# A Hybrid Strategy for Cold-Start Recommendation: Combining Moderate Y-Change and Error-Change Approaches

Chi Phan [1]
[1]Computer Science Department, [2]College of Science, [1,2]San Jose State University
{*chi.phan, Chi Phan*}*@sjsu.edu*

***Abstract*—The challenges in recommendation systems arise when new users join without any historical interaction data, known as the cold-start problem. As a result, the system cannot generate meaningful recommendations, resulting in diminished user experience and engagement. This fundamental barrier in recommendation systems has motivated many researchers to explore more sophisticated user onboarding strategies. One of the most recent successful solutions for this problem was introduced by Laanen and Frasincar [1], which extends the error-based active learning framework in strategic item selection rather than random querying. Their approach leverages Single Value Decomposition to tackle uncertainty in identifying the most informative items for new user profiling, significantly improving the efficiency for the initial learning phase. Using their research as a baseline, in this project, I investigated a hybrid methodology that integrates their two primary strategies: Y-change and error-change selection. My experiment utilizes a substantial dataset comprising 70,000 user-item interactions to asses the hybrid approach against established baselines including random selection and PopGini [2] methods. The results demonstrate that the proposed hybrid strategy achieves measurable improvements over traditional methods, suggesting that combining active learning strategies can enhance cold-start recommendation performance.**

***Index Terms*—active learning, cold-start problem, error-based learning, hybrid methodology, recommendation systems, singular value decomposition, user onboarding.**

## I. Introduction

With the expansion of the Internet and digital platforms, recommendation systems have become more important to users, who face an overwhelming volume of content and find it difficult to make informed decisions. These systems are now central to platforms in e-commerce, entertainment, and social media [3].

978-1-7281-6251-5/20/$31.00 ©2020 IEEE

One of the key challenges in recommendation systems is the cold-start problem, which occurs when the system lacks sufficient data on user or item interactions. This often leads to poor recommendations, user disengagement, and neglect of new content [4]. Traditional solutions rely on popularity-based heuristics or content-based filtering, but may fail to capture sophisticated user preferences. Hence, recent researchers have shifted their focus to active learning strategies to address cold-start situations by strategically selecting a small set of items to query user feedback. aiming to draw the most informative preferences with minimal input [5]. Among these, error-based methods have shown promising results that significantly alter model predictions. Motivated by these findings, this project builds on the framework introduced by Laanen and Fransincar [1], who proposed two such strategies: Y-change, which identifies items leading to the large shifts in predicted ratings, and error-change, which targets those reducing the model's generalization error. Both are instance-based and do not require additional external information, making them more efficient and robust.

While each method has strengths, they are complementary: Y-change explores the preference space, while error-change refines model accuracy. To leverage both, this project proposes a hybrid method that combines both methods with equal weighting. The goal is to strike a balance between information gain and prediction liability, particularly in early-stage, cold-start scenarios. This project has three main contributions: (1) reproducing the original Y-change and error-change strategies in a smaller subset of data, (2) implementing the proposed hybrid

approach, and (3) evaluating all strategies against established baselines such as random selection and PopGini, using a reduced-scale experimental setup [6].

## II. Dataset Description

To reproduce the baselines and propose the new hybrid method, I used the same dataset from De Bijenkorf, a Dutch luxury department store with an online platform serving over 100,000 visitors daily and a turnover of €250,000 from over 200,000 items [1]. The dataset comprises 2,563,878 user-product interactions recorded between July 14, 2015, and July 13, 2016. Each interaction is binary: a value of '1' indicates a purchase or net positive behavior (i.e., more purchases than returns), while a '0' indicates neutral or negative behavior (i.e., equal purchases and returns). These implicit signals are treated as positive or negative ratings for recommendations. To reduce computational cost and enable faster experimentation, I randomly sampled 70,000 interactions from the full dataset. This subset preserves the binary interaction structure and serves as the basis for reproducing and extending the original experimental setup.

## III. Methodology

In this section, the model and active learning strategies are described to address the cold-start problem. The formulas used replicate those of Laanen and Frasincar [1], with the key differences being a reduced sample size (70,000 interactions) and a focused evaluation of only the most effective strategies: moderate Y-change and error-change variants, as these were found to yield the lowest RMSE in the original study.

