Master's Thesis Presentation for the 142th Dissertation Evaluation

A Study of Effective Micro-video Recommendation via Explicit Interaction Separation Reflecting Novelty

참신성이 반영된 명시적 상호작용 세분화를 통한 마이크로 비디오 추천 개선 연구

Sangkeun Park

psk492@gmail.com

13st Nov. 2024

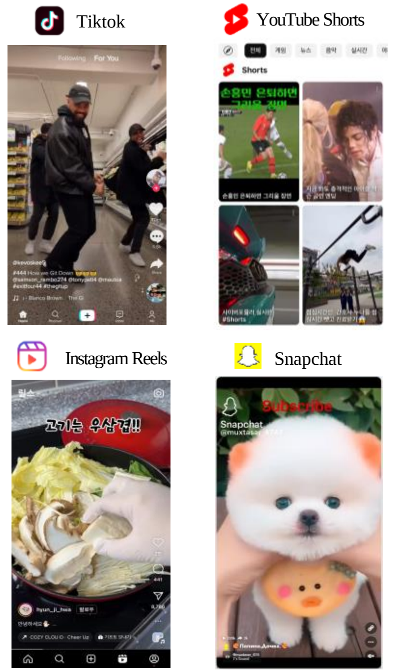
Contents

[1] S. Liu et al., "User Conditional Hashtag Recommendation for Micro-Videos," *2020 IEEE Int. Conf. Multimed*. Expo, London, UK, pp. 1-6, Jul. 2020.
[2] A. Chaudhary et al., "“Are you still watching?”: exploring unintended user behaviors and dark patterns on video streaming platforms.“ in *Proc. ACM Design. Interac. Sys. Conf.*, Australia, 2022, pp. 776-791.

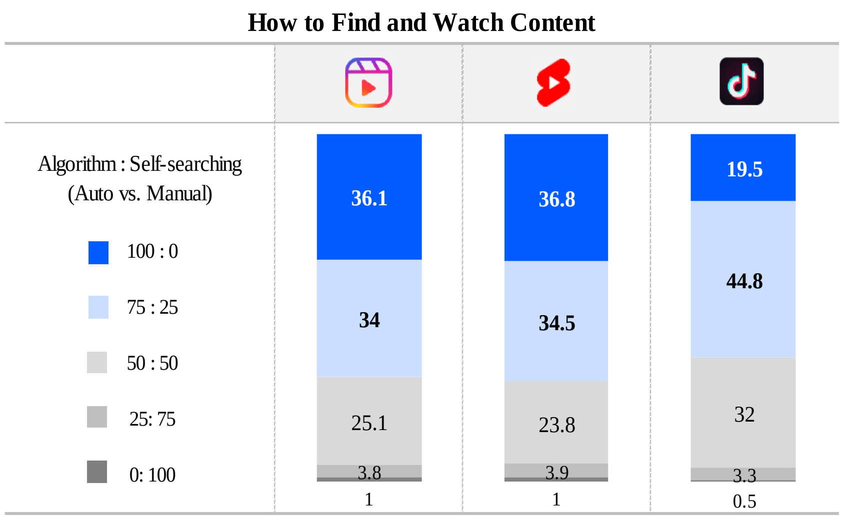
Introduction

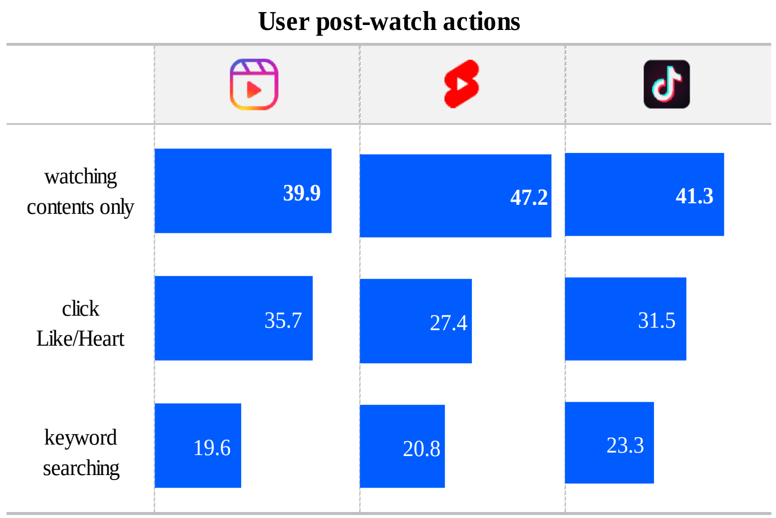
Micro-video Recommendation System (MvRS)

- Micro-videos, also known as short-form content, are brief media pieces created by individuals that showcase unique creativity and insight. They have rapidly increased in recent years due to their social nature and easy shareability on digital platforms [1].
- With boosting content availability, users struggle to navigate diverse topics leading to a reliance on automatic playback and indiscriminate viewing of undesired content without skipping, requiring a Micro-video Recommendation System (MvRS) [2].



