Crowdsourced Time-Sync Video Recommendation via Semantic-Aware Neural Collaborative Filtering

Zhanpeng Wu

School of Data and Computer Science
Sun Yat-Sen University, Guangzhou, China

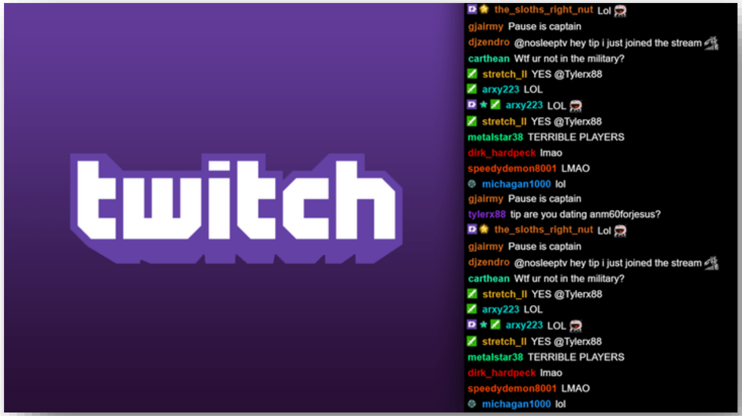
Joint work Yan Zhou, Di Wu, Yipeng Zhou, and Jing Qin

Background

- Time-synchronized Comment (TSC)

YouTube Live Chats

Bilibili



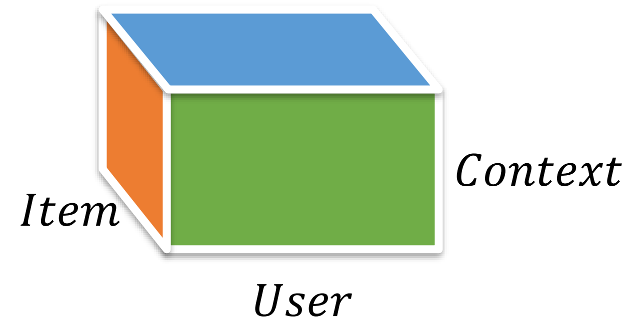
Twitch Chats

Nico Nico Douga

Background

- Traditional vs Context-aware

- Traditional Collaborative Filtering
  - $User \times Item \rightarrow Rating$
- Context-aware Recommendation
  - $User \times Item \times Context \rightarrow Rating$
  - Contextual information can be:
    - Location, Time, User comment, etc.



Background

- A Time-sync Video Example
  - Users with similar interests are more likely to get together and send a TSC

DC

MARVEL

Superhero Films

Available at https://www.bilibili.com/video/av22135056

Background

- TSC data → Sequence data
  - sort TSCs by timestamps
    - $\{TSC_1,\ TSC_2,\ \dots,\ TSC_T\}$
  - each TSC contains two components
    - $TSC \Leftrightarrow <user,\ content>$
  - user sequence
    - $\{user_1,\ user_2,\ \dots,\ user_N\}$

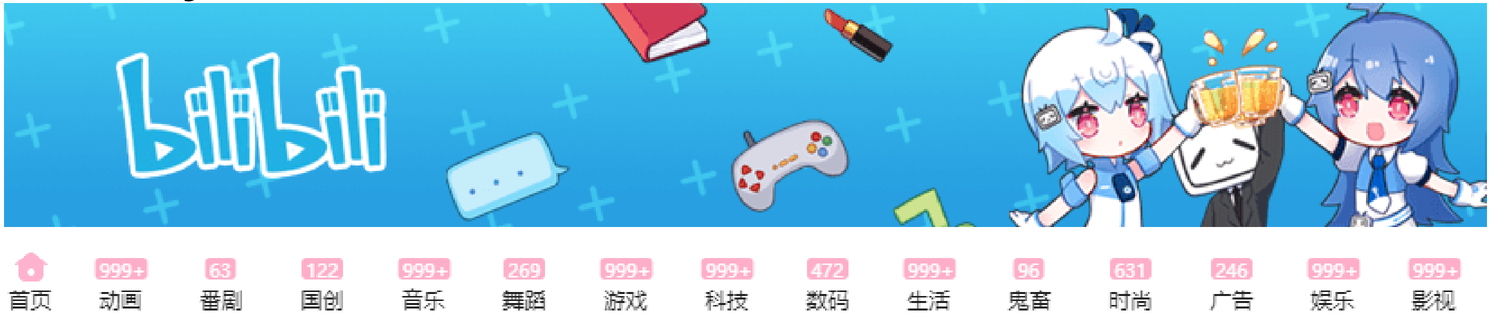
user representation

  - content sequence
    - $\{content_1,\ content_2,\ \dots,\ content_L\}$

video representation

Dataset

- Time-synchronized comment dataset



Category	Avg. TSCs / Video	Category	Avg. TSCs / Video
Anime	662	Guochuang	526
Dance	327	Live	864
Entertainment	597	Movie	688
Fashion	752	Music	677
Game	**1208**	Tech	776
Guichu	610		

