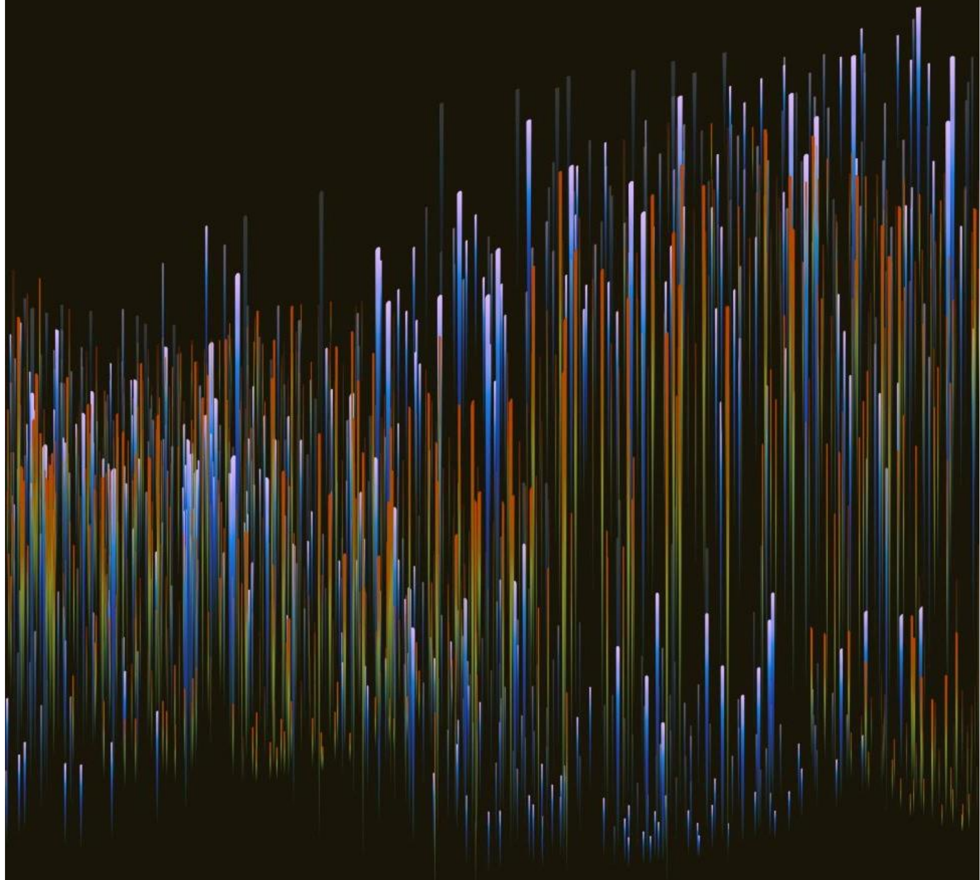
Context Preserving Data Augmentation for Sequential Recommendation

Benjamin Amankwata and Kenneth K. Fletcher

Computer Science Department

University of Massachusetts Boston



Outline

- Motivation
- Contributions
- Model
  - Data Augmentation (Strongly Restricted Permutations)
  - Item Representation Learning
  - CNNs for Sequential Recommendations
- Research Questions
- Experiments and Results
- Conclusions and Future Work

Motivation

1. Item ordering in user historical interaction (UHI) encodes important item relation signals crucial for sequential recommendation (SR)
2. Most UHI training data for SR is induced by an underling candidate generation process (UCGP) that suffers bias, distortion and noise
3. Rigidly adopting the item ordering the UCGP distorts fundamental item relation signals that are crucial for item representation learning in SR