Examples of Major Platforms for Micro-video





Source: Social Media Search Portal Trend Report 2023, OpenSurvey

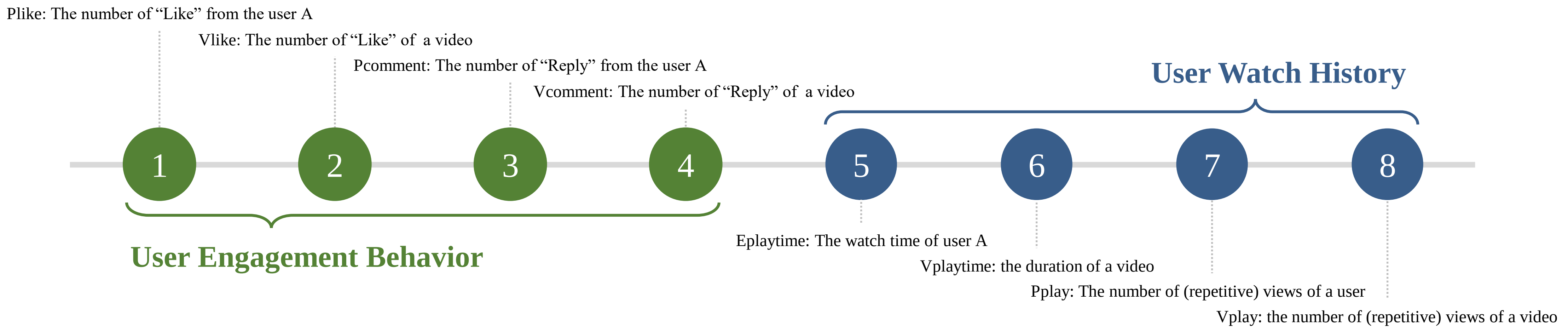
Statistics of Micro-video Usage in Major Platforms

[3] Z. Fu et al., “Deep Learning Models for Serendipity Recommendations: A Survey and New Perspectives,” *ACM Comput. Surv.*, vol. 56, no. 19, pp. 1-26, Aug. 2023
[4] M. Wilhelm et al., "Practical Diversified Recommendations on YouTube with Determinantal Point Processes," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manag.*, Torino, Italy, 2018, pp. 2165-2173.
[5] D. Klug et al., “Trick and Please. A Mixed-Method Study On User Assumptions About the TikTok Algorithm,” in *Proc. 13th ACM Web Sci. Conf.*, United Kingdom, 2021, pp. 84-92.
[6] M. Boeker et al., “An Empirical Investigation of Personalization Factors on TikTok,” in *Proc. ACM Web Conf.*, Lyon, France, 2022, pp. 2298-2309.

Objective and Standard Procedure

- The MvRS aims to enhance recommendation quality by introducing novel and unique items rather than relying solely on viewing history. Maintaining the freshness and originality of suggested content is a critical challenge in the recommendation system [3].
- Main platforms such as TikTok and YouTube employ MvRS, which utilizes various behavioral data, including replies, hashtag, and likes, by extracting features from content modality data to ensure originality and diversity while preserving user interest [4, 5, 6].

(Plike × Vlike) + (Pcomment × Vcomment) + (Eplaytime × Vplaytime) + (Pplay × Vplay)



TikTok places an emphasis on user behavior and engagement levels as well as individual watch history for recommendation.

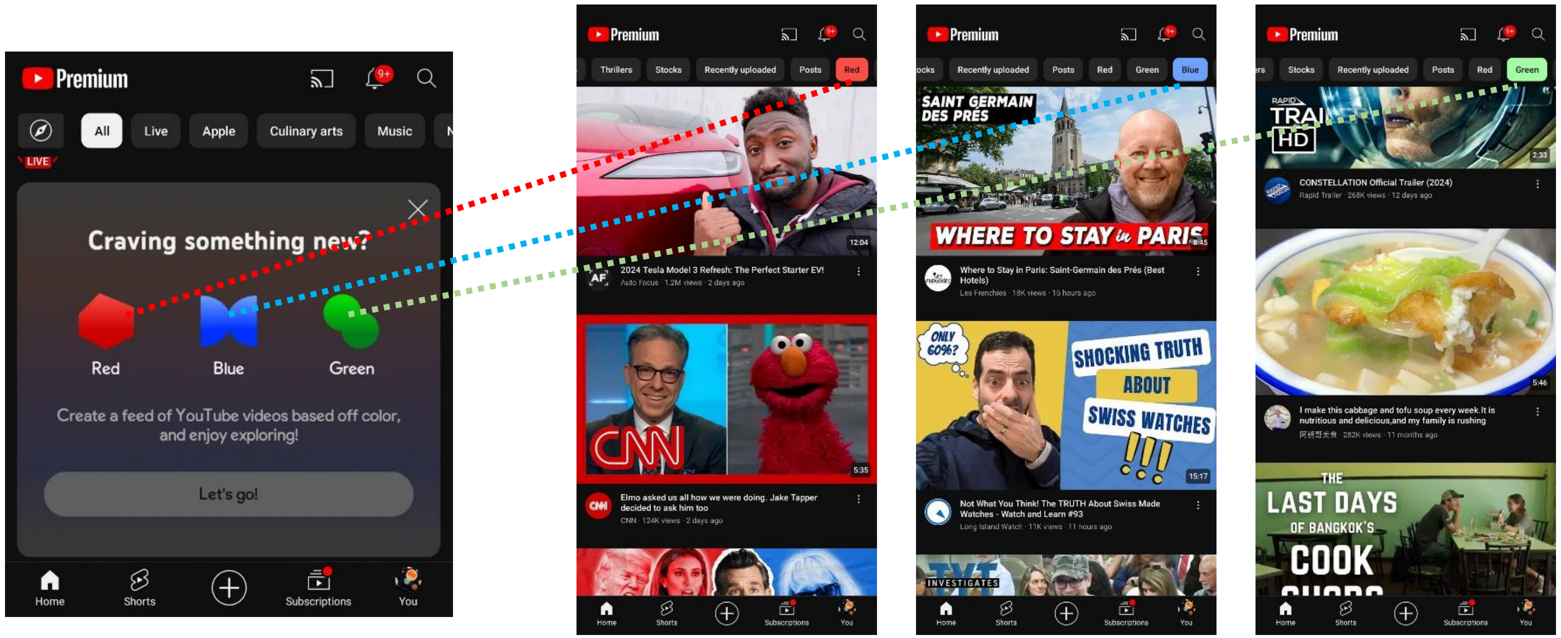
How Tiktok Reads Your Mind <https://www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html>
How Tiktok Recommends Videos #ForYou <https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you>

