VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering (SCOPE)

J-Component report

Programme: M. Tech (CSE with Specialization in BA)

Course Title: BIG DATA FRAMEWORKS

Course Code: CSE3120

Slot: F2

Title: Reinforcement Learning in Recommendation Systems

Team Members:

Shashank Singh | 21MIA1110

Dazzle A J | 21MIA1119

Faculty: G. Suganeshwari

Sign:

Date:

CONTENTS

Abstract

Reinforcement Learning (RL) has emerged as a promising approach for improving recommendation systems by enabling intelligent decision-making in dynamic and personalized environments. This research paper explores the application of RL techniques in recommendation systems, aiming to enhance the accuracy, relevance, and user satisfaction of recommendations.

The report begins by providing an overview of the fundamental components of recommendation systems, including user profiles, item catalogues, and feedback mechanisms. It emphasizes the limitations of traditional recommendation algorithms and the need for more adaptive and personalized approaches.

Next, the report delves into the concepts and methodologies of RL in the context of recommendation systems. It discusses the formulation of recommendation problems as Markov Decision Processes (MDPs) and the use of RL algorithms to learn optimal policies. Various RL algorithms such as Q-learning, deep Qnetworks, and policy gradient methods are explored, highlighting their strengths, limitations, and suitability for different recommendation scenarios.

Furthermore, the report addresses the challenges associated with RL in recommendation systems. It examines the exploration-exploitation trade-off and explores techniques for efficient exploration, including epsilon-greedy strategies and Thompson sampling. The paper also investigates the cold-start problem and presents methods for handling sparse or incomplete user data during the initial stages of recommendation.

Evaluation methodologies for RL-based recommendation systems are extensively discussed in the paper. It covers offline evaluation techniques, such as precision and recall, as well as online evaluation methods, including user engagement and conversion rates. The paper emphasizes the importance of conducting comprehensive evaluations to assess the effectiveness and impact of RL algorithms on user experience and business objectives.

Lastly, the paper explores potential future directions and research opportunities in the field of RL-based recommendation systems. It discusses the integration of deep learning techniques with RL, the incorporation of user feedback in the learning process, and the scalability of RL algorithms for large-scale recommendation systems.

Objective

The objective of applying Reinforcement Learning (RL) in recommendation systems is to improve the effectiveness, personalization, and user satisfaction of recommendation algorithms. Specifically, the objectives can be outlined as follows:

- **Enhanced Recommendation Accuracy:** The primary objective is to enhance the accuracy of recommendations by leveraging RL techniques. RL algorithms can learn from user interactions and feedback to optimize the recommendation policies over time, resulting in more precise and relevant recommendations.
- **Personalized Recommendations:** RL enables the creation of personalized recommendation systems by modelling individual user preferences. The objective is to develop algorithms that can adapt to each user's unique tastes and provide tailored recommendations, thereby improving user satisfaction and engagement.
- **Adaptive and Dynamic Recommendations:** RL allows recommendation systems to adapt to changes in user preferences and item availability over time. The objective is to develop algorithms that can continuously learn and update the recommendation policies based on evolving user behaviour and environmental dynamics.
- **Exploration-Exploitation Trade-off:** RL algorithms aim to strike a balance between exploration and exploitation. The objective is to design recommendation systems that explore new items or options to discover user preferences while also exploiting known preferences to provide recommendations that are likely to be appreciated by users.
- **Handling Cold-Start Problem:** The cold-start problem refers to situations where there is limited or no initial user data available for making recommendations. The objective is to develop RL-based approaches that can effectively address this problem by employing exploration techniques, leveraging auxiliary information, or utilizing transfer learning methods.
- **Improved User Engagement:** The ultimate objective is to enhance user engagement and satisfaction with recommendation systems. By employing RL, the aim is to deliver more relevant and personalized recommendations that align with users' preferences, leading to increased user satisfaction, longer interaction times, and higher conversion rates.

Introduction

Recommendation systems play a crucial role in various domains, including ecommerce, streaming platforms, and content delivery services, by assisting users in discovering relevant items or content based on their preferences. Traditionally, recommendation algorithms have relied on collaborative filtering, content-based filtering, or hybrid approaches. However, these methods often face challenges in adapting to dynamic user preferences, providing personalized recommendations, and effectively exploring the recommendation space.

