Model performance report

The Self-Attention Sequential Recommendation model (SASRec) (Kang and McAuley, 2018) was run on a specially preprocessed dataset (Garcia Ling *et al.*, 2022), using the parameters listed in Table 1..

Table < 1 > Default parameters for SASRec model

Parameter	**Value**	**Comments**
batch_size	128	Batch size
lr	0.001	Learning rate
maxlen	50	Maximum sequence length
hidden_units	50	Embedding size
num_blocks	2	Number of transformer layers
num_epochs	1000	Number of epochs
num_heads	1	Number of attention heads
dropout_rate	0.2	Dropout rate
l2_emb	0.0	Regularization parameter

Normalized Discounted Cumulative Gain (NDCG@10) and Hit Rate (HR) were used to evaluate the model's performance. For this task, each of the 5,680 sequences was associated with an array of 101 Item IDs. The first position on each array was reserved for the “Truth” label—the actual last Item ID from the corresponding sequence (from the test set). The remaining 100 positions were populated with randomly selected Item IDs from the dataset, ensuring these IDs were not part of the sequence under evaluation. The model assigned a relevance score to each of the 101 Item IDs in the array. These scores show how relevant each item is to the user based on their past interactions. After scoring, the array was sorted by these scores from highest to lowest. The effectiveness of the model was then evaluated by finding where the "Truth" label—the actual last Item ID from the test set sequence—appeared in this sorted array. Based on this position, the model calculated the NDCG@10 and HR metrics to determine how accurately it identified relevant items. The evaluation of the models was conducted at every 20th epoch during training, and values for both metrics on the validation and test sets were obtained.

The model with default parameters achieved maximum values of 0.30 for NDCG@10 and 0.44 for HR, which occurred at epoch number 140 (Figure 1).

