WaySeq: User Sequence Modeling for Sponsored Product Click & Conversion Prediction

Production Lessons from Wayfair

AdKDD 2026 Workshop

Optimizing Sponsored Products Ranking at Wayfair

BASELINE RANKER

Multi-Task Deep Neural Network

- Predicts both immediate engagement (CTR) and downstream conversion (CVR) , which are used in the final ranking formulas.
- Integrates customer, product, and request context features.
- Interactions modeled effectively by crossing layers.

Limitation: Relies primarily on aggregate statistics and static features.

THE GOAL

Sequence-Aware Personalization

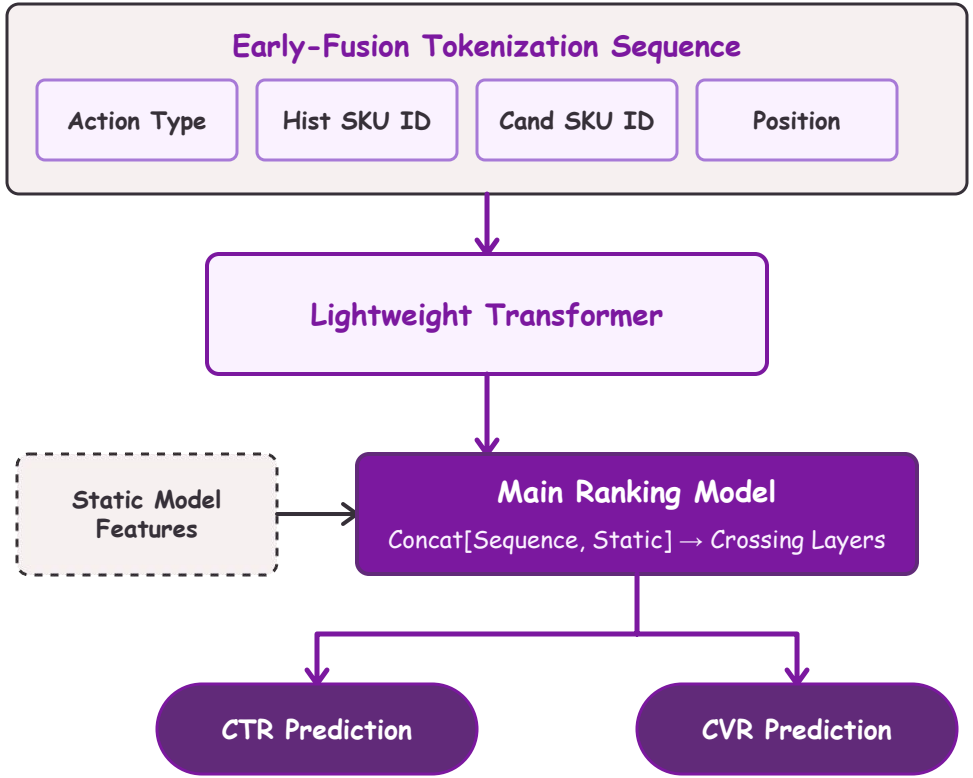
- Processes recent user interactions directly to capture real-time, short-term user intent.
- Enforces strict auction-time probability calibration.
- Meets highly demanding, real-time ad serving latency constraints.

Goal: Unlocking deeper personalization at scale.

WaySeq Integrates Sequence-Aware Multi-Task Ranking

First Production Sequence-Aware Framework

- **Augments Baseline:** Encodes user's multi-action behavior sequence alongside the candidate SKU being scored. Example actions: click, add to cart, favorite etc.
- **Early-Fusion Tokenization:** Formulates sequences by combining action type, historical SKU, candidate SKU, and position.
- **Transformer Backbone:** Processes tokens through a lightweight transformer.
- **Main Model Injection:** Injected directly into crossing layers to complement existing static features.



Learnings from Extensive Ablation Studies

1. Sequence Context Scaling

Scaling sequence context acts as a foundational driver for model comprehension, delivering consistent offline improvements:

- Longer sequence helps capture broader user intent windows.
- Minimum sequence length = 1 helps capture immediate customer interest within session.

2. SKU Representation

Replacing baseline embeddings with enriched concatenated representations significantly improved key offline ranking metrics:

- Combines visual features (visual embedding system) with deep semantic context (semantic embedding).
- Creates highly robust, multi-modal ad item representations.

3. Sequence Summarization

Candidate-Aware Summarization:

Highly effective soft-retrieval mechanism conditioned on the ad candidate with action bias.

- Candidate aware pooling:
  - ∑ (α_i * h_i)
  - **h_i: raw transformer output**
  - a_i: **attention Weights:** softmax over similarity between candidate SKU and historical events
- Action bias:
  - inject an action-type prior into the Candidate-Aware summary scores.
  - Multi-action histories contain asymmetric densities

Heavier Behavior-Aware Attention: Each action has their own Q,K,V → Added serving cost without consistent ranking-quality gains, possibly due to severe action imbalance between low-intent and high-intent behaviors.

WaySeq Delivers Significant Live A/B Business Gains

Live Business Performance

Our 5-week live test of WaySeq yielded exceptionally strong business results across key funnels:

- **+4.7%** lift in primary CTR, driving massive customer engagement.
- Customer CVR increased by **+0.8%**, streamlining the path to purchase.
- Ad spend grew by **+1.7%** as advertisers saw better relevance.
- Decreased overall CPC by **-2.8%**, improving merchant ROI.
- Successfully maintained low serving latency well under target SLA thresholds.

+4.7%

CTR Lift

Significant boost in click-through rate across core ad recommendations.

+5.9%

Orders / Impression

More orders generated per impression, enhancing marketplace efficiency.

+2.3%

ROAS Gain

Improved return on ad spend driven by highly targeted model relevance.

Practical Takeaways and Future Directions

The interface between user history, candidates, and production constraints matter more.

KEY LESSONS

Reliable Gains vs. Heavy Attention Failure

- **Broader behavior coverage** and stronger SKU representations provide consistent and reliable system gains.
- **Candidate-aware mechanism** ensures real-time context is highly preserved during inference.
- **Limitation:** Heavy action-specific attention fails under conditions of severe action imbalance.

FUTURE WORK

Unlocking New Sequence Horizons

- **Ultra-long behavior sequences:** Extending memory capacity via retrieval, lighter attention design etc.
- **Multimodal signals:** Enriching sequence tokens with visual and text signals from deep catalogs.
- **Generative pretraining:** Exploring innovative generative pretraining objectives for robust sequence modeling.

Q&A