Revisiting CORE for Session-based Recommendation: Dual-Attention and Hard-Negative Extensions

Liran Smadja, Itay Mizikov
Ben-Gurion University, Beer-Sheva, Israel
{smadjali, itaymizi}@post.bgu.ac.il

Project Repository: https://github.com/Itaymizik/RecSys---CORE

Abstract

Session-based recommendation addresses the problem of predicting the next item a user will interact with from the short sequence of actions observed in an ongoing session, often without reliable long-term user profiles. This setting is practically important because many recommendations must be delivered to anonymous or first-time users, and because user intent can shift across sessions. Recent surveys position session-based recommender systems as a distinct paradigm that emphasizes short-term yet dynamic preference modeling and timely adaptation to evolving session context.

This work is motivated by a specific modeling tension highlighted by CORE: when a session embedding is produced by a non-linear encoder, it may not lie in the same representation space as item embeddings, and this mismatch can lead to inconsistent scoring behavior during decoding. CORE proposes to unify the representation space for encoding and decoding via a representation-consistent encoder based on linear combination, together with a robust distance measuring strategy designed to reduce overfitting in a consistent representation space.

Building on these foundations, we revisit CORE from the perspective of representation consistency and propose two extensions that preserve its geometric principle. First, we introduce a Dual-Attention encoder with adaptive fusion of global session context and recent intent while keeping the session representation as a linear combination of item embeddings. Second, we propose a hard-negative-aware hybrid objective that combines cross-entropy with a pairwise ranking term focused on hard negatives. The remainder of this paper presents the background, formalizes these extensions, and outlines an empirical evaluation under the CORE protocol.

1 Introduction

Research Topic: Session-based Recommendation Systems

Recommender systems are widely used to mitigate information overload by helping users discover relevant content. In many real-world deployments, however, personalization cannot rely on a stable user profile: the platform may not know who the user is, the user may be new, or long-term histories may be inaccessible due to privacy, business constraints, or system design. Session-based recommendation focuses on this regime by tailoring suggestions to short-term needs inferred from an ongoing session typically a short sequence of interactions. This scenario is both practically relevant and technically challenging due to shifting intent and limited within-session evidence [1].

Recent surveys characterize session-based recommender systems as a specialized paradigm that captures short-term yet dynamic user preferences [2]. Modern approaches rely on learned representations and neural architectures including sequential networks and graph neural networks to encode sessions and score candidate items [3].

Within this landscape, CORE addresses a key geometric issue: when session embeddings are produced by non-linear encoders, they may not lie in the same representation space as item embeddings. Since decoding scores items by comparing session and item embeddings, this mismatch can manifest as an "inconsistent prediction" issue [4]. CORE resolves this by unifying the representation space through a representation-consistent encoder (forming session embeddings as linear combinations of item embeddings) and a robust distance measuring approach to prevent overfitting [4].

Building on CORE, we propose two extensions that preserve its geometric principle. First, we introduce a Dual-Attention encoder with adaptive fusion of global session context and recent intent while keeping the session representation as a linear combination of item embeddings. Second, we propose a hard-negative-aware hybrid training objective that combines cross-entropy with a pairwise ranking term focused on confusing items. Together, these improvements aim to enhance CORE's ranking quality while maintaining its representation consistency.

2 Related Work

The session-based recommendation problem has been studied for years and has accumulated a diverse set of approaches. A core thread across the literature is the attempt to infer a user's immediate intent from sparse within-session signals. The Recommender Systems Handbook chapter on session-based recommenders motivates the problem with practical scenarios and highlights why short sessions and absent long-term preference signals matter: recommendations must often be generated for anonymous or first-time users, and intent can differ substantially across sessions [1]. This framing also surfaces evaluation considerations and open challenges that remain active in the community.

Comprehensive surveys describe the evolution from earlier pattern-based and heuristic approaches to modern neural solutions. The survey by Wang et al. emphasizes that session-based recommender systems constitute a distinct paradigm, differing from content-based and collaborative filtering settings that typically model longer-term, relatively static preferences. In contrast, session-based models aim to capture short-term but dynamic preferences and adapt to changing session context [2]. This survey also argues for clearer problem statements and careful articulation of session-based characteristics and challenges, providing a useful reference point for positioning CORE-style methods within the broader field.

