Table 4: Performance comparison w.r.t. different base models enhanced by our proposed method TCPSRec.

Method	Meituan		Beauty	
	HR@10	NDCG@10	HR@10	NDCG@10
GRU4Rec	0.2320	0.1479	0.0640	0.0369
+TCPSRec	**0.2346**	**0.1511**	**0.0673**	**0.0382**
Caser	0.2433	0.1526	0.0677	0.0389
+TCPSRec	**0.2441**	**0.1553**	**0.0683**	**0.0394**
SASRec	0.2708	0.1745	0.0683	0.0405
+TCPSRec	**0.2797**	**0.1846**	**0.0841**	**0.0440**
TiSASRec	0.2711	0.1735	0.0694	0.0408
+TCPSRec	**0.2786**	**0.1808**	**0.0835**	**0.0433**

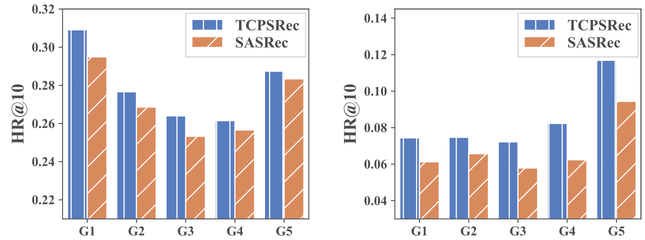
4.3 Further Analysis

In this section, we conduct further analysis on TCPSRec .

4.3.1 ***Ablation Study.*** To validate the contribution of each pre-training objective in TCPSRec, we perform an ablation study on four real-world datasets, *i.e.,* Meituan, Yelp, Beauty, and Sports. We present the experimental results in Figure 3, where "*w/o* GIM", "*w/o* LIM", "*w/o* CPM" and "*w/o* FPM" denote the variants by removing the global invariant modeling objective, the local invariant modeling objective, the coarse-grained periodicity modeling objective, and the fine-grained periodicity modeling objective in TCPSRec, respectively. From this figure, we can observe that the performance of TCPSRec significantly decreases when removing any pre-training objective. Meanwhile, all the four variants perform better than the base model SASRec.

Furthermore, the importance of these objectives varies on different datasets. The global invariant modeling objective (GIM) brings more improvement than the other objectives. It is because GIM is more basic than the other two objectives. In addition, the improvements achieved by coarse-grained periodicity modeling (CPM) and fine-grained periodicity modeling (FPM) vary greatly across different datasets, which implies that different periodicities might be needed in different scenarios. Overall, the ablation study indicates that all pre-training objectives contribute to the performance improvement of our approach.

4.3.2 ***Applying the Pre-training Strategy on Other Models.*** With the default settings, we adopt SASRec as the based model in TCPSRec (refer to Section 2). Actually, our proposed pre-training method is generally applicable for other sequential recommender systems, such as GRU4Rec [13], Caser [28], and TiSASRec [18]. Therefore, we conduct experiments to examine whether the proposed pre-training strategy can also improve these sequential recommender systems. We report the results on the industry dataset Meituan and the public dataset Beauty in Table 4. From this table, we can observe that the proposed method can consistently improve the performance of GRU4Rec, Caser, SASRec, and TiSASRec, which further shows the effectiveness of the proposed method. In addition, we find that the proposed temporal pre-training method can also improve the performance of TiSASRec, a time-aware sequential recommender system, suggesting that our approach can better leverage temporal information. The experimental results further indicate the usefulness of temporal information in improving the recommendation performance.



