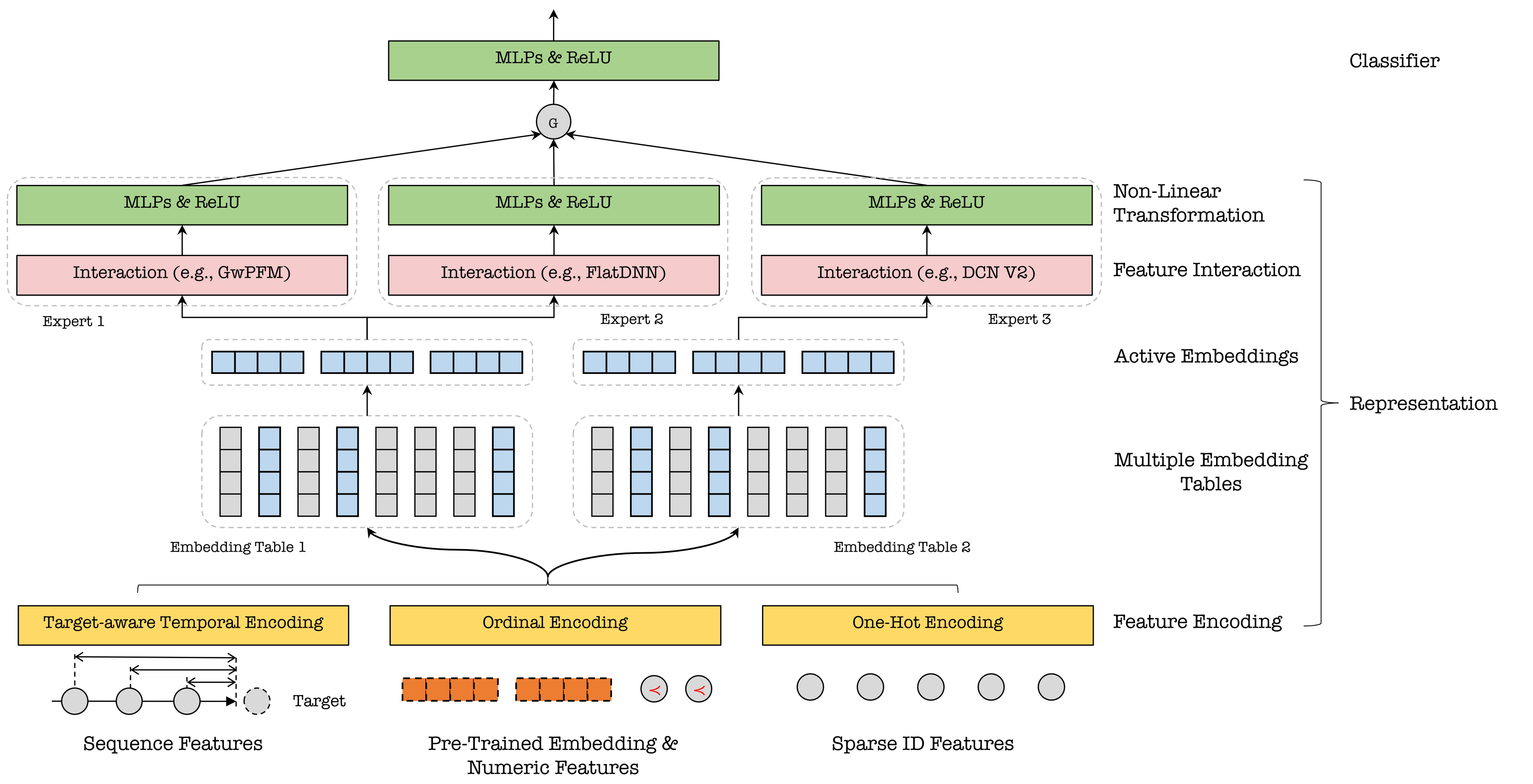
Ads Recommendation in a Collapsed and Entangled World

Junwei Pan, Wei Xue, Ximei Wang, Haibin Yu,
Xun Liu, Shijie Quan, Xueming Qiu, Dapeng Liu, Lei Xiao, Jie Jiang