More recent survey work focusing on neural architectures organizes session-based methods around how they represent and propagate dependencies among interactions. One recurring distinction is between sequential models, which treat the session primarily as an ordered sequence, and graph-based models, which represent the session (or its transitions) as a graph to capture richer relational signals beyond strict adjacency. The survey by Li et al. highlights that session-based recommendation specializes in capturing short-term preferences and enabling timely recommendations, and it uses this lens to organize and compare neural solutions across sequential and graph perspectives [3]. This organization is particularly relevant for CORE, which intentionally takes a minimalist stance on architecture while emphasizing representation-space consistency as the key design axis.

Within the family of neural session-based models, CORE is directly motivated by a representational mismatch: the encoder produces a session embedding and the decoder scores items by comparing that session embedding to item embeddings, yet typical non-linear encoders do not guarantee that these objects live in a consistent space. CORE proposes to make this consistency an explicit objective by construction [4]. Our work is positioned in this neighborhood: we treat CORE as a strong and simple baseline, and we use the surveys and handbook chapter above to ground both the problem definition and the design space in which CORE-like improvements can be explored.

3 Background

3.1 Problem Setting

We adopt the standard session-based recommendation setting. Let $\mathcal{V}$ denote the item set. A session s is an ordered sequence of interacted items,

$$s = [v_1, v_2, \ldots, v_m], \quad v_i \in \mathcal{V}, \tag{1}$$

where m is the session length. The goal is to predict the next interacted item v_{m+1} given the observed prefix $[v_1, \ldots, v_m]$. In many practical cases, user identity and long-term histories are not available or not reliable, and the model must infer short-term intent from the within-session interactions alone [1].

A common neural formulation introduces an item embedding table $\mathbf{E} \in R^{|\mathcal{V}| \times d}$, where each item v is mapped to an embedding $\mathbf{h}_v \in R^d$. Given a session, the model constructs a session representation $\mathbf{h}_s$ (encoding) and then scores candidate items (decoding) by comparing $\mathbf{h}_s$ with candidate item embeddings.

3.2 Representation Consistency and the CORE Motivation

A core observation in CORE is that the encoding-decoding interface can be geometrically inconsistent. Many session encoders are non-linear (for example, recurrent encoders), and thus the resulting session embedding is not guaranteed to lie in the same linear space spanned by item embeddings. Yet decoding frequently relies on direct similarity computations between $\mathbf{h}_s$ and item embeddings. CORE argues that this mismatch can lead to an "inconsistent prediction" issue: even when sessions share an objective, non-linear encoding can map them to different regions of space, producing unstable or inconsistent similarity behavior with respect to item embeddings [4].

CORE addresses this by unifying the representation space for encoding and decoding, with two key components: a representation-consistent encoder (RCE) and a robust distance measuring (RDM) strategy for decoding [4]. The emphasis is not on a more complex encoder architecture, but rather on ensuring that the scoring space is coherent by construction.

3.3 CORE: Representation-Consistent Encoder

Given a session $s = [v_1, \ldots, v_m]$, CORE first obtains item embeddings $\mathbf{h}_{s,i}$ for each interaction via the embedding table. The representation-consistent encoder defines the session embedding as a weighted linear combination of item embeddings:

$$\mathbf{h}_s = \sum_{i=1}^{m} \alpha_{s,i}\, \mathbf{h}_{s,i}. \tag{2}$$

The weights $\alpha_{s,i}$ reflect item importance within the session and are produced by a deep neural network operating on the sequence of item embeddings, but constrained so that $\mathbf{h}_s$ remains a linear combination in the original embedding space. CORE instantiates this idea with two simple variants [4]. In the mean-pooling variant, the encoder ignores order and importance and simply uses

$$\alpha_{s,i} = \frac{1}{m}, \tag{3}$$

so that the session representation is the average of its item embeddings. In the Transformer-based variant, CORE applies a stack of self-attention blocks over the sequence $[\mathbf{h}_{s,1}; \ldots; \mathbf{h}_{s,m}]$ to obtain contextualized token representations $\mathbf{F} \in R^{m \times d'}$, and then computes normalized importance weights via a learned vector $\mathbf{w} \in R^{d'}$:

$$\boldsymbol{\alpha} = \mathrm{softmax}(\mathbf{w}\, \mathbf{F}^\top). \tag{4}$$

