UNIVERSITEIT VAN AMSTERDAM

DUALRec: A Hybrid Sequential and Language Model Framework for Context-Aware Movie Recommendation

Yitong Li

Msc Behavioural Data Science

WHAT IS RECOMMENDER?

'A recommendation engine, also called a recommender, is an artificial intelligence (AI) system that suggests items to a user. Recommendation systems rely on big data analytics and machine learning (ML) algorithms to find patterns in user behavior data and recommend relevant items based on those patterns.' -- IBM (2024)

DUALREC FRAMEWORK

multi-modal LSTM + LLM Semantic Reasoning = Enhanced Personalized Recommendations

Development Environment & LLM Selection

Coding:

- Google Colab Pro
- A100 GPU
- 40GB GPU RAM
- Python 3

LLM Selction:

- DeepSeek V3-685B
- Mistral 8B
- DeepSeek-R1-Distill-Qwen-1.5B (for finetuning)
- Mistral 7B (for finetuning)

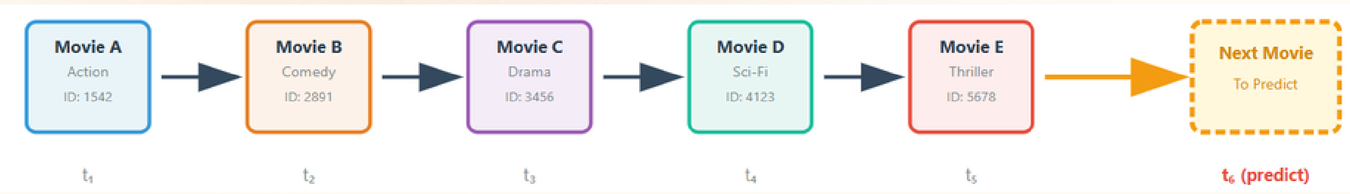
Dataset

- **MovieLens 1M dataset**
- Stable benchmark dataset with 1 million ratings from 6000 users on 4000 movies.
- Each rating entry includes user ID, movie ID, rating value (1-5), movie titles, movie genres, and timestamp.

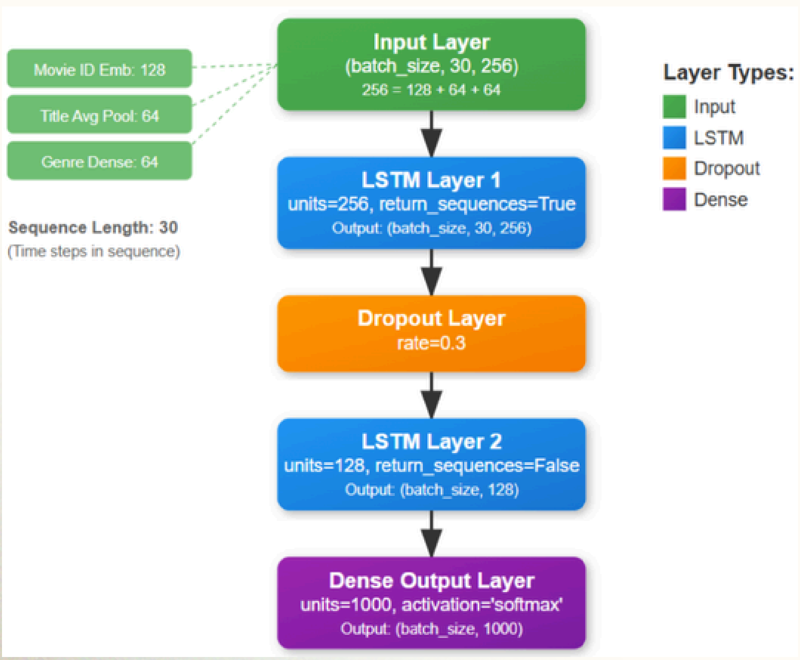
Data manipulation

- Movie filtering: Top 1,000 most frequently watched movies
- Movie ID → 128-dim embeddings
- Tokenized titles (Keras Tokenizer)
- Multi-label genre information- 18 movie genres
- Data split: 70% train / 15% validation / 15% test

