FitNExT: Leveraging Transformers with Context-Augmented Start Tokens to Generate Recommendations for New Users in Connected Fitness at Peloton

Shoya Yoshida*, O. Nganba Meetei

Peloton Interactive, 441 9th Ave. New York, NY. USA.

Abstract

Several studies have shown that the likelihood of long term adherence to fitness routines increases when an individual develops intrinsic motivation towards fitness. Connected fitness platforms, like Peloton, strive to make workouts enjoyable and convenient to help users nurture this intrinsic motivation. It is particularly important for new members, who haven't formed a habit yet, that their initial experience nudges them towards long term engagement. Personalized recommendation is a useful tool in enabling early habit formation, but it is a difficult task. Different members start at varying fitness levels and progress at different rates, on top of the lack of information due to the cold start problem. To address this challenge, we introduce the Fitness New user Experience Transformer (FitNExT) model. It combines the proven strength of transformer architecture in understanding sequential data with an innovative approach for contextualizing the start of a member's fitness journey which we refer to as Context Augmented Start Token (CAST). To train FitNExT, we focus on the sequence of the first few workouts when a member starts their journey on the Peloton platform, which by definition is short. CAST adds a feature rich token that clearly signifies the start of a member's journey and enables the transformer model to more efficiently learn the starting fitness state and the varying rates of progression from the limited sequential data. We use the outputs of this model to display rows of recommendations on the homescreen to ease new users into their fitness routine. Our offline evaluations show that our model FitNExT significantly outperforms heuristics-based approach and achieves a 17.4% relative increase over a version of the model trained without using CAST. Furthermore, we ran an online A/B test and observed that FitNExT powered recommendations positively influenced early user engagement by improving conversion by 11.8% and homescreen row reuse rate by 30.0%.

Keywords

Recommender Systems, Cold-Start Problem, Sequential Recommendations,

1. Introduction

Peloton is an interactive digital fitness platform accessible via dedicated fitness equipment such as bike and treadmill, as well as through mobile or TV apps, delivering live and on-demand fitness classes for users to work out anytime, anywhere. When it comes to fitness, research

Proceedings of BehavRec '23, the First International Workshop on Behavior Change and Persuasive Recommender Systems at the 17th ACM Conference on Recommender Systems (RecSys '23), Singapore, September 18-22, 2023

*Corresponding author.

shoya.yoshida@onepeloton.com (S. Yoshida); nganba.meetei@onepeloton.com (O. N. Meetei)
0009-0007-2316-287X (S. Yoshida); 0009-0007-7505-0810 (O. N. Meetei)

consistently shows that intrinsic motivation plays a pivotal role in forming and sustaining habits of exercising regularly in the long-term [1]. Making workouts enjoyable, convenient, and tailored to individual needs are key components in nurturing this motivation, especially for new users who are at the cusp of habit formation. Hence, careful personalization of class recommendations on the homescreen is crucial. It can guide new users to quickly discover classes they enjoy to help spark intrinsic motivation to promote long-term habit formation and ensure user satisfaction on the platform.

However, curating the homescreen with class recommendations for new users can be especially difficult. First, Peloton has a vast, growing content library of over 80 thousand workout classes led by dozens of unique instructors with various class types and difficulty levels. Second, we must address the cold-start problem of recommending these classes to new users with very little information. This is a notably tricky problem for connected fitness due to the inherent nature of fitness: users start at different fitness levels and also advance at different speeds. This cannot be overlooked, as influencing new users to take fitness classes of appropriate levels is critical to maintaining motivation. For example, users that take a class that is too difficult for them may become dejected, or users that take a class that is too easy may not find enough value in the platform. For beginners, it is imperative that we steer them through a gradual, healthy progression from beginner-friendly classes to later more advanced classes. For users who are new to the Peloton platform but not new to, for example, cycling, our recommender system needs to adapt to skip recommending beginner cycling classes. Some other users may need to start with beginner classes but may already be a seasoned athlete in a different fitness discipline and ramp up to more difficult classes quickly. In this case, the system would need to quickly detect the high rate that the user is leveling up and recommend harder classes each session. Hence, this problem of generating new-user-friendly recommendations in connected fitness is challenging as it requires the system to attain a deep understanding of all the different patterns of sequences of workouts that new users typically undertake. Then during inference, it must use that knowledge to quickly assess which pattern each particular individual falls under. Overall, unlike in fields like e-commerce where there is typically no clear upwards progression in customer behavior across sessions, connected fitness introduces an additional complexity of non-stationary fitness levels. The recommender system to successfully guide new users through the beginning of their fitness journey to form habits must quickly capture both the level aspect and user preferences from a limited sequence of workouts in a cold-start setting. This also makes using popular recommender models such as DLRM [2] infeasible as it is not able to richly encode the sequence nor adapt quickly in a cold-start setting.

