Report for milestone1

In this report, it recorded how TEAM3 developed the first model with given data for milestone1.

1. Learning: We finally decided to use a technique called “matrix factorization”, a simple embedding model Given the feedback matrix A ∈R(m×n), where m is the number of users and n is the number of movies, the model learns:

* A user embedding matrix U∈R(m×d), where row i is the embedding for user i.
* A movie embedding matrix V∈R(n×d), where row j is the embedding for movie j.

The number in the matrix indicates the ratings from the users predicted by the model for a given movie based on past records. The performance of the model is reported by mean square error, where using predicted rating to minus the actual rating and square the result can get the MSE. A smaller MSE indicates that the model is working as expected, vice versa.

Why the model does not perform as well as expected:

The learning part took approximately two hours on GCP with 20,000 data from Kafka steam and should be longer using colab.

1. Inference service: The matrix will output 20 movies with highest predicted rating for a given user, from high to low. The ranking indicates which movies will be considered to have a higher score than others based on the user’s previous preference, which means they are more recommended for the user to watch and give a score for it.

Link: <https://colab.research.google.com/drive/1iVHJefQsbeaO-fQPvrUnMpB2aLxcwZSY?usp=sharing#scrollTo=x9tz7IqIFmYr>

1. Team process and meeting notes:

Meeting notes are stored on local devices, written by Zeyu Li.

1. The first meeting on September 23rd:

Meeting in person. Divide further works. Discuss what algorithm should be used for training and the input data format to train the model.

Barry Li , Marcos Souto Jr. and Zeyu Li are responsible for building the model. Two models are planned, and one of them will be selected in further meeting to hand in for milestone 1.

Shanzid Shaiham and Kua Chen are responsible for collecting the data from Kafka and clear the data to make sure useful parts are preserved for training the model.

Two models are built with data collected. One of them will be used further.

1. Second meeting on September 28th:

Meeting in person. Select the ‘matrix factorization’ model for training and to hand in for the first milestone. Decide how many data should we feed into the model. Talk about the functions of the model and see if it can fulfill needs for the current stage and how to interpret the data we have for the moment.

Collect everyone’s idea on further actions for the feedback loop, how will the model grow when we get new data inputs. We made an agreement on the model growth further, as new data are feed into the model, we separate a smaller segment that only works on the new data and at the same time the bigger model is preserved.

The model is trained with 20,000 data inputs after the meeting.

1. Third meeting on September 30th:

Meeting in person. Deploy the model.