A Recommendation Equation in Tiktok

Introduction

Problem and Strategy

- Most MvRS models classify interactions based on watch time or history, categorizing them straightforwardly as 'positive' for viewing and 'negative' for skipping. This oversimplification complicates the interpretability of the model and degrades its performance.
- The automatic playback worsens this issue by causing indiscriminate repetitive viewing of specific interests. Therefore, categorizing viewed videos as 'novel' or 'ordinary' explicitly allows for considering novelty and serendipity, the fundamentals of recommendation.



Major signals for custom curation without viewing history in YouTube	
Clicks	Indicate **initial interest** but may not reflect actual satisfaction or full engagement with the content.
Survey Responses	Collects user ratings to gauge **satisfaction**, focusing on highly rated content (four to five stars, colors)
Sharing, Likes, Dislikes	Sharing and likes generally signal satisfaction, while dislikes indicate the opposite, with the system adjusting based on individual **user behavior**.

On YouTube's Recommendation System <https://blog.youtube/inside-youtube/on-youtubes-recommendation-system>
YouTube Tries RGB Color Feeds <https://9to5google.com/2024/02/04/youtube-color-feeds/>

A Survey and Application for Novelty in YouTube

Related Work

Learning Fine-grained User Interests (FRAME) [7]

- Graph Construction: Features extracted from **video clips are segregated** into two sets based on **complete views and skips**, from which **positive and negative interaction graphs are formed**.
- Drawbacks: Considering the **brief duration and autoplay characteristic** of micro-videos, **assuming that all users who watch an entire video have a favorable preference may be an uncertain oversimplification**.

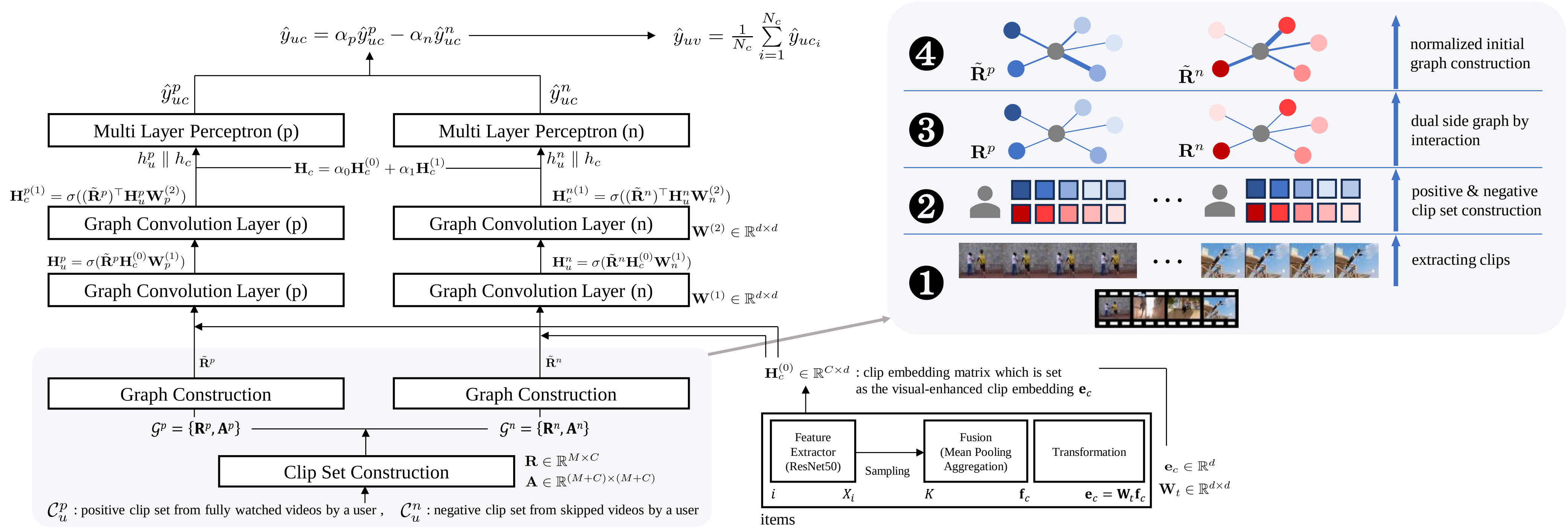
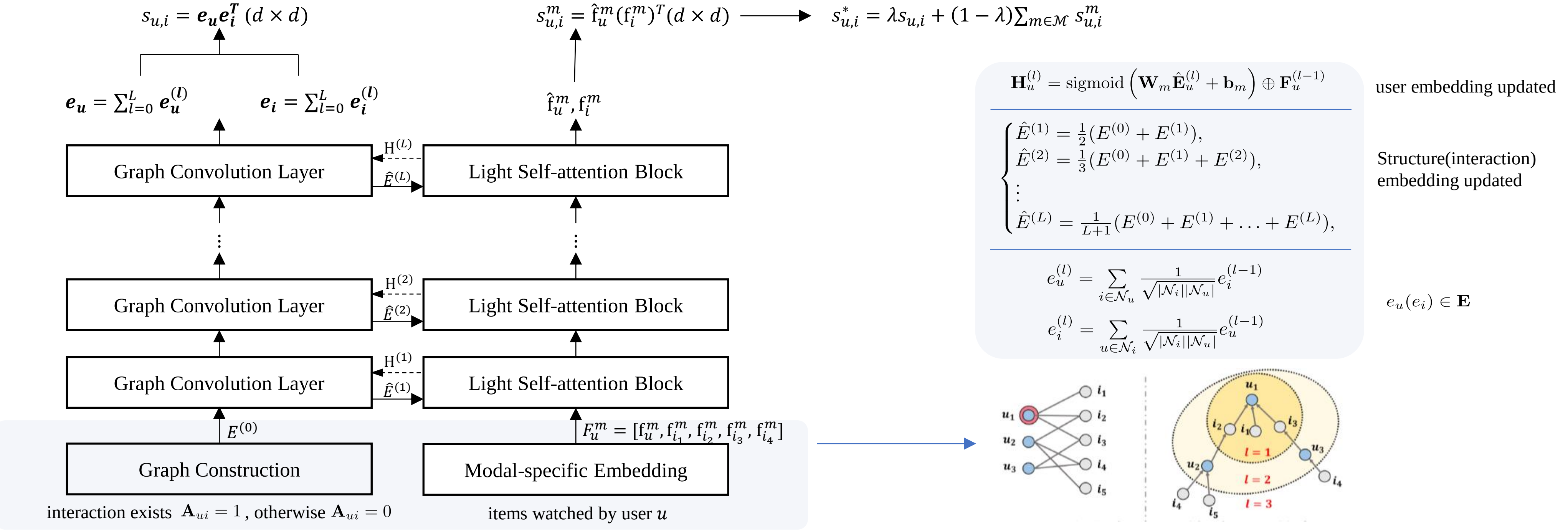


