05_content_hybrid_dev

April 20, 2025

[1]:
```
# Cell [1] - Imports and Setup
import pandas as pd
import numpy as np
import torch
import torch.nn as nn
import torch.optim as optim
from torch.utils.data import DataLoader
from sklearn.model_selection import train_test_split
from tqdm.notebook import tqdm
import matplotlib.pyplot as plt
import sys
from pathlib import Path
import math
import seaborn as sns


# Add project root to sys.path
project_root = Path.cwd().parent
if str(project_root) not in sys.path:
    sys.path.append(str(project_root))

# Import project modules
from src import config
from src.data.dataset import HybridDataset, create_mappings_and_unique_ids #␣
 ↪<<< Import HybridDataset
from src.models.hybrid import HybridNCF # Import the Hybrid model

# Set device
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
print(f"Using device: {device}")

# Set display options
pd.set_option('display.max_columns', 100)
pd.set_option('display.max_rows', 100)
sns.set_style("whitegrid")
print("Setup complete. Modules imported.")
```

```
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
```

```
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Using device: cpu
Setup complete. Modules imported.
```

[2]:
```python
# Cell [2] - Load Processed Data
interactions_path = config.PROCESSED_DATA_DIR / "interactions_final.parquet"
users_path = config.PROCESSED_DATA_DIR / "users_final.parquet"
items_path = config.PROCESSED_DATA_DIR / "items_final.parquet"

try:
    interactions_df = pd.read_parquet(interactions_path)
    users_features_df = pd.read_parquet(users_path) # User features (not␣
 ↪directly used by HybridNCF yet)
    item_features_df = pd.read_parquet(items_path)   # Item features␣
 ↪(presentation_id as col)
    print("Processed data loaded successfully.")
    print(f"Interactions shape: {interactions_df.shape}")
    print(f"Users shape: {users_features_df.shape}")
    print(f"Items shape: {item_features_df.shape}")

    # IMPORTANT: Ensure item_features_df has 'presentation_id' as index for the␣
 ↪dataset
    if 'presentation_id' in item_features_df.columns:
        item_features_df = item_features_df.set_index('presentation_id')
        print("Set 'presentation_id' as index for item_features_df.")
    elif item_features_df.index.name != 'presentation_id':
         raise ValueError("item_features_df must have 'presentation_id' as␣
 ↪index or column.")

    # Store item feature dimension
    ITEM_FEATURE_DIM = item_features_df.shape[1]
    print(f"Item feature dimension: {ITEM_FEATURE_DIM}")


except FileNotFoundError as e:
    print(f"Error loading processed files: {e}")
    print("Please ensure the preprocessing pipeline (run_preprocessing.py) has␣
 ↪run successfully.")
    raise e
except Exception as e:
    print(f"An unexpected error occurred during loading: {e}")
    raise e

print("\nInteractions Head:\n", interactions_df.head(3))
print("\nItem Features Head:\n", item_features_df.head(3))
```

```python
# Drop constant columns from item features if they exist (e.g., all zeros)
# These provide no information for the MLP
const_cols = item_features_df.columns[item_features_df.nunique() <= 1]
if len(const_cols) > 0:
    print(f"\nDropping constant item feature columns: {const_cols.tolist()}")
    item_features_df = item_features_df.drop(columns=const_cols)
    ITEM_FEATURE_DIM = item_features_df.shape[1]
    print(f"Updated item feature dimension: {ITEM_FEATURE_DIM}")
```

```
Processed data loaded successfully.
Interactions shape: (28466, 7)
Users shape: (25364, 9)
Items shape: (22, 22)
Set 'presentation_id' as index for item_features_df.
