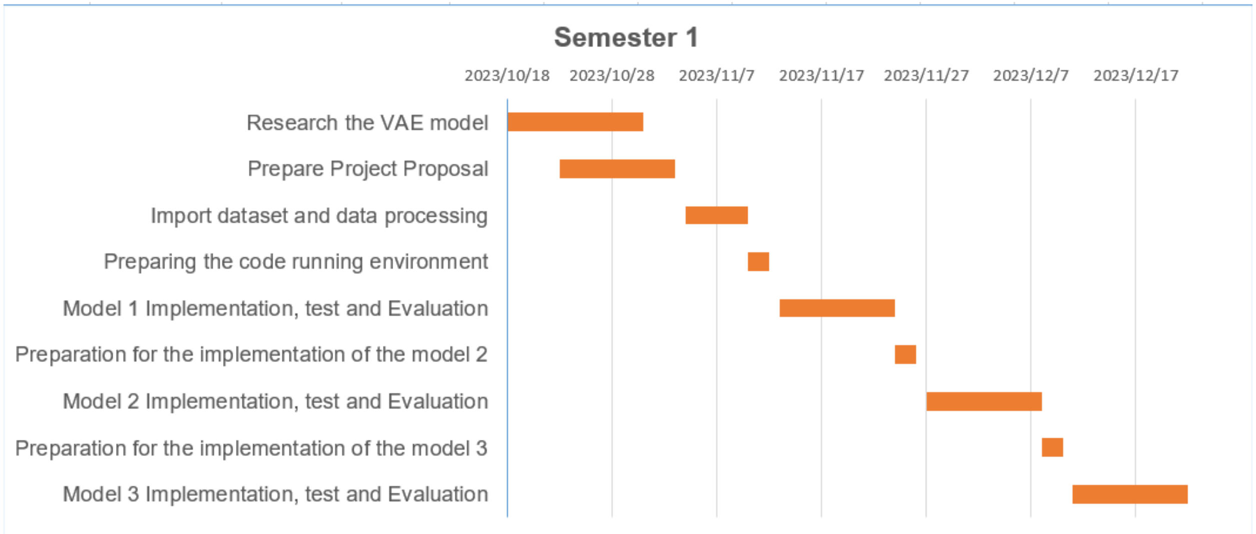

Figure 5 Schedule (Semester 1)

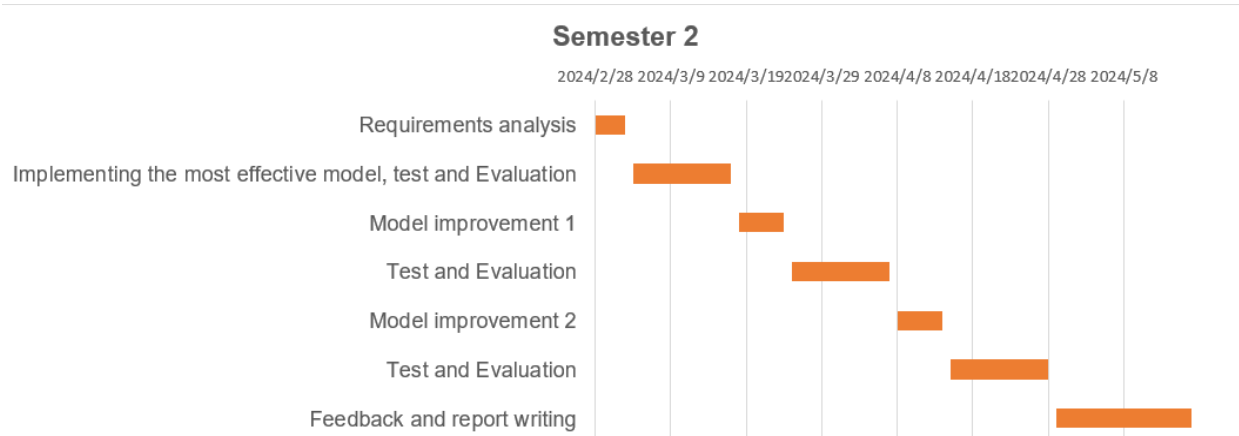

Figure 6 Schedule (Semester 2)

4.3 Project Version Management

Using GitHub to manage the several versions of project code. And the Git repository link is https://github.com/FOMOKN/Project.git.

4.4 Project Data Management

All references in the project will be saved and managed using Zotero.

To have a better grasp of the project's progress, both local and cloud-based storage management will be used. Use the Feishu cloud document folder to store documents including project proposal, weekly reports, progress report, final report, etc. Feishu link is https://y1jgvfzywn.feishu.cn/drive/folder/QYKhfak5OlAv5Hdz8UycoHu8neh.

