

Simplicity Wins: Linear Beats Deep in Next-Basket Recommendation



Recurrent Autoregressive Linear Model for Next-Basket Recommendation

Tereza Zmeškalová¹, Antoine Ledent², Martin Spišák³,
Pavel Kordik^{1,3}, and Rodrigo Alves^{1,3}

Czech Technical University in Prague¹

Singapore Management University²

Recombee³

Introduction

Next-basket recommendation aims to predict the (sets of) items that a user is most likely to purchase during their next visit, capturing both short-term sequential patterns and long-term user preferences. However, effectively modeling these dynamics remains a challenge for traditional methods, which often struggle with interpretability and computational efficiency, particularly when dealing with intricate temporal dependencies and inter-item relationships. In this paper, we propose **ReALM**, a **Re**curent **A**utoregressive **L**inear **M**odel that explicitly captures temporal item-to-item dependencies across multiple time steps. By leveraging a recurrent loss function and a closed-form optimization solution, our

approach offers both interpretability and scalability while maintaining competitive accuracy. Our main contributions are:

1. We introduce three ReALM variants: temporal, sequential, and hybrid, each tailored to different user behavior dynamics.
1. We evaluate ReALM against seven baselines on two real-world datasets. Crucially, ReALM matches state-of-the-art accuracy with **substantially lower** resource requirements.
1. We provide an interpretability analysis showing that ReALM recovers complex item relationships and temporal patterns.

Recurrent Behavioral Trends in NBR

Consider the example in Figure 1, which shows four users' shopping baskets over four periods (e.g., weeks or months) in a supermarket offering eight items. It highlights distinct behavioral patterns: fruit is purchased weekly due to perishability (users 1 and 4); milk is typically bought biweekly (users 3

and 4), but weekly when paired with cereals (user 2), revealing an inter-item dependency; diaper sizes progress from S to M to L (users 1 and 2), reflecting life-stage transitions; and users 1 and 3 alternate between granola bars and protein shakes, indicating dietary variety.

