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

Tencent 腾讯

KDD2024
BARCELONA, SPAIN

Overview

Three main challenges:

- **C1: Priors for Representation**

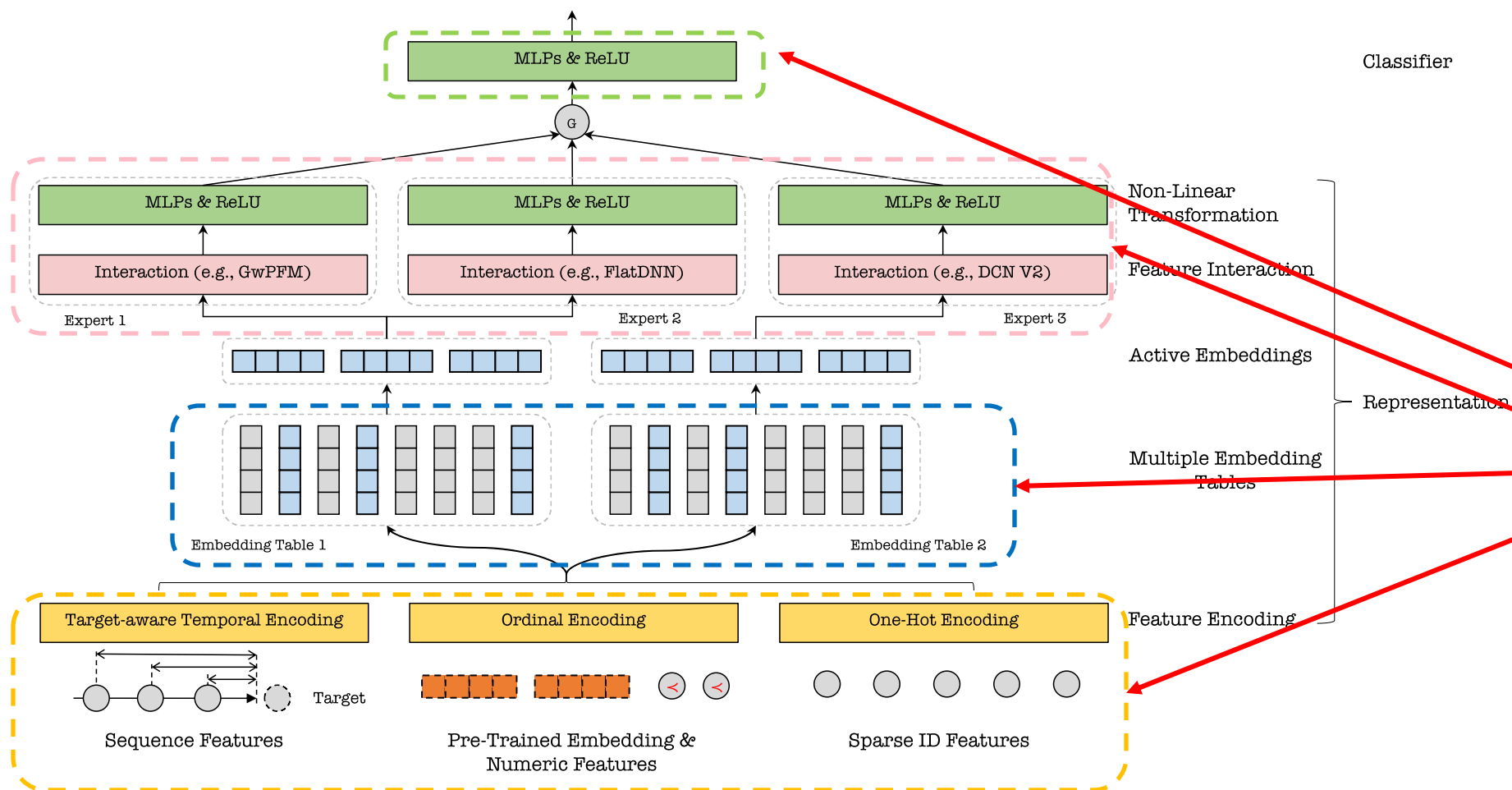- **C2: Dimensional Collapse**

- **C3: Interest Entanglement**

Others

- Model Training Techniques

- Analysis Tools

This work is a summary of many advances of Tencent Ads Science team in the last several years, some of them are published as individual papers (e.g., in ICML 2024, KDD 2024, WWW 2024, AAAI 2024, etc.).

Architecture

Our model follows the widely adopted Embedding & Explicit Interaction framework, which consists of four key modules:

- Classification towers
- Experts
- Multi-embedding lookup
- Feature encoding

Outline

Three main challenges:

- **C1: Priors for Representation**

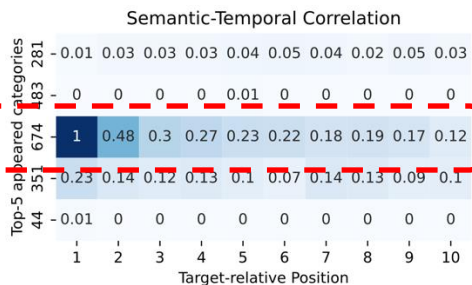- **C2: Dimensional Collapse**

- **C3: Interest Entanglement**

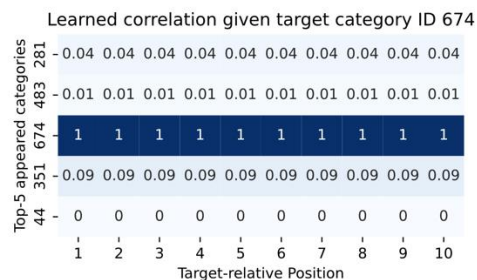# C1: Priors of Sequence Features

Target category: 674

Semantic correlation: behaviors with the same category as the target (674 here) are more correlated.