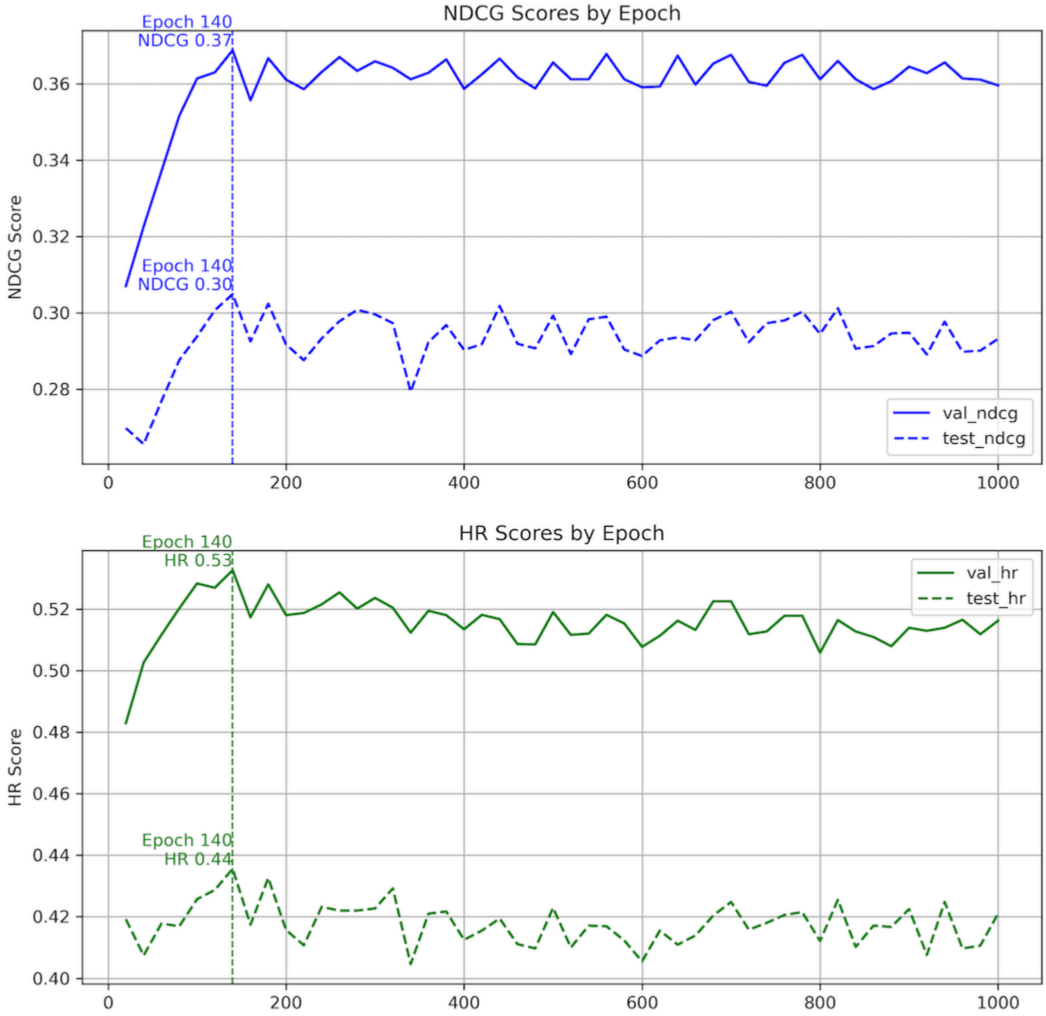

Fig < 1 > Performance metrics over epochs for the model with default parameters

The experiment continued by adjusting the embedding size parameters. Two different sizes were tested: one smaller than 50 and one larger, specifically 10 and 100. Since the first model peaked at epoch 140, it was decided to run the models with embedding sizes of 10 and 100 for only 300 epochs to save time (all the rest parameters remained unchanged). The obtained results (Table 2) led to the conclusion that reducing the embedding size to 10 resulted in a decline of the model's performance, while increasing the embedding size to 100 did not improve the NDCG and HR metrics. These findings suggest that an embedding size of 50 is optimal for capturing the characteristics of the Item IDs in a given dataset.

Table < 2 > Performance results of the experiments by embedding size

emb_size	# of epochs	test_ndcg	test_hr
10	300	0.28	0.43
50	1000	**0.30**	**0.44**
100	300	0.30	0.42

The observed dynamics in the metric values and consultation with a professor suggested that the default learning rate was too high, preventing the model from finding a local minimum. Consequently, three additional experiments were conducted using embedding sizes of 10, 50, and 100, with the learning rate reduced to 0.0001 (from the default 0.001), while all other parameters remained unchanged. Plots of all experiments are available in Appendix 1.

Changing the learning rate parameter led to improved NDCG@10 and HR for models with embedding sizes of 50 and 100; the former achieved the highest values for both metrics. However, the model with an embedding size of 10 showed poorer results (Table 3). This further demonstrated that a smaller embedding size leads to a decline in the model's performance, while increasing the embedding size does not yield any substantial improvements.

Table < 3 > Performance results of experiments with a learning rate of 0.0001

emb_size	**learning rate**	**# of epochs**	**test_ndcg**	**test_hr**
10	0.0001	1000	0.26	0.40
50	0.0001	1000	**0.32**	**0.45**
100	0.0001	2000	0.31	0.43

The obtained results indicate that, on average, the Item ID predicted by the model appears in the top 10 Item IDs in 45% of cases (HR metric). Furthermore, the predicted Item ID is, on average, ranked between the 7th and 8th positions within the top 10 (NDCG metric). This indicates that there is still room for further improvement in the model's performance.

References

Garcia Ling, C., HMGroup, E., Rim, F., inversion, Ferrando, J., Maggie, neuraloverflow, and xlsrin (2022) *H&M Personalized Fashion Recommendations* [online]. Available from: https://www.kaggle.com/competitions/h-and-m-personalized-fashion-recommendations.

Kang, W.-C. and McAuley, J. (2018) *Self-Attentive Sequential Recommendation*. In: *2018 IEEE International Conference on Data Mining (ICDM)* [online]2018 IEEE International Conference on Data Mining (ICDM). Singapore: IEEE, pp. 197–206. Available from: https://ieeexplore.ieee.org/document/8594844/ [Accessed 26 December 2024].