(a) Meituan **(b) Beauty**

Figure 4: Analysis of recommendation for different sparse-level users. $G5$ **denotes the group of users with the largest average number of interactions, and** $G1$ **is the opposite.**

4.3.3 ***Impact of Data Sparsity Levels.*** Traditional sequential recommender systems often suffer from data sparsity issues because they only rely on interaction sequences for training. To study the effect of the sparsity of interaction data on training, we group users according to the sparsity of their interactions and evaluate the performance of TCPSRec on different user groups separately. Concretely, we divide all users into five user groups according to their interaction number, each with the same total number of interactions (ensuring an equal amount of training data) but different numbers of users. Then, we train TCPSRec and SASRec separately on the interaction data of each user group, and compare their recommendation performance on these five user groups. The results are reported in Figure 4, where users in group $G1$ have the sparsest interaction data. It can be observed that TCPSRec outperforms SASRec on all user groups, which implies that the improvement brought by our method is stable and robust for users with different data sparsity levels. In addition, according to the results in Table 3, TCPSRec also achieves more improvement on the relatively sparse Beauty dataset than on the dense Meituan dataset. These findings show that the proposed TCPSRec can effectively alleviate the sparsity issue for sequential recommendation.

4.3.4 ***Impact on the Convergence Speed.*** To examine how pre-training affects the convergence speed, we further compare the loss of TCPSRec and SASRec during training on the recommendation task. Note that SASRec is randomly initialized, while TCPSRec corresponds to SASRec initialized by the pre-trained parameters. We adopt the early stopping strategy in the training process of recommendation tasks, *i.e.,* if HR@10 on the validation data does not increase for ten successive epochs, we consider that the model has converged and stop the training. Figure 5 shows the training process of SASRec and TCPSRec on Meituan and Beauty datasets. This figure shows that the proposed TCPSRec converges faster and achieves lower loss, while the SASRec needs more epochs to achieve similar performance. Specifically, on the Beauty dataset, SASRec requires more than 70 epochs of training to converge, while TCPSRec only needs about 20 epochs of fine-tuning to achieve better performance. The experimental results demonstrate that the proposed TCPSRec can significantly accelerate the convergence of recommendation models and achieve better performance.

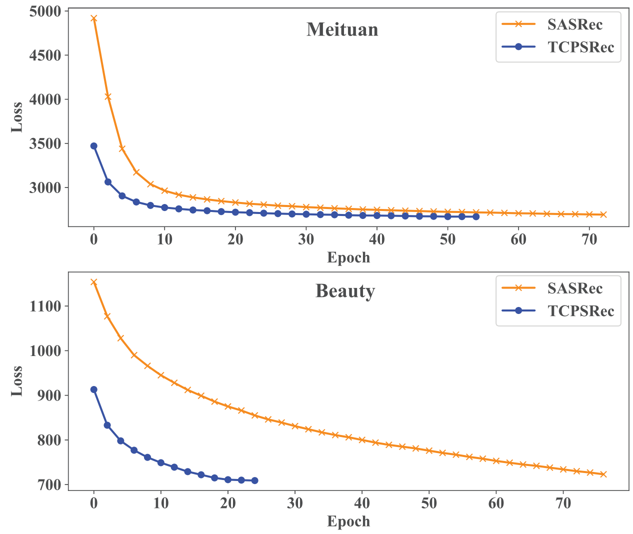


Figure 5: Analysis of convergence speed. SASRec is randomly initialized, while TCPSRec corresponds to the SASRec variant initialized by the pre-trained parameters.

5 RELATED WORK

In this section, we review the related work in two aspects, namely sequential recommendation and pre-training for recommendation.

5.1 Sequential Recommender Systems

Sequential recommender systems have gained success in capturing user preferences from sequential interaction data [33]. Early studies [11, 24] employ the Markov chain method to model the item transitions for capturing sequential relatedness. Subsequently, with the development of deep learning [17], Convolutional Neural Networks (CNN) [28], Recurrent Neural Networks (RNN) [13, 34] and other effective variants [14, 23] are applied to encode the user behavior sequences. Recently, significant attention has been dedicated to Transformer [30] based sequential models [3, 16, 20, 36]. For instance, SASRec [16] applies self-attention networks to model user behavior sequences, which effectively improves the recommendation performance. Furthermore, recent works [3, 36] introduce self-supervised learning into sequential user behavior modeling. Despite the remarkable success, the above methods seldom consider temporal information for modeling historical interactions. More recently, several methods [6, 8, 18, 32, 38] further leverage the temporal information for sequential recommendation. For example, TiSASRec [18] utilizes the time intervals between interactions to improve the self-attention mechanism of SASRec. TASER [38] takes into consideration both absolute time patterns and relative time intervals reflected by timestamps. However, these studies consider timestamps as auxiliary item features [6, 32] or use timestamps to develop time-aware models [8, 18, 38], which inevitably requires additional computational costs in the inference stage. Different from previous studies, our approach devises self-supervised objectives tailored for sequential recommendation, which can effectively capture temporal correlations for improving recommendation via pre-training.