Collected from https://www.bilibili.com till 2018/11/15

Dataset

- The fields we mainly used:
  - user_id, video_id, TSC_content, TSC_timestamp

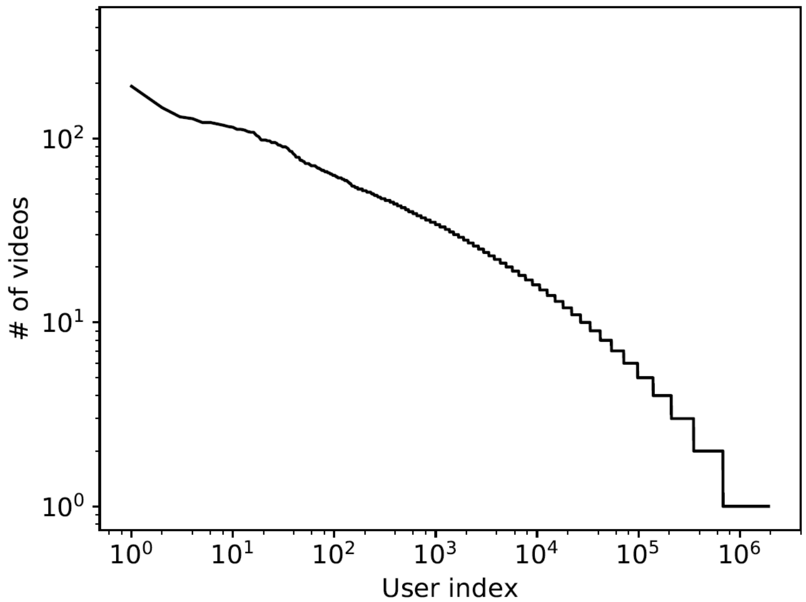
- For gaming category:

# of videos	2,637
# of users	57,294
# of TSCs	836,806
# of user-generated tags	3,483
Avg # of TSCs per user	14.61
Avg # of TSCs per video	317.33
Avg # of user-generated tags per video	7.01
Max/Min # of TSCs for a user	731/5
Max/Min # of TSCs for a video	4393/1
Max/Min # of user-generated tags for a video	14/1
Max/Min # of TSCs for a user leaving in a video	299/1

Collected from gaming category till 2018/12/15

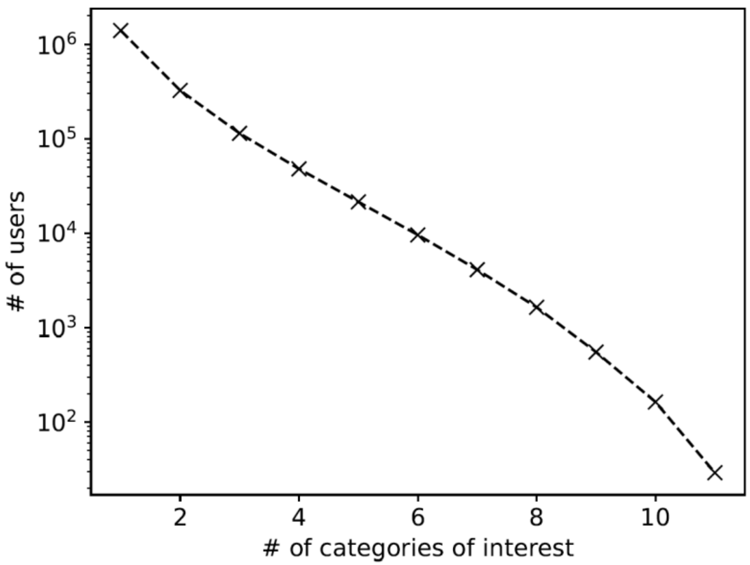
Measurement

- Basic statistics



User's viewing behavior is consistent with Power-Law.