### A. Matrix Factorization

To predict user preferences, I use the same matrix factorization model based on Singular Value Decomposition (SVD), implemented via the Surprise Python library [7]. Each user $u$ and item $i$ are represented in a shared latent factor space of dimensionality $n$, where interactions are modeled by the inner product of their respective vectors:

$$\hat{a}_{ui} = q_i^\top p_u \tag{1}$$

To improve accuracy, bias terms are incorporated, yielding the full prediction function:

$$\hat{a}_{ui} = q_i^\top p_u + \mu + b_u + b_i \tag{2}$$

where $\mu$ is the global average interaction value, and $b_u$, $b_i$ represent the user and item biases. The parameters are optimized by minimizing the regularized squared error over observed interactions. I reused the best-performing hyperparameters from the prior study: 100 latent factors, $\lambda_1 = 10^{-8}$ for bias regularization, and $\lambda_2 = 10^{-5}$ for regularization of the latent factors.

### B. Moderate Y-Change and Error-Change Strategies

To mitigate the cold-start problem, I adopted the same active learning strategy from Laanen and Frasincar [1] that ranks items for presentation to new users. In the setup, a hypothetical cold user $u_0$ is introduced to the system. For each candidate item $i_x$, a synthetic interaction $(i_x, u_0, y)$ is added to the training set $T$ using possible binary ratings $y \in \{0, 1\}$. The effectiveness of $i_x$ is then evaluated using a generalization error over a held-out test set $A_{\text{Test}}$.

The **moderate Y-change** strategy measures the average squared change in prediction outputs across the test set, defined as:

$$\hat{G}_{\Delta Y\text{-}M}(i_x) = -\frac{1}{2} \sum_{y \in \{0,1\}} \sum_{(u,i) \in A_{\text{Test}}} \left( \hat{f}_T(u,i) - \hat{f}_{T \cup (i_x, u_0, y)}(u,i) \right)^2 \tag{3}$$

Here, $\hat{f}_T(u,i)$ represents the predicted interaction based on the original training set $T$, and $\hat{f}_{T \cup (i_x,u_0,y)}(u,i)$ denotes the new prediction after adding the synthetic interaction. This criterion favors items that cause substantial, yet averaged, shifts in predicted outcomes.

In contrast, the **moderate Error-change** strategy quantifies the average reduction in squared prediction error concerning the true labels $a_{ui}$ in the test set:

$$\hat{G}_{\Delta E\text{-}M}(i_x) = \frac{1}{2} \sum_{y \in \{0,1\}} \sum_{(u,i) \in A_{\text{Test}}} \left( a_{ui} - \hat{f}_{T \cup (i_x, u_0, y)}(u, i) \right)^2 \quad (4)$$

This method prioritizes items that are expected to improve predictive accuracy when added to the model.

### C. Hybrid Strategy

I proposed a **hybrid strategy** that combines the moderate Y-change and Error-change approaches by summing their respective generalization errors for each candidate item:

$$\hat{G}_{\text{Hybrid}}(i_x) = \hat{G}_{\Delta Y-M}(i_x) + \hat{G}_{\Delta E-M}(i_x) \quad (5)$$

This formulation balances the potential informativeness of a candidate item (captured by output shifts) with its practical effect on reducing prediction error. I give equal weight to both components in this study as a principled middle ground.

## IV. Evaluation

The performance of each item selection strategy in cold-start scenarios is evaluated following the experimental architecture described in [1], with modifications suited to the reduced sample size and computational constraints. The evaluation focuses on five strategies: PopGini, moderate Y-change, moderate Error-change, random selection (averaged over 10 trials), and the proposed hybrid strategy combining Y-change and Error-change.

The performance for different numbers of shown items, $k = 2, 4, 6$, was assessed to simulate early-stage cold user interactions. The reason why smaller values of $k$ were selected is the reduced size of the sample data. For each $k$, I chose an appropriate set of cold users—those who have interacted with at least one item outside the selected $k$—to ensure reliable estimation of prediction error on the remaining interactions. This dynamic filtering of cold users is essential due to the sparsity of the reduced 70,000-sample dataset, and the resulting number of cold users varied slightly across strategies and item counts.

The evaluation metric is Root Mean Squared Error (RMSE), a standard measure for prediction accuracy in recommender systems. Table I presents the RMSE scores for each strategy and item count. The random strategy, used as a baseline, performs the worst overall, particularly at $k = 6$ where the model begins to overfit uninformative item interactions. PopGini shows moderate improvements over random selection, especially at $k = 4$, but fails to generalize consistently across all values of $k$ as shown in Fig.

