

Machine Learning Project: An Article Recommendation System

Gao Qikai

PID: 730412676

COMP 562 Spring 2020

University of North Carolina at Chapel Hill

1. Introduction

Recommendation systems are seen more commonly in our daily lives. By learning users' preferences from their reviews and reactions to numbers of contents like movies, articles, music and news, media service providers are able to recommend users contents they are more likely to enjoy. There could be relationships between numbers of contents, that whether a user likes a content will contribute to the decision that whether another content is recommended to the user. In this project, I used anonymous ratings data from the Jester online joke recommender system to find the linear relationship among 100 different articles, based on rated preferences from 24,938 users.

Note that although this project was meant to be done in group of four students, due to the pandemic shutting down physical instructions and myself being a semester exchange student, I was asked to leave UNC and go back to my home university in

Singapore. Thus, the twelve-hour time difference made it relatively hard to communicate with teammates and difficult to share data. Thus, having asked Mr. Jorge Silva for consent, I worked solo for this machine learning project.

2. The Data

The dataset used in this machine learning project is sourced from Jester online recommender system (Goldberg, n.d.). Each user rated numbers of articles from 100 choices, gave them continuous ratings from -10.00 to +10.00, where a higher ranking means the user likes the content more. The dataset is stored as Excel .xls format, where each row represents one user, and the 100 columns of one row represents the rating of that user gave to the specific article. When 99 is present in a cell, it means the user did not give a ranking for that article.

3. Methodology

Given 100 different articles, I chose one of the most reviewed, number 5, to be the testing data and other 99 articles to be the training data. Doing so will allow me to test the performance linear relationship predicted by the machine learning model by comparing the predicted data with the testing data. Using linear regression model, each of the 99 articles has a weight factor that decides how much this rating affects the prediction.

4. The Experiment and Results

Using linear regression model, having done gradient ascents and calculated the average differences between given Y values and predicted Y values, the algorithm is able to produce predictions with average error smaller than 5, which the rating ranges from -10.00 to +10.00.

The code and dataset for this project is available at

https://github.com/gaoqikai/COMP562_Linear_Regression.

Bibliography

Goldberg, K. (n.d.). *Anonymous Ratings Data from the Jester Online Joke Recommender System*. Retrieved from <https://goldberg.berkeley.edu/jester-data/>