In both variants, the resulting session embedding $\mathbf{h}_s$ is explicitly formed as a linear combination of item embeddings and therefore lies in the same representation space as the items. This consistency has a direct implication for common dot-product decoding: if one scores a target item v_t by $\mathbf{h}_s^\top \mathbf{h}_t$ while $\mathbf{h}_s$ is a weighted sum of item embeddings, then the score decomposes into a weighted sum of item-to-item dot products, which aligns session-to-item scoring with item-to-item similarity in the same space [4].

3.4 CORE: Robust Distance Measuring for Decoding

CORE further observes that simply operating in a consistent space can lead to overfitting of embeddings if the decoder is not regularized appropriately. To mitigate this, CORE revisits the widely used dot-product decoder from the perspective of tuplet loss optimization and introduces a controllable temperature margin together with dropout on candidate item embeddings [4]. Let $\mathbf{h}_s$ denote the session embedding and $\mathbf{h}_v$ the embedding of item v. CORE applies dropout to item embeddings, yielding $\mathbf{h}'_v$, and measures similarity via cosine similarity with temperature τ:

$$\ell = -\log \frac{\exp\big(\cos(\mathbf{h}_s, \mathbf{h}'_{v^+})/\tau\big)}{\sum_{v \in \mathcal{V}} \exp\big(\cos(\mathbf{h}_s, \mathbf{h}'_v)/\tau\big)}, \tag{5}$$

where v^+ is the ground-truth next item. This formulation replaces the fixed margin implicit in standard dot-product cross-entropy with a tunable temperature τ, and uses dropout directly on item embeddings to improve robustness in the consistent representation space. When embeddings are ℓ_2-normalized, cosine similarity and squared Euclidean distance are monotonically related, so this decoder can also be interpreted as operating on (negative) distances in the consistent space.

4 Methodology

4.1 Proposed Methodology

CORE highlights a geometric tension: decoding typically compares a session representation to item embeddings, yet common non-linear encoders do not guarantee that the session embedding lies in the same representation space as the item embeddings. CORE resolves this by constructing the session representation as a weighted linear combination of item embeddings (Eq. 2) and by using robust distance measuring strategies to reduce overfitting in the consistent space [4]. Our methodology follows the same design principle: we do not change the representational *form* of the session embedding (it remains a linear combination in the item space), but we modify how the importance weights are computed in the Transformer-based variant, and how training exposes the model to difficult negatives. Concretely, we build on CORE-trm [4] and introduce two complementary improvements:

Contribution 1: Dual-Attention with Adaptive Fusion. CORE-trm learns item importance weights using self-attention blocks and then applies a softmax-based weighting scheme. While this captures sequential patterns, it can under-emphasize a frequent phenomenon in sessions: the *last interactions* may reflect the most current short-term intent. We therefore compute two sets of importance weights, one capturing a global session context and one explicitly emphasizing the last-item-driven recent intent, and then fuse the resulting session representations using a learnable gate. This preserves representation consistency because the final session embedding is still a linear combination of item embeddings.

Contribution 2: Hard-Item Negatives with a Hybrid Objective. CORE and many session-based models are trained with a next-item classification objective that contrasts the positive next item against a large set of negatives. However, many negatives are trivially easy,

especially in large item catalogs. We therefore adopt *hard negative mining* by selecting the highest-scoring incorrect items under the current model, and we therefore adopt hard negative mining by selecting the highest-scoring incorrect items under the current model, and add a pairwise ranking term that explicitly encourages the positive item to score higher than these hard negatives. . This leaves inference unchanged (still based on similarity in the consistent space) but strengthens training signals on confusing items.

We denote our architecture-improved model as CORE-DA (Dual Attention), our training-improved model as CORE-HN (Hard Negatives), and the combined model as CORE-DAHN when both are enabled.

Relation to CORE-trm. Our architecture is a strict extension of CORE-trm. If we disable the recent-intent branch (using only the global attention weights), fix the gating mechanism to always select the global branch, and remove the residual mean-pooling path, the resulting encoder reduces to the Transformer-based representation-consistent encoder used in CORE-trm [4]. Similarly, if we set $\lambda = 0$ and do not mine hard negatives, the training objective reduces to the standard cross-entropy loss employed in CORE. Thus, CORE-DAHN recovers CORE-trm as a special case.