Temporal correlation: there is a strong temporal decaying pattern.
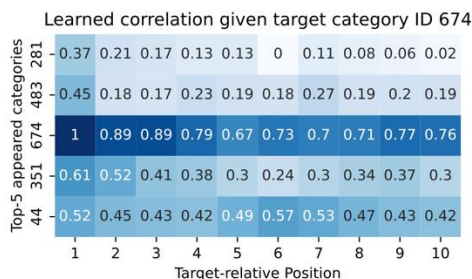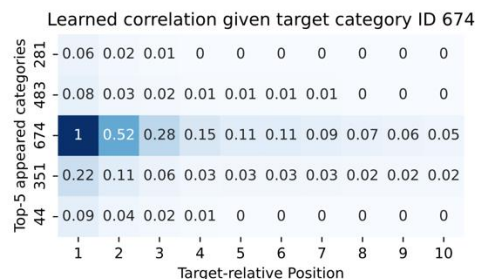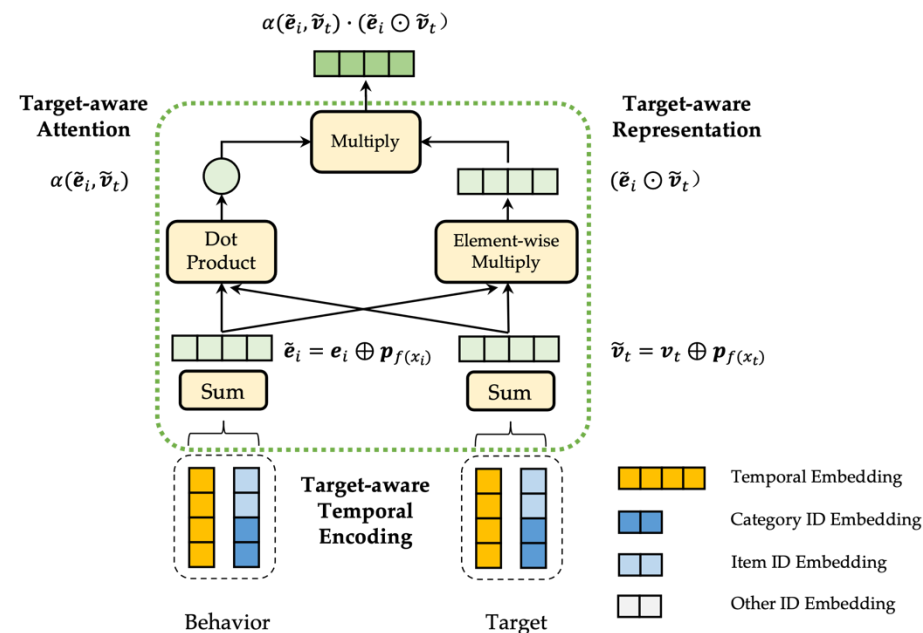

Semantic-Temporal Correlation

(a) Ground truth Correlation

Learned correlation given target category ID 674

(b) DIN's learned Correlation

Learned correlation given target category ID 674

(c) SASRec's learned Correlation

Learned correlation given target category ID 674

(d) TIM's learned Correlation

$\alpha(\tilde{e}_i, \tilde{v}_t) \cdot (\tilde{e}_i \odot \tilde{v}_t)$

Target-aware Attention
$\alpha(\tilde{e}_i, \tilde{v}_t)$

Target-aware Representation
$(\tilde{e}_i \odot \tilde{v}_t)$

Multiply

Dot Product

Element-wise Multiply

$\tilde{e}_i = e_i \oplus p_{f(x_i)}$ $\tilde{v}_t = v_t \oplus p_{f(x_t)}$

Sum   Sum

Behavior   Target

Target-aware Temporal Encoding

Temporal Embedding
Category ID Embedding
Item ID Embedding
Other ID Embedding

- There are strong semantic and temporal correlation between behaviors and the target.
- Existing SOTA methods fail to capture such correlation.

Our proposed TIN composes 3 key components: target-aware temporal encoding, target-aware attention and target-aware representation.

