Social Attentive Network for Live Stream Recommendation

Dung-Ru Yu
National Chiao Tung Univ., Taiwan
dry2@nctu.edu.tw

Chiao-Chuan Chu
Mediatek, Taiwan
Chiao-Chuan.Chu@mediatek.com

Hsu-Chao Lai
National Chiao Tung Univ., Taiwan
hsuchao.cs05g@nctu.edu.tw

Jiun-Long Huang
National Chiao Tung Univ., Taiwan
jlhuang@cs.nctu.edu.tw

ABSTRACT

Live streaming platforms not only provide live videos but also allow social interactions between viewers via real-time chatting. However, none of existing research has studied the social impact for recommending live streams. In this work, we formulate a new personalized recommendation problem by factoring in both video and social contents (chats). Accordingly, we 1) design a new attention network *ANSWER* to identify viewers' attention on video and social contents, and 2) rank the channels based on the attentive features. We collect a real dataset from Twitch for evaluation. The experimental results manifest that ANSWER outperforms baselines by at least 26.6% in terms of NDCG@5.

ACM Reference Format:
Dung-Ru Yu, Chiao-Chuan Chu, Hsu-Chao Lai, and Jiun-Long Huang. 2020. Social Attentive Network for Live Stream Recommendation. In *Companion Proceedings of the Web Conference 2020 (WWW '20 Companion), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3366424.3382679

1 INTRODUCTION

Live streaming platforms, such as Twitch, Facebook Gaming, YouTube Live, and Microsoft Mixer, have emerged as one of the most popular social services recently. For Twitch, the total watched time has reached 434 billion minutes[1] and the daily active viewers have grown up to 10 million in 2018 [6].

In addition to watching live videos, the great success of live streaming platforms may come from the live and social experiences. Most of the platforms support *chat rooms*, which allow people to interact with each other by sending real-time comments (as shown in Figure 1). Facebook reports that friends interact 10 times more on live videos than on traditional videos.[2] Indeed, social interactions bring viewers joyful experiences in live streaming. Therefore, if live stream recommenders only consider video contents, viewers may like the recommended videos but fail to enjoy interacting in chat rooms.

In this work, let $U = \{u_1, u_2, ..., u_N\}$ denote a viewer set and $V = \{v_1, v_2, ..., v_M\}$ denote a broadcasted live video set. A live video $v_m \in V$ consists of sequences of video frames and comments, whose embeddings are denoted as $\{\mathbf{f}_m^1, \cdots, \mathbf{f}_m^t, \cdots\}$ and $\{\mathbf{c}_m^1, \cdots, \mathbf{c}_m^t, \cdots\}$, respectively, where t denotes the timestamp. Since the story of movies or live videos changes frequently, they are usually cut into small segments in recommendation tasks to avoid enormous data processing [2]. Thus, we follow [2] to segment each live video v_m as $S_m = \{S_m^1, \cdots, S_m^p, \cdots, S_m^P\}$ with a time window d. The goal is to predict if a viewer will interact with the next segment of a live video (i.e., as viewer satisfaction) by considering both video and social contents (comments) in previous segments.

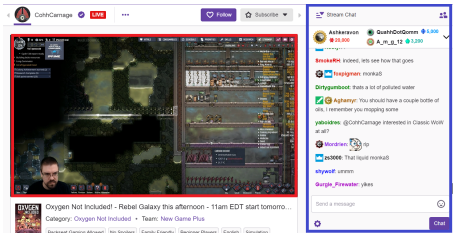

Figure 1: The blue rectangle illustrates a chat room.

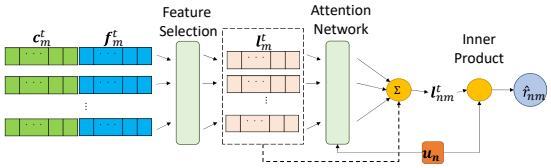

Figure 2: ANSWER illustration.

