GRU4RecBE: A Hybrid Session-Based Movie Recommendation System
(Student Abstract)

Michael Potter, [1] **Hamlin Liu ,**[*2] **Yash Lala ,**[*2] **Christian Loanzon ,**[*2] **Yizhou Sun** [2]

[1] Electrical and Computer Engineering Department, University of California, 420 Westwood Plaza, Los Angeles, CA 90095
[2] Computer Science Department, University of California, 404 Westwood Plaza, Los Angeles, CA 90095
{mlpotter124,hamlinliu250,yashlala,chrizon}@g.ucla.edu, yzsun@cs.ucla.edu

Abstract

We present a novel movie recommendation system, GRU4RecBE, which extends the GRU4Rec architecture with rich item features extracted by the pre-trained BERT model. GRU4RecBE outperforms state-of-the-art session-based models over the benchmark MovieLens 1m and MovieLens 20m datasets.

Introduction

Recommendation Systems (RSs) are critical for users to make choices in a massive and rapid growing market of content and products. Anonymous implicit feedback session-based RSs (SBRSs) capture the user's interest by leveraging the user's past item click(s) within a current click-session to predict the next item to recommend. Implicit feedback SBRSs are prevalent due to their privacy-preserving nature of not tracking an individual's entire interaction history over multiple sessions. This is a challenging task due to uncertainty and limited information in user behavior. Literature on SBRSs is vast (see Supplemental Materials), where benchmark deep learning methods (Sun et al. 2019) model a user's interest only with click information, but very few studies include item information (Wang et al. 2021; Hidasi et al. 2016b). We extend SBRSs by incorporating rich item features that reveal user behaviour patterns, along with item ids for movie recommendation. Particularly, we extend GRU4Rec (Hidasi et al. 2016a) with features extracted from text descriptions of the items of interest via the Bidirectional Encoder Representations from Transformers (BERT) architecture (Devlin et al. 2018). Extensive experiments on the benchmark MovieLens 1m and MovieLens 20m datasets show that our model outperforms state-of-the-art (SOTA) SBRS models.

Model Architecture

We extend the GRU4Rec architecture by utilizing movie ids, along with IMDb movie plots for SBRS modeling (c.f. Figure 1); the resulting method is termed as GRU4RecBE.
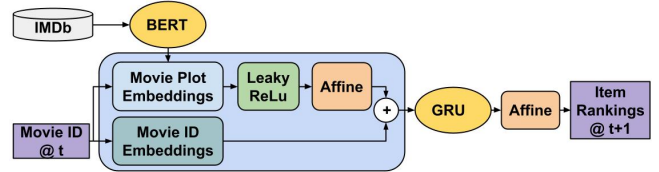

Figure 1: GRU4RecBE Model Architecture

Embedding Layers: We consider as input a sequence of T ids of the movies watched in a session, where each id is denoted by x_t, $t \in [1, \ldots, T]$. Given each id x_t, GRU4RecBE extracts latent representations via two embedding layers. The first layer receives x_t and extracts its id embedding $\mathbf{x}_t^{movie} \in \mathbb{R}^K$. The second receives the movie plot pertaining to x_t and extracts a plot embedding $\mathbf{x}_t^{plot} \in \mathbb{R}^B$. The parameters of the plot embedding layer are initialized with features extracted by the pre-trained BERT model (Devlin et al. 2018), i.e., the hidden representations extracted by the last layer from the [CLS] token of the movie plot.

Projection Layer: The plot embedding $\mathbf{x}_t^{plot}$ is projected into same dimensions as the id embedding via a leaky relu activation and an affine layer, resulting in the projected plot embedding $\mathbf{x}_t^{pp} \in \mathbb{R}^K$.

Input Aggregation: The id and plot embeddings are aggregated by addition to form $\mathbf{x}_t^{merge} \in \mathbb{R}^K$.

Gated Recurrent Unit - RNN Layer: A Gated Recurrent Unit (GRU) layer receives the merged embedding $\mathbf{x}_t^{merge}$, in which the hidden state $\mathbf{h}_t \in \mathbb{R}^D$ encodes the sequential session behavior up to time step t.

Decoder: The decoder is an affine layer, which receives $\mathbf{h}_t$ and predicts a real-valued recommendation score for each movie among all movies in the dataset.

Training GRU4RecBE

We employ the Bayesian Personalized Ranking (BPR)-Max (Hidasi and Alexandros 2018) loss function to train GRU4RecBE on session-histories and respective next movie selections (see Supplemental Materials on Objective Functions for detail). While training GRU4RecBE, we aim to capture the unique information carried by both the movie ids, as well as the movie plots. Therefore, we follow the optimization approach by Hidasi et al. (2016b), where we alter-