Our work introduces an innovative approach to address this challenge with the Fitness New user Experience Transformer (FitNExT) model. This model takes in as inputs user features and a sequence of workouts, where each workout is represented by the contextually rich metadata about the workout and the taken class. The training data leverages the beginning portions of the workout sequence of users who have become active members on the platform, thereby having the model learn the behaviors of exemplary new beginners to produce guiding recommendations for current new members. We also introduce and include in each of the workout sequences the Context-Augmented Start Token (CAST) for clearly marking the beginning of a user's fitness journey. It is analogous to the start token in Natural Language Processing (NLP) but richer in context. The predictions from the model are prominently displayed on the homescreen as rows of recommendations to nudge users to take the recommended classes to help them start their fitness journeys on auspicious paths. Our approach has shown excellent offline results and improved user engagement in an online A/B test, demonstrating tremendous value in influencing the users to take fitness classes of appropriate level.

2. Related Work

The transformer [3] models found great success in natural language processing tasks [4] and have also been seeing recent success in recommender system applications as well due to the sequential nature of user-item interaction history [5, 6, 7]. Transformers have been applied in click-through rate prediction by Alibaba with the Behavioral Sequence Transformer (BST) [8], as well as in session-based recommender systems, where the system needs to quickly adapt to user interactions [9, 10, 11, 12]. It's also worth noting that in these recent works, each item is typically represented by metadata instead of item IDs like in NLP to avoid item cold-starts.

3. Contributions

We design a recommender system that helps new users start a healthy habit as they embark on their fitness journey at Peloton. We adapt the BST model and apply it to the user cold-start scenarios in connected fitness services, which has been largely unexplored. The distinctiveness of our modeling challenge resides in capturing a user's fitness level and progression speed, as users graduate to more advanced classes. This requirement, which is not found in areas like e-commerce, presents itself within the time-sensitive context of the early habit formation stage in fitness. Our contributions are twofold. First, we design an approach to solve the challenge by training the FitNExT model on a new user dataset. Second, we introduce the Context-Augmented Start Token (CAST) to mark the beginning of a user's fitness journey, which helps the model get a better understanding of the user's start state. Both of these approaches help alleviate the new user cold-start problem. We also discuss how the model outputs are displayed as rows on the homescreen as recommendations to influence new user behavior.

4. Method / Approach

4.1. Training Data

We construct our training data $D = [H_1, H_2, ...H_N]$, where H_i is the beginning portion of the workout history of user i and N is the number of active users on the platform that have completed more than X workouts. Specifically, $H_i = [W_{i,1}, W_{i,2}..., W_{i,P}]$, where $W_{i,j}$ is the j-th workout completed by user i and we simply only include the first P workouts from each user. We use $X = 20$ and $P = 10$ in our experiments, and we employ the heuristic cutoff with X workouts to ingrain best new user behaviors from users that went on to become active into the training data. Each workout $W_{i,j}$ is denoted as a feature vector $W_{i,j} = [S_i; C_{i,j}; R_{k(i,j)}]$, where S_i is static user features that stays constant throughout the sequence such as user language settings, $C_{i,j}$ is contextual features about the workout for user i on their j-th workout such as used device, and

$R_{k(i,j)}$ is the metadata we have about class with index *k* that the user *i* took from our library on the *j*-th workout such as instructor or class type. For brevity, we will drop the explicit dependence of *k* on *i* and *j* from the notations going forward. Each of these feature vectors are represented by *s*, *c*, or *r* number of attributes, respectively; for example $R_k = \{A_{R_{k,1}}, A_{R_{k,2}}, ..., A_{R_{k,r}}\}$ where each $A_{R_{k,a}}$ is some feature attribute. It's also worth mentioning that we do not have our model learn any user ID or class ID embeddings to avoid cold-start problems. Furthermore, while no classes are ever the same on our platform and we continuously release new instructor-led workout classes, classes with similar metadata reoccur. This recurring nature of classes enables us to learn from historical workout data to apply to current new user recommendations.

