Final Term Report

Deepak
MA20BTECH11019
ma20btech11019@iith.ac.in

Tapishi Kaur
MA20BTECH11017
ma20btech11017@iith.ac.in

Safder Shakil
EM23MTECH11007
em23mtech11007@iith.ac.in

Abdul Waris
CE22RESCH11001
ce22resch11001@iith.ac.in

Abstract: Recommender Systems (RS) commonly leverage knowledge distillation, a technique where a compact student model is trained using insights from a larger, pre-existing teacher model to compress information. Recent research has highlighted the efficacy of transferring knowledge from the teacher's intermediate layer, significantly enhancing the quality of recommendations made by the student. However, current methods focus on transferring knowledge point-by-point at the individual representation level, which limits effectiveness as RS heavily relies on the relationships within the representation space. This study applies an existing topology distillation approach, guiding the student model by transferring the topological structure based on the relationships within the teacher's space. Initial observations suggest that teaching the entire topological structure to the student isn't consistently effective and can compromise performance due to the student's limited capacity. In our mid-term report, we had employed matrix multiplication-based BPR model and reduced the number of embeddings. In this latest report, we adopted a deep neural network recommender system. We applied Full Topology Distillation (FTD) while maintaining the same number of embeddings and reducing the number of layers in the model. Our findings indicate that FTD efficiently distills real-world recommender systems.

1. Problem Statement

The problem at hand is the inefficiency of contemporary recommender systems, which often demand substantial computational resources for high-accuracy recommendations. This inefficiency poses a significant challenge, particularly in resource-constrained environments and edge devices, where such systems struggle to operate effectively. The main objective of RecPress: Knowledge Distillation for Efficient Recommender Systems project is to employ knowledge distillation techniques to address this problem, with the aim of developing efficient recommender systems that balance accuracy and computational efficiency, thereby making them suitable for broader deployment.

2. Literature Recap

2.1. Topology Distillation for Recommender System

The paper introduces a general topology distillation approach for Recommender Systems (RS), which guides the student's learning by using the topological structure based on relational knowledge in the teacher's representation space. Two topology distillation methods are proposed:

1) FTD (Full Topology Distillation), which transfers the complete topology. FTD is suitable when the student has the capacity to learn all the teacher's knowledge.

2) HTD (Hierarchical Topology Distillation), which transfers the decomposed topology hierarchically. HTD is employed in scenarios where the student has limited capacity compared to the teacher.

Extensive experiments on real-world datasets consistently demonstrate that the proposed approach outperforms state-of-the-art competitors in RS. The paper suggests potential directions for advancing the topology distillation approach, including investigating layer selection and simultaneous distillation from multiple layers, extending topology distillation across different base models, and utilizing prior knowledge of user/item groups to improve the method's effectiveness by providing better supervision on relational knowledge.

3. Algorithm Description and Experiments

We studied the paper by "Topology Distillation for Recommender System SeongKu Kang, Junyoung Hwang, Wonbin Kweon, Hwanjo Yu".

- Using this paper as reference we tried to grasp the concepts of Topological Distillation. Compare FTD and HTD using a toy model. (results in MTR).

- We Compared the HTD vs FTD method using BPR (a prominent Matrix Factorization based recommender system) and CiteUlike dataset for top N recommendations.

- We are able clearly observe the conclusions made in TD for RecSys paper. Where Student model of smaller capacity or size performs well with knowledge distillation using HTD method and student model of significant size performs better when trained with FTD method.

- Our main is aim to study and Implement Topological distillation refered as Fully topological distillation(FTD) in paper, on Deep learning based recommender systems.

- To study further on Full Topological distillation, we implemented a Deep learning based Top n recommendation system, Neural Collaborative filtering.

- Used the famous recommender systems Dataset Movielens 100k for training.

- We also implemented the Full topological distillation method for the Neural Collaborative filtering model and Trained, experimented on the student models with less layers and smaller embeddings compared to parent model.