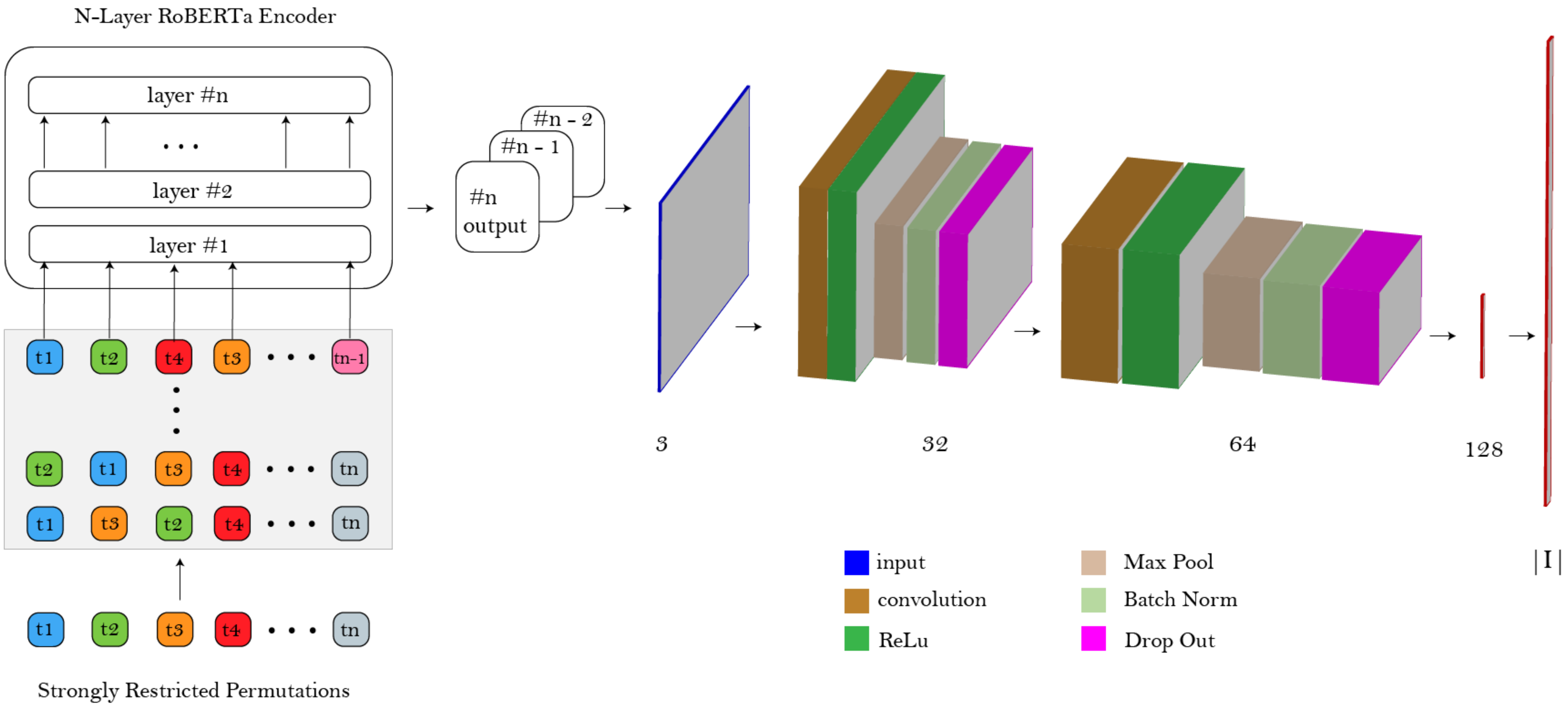
Contributions

1. We propose an effective self-supervised approach for recovering and amplifying item relation signals for superior item representation learning

2. We propose an efficient algorithm to create and sample diverse UHI sequences for data augmentation that preserves both local and global contextual information

3. We propose an approach for next-item recommendation that significantly outperforms the majority of state-of-the-art (SOTA) baselines on seven public datasets

Model: Overview



Model: Data Augmentation (SRP)

Objective:
Amplify item co-occurrence and interactional context signals while preserving local and global context.