Appendix 1

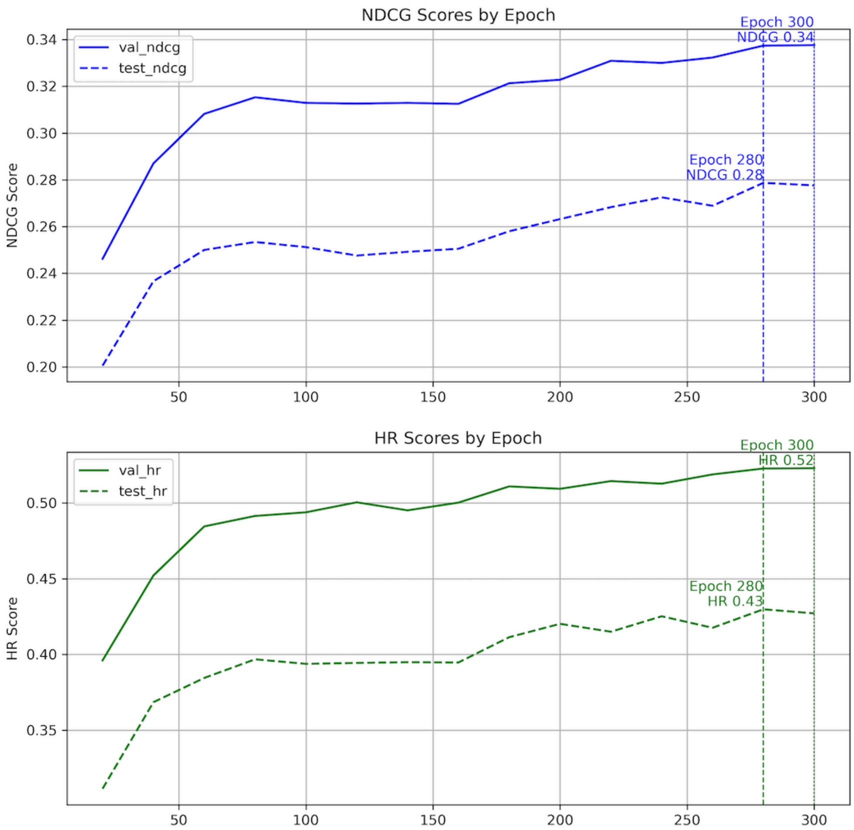

Fig < 1 > Experiment #2. embedding size=10; lr=0.001; epochs=300

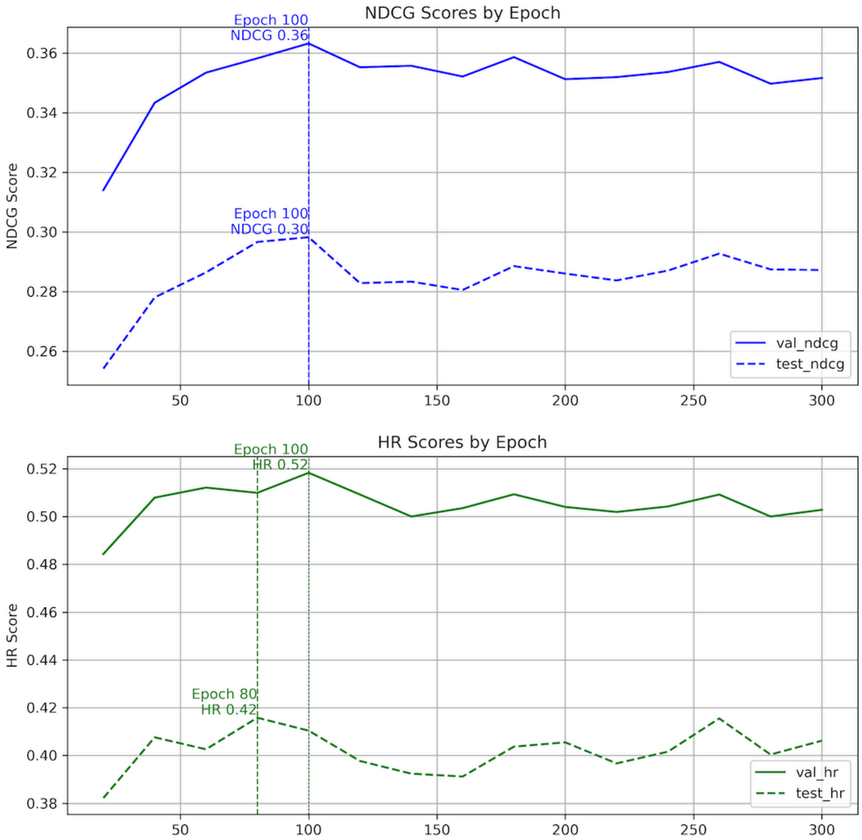