5.2 Pre-training for Recommendation

Since the success of pre-training models in natural language processing [7, 37], the pre-training technique has also gained significant attention in the field of recommender systems [39]. Recently, some works apply the pre-training technique to sequential recommender systems [27, 35, 40, 45] and graph-based recommender systems [9, 21, 31] to better learn the representation of users and items. As for pre-trained sequential recommender systems, most of existing studies adopt Transformer [30] architecture. For example, BERT4Rec [27] and BERT4SessRec [2] employ the BERT [7] to model user behavior sequences and capture the bidirectional information. Based on the mutual information maximization principle, S^3-Rec [45] designs four self-supervised learning objectives to capture the correlations between sequence data and contextual information. UPRec [35] incorporates user information, such as user attributes and social graphs, into pre-trained models for recommendation. PeterRec [40] further improves the fine-tuning efficiency for the pre-trained parameters. With the development of contrastive learning (CL) [15, 29], somes works [1, 3, 22, 36] utilize CL in multi-task learning framework to improve sequential recommendation. For instance, CL4SRec [36] takes advantage of data augmentation strategies to construct self-supervised tasks and extract meaningful interaction patterns with CL. More recently, ICLRec [3] further propose to leverage prototypical contrastive learning to improve sequencial recommendation. These methods seldom consider temporal information in constructing the self-supervised signals, which is less capable of modeling sequential characteristics for the recommendation task.

6 CONCLUSIONS

In this paper, we propose a novel **T**emporal **C**ontrastive **P**re-training method for **S**equential **Rec**ommendation, named **TCPSRec**. Based on the temporal intervals of interactions, we divided the interaction sequence into coherent subsequences, and designed temporal pre-training objectives accordingly. Specifically, TCPSRec modeled both the invariance and periodicity of user behavior. We model the invariance at item level, and consider both global invariance within a sequence and local invariance within a subsequence. Moreover, we modeled the periodicity at subsequence level, considering coarse-grained periodicity and fine-grained periodicity. To integrate the above motivations, we developed a unified contrastive learning appraoch with four pre-training objectives tailored for sequential recommendation. The experimental results on six real-world datasets demonstrate the effectiveness of the proposed TCPSRec.

As future work, we will further consider extending our pre-training method to incorporate more kinds of side information, such as social graphs or auxiliary attributes.

ACKNOWLEDGMENTS

This work was partially supported by Beijing Natural Science Foundation under Grant No. 4222027, National Natural Science Foundation of China under Grant No. 61872369, and Beijing Outstanding Young Scientist Program under Grant No. BJJWZYJH012019100020098. This work was also partially supported by Beijing Academy of Artificial Intelligence (BAAI) and Meituan. Xin Zhao is the corresponding author.

REFERENCES

[1] Shuqing Bian, Wayne Xin Zhao, Kun Zhou, Jing Cai, Yancheng He, Cunxiang Yin, and Ji-Rong Wen. 2021. Contrastive Curriculum Learning for Sequential User Behavior Modeling via Data Augmentation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 3737–3746.

[2] Xusong Chen, Dong Liu, Chenyi Lei, Rui Li, Zheng-Jun Zha, and Zhiwei Xiong. 2019. Bert4sessrec: Content-based video relevance prediction with bidirectional encoder representations from transformer. In *Proceedings of the 27th ACM International Conference on Multimedia*. 2597–2601.

[3] Yongjun Chen, Zhiwei Liu, Jia Li, Julian McAuley, and Caiming Xiong. 2022. Intent Contrastive Learning for Sequential Recommendation. In *Proceedings of the ACM Web Conference 2022*. 2172–2182.

[4] Mingyue Cheng, Fajie Yuan, Qi Liu, Xin Xin, and Enhong Chen. 2021. Learning Transferable User Representations with Sequential Behaviors via Contrastive Pre-training. In *ICDM*. IEEE, 51–60.