Reinforcement Learning (RL) has emerged as a promising approach to address these limitations and improve the effectiveness and personalization of recommendation systems. RL provides a framework for learning sequential decision-making policies through interactions with an environment to maximize long-term rewards. By modelling recommendation problems as Markov Decision Processes (MDPs), RL algorithms can dynamically learn and adapt recommendation policies based on user feedback and interactions.

The application of RL in recommendation systems brings several advantages. First, RL allows for personalized recommendations by learning user preferences over time and adapting the recommendation policies accordingly. This personalization enhances user satisfaction and engagement by providing tailored recommendations that align with individual tastes and preferences.

Second, RL algorithms enable exploration and exploitation in recommendation systems. They strike a balance between exploring new items or options to discover user preferences and exploiting known preferences to provide recommendations that are likely to be appreciated by users. This exploration exploitation trade-off helps mitigate the problem of providing only popular or redundant recommendations and promotes serendipitous discovery.

Third, RL facilitates adaptive and dynamic recommendations by continuously learning and updating recommendation policies based on evolving user behaviour and changes in item availability. This adaptability allows recommendation systems to respond to shifting preferences, trending items, and new user interactions, thereby improving the relevance and freshness of recommendations.

Literature Review

1. G. Pang, X. Zhu, K. Lu, Z. Peng and W. Deng, "**A simulator for reinforcement learning training in the recommendation field**," 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Exeter, United Kingdom, 2020, pp. 10371042, doi: 10.1109/ISPA-BDCloud-SocialComSustainCom51426.2020.00156.

   Abstract: Deep reinforcement learning (DRL) is an unsupervised learning method, which has great commercial value in recommendation scenarios where it is difficult to collect available labelled data. Although very few researchers have begun to do research on the integration of deep reinforcement learning and recommendation methods, the development of these researches is slow because it is difficult to build an online training environment. Therefore, this paper proposes an environment (i.e. user) simulator for training reinforcement learning models in the recommendation field (RL-E Simulator). On the one hand, the simulator builds a user state generation model based on the Generative Adversarial Network (GAN). On the other hand, we propose an rating model based on attention mechanism, and realizes the reward of the simulator to the actions of DRL-based recommendation (i.e. agent). Experimental results show that the simulator provides a low-cost training environment for the DRLbased recommendation.

   URL:
   https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9443735&i snumb er=9443670

2. M. Pham, H. Nguyen, L. Dang and J. A. Nieves, "**Compressive Features in Offline Reinforcement Learning for Recommender Systems**," 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 5719-5726, doi: 10.1109/BigData52589.2021.9671419.

   Abstract: In this paper, we develop a recommender system for a game that suggests potential items to players based on their interactive behaviours to maximize revenue for the game provider. Most of today's recommender systems in e-commerce and retail businesses are built based on supervised learning models and collaborative filtering, while our approach is built on a reinforcement-learning-based technique and is trained on an offline data set that is publicly available on an IEEE Big Data Cup challenge. The limitation of the offline data set and the curse of high dimensionality pose significant obstacles to solving this problem. Our proposed method focuses on improving the total rewards and performance by tackling these main difficulties. More specifically, we utilized sparse PCA to extract important features of user behaviors. Our Q-learning-based system is then trained from the processed offline data set. To exploit all possible information from the provided data set, we cluster user features to different groups and build an independent Q-table for each group. Furthermore, to tackle the challenge of unknown formula for evaluation metrics, we design a metric to self-evaluate our system's performance based on the potential value the game provider might achieve and a small collection of actual evaluation metrics that we obtain from the live scoring environment. Our experiments show that our proposed metric is consistent with the results published by the challenge organizers. We have implemented the proposed training pipeline, and the results show that our method outperforms current stateof-the-art methods in terms of both total rewards and training speed. By addressing the main challenges and leveraging the state-of-the-art techniques, we have achieved the best public leader board result in the challenge. Furthermore, our proposed method achieved an estimated score of approximately 20% better and can be trained faster by 30 times than the best of the current state-ofthe-art methods.

   URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9671419&isnumb er=9671273

3. X. He, K. Wang, H. Lu, W. Xu and S. Guo, "**Edge QoE: Intelligent Big Data Caching via Deep Reinforcement Learning,**" in IEEE Network, vol. 34, no. 4, pp. 8-13, July/August 2020, Doi: 10.1109/MNET.011.1900393. **Abstract:** In mobile edge networks (MENs), big data caching services are expected to provide mobile users with better quality of experience (QoE) than normal scenarios. However, the increasing types of sensors and devices are producing an explosion of big data. Extracting valuable contents for caching is becoming a vital issue for the satisfaction of QoE. Therefore, it is urgent to propose some rational strategies to improve QoE, which is the major challenge for content-centric caching. This article introduces a novel big data architecture consisting of data management units for content extraction and caching decision, improving quality of service and ensuring QoE. Then a caching strategy is proposed to improve QoE, including three parts: (1) the caching location decision, which means the method of deploying caching nodes to make them closer to users; (2) caching capacity assessment, which aims to seek suitable contents to match the capacity of caching nodes; and (3) caching priority choice, which leads to contents being cached according to their priority to meet user demands. With this architecture and strategy, we particularly use a caching algorithm based on deep reinforcement learning to achieve lower cost for intelligent caching. Experimental results indicate that our schemes achieve higher QoE than existing algorithms.

   URL:
   https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9146409&isnumb er=9146400

4. N. Yambem and A. N. Nandakumar, "**Optimizing Hadoop parameter for speedup using Q-Learning Reinforcement Learning**," 2021 Fourth

International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 2021, pp. 1-7, Doi: 10.1109/ICECCT52121.2021.9616965.

Abstract: Hadoop is the most popular open-source big data processing platform which is being used in many big data analytics applications. The performance of Hadoop can be fine-tuned for application performance requirements by adjusting the value of the some of the configuration parameters. Various methods have been proposed in literature for fine tuning the configuration parameters of Hadoop. The relation between the Hadoop performance tuning parameters and speed up is dependent on the nature of the applications and environment dynamics. Tuning the parameters without consideration of these dynamics results in sub optimal configurations and lower performance. Adaptive reinforcement learning using Q-Learning is proposed in this work to fine tune the configuration parameters with the objective of reducing the error between desired and achieved service level agreement (SLA).

URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9616965&isnumb er=9616621

5. M. Tracolli, M. Baioletti, V. Poggioni and D. Spiga, "**Effective Big Data Caching through Reinforcement Learning**," 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 2020, pp. 1499-1504, Doi: 10.1109/ICMLA51294.2020.00231. **Abstract:** In the era of big data, data volumes continue to grow in several different domains, from business to scientific fields. Sensors, edge devices, scientific applications and detectors generate huge amounts of data that are distributed for their nature. In order to extract value from such data requires a typical pipeline made of two main steps: first, the processing and then the data access. One of the main features for data access is fast response time, whose order of magnitude can vary a lot depending on the specific type of processing as well as processing patterns. The optimization of the access layer becomes more and more important while dealing with a geographically distributed environment where data must be retrieved from remote servers of a data lake. From the infrastructural perspectives, caching systems are used to mitigate latency and to serve better popular

data. Thus, the role of the cache becomes a key to have an effective and efficient data access. In this article, we propose a Reinforcement Learning approach, using the Q-Learning technique, to improve the performances of a cache system in terms of data management. The proposed method uses two agents with different objectives and actions to control the addition and the eviction of files in the cache. The aim of this system is to increase the throughput reducing, at the same time, the cache costs, such as the amount of data written, and network utilization. Moreover, we tested our method in a context of data analysis, with information taken from High Energy Physics (HEP) workflow.

URL:
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9356236&i snumb er=9356131

6. S. Bangari, S. Nayak, L. Patel and K. T. Rashmi, "**A Review on Reinforcement Learning based News Recommendation Systems and its challenges**," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 260-265, Doi: 10.1109/ICAIS50930.2021.9395812.

Abstract: Recommendation systems are helpful in both business perspective and user day to day life. These days online contents are generated in huge amount and due to this, users need a special recommendation application namely personalized News Recommendation, and it is highly challenging due to its dynamic nature. Therefore, getting a suitable and relevant news article for a user is difficult task. To address the above challenge Reinforcement Learning algorithms plays crucial role because these algorithms very much helpful in dealing with the dynamic environment and large space. This paper reviews the different Reinforcement algorithms namely Deep Qlearning network (DQN), Deep Deterministic Policy Gradient (DDPG) and Twin Delayed DDPG (TD3) to develop the news recommendation system and also mentioned the challenges faced by the reinforcement recommendation systems. In this study it was found that TD3 is best suited to develop the news recommendation system.

URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9395812&isnumb er=9395748

7. X. Gao and M. Qiu, "**Recommendation System Design for Social Media using Reinforcement Learning**," 2022 IEEE 9th International Conference on Cyber Security and Cloud Computing (CSCloud)/2022 IEEE 8th International Conference on Edge Computing and Scalable Cloud (EdgeCom), Xi'an, China, 2022, pp. 6-11, doi: 10.1109/CSCloud-EdgeCom54986.2022.00011.

   Abstract: Recommendation System plays an important role in capturing consumers' preference. In order to better utilize the effectiveness of the recommendation system, many social media platforms have developed algorithms to improve their performance. However, there are still a lot of platforms that keep recommending the same but useless contents to the customers, which even reduce the customers' interest in using these social media platforms. In response to this problem, we propose to transplant reinforcement learning to build efficient and effective recommendation system. Specifically, we decide to update our recommendation system frequently by taking all customers' behaviors into consideration to improve accuracy of the system's recommendation output. The experiment results show that our system can gain more than 100% profits and can converge to the optimal result quickly.

   URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9842879&isnumb er=9842867

8. K. S and G. K. Shyam, "**Review of Deep Reinforcement Learning-Based Recommender Systems**," 2022 Third International Conference on Smart

Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 2022, pp. 1-12, doi: 10.1109/ICSTCEE56972.2022.10099739. **Abstract:** Recommender systems (RS) are an indispensable technology that helps to address the issue of information explosion in the digital age. It can assist the customers by proposing personalized items and providers by increasing traffic to their website. Recommender systems can be applied to variety of business use cases including but not limited to electronic commerce applications and media recommendation (e.g., news, motion pictures, music, and videos). Deep reinforcement learning (DRL)-based RS has recently gained popularity as a research topic. Because of its interactive approach and autonomous learning ability, it frequently outperforms classic recommendation approaches, including deep learning-based methods. The objective of this work is to present a thorough overview of the state of the art for DRL implementations in recommendation systems. This paper starts with the background technologies involved in applying DRL in RS. Then, this paper examines recent advancements in DRL-based RS and discusses open issues. This review serves as an introduction to the topic of DRL in RS and its various facets. It identifies potential areas of research and provides valuable feedback to readers from academic and industry.

URL:
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10099739&isnum ber=10099487

9. Z. Yuyan, S. Xiayao and L. Yong, "**A Novel Movie Recommendation System Based on Deep Reinforcement Learning with Prioritized Experience Replay**," 2019 IEEE 19th International Conference on Communication Technology (ICCT), Xi'an, China, 2019, pp. 1496-1500, doi: 10.1109/ICCT46805.2019.8947012.
Abstract: A recommendation system plays an important role in information overload case by recommending personalized services to improve user experience. In this paper, a novel movie recommendation system based on deep reinforcement learning (DRL) framework is proposed. In proposed system model, the state information is preprocessed to overcome the problems of data sparsity and cold start. Specially, user's interest change is

captured using cross entropy and used to prioritize experience replay in a replay memory. The experiments verify that the proposed model can speed up the network update and improve recommendation accuracy.

URL:
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8947012&isnumb er=8946993

10. N. Guo, Z. Fu and Q. Zhao, "**Multimodal News Recommendation Based on Deep Reinforcement Learning**," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2022, pp. 279-284, Doi: 10.1109/ICSP54964.2022.9778361.

Abstract: Multimodal news recommendation is a challenging problem due to the rapid expansion of Internet information, bringing different levels of knowledge expression such as text, images, audio, and video, etc. In this paper, we propose a multimodal news recommendation method based on a deep reinforcement learning framework to represent user interests as multimodal information. The proposed method feeds the multimodal fusion feature into the rainbow agent of deep reinforcement learning to learn news representation. The experiment on the MIND and IM-MIND datasets gives the result of AUC, MRR, nDCG@5, and nDCG@10 scores, which outperform LSTUR, FIM, and DKN, NRMS, NPA, and DeepFM models. It shows that reinforcement learning is an excellent choice to fulfill recommendation tasks, and the multimodal fusion feature is effective for learning accurate news representations.

URL:
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9778361&isnumb er=9778231

Proposed System

Upon extensive understanding of the inferences, a better and higher metric value and accuracy can be achieved from the changes made in the original network architecture. Improving the candidate selection process or the off policy training can significantly boost the performance of the model. Keeping these insights in mind, our aim is to experiment further on these topics and adopt newer and faster models in these modules.