Item feature dimension: 21

Interactions Head:
    id_student presentation_id  total_clicks  interaction_days  \
0        6516       AAA_2014J          2791               159
1        8462       DDD_2013J           646                56
2        8462       DDD_2014J            10                 1

   first_interaction_date  last_interaction_date  implicit_feedback
0                     -23                    269           7.934513
1                      -6                    118           6.472346
2                      10                     10           2.397895

Item Features Head:
                 module_presentation_length  vle_prop_dataplus  \
presentation_id
AAA_2013J                               268           0.018957
AAA_2014J                               269           0.019802
BBB_2013J                               268           0.000000

                vle_prop_dualpane  vle_prop_externalquiz  vle_prop_folder  \
presentation_id
AAA_2013J                     0.0                    0.0              0.0
AAA_2014J                     0.0                    0.0              0.0
BBB_2013J                     0.0                    0.0              0.0

                vle_prop_forumng  vle_prop_glossary  vle_prop_homepage  \
presentation_id
AAA_2013J               0.071090           0.009479           0.004739
AAA_2014J               0.029703           0.009901           0.004950
BBB_2013J               0.059190           0.003115           0.003115

                vle_prop_htmlactivity  vle_prop_oucollaborate  \
presentation_id
```

```
AAA_2013J                        0.0                0.009479
AAA_2014J                        0.0                0.009901
BBB_2013J                        0.0                0.006231

                vle_prop_oucontent  vle_prop_ouelluminate  vle_prop_ouwiki  \
presentation_id
AAA_2013J                 0.322275                    0.0              0.0
AAA_2014J                 0.336634                    0.0              0.0
BBB_2013J                 0.009346                    0.0              0.0

                vle_prop_page  vle_prop_questionnaire  vle_prop_quiz  \
presentation_id
AAA_2013J                 0.0                     0.0       0.000000
AAA_2014J                 0.0                     0.0       0.000000
BBB_2013J                 0.0                     0.0       0.015576

                vle_prop_repeatactivity  vle_prop_resource  \
presentation_id
AAA_2013J                           0.0           0.450237
AAA_2014J                           0.0           0.460396
BBB_2013J                           0.0           0.735202

                vle_prop_sharedsubpage  vle_prop_subpage  vle_prop_url
presentation_id
AAA_2013J                     0.000000          0.028436      0.085308
AAA_2014J                     0.000000          0.029703      0.099010
BBB_2013J                     0.003115          0.118380      0.046729
```

[3]:
```python
# Cell [3] - Create Mappings and Hybrid Dataset

USER_COL = 'id_student'
ITEM_COL = 'presentation_id' # This is the index name now in item_features_df

# Create mappings from original IDs based on interactions data
user_id_map, item_id_map, unique_users, unique_items =␣
 ↪create_mappings_and_unique_ids(
    interactions_df, USER_COL, ITEM_COL
)
n_users = len(unique_users)
n_items = len(unique_items)

print(f"Number of unique users: {n_users}")
print(f"Number of unique items: {n_items}")

# Ensure item_features_df covers all items in the map
items_in_map_set = set(item_id_map.keys())
items_in_features_set = set(item_features_df.index)
```

```
if not items_in_map_set.issubset(items_in_features_set):
    missing = items_in_map_set - items_in_features_set
    raise ValueError(f"{len(missing)} items from interactions are missing in␣
 ↪item_features_df. E.g.: {list(missing)[:5]}")
if items_in_features_set != items_in_map_set:
     print(f"Warning: {len(items_in_features_set - items_in_map_set)} items in␣
 ↪features_df are not in interactions_df.")


# Split interactions for train/validation (simple random split for dev)
train_interactions, val_interactions = train_test_split(
    interactions_df, test_size=0.1, random_state=config.RANDOM_SEED
)

# Create Hybrid Datasets
# Pass the (potentially column-filtered) item_features_df
train_dataset_hybrid = HybridDataset(
    interactions_df=train_interactions,
    item_features_df=item_features_df,
    all_item_ids=item_features_df.index.tolist(),
    user_id_map=user_id_map,
    item_id_map=item_id_map,
    user_col=USER_COL,
    item_col=ITEM_COL,
    num_negatives=4
)

val_dataset_hybrid = HybridDataset(
    interactions_df=val_interactions,
    item_features_df=item_features_df,
    all_item_ids=item_features_df.index.tolist(),
    user_id_map=user_id_map,
    item_id_map=item_id_map,
    user_col=USER_COL,
    item_col=ITEM_COL,
    num_negatives=0
)

# Create DataLoaders
BATCH_SIZE = 512 # Adjust based on memory
train_loader_hybrid = DataLoader(train_dataset_hybrid, batch_size=BATCH_SIZE,␣
 ↪shuffle=True, num_workers=4, pin_memory=True)
val_loader_hybrid = DataLoader(val_dataset_hybrid, batch_size=BATCH_SIZE * 2,␣
 ↪shuffle=False, num_workers=4, pin_memory=True)

print(f"\nHybrid DataLoaders created. Batch size: {BATCH_SIZE}")
```

```
# Test a batch
print("\nSample batch from Hybrid Train DataLoader:")
for batch in train_loader_hybrid:
    users, items, feats, labels = batch
    print(" Users shape:", users.shape)
    print(" Items shape:", items.shape)
    print(" Feats shape:", feats.shape) # Should be (BATCH_SIZE,␣
 ↪ITEM_FEATURE_DIM)
    print(" Labels shape:", labels.shape)
    break
```

```
Number of unique users: 25364
Number of unique items: 22
Preparing HybridDataset…
Item features array created shape: (22, 21)
Dataset contains 25619 positive interactions.
