# E-commerce Session-based Recommendation System Design

Name: Yuanjie Yan

Student ID: z5484456

## 1. Scope

### 1.1 Background and Objectives

Many e-commerce users prefer anonymous browsing without creating accounts, yet still expect personalized recommendations. This project aims to provide accurate recommendations for these "guest" users by analyzing their session behavior patterns. Using the Diginetica dataset from CIKM Cup 2016[1], we predict which products users are likely to view or purchase next based on their current session activities.

### 1.2 Target Users

1. Anonymous browsers in e-commerce sites
2. Users who haven't logged in
3. New visitors exploring products
4. Privacy-conscious shoppers

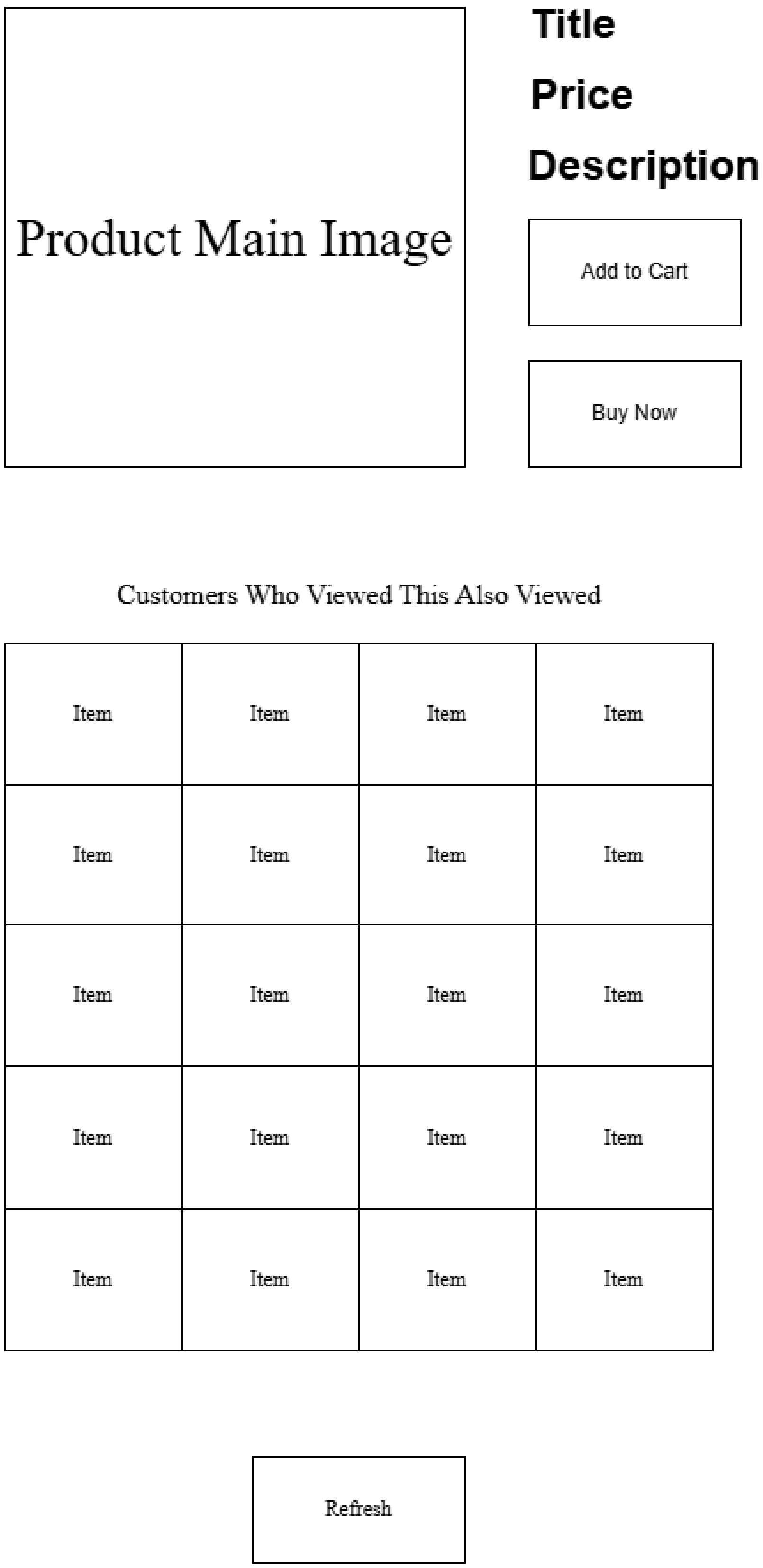
### 1.3 Recommendation Interface

Web-based design with 20 product recommendations per batch. Simple and direct interface:

1. "Customers who viewed this also viewed" section below product detail page
2. 4×5 grid display layout
3. Product details shown on mouse hover

4. One-click add to cart support

## 1.4 Interface Layout Mockup:



## 1.5 System Update Strategy

### 1.5.1 Real-time Updates:

- User behaviors collected in real-time through message queues
- Each session maintains a feature vector with dynamic updates

### 1.5.2 Model Updates:

- Batch retraining daily between 2-4 AM
- A/B testing to validate new model performance

### 1.5.3 Cold Start Handling:

- New items: Recommend based on category popularity
- New sessions: First 3 clicks recommend diverse popular items
- Quick strategy adjustment through user feedback

## 1.6 Business Value

The recommendation system generates revenue through:

- **Conversion Optimization**: Increase view-to-purchase conversion rate
- **Session Extension**: Keep users engaged longer on the platform
- **Cross-selling**: Recommend complementary products based on search context
- **Inventory Management**: Promote items based on stock levels and margins