Implementation

```
import argparse import os
os.environ['CUDA_LAUNCH_BLOCKING'] = "1" import numpy as
np import torch import torch.nn as nn import torch.nn.functional
as F from data import MovieLensDataset,
SyntheticMovieLensDataset from loss import loss_2s, loss_ce,
loss_ips from metric import Evaluator from model import
ImpressionSimulator, Nominator, Ranker from six.moves import
cPickle as pickle from torch.utils.data import DataLoader, Subset
from torchnet.meter import AUCMeter from MGA import
MGAEmbedding
import sys


parser = argparse.ArgumentParser()
#---------------MGA-------------------------------------
parser.add_argument('--n_epoch',   type=int,   default=12)
parser.add_argument('--n_hid',      type=int,  default=256)
parser.add_argument('--n_inp',      type=int,  default=128)
parser.add_argument('--clip',         type=int,   default=1.0)
parser.add_argument('--max_lr',  type=float, default=1e-3)
parser.add_argument('--dataset', default="ML")


#------------------main---------------------------------------------
parser.add_argument("--verbose", type=int, default=1, help="Verbose.")
parser.add_argument("--seed", type=int, default=0, help="Random seed.")
parser.add_argument("--loss_type", default="loss_2s") parser.add_argument("--device",
default="cuda") parser.add_argument("--alpha", type=float, default=1e-3, help="Loss
ratio.") parser.add_argument("--lr", type=float, default=0.05, help="Learning rate.")
parser.add_argument("--simulator_epoch", type=int, default=10, help="simulator_epoch")
parser.add_argument("--logging_epoch", type=int, default=40, help="simulator_epoch")
parser.add_argument("--ranker_epoch", type=int, default=10, help="simulator_epoch")
parser.add_argument("--train_epoch", type=int, default=40, help="simulator_epoch")
args = parser.parse_args() torch.manual_seed(0) torch.cuda.manual_seed(0) class
Logger(object):    def __init__(self,fileN ="Default.log"):
    self.terminal = sys.stdout
self.log = open(fileN,"a")     def
write(self,message):
self.terminal.write(message)
self.log.write(message)    def
flush(self):
    pass

filepath = "./data/" device = args.device dataset =
MovieLensDataset(filepath, device=device)
NUM_ITEMS = dataset.NUM_ITEMS
NUM_YEARS = dataset.NUM_YEARS
NUM_GENRES = dataset.NUM_GENRES
NUM_USERS = dataset.NUM_USERS
NUM_OCCUPS = dataset.NUM_OCCUPS
NUM_AGES = dataset.NUM_AGES NUM_ZIPS
= dataset.NUM_ZIPS
sys.stdout = Logger("./logs/MGA_epoch_{}_loss_{}_alpah_{}_lr_{}_train_epoch_{}_emb_{}.txt".format(args.n_epoch,
args.loss_type.split("_")[1], args.alpha, args.lr, args.train_epoch, args.n_inp)) print('Training with args:\n', args)
#---------------------------Generating pre-trained embeddings------------------- context_emb =MGAEmbedding(args,
filepath) user_embedding = context_emb.G.nodes['user'].data['inp'] item_embedding =
context_emb.G.nodes['movie'].data['inp'] year_embedding = context_emb.G.nodes['year'].data['inp']
gender_embedding = context_emb.G.nodes['gender'].data['inp'] age_embedding =
context_emb.G.nodes['age'].data['inp'] occupation_embedding =
context_emb.G.nodes['occupation'].data['inp'] zip_embedding = context_emb.G.nodes['zip-code'].data['inp']
print("load embedding complete! with size: ", 'user:', user_embedding.shape, 'movie:', item_embedding.shape,
  'year:', year_embedding.shape,'gender:', gender_embedding.shape , 'age:',age_embedding.shape ,
```

