My Goal?

I want to recommend healthy recipes to people. How? By checking their BMI index and calorie required for user. This is not something new. So what can you propose? Comparison of different algorithms and see which performs best.

Title – FeedMeRight – Comparison of Recipe Recommender systems.

Abstract:-----

With so many rapid changes happening around us, people are moving towards healthy lifestyle which includes choosing right food. Nowadays we are surrounded by overwhelming information which hinders the ability to choose right information. Recommendation system is a technique that would filter information and narrow it down based on our preferences and helps us to choose which we may like. Recommender systems have been widely used in e-commerce sites, social networking and entertainment industries. To help us in choosing healthy lifestyle recommender systems can be used in recipe domain to choose a right food for us not just based on user’s taste but also considering user’s lifestyle and calories required for user.

This thesis presents the design, implementation and evaluation of various recommender approaches within recipe domain. I will combine different recommendation techniques and use machine learning to recommend healthy recipes based on user profile. The overall goal is to discover which approach to recommend recipes offers better performance.

Keywords – Recommender system, Machine Learning, content-based, collaborative filtering, matrix factorization.

Contents

1. Introduction
   1. Motivation
   2. Problems Addressed and Challenges
   3. Thesis Contribution
   4. High level overview of proposed algorithm..(remove if repetitive)
2. Background

2.1 The World Wide Web

2.2 Information Retrieval

2.3 Information Filtering

2.4 Recommender System

2.4.1 Content-based Filtering

2.4.2 Collaborative Filtering

2.4.3 Hybrid approach

2.5 Recommendation Approaches

2.5.1 Basic approaches to find similar items

2.5.2 Vector Space Model

2.5.3 Extended Vector Space Model for Content-based Filtering

1. Implementation

3.1 Data Filtering

3.2 User-User CF (KNN)

3.3 Item-Item CF

3.4 Matrix Factorization

3.5

1. Experimental Results
   1. Results obtained from Content Based
   2. Results obtained from Collaborative Filtering
   3. Results obtained from Matrix Factorization
   4. Comparison
2. Experimental Analysis

5.1

1. Conclusion (& future work if I could think of something)

References

**Writeup reference - Health-aware Food Recommender System**

[1] Resnick and Varian. Acm press, recommender systems, volume 40., 1997. URL <http://doi.acm.org/10.1145/245108.245121>.

[bigdata] <https://www.oracle.com/big-data/guide/what-is-big-data.html>

[Khusro2016] <https://link.springer.com/chapter/10.1007%2F978-981-10-0557-2_112> Khusro2016

[recommender\_overview]<https://www.researchgate.net/publication/220604600_Recommender_Systems_An_Overview>

[contentbased]<https://www.researchgate.net/publication/236895069_Content-Based_Recommendation_Systems>

[contentbased\_architecture] <https://www.researchgate.net/publication/331063850_Recommender_Systems_Challenges_and_Solutions_Survey>

[CF] <http://files.grouplens.org/papers/FnT%20CF%20Recsys%20Survey.pdf> Reference for collaborative

filtering

[cf1] <https://link-springer-com.summit.csuci.edu/referenceworkentry/10.1007/978-3-319-32001-4_274-1> -Collaborative filtering by Ashrf (reference for memory based and model based collaborative filtering )

[7]<https://www.researchgate.net/publication/328231954_Comparative_analysis_of_recommender_systems_and_its_enhancements>

\cite{Resnick1997} \cite{CF} \cite{bigdata} \cite{Khusro2016} \cite{recommender\_overview} \cite{figures} \cite{contentbased} \cite{CF} \cite{cf\_figure}

Michael D. Ekstrand, Joseph A Konstan, coursera, Introduction to Recommender Systems: Non-Personalized and Content-Based https://www.coursera.org/learn/recommender-systemsintroduction/ lecture/ZkG45/summary-statistics-i.

## For similarity measures

<https://pdfs.semanticscholar.org/943a/e455fafc3d36ae4ce68f1a60ae4f85623e2a.pdf>

© Springer Science+Business Media Singapore 2016 1179

K.J. Kim and N. Joukov (eds.), Information Science and Applications (ICISA) 2016,

Lecture Notes in Electrical Engineering 376,

DOI: 10.1007/978-981-10-0557-2\_112

Recommender Systems: Issues, Challenges,

and Research Opportunities

Shah Khusro, Zafar Ali and Irfan Ullah

© Springer Science+Business Media Singapore 2016 1179

K.J. Kim and N. Joukov (eds.), Information Science and Applications (ICISA) 2016,

