Problem Statement

Steam is a video game distribution platform with over 120 million monthly users and 50,000 games, with 10,963 games added in 2022, a 14% increase from 2020. For gamers, the proliferation of new games poses the difficult question of what to play next. We developed a game recommendation system based on predicted playtime to answer this question.

Matrix Factorization Algorithm

The recommendation system enables users to determine how long they would spend playing a particular game. To achieve this, we employ matrix factorization. Matrix factorization is the optimal technique for discovering potential playtime for Steam users when predicting playtime based on minimal information. We use this technique to decompose the playtime matrix with rows representing games and columns representing users into two other matrices - user matrix and game matrix - in order to identify patterns in the dataset. The objective is to predict missing playtime values by modeling them as an interaction between user and item profiles. Our recommendation system is based on the premise that game playtime can be determined by a user's past gaming behavior and the attitudes of users with similar profiles. We assume that the latent features that lead users to enjoy specific games, such as the game genre or player mode, form a low-dimensional matrix, indicating that a limited number of variables determine game playtime. Using the loss on a validation set, we optimize the number of latent features, learning rate, and the l2 regularization parameter that prevents overfitting. The method permits us to predict playtime without knowledge of other user and game characteristics, as these are inferred by the algorithm that generates personal recommendations.

Data collection and methods

Using the Steam API, we have collected data on approximately 2,000 Steam users' owned games and playtime by generating random Steam user IDs. Then, we expanded our dataset with 6,000 additional users extracted from the friend lists of the original users. Our final dataset comprises a scarce matrix of 8,691 steam users and roughly 1,000 top games with 649,731 playtime records. An exploratory analysis revealed the log-normal distribution of user playtime. Therefore, we trained our model with logged playtime to avoid convergence issues due to outliers.

Results

Our final model displayed a mean squared error of 0.65 (log scale) on the validation set. We predicted playtime for a user's unplayed games by fitting new user profile weights using stochastic gradient descent, while maintaining the same game weights. Our program's output is the names of the top 10 unplayed games and their estimated playtime for a specific user. The poor performance of our algorithm in predicting outliers is a weakness that will be addressed in future iterations of our model.

Benefits to the Steam community

Our algorithm assists Steam users in discovering new games by providing information regarding prospective playtime. It also intends to boost sales for Valve, the app's proprietor, by providing more transparent recommendations.