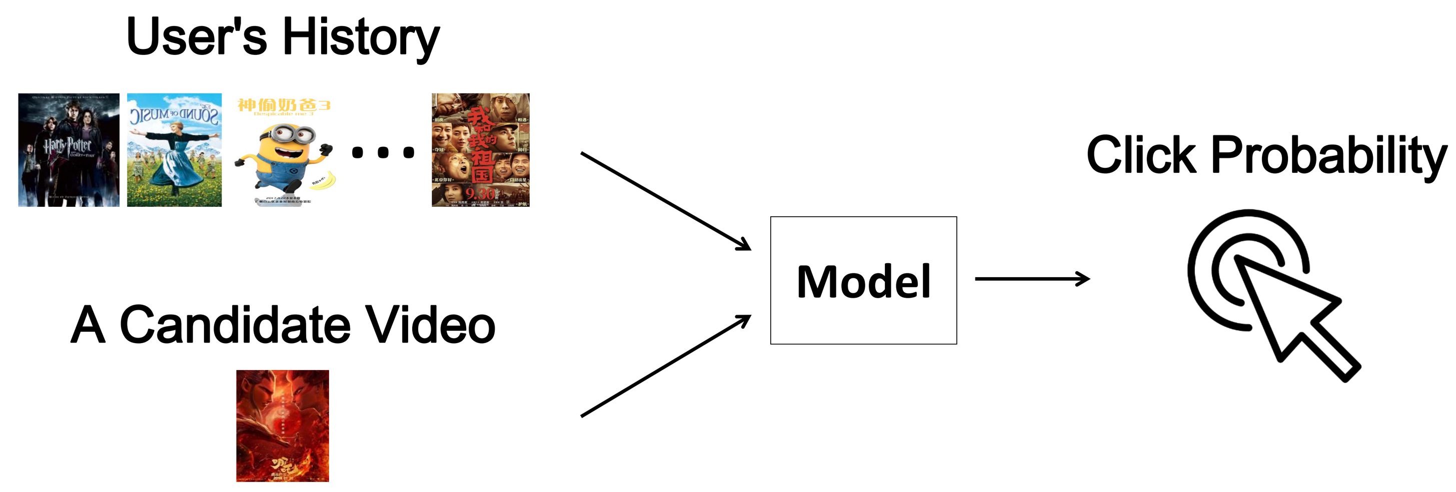
Exploring Content-based Video Relevance for Video Click-Through Rate Prediction

Xun Wang[1], Yali Du[2], Leimin Zhang[1], Xirong Li[3], Miao Zhang[4], and Jianfeng Dong[1*]

[1]Zhejiang Gangshang University [2]University of Technology Sydney [3]Renmin University of China [4]Zhejiang University

Introduction

In the hulu Challenge, given a list of videos that a user has viewed in history, participants are asked to predict whether the user will click a new candidate video, which is a standard **Click-Through Rate (CTR) prediction** problem.



Existing methods for video CTR problem have some **limitations:**

- Do not consider the video content (Cold-start problem)
- Lack explicit modeling of item-wise relevance

Dataset

The HULU challenge has two separated tracks: **TV-series** and **Movies**. Each track provides a dataset, and the dataset is composed of a bunch of **viewer records** and the corresponding **video content**.

viewer records:

0/1	7429	2588 1482 3454 925 7282
label	candidate video	viewer history

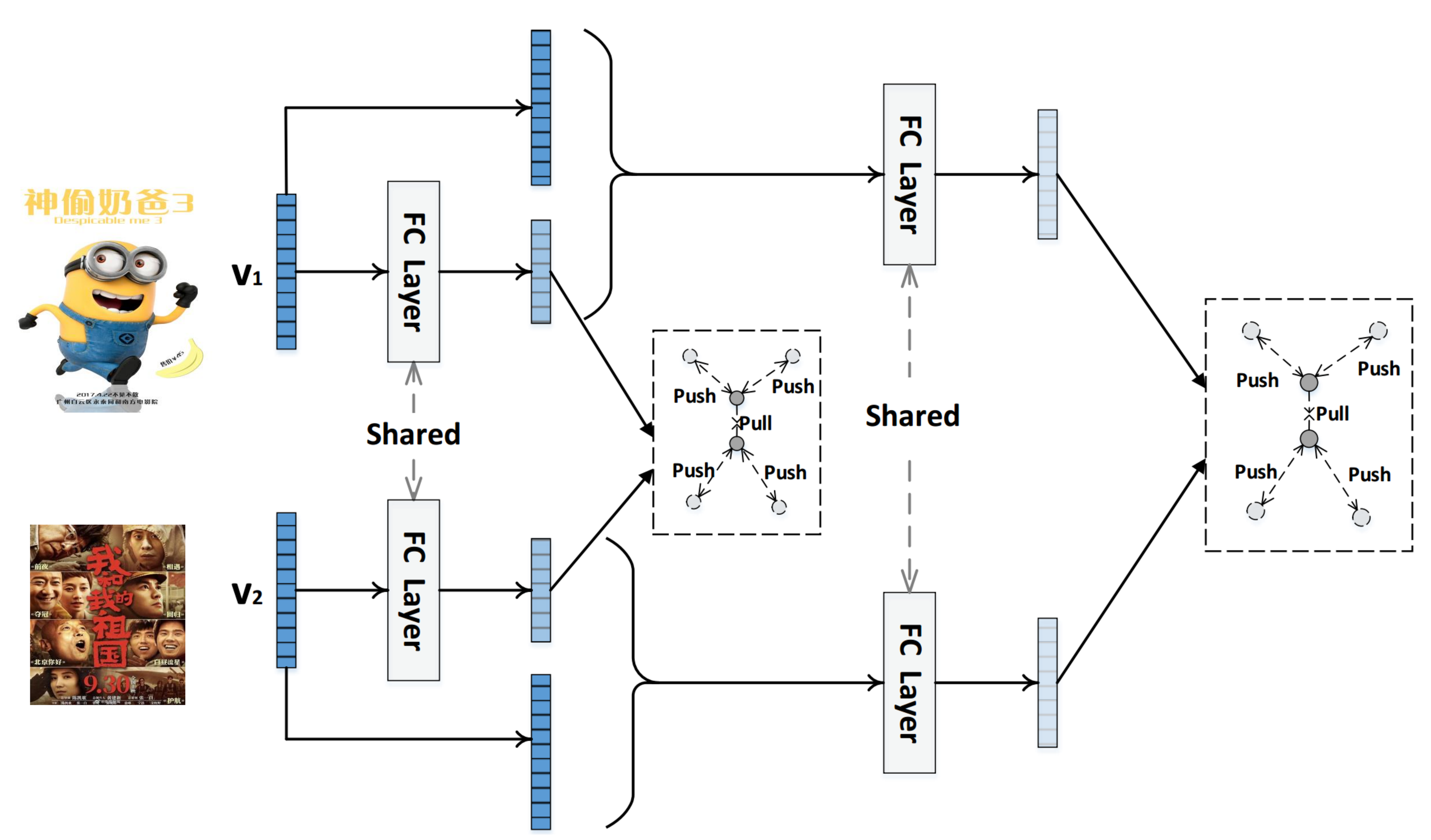
video content:

- visual
  - 2048-dim Inception-v3 feature
  - 512-dim R(2+1)D feature
- audio
  - 128-dim VGGish feature