KDD2024 BARCELONA, SPAIN

C1: Priors of Numeric & Embedding Features
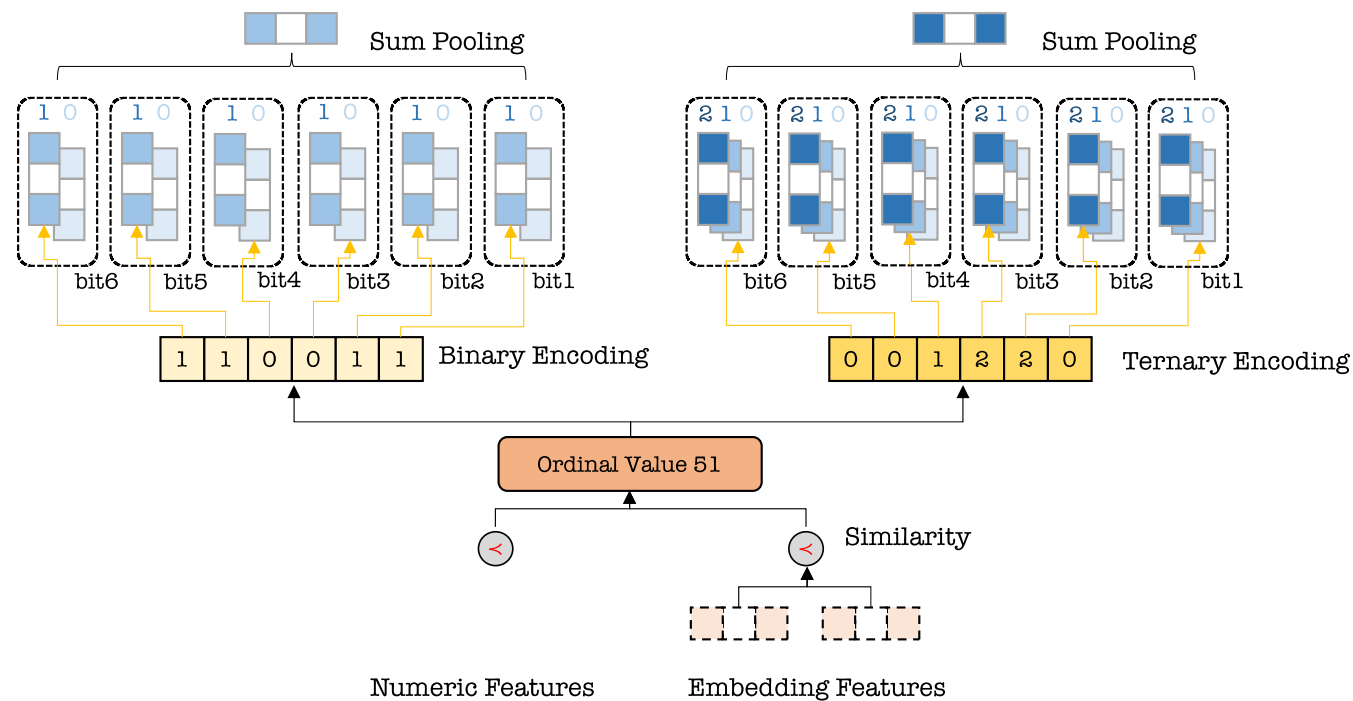
Numeric Features

Multiple Numeral Systems Encoding (MNSE):

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

$$\mathbb{B} = \text{func\_binary}(v), \mathbb{C} = \text{func\_ternary}(v), \ldots$$

Embedding Features

Compute a similarity score based on the embeddings

$$w_{\text{sim}}(u, i) = \text{sim}(\bar{e}_u, \bar{e}_i).$$

Treat the score as a numeric feature, use the MNSE method

Outline

Three main challenges:

- **C1: Priors for Representation**

- **C2: Dimensional Collapse**
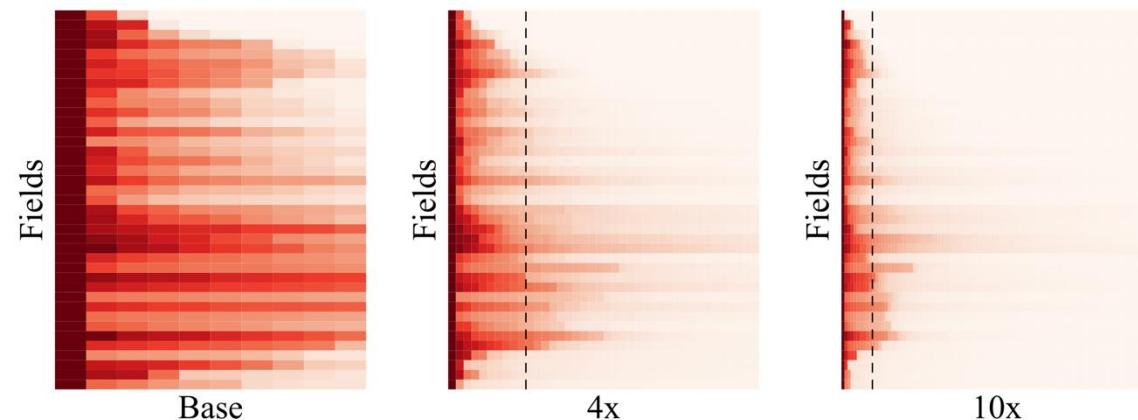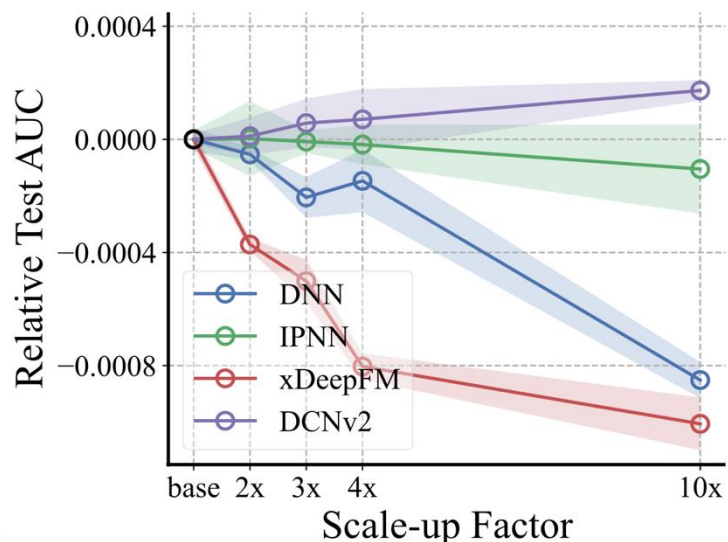
- **C3: Interest Entanglement**

C2: Dimensional Collapse of Embeddings

Large Recommendation System

How to scale up recommendation models w.r.t. #parameters?
- Most parameters in RecSys are embeddings (**99.999%**)
- Simply enlarge the embedding dimension does not always work.
- Singular spectrum analysis show dimensional collapse of embeddings

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.