Fig < 2 > Experiment #3. embedding size=100; lr=0.001; epochs=300

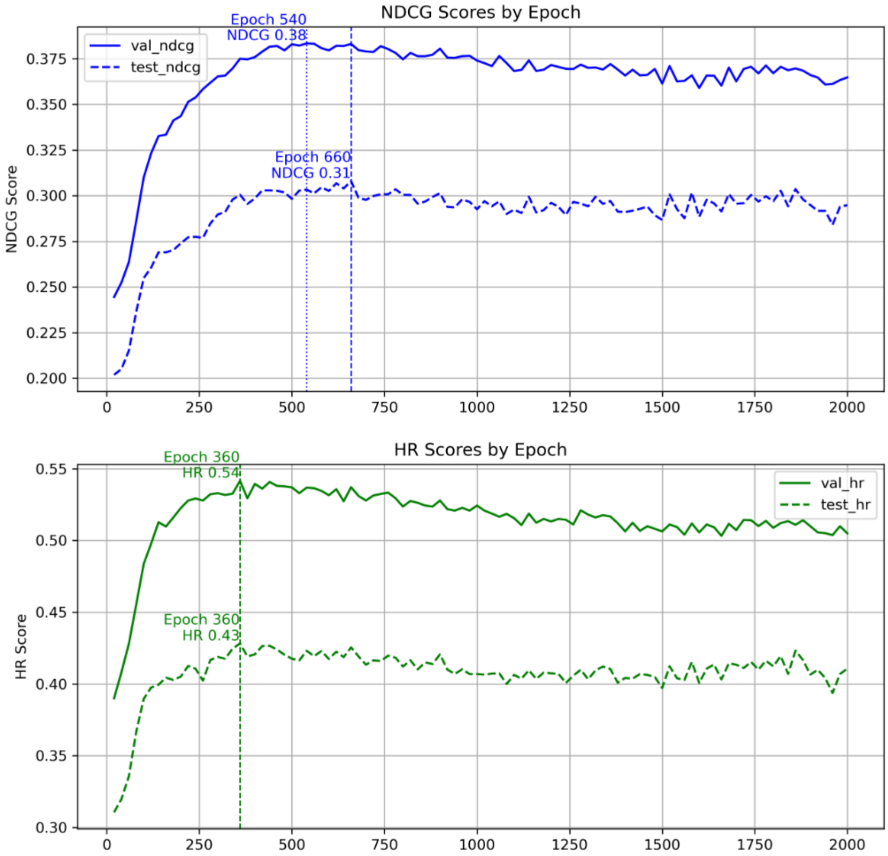

Fig < 3 > Experiment #4. embedding size=100; lr=0.0001; epochs=2000

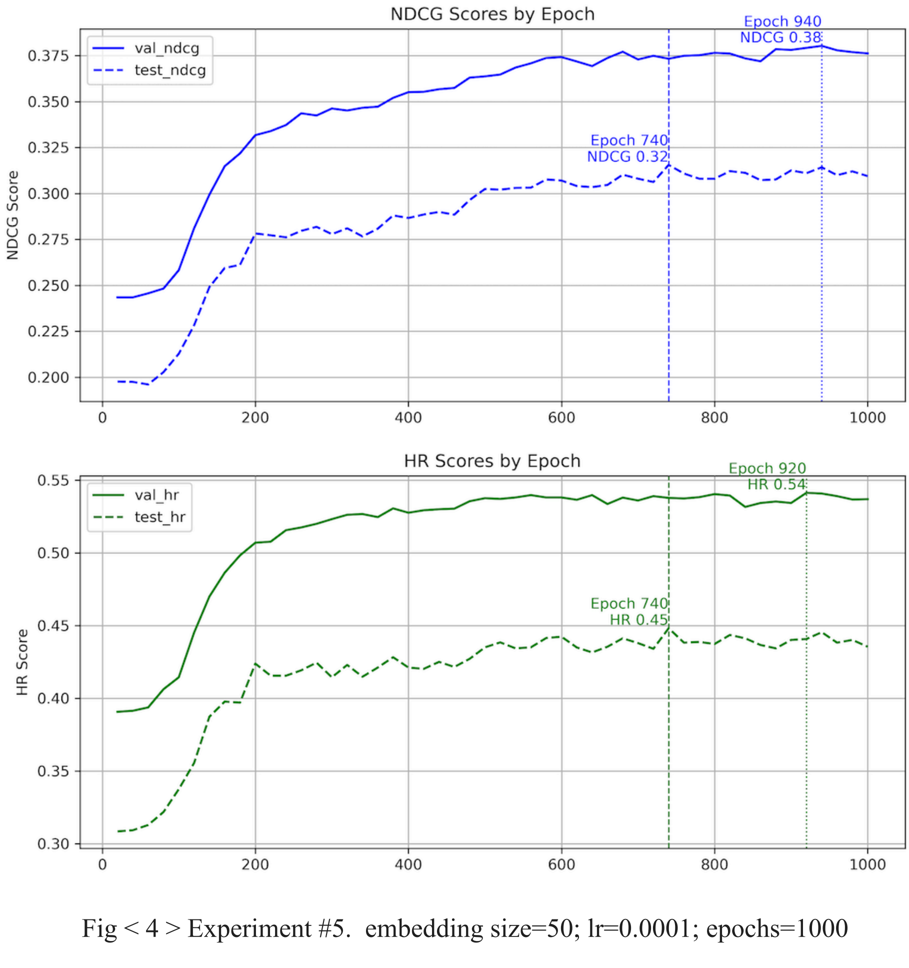