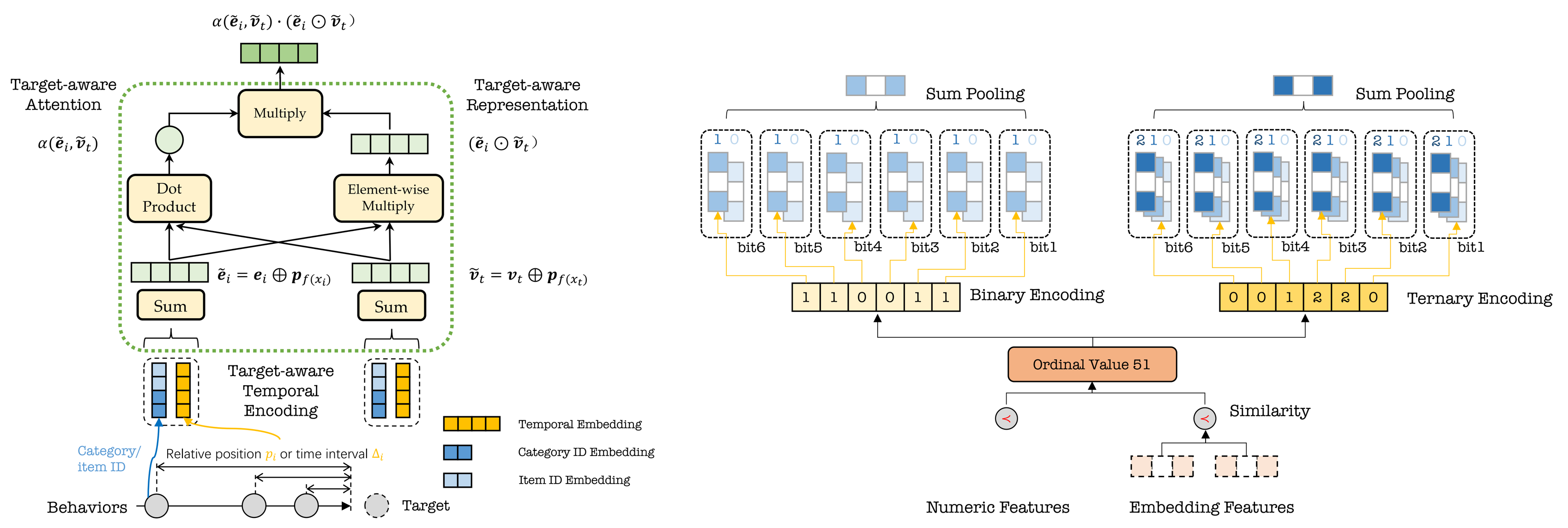
KDD2024 BARCELONA, SPAIN

Overview Architecture

- **Heterogeneous Mixture-of-Experts with Multi-Embedding**
  - 4 key components: Feature encoding, Multi-embedding lookup, Experts, Classification towers



Feature Encoding



- **Sequence Features**:
  - We propose Temporal Interest Module (TIM) [Tencent, WWW 2024] to capture the **semantic-temporal correlations** between behaviors and target.
  -

$$u_{\mathrm{TIM}} = \sum_{X_i \in \mathcal{H}} \alpha(\tilde{\mathbf{e}}_i, \tilde{\mathbf{v}}_t) \cdot (\tilde{\mathbf{e}}_i \odot \tilde{\mathbf{v}}_t)$$