Approach:

- Augment UHI from training dataset by sampling from strongly restricted permutations generated from the training data
- We consider permutations where each item in the UHI can be displaced up to k positions to the left or right of its original position

Our model is inspired by the distributional hypothesis in NLP as well as recent work in interactional context

Model: Data Augmentation (SRP)

Algorithm 1 Random Transposition Walk

Require: sequence S, displacement d, number of permutations to return $\#p$, number of candidate permutations to generate $\#c$

```
n ← |S|
A ← 0_{nxn}

for i ∈ {1, ..., n} do
    for j ∈ {1, ..., n} do
        if | i − j | ≤ d then
            a_{ij} ← 1
        end if
    end for
end for

Out ← {}
Candidates ← {}
C ← S

while | Candidates | ≤ #c do
    i_1, i_2 ← SAMPLE ( [1 ··· n] , 2)

    if σ(i_1) ∈ S_A AND σ(i_2) ∈ S_A then
        C ← SWAP (C, i_1, i_2)
        Candidates ← Candidates ∪ C
    end if
end while

Out ← SAMPLE ( Candidates , #p)
```

Model: Item Representation Learning

Objective:
Deepen the contextualized embeddings extracted from a transformer based encoder

Approach

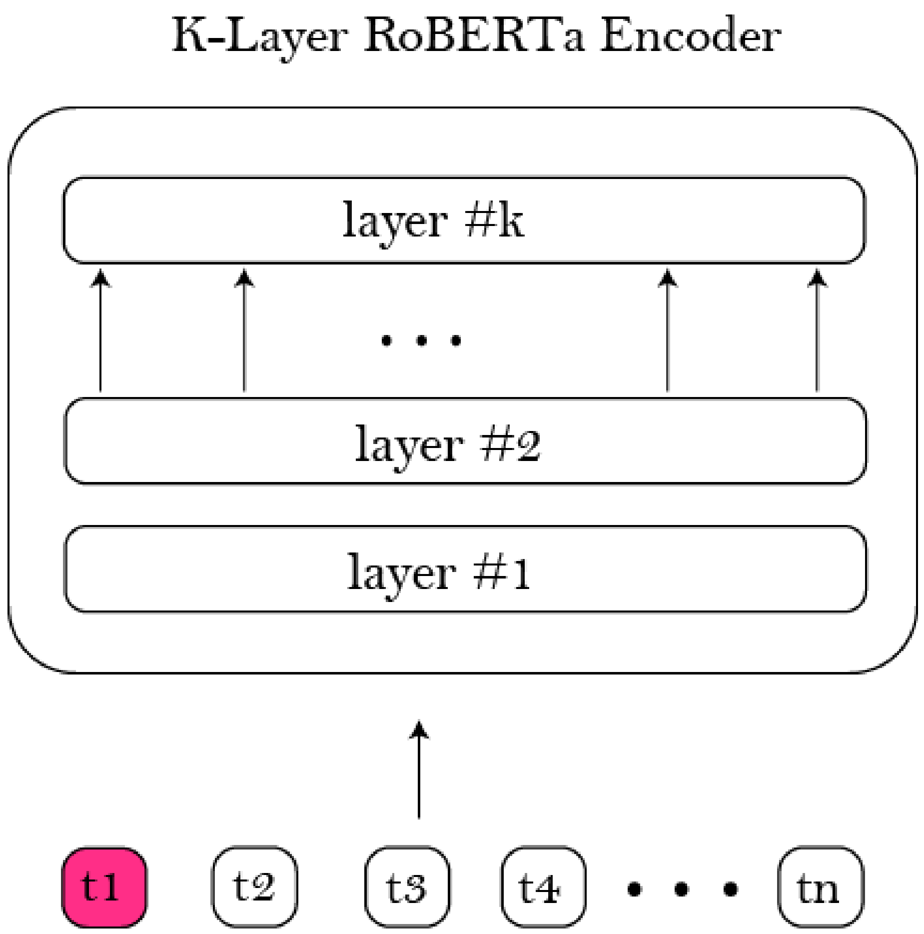
Given a sequence of n tokens (UHI) $T = \{t_1, \cdots t_n\}$,