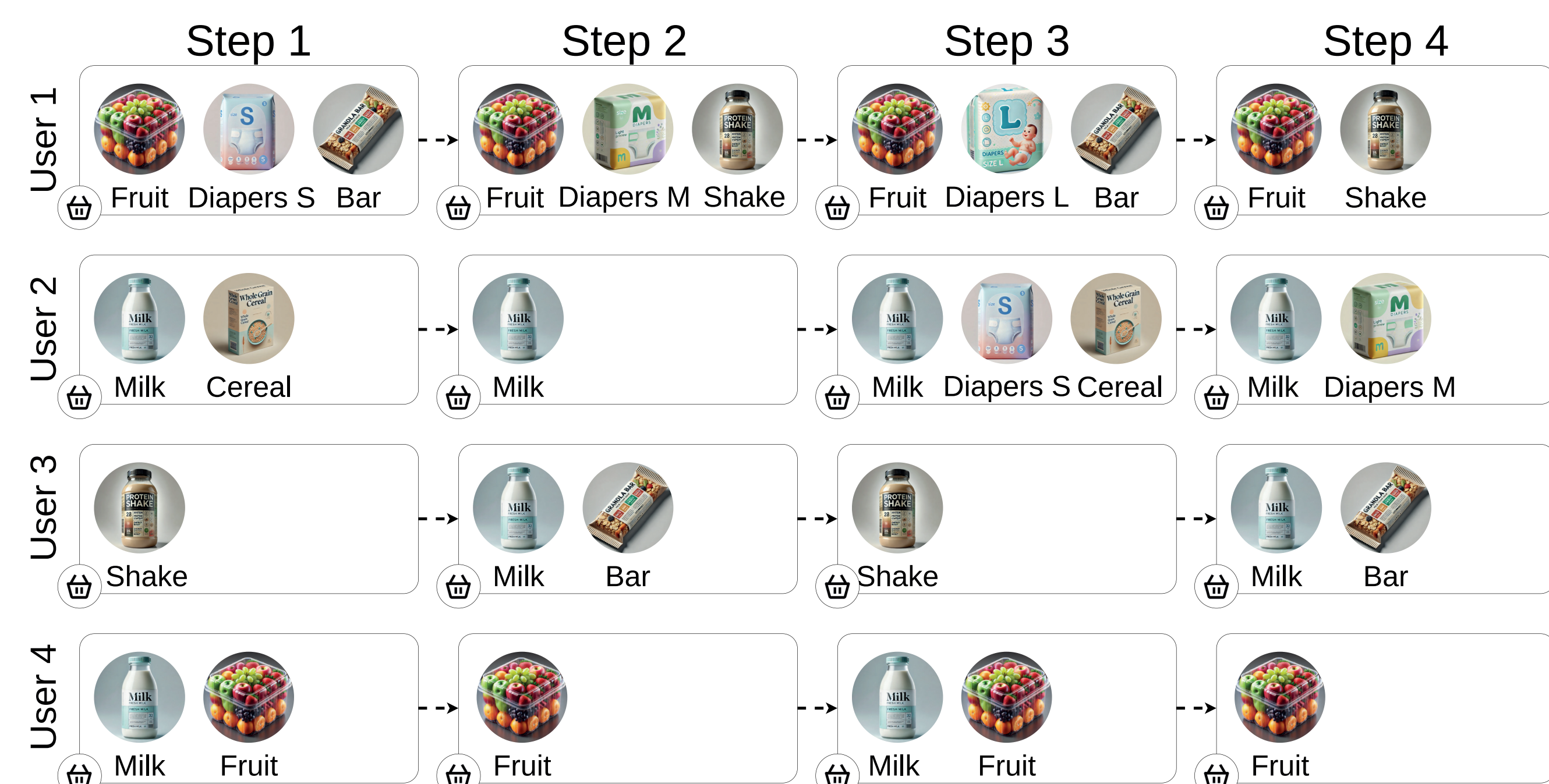


Figure 1 - Example of purchasing patterns in a grocery store.

ReALM

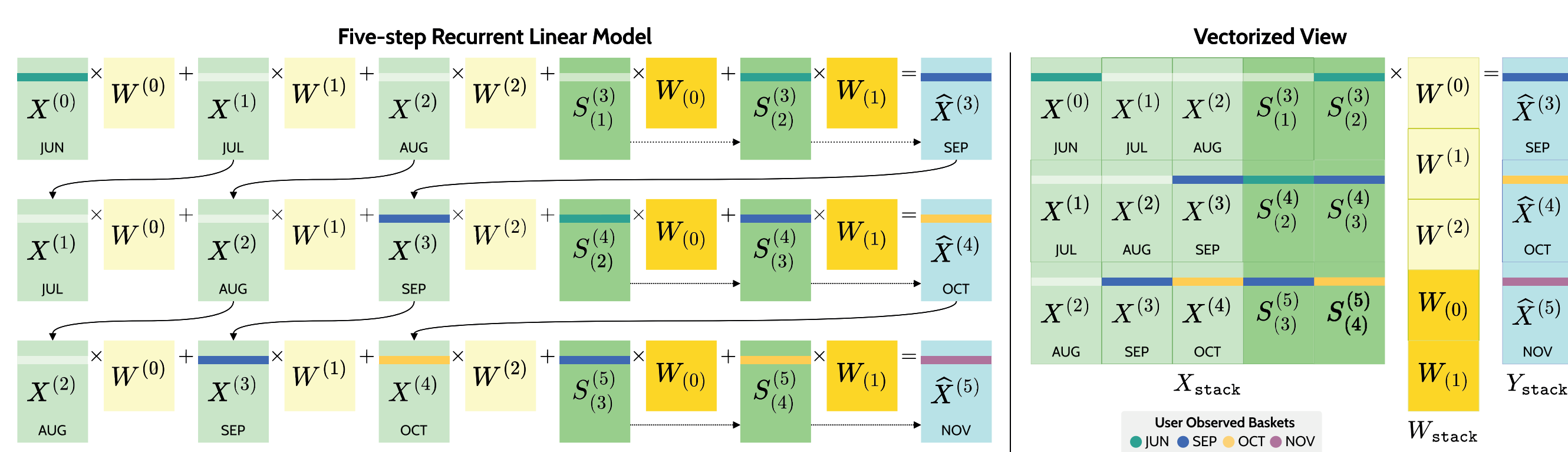


Figure 2 - Architecture of the five-step recurrent linear model (left) and its vectorized representation (right).