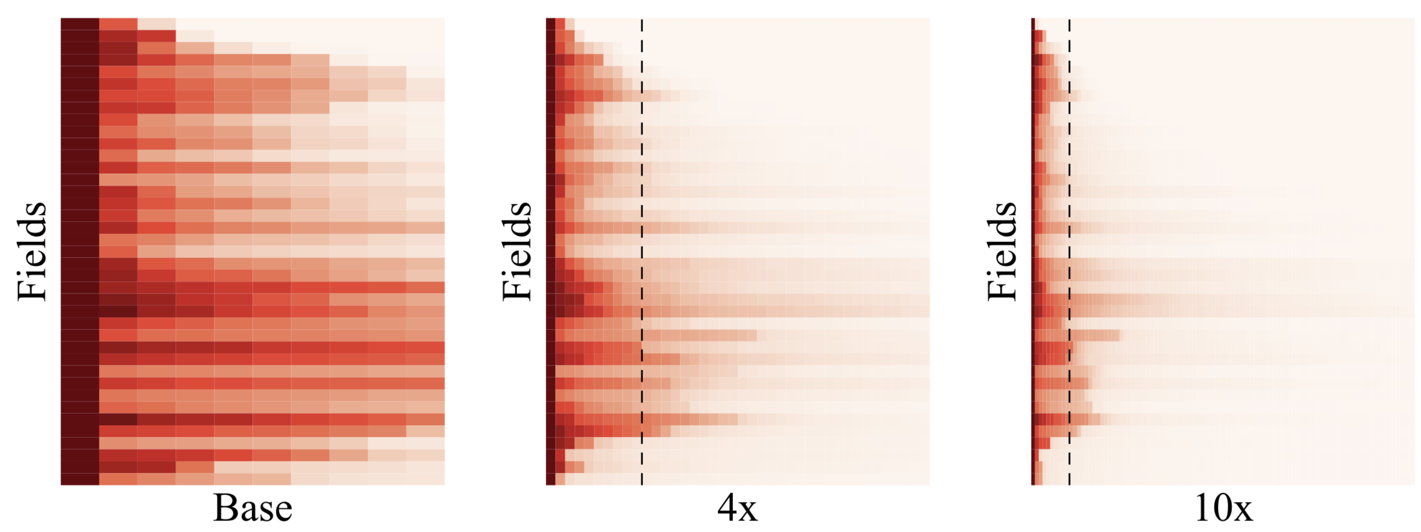
- **Numeric Features**:
  - Get the code of numeric features according to multiple numeral systems (i.e., binary, decimal) and then assign learnable embeddings to these codes.
  -

$$f_{\mathrm{MNS}}(v) = [\sum_{k=1}^{K_2} \mathbf{X}^{(2)}_{2k+\mathrm{B}_k}, \sum_{k=1}^{K_3} \mathbf{X}^{(3)}_{3k+\mathrm{C}_k}, \ldots, \sum_{k=1}^{K_n} \mathbf{X}^{(n)}_{nk+\mathrm{N}_k}]$$

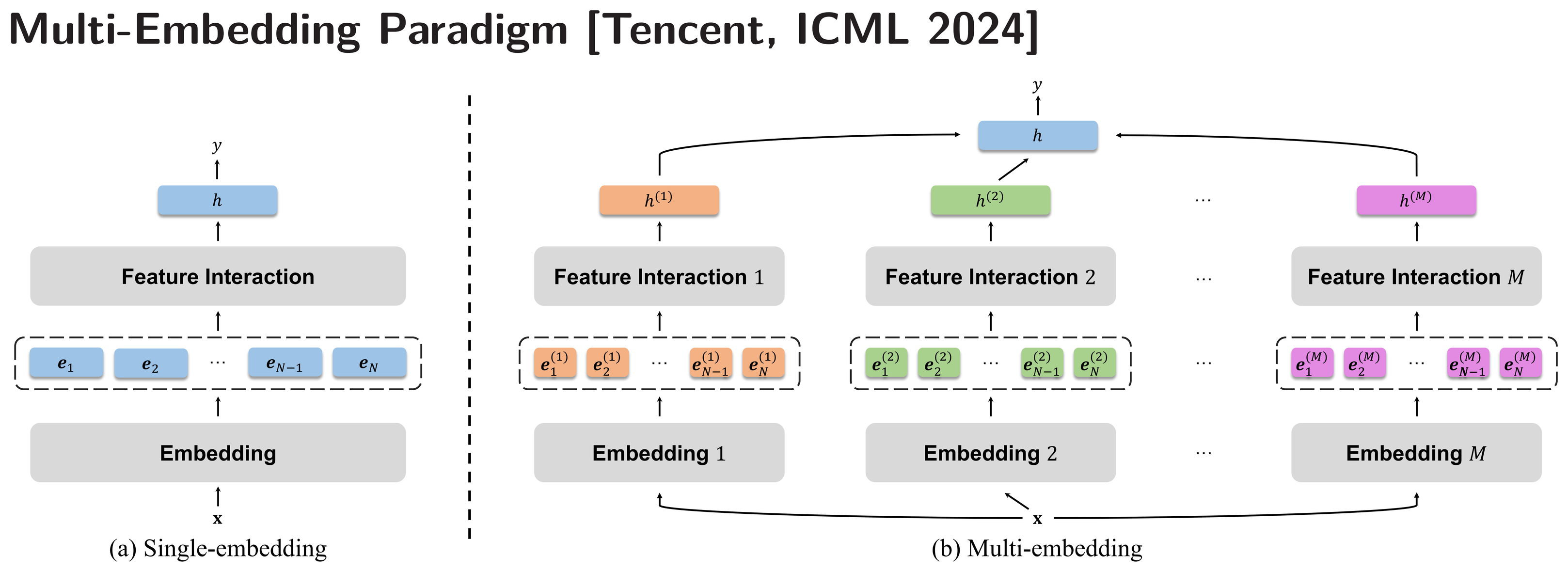
- **Embedding Features**:
  - Compute the similarity score between pre-trained embeddings of users and items, treat it as a numeric feature, and employ MNSE to encode it.