Related Work. The main differences of live stream recommendations are: 1) real-time contents and 2) social interactions in chat rooms. Note that live videos generate *real-time contents*. However, existing online *video recommenders* [2] analyze the contents of the whole video, which is not accessible in live streaming scenario. Plus, they do not consider the impact of *social interactions in chat rooms*. An alternative is to use *social-aware recommenders* [5], which leverages social links and interactions to learn social influence for personal satisfaction. However, they do not learn from video contents and hence not applicable to live streaming recommendations.

2 METHOD

We propose Attention Network for Social-aware Live Streaming Recommender System (ANSWER), as illustrated in Figure 2.

[1] https://bit.ly/2mBYCcq
[2] https://bit.ly/1V9oxkl

WWW '20 Companion, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7024-0/20/04.
https://doi.org/10.1145/3366424.3382679

Note that viewers may have different attention on different frames. For example, a viewer may like gaming scenes while another may like viewing streamers for interactions. Therefore, we design a novel attention network to learn viewer attention from video frame embeddings $\mathbf{f}$ and comment embeddings $\mathbf{c}$. The two embeddings of a live video v_m with the same timestamp t are first concatenated, denoted as $\mathbf{l}_m^t = \mathbf{f}_m^t \oplus \mathbf{c}_m^t$. To avoid the curse of high dimensionality, we adopt a feature selection layer as:

$$\mathbf{l'}_m^t = \phi(\mathbf{W}_1 \mathbf{l}_m^t), \tag{1}$$

where $\phi(\cdot) = max(0, \cdot)$ is the activation function, $\mathbf{W}_1$ is the weight matrix converting features into hidden layers, and $\mathbf{l'}_m^t$ is the low-dimensional latent feature vectors. We further design an attention network to learn viewer attention as:

$$\begin{aligned} a(n,m,t) &= \mathbf{w}^\intercal \phi(\mathbf{W}_u \mathbf{u}_n + \mathbf{W}_{l'} \mathbf{l'}_m^t + \mathbf{b}) \\ \alpha(n,m,t) &= \frac{\exp(a(n,m,t))}{\sum_{t'=0}^{d} \exp(a(n,m,t'))}, \end{aligned} \tag{2}$$

where $\mathbf{W}_u$ and $\mathbf{W}_{l'}$ are weight matrices of the attention network, and $\mathbf{b}$ is the bias vector. $\mathbf{w}$ projects the activated vectors into a score $a(n,m,t)$, denoting the personalized attention to frame f_m^t of viewer u_n. The scores are further normalized with the frames in the same segment by softmax function as Eq. 2. We derive the attentive live video embeddings $\mathbf{l'}_{nm}^t$ as:

$$\mathbf{l'}_{nm}^t = \sum_{t'=0}^{d} \alpha(n,m,t') \cdot \mathbf{l'}_m^{t'}, \tag{3}$$

which is the aggregation of embeddings of each frame in the same segment weighted by personalized attention. Therefore, those frames attracting a viewer the most will be highlighted by the attention.

Since viewers seldom leave their satisfaction during watching live streams, inspired by BPR [4], we rank the live videos by comparing pairs of distinct live videos. We first collect a training database $DB = \{(n,i,j) | u_n \in U, v_i \in I_n^+, v_j \in I_n^-\}$, where I_n^+ and I_n^- are the segment sets that viewer u_n has and has not left comments, respectively. Note that viewers leave comments when 1) they join discussions with other viewers (social reasons), or 2) they are entertained by the video contents and interact with the streamer (video reasons). The ranking objective is formulated as:

$$\min \sum_{(n,i,j) \in DB} -ln(\sigma(\hat{r}_{ni} - \hat{r}_{nj})) + \lambda_\theta \|\theta\|^2, \tag{4}$$

