Hierarchical Group-wise Ranking Framework for Recommendation Models

YaChen Yan, Liubo Li, Ravi Choudhary

July 31, 2025

Summary

Background

- CTR/CVR models can achieve improved performance by incorporating **ranking** objectives beyond **calibration** objectives.
  - Calibration Objective: Pointwise LogLoss
  - Ranking Objective: Pairwise/Listwise Ranking Loss
  - **Hybrid Objective**: Combining calibration loss with ranking loss

$$L = \alpha L_{\text{calibration}} + (1 - \alpha) L_{\text{ranking}}$$

- **Context-aware Data Sampling:**
  - Ranking loss is applied over item lists grouped by:
    - Recommendation request
    - Search query
    - User session

Background Cont.

- Existing research efforts mainly focused on the formulation and optimization of **ranking loss function**: RankNet, ListNet, Calibrated Softmax, JRC, ListCE.
- Recent work[1] demonstrates that incorporating ranking loss generates **larger gradients for negative samples**, thereby alleviating the **gradient vanishing issue** commonly observed when optimizing solely with pointwise logloss.

[1] Wu et al., *Understanding the Ranking Loss for Recommendation with Sparse User Feedback*, KDD 2024

Motivation

- Problems:
  - Context-aware data sampling **violates global data shuffling**, potentially degrading model performance.
  - Severe **sample and label sparsity** at the context level, particularly in CVR prediction tasks.
- Solution:
  - Design a **cross-context data sampling** strategy to preserve global data shuffling and enrich ranking signal.
  - Conduct **controllable hard negative sampling** based on gradient magnitude to reduce gradient variance and accelerate convergence.

Hierarchical Group-wise Negative Sampling

- **Intuition:**
  - Similar users tend to:
    - Be exposed to similar items
    - Share common interests
  - These users can provide more informative negatives for each other
- **Key Idea:**
  - Cluster users using Residual Vector Quantization (RVQ)
    - Can be **co-updated** with ranking model in streaming settings
    - Better clustering quality with **hierarchical structure**
  - Group user-item pairs based on the learned hierarchical user clusters
  - Applying listwise ranking loss
    - On each user-item interaction **groups**
    - Across each hierarchical cluster **levels**

Residual Vector Quantization

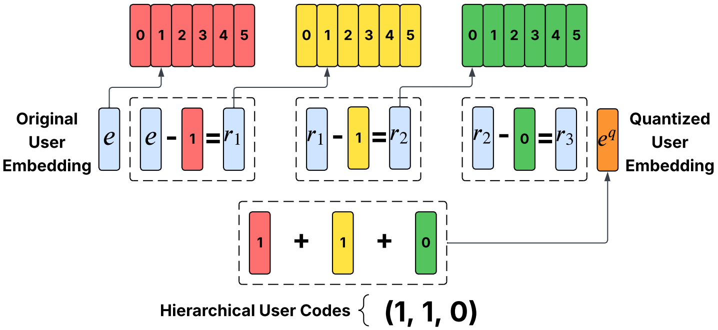


Figure: Recursive multi-level vector quantization

- **Quantized User Embeddings:** Discrete, semantic embeddings for similar user groups.
- **Hierarchical User Codes:** Trie-like structure; shared prefixes indicate user similarity.

Training of Residual Vector Quantizer

- **Objective:**
  - Hierarchical codes capturing user similarity semantics
  - Adaptive clustering responsive to evolving user interests
- **Modified Loss Formulation (compared to VQ-VAE):**
  - **Reconstruction Loss $\rightarrow$ Auxiliary Calibration Loss**
    - Applies logloss on auxiliary predictions ($\hat{y}^q$) computed from $\mathbf{e}_u^q$, $\mathrm{sg}(\mathbf{e}_i)$.
    - Promotes task-relevant semantics in the codebook space.
    - Facilitates adaptation to evolving user interests.
  - **Codebook Loss $\rightarrow$ EMA Update**
    - Enables smooth, streaming-friendly codebook update.
    - Improves codebook utilization.
  - **Commitment Loss $\rightarrow$ Omitted**
    - Limits embedding flexibility and dynamic cluster transitions.
    - Omission enables real-time adaptation to shifting user interests.

Visualization: Hierarchical Group-wise Negative Sampling

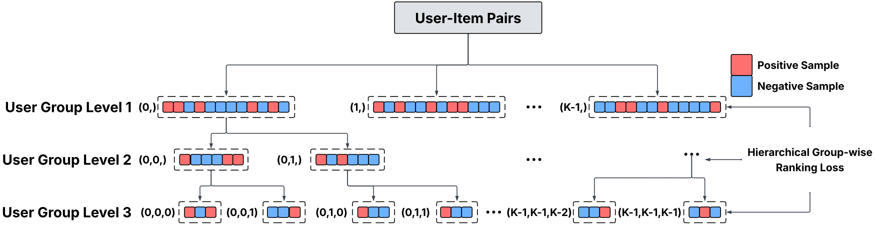


Figure: Hierarchical sampling is performed over multi-level nested user clusters. Items from similar users are grouped to create enriched sample lists.