We also prepend an extra item – the Context-Augmented Start Token (CAST) denoted by T_0 – at the beginning of each training sequence such that $H'_i = [T_0; W_{i,1}, W_{i,2}..., W_{i,P}]$. In Natural Language Processing, the inputs to the model are simply sequences of word tokens, and we set aside a special token to mark the start of a sentence. In Recommender Systems, instead of word tokens, the input can be a sequence of items, where each item is represented by multiple features and not solely by the item ID. Therefore, we observe that we can have a parallel representation of the start token in NLP in recommender systems where the start token in recommender systems is represented by metadata and augmented with contextual information. Thus, the CAST also has all the same features used for each of the classes, and for each categorical variable, it receives a special unique value [SEED] similar to [CLS] token in the BERT paper [4]. The values of the continuous features are set to 0, as we preprocess the data by normalizing them around a mean of 0. Please refer to figure 1 for a visualization of CAST. The point of the CAST is for the model to learn a separate unique representation to explicitly mark the beginning of their fitness journey. This helps the model efficiently learn the starting fitness state by achieving a granular understanding of how each feature relates to new user behaviors. For example, it is possible to learn that the embedding vector for the [SEED] in the duration feature embedding table is more similar to the embedding vector for a relatively easier 20 minute class that is suitable for new users than to that of a harder 60 minute class. The addition of the CAST into the sequence effectively functions as a springboard to let the model swiftly capture how new user sequences start. Overall, this training data lets the model grasp how a typical, successful new user fitness journey begins by learning from all the patterns that the current active users initially took to become active on the platform.

4.2. Training Method

Our model is trained using the next item prediction task with binary cross entropy loss. Namely, for user *i* and target class R_n, the probability of the user converting on the target class can be modeled as $P(R_n | H'_{i,j<n}; \theta)$ where θ denotes the model parameters. Our negative samples were fetched using the popularity-based random sampling strategy [13]. It's worth noting that during training and inference, in order to predict the first workout class that the user will take, the workout sequence is simply one item, which is the CAST. This training method lets the model learn about the first best classes that new beginners typically take on a feature-level by adjusting the embeddings learned for each feature of CAST. This empowers the model to guide new users even with zero workouts at first through a recommended path of classes to ramp up within the Peloton ecosystem.

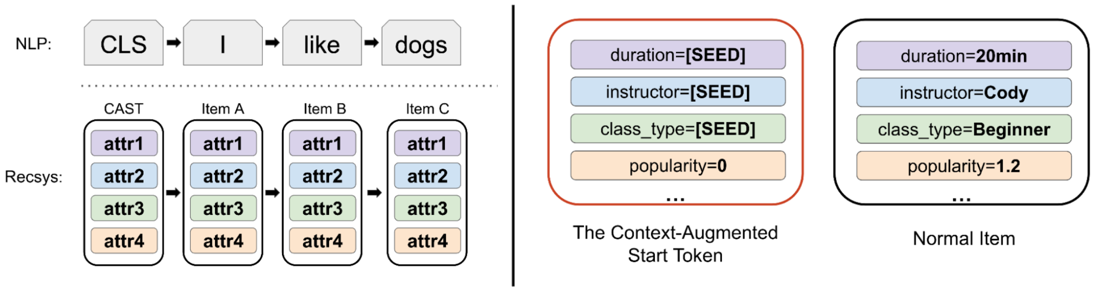

Figure 1: On the left: Visualization of the parallel between the start token in NLP and the proposed CAST in Recommender Systems. On the right: Visualization of what values of the features in CAST look like compared to those in a normal class in the sequence

4.3. FitNExT Model Architecture

Our FitNExT model adapts the Behavioral Sequence Transformer (BST), and it comprises three major components: the user feature encoder, the sequence encoder, and the target class encoder. The user feature encoder is a Multi-Layer Perceptron (MLP) to embed static user features, denoted by MLP_{user}. These features could include demographics or other non-sequential user information. The sequence encoder is a Transformer encoder, which captures the sequence of workouts undertaken by a new user. Each class in the sequence is first embedded using an MLP_{class} and these class representations are fed into the Transformer encoder and pooled, producing the workouts sequence embedding. The target class encoder reuses the same MLP_{class} used to embed each of the classes in the sequence encoder to maintain consistency across class representations. The user and sequence embeddings are concatenated and passed through another $MLP_{combine}$. Finally, we perform a dot product between the output of this MLP and the target class encoding and then pass the result through a sigmoid function to yield the final recommendation score. The visualization of this architecture is shown in figure 2.