Our methods

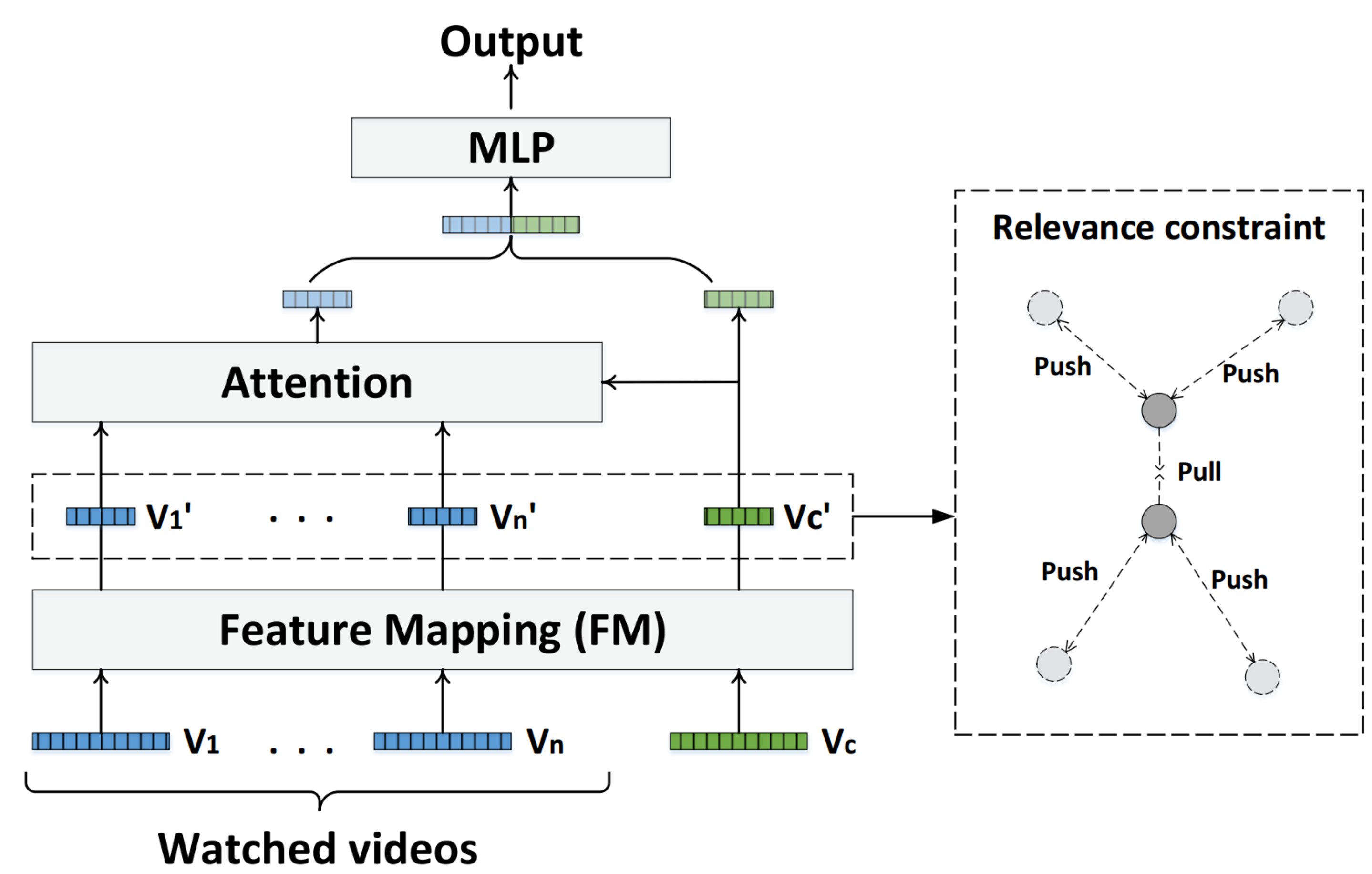
Method 1: Cascading Mapping Network (CMN)

We predict the **click probability score** by measuring the **relevance** between the candidate video and watched videos via CMN.



Method 2: Relevant-Enhanced Deep Interest Network (REDIN)

We improve deep interest network (DIN) by adding **explicit video relevance constraint** for model training.



Experiments

Performance comparison on the validation set. FM: feature mapping; RC: relevance constraint

	TV-series	Movies
random baseline	0.5000	0.5000
Consie similarity on original features	0.4060	0.5393
1-layer MLP	0.6792	0.5831
2-layer MLP	0.6546	0.6006
CMN	**0.6856**	**0.6071**
DIN	0.6160	0.6200
DIN + FM	0.6278	0.6311
DIN + FM + RC (REDIN)	**0.6533**	**0.6428**
DIN_{gru}	0.6175	0.6281
DIN_{gru} + FM	0.6246	0.6327
DIN_{gru} + FM + RC ($REDIN_{gru}$)	0.6316	0.6352
Late fusion	**0.7019**	**0.6528**

Our proposed solution is **more stable**. It gives scores of over 0.6 on both tracks, while the majority of teams only perform well on either track

	TV-series		Movies		sum of ranks
	AUC	rank	AUC	rank	
USTC_I_Know_U	0.6645	2	0.6523	1	3
this work	0.6022	4	0.6155	4	8
UESTC_cfm	0.6656	1	0.5858	7	8
MAGUS	0.5754	6	0.6520	2	8
potato	0.6510	3	0.5930	6	9
GrandRookie	0.5918	5	0.6124	5	10
XRGOGOGO	0.5000	12	0.6475	3	15
Distinc	0.5449	7	0.5732	10	17
Oases	0.5246	9	0.5838	8	17
MVAP	0.5400	8	0.5482	11	19
Dragon	0.5160	11	0.5755	9	20
MIDAS@CBVRP	0.5181	10	0.5337	12	22

Conclusions

- CMN by video relevance prediction is a simple but very effective model, which can be used for strong baseline for video CTR.
- The relevance constraint in REDIN may can be extended to other deep learning based CTR models.
- Late fusion of multiple models achieves the best performance.
- Exploring video relevance is promising for video click-through rate prediction.

dongjf24@gmail.com

WeChat