Output

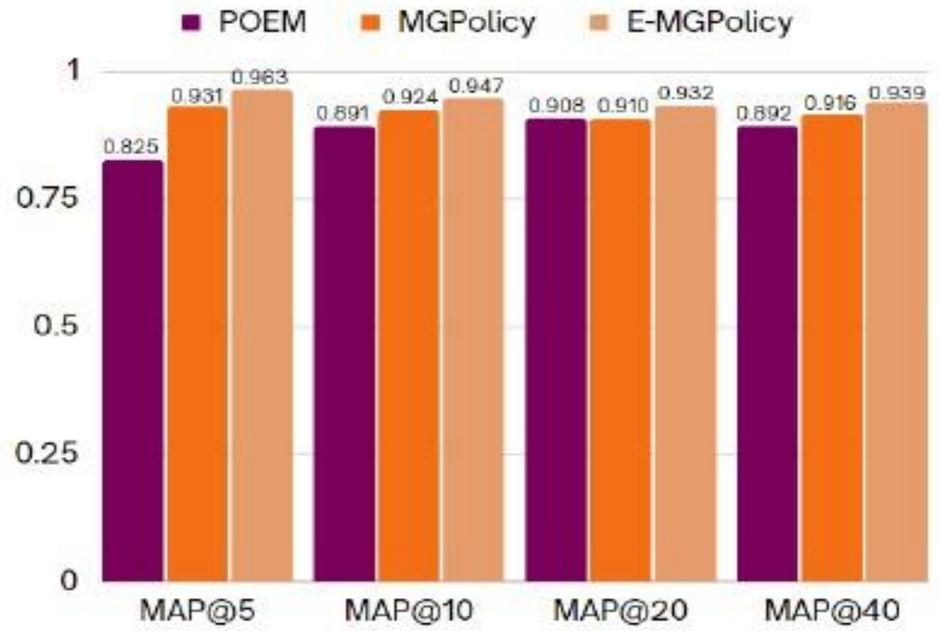

Figure 2 MAP@K Results

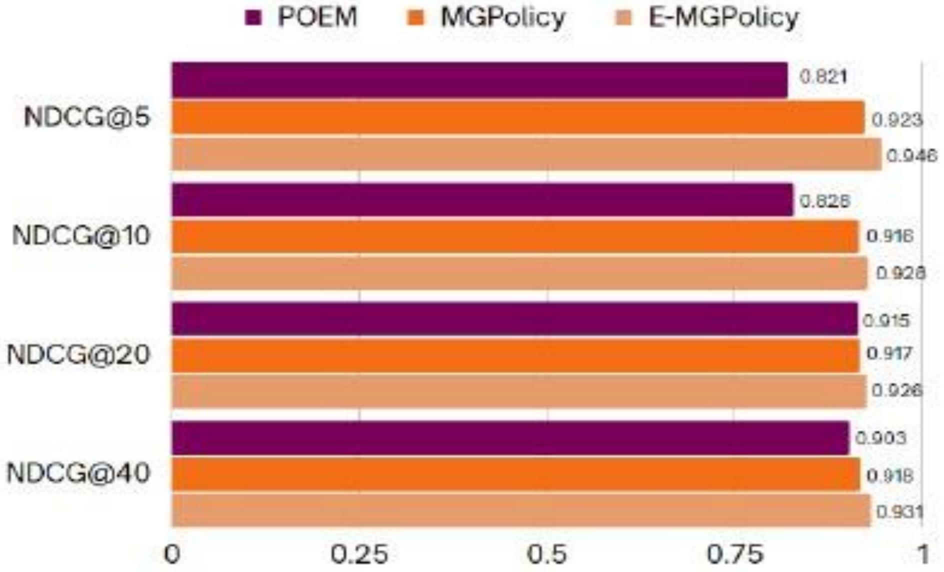

Figure 3 NDCG@K Results

Results & Discussion

As seen from Figures 2 and 3, It is clearly visible that the proposed methodology has improved on the metrics MAP@K and NDCG@K. We ran experiments for the Top-K recommendations, where K ranges from 5,10,20 and 40. This is to ensure that a different number of searches or preferences can accommodate for the change number of predictions. It also ensures that there has been no overfitting of the data as a larger number of recommendations are also showcasing a valid trend.

The results show us various inferences, such as the use of R2N2s instead of

RNNs. For the State representations, the MGPolicy work uses RNNs while our EMGPolicy Network adopts Residual Recurrent Neural Networks. They utilize a slightly longer memory chain for training and hence go deeper while understanding the features from the meta graph vector representations. We also experimented with LSTM Gates at the end but the results were unsatisfactory in comparison to GRU gates R2N2s.

Another inference from the study shows that increasing the efficiency of the Off-policy training can increase the recommendation generation process. This can result in a more unbiased method of training, slightly better than that of MGPolicy.