During inference, for brand new users where the sequence is simply just CAST, the trained model essentially outputs globally popular classes among new users as that is the best any model can do without further information. After a user takes the first workout, the transformer is able to richly encode the sequence combined with CAST to quickly adapt the next recommendations. Typically, if the user takes a beginner class as the first workout, the model will keep recommending beginner classes for a few more sessions before switching to more intermediate-level classes to help the users ramp up. On the other hand, if the user takes a very challenging 45 minute high-intensity class as the first class, the model swiftly understands that this user is an experienced athlete and shifts the recommendations to show similar challenging classes. Overall, the model leverages the various patterns of beginnings of fitness journeys it saw during training from the active forerunners of the platform to help guide the new members choose their next workout class of appropriate level at their stage.

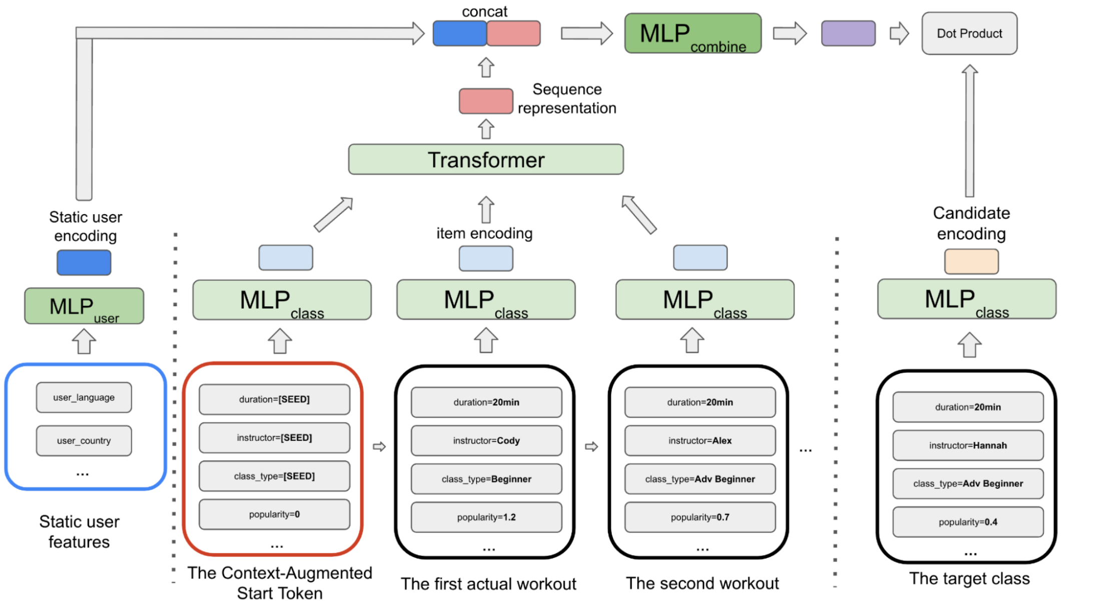

Figure 2: A Visualization of the overall architecture of the FitNExT model

5. Presenting the Recommendations to the Users

The class recommendations on the homescreen at Peloton are presented as rows, where each row is populated by some logical grouping of classes. Normal active members that have completed a certain number of workouts on the platform see myriad personalized rows of recommendations on the homescreen based on their workout history. However, before this work, new members only saw a limited amount of rows that were only powered by rules-based logic. With this work, we bridge the gap between new beginners to reaching the stable homescreen experience by producing meaningful rows of recommendation powered by machine learning to help new users commence their journey optimally. Figures 3 and 4 show two non-cherry picked examples of rows that were generated using the FitNExT model. All rows are given a title, which serves as an explanation for the set of recommendations to set the context. The examples are shown for the bike platform for a brand new user with zero workouts. The content of the rows are refreshed regularly and become more personalized with each session.