Lecture Notes in Electrical Engineering 376,

DOI: 10.1007/978-981-10-0557-2\_112

Recommender Systems: Issues, Challenges,

and Research Opportunities

Shah Khusro, Zafar Ali and Irfan Ullah

© Springer Science+Business Media Singapore 2016 1179

K.J. Kim and N. Joukov (eds.), Information Science and Applications (ICISA) 2016,

Lecture Notes in Electrical Engineering 376,

DOI: 10.1007/978-981-10-0557-2\_112

Recommender Systems: Issues, Challenges,

and Research Opportunities

Shah Khusro, Zafar Ali and Irfan Ullah

© Springer Science+Business Media Singapore 2016 1179 K.J. Kim and N. Joukov (eds.), Information Science and Applications (ICISA) 2016, Lecture Notes in Electrical Engineering 376, DOI: 10.1007/978-981-10-0557-2\_112 Recommender Systems: Issues, Challenges, and Research Opportunities Shah Khusro, Zafar Ali and Irfan Ullah

##### Hybrid type –

<https://www.bluepiit.com/blog/demystifying-hybrid-recommender-systems-and-their-use-cases/>

## BMI DOCS

Rerefences from Paper –

1. Health\_Aware\_Reco – [3]
2. Bringing healthy food – [6] – BMR Formula

<https://www.kbcc.cuny.edu/academicdepartments/physci/Documents/labmanuals/sci70/bmi.pdf>

Impact of food craving and calorie intake on body mass index (BMI) changes during an 18-month behavioral weight loss trial – Check on Springer Link

1. <https://www.cdc.gov/healthyweight/losing_weight/index.html>
2. <https://www.ncbi.nlm.nih.gov/books/NBK56068/table/summarytables.t4/?report=objectonly>
3. <https://health.gov/our-work/food-nutrition/2015-2020-dietary-guidelines/guidelines/appendix-7/>
4. 2002, "Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids," Food and Nutrition Board, Institute of Medicine.
5. Ainsworth B.E., 2002, January, "The Compendium of Physical Activities Tracking Guide," Prevention Research Center, Norman J. Arnold School of Public Health, Univ of SC.
6. Actual Formula with source -- <http://www.checkyourhealth.org/eat-healthy/cal_calculator.php>
7. [**http://health.howstuffworks.com/question457.htm**](http://health.howstuffworks.com/question457.htm)
8. [**http://www.bmi-calculator.net/bmr-calculator/bmr-formula.php**](http://www.bmi-calculator.net/bmr-calculator/bmr-formula.php)
9. [**http://www.bmi-calculator.net/bmr-calculator/harris-benedict-equation/**](http://www.bmi-calculator.net/bmr-calculator/harris-benedict-equation/)
10. Bough KJ, Rho JM. [Anticonvulsant mechanisms of the ketogenic diet](https://doi.org/10.1111/j.1528-1167.2007.00915.x). Epilepsia. 2007;48(1):43-58. doi:10.1111/j.1528-1167.2007.00915.x
11. Nordmann AJ, Nordmann A, Briel M, et al. [Effects of low-carbohydrate vs low-fat diets on weight loss and cardiovascular risk factors: a meta-analysis of randomized controlled trials](https://doi.org/10.1001/archinte.166.3.285). Arch Intern Med. 2006;166(3):285-93. doi:10.1001/archinte.166.3.285
12. Harvie M, Wright C, Pegington M, et al. [The effect of intermittent energy and carbohydrate restriction v. daily energy restriction on weight loss and metabolic disease risk markers in overweight women](https://doi.org/10.1017/S0007114513000792). Br J Nutr. 2013;110(8):1534-47. doi:10.1017/S0007114513000792

Source of categorizing diet lables.

1. <https://www.fda.gov/food/new-nutrition-facts-label/how-understand-and-use-nutrition-facts-label>  
     
   Read More <http://nutritiondata.self.com/tools/calories-burned#ixzz6KsEKVQFh>.
2. Read More <http://nutritiondata.self.com/tools/calories-burned#ixzz6KsEKVQFh>

<https://medium.com/@bond.kirill.alexandrovich/precision-and-recall-in-recommender-systems-and-some-metrics-stuff-ca2ad385c5f8> Recall And Precesion

## Content based references articles for diagrams and descripts –

<https://medium.com/towards-artificial-intelligence/content-based-recommender-system-4db1b3de03e7>

## Content based, Collaborative, Hybrid architecture and flow – youtube video –

<https://www.youtube.com/watch?v=TFi7WXpaIiY>

Converting formula from Cosine similarity to prediction ratings

<https://www.researchgate.net/post/How_would_you_recommend_shifting_from_a_Similarity_Scores_to_Rating_Prediction_system>

