Project Design - COMP9727

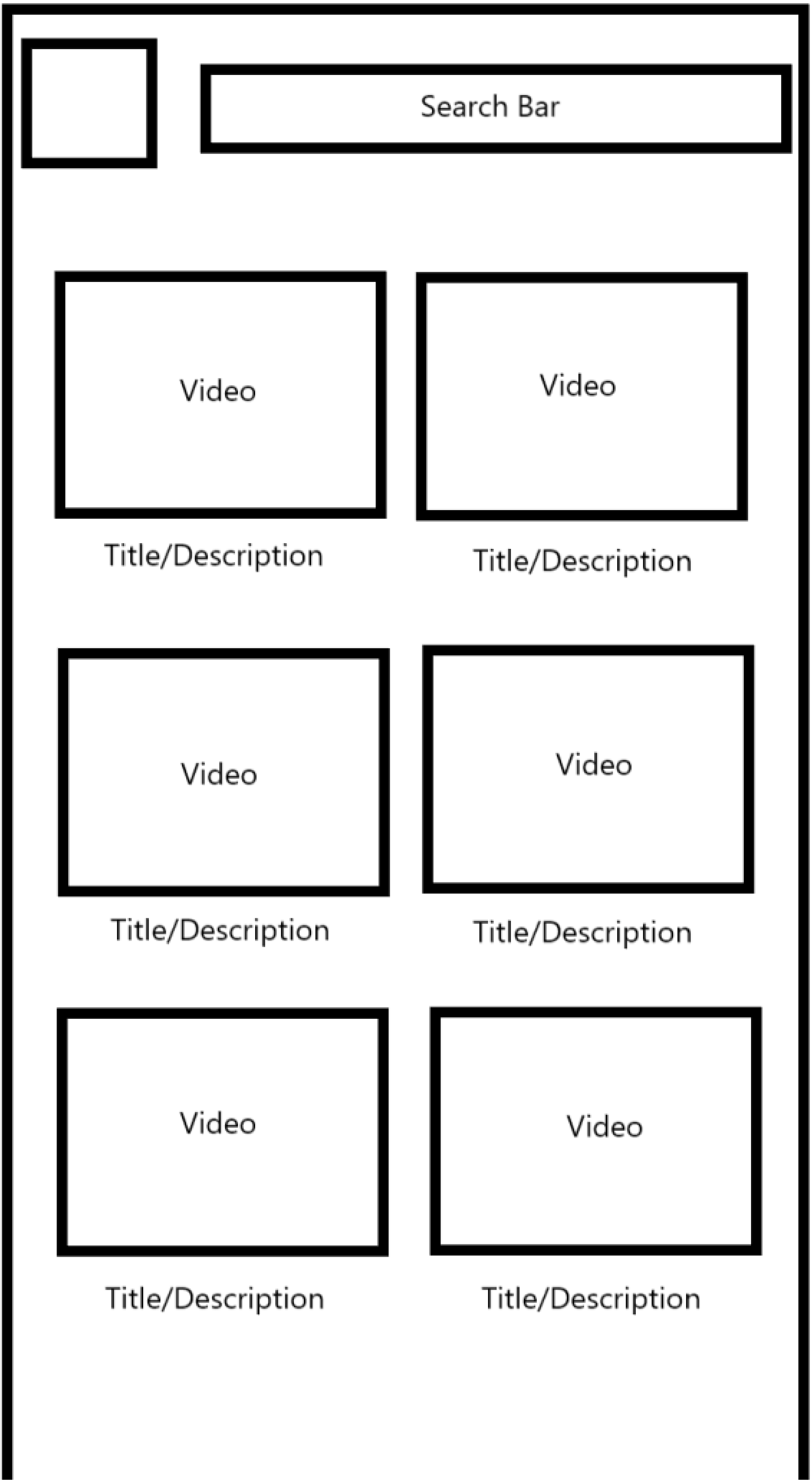
WenquWang z5539497

Scope

The system aims to recommend medium/short videos (about 60min or less) to viewers who aged 15-40 and seeking shattered entertainment, and extend their retention time as much as possible, by keep recommending videos that match their interests. Target user of this system could be either viewers who seek entertainment, or creators who wants to grow audiences, or companies who serves both and aim to monetize their digital content.