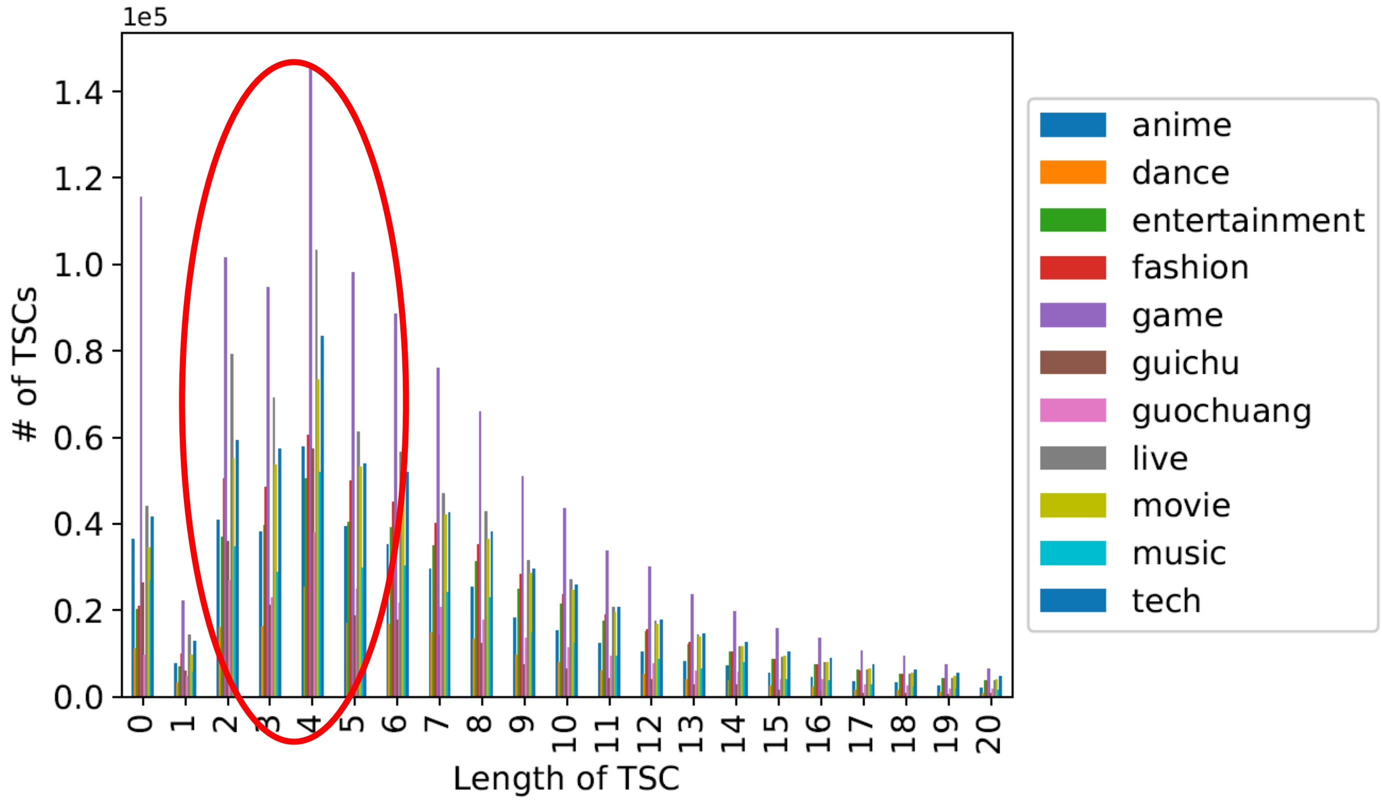
Users tend to focus on only one or two categories.



The number of categories that user focused is exponentially decreasing.

Measurement

- The length of the TSC is mostly distributed between 2 to 6



TSC often appears as a word or phrase.

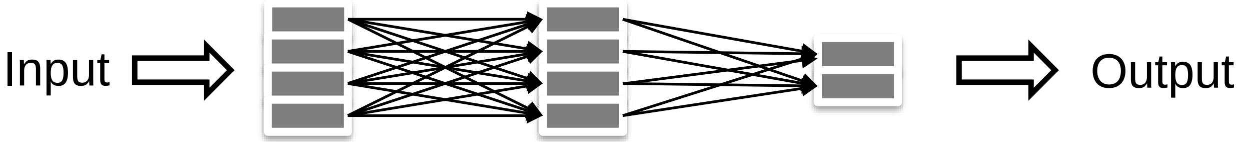
SACF Algorithm

- General Steps of SACF Algorithm
  - Semantic-aware Collaborative Filtering

- Step 1: Extract the user representations

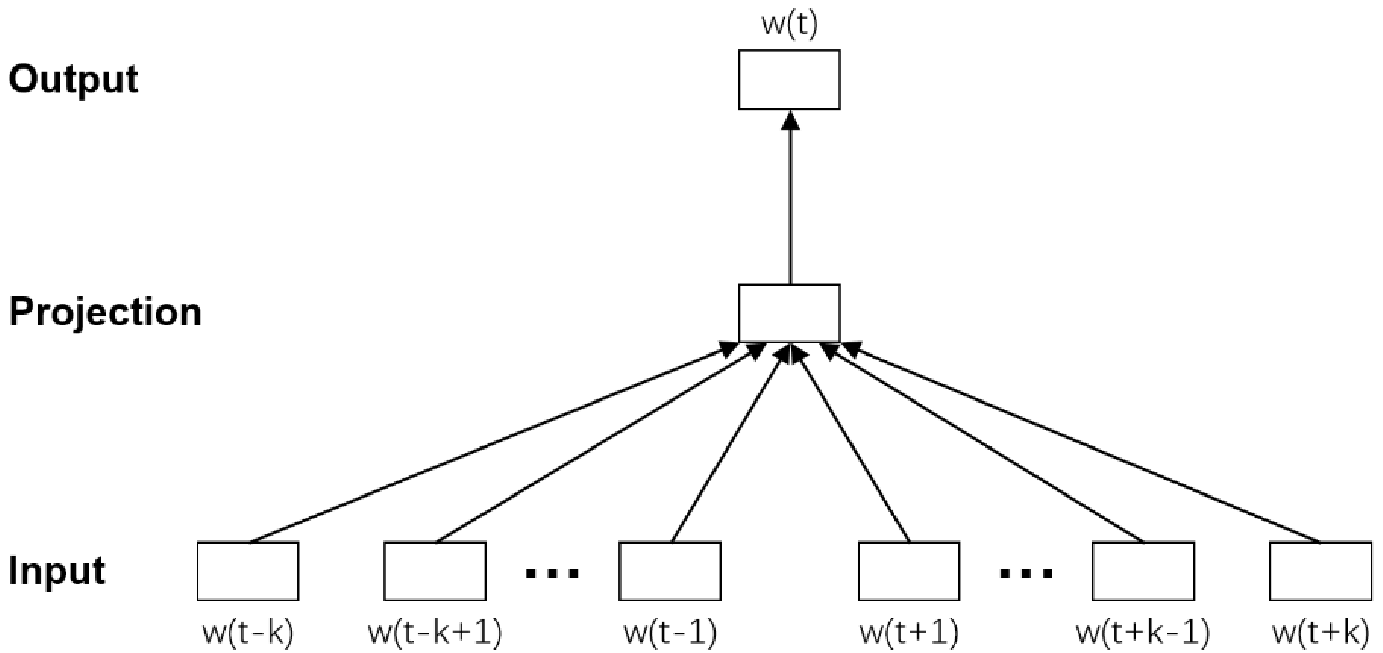
- Step 2: Extract the video representations

