Recommendation Engines

CSC - 9010 002

Fall – 2018

1. **Project Title:** Recommendation Engines

2. **Abstract:**

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. Mostly used in the digital domain, majority of today's E-Commerce sites like eBay, Amazon, Alibaba etc. make use of their proprietary recommendation algorithms to better serve the customers with the products they are bound to like.

In this research project, different types of recommendations, process strategies and different algorithms are outlined. This research also elucidates the advantages recommendation engines can offer to business.

3. **Related domain of study:**

Building recommendation models by,

- Collaborative filtering algorithm
- Amazon DSSTNE

4. **Algorithms:**

Recommendation engines are being supported by two popular versions of algorithms - collaborative filtering and cluster models.

Other algorithms including search-based methods and item-to-item collaborative filtering which focus on finding similar items, not similar customers. For each of the users purchased and rated items, the algorithm attempts to find similar items. It then aggregates the similar items and recommends them.

These recommendation systems leverage user activity patterns and predict what user could like in future based on the behavior patterns. In the context of recommendation systems, collaborative filtering is a method of making predictions about the interests of user by analyzing the taste of users which are similar to the said user. The idea of filtering patterns by collaborating multiple viewpoints is why it is called collaborative filtering.

Large data set 'Movielens' is broken down into multiple .csv files- links, movies, rating and tags. Each csv file is having entries based on UserID or Movie ID.
A comparison is done on the movie data based on 2 components ('easy watching' vs 'serious', 'CGI' vs 'dialog driven').

5. **Data Source URLs:**

1. https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0
   This website is a one-stop source to know about collaborative filtering for Recommendation engines. Different types of collaborative filtering approach have been explained with an example.
   Data Set 'Movielens' is downloaded from http://files.grouplens.org/datasets/movielens/

```
In [ ]: movies = pd.read_csv(path+'movies.csv')
        movies.head()
```

Out[ ]:

	movieId	title	genres
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
1	2	Jumanji (1995)	Adventure|Children|Fantasy
2	3	Grumpier Old Men (1995)	Comedy|Romance
3	4	Waiting to Exhale (1995)	Comedy|Drama|Romance
4	5	Father of the Bride Part II (1995)	Comedy

Steps for data analysis:

- Download MovieLens data and read it in pandas data frame.
- Install Surprise package by pip install scikit-surprise. Load the data into Dataset class
- Implement collaborative filtering algorithm after data preparation.
- Predictions compare to the actuals data, a plot graph to represent this.

2. https://github.com/amzn/amazon-dsstne/blob/master/FAQ.md
   This amazon labs github link explains different aspects of amazon library DSSTNE like its offerings and capabilities.
   This research project will only target to briefly explain how DSSTNE is used in recommendation models excluding the implementation.

6. Graphics:

- This project plans to lay down a scatter plot to compare the movie data based on 2 components ('easy watching' vs 'serious', 'CGI' vs 'dialog driven').
- A bar chart to show the distribution of rating values.

7. Current challenges:

- Finding more data sets that can be used for collaborative filtering algorithms.
- Understanding the implementation part of different types of algorithms under collaborative filtering.
- Understanding the basics of scikit-learn and getting hands-on.
  https://towardsdatascience.com/hands-on-introduction-to-scikit-learn-sklearn-f3df652ff8f2
  https://medium.com/dunder-data/from-pandas-to-scikit-learn-a-new-exciting-workflow-e88e2271ef62

8. Reference URLs:

1. https://www.computer.org/cms/Computer.org/magazines/whats-new/2017/07/mic2017030012.pdf

2. https://www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf
3. https://publications.computer.org/internet-computing/2017/09/27/amazon-all-the-research-you-need-about-its-algorithm-and-innovation/
4. https://towardsdatascience.com/5-advantages-recommendation-engines-can-offer-to-businesses-10b663977673
5. https://towardsdatascience.com/if-you-cant-measure-it-you-can-t-improve-it-5c059014faad

9. Domain or type of big data problems this project addresses:

Recommendation engines focused on serving online customers a personalized and customized content to each individual customer rather than solving a business problem. It uses existing information such as customer search and purchase data to analyze and recommend the products. 5 benefits that business can achieve from recommendation engines are increased revenue, customer satisfaction and personalization. The key to converting visitors to customers is to use real-time recommendations.