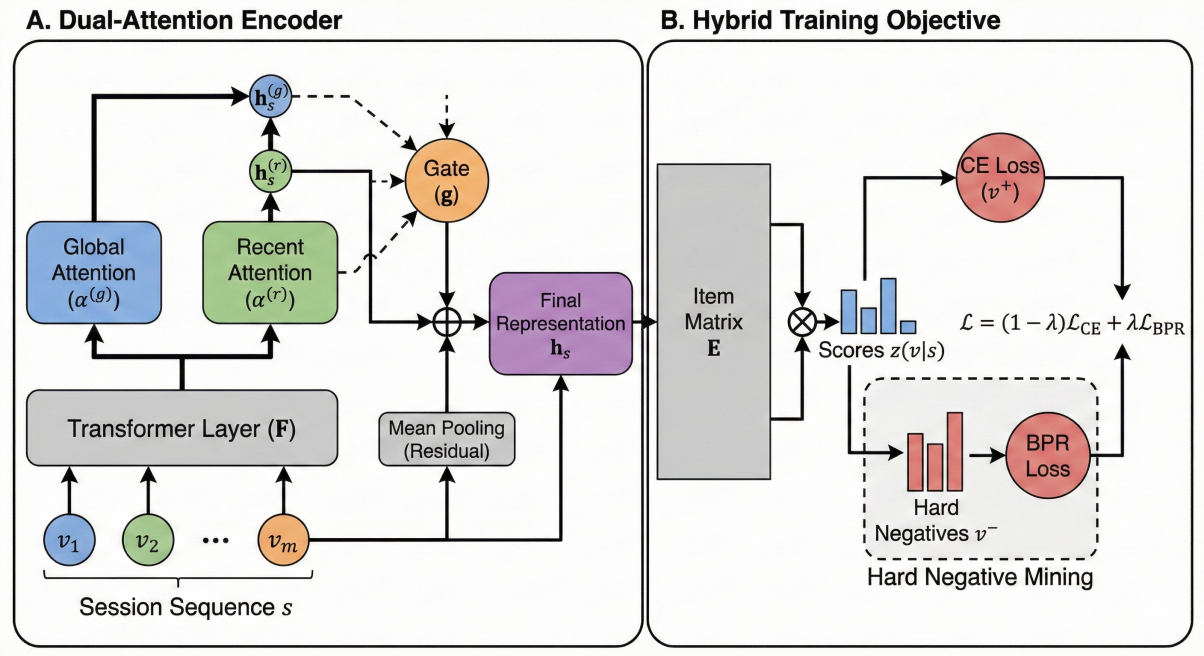

Figure 1: The overall architecture of CORE-DAHN. **(A) Dual-Attention Encoder:** Illustrates the parallel computation of global context ($\alpha^{(g)}$) and recent intent ($\alpha^{(r)}$), fused via an adaptive gate. **(B) Hybrid Training:** Shows the hard-negative mining process where the model is optimized using both Cross-Entropy on the positive item and BPR loss on hard negatives.

4.1.1 Dual-Attention Weighting in a Consistent Space

Let a session be represented by item embeddings $\mathbf{H}_s = [\mathbf{h}_{s,1}, \ldots, \mathbf{h}_{s,m}] \in R^{m \times d}$, where $\mathbf{h}_{s,i}$ is the embedding of item v_i. Following CORE-trm, we obtain contextualized representations $\mathbf{F} \in R^{m \times d'}$ using L Transformer self-attention blocks with positional encodings [4]. Instead of producing a single importance vector, we compute two attention score vectors:

Global-context scores. Following the CORE-trm aggregation style, we inject positional information explicitly when estimating global importance. Let $\mathbf{p}_i \in R^{d'}$ denote the positional

embedding of position i (or an equivalent position signal already used in the Transformer input). We compute global attention scores by an MLP over the concatenation of each contextualized token representation and its position signal:

$$e_i^{(g)} = \mathrm{MLP}_g\left([\mathbf{f}_i; \mathbf{p}_i]\right), \qquad \alpha_i^{(g)} = \mathrm{softmax}(e^{(g)})_i. \tag{6}$$

This branch aims to capture the overall session context, while remaining compatible with CORE's representation-consistent linear aggregation.