- Step 3: Learn the interactive function via DNN



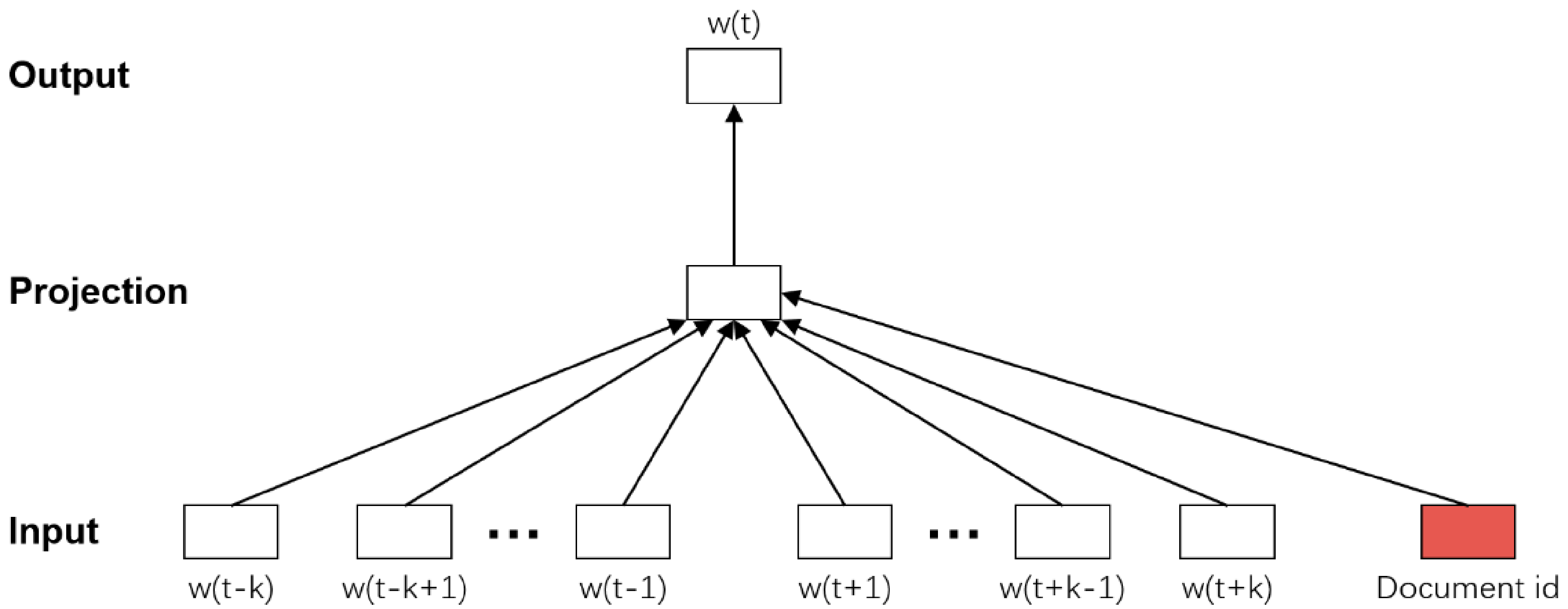
SACF Algorithm

- learn the representations from user sequences
  - Word2vec → User representations
  - word ⇔ user_id



