Will You Dance To The Challenge?
Predicting User Participation of TikTok Challenges

Lynnette Hui Xian Ng[1], John Yeh Han Tan[2], Darryl Jing Heng Tan[3], Roy Ka-Wei Lee[3]

[1]Carnegie Mellon University, [2]National University of Singapore, [3]Singapore University of Technology and Design

Email: lynnetteng@cmu.edu, johntyh@u.nus.edu, {darryl_tan, roy_lee}@sutd.edu.sg

Abstract—TikTok is a popular new social media, where users express themselves through short video clips. A common form of interaction on the platform is participating in "challenges", which are songs and dances for users to iterate upon. Challenge contagion can be measured through replication reach, i.e., users uploading videos of their participation in the challenges. The uniqueness of the TikTok platform where both challenge content and user preferences are evolving requires the combination of challenge and user representation. This paper investigates social contagion of TikTok challenges through predicting a user's participation. We propose a novel deep learning model, deepChallenger, to learn and combine latent user and challenge representations from past videos to perform this user-challenge prediction task. We collect a dataset of over 7,000 videos from 12 trending challenges on the ForYouPage, the app's landing page, and over 10,000 videos from 1303 users. Extensive experiments are conducted and the results show that our proposed deepChallenger (F1=0.494) outperforms baselines (F1=0.188) in the prediction task.

Index Terms—TikTok, social media mining, social contagion

I. INTRODUCTION

TikTok is a new video-based social media platform created by the Chinese company ByteDance. It is increasingly population among the younger demographic, with $\tilde{6}0\%$ of its users below the age of 34 [1] in July 2020. Its content features short video clips of under a minute. Upon opening the application, the app lands on the ForYouPage, which aggregates top content in the user's geographical region with recommended content based on the user's interaction (likes, shares) [2]. TikTok Challenges are video formats that users endlessly iterate upon, usually involving an easy to replicate song and dance. Popular challenges in a user's region appear in the ForYouPage to encourage users to join in with their fellow TikTokers.

Research Objectives. In this work, we predict the contagion of 12 TikTok Challenges that trend on the ForYouPage during the COVID19 lockdown. This work draws ideas from collaborative filtering, which exploit the binary similarities between users and items to recommend content to users [3], [4]. Research efforts in this area utilize video content

for recommendation on platforms like YouTube [5], [6] and enhance them with user feedback such as click-through rate and ratings [7], [8]. In this study, we extend the problem from predicting the viewership of a video to predicting user participation in a challenge. Challenge video content are not static as each user injects his own flavour into the challenge; performing a challenge is thus a combination of user's interests and the challenge topics.

We combine a user's previously uploaded videos and videos uploaded on the challenge to perform to predict whether a user will catch on the challenge contagion and participate in a challenge. With a total of 10,424 TikTok videos from 1303 users, we construct a deepChallenger model, a novel deep learning framework, which combines learned embeddings of challenge and user latent representations to perform this user-challenge participation predictive task. Our results show that the deepChallenger model outperforms (F1=0.494) the baseline models (F1=0.188).

II. BACKGROUND

As of August 2020, TikTok has about 100 million monthly active users [9]. Its content composes of either 15s or 60s short video clips. With the merger of music app Musical.ly [10], TikTok became more sound centric, allowing users to pick background sounds to match to their videos. Users create content by uploading a video to the application, along with adding video filters, background music and captions. Captions usually include popular hashtags that helps the video to get noticed through hashtag searches. Although TikTok has a following-follower structure, content is mainly consumed through the aggregated ForYouPage, rather than content solely produced by those a user is following.

Since the inception of the platform, studies have examined the causes of popularity of this new social media platform [11] and motivations of the young TikTok user demographic [12]. Other work examined communication of partisan views through a platform-unique feature called TikTok duets [13].