### My Writeup References –

1. Evaluating Collaborative Filtering Recommender Algorithms: A Survey - <https://ieeexplore.ieee.org/document/8550639?denied=>
2. Recommendation systems: Principles, methods and evaluation - <https://www.sciencedirect.com/science/article/pii/S1110866515000341#b0380>

### From Indian\_Cusine\_RecipeReco\_BasedonIngredients

#### I might not need point A

A. Web Scraping Web Scraping[6] (Scraping or Web Data Extraction or Web Harvesting[7]) is a system utilized to fetch a lot of information from sites whereby the information is extricated and spared to a nearby record in your PC or to a database in the table (spreadsheet) design. Web scraping is the way toward fetching information from sites. All the activity is completed by a bit of code which is known as a "spider". In the first place, it sends a "GET" question to a particular site. At that point, it parses a HTML record dependent on the got outcome. After it's done, the scrubber looks for the information you need inside the report, and, at long last, changes over it into the predefined format.

B. Data Cleaning Data cleansing or data cleaning[8] is the way toward distinguishing and adjusting (or expelling) degenerate or off base records from a record set, table, or database and alludes to recognizing fragmented, mistaken, off base or unessential parts of the data and afterward supplanting, altering, or erasing the filthy or undesirable data. Data cleansing might be performed intelligently with data wrangling instruments, or as group preparing through scripting.

C. Bags-of-Words The bag-of-words model [9] is a way of representing text data when modeling text with machine learning algorithms. The bags- of-words demonstrate is a rearranging portrayal utilized in natural language processing and information retrieval (IR)In this model, a content, (for example, a sentence or an archive) is spoken to as the sack (multi-set) of its words, dismissing syntax and even word request yet keeping variety.

Todo

1. Convert project into maven
2. Save each recipe information into json document from edamam API.. Loop through each unique recipe.
3. Will get data for each recipe. And I have data for user and recipe rating data.
4. From that I think I need to prepare user profile. Not sure how to calculate which ingredients each user likes. Check if we can get that info from allrecipes.com WebScrapor tool used in one of the paper was – spider plugin

### link for collaborative filtering.

<https://towardsdatascience.com/how-to-build-a-simple-recommender-system-in-python-375093c3fb7d>

## link to check if we can use it to evaluate recommender engine

<https://github.com/MaurizioFD/RecSys2019_DeepLearning_Evaluation>

Evaluation of recommender systems..

######## How the workflow should be –

1. Enter User Id:
   1. getRecipeFeatures() get called. It creates sparse matrix of all recipes based on it’s ingredients as features. Creates a .csv file for sparse matrix named recipe\_feature\_matrix.csv
   2. Whenever any new recipe will get added to the dataset, we need to update recipe\_feature\_matrix.csv file.
2. build\_user\_profile() will get called. It will read recipe\_feature\_matrix.csv wil get read.

In this method, user profile will get created for provided user\_id based on recipe\_feature\_matrix.csv file.

1. Based on user profile, similarity score will get calculated between user profile and all recipes. Sort the results in descending order.
2. Get the BMI of user and the required calorie range for user. And add 10 results from the sorted resultset such that calories of those recipes would fit in the required calorie range for user.

############ Evaluate the Engine ###############

Prof. thought on Evaluation

Recall -

and precision

we are defining what is good recommendation.

Cold start problem. In that case content based filtering would be more effective.

Full cross validation on train and test set.

To create vnv for kera –

1. Run conda create -n venv\_name and source activate venv\_name, where venv\_name is the name of your virtual environment.
2. Run conda install pip. This will install pip to your venv directory.
3. Find your anaconda directory, and find the actual venv folder. It should be somewhere like /anaconda/envs/venv\_name/.
4. Install new packages by doing /anaconda/envs/venv\_name/bin/pip install package\_name.

This should now successfully install packages using that virtual environment's pip!

Steps followed by me –

conda create –n conda\_venv

created environment location - C:\Users\jpall\Anaconda3\envs\conda\_venv

to activate virtual env in windows - C:\Users\jpall\Anaconda3\Scripts\activate.bat

files will be created at new place - C:\Users\jpall\Anaconda3\envs\conda\_venv\conda-meta