Illustration of the FRAME model

Related Work

Light Graph Transformer (LightGT) [8]

- Graph Construction: A bipartite graph is created from **video features and users with viewing histories**, facilitating the **learning of structural information through common neighbors**.
- Drawbacks: This approach **only considers users** who view records as **positive relations, disregarding those without as negative** and **overlooks obvious and valuable information** from user behaviors **such as skipping**.



Overall Workflow of the LightGT Model

Common Drawback and Solution

- Conventional Methods need to focus more on content novelty and viewing patterns rather than simply treating viewing history as positive or negative. Consequently, existing models exhibit significant limitations in effectively recommending micro-videos.
- By measuring the similarity of modalities from fully watched videos and constructing graphs that consider ordinary, novel, and negatively skipped content, the model recommends similar content from a novelty perspective using a specific scoring mechanism.

Method	Name / Year	Reference	Summary	Drawback
FRAME	Y. Shang / 2023	46th International ACM SIGIR Conference on Research and Development in Information Retrieval	• **segmenting the video into clips** and extracting features, **viewing the entire clip is considered a positive** interaction, while **skipping is considered a negative** interaction	• Assuming that all **users who watch an entire video** have a **favorable preference may be an uncertain oversimplification**
LightGT	Y. Wei / 2023	46th International ACM SIGIR Conference on Research and Development in Information Retrieval	• **learning structural information between common neighbor nodes** based on the viewing history of users and the **node embedding updated through a self-attention block**	• This approach **considers only views as positive interactions, ignoring valuable behavioral data like skipping**

Proposed Method

Notation and Definition

- $\mathcal{U}, \mathcal{I}, \mathcal{R}$ are collections of data, where the embeddings of user $\mathcal{U}$ and item $\mathcal{I}$ are denoted as $\mathbf{e}$, and the modality-specific features of $\mathcal{I}$ are represented by $\mathbf{f}$. The relationships formed between u and i, included in $\mathcal{R}$, contain preference labels y with watching time ratio.
- To utilize micro-video aspects, item i is divided into clips c, and clip set C_u are constructed based on interactions $\mathcal{R}$ with user u. The recommendation model is trained by minimizing loss, comparing labels y with predicted score $\hat{y}$ for the final prediction.

Symbol	Definition	Description
$\mathcal{U},\ \mathcal{I},\ \mathcal{R}$	$\mathcal{U} = \{u_1, u_2, ..., u_j\}$, $\mathcal{I} = \{i_1, i_2, ..., i_k\}$, $\mathcal{R} = \{(u, i, y) \mid u \in \mathcal{U}, i \in \mathcal{I}\}$	The set of users, items(videos), and user-item interactions
C_i	$C_i = \{c_1, c_2, \ldots, c_l\}$	The clip set of videos
$C_u{}^{\mathbf{intrct.}}$	$C_u{}^{\mathbf{intrct.}} = \cup_{i=1}^{k} \{c_{i,1}{}^{\mathbf{intrct.}}, c_{i,2}{}^{\mathbf{intrct.}}, \ldots, c_{i,l}{}^{\mathbf{intrct.}}\}$	The interacted clip set of users
$\mathbf{e},\ \mathbf{f}$	$\mathbf{e}_u, \mathbf{e}_i, \mathbf{f}_u^m, \mathbf{f}_i^m \in \mathbb{R}^d$, $m \in \mathcal{M} = \{v, a, t\}$	Embedding and Feature representation
$\hat{x}$	$\hat{x} = max_{\ c^{\mathbf{pos}},\, c^{\mathbf{nov}} \in C,\, c^{\mathbf{pos}} \neq c^{\mathbf{nov}}}\ Sim(c^{\mathbf{pos}},\ c^{\mathbf{nov}})$	Novelty score based on similarity
$y,\ \hat{y}$	$y = \frac{watch\ t.}{duration\ t.}$, $\hat{y} = \alpha \hat{x}^{\mathbf{Nov}} + \hat{y}^{\mathbf{Frm}}$	Preference label and predicted score

Proposed Method

Procedural Overview

- Each video is evenly split into four parts to utilize interactions precisely. Skipped clips are negative, and previous clips are positive. The similarity-based novelty scores categorize positive clips as ordinary or novel, creating interaction graphs.
- The novelty score $\hat{x}$ is derived using user and clip embeddings through Graph Convolutional Networks (GCNs). Aggregating the conventional FRAME score $\hat{y}$ for ordinary clips and the $\hat{x}$ with a novelty impact factor α yields the final score $\hat{y}_{u,v}$ at the video level.