TikTok challenges are another platform-unique feature, designed with a central dance or action to be replicated by other users and are branded with a challenge hashtag. They gain traction by inspiring their audiences to participate in them. The most popular challenges propagate widely through the app, reaching audiences across geographical boundaries. The #washyourhands challenge is a simple dance promoting washing of hands, aiming to reduce the spread of COVID19;

#attentionchallenge shows off gymnastics moves done at home due to the unavailability of community gyms during lockdown.

In the examining content contagion, previous work predicted popularity of video posts shared on social media platforms [14], [15] and hashtags given to Twitter posts [16], [17]. However, studies that examine user-video content are few, and are mostly survey-based studies [18]. Spread of ideas in TikTok differ from commonly studied social media platforms like Twitter. In the absence of platform-defined distribution processes like retweet dynamics as a measure of social contagion [19], [20], we measure social contagion through user participation rate. As users participate in a challenge, the higher chance it appears in another user's ForYouPage, which creates a cycle of participation. This study fills the research gap by investigating and predicting user's participation in TikTok challenges.

III. DATASET

We collected challenges and users videos data from TikTok using the Python API wrapper by davidteather[1]. This wrapper requires web credentials from a valid TikTok account. While conducting this study, some ethical issues must be taken into consideration. When a user posts a TikTok video, he can select one of the three audience types: everyone, friends only or himself. We only collected videos publicly available for everyone. In order to maintain user privacy, no user identity data was collected nor used in the analysis; a unique user identifier was used to correlate user videos.

A. Challenge data collection

We collected videos from 12 trending TikTok challenges that were initiated from 19 October to 19 December 2020, reflecting health campaigns and life surrounding the coronavirus pandemic. We created a TikTok account from Singapore and collected the top challenges from the ForYouPage. Challenges on the ForYouPage are recommended to users around the region for its engagement metrics, one of which is the number of users that watch the video till its end [21]. The challenges trending on the account's ForYouPage has two key categories: (1) challenges that aim to promote awareness on good hygiene habits to reduce the spread of the virus (eg #washyourhands); and (2) challenges that provide entertainment through dance and song during the quarantine period that governments enforced to control the pandemic (eg #BoredInTheHouse). These videos have a view count of 0.5 to over 2 million. We collected the following challenges: *#attentionchallenge, #blindinglightschallenge, #100wayschallenge, #BoredInTheHouse, #karenchallenge, #Cannibalchallenge, #renegadechallenge, #washyourhandschallenge, #ghencovychallenge, #godaddygo, #papertowelchallenge.*

We first searched for videos related to the challenge using the challenge hashtag, then downloaded the video by its ID. We collected around 1000 videos from each challenge, except when the crawler did not return the desired number of videos,

[1] https://github.com/davidteather/TikTok-Api

which we posit could be due to the crawler limitations and TikTok restrictions on the search query duration.

B. User data collection

Due to data collection constraints, we were unable to collect all the 7075 users present in the dataset. We thus collected the most representative users. We first analyzed the number of challenges each user performed in our Challenge dataset and collected users that performed at least 3 of our 12 challenges. We first searched for videos the user uploaded, then downloaded the video mp4 file. We excluded videos that contained the hashtag of the challenges we are focusing on. We ran this collection mechanism for a month, and collected TikToks from 1303 users, with 8 videos each. In total, we collected 10,424 videos from the selected users.

IV. PROPOSED MODEL

In this section, we elaborate on our proposed deepChallenger framework. The intuition of our proposed framework is to learn the latent representations of the users and challenges using two proxy tasks, and effectively combine the learned embeddings to predict user's involvement in TikTok challenges. In the subsequent sections, we describe the two main components of our framework: (a) challenge representation learning, and (b) user representation learning. Finally, we present how the learned representations are used to perform the user-challenge participation classification task.

A. Challenge Representations Learning

To learn the latent representation of challenges, we first formulate a proxy task: "*Given a TikTok video, predict which challenge it belongs to.*" Figure 1(a) illustrates our proposed approach to perform this proxy task, in which we encode an input challenge video v. Specifically, the video encoding models two aspects of v: the visual sequential image frames and the textual caption.