to install new packages –

C:\Users\jpall\Anaconda3\envs\conda\_venv>pip install tensorflow

### Links of Articles –

1. Theory - Collaborative filtering nearest neighborhood and matrix factorization - <https://towardsdatascience.com/intro-to-recommender-system-collaborative-filtering-64a238194a26>
2. Theory –Content based and collaborative based and matrix factorization and Evaluating Algortihms - <https://towardsdatascience.com/recommendation-systems-models-and-evaluation-84944a84fb8e>
3. Theory – Evaluating Recommender Systems –Actually everything – Content, Collaborative, matrix <https://medium.com/fnplus/evaluating-recommender-systems-with-python-code-ae0c370c90be>
4. Understanding Matrix Factorization - <http://nicolas-hug.com/blog/>
5. Evaluation code using surprise - <https://towardsdatascience.com/evaluating-a-real-life-recommender-system-error-based-and-ranking-based-84708e3285b>
6. <https://www.offerzen.com/blog/how-to-build-a-content-based-recommender-system-for-your-product>
7. Collaborative with evaluation by simple code - <https://www.ethanrosenthal.com/2015/11/02/intro-to-collaborative-filtering/>
8. Surprise basic good example of collaborative filtering- <https://realpython.com/build-recommendation-engine-collaborative-filtering/>
9. Tensorrrec all details.. huge blog with code -<https://towardsdatascience.com/getting-started-with-recommender-systems-and-tensorrec-8f50a9943eef>
10. Surprise all algorithm code in notebook –

<https://github.com/susanli2016/Machine-Learning-with-Python/blob/master/Movielens%20Recommender%20Metrics.ipynb>

<https://nbviewer.jupyter.org/github/NicolasHug/Surprise/blob/master/examples/notebooks/KNNBasic_analysis.ipynb>

1. <https://towardsdatascience.com/building-and-testing-recommender-systems-with-surprise-step-by-step-d4ba702ef80b>
2. Hybrid Implementations –

<https://www.kaggle.com/rounakbanik/movie-recommender-systems>

<https://www.kaggle.com/robottums/hybrid-recommender-systems-with-surprise>

1. Writeup reference pointwise - <https://github.com/jalajthanaki/Movie_recommendation_engine/blob/master/Intro_recommendation_system.ipynb>
2. # ## Source - <https://towardsdatascience.com/evaluating-a-real-life-recommender-system-error-based-and-ranking-based-84708e3285b>
3. ## Recall Precision

<https://towardsdatascience.com/recommendation-systems-models-and-evaluation-84944a84fb8e>

1. Matrix factorization

<https://cloud.google.com/solutions/machine-learning/recommendation-system-tensorflow-overview>

#### Command to convert Jupyter ntebook to python file –

jupyter nbconvert --to python allrecipes\_collaborative\_item\_item.ipynb

###### Tools to use to link references in the report

1. Mendeley
2. Jabref

############## Papers with summary ===🡺 what Should I follow –

1. Health Aware Food Recommender System – Extends Matrix Factorization by adding user’s BMI factor. Algo details in 2.
2. Using Tags and Latent factor in a Food Recommender System – matrix factorization without adding BMI factor. Best explanation about offline evaluation.

################# THOMS Review #############################

1. Motivation – www - should we replace it with http? –The main difference between WWW and HTTP is that they refer to different concepts. Simply put, HTTP is the protocol that enables communication online, transferring data from one machine to another. WWW is the set of linked hypertext documents that can be viewed on web browsers (such as Firefox, le Chrome, and more).
2. it did not help in changing user's behaviour towards helathy lifestyle. How healthy behavior would measured by the recommender

########### FLOW #####################

1. HI, I am Pallavi Chavan , and today I will be talking about FeedMeRight - Recipe recommender system. Before coming to masters program, I was working in BigData and Hadoop technology. The data I was working with was all about Credit card information and building reports on top of it by processing data. So my work experience inspired me to work on data and then I presented my first graduate seminar on bigdata and Hadoop technology.
2. Bigdata slide – Bigdata is any data which generates at high volume, high velocity with variety. It can be structured or non-structured. This image represents how data started o grow. The amount of data generation was very less before internet. As technology grows the amount of data generation started to grow exponentially.
3. One of great application of bigdata is to build recommender systems.

Build some point - why recipe recommender???

**Health factor -**

The research of World Health Organization shows that, globally there are 39 % adults are overweighed and 13 % adults are obese. Overweight and obesity cause the chronic diseases like diabetics, blood pressure. In order to maintain our weight, people need to consume the proper amount of calories based on person’s need. Generally, lemon people don’t know how much calories we are consuming from any meal. And our busy lifestyles don’t give us enough time to think about it. Thanks to the internet to provide information in our need very quickly but there are many resources are available that we don’t understand which one to choose. Which one is good for me? The result is people choose unhealthy and easy options. So, this problem can be solved by introducing recommender systems in recipe domain as recommender systems narrow down the information based user’s eating history and health.

**Literature Review –**

The work has been already done in this area. The research of Freyne and Berkovsky uses content base technique to recommend recipes. In this technique, only user’s taste has been considered.

My Questions for myself?

* 1. How to calculate calorie intake per meal?
  2. Range for adults height and weight?
  3. result set – replace slight increase with percentage.