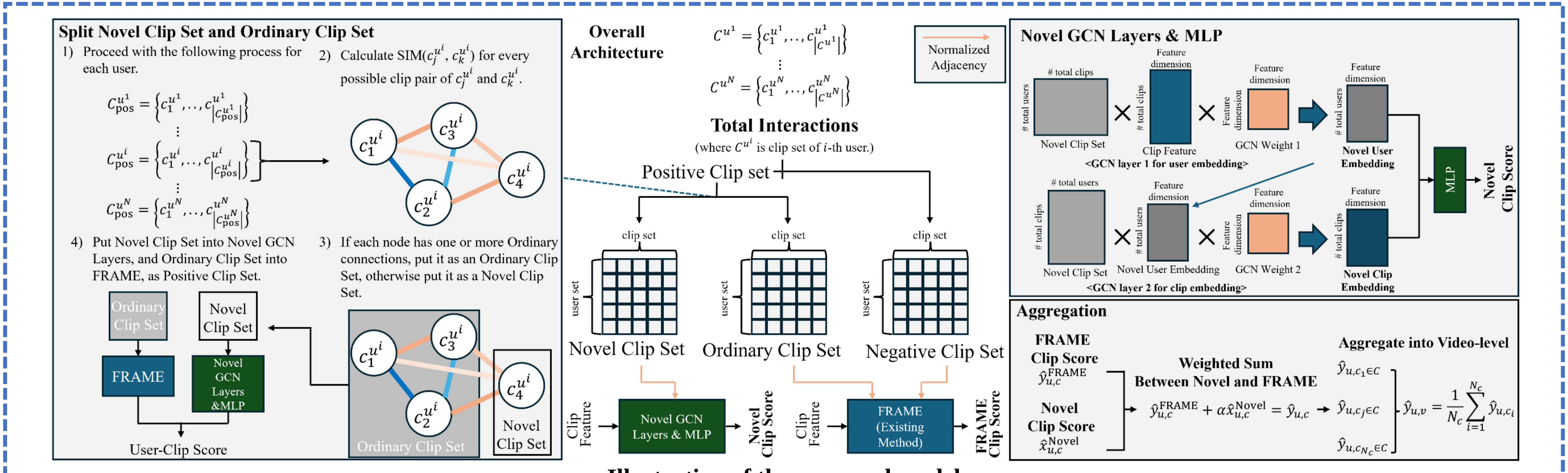
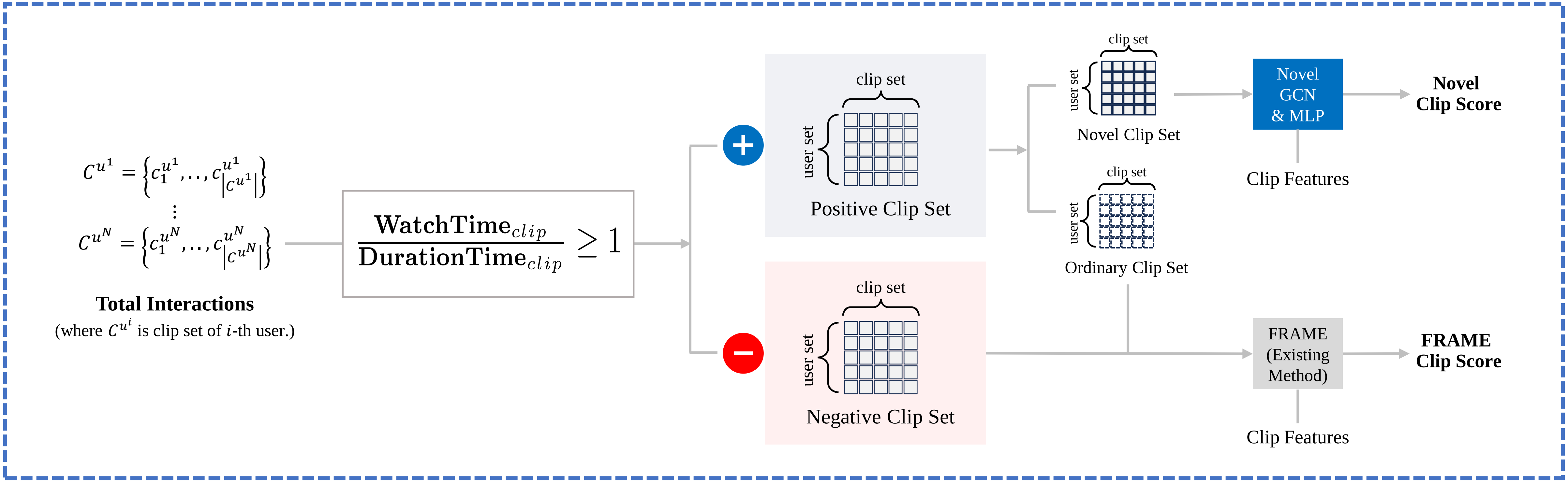


