Improving YouTubeDNN via Attention

Qi Hong [1] Mingrui Chen [2] Supervisor: Prof. Min Chi [3*]

[1]Beijing Institute of Technology [2]Huazhong University of Science and Technology [3]North Carolina State University

Introduction

Recommendation system is becoming more and more common in people's daily life, especially in streaming media. Therefore, designing a useful recommendation system is very necessary.

The contribution of this poster is summarized as follows:

- Inspired by [1], we add an attention network to the basic YouTubeDNN model for user's history behavior sequences.
- We change the original MLP structure to get a better performance.

Datasets

MovieLens-1M data set, including 1,664 users' comments on 943 movies.

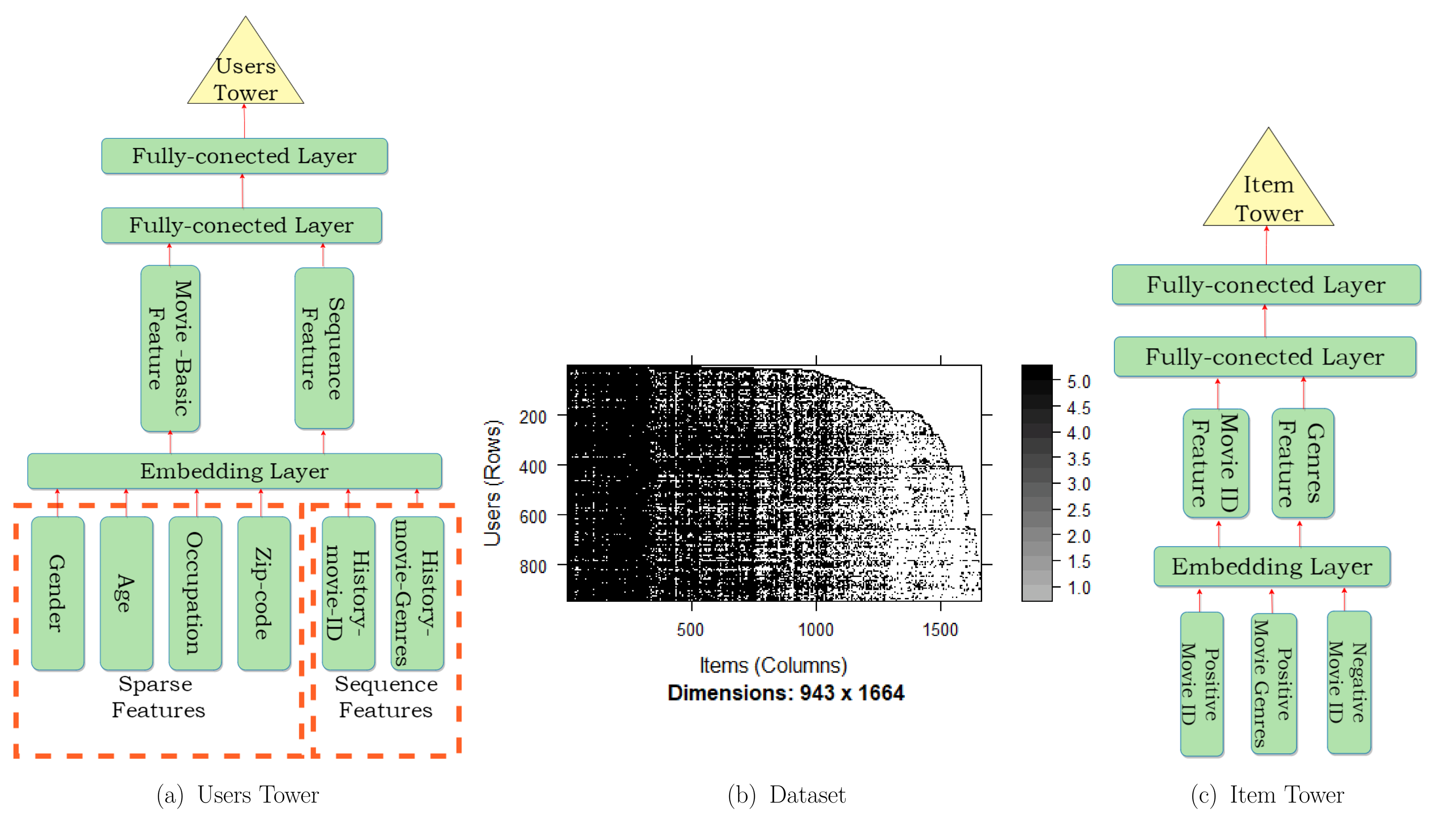

Figure 1. The MovieLens-1M Dataset

Two categories of features: *SparseFeature* and *SequenceFeature*.

- **Sparse Features**: Converted into a continuous integer value by **labelencoding** and outputs an **embedding vector**, such as user ID.
- **Sequence features**: Average each element after embedding and output an **embedding vector**, such as history.

Experiment Setup

Overall setup for both models are the same, i.e.,

- **User's MLP Layers:** 128+ReLU -> 256+ReLU -> 64+ReLU -> 32
- **Learning Rate:** 10^{-4}
- **Weight Decay:** 10^{-6}
- **Activation Unit's MLP Layers:** 256+ReLU -> 128+ReLU -> 1