As seen from the graphs, the best values have been achieved for top-5 recommendation space, while a small drop is observed in the top-10 recommendation area. The top 20 recommendation space still remains a hinderance in the values achieved, primarily due to the invariancies and incorrectly ranked recommendations. Given the importance of recommendations in the top 20 space, a small fault in the distribution is resulting in a drop of metric values. As for top 40 recommendation space, since the farther we move from the top ranked recommendations, their significance towards the final metric value decreases. Hence a major redistribution in the space is not going to have the same effect as a one occurring in the top 20 recommendation space.

Conclusion

In this thesis, we presented the Enhanced-Meta Graph Policy Network that incorporate the use of enhanced off-policy learning coupled with effective state representations extrapolated from meta graphs of user item interactions. An effective candidate selection algorithm is used for ensuring the correctness of the off-policy training. As part of our study, we also present some of the latest works in the field and the need for reinforcement learning models to alleviate the effectiveness of recommendation models. The proposed model has reached a comparably higher metric values for MAP and NDCG ranking metrics in comparison to the already existing MGPolicy and POEM networks. This work is done in the hopes that it can motivate future researchers in the field to take insights from the results achieved. We also aimed at reducing the sparsity and diversity issue in the Movie Lens dataset, and the new results show the proof of this attempt.

References

[1] G. Pang, X. Zhu, K. Lu, Z. Peng and W. Deng, "**A simulator for reinforcement learning training in the recommendation field**," 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Exeter, United Kingdom, 2020, pp. 10371042, doi: 10.1109/ISPA-BDCloud-SocialCom-SustainCom51426.2020.00156.

[2] M. Pham, H. Nguyen, L. Dang and J. A. Nieves, "**Compressive Features in Offline Reinforcement Learning for Recommender Systems**," 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 5719-5726, doi: 10.1109/BigData52589.2021.9671419.

[3] X. He, K. Wang, H. Lu, W. Xu and S. Guo, "**Edge QoE: Intelligent Big Data Caching via Deep Reinforcement Learning,**" in IEEE Network, vol. 34, no. 4, pp. 8-13, July/August 2020, Doi: 10.1109/MNET.011.1900393.

[4] N. Yambem and A. N. Nandakumar, "**Optimizing Hadoop parameter for speedup using Q-Learning Reinforcement Learning**," 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 2021, pp. 1-7, Doi: 10.1109/ICECCT52121.2021.9616965.

[5] M. Tracolli, M. Baioletti, V. Poggioni and D. Spiga, "**Effective Big Data Caching through Reinforcement Learning**," 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 2020, pp. 1499-1504, Doi: 10.1109/ICMLA51294.2020.00231.

[6] S. Bangari, S. Nayak, L. Patel and K. T. Rashmi, "**A Review on Reinforcement Learning based News Recommendation Systems and its challenges**," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 260-265, Doi: 10.1109/ICAIS50930.2021.9395812.

[7] X. Gao and M. Qiu, "**Recommendation System Design for Social Media using Reinforcement Learning**," 2022 IEEE 9th International Conference on Cyber Security and Cloud Computing (CSCloud)/2022 IEEE 8th International Conference on Edge Computing and Scalable Cloud (EdgeCom), Xi'an, China, 2022, pp. 6-11, doi: 10.1109/CSCloud-EdgeCom54986.2022.00011.

[8] K. S and G. K. Shyam, "**Review of Deep Reinforcement Learning-Based Recommender Systems**," 2022 Third International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 2022, pp. 1-12, doi: 10.1109/ICSTCEE56972.2022.10099739.

[9] Z. Yuyan, S. Xiayao and L. Yong, "**A Novel Movie Recommendation System Based on Deep Reinforcement Learning with Prioritized Experience Replay**," 2019 IEEE 19th International Conference on Communication Technology (ICCT), Xi'an, China, 2019, pp. 1496-1500, doi: 10.1109/ICCT46805.2019.8947012.

[10] N. Guo, Z. Fu and Q. Zhao, "**Multimodal News Recommendation Based on Deep Reinforcement Learning**," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2022, pp. 279-284, Doi: 10.1109/ICSP54964.2022.9778361.

[11] Xiangmeng Wang, Qian Li, Dianer Yu, Zhichao Wang, Hongxu Chen, Guandong Xu, **MGPolicy: Meta Graph Enchaned Off Policy Learning for Recommendations**