Generating 4 negative samples per positive.
HybridDataset preparation complete.
Preparing HybridDataset…
Item features array created shape: (22, 21)
Dataset contains 2847 positive interactions.
HybridDataset preparation complete.

Hybrid DataLoaders created. Batch size: 512

Sample batch from Hybrid Train DataLoader:
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
 Users shape: torch.Size([512])
 Items shape: torch.Size([512])
 Feats shape: torch.Size([512, 21])
 Labels shape: torch.Size([512])
```

[4]:
```
# === New Cell: Instantiate and Train HybridNCFRecommender ===
from src.models.hybrid import HybridNCFRecommender # Import the wrapper

# Define hyperparameters for the wrapper
```

```
CF_EMBEDDING_DIM_WRAP = 32
CONTENT_EMBEDDING_DIM_WRAP = 16
CONTENT_ENCODER_HIDDEN_WRAP = [32, 16]
FINAL_MLP_LAYERS_WRAP = [64, 32, 16]
DROPOUT_WRAP = 0.2
LEARNING_RATE_WRAP = 0.001
EPOCHS_WRAP = 10 # Match previous training
WEIGHT_DECAY_WRAP = 1e-5
BATCH_SIZE_WRAP = 512 # Match previous batch size
NUM_NEGATIVES_WRAP = 4 # Match previous negative samples

print("\n--- Initializing HybridNCFRecommender ---")
hybrid_recommender = HybridNCFRecommender(
    user_col=USER_COL, # Defined earlier
    item_col=ITEM_COL, # Defined earlier
    cf_embedding_dim=CF_EMBEDDING_DIM_WRAP,
    content_embedding_dim=CONTENT_EMBEDDING_DIM_WRAP,
    content_encoder_hidden_dims=CONTENT_ENCODER_HIDDEN_WRAP,
    final_mlp_layers=FINAL_MLP_LAYERS_WRAP,
    dropout=DROPOUT_WRAP,
    learning_rate=LEARNING_RATE_WRAP,
    epochs=EPOCHS_WRAP,
    batch_size=BATCH_SIZE_WRAP,
    num_negatives=NUM_NEGATIVES_WRAP,
    weight_decay=WEIGHT_DECAY_WRAP,
    device='auto'
)

# Train the model using the 'fit' method
# Pass interactions data AND the item features DataFrame
print("\n--- Training HybridNCFRecommender ---")
# Ensure interactions_df, item_features_df are defined and correct
# Fit on the full interactions data intended for this model instance
# hybrid_recommender.fit(train_interactions, item_features_df) # Option 1: Fit␣
 ↪on dev split
hybrid_recommender.fit(interactions_df, item_features_df)      # Option 2: Fit␣
 ↪on full data


print("\n--- HybridNCFRecommender Training Complete ---")
```

```
--- Initializing HybridNCFRecommender ---
Initialized HybridNCFRecommender
Using device: cpu

--- Training HybridNCFRecommender ---
```

```
Fitting HybridNCFRecommender…
 Mapped 25364 users and 22 items.
 Determined item feature dimension: 21
Initializing HybridNCF Network…
Initializing ContentEncoder Model…
 Input Dim: 21
 Hidden Dims: [32, 16]
 Output Embedding Dim: 16
 Layer Dimensions: [21, 32, 16, 16]
ContentEncoder Model Initialized.
HybridNCF Network Initialized.
Preparing HybridDataset…
Item features array created shape: (22, 21)
Dataset contains 28466 positive interactions.
Generating 4 negative samples per positive.
HybridDataset preparation complete.