# 2. Dataset(s)

## 2.1 Diginetica Dataset

CIKM Cup 2016 e-commerce search engine data including:

- 1.24 million user interactions
- 184,047 unique items
- 204,789 unique sessions
- Time span: 5 months

## 2.2 Data Structure

Main fields:

- SessionID - Unique session identifier
- UserID - Anonymized user ID (only 30% of interactions have this)
- ItemID - Product identifier
- TimeFrame - When the interaction occurred
- EventType - View or purchase event

## 2.3 Data Processing

1. **Session Segmentation**: New session after 30 minutes of inactivity
2. **Filtering Rules**:
   - Remove sessions with only 1 interaction
   - Remove items appearing less than 5 times
   - Filter test sessions occurring before training period
3. **Dataset Split**:
   - Training set: First 4 months
   - Validation set: Week 3 of month 5
   - Test set: Last week of month 5

## 2.4 Data Advantages

This dataset offers several benefits:

- Contains both view and purchase events (not just clicks)
- Includes search query context
- More realistic e-commerce scenario with actual conversions
- Anonymized but includes some user IDs for hybrid approaches

## 2.5 Data Limitations

The Diginetica dataset, while comprehensive, has several limitations that may impact our design and evaluation:

Limited User Identifiers: Only ~30% of interactions include a valid UserID, constraining the effectiveness of hybrid collaborative - filtering methods that rely on persistent user histories.

Short Time Span: The data covers just five months, which may not capture seasonal effects or long - term behavioral shifts.

Pre-processing Bias: The dataset was pre-cleaned and pre-split by the CIKM Cup organizers, potentially introducing selection biases and reducing its representativeness of raw e-commerce traffic.

# 3. Method(s)

## 3.1 Solution 1: Basic SASRec

Using Transformer for sequential recommendation, with the core idea being to learn user click patterns through self-attention mechanism.

**Why choose this**:

- Captures long-range dependencies
- Fast training speed
- Proven effectiveness

## 3.2 Solution 2: Incorporating Temporal Information

User click speed has meaning:

- Rapid consecutive clicks → Comparing products
- Long intervals → Interest shift

Therefore, encode time intervals into the model:

- 0-10 seconds: Closely related
- 10-60 seconds: Generally related
- Over 1 minute: Weakly related

## 3.3 Solution 3: Hierarchical Modeling

Split long sequences into segments:

1. First examine local behavior (e.g., clicked 5 phones consecutively)
2. Then examine overall path (phone → earphones → phone case)

This captures both details and global patterns.

## 3.4 Hybrid Strategy

Different methods for different session lengths:

- Short sessions (<5 clicks): Popular items and similar products
- Medium sessions (5-10 clicks): Sequential model and collaborative filtering
- Long sessions (>10 clicks): Rely entirely on sequential model
- Knowledge-based Fallback: If model confidence falls below a threshold (e.g. top - 20 scores < 0.3), revert to an item based collaborative filtering approach, recommending the most popular or co-viewed items as a baseline.

# 4. Evaluation

## 4.1 Performance Metrics

1. **Recall@20** - Percentage of target items appearing in top-20 recommendations
2. **MRR@20** - Mean Reciprocal Rank for purchase predictions
3. **NDCG@20** - Normalized Discounted Cumulative Gain
4. **Diversity** - Avoid recommending only similar items
5. **Purchase Precision** - Accuracy of purchase predictions vs. view predictions

## 4.2 Performance Requirements

- Response time < 100ms
- Single machine QPS > 1000
- Memory usage < 4GB

## 4.3 User Testing

Recruit team members (4~5 people) as testers using the web-based mockup interface:

- Each tester completes 3 representative shopping tasks on the simulated UI mockup
- Record click behavior and completion time
- Post-test questionnaire and interview
- All feedback is gathered on the mockup (simulated user interface) before backend integration

Sample test tasks:

1. "Buy a mouse for roommate"

2. "Check out recent digital products"
3. "Just browse to kill time"

## 4.4 Metric Trade-off Strategy

### 4.4.1 Priority Ranking:

1. MRR weight 40% - Most important, directly affects CTR
2. Response latency must be <100ms - Hard constraint
3. Recall weight 30% - Ensure coverage
4. Diversity weight 20% - Avoid recommendation fatigue
5. User satisfaction 10% - Qualitative reference

Ultimately, model selection will prioritize maximizing MRR first, then incorporate Diversity as the secondary criterion to balance relevance with variation.

### 4.4.2 Conflict Resolution:

- When Recall conflicts with Diversity, maintain minimum diversity threshold of 0.3
- Dynamically adjust weights based on weekly A/B test results
- Negative user feedback triggers immediate strategy adjustment

## 4.5 Project Highlights

1. **Real Purchase Data**: Unlike click-only datasets, includes actual purchase behavior
2. **Search Context Integration**: Leverages query information for better recommendations
3. **Practical Evaluation**: Separate metrics for view and purchase predictions
4. **Industry Relevance**: Based on real e-commerce search engine data

## 4.6 Expected Challenges

1. **Sparse Purchase Data**: Only 6.9% of sessions contain purchases
2. **Anonymous Users**: 70% of interactions lack user IDs
3. **Search Query Integration**: How to effectively use query information

## 4.7 Backup Plans

If model performance is poor:

- Focus on high-confidence sessions with purchases
- Use query similarity for cold start
- Implement ensemble methods combining view and purchase models

If computational resources limited:

- Use subset of most active items
- Implement efficient sparse matrix operations
- Apply dimensionality reduction techniques

# Reference

[1] Q. Liu et al., "CIKM Cup 2016 Track 2 Diginetica Dataset," CIKM 2016, Indianapolis, USA. Available: https://cikm2016.cs.iupui.edu/cikm-cup/