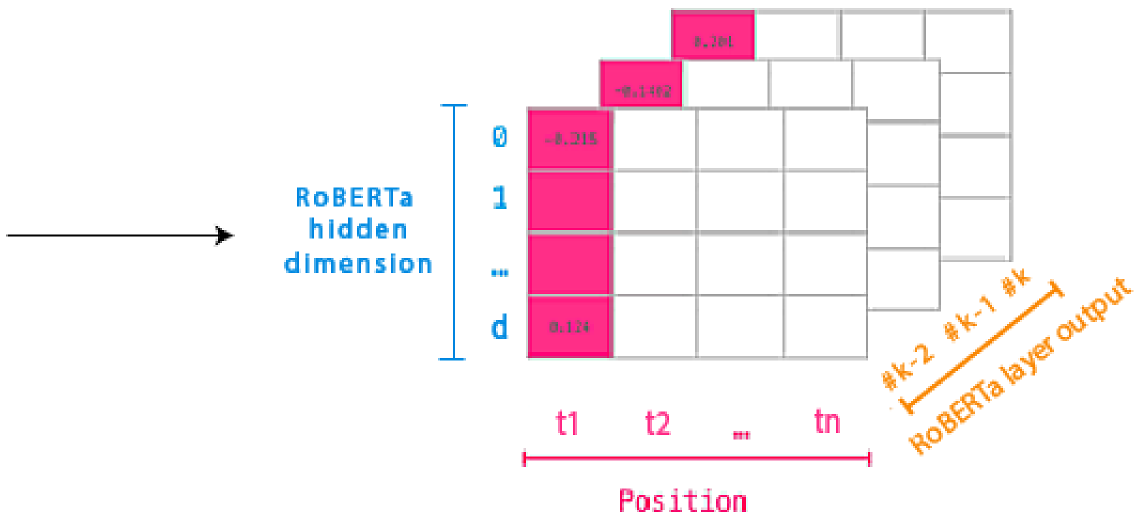
and a K-layer transformer encoder which outputs $O \in \mathbb{R}^{n\,*\,d}$ at each layer,

we obtain a final representation $R \in \mathbb{R}^{3\,*\,n\,*\,d}$ of the sequence by stacking the outputs of the final three layers of the encoder.