Recent-intent scores. To explicitly emphasize recency, we use the last valid token representation as a *query* signal. Let $\ell(s)$ be the last valid index in the padded sequence (i.e., the last interacted item in the session), and let $\mathbf{f}_{\ell(s)}$ denote its contextualized representation. We compute recent-intent scores by conditioning each token on this last-item query:

$$e_i^{(r)} = \mathrm{MLP}_r\left([\mathbf{f}_i; \mathbf{f}_{\ell(s)}]\right), \qquad \alpha_i^{(r)} = \mathrm{softmax}(e^{(r)})_i. \tag{7}$$

Intuitively, this branch prioritizes past interactions that best align with the most recent interaction, capturing late-session intent shifts without changing the representation-consistent form of the encoder.

Session representations and gating. Using the same (dropout-perturbed) item embeddings as CORE, we compute two session embeddings:

$$\mathbf{h}_s^{(g)} = \sum_{i=1}^{m} \alpha_i^{(g)}\, \hat{\mathbf{h}}_{s,i}, \qquad \mathbf{h}_s^{(r)} = \sum_{i=1}^{m} \alpha_i^{(r)}\, \hat{\mathbf{h}}_{s,i}. \tag{8}$$

We then fuse them by a learnable gate $\mathbf{g} \in R^2$:

$$\mathbf{g} = \mathrm{softmax}\Big(\mathrm{MLP}_{\mathrm{gate}}([\mathbf{h}_s^{(g)}; \mathbf{h}_s^{(r)}])\Big), \qquad \mathbf{h}_s^{(\mathrm{trm})} = g_1 \mathbf{h}_s^{(g)} + g_2 \mathbf{h}_s^{(r)}. \tag{9}$$

Finally, to preserve CORE's emphasis on stable consistent representations, we include the CORE-ave residual path and normalize:

$$\mathbf{h}_s^{(\mathrm{ave})} = \sum_{i=1}^{m} \alpha_i^{(\mathrm{ave})}\, \hat{\mathbf{h}}_{s,i}, \qquad \mathbf{h}_s = \mathrm{norm}\Big(\mathbf{h}_s^{(\mathrm{trm})} + \mathbf{h}_s^{(\mathrm{ave})}\Big), \tag{10}$$

where $\alpha^{(\mathrm{ave})}$ is uniform (mean pooling) as in CORE-ave [4]. This construction keeps $\mathbf{h}_s$ in the item embedding space by design.

4.1.2 Hard Negative Mining with a Hybrid CE+BPR Objective

Given the final session embedding $\mathbf{h}_s$ and item embedding matrix $\mathbf{E}$, we score all items by cosine-style similarity via normalized embeddings (equivalently a dot product after ℓ_2 normalization), scaled by a temperature τ:

$$z(v \mid s) = \frac{\mathbf{h}_s^\top \mathbf{h}_v}{\tau}, \qquad \mathbf{h}_v \in \mathbf{E}. \tag{11}$$

The standard next-item loss is the cross-entropy over the full catalog:

$$\mathcal{L}_{\mathrm{CE}} = -\log \frac{\exp(z(v^+ \mid s))}{\sum_{v \in \mathcal{V}} \exp(z(v \mid s))}, \tag{12}$$

where v^+ is the ground-truth next item.

Algorithm 1 Training Step for CORE-DAHN (Dual Attention + Hard Negatives)

Require: Batch of sessions $\{s_b\}$ with positives $\{v_b^+\}$, item embeddings $\mathbf{E}$, temperature τ, K, λ

Compute contextual token representations $\mathbf{F}_b \leftarrow \text{Transformer}(\mathbf{H}_{s_b})$

Compute $\alpha_b^{(g)}, \alpha_b^{(r)}$ and $\mathbf{h}_{s_b}^{(g)}, \mathbf{h}_{s_b}^{(r)}$

Fuse with gating to obtain $\mathbf{h}_{s_b}^{(\text{trm})}$; add CORE-ave residual to get $\mathbf{h}_{s_b}$