Formally, ReALM optimizes the following loss function:

$$\mathcal{L}(\theta) = \sum_{t=T-L}^{T-1} \left\| X^{(t)} - \left(\sum_{l=0}^{L-1} X^{(t-L+l)} W^{(l)} + \sum_{k=0}^{K-1} S_{(t-K+k)}^{(t)} W_{(k)} \right) \right\|_F^2.$$

Alternating Least Squares: It can be optimized using Alternating Least Squares (ALS), where each step has the following closed-form solution:

$$\widehat{W}^{(i)} = \left(\lambda I + \sum_{t=T-L}^{T-1} (X^{(t-L+i)})^\top X^{(t-L+i)} \right)^{-1} \sum_{t=T-L}^{T-1} (X^{(t-L+i)})^\top \mathbb{X}^{(i,t)},$$

$$\text{where } \mathbb{X}^{(i,t)} = X^{(t)} - \sum_{l=0, l \neq i}^{L-1} X^{(t-L+l)} W^{(l)} - \sum_{k=0}^{K-1} S_{(t-K+k)}^{(t)} W_{(k)},$$

Approx. Closed-form

- 1: **Input** $X_{\text{stack}}, Y_{\text{stack}}, \ell_2$ -regularization λ
- 2: $G \leftarrow X_{\text{stack}}^\top X_{\text{stack}} + \lambda I$
- 3: Compute fill-in reducing permutation P for matrix G
- 4: Compute sparse $\text{LDL}^\top \approx \text{PGP}^\top$
- 5: Compute sparse $K \approx L^{-1}$
- 6: $F_1 \leftarrow \text{KP}$
- 7: $F_2 \leftarrow D^{-1} F_1$
- 8: **Return** $\widehat{W}_{\text{stack}} = F_1^\top F_2 X_{\text{stack}}^\top Y_{\text{stack}} \approx \widehat{W}_{\text{stack}}$.

Performance Evaluation

We compare the performance of our method, ReALM, against seven baselines. ReALM achieves high recommendation accuracy, ranking as the most accurate model on the TaFeng dataset and second on Recombee-NBR1, nearly matching the top performer, DNNTSP. However, ReALM's sparse approximate closed-form solution reaches this performance orders of magnitude faster (**41 seconds** on a CPU versus **16 hours** on an A100 GPU), demonstrating efficiency unlocked by avoiding gradient descent.

Results

Dataset	Model	Recall@10	Recall@20	Recall@100	NDCG@10	NDCG@20	NDCG@100	Time
TaFeng	FPMC	0.0954	0.1275	0.2292	0.0911	0.1065	0.1401	400s
	TifuKNN	0.1057	0.1639	0.2782	0.1123	0.1303	0.1826	582s
	UP-CF	0.0706	0.0900	0.1483	0.0775	0.0878	0.1070	46s
	Sets2Sets	0.0891	0.1317	0.2220	0.0847	0.1053	0.1371	87000s ($\approx$ 24h)
	DNNTSP	0.0995	0.1447	0.1447	0.0952	0.1100	0.1088	8650s ($\approx$ 2h)
	Most Popular	0.0959	0.1259	0.2209	0.1324	0.1456	0.1772	N/A
	Last Basket	0.0513	0.0587	0.0609	0.0521	0.0573	0.0575	N/A
	ReALM-TS	0.1331	0.1651	0.2518	0.1664	0.1742	0.2026	695s
	ReALM-T	0.1053	0.1303	0.2163	0.1325	0.1372	0.1652	184s
	ReALM-S	0.1146	0.1438	0.2515	0.1481	0.1549	0.1902	433s
Recombee NBR1	FPMC	0.3110	0.4300	0.6869	0.2689	0.3207	0.4029	5300s ($\approx$ 1.5h)
	TifuKNN	0.4460	0.5513	0.7235	0.4309	0.4755	0.5296	721s
	UP-CF	0.2558	0.2987	0.4492	0.2709	0.2911	0.3379	826s
	Sets2Sets	0.4175	0.5213	0.6439	0.3948	0.4352	0.4752	153000s ($\approx$ 42h)
	DNNTSP	0.4673	0.5752	0.7311	0.4449	0.4932	0.5444	58220s ($\approx$ 16h)
	Most Popular	0.0686	0.1120	0.3965	0.0601	0.0785	0.1640	N/A
	Last Basket	0.2982	0.3092	0.3087	0.2890	0.2950	0.2951	N/A
	ReALM-TS	0.2814	0.3630	0.6263	0.2548	0.2865	0.3686	81s
	ReALM-T	0.4609	0.5600	0.7239	0.4473	0.4854	0.5379	41s
	ReALM-S	0.3415	0.4218	0.6349	0.3580	0.3870	0.4489	27s