- Samples are drawn from different cluster depths in RVQ hierarchy.
  - Shallow: easier negatives (coarse-grained clusters)
  - Deep: harder negatives (fine-grained clusters)
- Listwise Cross-Entropy loss is applied
  - On each user-item interaction groups
  - Across each hierarchical cluster levels

Model Architecture

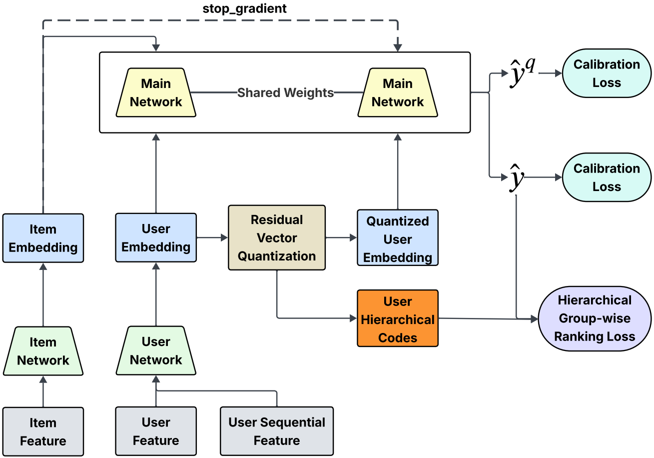


Figure: Model architecture with RVQ and Hierarchical Group-wise Ranking

Multi-Objective Training

- **Calibration Objective:**
  - Logloss applied on $\hat{y}$ from original embedding $\mathbf{e}_u$, $\mathbf{e}_i$.
- **Auxiliary Calibration Objective:**
  - Logloss applied on $\hat{y}^q$ from:
    - Quantized user embedding $\mathbf{e}_u^q$ (via STE)
    - Stop-gradient item embedding $\text{sg}(\mathbf{e}_i)$
- **Hierarchical Group-wise Ranking Objective:**
  - Listwise Cross-Entropy loss is **applied across RVQ levels**.
  - Each level's loss is **weighted by learned uncertainty**.

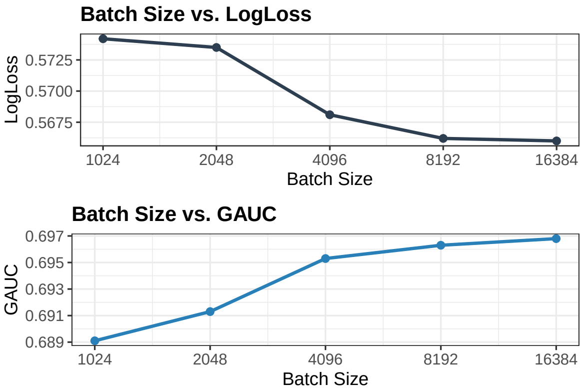
Objective Function

$$
\begin{aligned}
\mathcal{L} = \mathcal{L}_{\text{logloss}}(\hat{y}, y) + \lambda \mathcal{L}_{\text{logloss}}(\hat{y}^q, y) \\
+ \sum_{l=1}^{L} \left(\frac{1}{2\sigma_l^2} \mathcal{L}_{\text{listce}}^{(l)} + \log \sigma_l \right)
\end{aligned}
$$

Model Performance Comparison

Objective	KuaiRand			Taobao		
	LogLoss	AUC	GAUC	LogLoss	AUC	GAUC
LogLoss	0.5735	0.7510	0.6911	0.2011	0.6420	0.5708
LogLoss + PairwiseLogistic	0.5723	0.7524	0.6921	0.2002	0.6435	0.5728
LogLoss + SoftmaxCE	0.5727	0.7520	0.6920	0.2005	0.6428	0.5720
LogLoss + ListCE	0.5709	0.7537	0.6932	0.1995	0.6443	0.5734
JRC	0.5713	0.7533	0.6930	0.1993	0.6540	0.5732
GroupCE (proposed)	**0.5681**	**0.7556**	**0.6953**	**0.1982**	**0.6556**	**0.5745**

Effect of Batch Size on Model Performance



- LogLoss and GAUC improve significantly as **batch size increases**.
- Larger batches enable **more diverse cross-context negative samples**, enhancing training signal.

Summary

- We propose a **hierarchical group-wise sampling** strategy for cross-context hard negative mining, based on a learned user hierarchy derived from Residual Vector Quantization (RVQ).
- We introduce a **hierarchical listwise ranking loss** to capture coarse-to-fine-grained preference signals across multiple granularity levels of user similarity.

Thank You!