The first example in figure 3 is a row titled "Get Started with Cycling Classes," and it essentially shows the model's best recommendations for the new user. This row is displayed at the very top of the homescreen upon logging in to grab attention, and the title is consciously worded so that it is clear that these recommendations are what members should start their journey with. With this row, by definition of how we trained the model, we nudge the new users to take one of the classes that the model has predicted are the best recommendations for the new user to keep being engaged on the platform and form a habit. Effectively, these recommendations are similar to the ones that current active members historically took to start their fitness journey, and it signifies similar classes in the row have delighted and helped spark motivation in the

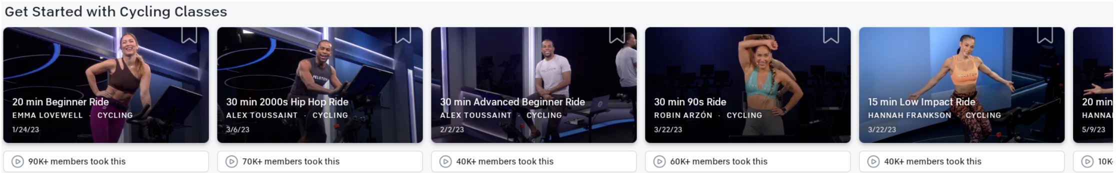

Figure 3: An example of the "Get Started with Cycling Classes" row

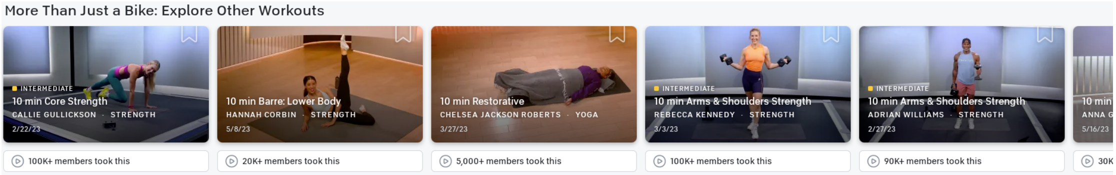

Figure 4: An example of the "More than Just a Bike: Explore Other Workouts" row

past. Since figure 3 is from a user without any workouts, the model generates a great variety of popular choices of first workouts, ranging from beginner-friendly classes such as the "20 min Beginner Ride" and the "15 min Low Impact Ride" to intermediate level classes such as "30 min Advanced Beginner Rides" to slightly harder classes such as the "30 min 2000s Hip Hop Ride." As new users take more classes, the model will learn their fitness level and progression late to continue guiding their journey with more focused class recommendations. The only explicit filter we apply is to only allow each instructor to appear up to twice in this row to help users explore our 50+ instructors, all with unique teaching styles.

The second row "More than Just a Bike: Explore Other Workouts" is aimed at educating the new users to become aware of all the different types of other classes they can take on their new bike other than cycling. To generate this row, the candidate classes excluding cycling classes are scored by the model and we limit each class type to appear only up to twice to allow for a wide variety of classes to be surfaced.

6. Results

6.1. Offline Evaluation

In the offline setting, we compared the MAP@5 of our FitNExT model with CAST against two baselines: popularity-based heuristics and the FitNExT model trained without CAST as an ablation study of the impact of CAST. In order to remove the effect of the CAST, the categorical values of CAST were replaced with the 0s, which is a separate special index used in each embedding table for the non-existent workouts; it is the index used to pad the rest of the sequence not yet utilized during training in next item prediction. The continuous values were also replaced with 0s. We employed a 90/5/5 split between training, validation, and test users, and the MAP@K was calculated by predicting what test users took on the last day of workouts, which was held out for all users during training. Our FitNExT model without CAST had a comfortable 35X improvement over the heuristics, and our FitNExT model with CAST achieved an additional 17.4% relative improvement over the FitNExT model without CAST. These results evidence the value in leveraging previous new users' workout histories and the proposed CAST concept.

7. Conclusion

In this work, we tackled the challenge of designing a recommender system that guides new users through the beginning of their fitness journey to help them form a habit of exercising. We designed a model that is able to quickly learn their fitness level to predict classes of appropriate difficulty and user preferences, and we generated rows of recommendations to show on the homescreen to influence new user behavior. Specifically, we proposed a novel approach of applying the transformer architecture that learns from a dataset of the beginning portion of active, successful members' workout histories so that the model learns about all the different patterns a fitness journey can start and progress. We also introduce the concept of a Context-Augmented Start Token (CAST) to represent the start of a user's fitness journey, which tremendously helps the model efficiently learn about the starting state of each sequence of workouts. Our approaches are validated in offline evaluation by the significant increase in MAP@5 compared to baselines including the model trained without CAST, and our successful online A/B test further solidified the value. The positive results reinforce the value of sequence-based recommendation systems in platforms where the user journeys start differently, evolve over time, and require guidance based on actions of previous new users. Future work will focus on enhancing the model with additional contextual information and exploring multi-task learning to predict classes that users can take across different platforms we offer.