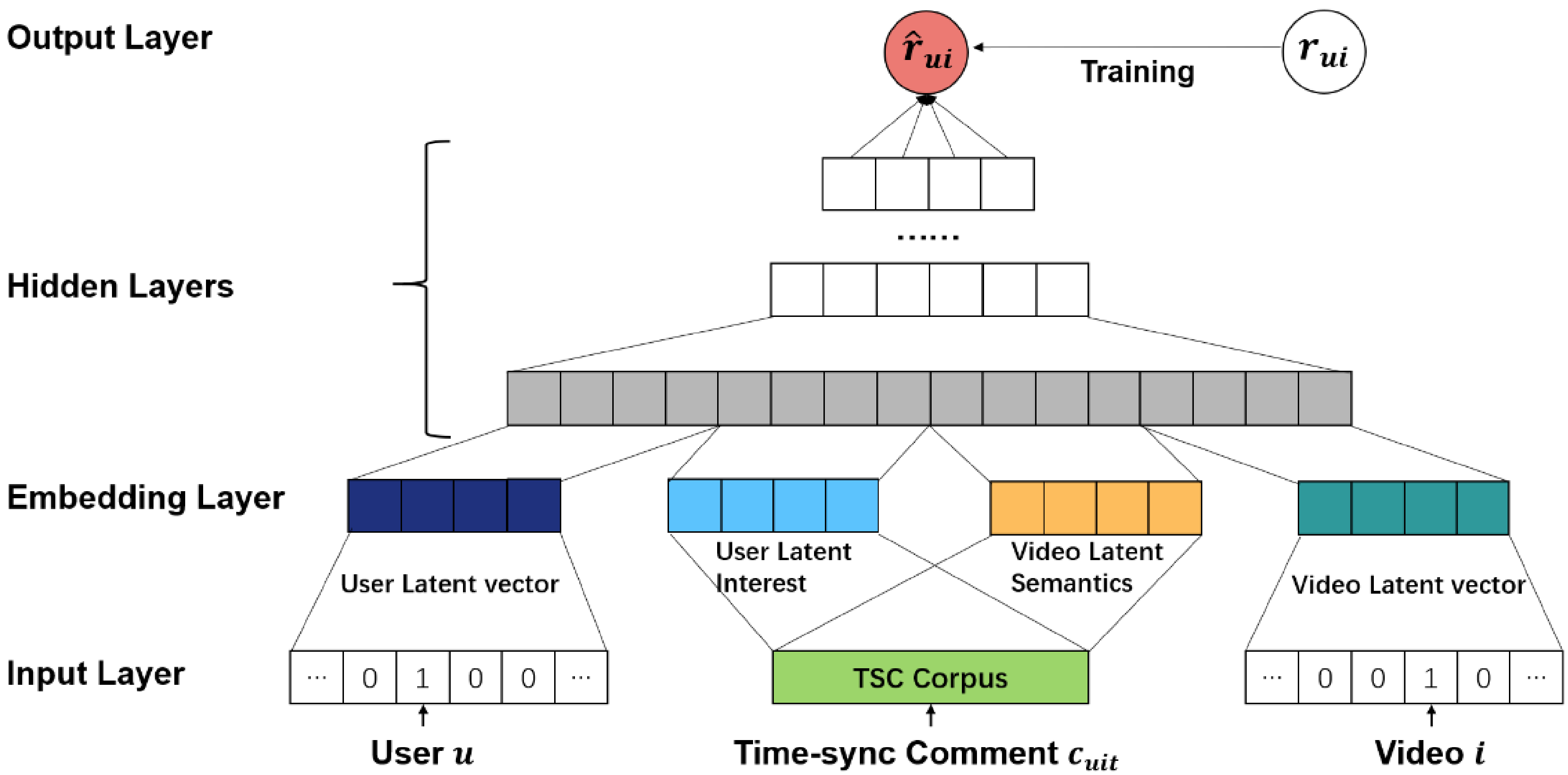
SACF Algorithm

- learn the representations from content sequences
  - Doc2vec → Video representations
  - word ⇔ TSC_id and doc ⇔ video_id