Illustration of the proposed model

Proposed Method

Details

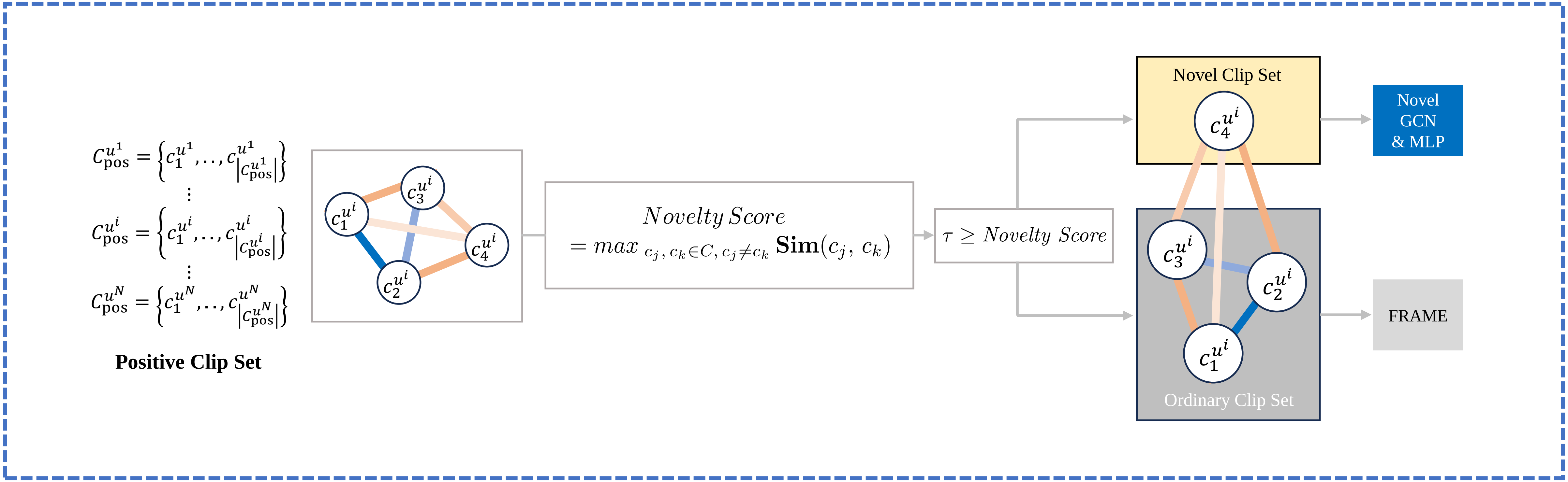
- In the context of user interactions, if the viewing duration for each clip is less than the total length, a skip has occurred. All clips before the clip that occurred skipping actions should have positive interactions, whereas the skipped clip is deemed negative.
- Positive interactions are categorized based on novelty into novel clips and ordinary clips. The negative clips from the previous step and ordinary clips that were just categorized are then utilized as inputs for the existing FRAME model.



Proposed Method

Details

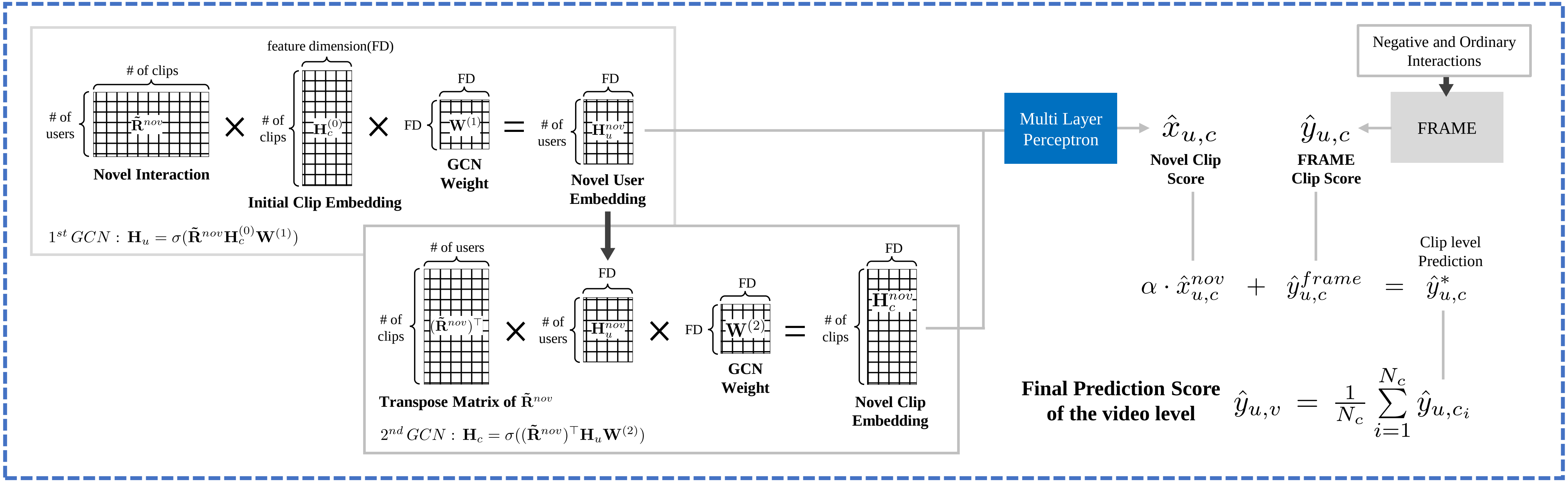
- The novelty score is determined by calculating the cosine similarity between the feature vectors of clip pairs within the positive clip set, where an output value closer to 0 indicates a higher novelty of the clip.
- If the calculated score is equal to or greater than a threshold t=0.75, the clip is classified as an ordinary clip; otherwise, a novel clip. This classification allows for the extraction of focused user interest in novel clips.