where $\hat{r}_{ni}$ is the satisfaction score of viewer u_n to live video v_i, $ln(\cdot)$ is the log-likelihood function, $\sigma(\cdot)$ is the sigmoid function mapping $\hat{r}_{ni}$ to a value between 0 and 1, θ is the model parameter set, and λ_θ controls the weight of the sparsity regularization term. The goal is to minimize the total ranking error, which increases when $\hat{r}_{ni} - \hat{r}_{nj} < 0$. Equipped with the attentive live video embeddings $\mathbf{l'}_{nm}^t$, the viewer satisfaction is defined as:

$$\hat{r}_{ni} = \mathbf{u}_n^\intercal \mathbf{l'}_{ni}^t = \mathbf{u}_n^\intercal \sum_{t'=0} \alpha(n,i,t') \cdot \mathbf{l'}_i^t, \tag{5}$$

which is the inner product of viewer embedding and $\mathbf{l'}_{nm}^t$. Stochastic gradient descent is applied to optimize the objective.

Table 1: Top-K performance (NDCG@K).

	$K = 5$	$K = 10$	$K = 15$	$K = 15$
Random	0.13	0.21	0.28	0.30
Popular	0.35	0.40	0.43	0.75
Visited	0.40	0.45	0.49	0.52
ANSWER-V	0.39	0.43	0.48	0.50
ANSWER-C	0.03	0.18	0.28	0.31
ANSWER	0.55	0.57	0.58	0.59

3 EXPERIMENTAL RESULTS

Due to the lack of open dataset, we crawl data from a popular team STEAM·GIFTS on Twitch from 2018/10/01 to 2018/12/31. It includes 8,192 viewers, 31 channels, 220 live videos, 475K comments, and over 54K watching minutes. The widely-used leave-one-out evaluation protocol for top-K recommendation is adopted [4]. We use top-K Normalized Discounted Cumulative Gain (NDCG@K) to evaluate the ranking results. The embeddings of videos and social contents are pre-trained by [1] and [3], respectively.

We compare with the following baselines to justify the effectiveness. `Random` recommends video to each viewer randomly. `Popular` recommends videos to viewers based on the popularity, which is the total number of interactions. `Visited` recommends the video of the latest channel visited by each viewer. `ANSWER` is our proposed method, while `ANSWER-V` and `ANSWER-C` considers only video and social contents (chats), respectively.

Table 1 compares top-K performance of baselines. `ANSWER` outperforms other baselines by at least 26.6% in terms of NDCG@5, which manifests that, instead of popular or visited channels, viewers tend to interact with streamers providing interesting video contents or interested viewer communities. By comparing with `ANSWER-V` and `ANSWER-C`, the performance degrades at least 18.3% of the performance for all K. That is, both of the contents are significant for learning viewer satisfaction in live streaming platforms.

4 CONCLUSION

In this work, we make the first attempt to study the social interactions in live streaming platforms for recommendations. A new recommendation problem is formulated, and we design ANSWER that identifies personalized attentions among videos and chats to make fine-grained recommendations. The experimental results on the Twitch dataset crawled by ourselves manifest that ANSWER outperforms other baselines by at least 26.6% in terms of NDCG@5.

ACKNOWLEDGMENTS

This work is supported in part by MOST in Taiwan via grants 106-2221-E-009-152-MY3, 108-2218-E-009-049 and 108-2627-H-009-001.

REFERENCES

[1] S. Abu-El-Haija et al. 2016. Youtube-8m: a large-scale video classification benchmark. *arXiv* (2016).
[2] Y. Deldjoo et al. 2016. Using visual features and latent factors for movie recommendation. In *ACM RecSys*.
[3] Q. V. Le and T. Mikolov. 2014. Distributed Representations of Sentences and Documents. In *IEEE ICML*. 1188–1196.
[4] S. Rendle et al. 2009. Bpr: bayesian personalized ranking from implicit feedback. In *UAI*.
[5] P. Sun et al. 2018. Attentive recurrent social recommendation. In *ACM SIGIR*.
[6] D. Y. Wohn et al. 2018. Explaining viewers' emotional, instrumental, and financial support provision for live streamers. In *ACM CHI*.