SACF Algorithm

- Semantic-aware Video Recommendation



Evaluation Metrics

- Hit Ratio (HR)
  - focus on Precision

$$\mathrm{HR} = \frac{\sum_{u \in u} |T(u) \cap R(u)|}{\sum_{u \in u} |T(u)|}$$

- Normalized Discounted Cumulative Gain (NDCG)
  - focus on Ranking Quality

$$\mathrm{DCG}_q = \sum_{z=1}^{q} \frac{h(z)}{\log_2(i+1)}$$

$$\mathrm{NDCG}_u = \frac{\mathrm{DCG}_u}{\mathrm{DCG}^*}$$

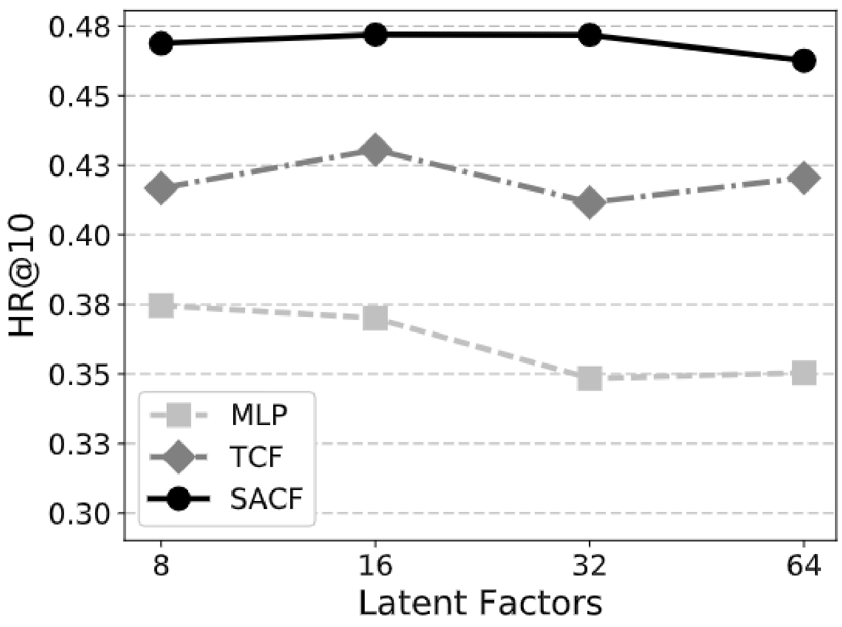
$$\mathrm{NDCG} = \frac{1}{n} \sum_{u \in u} \mathrm{NDCG}_u$$

※ He X, Liao L, Zhang H, et al. Neural collaborative filtering

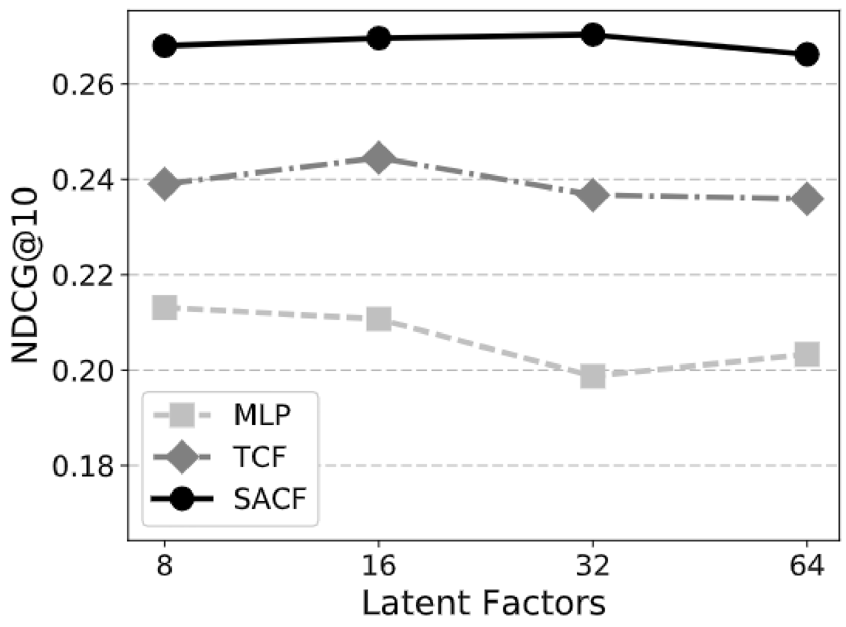
Alg.	Description
MLP	Neural Network Recommendation Model ※
TCF	Tag Collaborative Filtering (with contextual info)

Experiment 1: Performance Comparison

- Performance under different latent factors (LF)
  - The size of the last hidden layer can be considered as the number of latent factors