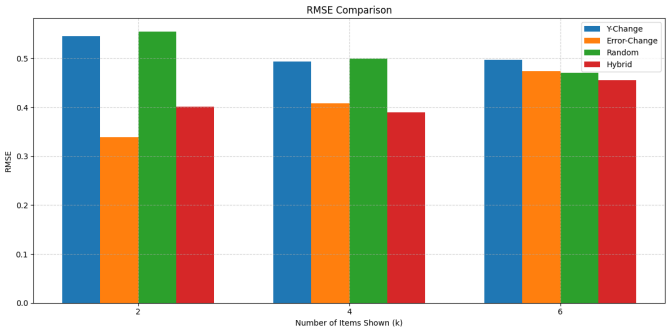


Fig. 1: RMSE Comparison: Y-Change, Error-Change, Random, Hybrid

TABLE I: RMSE for Different Strategies and Items Shown ($k$)

| Strategy | 2 items | 4 items | 6 items |
|---|---|---|---|
| Random (avg) | 0.4638 | 0.4889 | 0.5343 |
| PopGini | 0.4712 | 0.4578 | 0.5211 |
| Y-change (mod) | 0.5454 | 0.4933 | 0.4974 |
| Error-change (mod) | **0.3391** | **0.4078** | 0.4736 |
| Hybrid | 0.4752 | 0.4883 | **0.4713** |

The error-change strategy yields the lowest RMSE values at both $k = 2$ and $k = 4$, achieving 0.3391 and 0.4078, respectively. This suggests that selecting items based on expected reduction in prediction error is highly effective when the number of interactions is small. The Y-change strategy, which prioritizes items that cause significant shifts in output predictions, performs better than random and PopGini but does not match the accuracy of error-change. Its best performance is observed at $k = 4$ with an RMSE of 0.4933.

The hybrid strategy, which equally combines the moderate Y-change and error-change scores, performs competitively across all settings, achieving an RMSE of 0.4752 at $k = 2$, 0.4883 at $k = 4$, and its

best score of 0.4713 at $k = 6$. Interestingly, while the hybrid approach does not achieve the lowest RMSE in any single case, it consistently ranks near the top, reflecting its role as a balanced compromise between exploration and exploitation.

The reason the hybrid strategy does not outperform error-change at smaller values of $k$ may lie in the fixed equal weighting ($\alpha = 0.5$) used to combine the two scoring functions. While Y-change captures the informativeness of an item by assessing its impact on predictions, it may introduce unnecessary variance when added to the more stable Error-change score. A potential direction for future work is to optimize the weighting factor $\alpha$ dynamically, based on the number of items shown, user-specific interaction history, or statistical properties of the ranking scores.

## V. Limitations And Future Work

While RMSE is a widely used metric for evaluating prediction quality in recommender systems, it may not fully capture the effectiveness of active learning strategies in this cold-start setup. In particular, RMSE assumes a continuous rating scale and penalizes squared deviations between predicted and actual values. However, this dataset contains binary interactions (0 or 1), which represent implicit feedback, such as purchases or returns, rather than explicit ratings. In such contexts, a slight deviation from the actual binary label (e.g., predicting 0.4 for a true value of 1) is penalized just as heavily as in continuous-valued systems, even though the recommendation may still be helpful in practice.

Moreover, RMSE does not account for the informativeness or diversity of the selected items. An item that significantly enhances the model's understanding of a user's preferences, even if it is not accurately predicted, may still be valuable for long-term personalization. Active learning strategies aim to maximize such information gain, but RMSE captures only short-term predictive accuracy. As a result, methods like Y-change, which prioritize shifts in model behavior, may appear less effective under RMSE even though they serve their intended purpose.

For future work, I propose incorporating alternative evaluation metrics such as Precision, Recall, or AUC, which are more suited for binary recommendation tasks. These metrics better reflect ranking quality and user relevance. Additionally, integrating real-world user feedback after active item presentation, such as rating confirmations or click-through behavior, would allow for a more meaningful evaluation of strategy performance in practical deployments. Such feedback could also inform adaptive weighting in hybrid strategies or guide reinforcement learning-based recommendation models.

Another potential evaluation metric for active learning strategies is the Jaccard Index [8], which measures the similarity between recommended item sets across users. This metric is especially useful in understanding whether different strategies yield consistent item selections for cold users, as well as in the binary values format. A high Jaccard score indicates agreement between users or methods in item selection, which may suggest redundancy, whereas a lower score could reflect diversity or personalization.

However, in the current setup, the effectiveness of Jaccard as a comparative metric is limited by both the small sample size (70,000 interactions) and the small number of items shown ($k = 2, 4, 6$). As a result, the absolute values of the Jaccard Index are minimal, typically below 0.01, as shown in II, making comparisons difficult. These low values may not necessarily reflect poor performance but rather the sparsity of overlap among candidate items when only a few items are selected per user. In future work, applying this metric to larger datasets with more cold users and a wider range of item exposure may yield more actionable insights into selection consistency across strategies.