Datasets	Metrics	POP	GRU4Rec	GRU4Rec+	SR-GNN	Caser	SASRec	NextItNet	BERT4Rec	**GRU4RecBE**
	HIT@10	0.1358	0.6207	0.6351	0.6449	0.6692	0.6629	0.6755	0.6970	**0.7083**
	HIT@5	0.0715	0.4673	0.5103	0.5169	0.5353	0.5434	0.5551	0.5876	**0.5954**
	HIT@1	0.0141	0.1583	0.2092	0.2348	0.2194	0.2351	0.2593	0.2863	**0.2944**
ML-1m	NDCG@10	0.0621	0.3627	0.4064	0.4231	0.4268	0.4368	0.4368	0.4818	**0.4894**
	NDCG@5	0.0416	0.3196	0.3705	0.3816	0.3832	0.3980	0.3983	0.4454	**0.4529**
	MRR	0.0627	0.3041	0.3462	0.3683	0.3648	0.3790	0.3817	0.4254	**0.4324**
	HIT@10	0.1378	0.5844	0.6524	0.4913	0.5427	0.7136	0.7313	0.7473	**0.7908**
	HIT@5	0.0805	0.4657	0.5118	0.3578	0.3804	0.5727	0.6009	0.6323	**0.6759**
	HIT@1	0.0221	0.1459	0.2021	0.1363	0.1232	0.2544	0.2832	0.3440	**0.3638**
ML-20m	NDCG@10	0.0695	0.3637	0.4087	0.2931	0.3062	0.4665	0.4929	0.5340	**0.5670**
	NDCG@5	0.0511	0.3090	0.3630	0.2517	0.2538	0.4208	0.4505	0.4967	**0.5297**
	MRR	0.0709	0.2967	0.3476	0.2516	0.2529	0.4026	0.4303	0.4785	**0.5063**

Table 1: Hit, NDCG, and MRR for benchmark recommender systems on the ML-1M and ML-20m datasets.

nate between optimizing the weights of the plot embedding and id embedding layers at each epoch.

Datasets

We use the MovieLens (ML) 1m and ML 20m datasets (Harper and Konstan 2015) for evaluation, where we represent each user as a session. ML is a popular benchmark dataset for evaluating RS, which contains user id, movie id, and timestamp for each user click. We use the publicly available ImdbId identifier and IMDbPY [1] to acquire movie plot summaries for rich feature information.

Experimental Setup

We use the conventional *leave-one-out* evaluation for the next movie recommendation task. For each user in the dataset, we hold out the last movie of the session as the test set, hold out the second to last movie as the validation set, and the remaining movies for the training set. The maximum sequence length is set to 200 movies as in Sun et al. (2019). For fair evaluation, we follow the common convention (Sun et al. 2019) of pairing each ground-truth movie in the test set with 100 random negative movies which the user has not interacted with. The negative movies are sampled according to their popularity without replacement. Therefore, we rank the ground-truth movie amongst the 100 random negative movies for each user. We compare GRU4RecBE with several SOTA methods (see Supplemental Materials on Baseline Methods for detail). Our code is publicly available at https://github.com/yashlala/attentive-session-based-recs.

Experimental Results

Table 1 summarizes the best evaluations of all baseline models as well as GRU4RecBE on the ML-1M and ML-20m datasets. We show that using rich features significantly improves performance for SBRSs across all metrics such as Hit rate, Normalized Discount Cumulative Gain (NDCG), and Mean Reciprical Rank (MRR). Furthermore, as the dataset size increases, the margin of improvement between GRU4RecBE and the next best method increases. We further

---

[1] https://imdbpy.github.io/

demonstrate example set of recommendations in the ML-1M dataset using GRU4RecBE in the supplementary material.

Conclusion and Future Work

Previous work in the field of SBRSs has largely focused on the user's click history, with few studies including item information (Wang et al.2021). Our major contribution is successfully establishing the utility of movie embedding information via transformers to these systems. We demonstrate that GRU4RecBE outperforms SOTA feature-agnostic recommender systems, especially when applied in conjunction with unconventional training techniques such as alternating optimizers. As future work, utilizing attention mechanisms along with feature information could improve the model performance and provide better interpretability.

References

Devlin, J.; Chang, M.; Lee, K.; and Toutanova, K. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR*, abs/1810.04805.

Harper, F. M.; and Konstan, J. A. 2015. The MovieLens Datasets: History and Context. *ACM Trans. Interact. Intell. Syst.*, 5(4).

Hidasi, B.; and Alexandros, K. 2018. Recurrent Neural Networks with Top-k Gains for Session-Based Recommendations. In *Proceedings of the 27th ACM International*, CIKM '18, 843–852. New York, NY, USA. ISBN 9781450360142.

Hidasi, B.; Karatzoglou, A.; Baltrunas, L.; and Tikk, D. 2016a. Session-based Recommendations with Recurrent Neural Networks. arXiv:1511.06939.

Hidasi, B.; Quadrana, M.; Karatzoglou, A.; and Tikk, D. 2016b. Parallel Recurrent Neural Network Architectures for Feature-Rich Session-Based Recommendations. In *Proceedings of the 10th ACM*, RecSys '16, 241–248.

Sun, F.; Liu, J.; Wu, J.; Pei, C.; Lin, X.; Ou, W.; and Jiang, P. 2019. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. In *Proceedings of the 28th ACM International*, CIKM '19.

Wang, S.; Cao, L.; Wang, Y.; Sheng, Q. Z.; Orgun, M. A.; and Lian, D. 2021. A Survey on Session-Based Recommender Systems. *ACM Comput. Surv.*, 54(7).