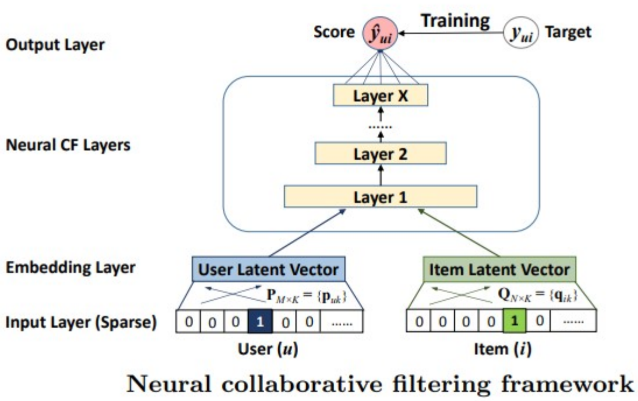
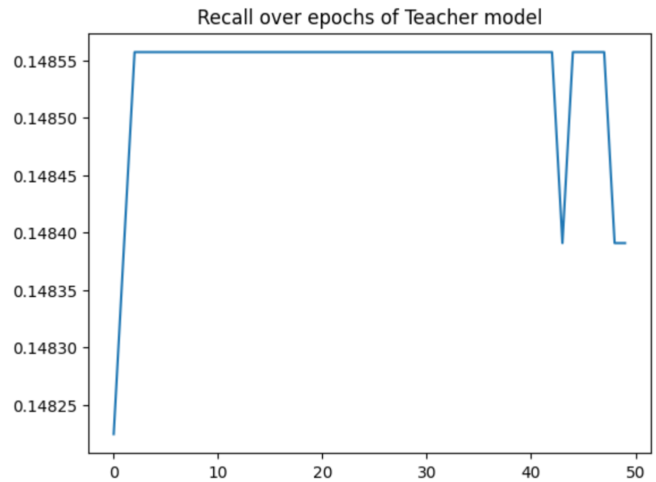
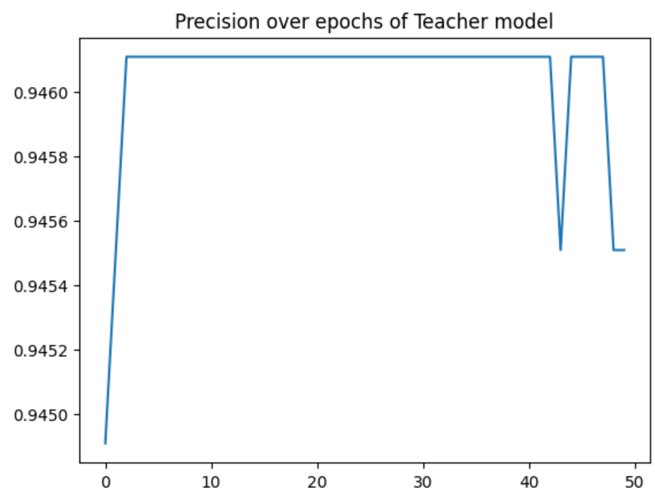


Figure 1. Neural Collaborative Filtering



(a) Recall value for Teacher model at each epoch



(b) Precision value for Teacher model at each epoch

Data Description

In this study, we utilized the renowned MovieLens 100k dataset, encompassing a vast array of data points:

- The dataset comprises information on 62,423 films, accumulating a total of 25,000,095 ratings and 1,093,360 tag applications.
- Over a period spanning from January 9, 1995, to November 21, 2019, the dataset was generated by 162,541 users.
- The dataset's creation date is November 21, 2019.

The selection process for inclusion involved random sampling of individuals, ensuring that each selected user had provided ratings for a minimum of 20 films. Notably, no demographic information was available for the users, and their identification is solely represented by an ID, lacking additional accompanying data.