Finding 1 (Interaction-Collapse Theory). In feature interaction of recommendation models, fields with low-information-abundance embeddings constrain the information abundance of other fields, resulting in collapsed embedding matrices.

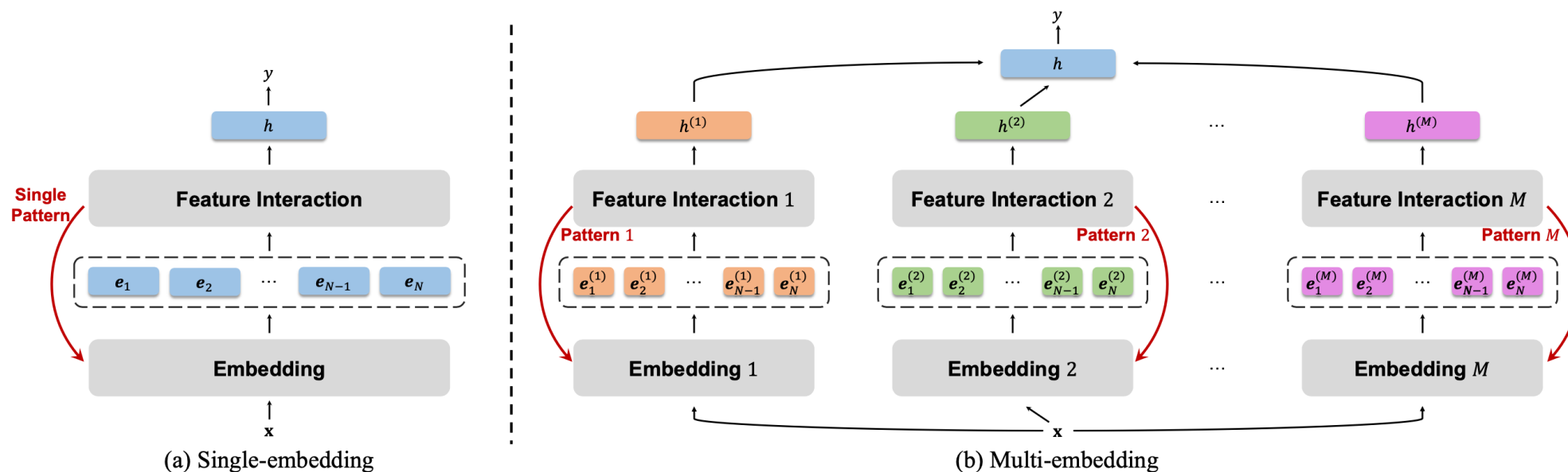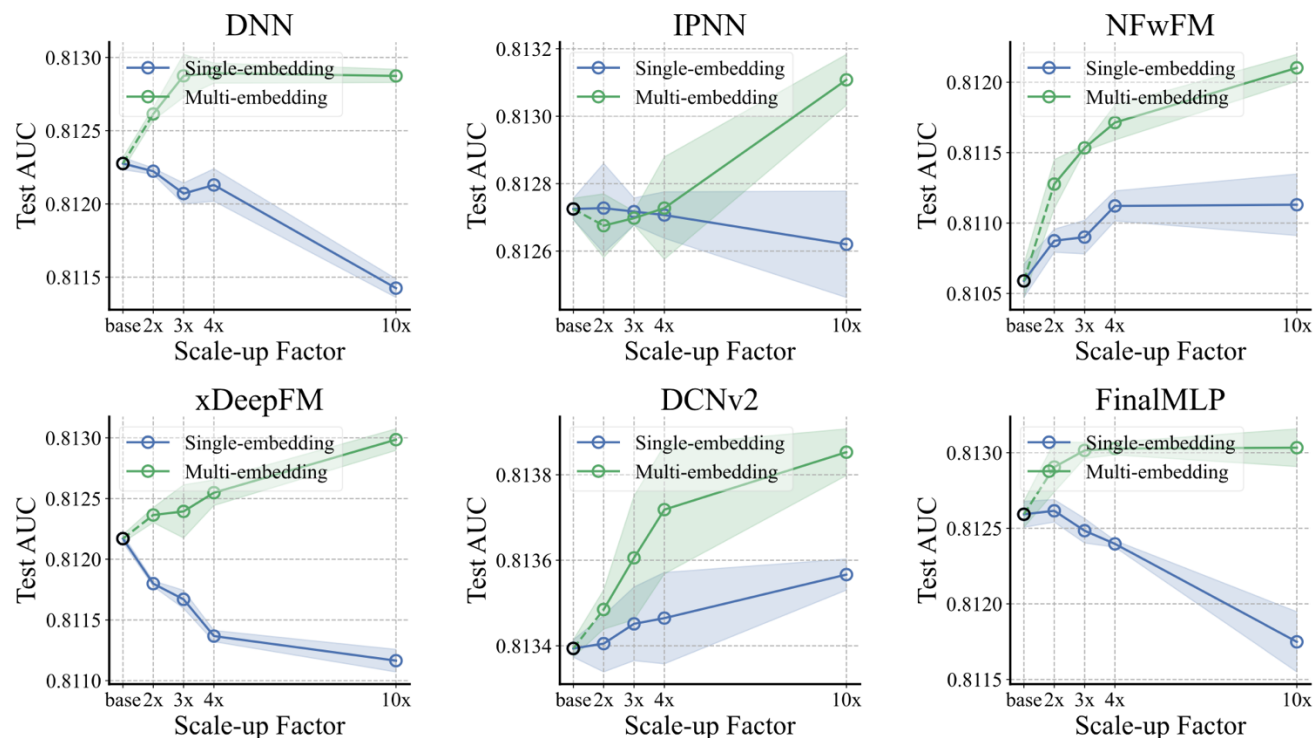# C2: Dimensional Collapse - Architecture