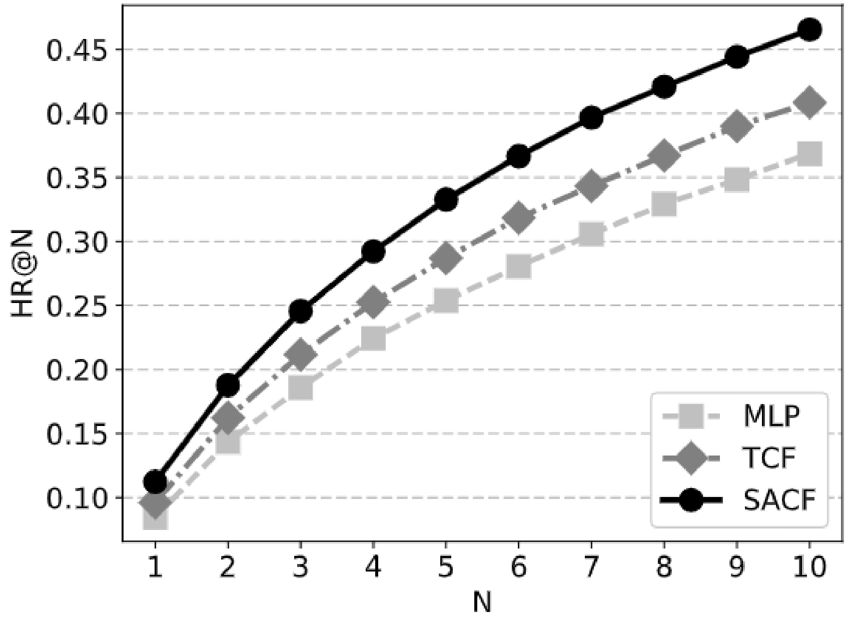
(a) HR@10, ES=64



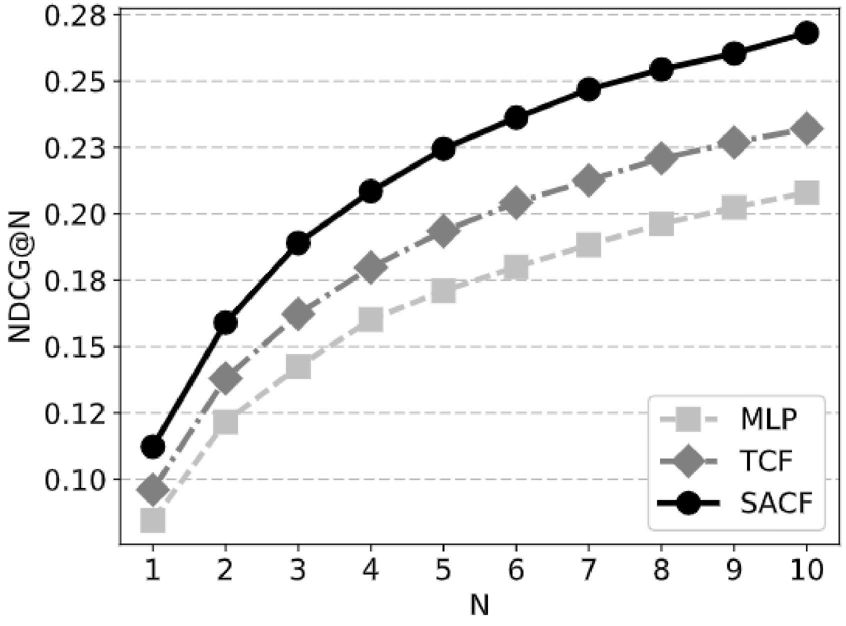
(b) NDCG@10, ES=64

Experiment 1: Performance Comparison

- Top-N Performance Evaluation
  - Recommended list length range from 1 to 10



(a) HR@N, ES=64, LF=8

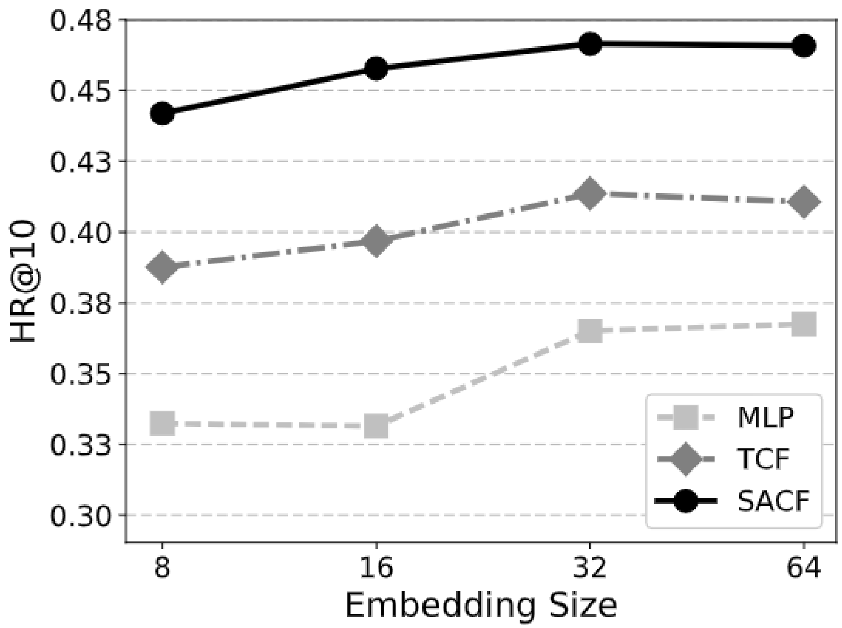


(b) NDCG@N, ES=64, LF=8

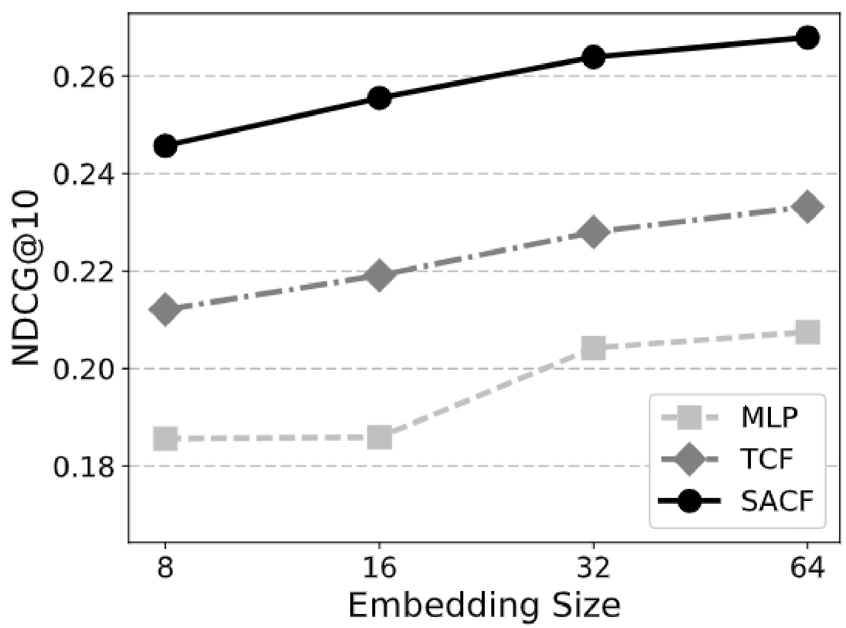
Performance curves follow: MLP < TCF < SACF