where d is the embedding dimension of each encoder layer

Model: Item Representation Learning

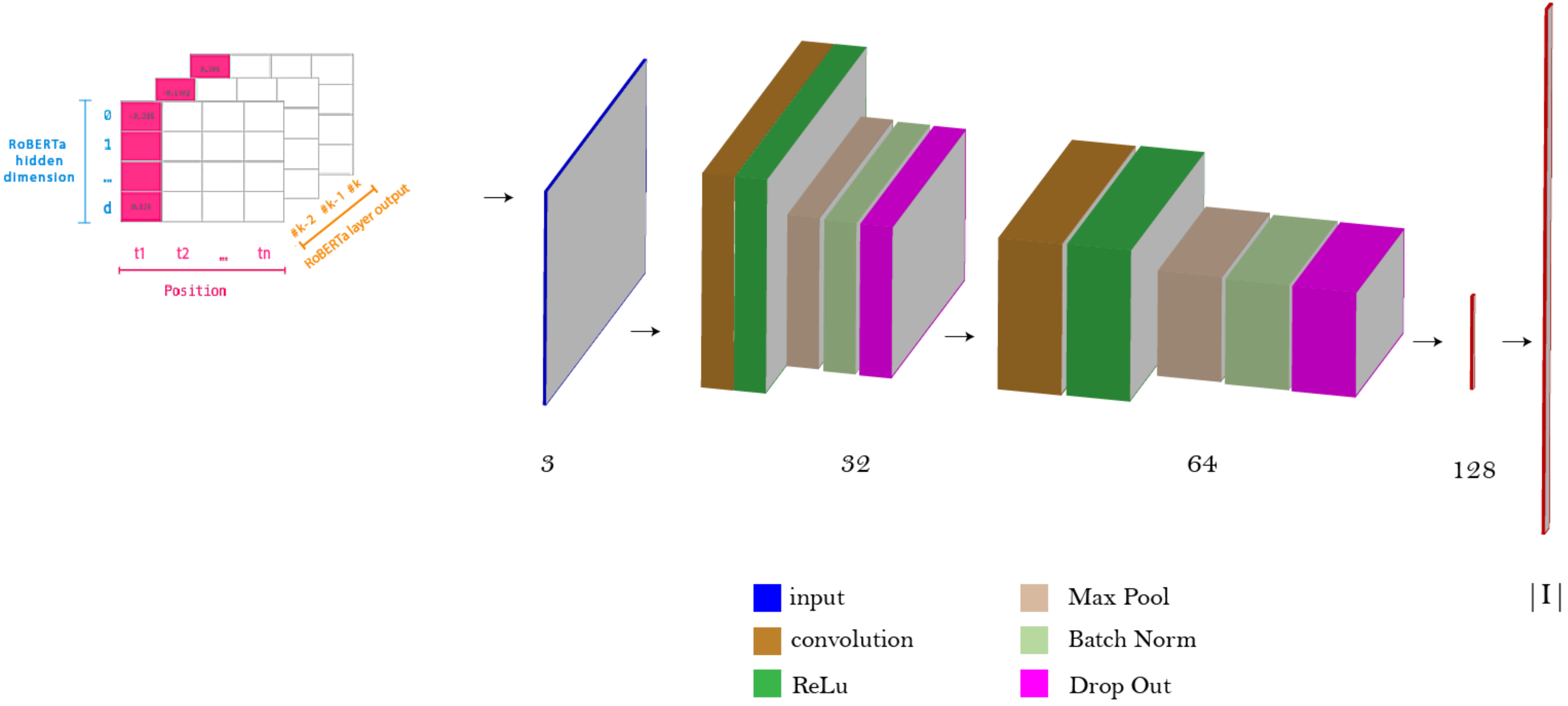




Model: CNN for Sequential Recommendation

- The input to the CNN is the R^{3*n*d} representation obtained from the Transformer encoder
- The output of each of the three final layers which are treated as separate channels of the input.
- The output of the CNN is a compact 128 vector
- A detailed description of the CNN architecture is presented next

Model: CNN for Sequential Recommendation



Experiments: Datasets

Table 1: Dataset statistics (after pre-processing)

Dataset	#users	#items	#interactions
ML25M	130,004	7,984	8,895,228
4SQUARE	227,301	14,153	7,183,758
YELP	478,630	7,239	2,714,935
LASTFM-1k	414,854	13,430	5,242,483
30M	22,162	2,026	243,428
NOWPLAYING	19,558	1,726	173,465
AOTM	12,670	1,094	83,571

YELP: A subset of Yelp's businesses, reviews, and user data across 8 metropolitan areas in the USA and Canada

4SQUARE: Global check-in data from Foursquare. It contains check-ins users on cities across 77 countries.

MovieLens 25M: User-Movie interaction activity from a movie recommendation service

LASTFM 1K: Whole listening habits for nearly 1,000 users.

AOTM: The AOTM dataset was collected from the Art-of-the-Mix platform and is publicly available.

30MUSIC: is a collection of listening and playlists data retrieved from Internet radio stations through Last.fm API.

NOWPLAYING: Music listening behavior of users created from mining information from social networks

Research Questions

- **RQ1** How does COMBERT perform on sequential recommendation compared to existing SOTA models?
- **RQ2** What is the contribution of the proposed SRP data augmentation approach to the overall performance of COMBERT?
- **RQ3** Can existing SOTA models for sequential recommendation benefit from our proposed data augmentation approach?

RQ1: COMBERT compared to SOTA SR models?

- COMBERT, our proposed model outperforms existing baselines on all seven datasets

- The average performance improvement across all the baselines is 94% for MRR and 80% for Recall.

Table 2: MRR@20 baseline comparison

	GRU4REC	CASER	BERT4REC	SASREC
30M	65.03%	38.01%	77.44%	126.92%
NOWP	28.80%	22.80%	74.26%	72.99%
AOTM	100.00%	61.54%	110.00%	110.00%
YELP	153.33%	90.00%	245.45%	216.67%
LASTFM-1K	100.00%	21.02%	104.30%	216.67%
ML25M	91.30%	51.72%	120.00%	120.00%
4SQUARE	45.89%	18.33%	74.59%	74.59%
Average	**83.48%**	**43.35%**	**115.15%**	**133.98%**

Table 3: RECALL@20 baseline comparison

	GRU4REC	CASER	BERT4REC	SASREC
30M	64.39%	38.91%	115.57%	159.66%
NOWP	38.17%	33.81%	134.67%	119.25%
AOTM	69.14%	48.91%	98.55%	77.92%
YELP	101.54%	63.75%	142.59%	147.17%
LASTFM-1K	88.17%	20.27%	86.17%	196.61%
ML25M	63.83%	43.93%	79.07%	81.18%
4SQUARE	27.23%	11.47%	47.72%	52.35%
Average	**64.64%**	**37.29%**	**100.62%**	**119.16%**

RQ2: contribution of SRP to the overall performance of COMBERT

Table 5: Results for **MRR@20**. Bold face indicates the best results, underline indicates runner up and ties are marked with *.

