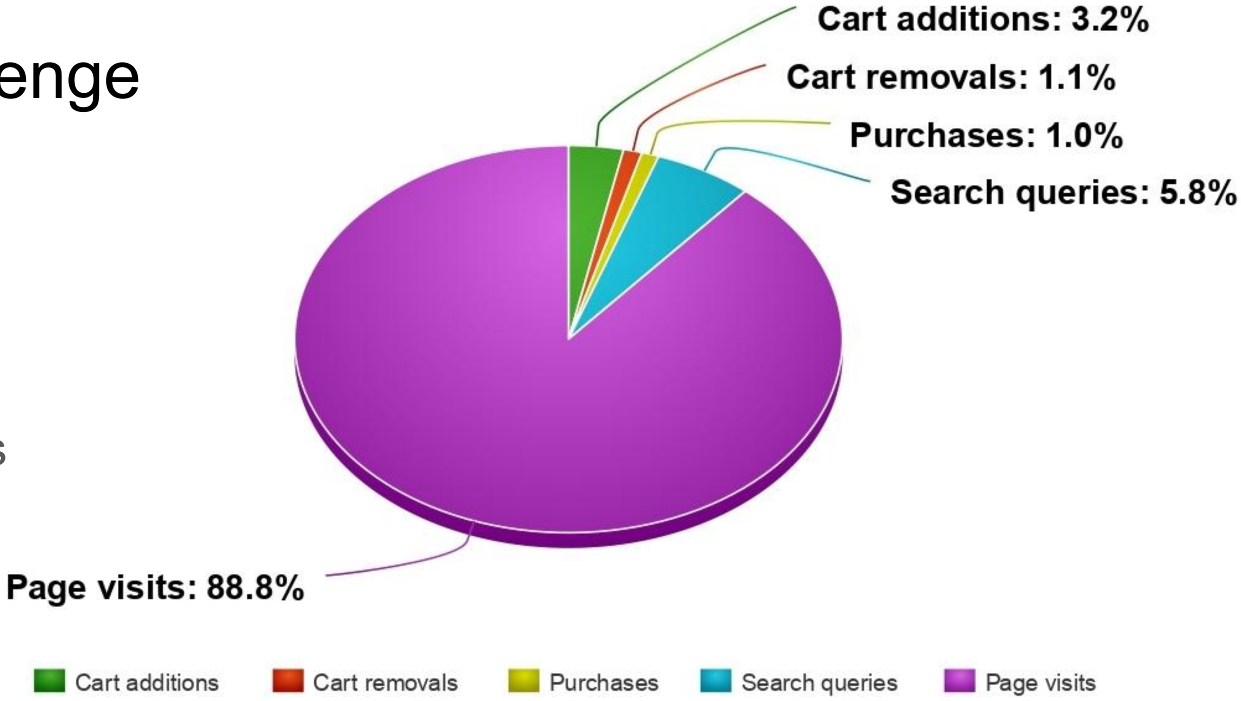
# Blending Sequential Embeddings, Graphs, and Engineered Features:
## 4th Place Solution in RecSys Challenge 2025

Sergei Makeev, Alexandr Andreev, Vladimir Baikalov,
Vladislav Tytskiy, Aleksei Krasilnikov and Kirill Khrylchenko

# About the Challenge

Data:

- 18.7M clients
- 1.3M SKUs
- 12.7M web pages



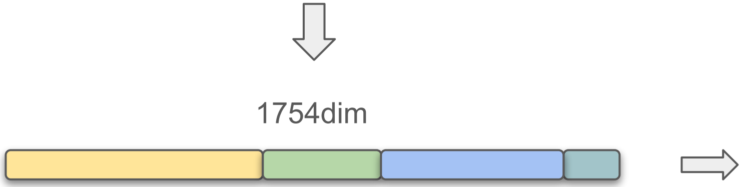

Goal:

- **Universal Behavioral Profile** - single embedding for churn, propensity category and propensity SKU prediction for each user

# Solution overview

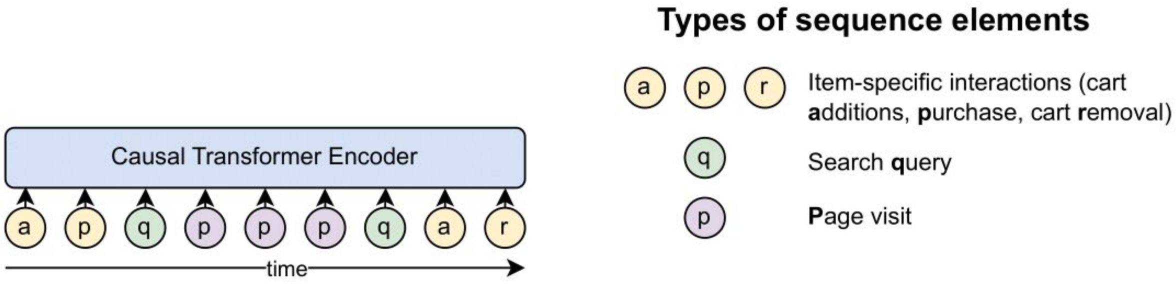
We stacked embeddings from four models:

1. **Sequential encoder** 1024dim
2. **TwHIN GNN** 256dim
3. **Handcrafted features** 346dim
4. **Deep Cross Network** 128dim

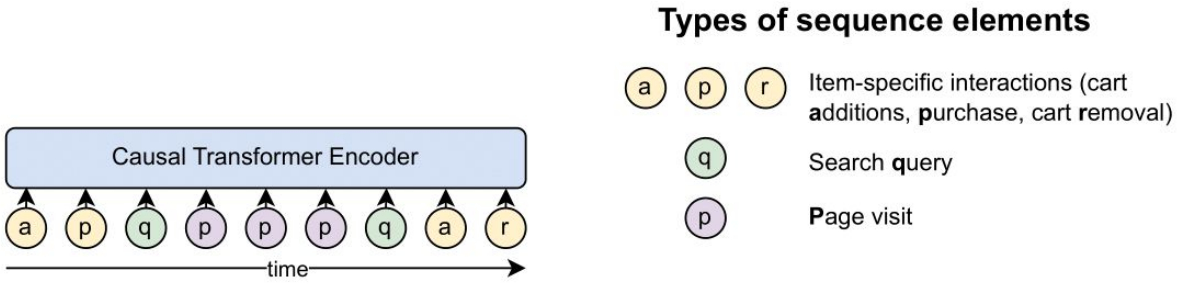



Universal Behavioral Profile

# Sequential Encoder



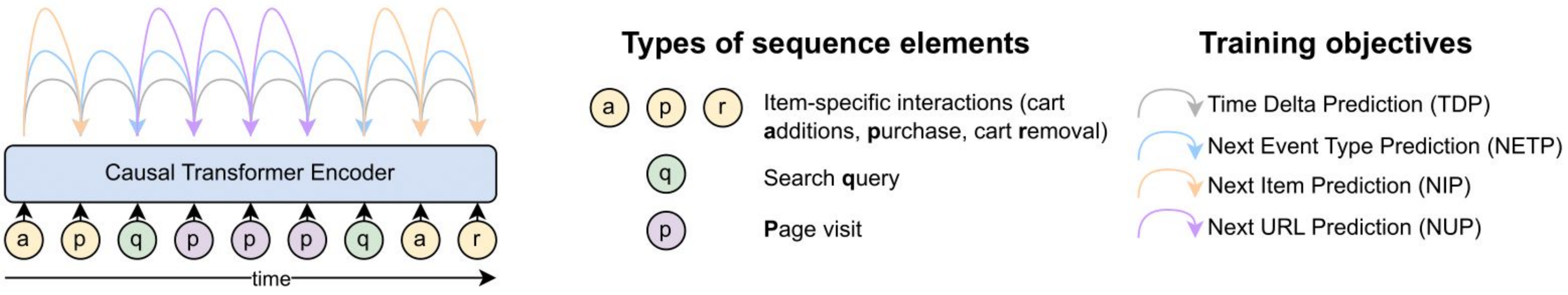
# Sequential Encoder




Types of user interactions:

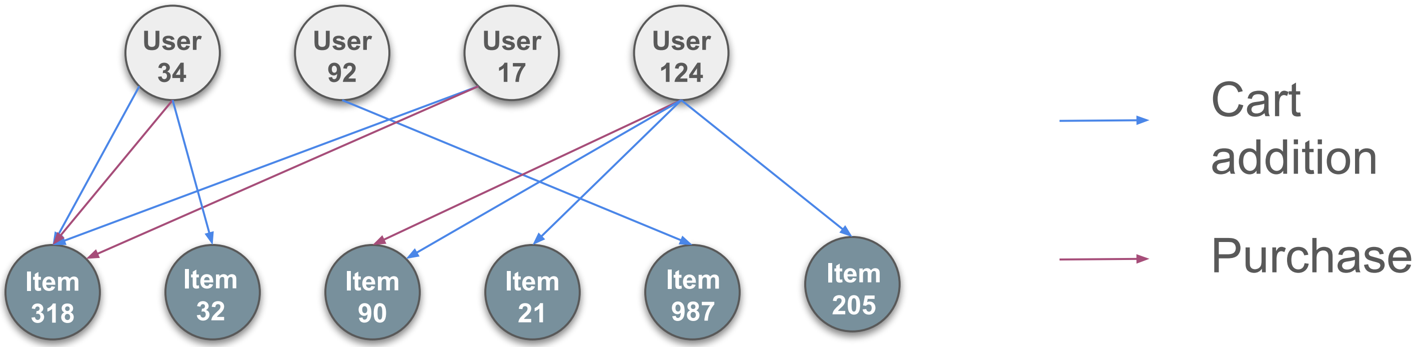
1. **Cart additions, cart removals, purchases**: timestamp, client ID, SKU ID, price, name, category ID
2. **Search queries**: timestamp, client ID, query
3. **Page visits**: timestamp, client ID, URL ID

# Sequential Encoder



# TwHIN

- Learnable embeddings for 1.5M most active users (including 1M users for evaluation)
- Item encoder from Sequential Encoder
- Cart additions and purchases as edges




Learnable parameters: 1.5M user embeddings + 2 relation embeddings
+ item encoder parameters

# TwHIN

**source embedding** = normalize(i-th user embedding)
**target embedding** = normalize(j-th item embedding + relation embedding ( ))

**probability of an edge:** $p(e) = \sigma(\text{source embedding} \times \text{target embedding})$

$$\mathcal{L}_{\mathrm{TwHIN}}(\mathcal{B}) = -\frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}}\left(\log(p(e_{ii})) + \sum_{j\in\mathcal{B},j\neq i}\log(1-p(e_{ij}))\right),$$