Fig < 4 > Experiment #5. embedding size=50; lr=0.0001; epochs=1000

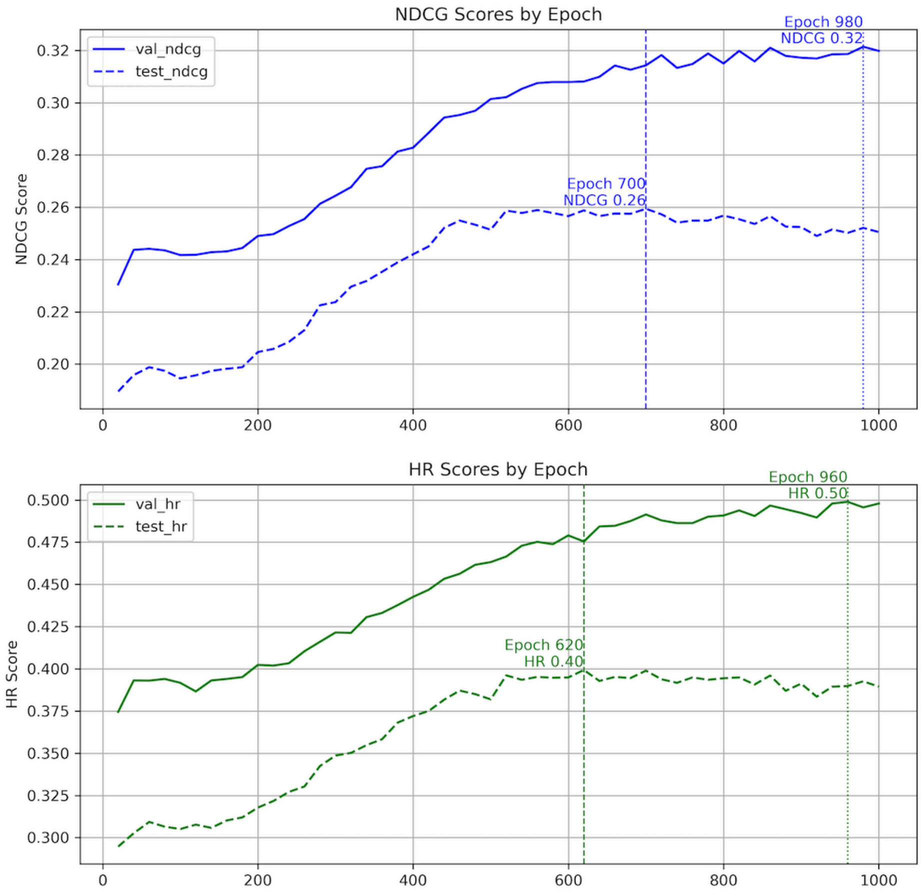

Fig < 5 > Experiment #6. embedding size=10; lr=0.0001; epochs=1000