The dataset includes several pertinent files, namely genome-scores.csv, genome-tags.csv, links.csv, movies.csv, ratings.csv, and tags.csv, as documented by Maxwell et al. (2015).

Training

1. We conducted training on the primary teacher NCF model, comprising 7 layers and an embedding dimension of 200, over a span of 50 epochs. For reference, please consult Figures 2 and 3 for precision and recall metrics pertaining to this architecture. The training employed BCElogitLoss.

2. Additionally, the child model, trained through Full

Topological Distillation utilizing the primary teacher model, constitutes an NCF model with 2 layers and an embedding size of 200. In this scenario, both BCElogitLoss and Topological loss are employed as loss functions, consequently resulting in higher loss values depicted in the graphs.

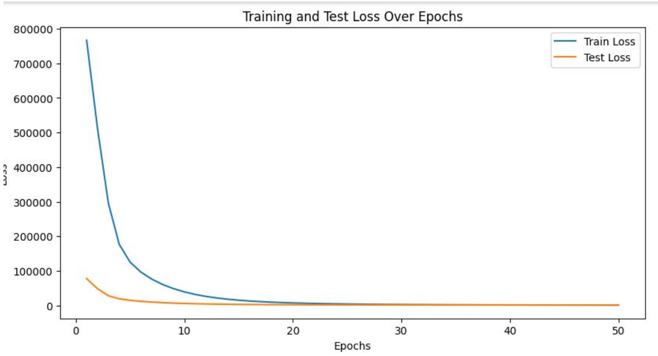


Figure 3. Training and Testing loss during FTT distillation.

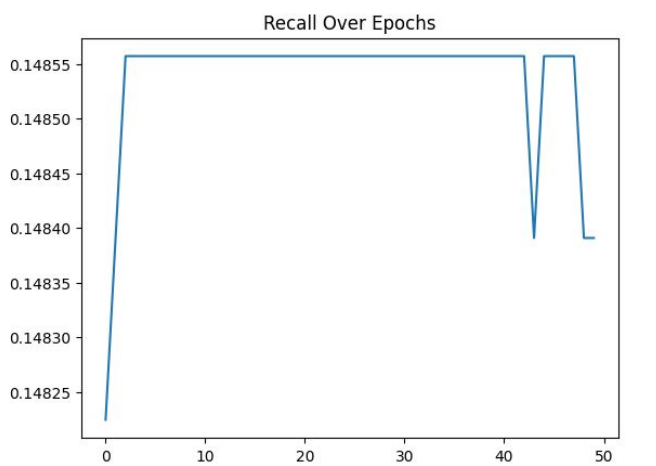


Figure 4. Recall of child model at each epoch of training.

Model Descriptions

The baseline model employed is a basic recommendation system utilizing the user-item rating matrix exclusively to forecast ratings. In contrast, the FTD model represents a topology distillation framework aimed at emulating the structure of a teacher model. FTD entails the transfer of the entire topology and is implemented when the student model possesses adequate capacity to assimilate the entirety of the teacher's knowledge.

On the other hand, the HTD model operates as a hierarchical topology distillation approach, endeavoring to replicate the teacher model's structure across multiple tiers.