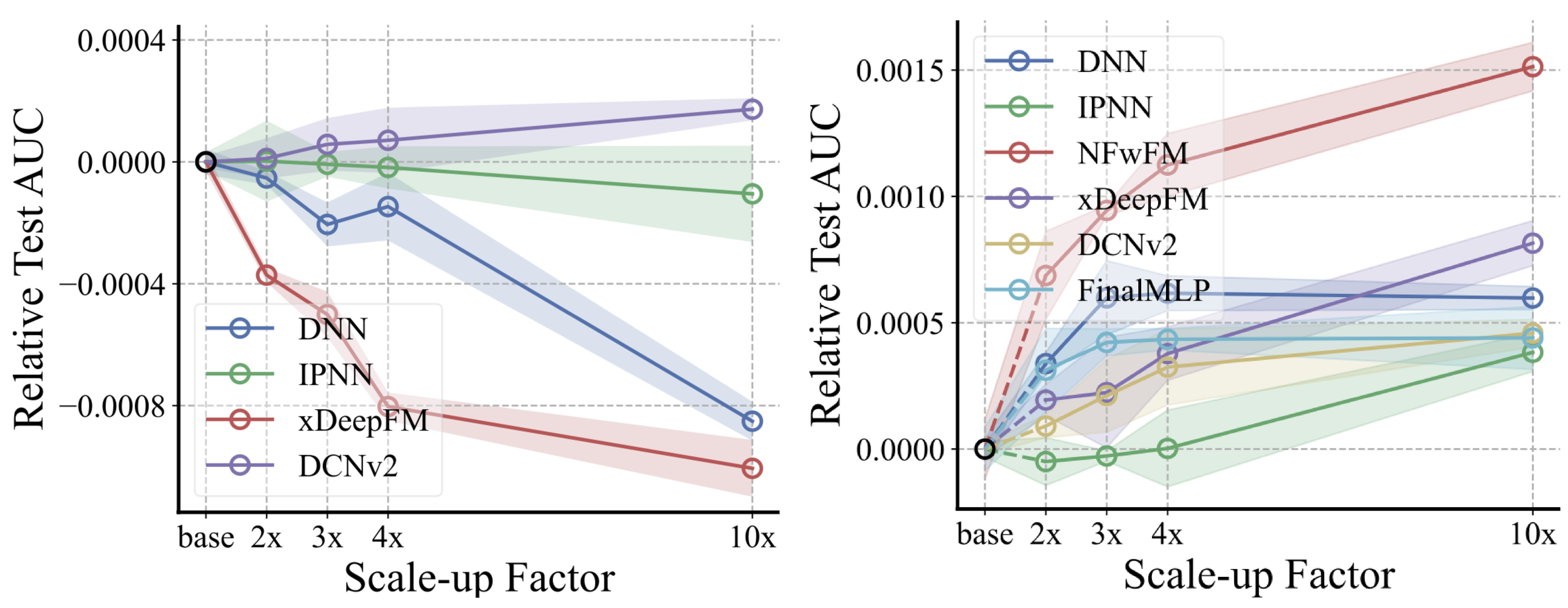
Dimensional Collapse

- **Dimensional Collapse of Embeddings**
  - Visualized by the singular spectral analysis.
  - Many singular values are very small, indicating that embeddings of many fields end up spanning a lower-dimensional subspace instead of the entire available embedding space.