		30M	NOWP	AOTM	YELP	LASTFM-1K	ML25M	4SQUARE
GRU4REC	w/o SRP	0.143	0.184	0.021	0.015	0.095	0.023	0.146
	w/ SRP	0.169	0.186	0.022	0.022	0.118	0.024	0.143
	% Δ	18.18%	1.09%	4.76%	46.67%	24.21%	4.35%	-2.05%
CASER	w/o SRP	0.171	0.193	<u>0.026</u>	0.02	0.157	0.029	<u>0.18</u>
	w/ SRP	0.141	**0.235***	0.023	0.018	0.167	<u>0.033</u>	0.176
	% Δ	-17.54%	21.76%	-11.54%	-10.00%	6.37%	13.79%	-2.22%
BERT4REC	w/o SRP	0.133	0.136	0.02	0.011	0.093	0.02	0.122
	w/ SRP	0.162	0.184	0.021	<u>0.027</u>	**0.189***	0.023	0.166
	% Δ	21.80%	35.29%	5.00%	145.45%	103.23%	15.00%	36.07%
SASREC	w/o SRP	0.104	0.137	0.02	0.012	0.06	0.02	0.122
	w/ SRP	<u>0.175</u>	0.18	0.019	0.018	0.169	0.021	0.167
	% Δ	68.27%	31.39%	-5.00%	50.00%	181.67%	5.00%	36.89%
COMBERT	w/o SRP	0.19	0.188	0.021	0.015	0.138	0.038	**0.214***
	w/ SRP	**0.236**	**0.237***	**0.042**	**0.038**	**0.19***	**0.044**	**0.213***
	% Δ	24.21%	26.06%	100.00%	153.33%	37.68%	15.79%	-0.47%

RQ2: contribution of SRP to the overall performance of COMBERT

Table 6: Results for **RECALL@20**. Bold face indicates the best results, underline indicates runner up and ties are marked with *.

		30M	NOWP	AOTM	YELP	LASTFM-1K	ML25M	4SQUARE
GRU4REC	w/o SRP	0.278	0.338	0.081	0.065	0.186	0.094	0.382
	w/ SRP	0.326	0.351	0.079	0.085	0.233	0.092	0.378
	% Δ	17.27%	3.85%	-2.47%	30.77%	25.27%	-2.13%	-1.05%
CASER	w/o SRP	<u>0.329</u>	0.349	<u>0.092</u>	0.08	0.291	0.107	<u>0.436</u>
	w/ SRP	0.282	<u>0.427</u>	0.091	0.072	0.317	<u>0.12</u>	0.426
	% Δ	-14.29%	22.35%	-1.09%	-10.00%	8.93%	12.15%	-2.29%
BERT4REC	w/o SRP	0.212	0.199	0.069	0.054	0.188	0.086	0.329
	w/ SRP	0.301	0.333	0.081	<u>0.097</u>	<u>0.341</u>	0.091	0.398
	% Δ	41.98%	67.34%	17.39%	79.63%	81.38%	5.81%	20.97%
SASREC	w/o SRP	0.176	0.213	0.077	0.053	0.118	0.085	0.319
	w/ SRP	0.326	0.34	0.078	0.077	0.317	0.089	0.394
	% Δ	85.23%	59.62%	1.30%	45.28%	168.64%	4.71%	23.51%
COMBERT	w/o SRP	0.381	0.376	0.074	0.068	0.315	0.141	0.459
	w/ SRP	**0.457**	**0.467**	**0.137**	**0.131**	**0.35**	**0.154**	**0.486**
	% Δ	19.95%	24.20%	85.14%	92.65%	11.11%	9.22%	5.88%

