MLND-Capstone project

Book Recommender System Proposal

***1. Domain background**

Recommender systems have become increasingly popular in recent years, and are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. There are also recommender systems for experts,collaborators,jokes, restaurants, garments, financial services,life insurance, romantic partners , and Twitter pages.*(1)

***2.Problem statement**

As websites are using recommender systems widely, online book buyers are beginning to see books they did not know existed but were 'recommended' to fit their individual tastes. As a avid reader,I want to build a book recommender system.

***3.Datasets and inputs**

Collected by Cai-Nicolas Ziegler in a 4-week crawl (August / September 2004) from the Book-Crossing community with kind permission from Ron Hornbaker, CTO of Humankind Systems. Contains 278,858 users (anonymized but with demographic information) providing 1,149,780 ratings (explicit / implicit) about 271,379 books.

BX-Users Contains the users. Note that user IDs (`User-ID`) have been anonymized and map to integers. Demographic data is provided (`Location`, `Age`) if available. Otherwise, these fields contain NULL-values.

BX-Books Books are identified by their respective ISBN. Invalid ISBNs have already been removed from the dataset. Moreover, some content-based information is given (`Book-Title`, `Book-Author`, `Year-Of-Publication`, `Publisher`), obtained from Amazon Web

Services. Note that in case of several authors, only the first is provided. URLs linking to cover images are also given, appearing in three different flavours (`Image-URL-S`, `Image-URL-M`, `Image-URL-L`), i.e., small, medium, large. These URLs point to the Amazon web site.

BX-Book-Ratings Contains the book rating information. Ratings (`Book-Rating`) are either explicit, expressed on a scale from 1-10 (higher values denoting higher appreciation), or implicit, expressed by 0.

***4. Problem solution**

Recommendations made by websites are not made at random, but are based on other similar readers' preferences or purchase histories. Recently this phenomenon is becoming a powerful marketing tool that retailers deploy to meet reader expectations and generate sales through up-selling or cross-selling. And the engine to generate recommendations is powered by algorithm based Recommendation Systems using an array of techniques such as Collaborative Filtering and Markov Chains.

***5.Benchmark model**

The goal of this project is to build a Recommendation System for book buyers through Collaborative Filtering, Collaborative filtering produces recommendations based on the knowledge of users' attitude to items, that is it uses the "wisdom of the crowd" to recommend items.

***6.Evaluation Metrics**

A distance metric commonly used in recommender systems is cosine similarity, where the ratings are seen as vectors in n-dimensional space and the similarity is calculated based on the angle between these vectors. Cosine similarity is recommended to use when there is a sparse data. A problem with Cosine similarity is that it does not consider the differences in the mean and variance of the ratings made to items. On the other hand In our case as we have a dataset with high sparsity so We are going to use cosine similarity .

***7.Project design (workflow)**

(1)Data cleaning with pandas

(2)sframe data visualization with graph lab

(3)cosine similarity to build recommender

(4) cross-validation

(5)Evaluating Recommender with RMSE and F1

(6)Adjust recommender parameters with grid search Model Comparisions to select the best model.

(7)evaluate our final model with the best parameters inputted

(8)Comparison of Final Model with the Previous Models by RMSE

(9)Comparison of Final Model with the Previous Models by Precision , Recall and Cutoff

(10)The final Recommender Model

refrences

(1) https://en.wikipedia.org/wiki/Recommender_system

(2) http://www2.informatik.uni-freiburg.de/~cziegler/BX/