[5] Junsu Cho, Dongmin Hyun, SeongKu Kang, and Hwanjo Yu. 2021. Learning heterogeneous temporal patterns of user preference for timely recommendation. In *Proceedings of the Web Conference 2021*. 1274–1283.

[6] Qiang Cui, Chenrui Zhang, Yafeng Zhang, Jinpeng Wang, and Mingchen Cai. 2021. ST-PIL: Spatial-Temporal Periodic Interest Learning for Next Point-of-Interest Recommendation. In *CIKM*. 2960–2964.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[8] Ziwei Fan, Zhiwei Liu, Jiawei Zhang, Yun Xiong, Lei Zheng, and Philip S Yu. 2021. Continuous-time sequential recommendation with temporal graph collaborative transformer. In *CIKM*. 433–442.

[9] Bowen Hao, Jing Zhang, Hongzhi Yin, Cuiping Li, and Hong Chen. 2021. Pre-training graph neural networks for cold-start users and items representation. In *WSDM*. 265–273.

[10] F Maxwell Harper and Joseph A Konstan. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)* 5, 4 (2015).

[11] Ruining He and Julian McAuley. 2016. Fusing similarity models with markov chains for sparse sequential recommendation. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 191–200.

[12] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web*. 507–517.

[13] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based recommendations with recurrent neural networks. *ICLR* (2016).

[14] Jin Huang, Wayne Xin Zhao, Hongjian Dou, Ji-Rong Wen, and Edward Y Chang. 2018. Improving sequential recommendation with knowledge-enhanced memory networks. In *SIGIR*. 505–514.

[15] Ashish Jaiswal, Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Debapriya Banerjee, and Fillia Makedon. 2020. A survey on contrastive self-supervised learning. *Technologies* 9, 1 (2020), 2.

[16] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *ICDM*. IEEE, 197–206.

[17] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444.

[18] Jiacheng Li, Yujie Wang, and Julian McAuley. 2020. Time interval aware self-attention for sequential recommendation. In *Proceedings of the 13th international conference on web search and data mining*. 322–330.

[19] Yicong Li, Hongxu Chen, Xiangguo Sun, Zhenchao Sun, Lin Li, Lizhen Cui, Philip S Yu, and Guandong Xu. 2021. Hyperbolic hypergraphs for sequential recommendation. In *CIKM*. 988–997.

[20] Yang Li, Tong Chen, Peng-Fei Zhang, and Hongzhi Yin. 2021. Lightweight self-attentive sequential recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 967–977.

[21] Yong Liu, Susen Yang, Chenyi Lei, Guoxin Wang, Haihong Tang, Juyong Zhang, Aixin Sun, and Chunyan Miao. 2021. Pre-training graph transformer with multimodal side information for recommendation. In *Proceedings of the 29th ACM International Conference on Multimedia*. 2853–2861.

[22] Ruihong Qiu, Zi Huang, Hongzhi Yin, and Zijian Wang. 2022. Contrastive learning for representation degeneration problem in sequential recommendation. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 813–823.

[23] Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. 2017. Personalizing session-based recommendations with hierarchical recurrent neural networks. In *proceedings of the Eleventh ACM Conference on Recommender Systems*. 130–137.

[24] Steffen Rendle. 2010. Factorization machines. In *2010 IEEE International conference on data mining*. IEEE, 995–1000.

[25] Paul Resnick and Hal R Varian. 1997. Recommender systems. *Commun. ACM* 40, 3 (1997), 56–58.

[26] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2011. Introduction to recommender systems handbook. In *Recommender systems handbook*. Springer, 1–35.

[27] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 1441–1450.

[28] Jiaxi Tang and Ke Wang. 2018. Personalized top-n sequential recommendation via convolutional sequence embedding. In *Proceedings of the eleventh ACM international conference on web search and data mining*. 565–573.

[29] Aaron Van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv e-prints* (2018), arXiv–1807.