References

[1] P. J. Teixeira, E. V. Carraça, D. Markland, M. N. Silva, R. M. Ryan, Exercise, physical activity, and self-determination theory: A systematic review, International Journal of Behavioral Nutrition and Physical Activity 9 (2012) 78. URL: https://doi.org/10.1186/1479-5868-9-78. doi:10.1186/1479-5868-9-78.

[2] M. Naumov, D. Mudigere, H.-J. M. Shi, J. Huang, N. Sundaraman, J. Park, X. Wang, U. Gupta, C.-J. Wu, A. G. Azzolini, D. Dzhulgakov, A. Mallevich, I. Cherniavskii, Y. Lu, R. Krishnamoorthi, A. Yu, V. Kondratenko, S. Pereira, X. Chen, W. Chen, V. Rao, B. Jia, L. Xiong, M. Smelyanskiy, Deep Learning Recommendation Model for Personalization and Recommendation Systems, 2019. URL: https://arxiv.org/abs/1906.00091v1.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is All you Need, in: Advances in Neural Information Processing Systems, volume 30, Curran Associates, Inc., 2017. URL: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

[4] J. Devlin, M.-W. Chang, K. Lee, K. N. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018. URL: https://arxiv.org/abs/1810.04805.

[5] W.-C. Kang, J. McAuley, Self-Attentive Sequential Recommendation, IEEE Computer

Society, 2018, pp. 197–206. URL: https://www.computer.org/csdl/proceedings-article/icdm/2018/08594844/17D45Xq6dBh. doi:10.1109/ICDM.2018.00035.

[6] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, P. Jiang, BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 1441–1450. URL: https://doi.org/10.1145/3357384.3357895. doi:10.1145/3357384.3357895.

[7] S. Zhang, Y. Tay, L. Yao, A. Sun, Next Item Recommendation with Self-Attention, 2018. URL: https://arxiv.org/abs/1808.06414v2.

[8] Q. Chen, H. Zhao, W. Li, P. Huang, W. Ou, Behavior sequence transformer for e-commerce recommendation in Alibaba, in: Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data, DLP-KDD '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 1–4. URL: https://doi.org/10.1145/3326937.3341261. doi:10.1145/3326937.3341261.

[9] X. Chen, D. Liu, C. Lei, R. Li, Z.-J. Zha, Z. Xiong, BERT4SessRec: Content-Based Video Relevance Prediction with Bidirectional Encoder Representations from Transformer, in: Proceedings of the 27th ACM International Conference on Multimedia, MM '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 2597–2601. URL: https://doi.org/10.1145/3343031.3356051. doi:10.1145/3343031.3356051.

[10] G. de Souza Pereira Moreira, S. Rabhi, R. Ak, B. Schifferer, End-to-End Session-Based Recommendation on GPU, in: Proceedings of the 15th ACM Conference on Recommender Systems, RecSys '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 831–833. URL: https://doi.org/10.1145/3460231.3473322. doi:10.1145/3460231.3473322.

[11] G. de Souza Pereira Moreira, S. Rabhi, J. M. Lee, R. Ak, E. Oldridge, Transformers4Rec: Bridging the Gap between NLP and Sequential / Session-Based Recommendation, in: Proceedings of the 15th ACM Conference on Recommender Systems, RecSys '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 143–153. URL: https://doi.org/10.1145/3460231.3474255. doi:10.1145/3460231.3474255.

[12] G. d. S. P. Moreira, S. Rabhi, R. Ak, M. Y. Kabir, E. Oldridge, Transformers with multi-modal features and post-fusion context for e-commerce session-based recommendation, 2021. URL: https://arxiv.org/abs/2107.05124v1.

[13] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed Representations of Words and Phrases and their Compositionality, in: Advances in Neural Information Processing Systems, volume 26, Curran Associates, Inc., 2013. URL: https://papers.nips.cc/paper_files/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html.