Activity Recommendation System
–Tackling the User Cold-Start Problem

蔡彥竹
ID: b11902003
Email: b11902003@csie.ntu.edu.tw
Contribution:
Research, GUI implementation

劉蕃熙
ID: b12902031
Email: l.felicity@gmail.com
Contribution:
Research, Model implementation

李瑞恩
ID: b12902067
Email: b12902067@csie.ntu.edu.tw
Contribution:
Research, Evaluation implementation

1. Introduction

As college students, we are eager to explore international opportunities such as scholarships, exchange programs, forums, and conferences. However, these opportunities are scattered across various websites, often poorly organized and lacking personalized recommendations. For instance, on platforms like Opportunities Circle, filtering options are limited to countries or funding types. This often results in a long list of irrelevant activities, forcing users to manually sift through the content.

To improve user experience, we aim to design a recommendation system that suggests tailored opportunities based on users' profiles and past interactions. Given that these platforms frequently encounter new users who use the system sporadically, our focus is on addressing the user cold-start problem.

2. Related Work

2.1. Contrastive learning for cold-start recommendation.

This paper addresses the item cold-start problem by encouraging the feature representations of cold-start items to preserve collaborative signals. To achieve this, they design a contrastive learning objective that maximizes mutual information between item content features and collaborative embeddings. The framework, CLCRec, utilizes multi-layer perceptrons as feature encoders and leverages cosine similarity within contrastive loss to align representations.

Unlike this paper, which targets item cold-start, our project focuses on user cold-start. We initialize user vectors using a one-layer perceptron with one-hot encoded inputs.

2.2. MeLU: Meta-Learned User Preference Estimator for Cold-Start Recommendation

MeLU leverages meta-learning, which allows it to rapidly adapt to new tasks with few examples. Unlike collaborative filtering that relies on similar users, MeLU estimates preferences based on the individual user's item-consumption history. This paper also introduces an evidence candidate selection strategy based on MeLU, designed to identify distinguishing items for customized preference estimation, enhancing initial recommendation performance for new users.

2.3. Task-adaptive Neural Process for User Cold-Start Recommendation

TaNP is a meta-learning recommendation model that treats each user's preference as a stochastic process. It generates a predictive distribution based on observed user interactions and uses a task-adaptive mechanism to tailor global knowledge to each user, improving recommendation accuracy without relying heavily on gradient-based updates. The main difference between TaNP and our model is that TaNP adapts from some data, while our method is built to work with no data at all.

3. Methodology

3.1. How Do We Solve the Problem

The primary challenge we encountered in this project was addressing the system cold-start problem, where no prior data is available—a combination of both user cold-start and item cold-start scenarios. Our initial plan was to construct user and item vectors using TF-IDF and Latent Semantic Indexing (LSI). However, after carefully weighing the advantages and limitations of this method, we opted instead to use a large language model (LLM) to generate synthetic training and testing data. Details on how LLM is used to create the dataset can be found in Sections 3.3.1 and 4.3.1.

Comparison of the Two Approaches:

1) The LSI-based method requires no existing data, making it suitable for pure cold-start situations. However, without any training data, LSI

alone cannot provide personalized recommendations.

2) LLM-generated data may suffer from quality and consistency issues, which are further discussed in Section 5.1.

In our system, if the initial recommendation is poor, the new users are unlikely to return. Therefore, delivering high-quality recommendations on the first attempt is critical, which is why we chose the LLM-based approach.

3.2. Recommendation Process

Our system recommends five items at a time by ranking items based on the top five dot products between the user embedding and item embeddings.

Note: To improve real-time performance, we only update user embeddings upon receiving new ratings. Item embeddings are updated in batch during system downtime.

3.3. Training Process

3.3.1. Train data generation.
We simulate 4,000 user profiles, each with attributes such as user ID, age group, major, preferred countries, and one to three interests—one of which always aligns with their major. Age is categorized into four levels: freshman/sophomore, junior/senior, master's, and PhD. We then ask Gemini to randomly rate 15 items based on each pseudo profile.