RQ3: contribution of SRP to SOTA SR models

Table 4: average baseline improvement with SRP

	GRU4REC	CASER	BERT4REC	SASREC
MRR@20	13.89%	0.09%	51.69%	52.60%
RECALL@20	10.22%	2.25%	44.93%	55.47%

Conclusion

- In this work, we proposed a two-stage technique to recover and amplify item relation signals which are compromised in training data for SR
  - we recover compromised signals with a novel data augmentation process that involves generating synthetic UHI from observed UHI samples
  - In the second stage we employ a self-supervised pre-training approach that leverages the recovered item relation signals to learn superior item representations.
- We develop a novel CNN-based SR model as a downstream task to utilize our learned item representations.
- We demonstrate that our approach is able to recover item relational dynamics distorted by the candidate generation process
- Our proposed SR model significantly outperforms four SOTA models for SR and an ablation study confirms that our item representation learning approach is a significant contributor to the model's superior performance.
- we find that the majority of our compared SOTA baselines also enjoy a significant performance boost from utilizing our learned item representations

Future Work

- Extend this work to explore the properties of item relation signals that are recovered/recoverable.
- Explore neural approaches to optimize item relation signal recovery.

References

1. Amankwata, B., Fletcher, K.K.: Contexts embedding for sequential service recommendation. In: 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC). pp. 1087–1092. IEEE (2022)
2. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learningto align and translate. CoRR abs/1409.0473 (2015)
3. Blumberg, O.: Permutations with interval restrictions. Ph.D. thesis, PhD thesis, Stanford University (2012)
4. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.E.: A simple framework for contrastive learning of visual representations. ArXiv abs/2002.05709 (2020)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: NAACL (2019)
6. Diaconis, P., Graham, R.L., Holmes, S.P.: Statistical problems involving permutations with restricted positions (1999)
7. Grbovic, M., Radosavljevic, V., Djuric, N., Bhamidipati, N.L., Savla, J., Bhagwan, V., Sharp, D.: E-commerce in your inbox: Product recommendations at scale. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2015)
8. Harris, Z.S.: Distributional structure (1954)
9. He, R., McAuley, J.: Fusing similarity models with markov chains for sparse sequential recommendation. 2016 IEEE 16th International Conference on Data Mining (ICDM) pp. 191–200 (2016)
10. Hidasi, B., Karatzoglou, A., Baltrunas, L., Tikk, D.: Session-based recommendations with recurrent neural networks. CoRR abs/1511.06939 (2016)

References

Hidasi, B., Quadrana, M., Karatzoglou, A., Tikk, D.: Parallel recurrent neural network architectures for feature-rich session-based recommendations. Proceedings of the 10th ACM Conference on Recommender Systems (2016)
12. Kalchbrenner, N., Espeholt, L., Simonyan, K., van den Oord, A., Graves,
A., Kavukcuoglu, K.: Neural machine translation in linear time. ArXiv abs/1610.10099 (2016)
13. Kang, W.C., McAuley, J.: Self-attentive sequential recommendation. 2018 IEEE International Conference on Data Mining (ICDM) pp. 197–206 (2018)
14. Lastfm: Last.fm dataset - 1k users. http://ocelma.net/ MusicRecommendationDataset/lastfm-1K.html, accessed: 2020-12-30
15. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M.,Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. ArXiv abs/1907.11692 (2019)
16. Liu, Z., Chen, Y.G., Li, J., Yu, P.S., McAuley, J., Xiong, C.: Contrastiveself-supervised sequential recommendation with robust augmentation. ArXiv abs/2108.06479 (2021)
17. McFee, B., Lanckriet, G.: The natural language of playlists. In: ISMIR (2011)
18. Miaschi, A., Dell'Orletta, F.: Contextual and non-contextual word embeddings: an in-depth linguistic investigation. In: REPL4NLP (2020)
19. Mikolov, T., Chen, K., Corrado, G.S., Dean, J.: Efficient estimation of word representations
in vector space. In: ICLR (2013) 20. MovieLense: Movielens 25m dataset. https://grouplens.org/datasets/ movielens/25m/, accessed: 2020-12-30
21. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation.In: EMNLP (2014)
22. Peters, M.E., Neumann, M., Zettlemoyer, L., tau Yih, W.: Dissecting contextual word embeddings: Architecture and representation. In: EMNLP (2018)
23. Radford, A., J´ozefowicz, R., Sutskever, I.: Learning to generate reviews and discoveringsentiment. ArXiv abs/1704.01444 (2017)
24. Rendle, S., Freudenthaler, C., Schmidt-Thieme, L.: Factorizing personalized markov chains for next-basket recommendation. In: WWW '10 (2010)