Figure 7. Architectures of single-embedding (left) and multi-embedding (right) models.

- Instead of scaling up the model by enlarging the embedding dimensions, we scale by **introducing more embedding tables**: the **Multi-Embedding paradigm**.
- *Less influenced* by the interaction-collapse theory and *mitigate embedding collapse* while *keeping* the original interaction modules

C2: Dimensional Collapse - Results

- *Can be applied to all existing CTR models*
- *Compared to Single-Embedding, Multi-Embedding paradigm, successfully achieves* *scale-up regarding #params*
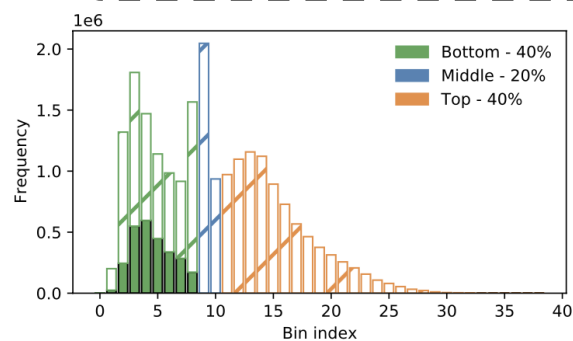
Outline

Three main challenges:

- **C1: Priors for Representation**

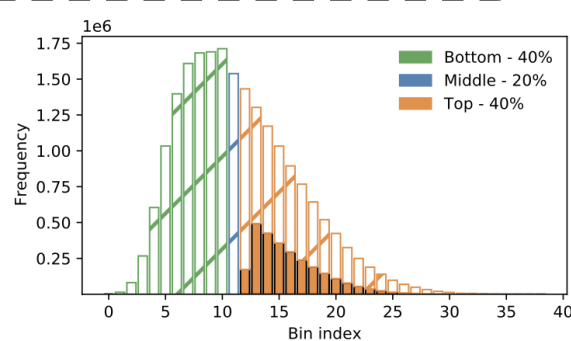- **C2: Dimensional Collapse**

- **C3: Interest Entanglement**
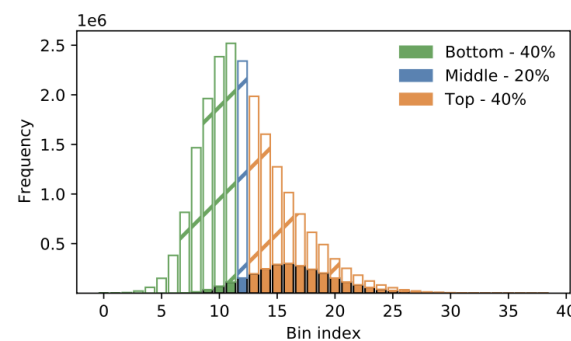
C3: Interest Entanglement

- Consider two tasks: like and finish.
- We select a set of (user, item) pairs with conflicting embedding distance (interest) according two individual tasks' embeddings.
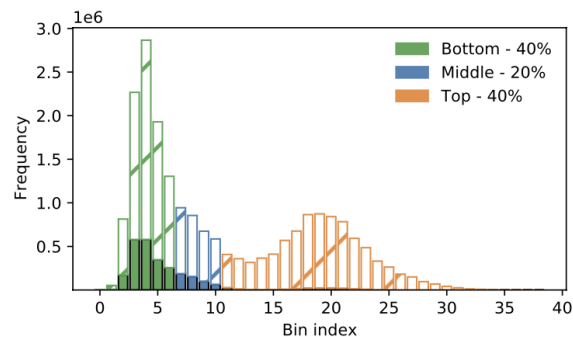

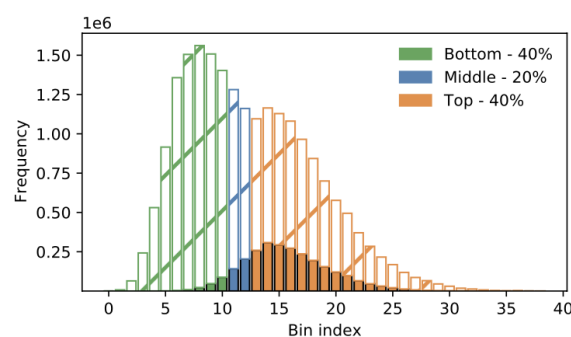
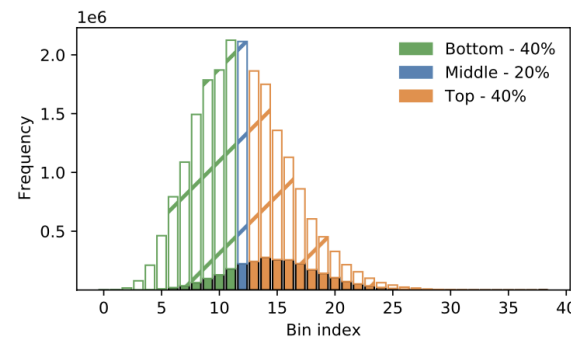(a) Embeddings of Single Task Like