Compute logits $z(v \mid s_b) = \mathbf{h}_{s_b}^\top \mathbf{h}_v / \tau$

$\mathcal{L}_{\text{CE}} \leftarrow$ cross-entropy over $z(\cdot \mid s_b)$ and v_b^+

Mine hard negatives $\mathcal{N}_K(s_b) \leftarrow \text{TopK}(\{z(v \mid s_b) : v \neq v_b^+\})$

$\mathcal{L}_{\text{BPR}} \leftarrow \frac{1}{K} \sum_{v^- \in \mathcal{N}_K(s_b)} \log(1 + \exp(z(v^- \mid s_b) - z(v_b^+ \mid s_b)))$

return $\mathcal{L} = (1-\lambda)\mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{BPR}}$

To focus learning on confusing alternatives, we mine hard negatives by selecting the top-K highest-scoring *items* under the current model, excluding the ground-truth next item:

$$\mathcal{N}_K(s) = \underset{v \in \mathcal{V} \setminus \{v^+\}}{\text{argTopK}} \; z(v \mid s). \tag{13}$$

In practice, the TopK selection is treated as a non-differentiable operator (stop-gradient): we first compute scores under the current model and then use the selected indices as negatives for the pairwise term.

We then add a pairwise ranking term over these negatives:

$$\mathcal{L}_{\text{BPR}} = \frac{1}{K} \sum_{v^- \in \mathcal{N}_K(s)} \log\left(1 + \exp\left(z(v^- \mid s) - z(v^+ \mid s)\right)\right), \tag{14}$$

which encourages the positive item to score higher than mined hard negatives (a smooth separation objective rather than a hard margin).

4.2 A Minimal Running Example

Consider a session $s = [v_1, v_2, v_3]$ where v_3 is the most recent interaction. CORE-style encoding constructs $\mathbf{h}_s$ as a weighted sum of $\mathbf{h}_{s,1}, \mathbf{h}_{s,2}, \mathbf{h}_{s,3}$. In CORE-DA, the global attention may assign higher weight to v_2 if it best represents the session's overall context, while the recent-intent attention can place extra weight on v_3 to reflect a late-stage shift in intent. The gating module then adapts the mixture based on how consistent the two signals are: when global and recent views agree, the gate tends to average them; when the last item indicates a clear pivot, the gate can emphasize the recent view. During training with CORE-HN, if the model currently ranks an incorrect item u very highly for this session, u will be selected into $\mathcal{N}_K(s)$ and will contribute explicitly to the pairwise penalty in Eq. (**??**), pushing the model to rank v^+ above u under the same session embedding $\mathbf{h}_s$.

5 Empirical Evaluation

5.1 Research Questions and Hypotheses

Our empirical study is structured around three questions. First, we ask whether improving importance-weight estimation via dual attention yields gains over CORE-trm across standard benchmarks. Second, we ask whether hard-negative-aware training improves ranking quality beyond standard cross-entropy training, particularly by reducing confusions among high-scoring incorrect items. Third, we analyze whether the two improvements are complementary, i.e., whether combining them outperforms applying each in isolation.

5.2 Preparing Experimental Data

To ensure comparability with CORE, we follow the same evaluation protocol and public benchmarks as reported in the CORE study, which includes five datasets and a standard train/validation/test split ratio of 8:1:1 [4]. We adopt the same session construction conventions and filtering rules as described in the CORE reference implementation and paper, and we report results per dataset under the same top-K evaluation settings (e.g., $K = 20$ as in CORE).

5.3 References to the Open-Source Code

Our experiments build on the open-source CORE implementation released by the authors of CORE [4]. We implement CORE-DA, CORE-HN, and CORE-DAHN as modular extensions to the CORE-trm codepath.

5.4 Performance Evaluation Metrics

We focus on top-K ranking metrics standard in session-based recommendation. Because next-item prediction has a single ground-truth item per test session, Recall@K coincides with Hit Rate@K: it measures the fraction of test sessions for which the target item appears in the top-K recommendation list. We also report MRR@K (mean reciprocal rank), which additionally rewards placing the target item higher in the ranked list [3]. Following CORE, our primary metrics are Recall@20 and MRR@20 [4].

