Book Recommendation System

Timeline: October 5th - December 15th

Xiaohui Chen - Rajendra Bhagroo - Shelly Gupta

NYIT
College of Engineering & Computing Sciences

Problem & Goal

Problem & Goal

What do you want to solve? - Why do you think it is important? - What results do you expect?

Problem

- Users need a recommendation system that is relevant in regards to a users rating history.

Goal

- Our goal is to use collaborative filtering and information retrieval techniques to recommend books users are likely to read.

- We think this is important because we are avid readers who experience sub-optimal recommendations based on book categories rather than on the book contents.

- We expect to accurately recommend at least three books based on previous user ratings.

Formalization Into IR Related Task

How Is Our Project Related To IR?

Formalization Into IR Related Task

How is our project related to IR?

Collaborative Filtering & Recommender System

- For our project we will implement a collaborative filtering recommendation system. We will apply ALS (alternating least squares) algorithm in Spark.ml library to do model-based collaborative filtering.
- Unlike user collaborative filtering and item collaborative filtering, model-based collaborative filtering does not directly predict the new items or give recommendations for different users, instead it will train a model based on user preference data and use the model to do predictions and compute recommendations.
- Matrix factorization, take latent factors into account -> predict ratings for new items
- Throw new data into training set and update the ALS Model each time.

Data Plan

Data Plan

What kind of data? - Where and how do you get the data?

What Kind Of Data?

- CSV Data
  - Book-Crossing data set
    - Information about...
      - Users (User-ID, Location, Age)
      - Books (ISBN, Book-Title, Book-Author, Year-Of-Publication, Publisher)
      - Book Ratings (0-10)

User-ID	Location	Age
1	nyc, new york, usa	NULL
2	stockton, california, usa	18
3	moscow, yukon territory, russia	NULL
4	porto, v.n.gaia, portugal	17
5	farnborough, hants, united kingdom	NULL

User-ID	ISBN	Book-Rating
276725	034545104X	0
276726	0155061224	5
276727	0446520802	0
276729	052165615X	3
276729	0521795028	6
276733	2080674722	0
276736	3257224281	8
276737	0600570967	6
276744	038550120X	7

* Data needs to be cleaned before we process with it.

ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher
0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press
0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada
0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial
0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux
0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company
0399135782	The Kitchen God's Wife	Amy Tan	1991	Putnam Pub Group
0425176428	What If?: The World's Foremost Military Historia...	Robert Cowley	2000	Berkley Publishing Group
0671870432	PLEADING GUILTY	Scott Turow	1993	Audioworks
0679425608	Under the Black Flag: The Romance and the Re...	David Cordingly	1996	Random House

Where Will We Get Our Data?

- http://www2.informatik.uni-freiburg.de/~cziegler/BX/
  - Improving Recommendation Lists Through Topic Diversification, Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, Georg Lausen; *Proceedings of the 14th International World Wide Web Conference (WWW '05),* May 10-14, 2005, Chiba, Japan. *To appear.*

Schedule

Schedule

November 2nd - November 9th (Week 1)

- Project Setup (Github, Jupyter Notebook, Etc)
- Refine Project Plan
- Book-Crossing Dataset - Processing Into dataframe & Cleaning data

November 9th - November 16th (Week 2)

- Construct our model
- Use Spark Library to manipulate Book Crossing Dataset & Finalize our model

November 16th - November 23rd (Week 3)

- Begin User Interface Design
- Tweak model & prepare interface from model to UI

November 23rd - November 30th (Week 4)

- Finish UI
- Merge UI with Model
- Begin technical documentation

November 30th - December 7th (Week 5)

- Testing & Debugging of model and interface
- Refine & finish technical documentation