3.3.2. Loss Function.

$$
\begin{aligned}
L = -& \sum_{(u,i,j) \in D} \ln \sigma(z_u^\top z_i - z_u^\top z_j) \\
& + \lambda_1 \sum_{u \in U} \| z_u - (M x_u + \mathbf{b}) \|^2 \\
& + \lambda_3 \sum_{i \in I} \| z_i - \text{lsi}_i \|^2 \\
& + \lambda_2 \left(\|M\|^2 + \|\mathbf{b}\|^2 + \sum_{u \in U} \|z_u\|^2 + \sum_{i \in I} \|z_i\|^2 \right)
\end{aligned}
\tag{1}
$$

z_u: user embedding
z_i: item embedding
x_u: One-hot encoded user profile
lsi_i: TF-IDF-derived LSI vector for item (unigram)
M: linear transformation vector
$\mathbf{b}$: bias term

Our loss function is based on Bayesian Personalized Ranking (BPR). To address the cold-start issue, we introduce a regularization term that constrains the divergence between the initial embedding and the updated embedding after BPR optimization. This encourages the consistency between the pre-trained embeddings (e.g.,

derived from profile/TF-IDF features) and the final embeddings learned through BPR. As a result, the system can generate reasonable recommendations for new users, even before they accumulate sufficient interaction data.
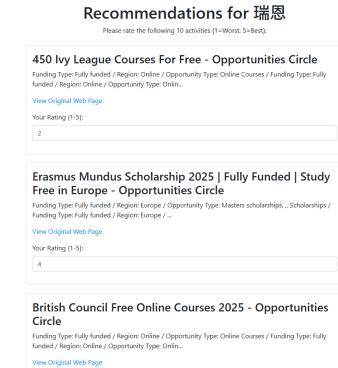
4. Experiment

4.1. Experiment design

We designed our experiments to evaluate two main aspects:

1) The overall performance of our model compared to a baseline model.
2) The performance of both models across different rounds of recommendations.

The baseline model was implemented using the original BPR algorithm. We evaluated both models using two metrics–NDCG@5 and Kendall's Tau-b. For the NDCG@5 evaluation, each user completed three rounds of rating, with five items recommended per round by each model.

4.2. Results of Our System

4.3. Evaluation Results

4.3.1. NDCG@5 in Test data.

TABLE 1. NDCG@5 IN TEST DATA

Model	Round 1	Round 2	Round 3
Our Model	0.6899	0.6688	0.6312
Baseline Model	0.6105	0.6977	0.6427
Mean (Our Model)	0.6634 (std = 0.2090)		
Mean (Baseline Model)	0.6502 (std = 0.1984)		

For the test data, we generated 200 pseudo user profiles and used Gemini to rate activities based on these profiles.

In Round 1, where users are new to the system, our model outperformed the baseline. This demonstrates the effectiveness of our model in handling user cold-start scenarios. In Round 2, the baseline showed a noticeable

improvement, likely because it had adapted to the users' preferences through the initial ratings. By Round 3, both models experienced a performance decline, which we attribute to a limited number of suitable items per user and our system's policy of not repeating recommendations. As a result, high-quality options were depleted over time.

While the baseline outperformed our model in Round 2, we believe this is due to a design choice in our model: the updated user vector is constrained to remain close to its initial representation. The baseline, by contrast, updates user vectors more freely based on new feedback, allowing it to adapt more aggressively and potentially achieve better short-term personalization.

4.3.2. NDCG@5 in A/B Test.

TABLE 2. NDCG@5 EVALUATION FOR A/B TEST

Model	Round 1	Round 2	Round 3
Our Model	0.8803	0.8983	0.8734
Baseline Model	0.8410	0.9031	0.8727
Mean (Our Model)	0.8840 (std = 0.1268)		
Mean (Baseline Model)	0.8722 (std = 0.1287)		

We conducted an A/B test with 10 real users, comparing the performance of our model to the baseline. Key findings include:

1) Round 1: Our model outperformed the baseline, supporting its strength in the cold-start phase.
2) Round 2: Both models improved, with the baseline slightly better than our model.
3) Round 3: Both models showed slight performance drops, likely due to a scarcity of unrecommended relevant items.

Overall, our model maintained slightly better average performance across all rounds. For further details on the user study, please refer to Section 4.4.

4.3.3. Kendall's Tau-b in Test Data.

$$\tau_b = \frac{C - D}{\sqrt{(C + D + T_x)(C + D + T_y)}} \quad (2)$$

TABLE 3. KENDALL'S τ_b COMPARISON

Model	Kendall's τ_b
Our Model	0.2920
Baseline Model	0.2472

We use Kendall's Tau-b to evaluate how similar the Gemini ratings are to the recommendations generated by our model. This is done by comparing the pairwise ordering of item scores, and checking whether the model preserves the correct relative ranking between item pairs.