5.5 Benchmarking Methods and Data

We compare to CORE-ave and CORE-trm [4]. Our aim is not to re-argue the full landscape, but to position the proposed improvements relative to CORE and to the strong baselines already considered in CORE.

5.6 Experimental Setup

All configurations are trained under a fixed, reproducible setup with a single random seed (`seed` = 2020). Unless otherwise noted, we use 20 epochs with early stopping (`stopping_step` = 5), batch size 2048, and Adam with learning rate 10^{-3}. For CORE-HN and CORE-DAHN, we enable hard negatives with `hard_neg_k` = 5 and mixing coefficient $\lambda = 0.2$ (Eq. **??**). We evaluate once on the held-out test split using Recall@20 and MRR@20.

5.7 Statistical Tests for Model Evaluation

This draft reports single-run results (one seed per configuration). A complete statistical analysis (mean±std across multiple seeds, paired tests, and/or confidence intervals) is left to future work, and is particularly important because the observed differences are often within a few tenths of a percentage point.

5.8 Ablation Study

We isolate each contribution by comparing:
(i) CORE-DA against CORE-TRM to measure the effect of the dual-attention encoder.
(ii) CORE-HN against CORE-TRM to measure the effect of hard-negative-aware training.
(iii) CORE-DAHN to test whether the two improvements are complementary.

Table 1: Main results on CORE benchmarks (single run with `seed` = 2020). Metrics are Recall@20 / MRR@20 (%). Best MRR@20 per dataset is bolded.

Dataset	CORE-ave	CORE-trm	CORE-DA	CORE-HN	CORE-DAHN
Diginetica	50.28 / 17.99	52.87 / 18.66	52.90 / 18.66	52.52 / 18.46	52.67 / **18.69**
Nowplaying	20.24 / 6.64	21.73 / 7.45	21.77 / **7.67**	21.70 / 7.43	21.46 / 7.64
RetailRocket	58.58 / 37.18	61.80 / 38.59	61.95 / 38.80	61.82 / 38.69	61.83 / **38.85**
Tmall	45.02 / 31.62	44.99 / 31.63	45.01 / 31.63	44.92 / **31.76**	44.92 / 31.75
Yoochoose	59.02 / 25.07	64.58 / 28.21	64.60 / 28.29	64.56 / **28.31**	64.43 / 28.19

5.9 Results

Table 1 reports single-run test results across five CORE benchmarks. We report Recall@20 and MRR@20 as percentages. Best MRR@20 per dataset is highlighted.

Overall baselines. Across all datasets, CORE-trm substantially improves upon CORE-ave, consistent with the role of attention-based importance weighting in CORE's representation-consistent space. The gap is particularly pronounced on larger benchmarks such as Yoochoose and RetailRocket, where order-aware weighting improves both Recall@20 and MRR@20.

Effect of Dual-Attention (CORE-DA). Dual attention yields small but generally positive rank-quality gains (MRR@20) over CORE-trm on four out of five datasets (Nowplaying, RetailRocket, Tmall, Yoochoose), with the largest observed increase on Nowplaying and RetailRocket (about +0.2 absolute MRR points). On Diginetica, CORE-DA matches CORE-trm on MRR@20 and slightly improves Recall@20.

Effect of Hard Negatives (CORE-HN). Hard-negative-aware training exhibits a dataset-dependent trade-off. It improves MRR@20 on Tmall and Yoochoose (and slightly on RetailRocket), but can reduce MRR@20 on Diginetica and Nowplaying. In several cases, hard negatives slightly decrease Recall@20 while sharpening the ranking of the correct item when it remains in the top-K list, consistent with the intuition that focusing on hard confounders can reallocate probability mass away from easy negatives.

Combined model (CORE-DAHN). The combined model achieves the best MRR@20 on Diginetica and RetailRocket, indicating that the two mechanisms can be complementary in some regimes. However, complementarity is not universal: on Nowplaying, CORE-DA remains best; on Tmall and Yoochoose, CORE-HN remains best in MRR@20, while CORE-DAHN does not consistently add further gains and can slightly reduce Recall@20. This suggests that the interaction between the dual-attention encoder and hard-negative mining is sensitive to dataset characteristics and hyperparameter settings (e.g., λ and K).

5.10 Discussion

Our improvements were designed to strengthen CORE-trm along two orthogonal axes while preserving CORE's geometric principle.