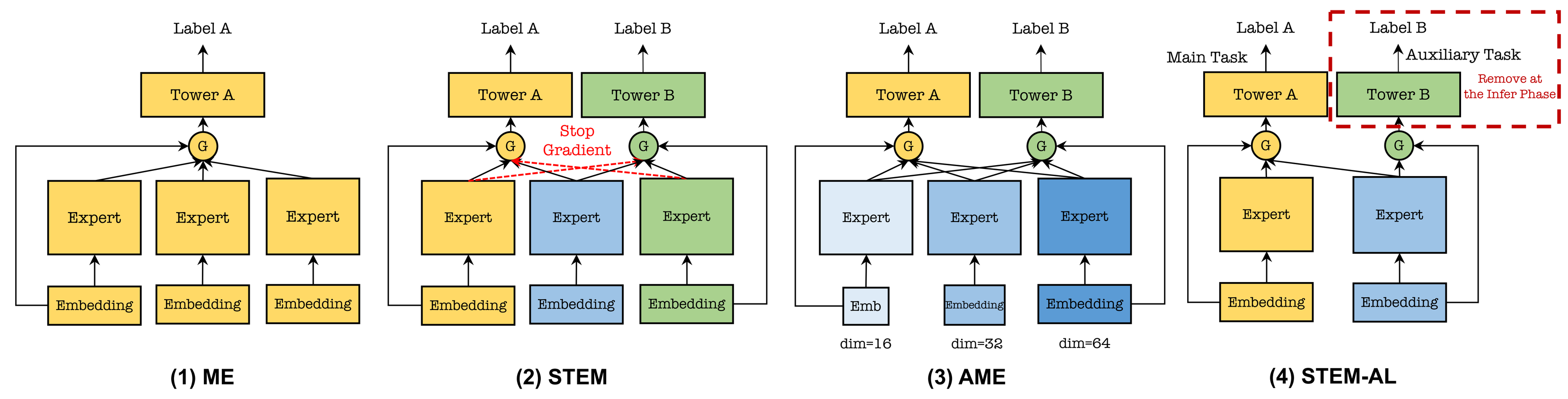
- **Multi-Embedding Paradigm [Tencent, ICML 2024]**