[30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[31] Hui Wang, Kun Zhou, Wayne Xin Zhao, Jingyuan Wang, and Ji-Rong Wen. 2022. Curriculum Pre-Training Heterogeneous Subgraph Transformer for Top-N Recommendation. *ACM Trans. Inf. Syst.* (mar 2022). https://doi.org/10.1145/3528667

[32] Jianling Wang, Raphael Louca, Diane Hu, Caitlin Cellier, James Caverlee, and Liangjie Hong. 2020. Time to Shop for Valentine's Day: Shopping Occasions and Sequential Recommendation in E-commerce. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 645–653.

[33] Shoujin Wang, Liang Hu, Yan Wang, Longbing Cao, Quan Z. Sheng, and Mehmet Orgun. 2019. Sequential Recommender Systems: Challenges, Progress and Prospects. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 6332–6338. https://doi.org/10.24963/ijcai.2019/883

[34] Chao-Yuan Wu, Amr Ahmed, Alex Beutel, Alexander J Smola, and How Jing. 2017. Recurrent recommender networks. In *Proceedings of the tenth ACM international conference on web search and data mining*. 495–503.

[35] Chaojun Xiao, Ruobing Xie, Yuan Yao, Zhiyuan Liu, Maosong Sun, Xu Zhang, and Leyu Lin. 2021. UPRec: User-Aware Pre-training for Recommender Systems. *arXiv preprint arXiv:2102.10989* (2021).

[36] Xu Xie, Fei Sun, Zhaoyang Liu, Shiwen Wu, Jinyang Gao, Jiandong Zhang, Bolin Ding, and Bin Cui. 2022. Contrastive learning for sequential recommendation. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 1259–1273.

[37] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems* 32 (2019).

[38] Wenwen Ye, Shuaiqiang Wang, Xu Chen, Xuepeng Wang, Zheng Qin, and Dawei Yin. 2020. Time matters: Sequential recommendation with complex temporal information. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1459–1468.

[39] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Jundong Li, and Zi Huang. 2022. Self-Supervised Learning for Recommender Systems: A Survey. *arXiv preprint arXiv:2203.15876* (2022).

[40] Fajie Yuan, Xiangnan He, Alexandros Karatzoglou, and Liguang Zhang. 2020. Parameter-efficient transfer from sequential behaviors for user modeling and recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1469–1478.

[41] Xu Yuan, Hongshen Chen, Yonghao Song, Xiaofang Zhao, and Zhuoye Ding. 2021. Improving Sequential Recommendation Consistency with Self-Supervised Imitation. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, Zhi-Hua Zhou (Ed.). International Joint Conferences on Artificial Intelligence Organization, 3321–3327. Main Track.

[42] Wayne Xin Zhao, Junhua Chen, Pengfei Wang, Qi Gu, and Ji-Rong Wen. 2020. Revisiting Alternative Experimental Settings for Evaluating Top-N Item Recommendation Algorithms. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2329–2332.

[43] Wayne Xin Zhao, Yupeng Hou, Xingyu Pan, Chen Yang, Zeyu Zhang, Zihan Lin, Jingsen Zhang, Shuqing Bian, Jiakai Tang, Wenqi Sun, Yushuo Chen, Lanling Xu, Gaowei Zhang, Zhen Tian, Changxin Tian, Shanlei Mu, Xinyan Fan, Xu Chen, and Ji-Rong Wen. 2022. RecBole 2.0: Towards a More Up-to-Date Recommendation Library. In *Proceedings of the 31th ACM International Conference on Information & Knowledge Management*.

[44] Wayne Xin Zhao, Shanlei Mu, Yupeng Hou, Zihan Lin, Yushuo Chen, Xingyu Pan, Kaiyuan Li, Yujie Lu, Hui Wang, Changxin Tian, et al. 2021. Recbole: Towards a unified, comprehensive and efficient framework for recommendation algorithms. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 4653–4664.

[45] Kun Zhou, Hui Wang, Wayne Xin Zhao, Yutao Zhu, Sirui Wang, Fuzheng Zhang, Zhongyuan Wang, and Ji-Rong Wen. 2020. S3-rec: Self-supervised learning for sequential recommendation with mutual information maximization. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 1893–1902.