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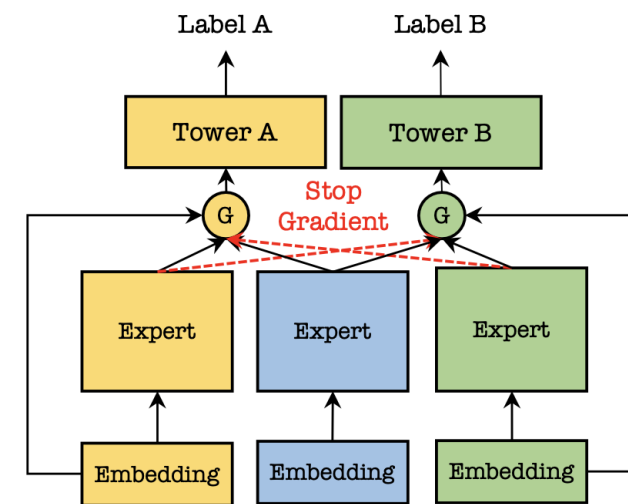
(c) Shared-Embeddings in PLE

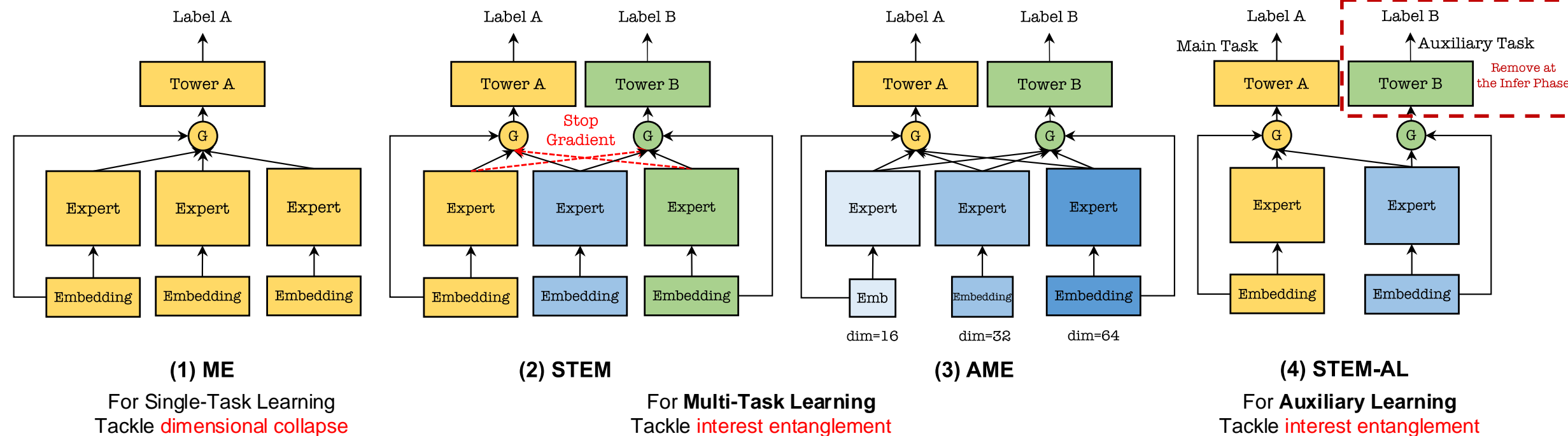(d) Like-Specific Embeddings in STEM

(e) Finish-Specific Embeddings in STEM

(f) Shared Embeddings in STEM

(2) **STEM**

Shared-and Task-specific Embedding paradigm.