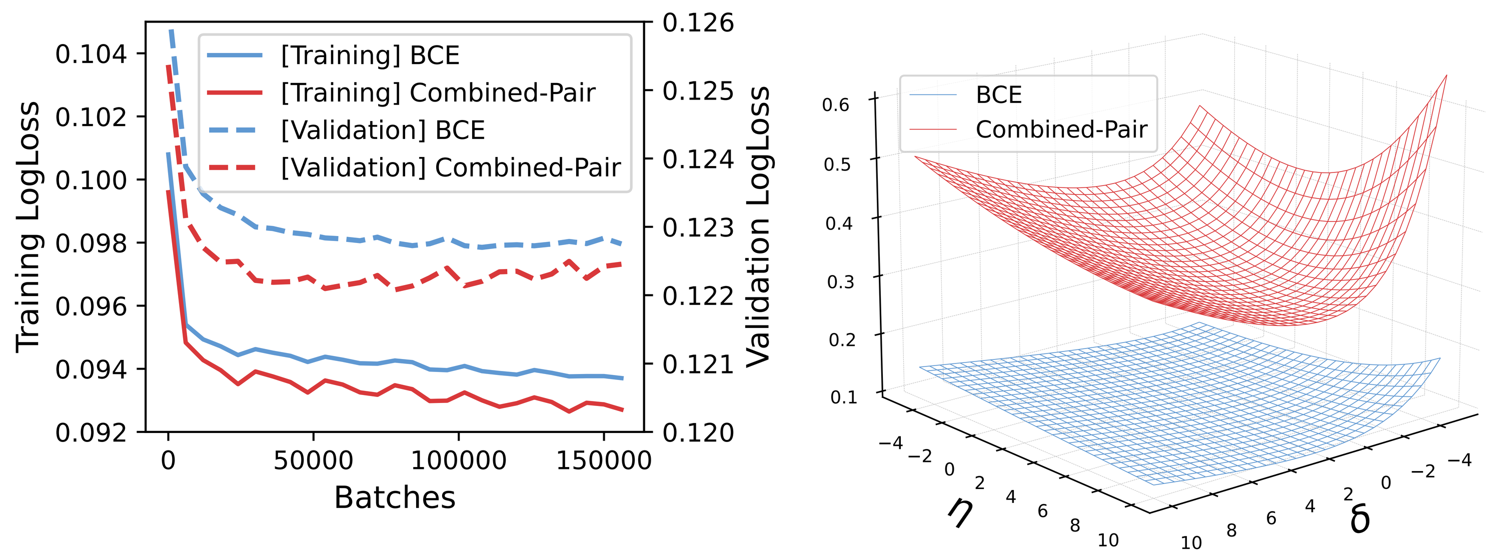
Interest Entanglement

- **Multi-Task Learning**
  - Shared and Task-specific Embedding (STEM) [Tencent, AAAI 2024]
  - Asymmetric Multi-Embedding (AME)
- **Auxiliary Learning**
  - Shared and Task-specific Embedding for Auxiliary Learning (STEM-AL)



Model Training

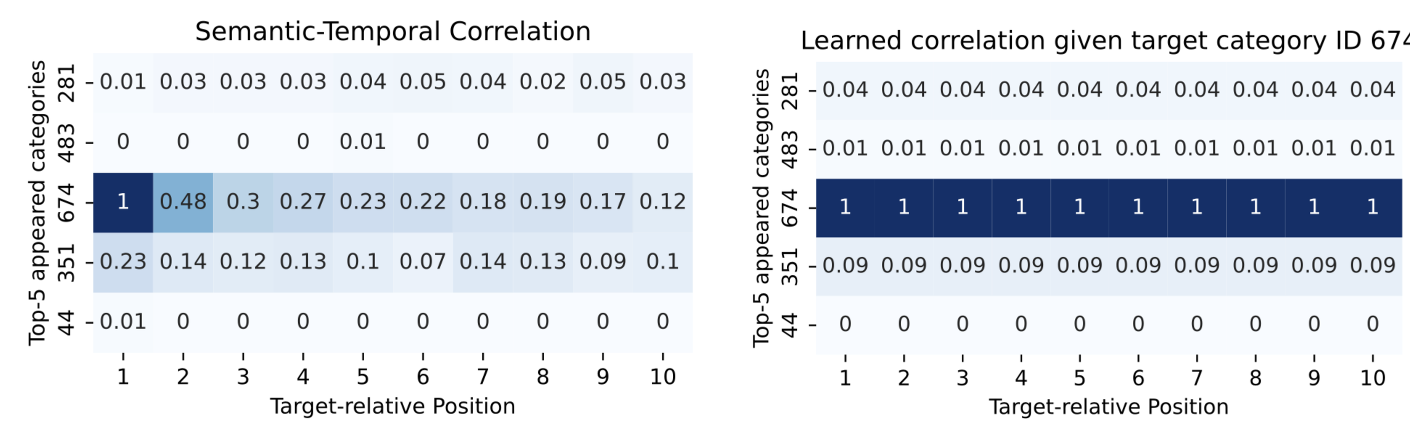
- **Auxiliary Ranking Loss** [Tencent, KDD 2024]: We find when only using Binary Cross Entropy loss in sparse positive scenarios (such as CTR prediction), negative samples suffer from **gradient vanishing**. Introducing ranking loss mitigates this problem, resulting in better classification ability.



- **Repeated Exposure**: We employ **Inverse Propensity Weighting** to down weight the positive samples of repeated exposure.