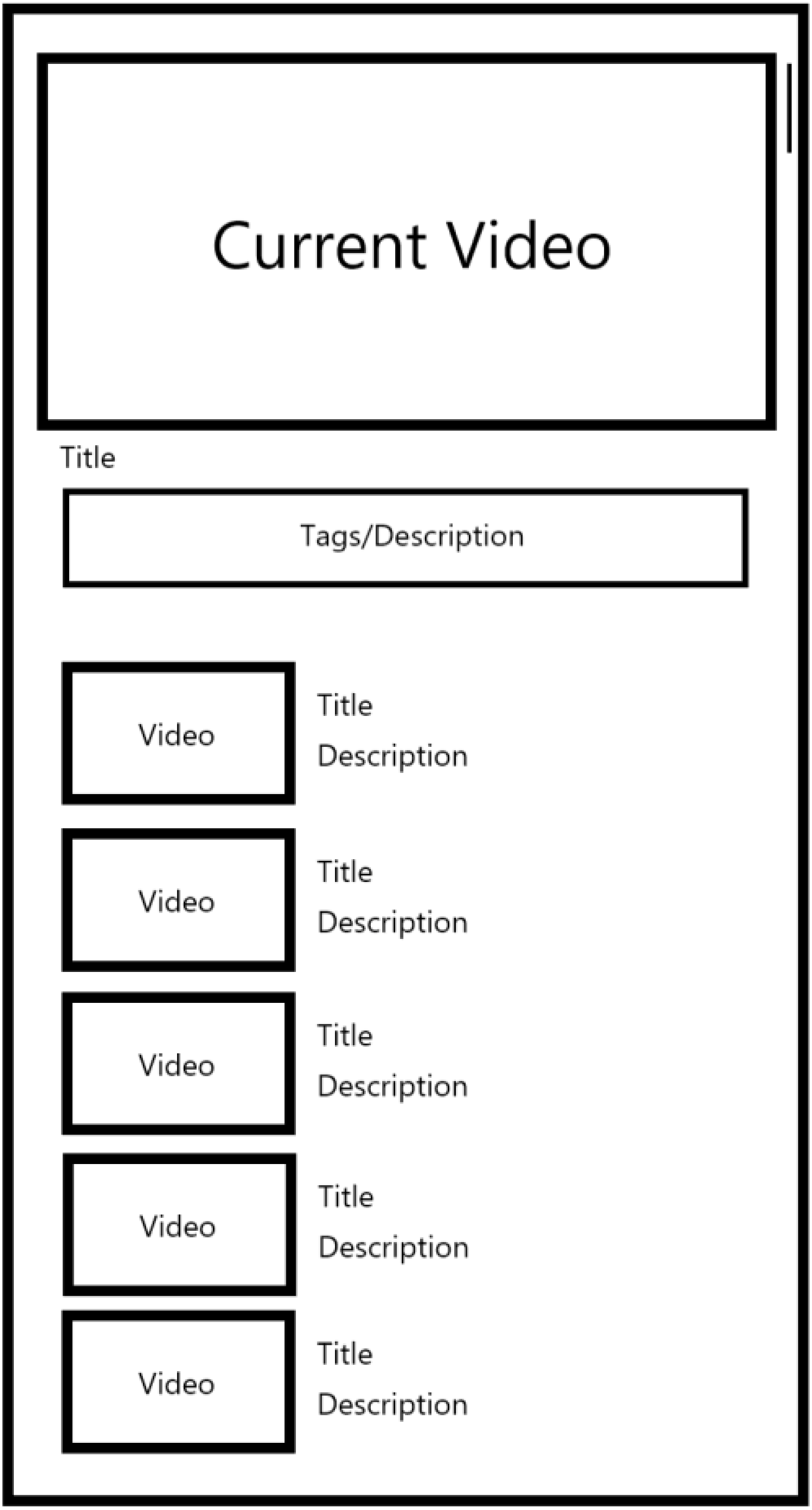
Possible user interface as below:

Current user interface only applies to mobile device.



Search Bar
Video
Video
Title/Description
Title/Description
Video
Video
Title/Description
Title/Description
Video
Video
Title/Description
Title/Description

This is the home page. Items presented to user who stays/browse at this page should be around 6-8 depending on their screen size. Numbers of recommended videos should be 12-16. User can click on video or title/description on this page, which will jump to next interface below, or scroll up or down on this page, system will keep generate new recommendations. User can also search on the search bar on the top of this page.


Current Video
Title
Tags/Description
Video
Title
Description
Video
Title
Description
Video
Title
Description
Video
Title
Description
Video
Title
Description

This is the detail/video page. User can view full video, read description, follow creator on this page. Also, they can see the 'official' tag listed below the video, and click on those to start a search for this tag. Items presented to user who stays/browse at this page should be around 4-5 depending on their screen size. Numbers of recommended videos should be 12-16. System will give a new group of recommendation (same size) if viewer browse through all recommendation without clicking one.

The system should be able to update user profile in real-time, every time a user watches, likes, comments, shares, finishes a broadcast, and scrolls for a certain duration, it will be transmitted back in real time to instantly rearrange the subsequent candidates. For cold start, the system will collect onboarding preferences tags directly from new user. If the user skip this process, then present multiple videos from different classification, and gradually build user profile on daily or hourly incremental data. Details described in Methods section.

Revenue can come from ads, stream tipping, commercial, or subscriptions which can have access to more videos or skip adds.

Dataset

We chose MicroLens – A Content-Driven Micro-Video Recommendation Dataset at Scale [arXiv:2309.15379] as main dataset to train and test the system. It's a large-scale open source short video recommendation dataset maintained by Westlake University since 2023.

