Chapter 7

Conclusion

This dissertation explorer the concept of recommender systems for Scratch. It briefly describes the motivation for the study and focuses on the problem of content discovery, self-directed learning on the Scratch platform. The idea of a recommender system might look like a simple problem but it requires an understanding of the underlying data analysis concepts. Various different recommender systems used in different fields are discussed. Specifically the once in the area of e-learning and online teaching, which have gained popularity and are used extensively. Different studies have attempted to develop a recommender system for Scratch in their own teaching context.

Over the period, various research has been done in the area of a recommender system. Various techniques have been proposed for building recommend systems for different applications. This thesis discusses the standard techniques in recommender system space and their uses. Content-based filtering, collaborative filtering and hybrid approach are most widely used and accepted as standard amongst the community. A Collaborative filtering process is used and applied on the Scratch dataset to build an item-item collaborative filtering recommender system. Basic tasks in a recommender system like generating a recommendation list and calculating the prediction are detailed for the dataset.

For generating a recommendation list in an item-item collaborative technique, the first step is similarity computation. Three different techniques are used for similarity computation. Cosine based similarity, correlation-based similarity and adjusted cosine based similarity are the three well-known techniques for computing an item similarity. All these three techniques were applied to the dataset to get the item similarity matrix for all the project a user has rated. Once the matrix was generated prediction calculation was done based on the weighted sum method. Prediction output gave us the predicted recommendation rating for the user. Once the prediction was generated. the projects were recommended to the user according to the score.

Different methods of evaluation for the accuracy of recommendation are discussed. For measuring the accuracy of the recommendation we use MAE as the evaluation metrics. On comparing the three similarity calculation methods, it is observed that the cosine similarity is the best performing one in terms of MAE.

When we look at the top 5 recommended items similar to the supplied item i.e our seed project, we observe some similarities between the recommended items. All the items had a similar set of script blocks and sprite usages. This shows the correctness of the similarity computation blocks.

To demonstrate this a client/server model was developed. In the scripts, we use standard Scratch API and Python data analysis libraries and ML libraries for model building.

Although this thesis demonstrates a working model of a recommendation system for scratch, much more work is needed to validate the usefulness of the methods. Many more useful methods could be developed by considering the user's behaviour over time. This method of recommendation is nor prone to general problems faced by recommender systems. Limitations like over specialisation problem, cold start problem and concentration bias limit the work of the proposed method. Moreover, the quality of the data available also affects largely on the prediction results. If the data has missing values for the user ratings, predictions cannot be calculated correctly by the algorithms.

Future work for this study would involve qualitatively testing the proposed method for the usefulness of the system. Various testing methodologies could be adopted like a/b testing, user surveys or other novel techniques as discussed in [44]Knijnenburg. Also, more data could be extracted from the datasets to gain more information about the user. like a text analysis or sentiment analysis could be done on the comments data on the projects for understanding the user preference and taste. A more complex analysis could be a social network graph analysis on the follower -followee relationship to understand the user interactions and taste. A heuristics-based recommendation could be a fast and effective alternative if developed correctly.