Proposed Method

Details

- GCN is applied by constructing a user-clip graph marking novel clips, transforming it into a user embedding matrix, obtaining clip embeddings via $\mathbf{H}_u^{nov} = \sigma(\tilde{\mathbf{R}}^{nov}\mathbf{H}_c^{(0)}\mathbf{W})$ and concatenating these embeddings with an MLP to derive the novel clip prediction score $\hat{x}$.
- The score for each clip is calculated as $\alpha\hat{x} + \hat{y}$, where α is a novelty impact parameter adjusting the weight assigned to the similarity of novel clips, a value determined by the user. The final prediction is derived by taking the average scores of clips within a video.



Experimental Settings

Dataset

- The experiments were conducted with two real-world datasets collected from actual platforms to compare with previous studies.

Dataset	#Interactions	#Users	#Videos	V	A	T
MicroVideo-A	342,694	12,739	58,291	128	-	-
Tiktok	726,065	36,656	76,085	128	128	128

Configuration

- Maximum 30 epochs in 10 iterations for each dataset
- Data Split: (Train : Validation : Test) = (0.8 : 0.1 : 0.1), Optimizer: Adam, Learning Rate: 1e-2 to 1e-6
- Environments: NVIDIA GeForce RTX 3090, Pytorch 2.0

Evaluation Measures

- Recall@k: The ratio of actual user-preferred items included in the top k items recommended by the model.
- NDCG@k: The ratio of the top k recommended items that the model has ranked in the same order as the user actual preferences.

Experimental Results

Performance Comparison

- The proposed model effectively aggregates user preference signals by considering novel interactions, achieving higher evaluation metric scores than comparison methods.

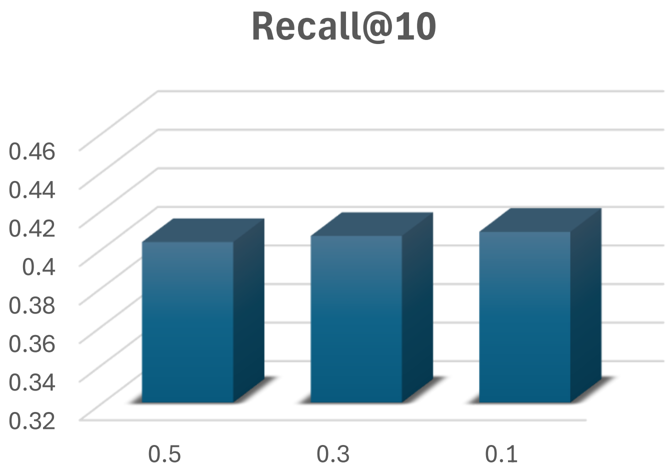
Dataset	**MicroVideo-A**				**Tiktok**			
Method	Recall@10	NDCG@10	Recall@20	NDCG@20	Recall@10	NDCG@10	Recall@20	NDCG@20
Proposed	**0.4085 ± 0.0058**	**0.5667 ± 0.0101**	**0.5384 ± 0.0036**	**0.6925 ± 0.009**	**0.2517 ± 0.0062**	**0.0910 ± 0.0057**	**0.5009 ± 0.0033**	**0.1531 ± 0.0060**
FRAME	0.3556 ± 0.0163	0.4716 ± 0.0341	0.4918 ± 0.0143	0.6031 ± 0.0317	0.1493 ± 0.0041	0.0585 ± 0.0018	0.4186 ± 0.0041	0.1247 ± 0.0019
LightGT	0.0202 ± 0.0024	0.0132 ± 0.0014	0.0529 ± 0.0092	0.0275 ± 0.0047	0.1186 ± 0.0038	0.0632 ± 0.0021	0.3945 ± 0.0039	0.1194 ± 0.0020

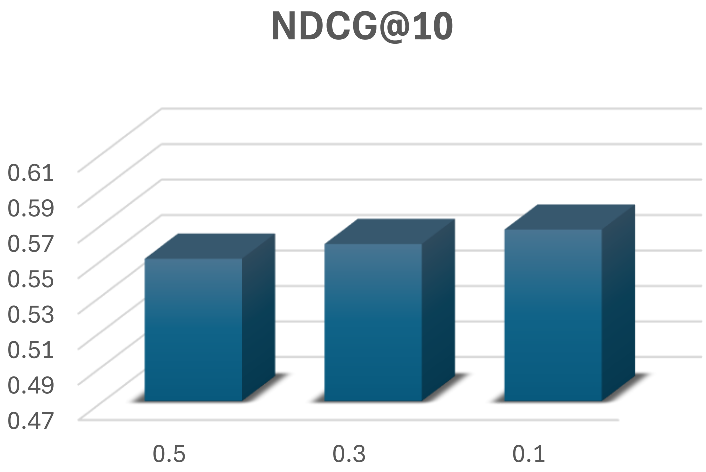
Experimental Results

In-depth Analysis

- The experiment conducted on the proposed model using the MicroVideo-A dataset aimed to investigate the impact of changes in the novelty impact parameter α on performance. The results indicated that the model achieved optimal performance with the small α.
- Conversely, applying an excessive weight to novelty diminished the performance, suggesting that solid novelty recommendations might confuse the model. This finding implies that a high emphasis on novelty can potentially degrade the model's effectiveness.