Experiment 2: Embedding Size

- The embedding size determines the ability of the feature to describe the system



(a) HR@10, LF=8

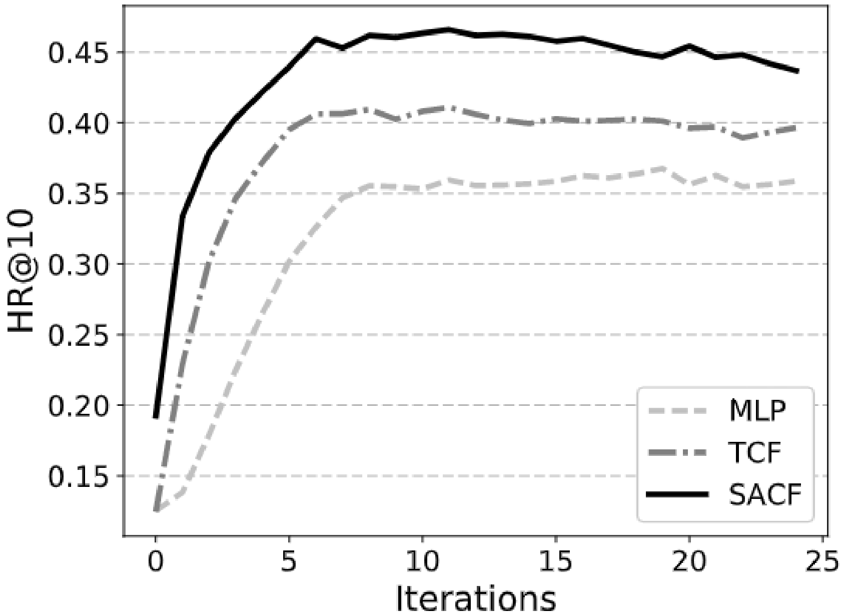


(b) NDCG@10, LF=8

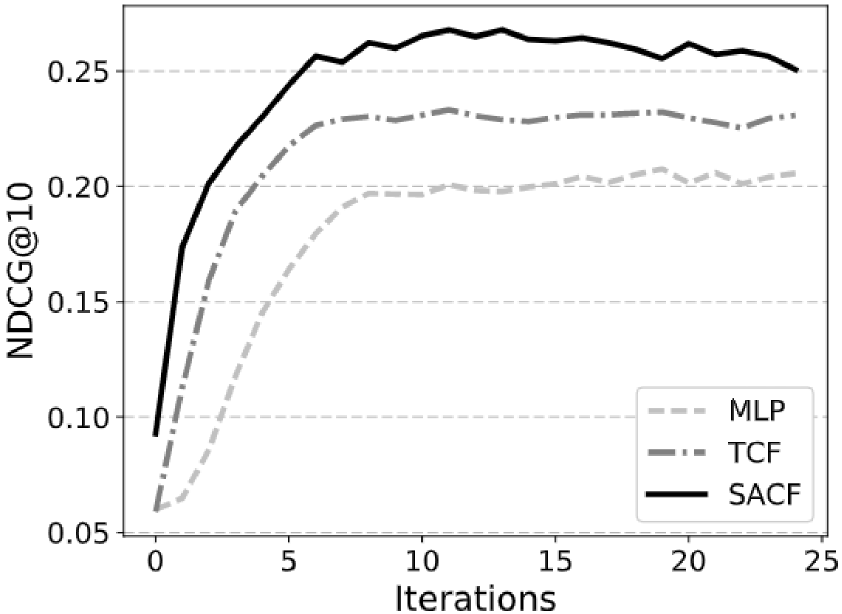
First rise and then stabilize

Experiment 3: Training Iterations

- Under-fitting → Over-fitting



(a) HR@10, ES=64, LF=8



(b) NDCG@10, ES=64, LF=8

Excessive iterations impair final performance

Future Work

- Real-time Recommender System
  - Real-time scenes are increasing
  - user experience will be better in real-time

- Crowdsourced Heterogeneous Information Fusion
  - The source of contextual information is more abundant
  - User profile can be constructed more accurately

- …

Q&A

THANKS

Crowdsourced Time-Sync Video Recommendation via Semantic-Aware Neural Collaborative Filtering

Zhanpeng Wu

Sun Yat-sen University, Guangzhou, China

wuzhp5@mail2.sysu.edu.cn