Model Stage 1- Sequential Behaviour Modelling



- Group interactions by user and sort chronologically
- Create sequences of 30 consecutive movies with the 31st as target
- Tokenize movie titles (max 10 tokens) → 64-dim embeddings → global average pooling
- Generate 30 sequences per user using sliding window approach
- Genre as binary multi-hot vectors, 18 categories



- **Input:** Sequence of 30 × d (where d = combined feature dimensions)
- **Layer 1:** 256 units, returns sequences (local dependencies)
- **Layer 2:** 128 units, final hidden state, return global representation of user long-term preference.
- **Dropout:** 0.3 for both layers
- **Output:** Softmax over 1K movies → top-1 prediction

Model Stage 2- LLM based recommendation generation

Example LLM Recommendations:

Based on the user's preference for animated, family-friendly films with adventurous and musical elements, here are three recommendations that align with their viewing history:

1. Tarzan (1999) Genres: Animation, Adventure, Children's, Musical
2. The Emperor's New Groove (2000) Genres: Animation, Adventure, Children's, Comedy
3. Lilo & Stitch (2002) Genres: Animation, Children's, Comedy, Science Fiction

- **Free form text generation**
- Structured prompts, no settled list of movies to choose for recommendation

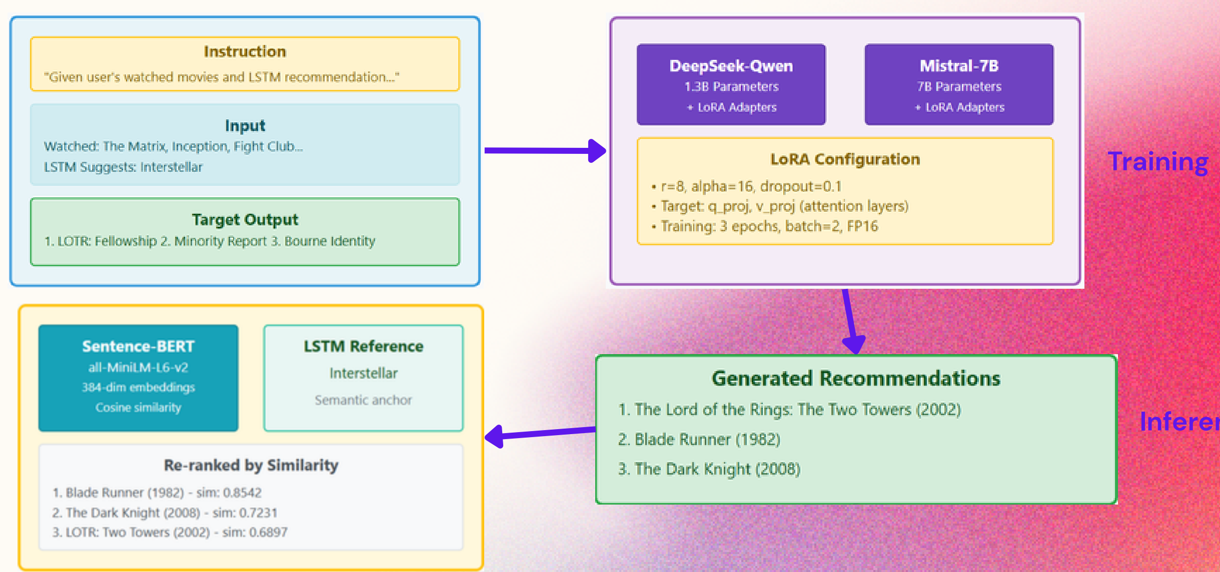
Model Stage 3- LoRA Fine-Tuning and Post-Generation Optimization

Method:

- Low-Rank Adaptation (LoRA) for parameter efficient fine tuning
- Two model variants (DeepSeek-Qwen 1.3B and Mistral-7B)
- **Instruction-following text generation**

Finetuning Data Construction Process:

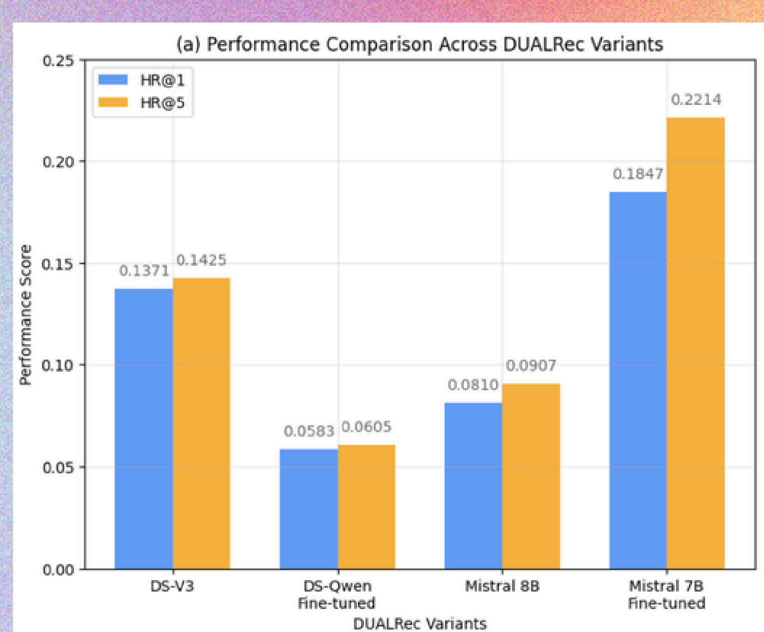
- For each user, take first 5 movie viewing sequence as input
- Use the last 5 movies the user actually watched as the supervised learning window
- Combined with top-1 prediction from Stage 1 LSTM



Result &Discussion

- **Base Model Quality > Fine-tuning:** DeepSeek V3 (pre-trained) outperformed fine-tuned DeepSeek-Qwen
- **Instruction-Following Architecture Advantage:** Mistral-7B achieved highest HR@1 and NDCG@1
- **Parameter Efficiency:** Fine-tuned Mistral-7B > Mistral-8B. Task-specific adaptation more effective than scaling model size
- **Limited Top-K Improvement:** Minimal HR@1 → HR@5 gains across DUALRec variants
- **Traditional models (NARM, STAMP):** Optimize discriminative spread across candidates
- **LLM-based:** Generate semantically cohesive but homogeneous recommendations

Model	HR@1	HR@5	NDCG@1	NDCG@5
MostPop	0.0004	0.0070	0.0004	0.0053
SLIM	0.1270	0.3600	0.1270	0.2530
FPMC	0.1132	0.3748	0.1132	0.2464
NARM	0.1692	0.5230	0.1692	0.3501
STAMP	0.1584	0.5078	0.1584	0.3367
GCE-GNN	0.1312	0.4748	0.1312	0.3044
MCPRN	0.1434	0.4758	0.1434	0.3157
HIDE	0.1498	0.4998	0.1498	0.3256
Atten-Mixer	0.1490	0.4932	0.1490	0.3216
UniSRec	0.0508	0.2508	0.0508	0.1459
NIR	0.0572	0.2326	0.0572	0.1436
DUALRec – DS V3	**0.1371**	**0.1425**	**0.1371**	**0.1406**
DUALRec – Mistral 8B	**0.0810**	**0.0907**	**0.0810**	**0.0871**
DUALRec – DS Qwen Finetuned	**0.0583**	**0.0605**	**0.0583**	**0.0597**
DUALRec – Mistral 7B Finetuned	**0.1847**	**0.2214**	**0.1847**	**0.2078**



Link to Full Paper: https://arxiv.org/abs/2507.13957