For the visual sequential image frames, we first breakdown the video into 50 image frames and input each image frame into a pre-trained VGG16 [22] model to extract the frame's embedding e_f^i. Next, we concatenate all the image frames' embeddings to form a visual embedding of the video, i.e., $x_i^c = (e_{f1}^i \oplus e_{f2}^i \cdots \oplus e_{f50}^i)$. Each image frame is pre-processed by resizing to (224,224) and normalized.

To model the video's textual caption, we first tokenize the caption using default tokenizer of BERT language model [23] and extract the pre-trained BERT embeddings: $E_B = \{e_1^b, e_2^b, \ldots, e_n^b\}$, $e_i^b \in R^{768}$. A pre-trained BERT model recieves these embeddings as an input, and outputs the [CLS] token caption embedding as x_t^c.

Finally, we represent a video embedding as the concatenation of its visual embedding and caption embedding, i.e., $x_v^c = (x_i^c \oplus x_t^c)$. The video embedding x_v^c is then fed into a feed forward layer to predict the video's challenge class label.

After training the proposed model in Figure 1(a) with the proxy task, we learn a set of video embeddings that are predictive of the challenge videos. The underlying intuition is
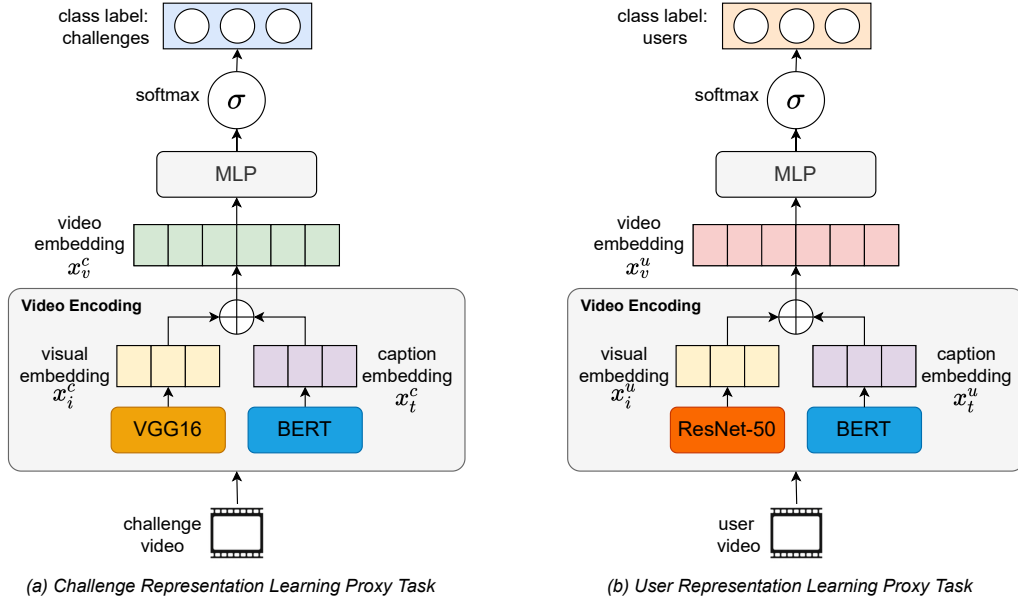
Fig. 1: Proxy tasks for (a) challenge and (b) user representation learning

that learned video embeddings contain latent attributes that are descriptive of the challenges. Therefore, for each challenge, we denote the challenge's latent presentation as:

$$x^c = (x^c_{v1} \oplus x^c_{v2} \cdots \oplus x^c_{vn}) \tag{1}$$

where x^c is a concatenation of the learned video embeddings of the challenge-related videos, and n is the number of videos that belong to the challenge. The challenge representation will be used in the user-challenge participation classification task described in Section IV-C.