Methods

1. **Based Model YouTubeDNN(YDNN)[2]**

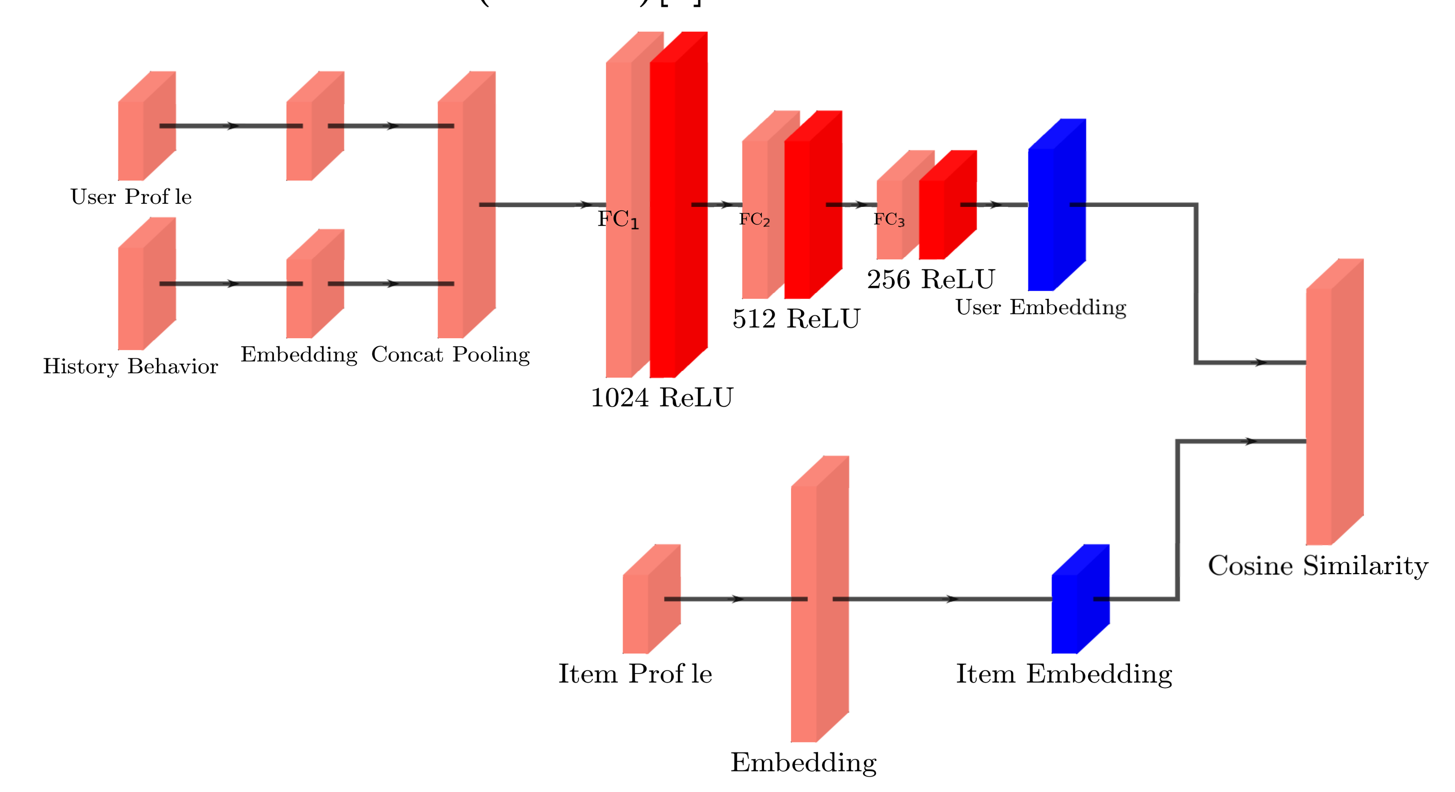

Figure 2. YouTubeDNN Structure

2. **Improved Model: YouTubeDNN With Attention(YDNNA)**