Dual attention improves rank quality with minimal disruption. CORE-DA tends to increase MRR@20 while leaving Recall@20 nearly unchanged relative to CORE-trm. This pattern is consistent with the intended behavior: by explicitly modeling global session context and recent intent and adapting their fusion via a gate, the model can place the true next item higher in the ranked list without fundamentally changing the candidate set recalled into the top-K.

Hard negatives sharpen decisions but can hurt recall. CORE-HN often improves MRR@20 (notably on Tmall and Yoochoose) while slightly reducing Recall@20. This is consistent with the training signal: mining top-scoring incorrect items and penalizing them more strongly can improve separation from confounders, but may also make the distribution more peaked around a narrower set of candidates and thus slightly harm top-K coverage.

Why the combined model is not always additive. CORE-DAHN wins on Diginetica and RetailRocket in MRR@20, but does not dominate elsewhere. A plausible explanation is that hard-negative mining is performed using the current model's scoring function; when the encoder changes (dual attention + gating + residual), the set of mined negatives and the gradients induced by the hybrid objective can interact with the gating dynamics. If the gate overreacts to mined confounders, it may bias the fusion toward representations that are locally good for separating a few hard negatives but suboptimal for global top-K coverage, especially in datasets with highly popular items or frequent intent shifts. This motivates future tuning and analyses of λ, K, and gate behavior per dataset.

6 Conclusions and Future Work

We revisited CORE through the lens of its original motivation: ensuring representation consistency between session encoding and item decoding. Building on CORE-trm, we introduced two improvements that preserve this principle. First, we proposed a Dual-Attention mechanism with adaptive gating that fuses global context with explicit recent-intent emphasis while keeping the session embedding as a linear combination of item embeddings. Second, we proposed a hard-negative-aware hybrid objective that strengthens learning against confusing items by combining cross-entropy with a pairwise ranking term.

Empirically, Dual-Attention (CORE-DA) yields consistent, modest improvements in rank quality (MRR@20) across most benchmarks while maintaining similar Recall@20 to CORE-trm. Hard negatives (CORE-HN) provide dataset-dependent benefits, improving MRR@20 on Tmall and Yoochoose but slightly harming some datasets. The combined approach (CORE-DAHN) achieves the best MRR@20 on Diginetica and RetailRocket, but does not universally dominate, suggesting that complementarity depends on dataset characteristics and hyperparameters.

6.1 Limitations

This article reports single-run results with a fixed random seed and therefore does not establish statistical significance. Moreover, hard negative mining introduces additional computation due to top-K selection; practical deployments on very large catalogs may require approximate retrieval or cached candidate sets.

Hard negatives and false negatives. Hard-negative mining may occasionally select unobserved yet relevant items as negatives (false negatives). In this work we follow the standard implicit-feedback assumption and only exclude the ground-truth next item; exploring additional filtering or debiasing strategies is left for future work.

6.2 Future work

Two directions appear especially promising. First, we will run multiple seeds per configuration and report mean±std with statistical tests to verify that the observed gains are robust. Second, we will explore dataset-adaptive hard-negative settings (e.g., tuning λ and K, or scheduling λ over training) and analyze gate behavior to better understand when dual attention and hard-negative mining are complementary. Finally, we plan to explore multi-interest consistent-space variants

and graph-informed weighting mechanisms that influence importance estimation while preserving CORE's representation principle [4, 3].

References

[1] D. Jannach, M. Quadrana, and P. Cremonesi, "Session-based recommender systems," in *Recommender Systems Handbook*, pp. 301–335, Springer, 3rd ed., 2022.

[2] S. Wang, L. Cao, Y. Wang, Q. Z. Sheng, M. A. Orgun, and D. Lian, "A survey on session-based recommender systems," *ACM Computing Surveys*, 2021.

[3] Z. Li, C. Yang, Y. Chen, X. Wang, H. Chen, G. Xu, L. Yao, and Q. Z. Sheng, "Graph and sequential neural networks in session-based recommendation: A survey," *arXiv preprint*, 2024.

[4] Y. Hou, B. Hu, Z. Zhang, and W. X. Zhao, "CORE: Simple and effective session-based recommendation within consistent representation space," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2022.