B. User Representations Learning

We adopt a similar approach to learn the user's latent presentation. We formulate a proxy task: "*Given a video, predict which user it belongs to.*" Figure 1(b) illustrates our proposed approach to perform the proxy task, where we first encode an input user video v by modeling its visual sequential image frames and textual caption.

Differing from the video encoding used in challenge representation learning, we adopt a pre-trained ResNet-50 [24] to model the image frames and extracted the image frame's embedding e^i_f. ResNet-50 is preferred over VGG16 because through empirical experiments, ResNet-50 yields better results for this proxy task. Next, we concatenate all the image frames' embeddings to form a visual embedding of the video, i.e., $x^u_i = (e^i_{f1} \oplus e^i_{f2} \cdots \oplus e^i_{f50})$.

The video's textual caption is modeled using the same approach as described in earlier section. We use the output from the BERT's [CLS] token caption embedding as x^u_t.

The user's video embeddings are represented as concatenation of it's visual embedding and caption embedding, i.e., $x^u_v = (x^u_i \oplus x^u_t)$. The video embedding x^u_v is input into a feed forward layer to predict the user class label of the video.

Using the same intuition for the challenge latent representation learning, the video embeddings learned with the user representation learning proxy task would contain latent attributes that are descriptive of the users. Therefore, for each user, we denote the user's latent presentation as:

$$x^u = (x^u_{v1} \oplus x^u_{v2} \cdots \oplus x^u_{vm}) \tag{2}$$

where x^u is a concatenation of the learned video embeddings of the videos posted by the users, and m is the number of videos from the users. The user representation will be used in the user-challenge participation classification task described in Section IV-C.

C. User-Challenge Participation

Finally, we perform the binary classification task: "*Given a user and a challenge, predict whether the user will participate in the challenge.*" The challenge embedding x^c and user embedding x^v are fed into a feed-forward layer for user-challenge participation prediction. Our model's intuition is that the challenge and user embeddings would capture latent features that are predictive of the user's preferences towards challenge participation. We note that user-posted videos that belongs to the challenge video dataset are excluded when learning the user representation. This prevents bias from using overlapping training data to model both challenge and user representations.

D. Implementation Details

We used the same embedding size for pre-trained BERT embeddings and the same number of video frames $f = 50$ for video embeddings. Additionally, we added a dropout layer with 20% dropout for the fully connected layer. We use the ADAM optimizer [25] with a learning rate of 0.001 to train

TABLE I: Performance on challenge and user representation learning proxy tasks

Model	Macro-Prec	Macro-Rec	Macro-F1
Challenge Representation Learning			
VGG16	0.248	0.186	0.153
VGG16 + BERT	**0.660**	**0.563**	**0.494**
ResNet-50	0.257	0.174	0.125
ResNet-50 + BERT	0.583	0.486	0.385
User Representation Learning			
VGG16	0.163	0.558	0.163
VGG16 + BERT	0.186	0.749	0.188
ResNet-50	0.194	0.725	0.195
ResNet-50 + BERT	**0.197**	**0.733**	**0.197**

TABLE II: Experimental results for user-challenge participation prediction task

Model	Macro-Prec	Macro-Rec	Macro-F1
VGG16	0.017	0.200	0.083
VGG16 + BERT	0.188	0.750	0.188
ResNet-50	0.050	0.059	0.050
ResNet-50 + BERT	0.187	0.750	0.188
deepChallenger	**0.494**	**0.933**	**0.494**

our models. We used a batch size 4 to optimise for the large model size, with 30 epochs of training optimised for early stopping and a three-fold cross-validation.