Interpretability

We briefly discuss interpretability properties directly extracted from the weight matrices of ReALM. Specifically, a i, j entry of the i -th weight matrix encodes a score related to the likelihood of

including item- j in the next basket given that item- i was present at step- i . Figure 3 shows bar plots illustrating different behavioral patterns, focusing on the ReALM-T variant on the TaFeng dataset.

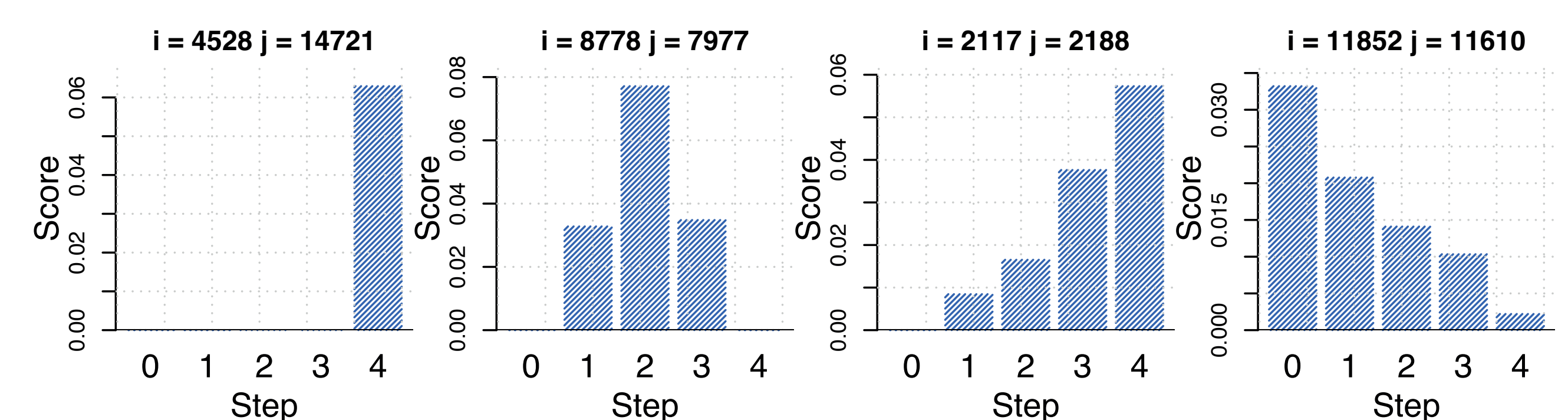


Figure 3: Bar plots presenting interpretability of ReALM-T.

Conclusion and Future Work

- We introduced ReALM, a family of NBR models that combines accuracy, efficiency, and interpretability.
- ReALM captures short- and long-term patterns through a recurrent loss and uses an approximate closed-form solution to accelerate training.
- On two real-world datasets, ReALM matches deep-learning baselines in accuracy while training much faster, scaling linearly, and enabling interpretable sequential insights.
- Future work could incorporate context (e.g., location, device) and content-based features.



Paper
(ACM Digital Library)



Repository
(Dataset, Codes For Reproducibility)