It is currently the largest and most complete short video recommendation dataset with million class videos and billion class interactions, which is sufficient to support recommend system we aim to build. It also have 50k subset version for simple training and implementation.

A simple view of subset version of this dataset:

```
MicroLens-50K/
├── interactions.parquet (marks user actions)
├── users.csv (user profile)
├── videos_meta.csv  (information of video, including tags)
├── likes_and_views.txt
├── multimodal/
│   ├── cover_images/ (jpg format pictures)
│   ├── full_videos/
│   ├── audio_tracks/
│   └── subtitles/
└── extracted_features/ (official extracted features)
```

The main interaction file contains detailed behaviour logs such as user IDs, video IDs, timestamps, watch time, completion flags, likes, shares and comments, allows sequential modelling, labelling by complete ratio and multi-task learning in ranking stage. In actual system implementation, we can classify videos using fields in videos_meta.csv and word conclusion generated from cover_image (OCR if possible, or image-to-word by multimodel LLM).

About potential problem with this dataset, since it's sourced from a single platform, it may not cover private or highly noisy content. For privacy and copyright reasons, data are anonymised and cleaned, so real-world "dirty data" and adversarial behaviours won't appear. System build on such dataset may fails to generalise to production.

Methods

For classification, we can use OCR/CLIP/VideoMAE/LLM to process cover images and frames randomly extracted from video, generate semantic vectors and label/categories (label, or tags added by creator should also taken into account). This can help with cold start with newly uploaded videos without tags, result (tags) generated by classification can be compaired with ones in videos_meta.csv.

For recommendation, we can try:

Content-based recommendation, use the attributes of the item itself (in this case tags from classification) to match the user's profile. It can better deal with the cold start problem, and only need tags to make recommendations, can be effective for scenarios where users have clear interests or their interests change quickly.

Collaborative filtering, recommendations made based on the historical behavior similarities between users and items. We may try implementations include neighborhood-based methods and matrix decomposition. It can effectively find the implicit patterns of user interests and improve recommendation personalization, performs well when there is relatively rich interaction data, exactly what we have in the dataset. However, it faces the cold start problem when facing new users or new items.

Hybrid recommendation, combines multiple recommendation mechanisms, such as content, collaborative filtering, and knowledge-driven. We may try methods including weighted fusion, cascade, and feature splicing input models.

Thses methods can also combine with simple rule-based recommendation, using domain knowledge, expert rules, explicit constraints to make recommendations. Restricting certain content to only be recommended to a specific user group. This can be used as an auxiliary supplement to all above.

The baseline implementation can be: for each user, calculate the content similarity score and collaborative filtering score, combined according to a certain weight, and select TopN recommendations. Regarding the time sequence of user behavior, introduce sequence models such as RNN and Transformer to explore interest evolution, such as short-term interest modeling. Regarding contextual information, introduce contextual features to participate in modeling to improve the relevance and diversity of recommendations.

Evaluation

We can try a variety of classic top-N recommendation indicators and per-user personalized indicators, such as:

Precision@N / Recall@N: measures how many of the top N items in the recommendation list are actually liked (or watched for a certain ratio). Calculate these for each user and then find the average, can measure the fairness and personalization effect of different user groups.

NDCG@N: Ranking of recommended content and give higher rewards to highly relevant items that are ranked higher

MAP and MRR : reflect the quality of the recommendation list as a whole, suitable for application situations where users only pay attention to the first few items or click(like/watch) on first group of recommendation.

Also measure coverage, diversity and coverage of recommended items should be increased to avoid recommendations being concentrated on a few popular items.

If we get the chance to obtain real user feedback based on user interaction, we can measure click through rate and dwell time/ratio, which can reflect the percentage of users clicking on recommended content and the ability of recommended content to attract users to continue browsing.

Also, if system generate recommendation without given user actions, we can mark user actions in interactions.parquet as ground truth and evaluate model by how much recommendation overlaps with actual user action.

The best model is usually the one that performs well in the primary indicator and has no obvious shortcomings in the secondary indicators. We can use multi-objective optimization, or first select the model based on the primary indicator, and then use the secondary indicator for further fine-tuning. If we managed to obtain actual user feedback based on user interaction, completion rate and average viewing time will then set as the primary indicators, and other indicators will be used as constraints or secondary goals.

User study design

Invite 10-20 target users, covering different interests/experience levels, provide an interactive recommendation system mock up that can actually play videos(from dataset), and display multiple recommendation lists. These users will be asked to browse the recommended content according to their own interests, watch, like or dislike videos,and do simulated comments and sharing operations. We will then Record user operation behaviors such as clicks, browsing time, selection tendencies for further fine-tuning the system.
In this study we shall use click-through rate and dwell time as the main indicator. For auxiliary supplement we can conduct questionnaires to subjectively score the relevance, personalization, and richness of the recommended content.

Also, we can measure: the time before the user starts watching the first video, the time before the user watches a video in its entirety for the first time, and the ratio of the time the user spends watching videos to the time they spend browsing the recommended list, to evaluate how user is attracted by our system, to further maximize their retention time.