Analysis Tools

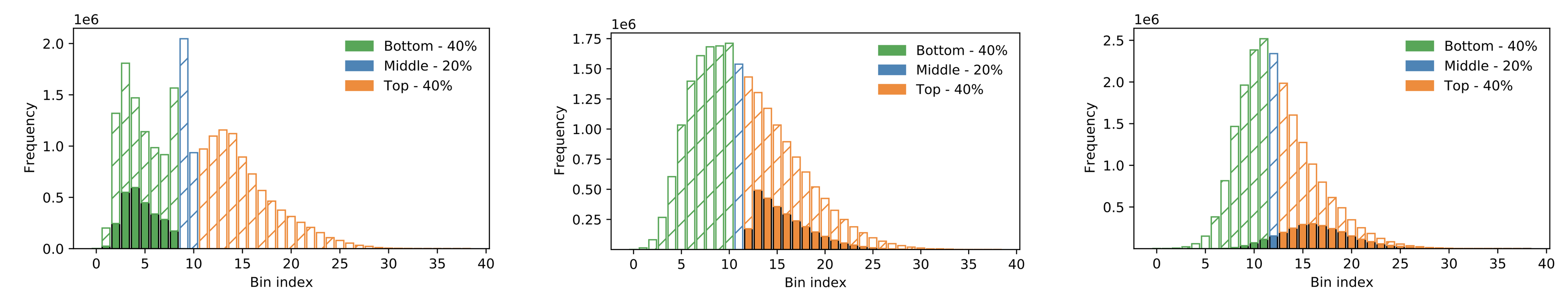
- **Analysis of Behavior-Target Correlation**



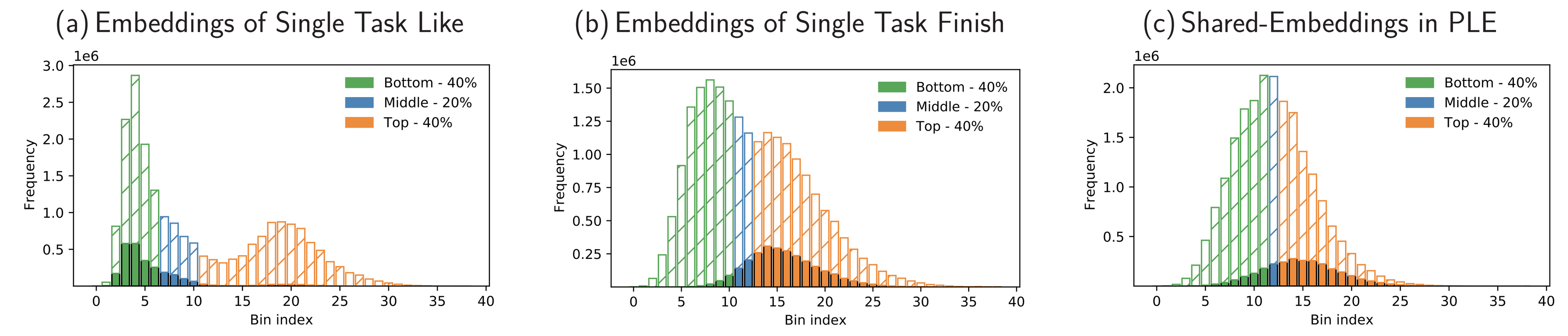
(a) Ground truth Correlation (b) DIN's learned Correlation

(c) SASRec's learned Correlation (d) TIM's learned Correlation

- **Analysis of Interest Entanglement**



(a) Embeddings of Single Task Like (b) Embeddings of Single Task Finish (c) Shared-Embeddings in PLE



(d) Like-Specific Embeddings in STEM (e) Finish-Specific Embeddings in STEM (f) Shared Embeddings in STEM

References

- Temporal Interest Network for User Response Prediction. WWW 2024.
- On the Embedding Collapse when Scaling up Recommendation Models. ICML 2024.
- STEM: Unleashing the Power of Embeddings for Multi-task Recommendation. AAAI 2024.
- Understanding the Ranking Loss for Recommendation with Sparse User Feedback. KDD 2024.