TABLE II: Jaccard Index Across Strategies and Items Shown ($k$)

| **Strategy** | **2 items** | **4 items** | **6 items** |
|---|---|---|---|
| PopGini | 0.0131 | 0.0070 | 0.0059 |
| Y-change (mod) | 0.0003 | 0.0011 | 0.0026 |
| Error-change (mod) | 0.0006 | 0.0005 | 0.0003 |
| Hybrid | 0.0004 | 0.0012 | 0.0022 |

Furthermore, to allow future adaptation and fine-tuning of the hybrid strategy, future work may focus on tuning the weight $\alpha$ of the scoring function:

$$\hat{G}_{\text{Hybrid}}(i_x) = \alpha \cdot \hat{G}_{\Delta Y\text{-}M}(i_x) + (1-\alpha) \cdot \hat{G}_{\Delta E\text{-}M}(i_x), \quad \alpha \in [0,1] \tag{6}$$

The parameter $\alpha$ enables control over the balance between exploration (Y-change) and exploitation (Error-change). In this project, the fixed $\alpha = 0.5$ represents equal weighting, but future research may investigate dynamic or user-adaptive tuning of $\alpha$ based on interaction context or feedback effectiveness.

## VI. Conclusion

In this project, I reproduced and extended prior research on active learning strategies for alleviating the cold-start problem in recommender systems. Using a reduced version of the De Bijenkorf dataset (70,000 interactions), the project implemented two error-based strategies—moderate Y-change and moderate Error-change—and introduced a hybrid method that combines both via equal weighting.

Each strategy was evaluated using Root Mean Squared Error (RMSE) over varying numbers of shown items ($k = 2, 4, 6$), and compared against PopGini and a random baseline. The Error-change strategy achieved the lowest RMSE scores, particularly at $k = 2$ and $k = 4$, confirming its effectiveness at optimizing short-term prediction accuracy. While the hybrid method did not outperform Error-change in absolute RMSE, it consistently performed competitively across all values of $k$, highlighting its potential as a balanced trade-off between informativeness and precision.

Beyond RMSE, I also examined the Jaccard Index as a measure of strategy similarity. Although the Jaccard values were very low across all strategies, this is likely due to the small dataset size and limited number of shown items. As such, these values are challenging to interpret in the current setting; however, future work could explore their utility in larger-scale experiments.

The limitations of relying solely on RMSE for evaluation are also emphasized, particularly in the context of binary, implicit feedback data. Alternative metrics, such as Precision, Recall, and AUC, should be incorporated in future work to capture user-perceived relevance better. Moreover, real-world deployment scenarios could benefit from adaptive weighting of the hybrid strategy, where the $\alpha$ parameter is tuned based on feedback or item characteristics.

Overall, my findings suggest that error-based strategies are effective for cold-start recommendations in active learning strategies, and that hybrid methods offer a promising direction for balancing exploration and accuracy. Future work will focus on dynamic adaptation, larger-scale validation, and evaluation using real-time user interactions.

## References


[1] R. Laanen and F. Frasincar, "Alleviating the cold-start problem in recommender systems using error-based learning," in *2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2024, pp. 310–315.

[2] T. Geurts and F. Frasincar, "Addressing the cold user problem for model-based recommender systems," in *Proceedings of the International Conference on Web Intelligence (WI '17)*. ACM, 2017, pp. 745–752.

[3] F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*, 10 2010, vol. 1-35, pp. 1–35.

[4] J. Bobadilla, F. Ortega, A. Hernando, and J. Bernal, "Generalization of recommender systems: Collaborative filtering extended to groups of users and restricted to groups of items," *Expert Systems with Applications*, vol. 39, no. 1, pp. 172–186, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417411009675

[5] N. Rubens, M. Elahi, M. Sugiyama, and D. Kaplan, *Active Learning in Recommender Systems*. Boston, MA: Springer US, 2015, pp. 809–846. [Online]. Available: https://doi.org/10.1007/978-1-4899-7637-6_24

[6] T. Geurts and F. Frasincar, "Addressing the cold user problem for model-based recommender systems," in *Proceedings of the International Conference on Web Intelligence*, ser. WI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 745–752. [Online]. Available: https://doi.org/10.1145/3106426.3106431

[7] N. Hug, "Surprise: A python library for recommender systems," *Journal of Open Source Software*, vol. 5, no. 52, p. 2174, 2020. [Online]. Available: https://doi.org/10.21105/joss.02174

[8] P. Jaccard, "Étude comparative de la distribution florale dans une portion des alpes et des jura," *Bulletin de la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 547–579, 1901.