$\mathcal{B}$ is a training batch

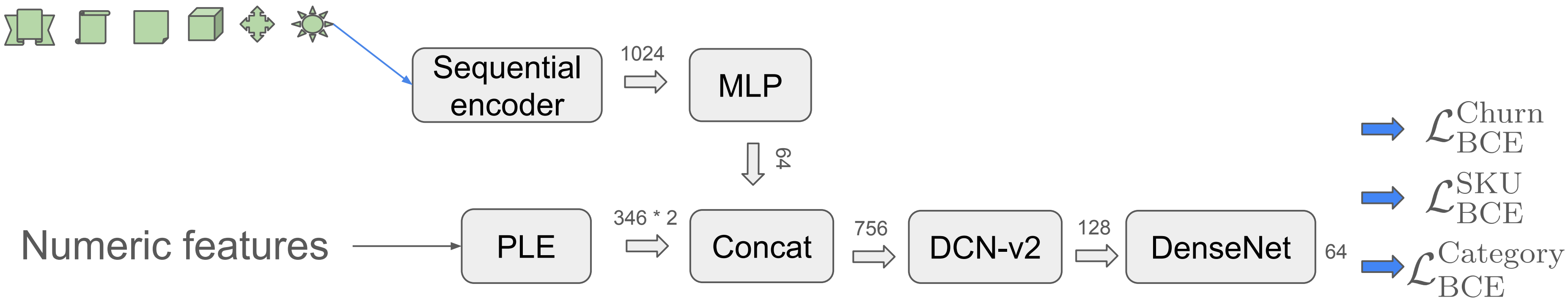
# Handcrafted Features

1. Basic Features:
    a. Interaction counters, price averages, cumulative statistics, frequencies, etc. $\Rightarrow$ $41 \times 5 = 205$
    b. Calculated over 7, 14, 28, 60, inf days
2. Cluster-based features:
    a. Aggregates over item name clusterings
    b. Aggregates over price groups $\Rightarrow$ $141$
    c. Exponential weighting

For Universal Behavioral Profile, we used a standardized vector of numeric features

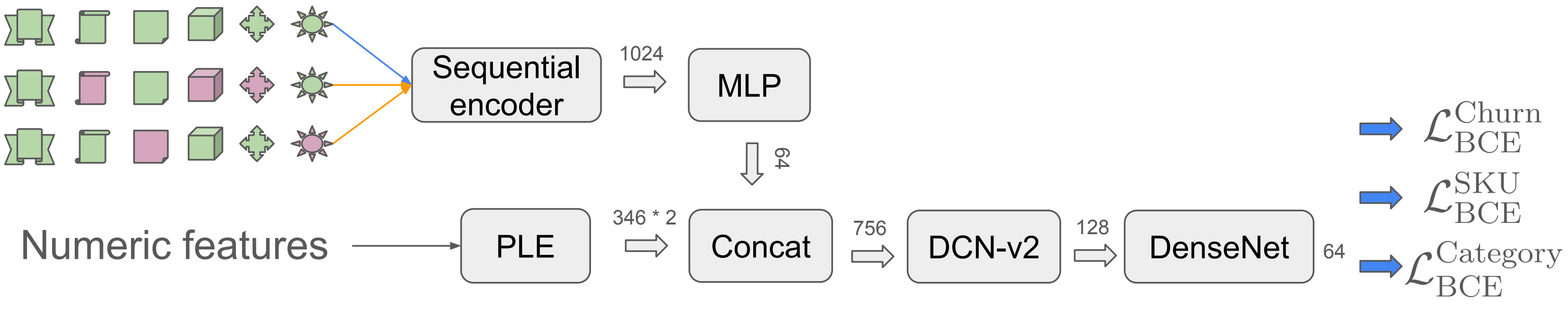
# Deep Cross Network

- Models feature interactions
- Incorporates Contrastive Learning



# Deep Cross Network

- Models feature interactions
- Incorporates Contrastive Learning




$$\mathcal{L}_{\mathrm{SSL}}^{\mathrm{CL}}(p'_u, p''_u) = -\log \frac{e^{\langle p'_u, p''_u \rangle / \tau}}{e^{\langle p'_u, p''_u \rangle / \tau} + \sum_{n \in \mathcal{N}} e^{\langle p'_u, p''_n \rangle / \tau}},$$ $\mathcal{N}$ - in-batch negatives

# Key ablation insights

- All four models contribute meaningfully
- For sequential encoder, NIP is the most critical training objective
- Scaling sequential encoder size yields improvements
- Both purchases and cart additions are important for TwHIN GNN
- Contrastive learning benefits sparse targets