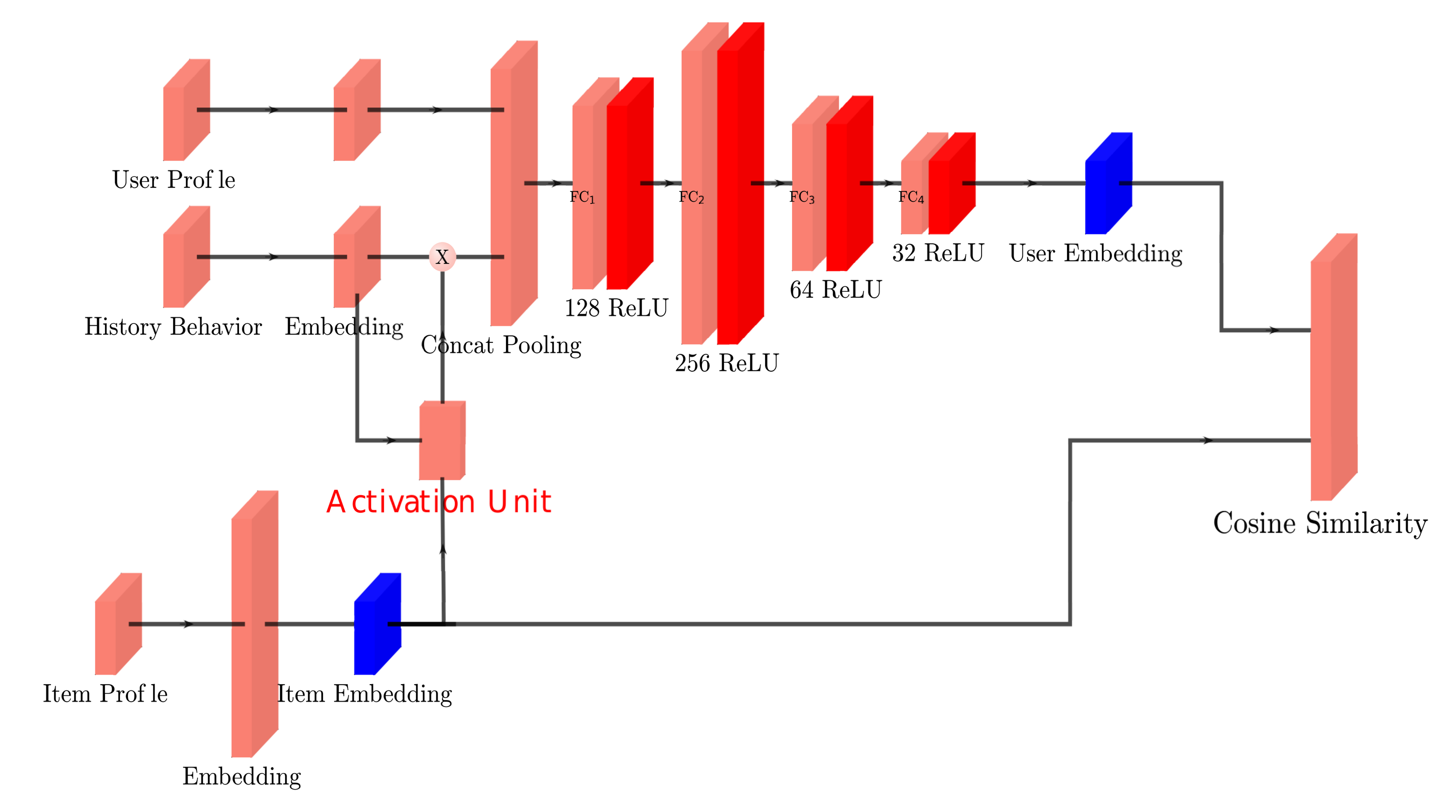

Figure 3. YDNNA Structure

3. **Activation Unit**

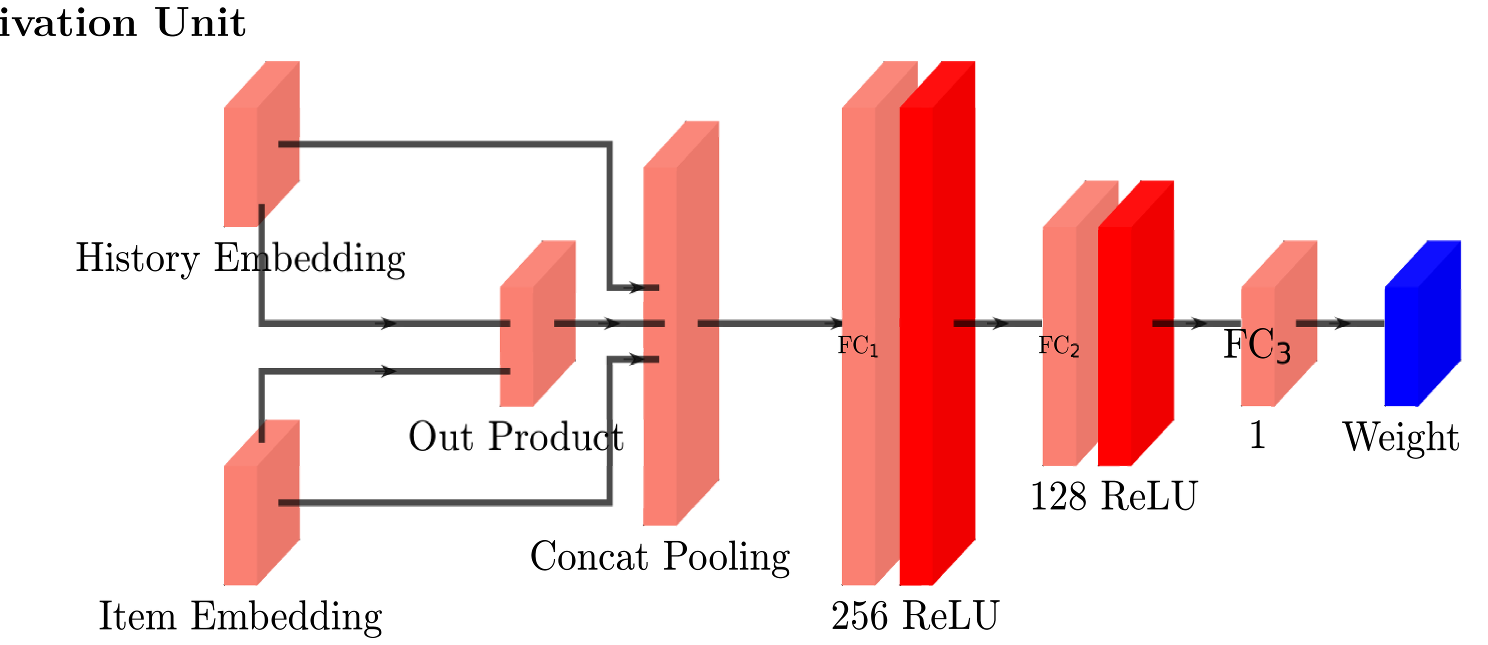

Figure 4. Activation Unit Structure

Results

Metric: Precison and **Recall** are chosen to have a evaluation.

Loss, Recall and Precision: All experiments are repeated 5 times and the average results are illustrated. Set TopK=3, 5, 10, 50, 100 respectively.