V. EXPERIMENTS

A. Experimental Settings

Baselines. As we are proposing a novel prediction task, there are no known baseline that predicts user-challenge participation. Therefore we designed the following baselines that utilized features of user's posted videos to predict the user's challenge participation:

- *VGG16*: We utilize the visual embeddings of user's videos to train a deep neural network (DNN) classifier to predict if the user would would participate in a given challenge. The visual embeddings are extracted using a pre-trained *VGG16* [22].
- *VGG16+BERT*: *VGG16* baseline enhanced with caption embeddings learned using pre-trained BERT [23].
- *ResNet-50*: We adopt the similar approach used in *VGG16* baseline but adopted the pre-trained *ResNet-50* [24] to extract the visual embeddings of the user's videos.
- *ResNet-50+BERT*: *ResNet-50* baseline enhanced with caption embeddings learned using pre-trained BERT.

Training and Testing Set. In our experiments, we adopt an 80-20 split, where for each experimental run, 80% of the videos are used for training with the remaining 20% used for testing. This setting also applies to the challenge and user latent representation learning proxy tasks.

Evaluation Metrics. We use the macro averaging precision (Macro-F1), recall (Macro-Rec) and F1 score (Macro-F1) as the evaluation metrics. Three-fold cross-validation is used in our experiments and the average results are reported.

B. Experimental Results

In Table I, we report the performance on the proxy tasks. We observed that models that utilized textual caption information yield better performance on both proxy tasks. Specifically, we observe that the VGG16+BERT model performs better for the challenge representation learning task, while ResNet-50+BERT achieves better performance for the user representation learning task. Motivated by these empirical results, we

adopt the superior model configurations to learn the challenge and user representations in our deepChallenger framework.

Table II shows the experimental results for the user-challenge participation prediction task. We observe that deepChallenger outperforms the best baseline (F1=0.494 vs F1=0.188). We note that the baseline models with caption embeddings (i.e., models with BERT) outperforms the baseline models that utilized only visual embeddings (i.e., VGG16 and ResNet-50). Nevertheless, the *VGG16+BERT* and *ResNet-50+BERT* baselines are observed to achieve similar performance. The main difference between deepChallenger and the baselines is that deepChallenger learned and utilized the challenge embeddings for the final prediction task. Thus, the superior performance of deepChallenger underscores the importance of incorporating challenge representations in predicting user challenge adoption.

VI. DISCUSSION AND CONCLUSION

In this study, we study social contagion of TikTok challenges by predicting whether users will participate in a challenge, given their past video uploads and the challenge videos. Both factors are continually evolving across time. We used the top 1000 trending challenge videos to characterize challenges as these videos are most likely to appear on many user's ForYouPage. We use the latest 8 videos of each user to characterize a user's interests.

Our proposed deepChallenger model performs 2.5 times better (F1=0.494) than the user-challenge participation baselines, which only uses user past data to predict whether a user will participate in the challenge. We posit that the low F1 score for the baseline task (F1=0.188) could be due to the data sparsity: there is an overwhelming number of users compared to the number of videos each user is represented by.

The user representation learning task characterizes what type of video a user will upload, while the challenge representation learning proxy task characterizes the types of videos that are common of each challenge. In investigating whether a user will adopt a certain social challenge, it is not sufficient to look only at a user's past videos (F1=0.188), but we must look at the nature of the challenge videos. A user is most likely to participate in a challenge if the challenge nature suits his interests. The increase in model performance upon adding challenge representation learning shows that challenge actions must match user's interests (gleaned from past videos) for a user to catch the challenge virus.

As with any study, there are several limitations of this work. Since TikTok does not have a native API for data

collection, we used a wrapper available on Github, which is highly dependent on the website structure. As a fast-developing platform, TikTok's webpage structure constantly changes, during which data collection is impacted and we had to patch the API before resuming collection.

Our study on TikTok social contagion is part of a broader topic of user engagement on social media. TikTok presents a new medium for viral content adaptation, namely through video for content spread. For future work, we hope to characterize the propagation and replication reach of ForYouPage challenges, dances and songs on this wildly popular platform.