MicroVideo-A	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$
Recall@10	0.4085 ± 0.0058	0.4064 ± 0.0057	0.4032 ± 0.0069
NDCG@10	0.5667 ± 0.0101	0.5585 ± 0.0286	0.5502 ± 0.0249



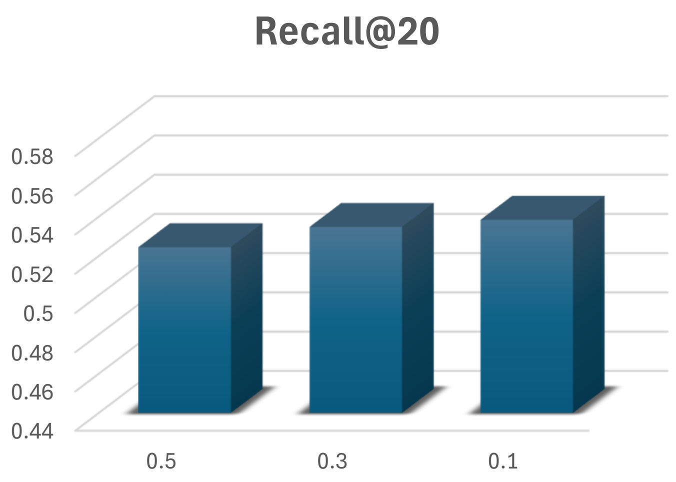


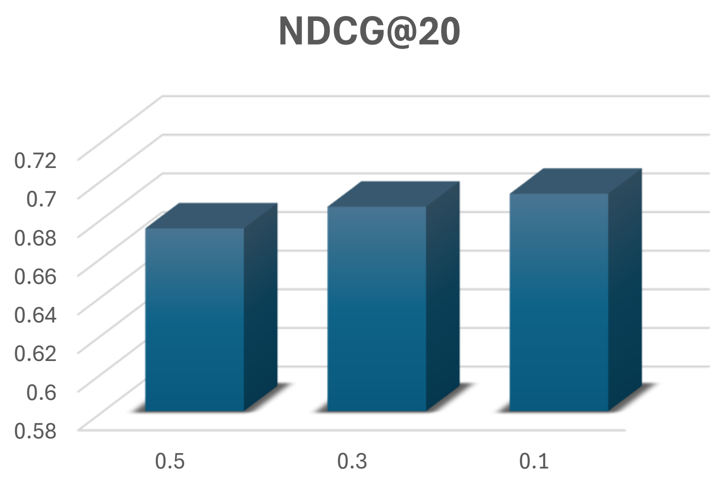
Experimental Results

In-depth Analysis

- Experimental results indicate that variations in the novelty impact parameter α do not significantly influence the performance. Consequently, the proposed model exhibits robustness to changes in α, allowing for relatively stable outputs under arbitrary settings.
- However, it is observed that performance tends to degrade when tremendous values are assigned to α. Therefore, it is empirically suggested to set α to approximately 0.1.

MicroVideo-A	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$
Recall@20	0.5384 ± 0.0036	0.5347 ± 0.0114	0.5244 ± 0.0125
NDCG@20	0.6925 ± 0.009	0.6858 ± 0.0235	0.6746 ± 0.0247





Conclusion and Contribution

- The research on the MvRS focuses on creating and verifying methods to suggest novel contents by analyzing user behaviors and item features. This will improve interaction models and data analysis to achieve higher recommendation performance.
- The unique characteristics of micro-videos, such as autoplay, complicate user-item interaction detection. To address this, the proposed method captures user preference signals more precisely by distinguishing ordinary and novel items based on their features.
- The proposed model surpassed existing methods and demonstrated effectiveness by performing better experiment results on real-world datasets. Additionally, further in-depth analysis provided insights into the model's robustness.

Achievement

- Conferences
  - A Survey of Recent Advances in Micro-video Recommendation Systems, ***IEIE***, 2024
  - A Simple Localization and Classification Framework Based on Template Matching and Self-Attention for TAO Treatment Diagnosis, ***ICCE***, 2024
- Competition - Medical Technology for Patients with Thyroid-associated Ophthalmopathy, CAU AI Technology-based Employment and Start-up Workshop, 2023, 3rd Prize
- Events - 2023 SW Education Festival [CAU Booth], Ilsan, Korea, 2023 , CAU AI Symposium [AutoML Lab Booth], Seoul, Korea, 2023

Thank you

Reference

[1] S. Liu et al., "User Conditional Hashtag Recommendation for Micro-Videos," *2020 IEEE Int. Conf. Multimed*. Expo, London, UK, pp. 1-6, Jul. 2020.

[2] A. Chaudhary et al., "“Are you still watching?”: exploring unintended user behaviors and dark patterns on video streaming platforms." in *Proc. ACM Design. Interac. Sys. Conf.*, Australia, 2022, pp. 776-791.

[3] Z. Fu et al., “Deep Learning Models for Serendipity Recommendations: A Survey and New Perspectives,” *ACM Comput. Surv.*, vol. 56, no. 19, pp. 1-26, Aug. 2023

[4] M. Wilhelm et al., "Practical Diversified Recommendations on YouTube with Determinantal Point Processes," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manag.*, Torino, Italy, 2018, pp.2165-2173.

[5] D. Klug et al., “Trick and Please. A Mixed-Method Study On User Assumptions About the TikTok Algorithm,” in *Proc. 13th ACM Web Sci. Conf.*, United Kingdom, 2021, pp. 84-92.

[6] M. Boeker et al., “An Empirical Investigation of Personalization Factors on TikTok,” in *Proc. ACM Web Conf.*, Lyon, France, 2022, pp. 2298-2309.

[7] S. Yu et al. "Learning Fine-grained User Interests for Micro-video Recommendation," in *Proc. 46th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Taipei, Taiwan, 2023, pp. 433-442.

[8] Y. Wei et al., “LightGT: A Light Graph Transformer for Multimedia Recommendation,” in *Proc. 46th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr*., Taipei, Taiwan, 2023, pp. 1508-1517.