C3: Interest Entanglement, Architecture

(1) ME
For Single-Task Learning
Tackle dimensional collapse

(2) STEM
For **Multi-Task Learning**
Tackle interest entanglement

(3) AME

(4) STEM-AL
For **Auxiliary Learning**
Tackle interest entanglement

Multi-Task or Multi-Domain Learning
- STEM: One embedding table for each task: too many embedding tables!
- AME (Asymmetric Multiple-Embedding): Fixed number of embeddings, with different dimension to disentangle them.
- Crocodile: Fixed number of embeddings, with an explicit disentangle loss.
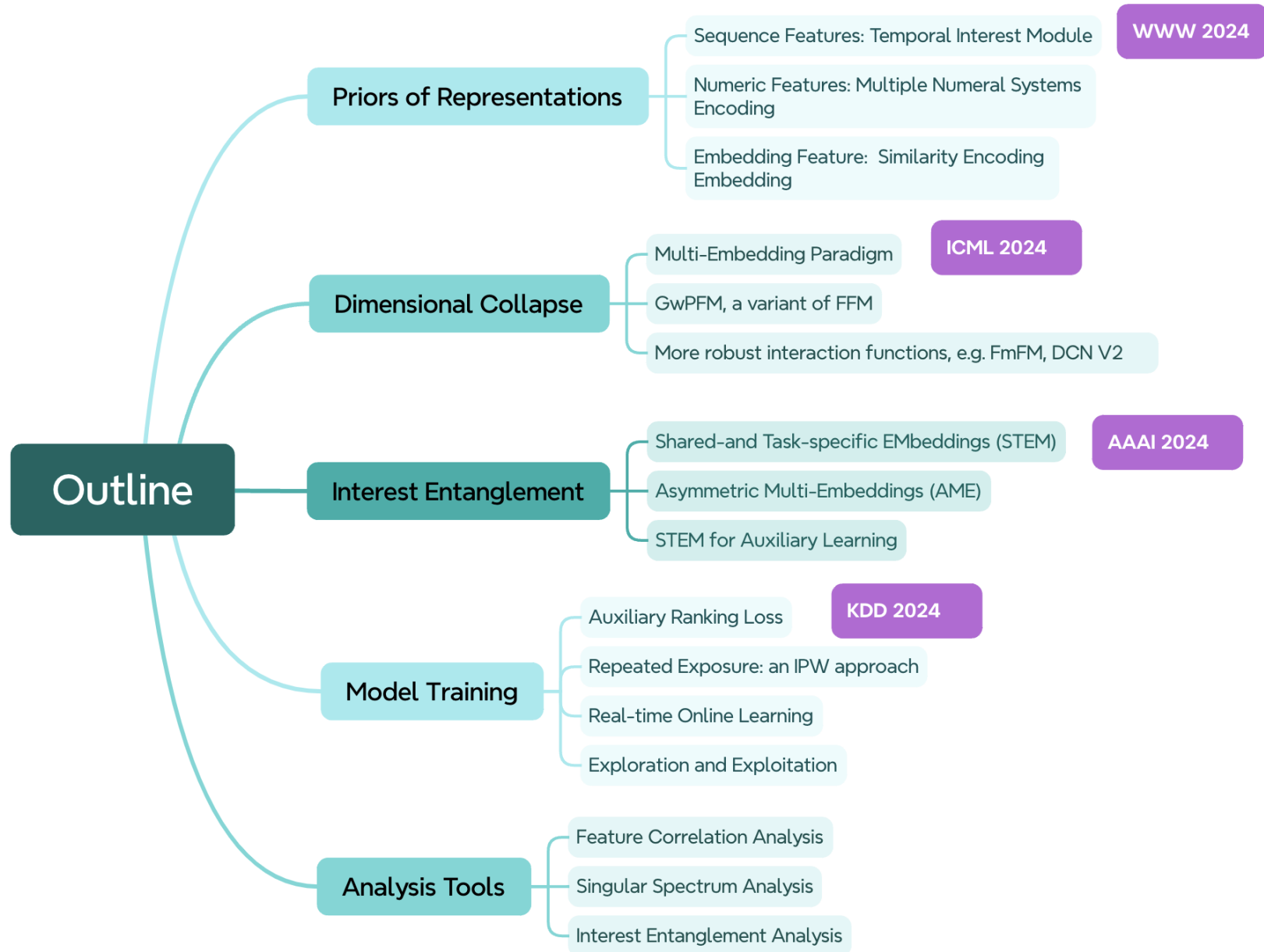
Auxiliary Learning
- STEM-AL: One embedding for the main task, and one shared embedding for all tasks.

STEM: Unleashing the Power of Embedding for Multi-Task Recommendation. AAAI 2024.
Crocodile: Cross Experts Covariance for Disentangled Learning in Multi-Domain Recommendation. arxiv 2024.

Overview

Outline

Priors of Representations — WWW 2024
- Sequence Features: Temporal Interest Module
- Numeric Features: Multiple Numeral Systems Encoding
- Embedding Feature: Similarity Encoding Embedding

Dimensional Collapse — ICML 2024
- Multi-Embedding Paradigm
- GwPFM, a variant of FFM
- More robust interaction functions, e.g. FmFM, DCN V2

Interest Entanglement — AAAI 2024
- Shared-and Task-specific EMbeddings (STEM)
- Asymmetric Multi-Embeddings (AME)
- STEM for Auxiliary Learning

Model Training — KDD 2024
- Auxiliary Ranking Loss
- Repeated Exposure: an IPW approach
- Real-time Online Learning
- Exploration and Exploitation

Analysis Tools
- Feature Correlation Analysis
- Singular Spectrum Analysis
- Interest Entanglement Analysis

KDD2024
BARCELONA, SPAIN

Q & A

Reference papers by our team:

- 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.

Recommender

SCAN ME

KDD2024
BARCELONA, SPAIN

Outline

Three main challenges:

- **C1: Priors for Representation**

- **C2: Dimensional Collapse**

- **C3: Interest Entanglement**

Others

- Model Training

- Analysis Tools

Model Training

- Auxiliary Ranking Loss: Gradient Vanishing of Negative Samples.
- Repeated Exposure: Inverse-Propensity Weighting
- Real-time Online Learning: tackling the delayed feedback issue.
- Gaussian Process for Exploitation and Exploration.