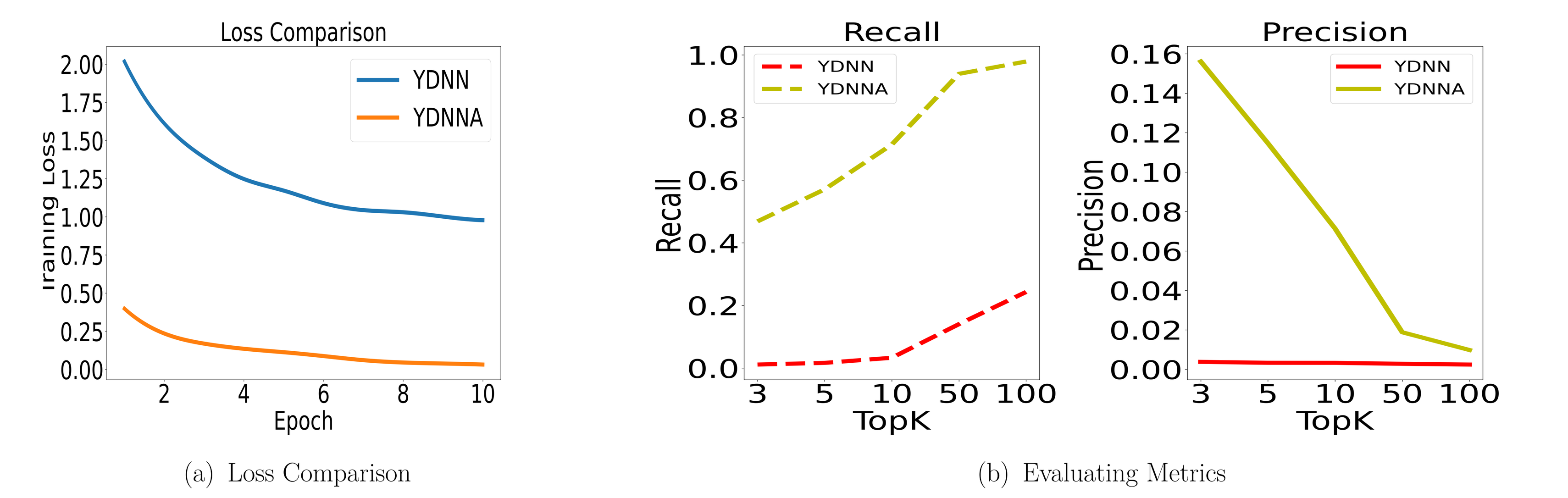

Figure 5. Loss and Metrics Comparison

Example: TopK=3

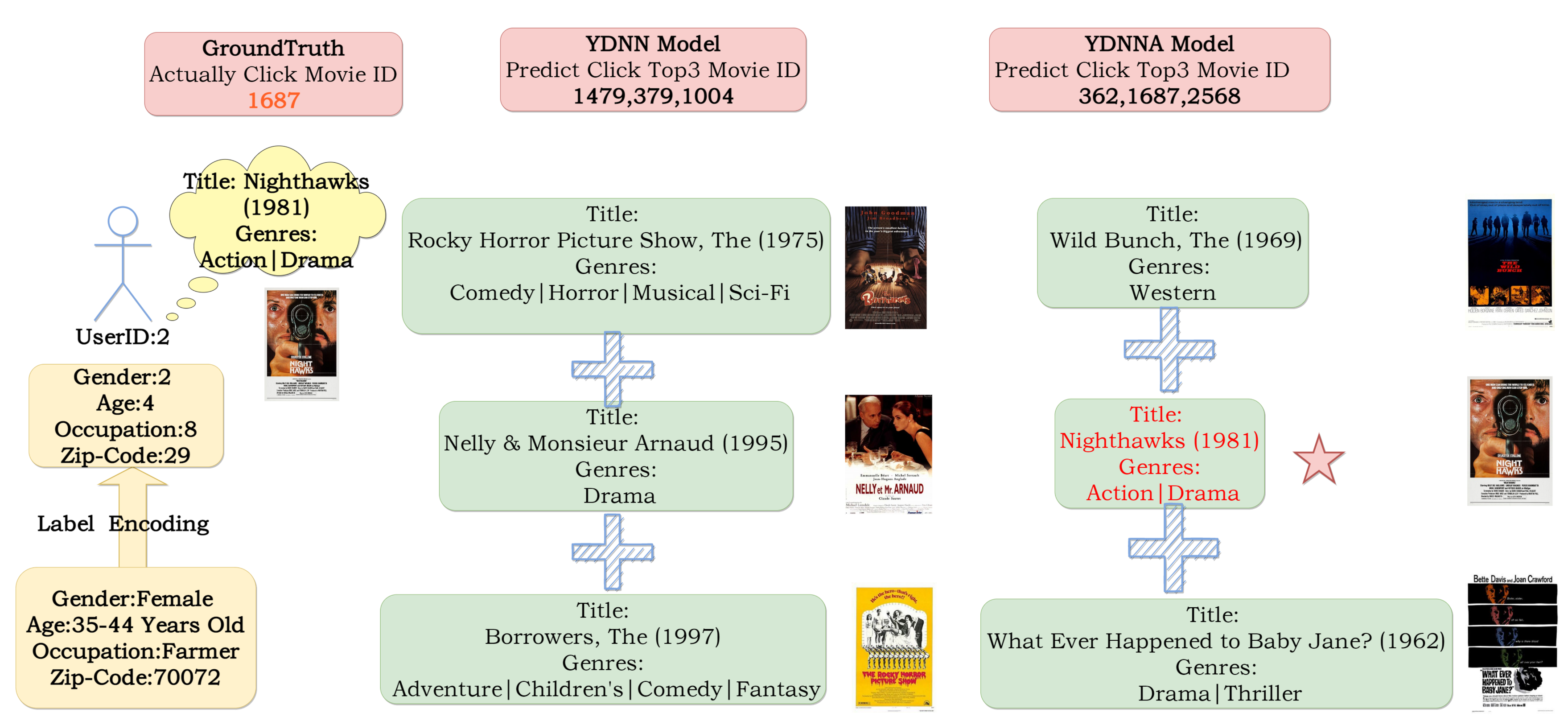

Figure 6. Example

Conclusion

- **Faster:** The YDNNA model can reach the convergence within only 10 epochs, which is faster than original YouTube DNN model.
- **More Accurate:** In the testing process, no matter how many TopK items are selected, our model can always perform better.

References

[1] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. Deep interest network for click-through rate prediction. pages 1059–1068, 2018.

[2] Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. pages 191–198, 2016.