EECS 349 Final Report

Group62: Jianyou Fang, Jianbo Liu, Xiao Pan, Guanglong Xu
NetID: JFS5405, JLH2293, XPU1036, GXO2990

Introduction

Our task is to provide a recommendation list of games to players base on player's reviews on games he/she has played in the past, along with other features like game category and game level. We will apply a few machine learning algorithm to the dataset collected from several game website to achieve better and more accurate recommendation. With the recommendation list, user could not only save lots of money on other games they might not interest in, but also saving a great amount of time to search for games fit for them. That's why we consider this is a pretty important task for us to explore.

Dataset

We get the dataset from metacritic.com by crawling and scraping data by Scrapy and the dataset is from 2012 to now. The total number of our data is about 37000 user reviews and 1397 game data, and the data have features of game title, game critic score, game user score, game platform, game genre, publisher and developer of the game, game rating, user name, user score on single game, user's average review score. The attributes of each game is title, rating, developer, genre and metascore.

Training & Test

We trained multiple algorithms at the same time and use 10-fold cross-validation to assure the models' generalizability. We also use different parameters to find the optimal model. We convert the 0-10 rating to 0-5 points.

Results And Analysis

We use scikit-learn to analyse different machine learning algorithms. We tested 8 algorithms: K-NN, Multi-layer Perceptron, Decision Tree classifier, random forest, Naive Bayes Classifier, Quadratic discriminant analysis, Support Vector Classification and ada boosting. We chose these theod because we think they are good at the problem we are facing and will suit the dataset we are using. 10-fold cross validation was used to find the best algorithm with best parameters. After running each classifier and analyzed each of the result, we found decision trss was the best algorithm to use for our propose. However due to the limitation on the hardware, we are not able

to get more data thus even the Decision tree model didn't yield a super high accuracy. Table 1 shows different performance of several algorithm.

KNN	MLP	DT	RF	AB	NB	QDA	SVC
0.5	0.56	0.58	0.55	0.60	0.10	0.038	0.55

Table 1 Accuracy of different algorithm

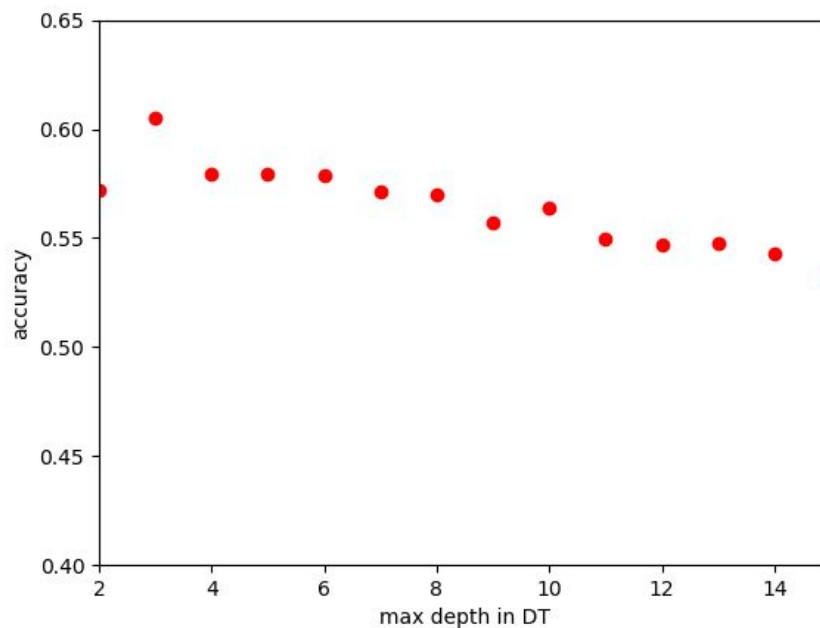There for continue to tweak on the DT method. We changed the max depth of the algorithm and we get Figure 1.

Figure 1 Change max depth in DT

We can see an peak at max_depth = 3. We got 61% accuracy by changing one parameter.

Future Improvement

We need to do is get more accurate data to achieve more accurate prediction since our current highest prediction accuracy is 61%. Furthermore, the complexity of human nature and people's social circle could also affect whether people like a game or not. Meanwhile, the decision tree model experiments show that our accuracy reach the peak while the max depth is 3. It indicates that our feature is too simple to make accurate prediction. We should also try other models the RNN to run our data to check we could get a more accurate prediction.

Who did what

Jianyou Fang: Wrote script to get data from several game website. Participated in the discussion of Project Proposal, Project Status Report and Final Report. Wrote the project proposal. Helped try to find the best feature for the prediction.

Jianbo Liu: Running data through weka. Wrote python code and implemented several model such as decision tree, random forest to analysis the data and make predictions. Optimized the accuracy of machine learning algorithm. Assisted develop the website for the final project. Wrote the final report.

Xiao Pan: Formatted the collected data and wrote python code and implemented several model such as KNN, Multi-level Perceptron to analysis the data and make predictions. Wrote project proposal and project status report.

Guanglong Xu: Developed the website for the final project report. Running data through weka. Proposed several machine learning model and analysis the result of each model. Optimized the accuracy of machine learning algorithm. Wrote the Project Proposal, Project Status Report, and Final report.