4.5 Project Deliverables

- Project Proposal
- Ethic Forms
- Weekly Report
- Progress Report
- Slides
- Finial Report
- Project Code

5 Professional Issues and Risk:

5.1 Risk Analysis

Risk ID	Potential Risk	Cause ID	Potential Causes	Severity	Likelihood	Risk	Mitigation ID	Mitigation
R1.1	Missed deadline	C1.1.1	Illness	1	3	3	M1.1.1	Register exceptional circumstances if ill.
		C1.1.3	Poor time management	4	3	12	M1.1.2	Make a Gantt plan early
R2.1	Loss of data	C2.1.1	Poor version control	4	4	16	M2.1.1	Implement version control strategy at start.
R2.2	Insufficient data	C2.2.1	Difficulties in data collection	3	2	6	M2.2.1	Broaden data sources and increase data collection channels
R2.3	Data leak	C2.3.1	Security hole	4	3	12	M2.3.1	Strengthen security measures and conduct regular system security checks
R3.1	Software bugs	C3.3.1	Non-modular design	1	3	3	M3.3.1	Create highly modular desigh before implementation
		C3.3.2	Poor test plan	4	3	12	M3.3.2	Create test plan at start
R4.1	Model overfitting	C4.1.1	Poor choice of model parameters	4	3	12	M4.1.1	Perform cross-validation and select appropriate model complexity
R5.1	System performance issues	C5.1.1	Poor choice of model parameters	4	3	12	M5.1.1	Perform cross-validation and select appropriate model complexity

Figure 7 Risk Analysis

5.2 Professional Issues

- Legal issues
  - Involves user data and therefore needs to comply with data protection laws, such as GDPR, to ensure the security and privacy of user data
  - The book datasets used may have copyright implications and there is a need to ensure fair use and respect for the intellectual property rights of the original authors.
- Social and ethical issues
  - When processing user reading preferences and feedback data, it is necessary to ensure that user privacy is not violated.
  - Ensure that algorithms do not direct users to favor specific books or publishers to avoid potential bias and manipulation.
- Environmental issue

- Although the project is primarily a digital product, energy efficiency and sustainability should be considered in the use of hardware, such as optimizing algorithms to reduce the use of computing resources.

- Code of Professional Conduct
  - Follow the codes of conduct of professional organizations such as BCS (British Computer Society), ACM (American Computer Society), etc.
- Personal values and standards
  - Report the results of the study honestly, without exaggerating the performance of the recommender system

6 References

[1] Möbius, 'Book Recommendation Dataset'. Accessed: Dec. 23, 2023. [Online]. Available: https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset

[2] P. Kerudi, T. Srujan, and R. C K, 'BOOK RECOMMENDATION SYSTEM: A SYSTEMATIC REVIEW AND RESEARCH ISSUES', *Huazhong Keji Daxue Xuebao Ziran Kexue BanJournal Huazhong Univ. Sci. Technol. Nat. Sci. Ed.*, vol. 50, p. HST-0621, Jun. 2021.

[3] H. Alharthi, D. Inkpen, and S. Szpakowicz, 'A survey of book recommender systems', *J. Intell. Inf. Syst.*, vol. 51, no. 1, pp. 139–160, Aug. 2018, doi: 10.1007/s10844-017-0489-9.

[4] D. Liang, R. G. Krishnan, M. D. Hoffman, and T. Jebara, 'Variational Autoencoders for Collaborative Filtering', arXiv.org. Accessed: Oct. 24, 2023. [Online]. Available: https://arxiv.org/abs/1802.05814v1

[5] M. Weimer, A. Karatzoglou, Q. V. Le, and A. Smola, 'COFIRANK Maximum Margin Matrix Factorization for collaborative ranking', in *Proceedings of the 20th International Conference on Neural Information Processing Systems*, in NIPS'07. Red Hook, NY, USA: Curran Associates Inc., Dec. 2007, pp. 1593–1600.

[6] J. Weston, S. Bengio, and N. Usunier, 'WSABIE: scaling up to large vocabulary image annotation', in *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume Three*, in IJCAI'11. Barcelona, Catalonia, Spain: AAAI Press, Jul. 2011, pp. 2764–2770.

[7] D. P. Kingma and M. Welling, 'Auto-Encoding Variational Bayes'. arXiv, Dec. 10, 2022. doi: 10.48550/arXiv.1312.6114.

[8] G. Zhang, Y. Liu, and X. Jin, 'Adversarial Variational Autoencoder for Top-N Recommender Systems', in *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Nov. 2018, pp. 853–856. doi: 10.1109/ICSESS.2018.8663730.

[9] M. Krishnathasan, 'Movie Recommendation System Using Concurrent Hybrid Variational Autoencoders', in *2021 21st International Conference on Advances in ICT for Emerging Regions (ICter)*, Dec. 2021, pp. 1–5. doi: 10.1109/ICter53630.2021.9774813.

[10] I. Shenbin, A. Alekseev, E. Tutubalina, V. Malykh, and S. I. Nikolenko, 'RecVAE: a New Variational Autoencoder for Top-N Recommendations with Implicit Feedback', in *Proceedings of the 13th International Conference on Web Search and Data Mining*, Jan. 2020, pp. 528–536. doi: 10.1145/3336191.3371831.