--- Starting HybridNCF Training (10 Epochs) ---
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes

Epoch 1/10:   0%|          | 0/278 [00:03<?, ?it/s]

Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Epoch 1/10 - Training Loss: 0.6888
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
```

```
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes

Epoch 2/10:   0%|          | 0/278 [00:03<?, ?it/s]

Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Epoch 2/10 - Training Loss: 0.5201
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes

Epoch 3/10:   0%|          | 0/278 [00:03<?, ?it/s]

Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
```

```
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Epoch 3/10 - Training Loss: 0.4962
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes

Epoch 4/10:   0%|          | 0/278 [00:03<?, ?it/s]

Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Epoch 4/10 - Training Loss: 0.4899
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes

Epoch 5/10:   0%|          | 0/278 [00:03<?, ?it/s]

Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
```

```
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Epoch 5/10 - Training Loss: 0.4886
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes

Epoch 6/10:   0%|          | 0/278 [00:03<?, ?it/s]

Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Epoch 6/10 - Training Loss: 0.4864
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
```

```
Epoch 7/10:   0%|          | 0/278 [00:03<?, ?it/s]
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Epoch 7/10 - Training Loss: 0.4724
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Epoch 8/10:   0%|          | 0/278 [00:03<?, ?it/s]
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Epoch 8/10 - Training Loss: 0.4430
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
```

```
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes

Epoch 9/10:   0%|          | 0/278 [00:03<?, ?it/s]

Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Epoch 9/10 - Training Loss: 0.4113
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes

Epoch 10/10:   0%|          | 0/278 [00:03<?, ?it/s]

Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Loading .env from: /Users/mohit/Desktop/everything/ATLAS/Semester
4/Pinnacle/recsys_final/.env
Database URI configured: Yes
Epoch 10/10 - Training Loss: 0.3819
```

```
--- HybridNCF Training Finished ---

--- HybridNCFRecommender Training Complete ---
```

[5]:
```python
# Cell [7] - Evaluate Hybrid Model (Corrected WITH Wrapper)

import pandas as pd
import numpy as np
import torch
from pathlib import Path
import sys

# --- Ensure project root is in sys.path ---
project_root = Path.cwd().parent
if str(project_root) not in sys.path:
    sys.path.append(str(project_root))
# -----------------------------------------

# --- Import necessary functions/classes ---
from src import config
from src.data import preprocess  # For time_based_split
from src.evaluation.evaluator import RecEvaluator
# ------------------------------------------

# --- Ensure necessary variables/data are defined ---
# --- MODIFIED CHECK: Check for the wrapper instance ---
if 'hybrid_recommender' not in locals():
    raise NameError("HybridNCFRecommender instance 'hybrid_recommender' not␣
 ↪defined. Run the training cell first.")
# ----------------------------------------------------
if 'user_id_map' not in locals(): raise NameError("'user_id_map' not defined.␣
 ↪Run cell [3] first.")
if 'item_id_map' not in locals(): raise NameError("'item_id_map' not defined.␣
 ↪Run cell [3] first.")
# --- MODIFIED CHECK: Use item_features_df loaded earlier ---
if 'item_features_df' not in locals() or not isinstance(item_features_df, pd.
 ↪DataFrame):
    raise NameError("'item_features_df' not defined or not a DataFrame. Run␣
 ↪cell [2] first.")
if item_features_df.index.name != 'presentation_id': # Check index on the␣
 ↪correct variable
   raise ValueError("item_features_df (from cell [2]) must have␣
 ↪'presentation_id' as index.")
# ------------------------------------------------

# --- Load/Recreate the CORRECT Time-Based Train/Test Split ---
# (This section remains the same)
```

```
print("Loading/Recreating time-based split for evaluation...")
# Use interactions_df if already loaded, otherwise load it
if 'interactions_df' not in locals() or not isinstance(interactions_df, pd.
 ↪DataFrame):
    interactions_path_eval = config.PROCESSED_DATA_DIR / "interactions_final.
 ↪parquet"
    if not interactions_path_eval.exists():
         raise FileNotFoundError(f"Cannot find {interactions_path_eval}. Run␣
 ↪preprocessing first.")
    interactions_df_eval = pd.read_parquet(interactions_path_eval) # Use a␣
 ↪different name
else:
    interactions_df_eval = interactions_df # Use the one already loaded

# --- USE CONFIG VALUE ---
TIME_THRESHOLD = config.TIME_SPLIT_THRESHOLD
train_df_eval, test_df_eval = preprocess.time_based_split(
    interactions_df=interactions_df_eval,
    user_col='id_student',
    item_col='presentation_id',
    time_col='last_interaction_date',
    time_unit_threshold=TIME_THRESHOLD
)
print(f"Time-based split ready. Train: {train_df_eval.shape}, Test:␣
 ↪{test_df_eval.shape}")
# ----------------------------------------------------------------------

# --- Item Features are already loaded in item_features_df from cell [2] ---
print("Using item_features_df loaded in cell [2].")