The Kendall's Tau-b score of our model is around 0.3, which is higher than the baseline model's score of approximately 0.25. This indicates that our model is more effective in recommending activities for cold-start users. Moreover, the p-value analysis shows that about one-third of users have a p-value lower than 0.05, suggesting that the recommendations for these users are statistically significant and unlikely to be random.

Besides, we found that most users have a Kendall's Tau-b higher than 0.5, indicating strong agreement between the predicted rankings and the ground truth. However, about 10% of users have a Tau-b close to 0, meaning that the model fails to capture their preferences. This suggests that while our model works well for the majority of users, a small group of users may need to be re-ranked after BPR training to improve recommendation quality.

4.4. User Study

After conducting an A/B test with real users, we observed that participants were able to clearly perceive the advantages of our model in the first round of recommendation, where it more effectively suggested useful activities. In contrast, the baseline model showed its strength in the second round, where users were more likely to notice improvements based on their initial feedback.

Additionally, some users reported that the system recommended too many online courses. Upon examining the dataset, we found that this was highly correlated with the data distribution—approximately one-third of all activities were online courses. To improve user experience, we are considering introducing filters or limiting the proportion of online course recommendations in future iterations of our system.

5. Future Outlook

5.1. Improve Training Data Quality

Current issues with our training data include:

1) LLM-generated train data may not reflect real user behavior and have too much noise.
2) The diversity in interest categories is limited.
3) Random item selection may be too homogeneous, reducing user profiling depth.

5.2. Item Bias and Popularity Skew

We observed that online courses are too frequently recommended to new users in the first round of recommendation. Potential causes include:

1) LLM-generated train data tends to give online courses high ratings.
2) Imbalanced item distribution on the platform.

We plan to investigate and correct for these biases.

5.3. Personalization Improvements

Our model currently generalizes user interests based on the user's major and age group. Users with atypical preferences may receive suboptimal recommendations. We aim to develop finer-grained personalization strategies in future work.

5.4. Incorporate More Advanced Models

We attempted to replicate methods from other models focusing on cold start problems, but encountered compatibility issues. Future work will focus on studying and integrating these techniques to enhance our model's performance.

5.5. Deployment and Integration

Our ultimate goal is to deploy our system on current opportunity aggregation websites. We aim to provide better user experience and attract more users, making sure international opportunities are available to all aspiring young talents.

6. Conclusions

In this project, we developed a recommendation system capable of operating without any prior user interaction data. Our approach leverages basic demographic information—such as age group and academic major—to generate personalized recommendations for new users. With just a few rounds of feedback, the system quickly adapts to user preferences, making it well-suited for platforms that lack historical interaction data.

In this project, we learned how to handle scenarios where no training data is available, and how to select appropriate evaluation methods to assess the performance of our model. When modifying the loss function, we incorporated ideas from related papers as well as knowledge learned from class, in order to develop a new model for the user cold-start problem.

Looking ahead, our ultimate goal is to deploy this system on existing opportunity aggregation platforms, enhancing user experience and making international opportunities more accessible to aspiring young talents. To further improve recommendation quality, we plan to reduce bias in training data by diversifying interest categories and limiting the over-representation of online courses. Additionally, we aim to incorporate more advanced personalization methods—such as MeLU and other user modeling techniques—to deliver even more relevant and effective recommendations.

7. References

1) YinweiWei, XiangWang, Qi Li, Liqiang Nie, Yan Li, Xuanping Li, and Tat-Seng Chua. 2021. Contrastive Learning for Cold-Start Recommendation. In Proceedings of the 29th ACM International Conference on Multimedia (MM' 21), October 20–24, 2021, Virtual Event, China. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3474085.3475665

2) Hoyeop Lee, Jinbae Im, Seongwon Jang, Hyunsouk Cho, Sehee Chung. 2019. MeLU: Meta-Learned User Preference Estimator for Cold-Start Recommendation. In The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD ' 19), August 4–8, 2019, Anchorage, AK, USA. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3292500.3330859

3) Xixun Lin, Jia Wu, Chuan Zhou, Shirui Pan, Yanan Cao, Bin Wang. 2021. Task-adaptive Neural Process for User Cold-Start Recommendation. In Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145