WEEK 1

Objective : To develop a Graph Neural Network for Social Recommendation

Topic : Learn Recommender System – Implement Matrix Factorization Method

Recommender System Basics –

A recommendation system helps users to find compelling content in a large catalog of items. It recommends the user certain item based on users taste, which rely on learning an appropriated embedding representation of the queries and items. However, a recommendation system also display items that users might not have thought to search for their own. This is because there are three types of recommendation systems which consider different parameters for their evaluation. They are –

- Content – Based Filtering
- Collaborative Filtering
- Latent factor

1. **Content Based Filtering –** Content-based filtering is the system that recommends items similar to what the user likes, based on their previous actions and explicit feedback. The main idea behind this is to create "Item" and "User" profiles which collectively matches with the catalog and accordingly recommends items to the users. The model captures the specific interest of a user for recommending items. But the model has only limited ability to expand the users' existing interest.
2. **Collaborative Filtering –** To overcome the constraints of Content-based filtering, Collaborative Filtering takes into consideration the likes and dislikes of similar users along with the approaches of content based filtering method. Collaborative filtering allows for serendipitous recommendations as it uses similarities between users and items simultaneously to provide recommendations.

Methods Used :

Matrix Factorization – It is a class of collaborative filtering algorithm which is very effective and efficient. It became well-known during the Netflix Prize challenge because of its efficiency and effectiveness. Matrix Factorization is the method by which a matrix is decomposed to its constituent elements which makes complex computational matrix problems easier.

In this situation, a matrix R (u × i), which is a user is divided into two matrix X (u × k) and Y (k × i). The matrix X is a user-concept matrix and Y is an item-concept matrix. To compute the rating of a user for an item we take the dot product of two matrices-

$$\mathbf{r_{ui} = X_u^T . Y_i}$$

The loss function regarding the accuracy –

$$\boldsymbol{L = \sum_{u,i}(r_{ui} - x_u^T . y_i)^2}$$

In order to go through with matrix factorization, we have studied the working and basics of three well-known algorithms. Those are –

1) Single Value Decomposition (SVD)
2) Alternating Least Square (ALS)
3) Weighted Alternating Least Square (WALS)

Single Value Decomposition (SVD) – In SVD, a complex matrix M(m × n) is decomposed into three smaller matrix U (m × r),∑(r × r),V^* (r × n) where U and V are orthonormal matrices, ∑ is a diagonal decreasing matrix and r is the rank of the matrix M. It is used to decrease the dimensionality of the matrix and also to deal with the sparsity of the user-item matrix M.

Alternating Least Square method – ALS method or Alternating Least Square method of matrix factorization is one of the iterative methods in order to find two matrices X and Y that best approximates the rating matrix R. In this method, the loss function is minimized with respect to either the row or column factor while keeping the other factor as a constant.

Weighted Alternating Least Squares (WALS) method – In this method, weights are introduced in the loss function. These weights vary for the zero and the non-zero entries of the matrix. The weights are generally calculated by the sum of the non-zero entries of a row in order to normalize the entries of users who rated a different number of items. This method is usually more efficient than ALS method.

Dataset :

The dataset rating.csv includes the rating information of movies. It includes five columns which are :

- userid
- productid
- categoryid
- rating
- helpfulness

  The figures provide information about a user rating a specific product of a particular category and the helpfulness of the rating to recommend further items to the user. The model is trained with the given stats of userid, productid and rating so as to build a proper recommender system.

Packages :

- "Surprise" is the main package which contains a set of built-in algorithms and datasets. It is a Python scikit for building and analyzing recommender system that deals with rating data explicitly.
- "Pandas" library is used for reading the dataset, manipulating and analyzing it.
- "SVD" package is used for Matrix Factorization.

    - Functions rmse() and mae() are used to calculate accuracy of the model.

Findings :

Value Table –

Analyzed dataset for the following algorithms

Algorithm used	RMSE	MAE
SVD	0.977	0.745
KNNBasic	1.068	0.795
KNNWithMeans	1.053	0.784

KNNBaseline	1.016	0.757
CoClustering	1.035	0.761
BaselineOnly	0.972	0.747
NormalPredictor	1.379	1.052

Experimental result :

From the above experiment we find that BaselineOnly and SVD algorithms have the least errors. So both SVD and Baseline are best contenders for building a recommendation system with the given dataset.

Conclusion :

So, at the conclusion of the first week we have gone through matrix factorization and the different methods to do matrix factorization. We have also found out the accuracy of various algorithms through the RMSE and MAE functions. We have also become familiarized to certain packages such as pandas and surprise. Although the goal was hard at the beginning, we are trying to grasp it as much as we can and we are enthusiastic to learn more in the coming weeks.