$$L^{CP} = \alpha L_{BCE} + (1-\alpha) L_{RankNet}$$
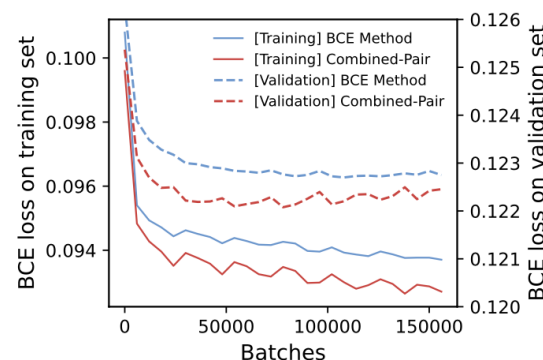
(a) BCE Loss Dynamics along epochs. (b) Loss Landscapes.
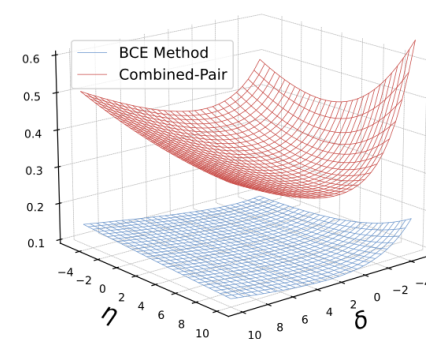
Outline

Three main challenges:

- **C1: Priors for Representation**

- **C2: Dimensional Collapse**

- **C3: Interest Entanglement**

Others

- Model Training

- Analysis Tools

Analysis Tools

Feature Correlation

Ground truth: mutual information.

Definition 1 (Category-wise Target-aware Correlation (CTC)). CTC is defined as the mutual information [18, 26] between behaviors with category c_i whilst occurring at position p: $X_{C(X)=c_i \wedge P(X)=p}$ and the user response label [1] on the target item with category c_t: $\mathcal{Y}_{C(Y)=c_t}$. Formally,

$$\text{Cor} = \text{MI}\left(X_{C(X)=c_i \wedge P(X)=p}, \mathcal{Y}_{C(Y)=c_t}\right), \quad (1)$$

Learned: $\overline{\text{Cor}}(f(X_i), \boldsymbol{e}_i, \boldsymbol{v}_t) = e^z \cdot \|\boldsymbol{r}\|_2.$
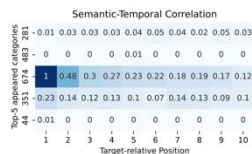
(a) Ground truth Correlation
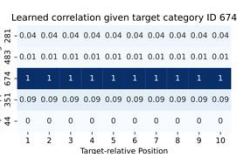
(b) DIN's learned Correlation
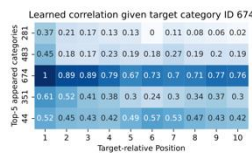
(c) SASRec's learned Correlation
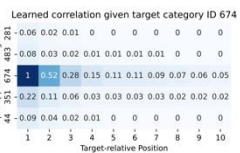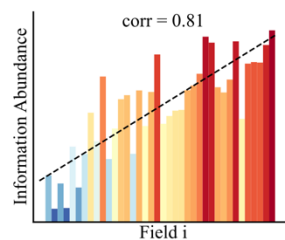
(d) TIM's learned Correlation

Dimensional Collapse

Definition 3.1 (Information Abundance). Consider a matrix $\boldsymbol{E} \in \mathbb{R}^{D \times K}$ and its singular value decomposition $\boldsymbol{E} = \boldsymbol{U\Sigma V} = \sum_{k=1}^{K} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\top}$, then the *information abundance* of $\boldsymbol{E}$ is defined as
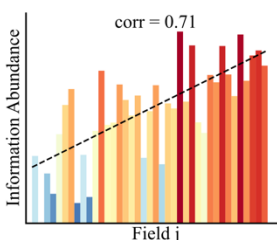
$$\text{IA}(\boldsymbol{E}) = \frac{\|\boldsymbol{\sigma}\|_1}{\|\boldsymbol{\sigma}\|_\infty},$$
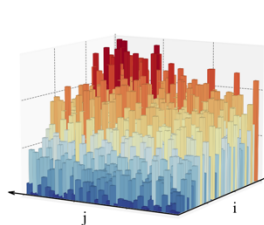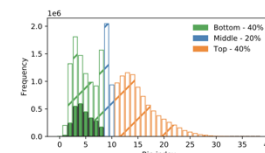

(a) $\text{IA}(\boldsymbol{E}_i^{\rightarrow j})$.

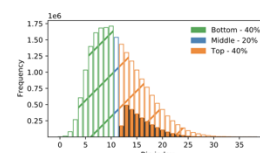(b) $\sum_{j=1}^{N} \text{IA}(\boldsymbol{E}_i^{\rightarrow j})$.

(c) $\sum_{i=1}^{N} \text{IA}(\boldsymbol{E}_i^{\rightarrow j})$.
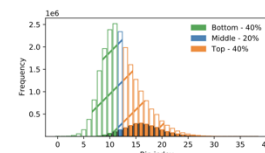
Interest Entanglement

1. Select a set of contradictory user item pairs whose embedding distance are among the top-40% regarding single task A embedding the bottom-40% regarding single B embedding .
2. Plot the distance distribution of these pairs regarding embeddings of MTL models.
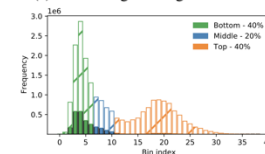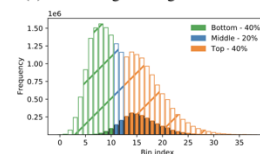

(a) Embeddings of Single Task Like
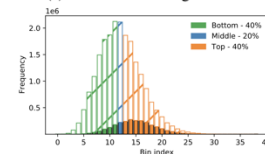
(b) Embeddings of Single Task Finish

(c) Shared-Embeddings in PLE

(d) Like-Specific Embeddings in STEM

(e) Finish-Specific Embeddings in STEM

(f) Shared Embeddings in STEM