# ----------------------------------------------------------------------

# --- NO WRAPPER NEEDED HERE - Model is already wrapped ---
# --- (Delete the old HybridEvaluatorWrapper class definition if it's still␣
 ↪here) ---
# --- (Delete the old hybrid_eval_wrapper = ... line if it's still here) ---
# ----------------------------------------------------------------------
print("Using the trained 'hybrid_recommender' instance directly.")


# --- Initialize Evaluator and Evaluate ---
if test_df_eval.empty:
    print("\nCannot evaluate Hybrid model: Test data (time-split) is empty.")
# --- Use item_features_df loaded from cell [2] ---
elif item_features_df.index.name != 'presentation_id':
    print("\nError: item_features_df must have 'presentation_id' set as index␣
 ↪for evaluator.")
```

```
else:
    print(f"\nInitializing evaluator with Train: {train_df_eval.shape}, Test:␣
 ↪{test_df_eval.shape}")
    hybrid_evaluator = RecEvaluator(
        train_df=train_df_eval,
        test_df=test_df_eval,
        # --- Pass the correctly loaded/indexed item_features_df ---
        item_features_df=item_features_df,
        user_col='id_student',
        item_col='presentation_id',
        k=config.TOP_K
    )

    # --- MODIFIED EVALUATION CALL: Use the wrapper instance ---
    print("\n--- Starting Evaluation of HybridNCFRecommender ---")
    # Use the 'hybrid_recommender' variable from the training cell
    # --- TRY REDUCING n_neg_samples FIRST ---
    print("Evaluating with n_neg_samples=20 for speed test...")
    hybrid_results = hybrid_evaluator.evaluate_model(hybrid_recommender,␣
 ↪n_neg_samples=20) # Reduced samples
    # ------------------------------------------------------

    print("\nHybrid Model Evaluation Results (n_neg_samples=20):") # Updated␣
 ↪print
    print(hybrid_results)

    # --- Optional: Run with full samples if the reduced one was fast enough ---
    # print("\n--- Starting Evaluation of HybridNCFRecommender␣
 ↪(n_neg_samples=100) ---")
    # hybrid_results_full = hybrid_evaluator.evaluate_model(hybrid_recommender,␣
 ↪n_neg_samples=100)
    # print("\nHybrid Model Evaluation Results (n_neg_samples=100):")
    # print(hybrid_results_full)
    # ------------------------------------------------------------------------

# ------------------------------------------------------------------
```

```
Loading/Recreating time-based split for evaluation…
Performing time-based split…
Original interactions shape: (28466, 7)
Splitting based on time threshold: last_interaction_date <= 250
 Initial train size: 22892, Initial test size: 5574
Filtered 4836 interactions from test set (users/items not in train).
Final Training set shape: (22892, 7)
Final Test set shape: (738, 7)
Users in Train: 20701, Users in Test: 731
Items in Train: 22, Items in Test: 13
```

```
Time-based split ready. Train: (22892, 7), Test: (738, 7)
Using item_features_df loaded in cell [2].
Using the trained 'hybrid_recommender' instance directly.

Initializing evaluator with Train: (22892, 7), Test: (738, 7)
Evaluator initialized with 22 unique candidate items.
Stored 20701 training interactions for filtering.
Prepared test data for 731 users.

--- Starting Evaluation of HybridNCFRecommender ---
Evaluating with n_neg_samples=20 for speed test…

--- Evaluating Model: HybridNCFRecommender ---
Total test users: 731. Evaluating 731 users known by the model.
Evaluating users:   0%|          | 0/731 [00:00<?, ?it/s]


--- Evaluation Results (K=10) ---
Precision@10: 0.0900
Recall@10: 0.8912
NDCG@10: 0.4698
n_users_evaluated: 731.0000
n_users_skipped: 0.0000
------------------------------

Hybrid Model Evaluation Results (n_neg_samples=20):
{'Precision@10': 0.09001367989056087, 'Recall@10': 0.8912448700410397,
'NDCG@10': 0.4697835952371101, 'n_users_evaluated': 731}
```