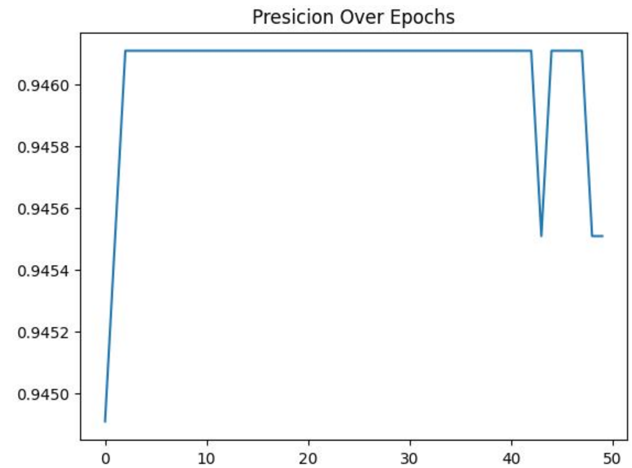
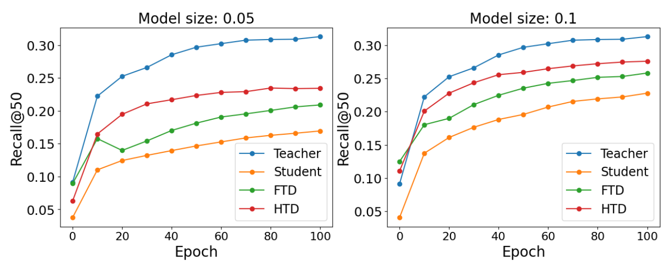
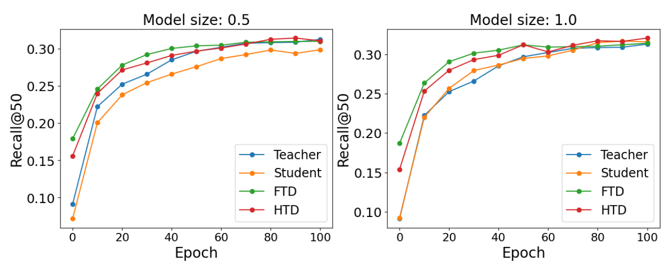


Figure 5. Precision of child model at each epoch of training.



(a) Recall value for models with 10 and 20 parameters respectively



(b) Recall value for models with 100 and 200 parameters respectively

HTD follows a hierarchical transfer of decomposed topology, specifically designed for scenarios where the student's capacity is considerably restricted compared to the teacher.

Additionally, the provided code snippet encompasses a dictionary termed `history_dict`. This data structure stores the training and validation losses of the three models across various student dimensions.

To replicate the results reported in the paper, we use the Python implementation of topology distillation provided in the following Colab notebook: https://colab.research.google.com/drive/1SkuA8i_LHzNaK8WbKr5RhkGeFuaf38wK?usp=sharing#scrollTo=m8CvQh3L5tJy

4. Conclusions

- Precision stabilizes remarkably at approximately 96.55% after a brief initial phase, maintaining a consistent level, paralleled by the behavior of the recall value.
- The model exhibits substantial decreases in losses within a few iterations, signifying consistent learning and subsequent steadiness in the trend.
- Comparative analysis against the Hierarchical Topology Distillation (HTD) model reveals FTD's superiority in accuracy when employing a substantial child model, exceeding 40%. FTD achieves higher accuracy levels in this context.

5. References

1. Topology Distillation for Recommender System SeongKu Kang, Junyoung Hwang, Wonbin Kweon, Hwanjo Yu
2. Distillation from Heterogeneous Models for Top-K Recommendation SeongKu Kang, Wonbin Kweon, Dongha Lee, Jianxun Lian, Xing Xie, Hwanjo Yu
3. DE-RRD: A Knowledge Distillation Framework for Recommender System SeongKu Kang et al (2020).
4. Neural Collaborative Filtering (2017) Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, Tat-Seng Chua
5. Wide & Deep Learning for Recommender Systems (2016) Heng-Tze Cheng et al.
6. DeepFM: A Factorization-Machine based Neural Network for CTR Prediction (2017) Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, Xiuqiang He
7. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer (2019) Fei Sun et al.
8. Distilling the Knowledge in a Neural Network (2015) Geoffrey Hinton et al
9. Knowledge Distillation: A Survey (2021) Jianping Gou et al.
10. A novel Enhanced Collaborative Autoencoder with knowledge distillation for top-N recommender systems. Pan, Yiteng et al. Neurocomputing 2019
11. Scene-adaptive Knowledge Distillation for Sequential Recommendation via Differentiable Architecture Search (2021) Lei chen et al.
12. F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4: 19:1–19:19. https://doi.org/10.1145/2827872