References

25. Sahlgren, M., Holst, A., Kanerva, P.: Permutations as a means to encode order in word space (2008)
26. Schuster, M., Nakajima, K.: Japanese and korean voice search. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) pp.5149–5152 (2012)
27. Sennrich, R., Haddow, B., Birch, A.: Neural machine translation of rare words with subword units. ArXiv abs/1508.07909 (2016)
28. Song, Y., Lee, J.: Augmenting recurrent neural networks with high-order usercontextual
preference for session-based recommendation. ArXiv abs/1805.02983(2018)
29. Song, Y., Wang, J., Liang, Z., Liu, Z., Jiang, T.: Utilizing bert intermediate layers for aspect based sentiment analysis and natural language inference. ArXiv abs/2002.04815 (2020)
30. Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., Jiang, P.: Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. Proceedings of the 28th ACM International Conference on Information and Knowledge Management (2019)
31. Tang, J., Wang, K.: Personalized top-n sequential recommendation via convolutional sequence embedding. Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (2018)
32. Teufel, S.: An Overview of Evaluation Methods in TREC Ad Hoc Information Retrievaland TREC Question Answering, pp. 163–186. Springer Netherlands, Dordrecht
(2007). https://doi.org/10.1007/978-1-4020-5817-2 633. Tian, Y., Sun, C., Poole, B., Krishnan, D., Schmid, C., Isola, P.: What makes for good views for contrastive learning. ArXiv abs/2005.10243 (2020)
34. Turrin, R., Quadrana, M., Condorelli, A., Pagano, R., Cremonesi, P.: 30music listening and playlists dataset. In: RecSys Posters (2015)
35. Vasile, F., Smirnova, E., Conneau, A.: Meta-prod2vec: Product embeddings using side-information for recommendation. Proceedings of the 10th ACM Conferenceon Recommender Systems (2016)
36. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser,L., Polosukhin, I.: Attention is all you need. In: NIPS (2017)

References

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: NIPS (2017)

37. Vedaldi, A., Lenc, K.: Matconvnet: Convolutional neural networks for matlab. Proceedings of the 23rd ACM international conference on Multimedia (2015)

38. Wu, J., Wang, X., Feng, F., He, X., Chen, L., Lian, J., Xie, X.: Self-supervised graph learning for recommendation. Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (2021)

39. Wu, J.: Introduction to convolutional neural networks (2017)

40. Wu, J., Cai, R., Wang, H.: D´ej`a vu: A contextualized temporal attention mechanism for sequential recommendation. Proceedings of The Web Conference 2020 (2020)

41. Wu, S., Tang, Y., Zhu, Y., Wang, L., Xie, X., Tan, T.: Session-based recommendation with graph neural networks. In: AAAI (2019)

42. Xie, X., Sun, F., Liu, Z., Gao, J., Ding, B., Cui, B.: Contrastive pre-training for sequential recommendation. ArXiv abs/2010.14395 (2020)

43. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. ArXiv abs/1505.00853 (2015)

44. Yang, D., Zhang, D., Zheng, V., Yu, Z.: Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns. IEEE Transactions on Systems, Man, and Cybernetics: Systems 45, 129–142 (2015)

45. Yang, J., Zhao, H.: Deepening hidden representations from pre-trained language models for natural language understanding. ArXiv abs/1911.01940 (2019)

46. Yelp: The dataset. https://www.yelp.com/dataset, accessed: 2020-12-30

47. Zangerle, E., Pichl, M., Gassler, W., Specht, G.: nowplaying music dataset: Extracting listening behavior from twitter. In: WISMM ’14 (2014)

48. ZhangWei, ChenZeyuan, ZhaHongyuan, WangJianyong: Learning from substitutable and complementary relations for graph-based sequential product recommendation. ACM Transactions on Information Systems (2021)