REFERENCES

[1] H. Tankovska, "U.s. tiktok audiences by age and gender 2020," Jan 2021. [Online]. Available: https://www.statista.com/statistics/1166153/usa-tiktok-user-distribution-age-gender/

[2] TikTok, "How tiktok recommends videos foryou," Aug 2019. [Online]. Available: https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you

[3] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, 2001, pp. 285–295.

[4] J. Lv, B. Song, J. Guo, X. Du, and M. Guizani, "Interest-related item similarity model based on multimodal data for top-n recommendation," *IEEE Access*, vol. 7, pp. 12 809–12 821, 2019.

[5] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston *et al.*, "The youtube video recommendation system," in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 293–296.

[6] S. Baluja, R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, and M. Aly, "Video suggestion and discovery for youtube: taking random walks through the view graph," in *Proceedings of the 17th international conference on World Wide Web*, 2008, pp. 895–904.

[7] T. Mei, B. Yang, X.-S. Hua, and S. Li, "Contextual video recommendation by multimodal relevance and user feedback," *ACM Transactions on Information Systems (TOIS)*, vol. 29, no. 2, pp. 1–24, 2011.

[8] B. Yang, T. Mei, X.-S. Hua, L. Yang, S.-Q. Yang, and M. Li, "Online video recommendation based on multimodal fusion and relevance feedback," in *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, ser. CIVR '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 73–80. [Online]. Available: https://doi.org/10.1145/1282280.1282290

[9] sherman4949, "Tiktok reveals detailed user numbers for the first time," Aug 2020. [Online]. Available: https://www.cnbc.com/2020/08/24/tiktok-reveals-us-global-user-growth-numbers-for-first-time.html

[10] J. Russell, "Short video service musical.ly is merging into sister app tiktok," Aug 2018. [Online]. Available: https://techcrunch.com/2018/08/02/musically-tiktok/

[11] L. Xu, X. Yan, and Z. Zhang, "Research on the causes of the "tik tok" app becoming popular and the existing problems," *Journal of Advanced Management Science*, vol. 7, no. 2, 2019.

[12] J. Ahlse, F. Nilsson, and N. Sandström, "It's time to tiktok: Exploring generation z's motivations to participate in# challenges," 2020.

[13] J. C. Medina Serrano, O. Papakyriakopoulos, and S. Hegelich, "Dancing to the partisan beat: A first analysis of political communication on tiktok," in *12th ACM Conference on Web Science*, ser. WebSci '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 257–266. [Online]. Available: https://doi.org/10.1145/3394231.3397916

[14] P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun, "Who should share what? item-level social influence prediction for users and posts ranking," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 2011, pp. 185–194.

[15] W. Ding, Y. Shang, L. Guo, X. Hu, R. Yan, and T. He, "Video popularity prediction by sentiment propagation via implicit network," in *Proceedings of the 24th ACM international on conference on information and knowledge management*, 2015, pp. 1621–1630.

[16] L. Yang, T. Sun, M. Zhang, and Q. Mei, "We know what@ you# tag: does the dual role affect hashtag adoption?" in *Proceedings of the 21st international conference on World Wide Web*, 2012, pp. 261–270.

[17] N. Ben-Lhachemi *et al.*, "Using tweets embeddings for hashtag recommendation in twitter," *Procedia Computer Science*, vol. 127, pp. 7–15, 2018.

[18] J. G. Myrick, "Emotion regulation, procrastination, and watching cat videos online: Who watches internet cats, why, and to what effect?" *Computers in Human Behavior*, vol. 52, pp. 168–176, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563215004343

[19] R. Kobayashi and R. Lambiotte, "Tideh: Time-dependent hawkes process for predicting retweet dynamics," 2016. [Online]. Available: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13026

[20] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.aap9559

[21] B. Feldman, "Unraveling the mystery of the tiktok 'for you' page," Nov 2019. [Online]. Available: https://nymag.com/intelligencer/2019/11/how-to-get-on-the-tiktok-for-you-page.html

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[23] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.