IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-7, Issue-4, April 2019 E-ISSN: 2347-2693

A Framework of Computational Methods for Recommender Engine

Aishwarya Rajamani[1*], Alpha Vijayan[2]

[1,2]Dept. of Computer Science, New Horizon College of Engineering, Bangalore, India

[*]*Corresponding Author: aishraja12@gmail.com, Tel.: +919980577928*

DOI: https://doi.org/10.26438/ijcse/v7i4.4448 | Available online at: www.ijcseonline.org

Accepted: 12/Apr/2019, Published: 30/Apr/2019

***Abstract*—** Personalization of services can be proclaimed as the fulcrum of today's industries. Recommender engines are vastly employed to lend personal relevance to the consumers of services. Recommenders are a class of systems built upon the semantics derived off of consumer profiling, quantitative representations of preferences and so on. They draw a lot of their underlying mechanisms from the field of computational methods, that is, the use of mathematical models and methods to effect relevant suggestions. This paper proposes a framework for the creation of a recommender engine that is capable of predicting a rating matrix as well as providing suggestions. The key advantage introduced by the framework is that in addition to collaborative methods it also employs the techniques of topic modelling and fuzzy logic to find latent topics of interest that are not evidently visible as well as latent groupings of particulars modelled. This framework is a three-component engine consisting of an exploratory module augmented with statistical analysis, visualizations and algorithmic analysis, a latent factor analysis module built on the principles of topic modelling, fuzzy c means algorithm and a prediction-suggestion module built on the concept of singular value decomposition. Thus, this engine can be used in emerging problem domains of relevance to today's society.

***Keywords*—**Natural Language Processing, Topics Modelling, Recommender Systems, Latent Factor Analysis, Framework

I. INTRODUCTION

The concept of recommendation has grown to be ubiquitous owing to the increasing need for personalized services and products. This benefits industries immensely as customized services are consumed in greater volumes. The fact is evidently corroborated by the wealth of statistical researches performed on the topic.

This paper proposes a multi-algorithm methodical approach for the purpose of creation of a recommendation engine. For the objective at hand, the exposition is done here in several sections. In section II is discussed various papers on the matter of recommenders. The Sections III, IV deal with methods and proposed framework respectively. Finally, in section V a concluding discussion is included.

II. LITERATURE REVIEW

The most common knowledge suggestion systems are e-retailers, music streaming services and on-demand video libraries. A formal and well-known classification here is: Applications that employ the essential semantics of resources under consideration for the process of suggestion, also popularly known as content based systems; Applications that make use of particular quantized preferences of the consumers, also known as collaborative or group based systems; Applications that utilize both of the mentioned approaches. Over the years systems have become more nuanced technologically with the advent of several ensemble approaches. Some of the systems are discussed in this section.

Users of the internet have taken to swapping things owned by them thereby acquiring new products while finding people who may be interested in their offerings. Books are one such commodity of common exchange. One such recommender suggests to people reading materials that may be to their liking by means of a three step calculation. A group based method of arriving at the quantitative expression of preference is used along with linear algebra methods and the technique of regression [1].

A pressing issue among mobile phone users is choice of plan and services. One research work that addresses the issue makes use of third approach listed as a category of classification of recommenders. The use cases addressed are the suggestion of cellular devices for purchase and the suggestion of cellular plans just weeks before expiry. For the purpose fuzzy mathematical technique is used in arriving at similarity of products and the consumers, which is a typical step in recommenders [2].

Implementation of personalization in the field of mobile applications is a very useful use case. Such an application is described wherein the choice of metric is three-fold: the time of recent visitation, the number of visits and the span of the same. Also, the system has done away with manual inputs and automatically retrieves consumer usage stats, computes the suggestions based on a variation of simple linear regression [3].

Another domain of need for personalization is travel. And, often blind quantitative factor based approaches are ineffective. Thus, a context driven methodology is proposed by making use of five factors: place, climatic situation, time, automatically derived emotion quotient from consumer behaviour online, and manual consumer inputs [4].

A survey records that owing to the unstructured nature of data that is available on the internet much cleaning and processing through techniques that model text quantitatively as distributions is necessary [5]. More specifically on the subject of recommenders another paper speaks about the use of microblog data for the purpose of providing customized suggestions [6]. Here also, the authors have stressed the importance and procedural description of cleansing such data.

Customization does not refer to direct suggestions alone. There is much work being done to make IT services more linguistically secular. It is observed in the case of this paper which offers querying in a vernacular language by the use of natural language processing [7].

III. METHODS

A. Graph Based Methods

To cluster is to group objects into classes and in the field of computation and machine learning, clustering is an on-the-fly method to learn the groupings of objects. To co-cluster is to group the rows and columns of a matrix at the same time. A set of rows that are like each other, across the columns, in some way, are grouped together.

In graph theory, spectral methods deal with the properties of connected nodes in relation to the corresponding vectors of linear transformation. The work on solving the problem of huge numbers of missing entries in recommendation system matrices makes use of spectral co-clustering for the purpose [8]. Study has been performed on a popular dataset pertaining to the field of cinema.

The general method requires the computation of a normalized matrix, say, Nm. It is found to be:

$$Nm = [Dx]^{-1/2} * N * [Dy]^{-1/2} \quad (1)$$

Where, Dx and Dy are diagonal degree matrices and N is adjacency matrix formulated from the problem.

Another novel method proposed by a group hailing from the Artificial Intelligence Lab from the University of Arizona, USA, is based on a graph model [9]. Here, the essential subjective factors are considered together with quantitative factors. The method of suggestion is based on the computation of a graph that models consumers, the commodity and the old purchase details. Results are given out by performing a traversal of the constructed graph.

B. Decomposition Methods

To decompose an entity is to reduce it into smaller, meaningful components. It can also be called as factorization where the entity is broken down into its constituent, sub-factors. In the universe of matrices, such decomposition methods are prevalent. In recommenders too, as matrices are used extensively to represent the involved entities and their defining relationship through quantitative surrogates, the factorization is often helpful in processing them.

A study makes use of a specific algorithm known as Singular Value Decomposition in association with the probabilistic Naïve Bayes method to arrive at a prediction system [10]. The general formula that arrives at a factored matrix by Singular Value Decomposition is given by:

$$Matrix = J * \square * K^T \quad (2)$$

Where, J represents an orthogonal matrix of row values; K represents the orthogonal matrix of columns and K^T, is the transpose of the K matrix. $\square$ is the diagonal matrix consisting of non-negative values.

C. Bayesian Methods

These methods which are based on the theorem proposed by Thomas Bayes in the late 1700s are widely applied today and probabilistic in nature. It is used in the process of predicting future outcomes and based on the concept of conditional occurrences observed.

A study makes use of Bayes theory to create a recommender for music [11]. In addition to Bayesian methods, fuzzy methods are also employed. Finally, prediction is done on the basis of utility matrix.

D. Linear Algrabaic Methods

Linear equations and the methods to solve them come under the head of linear algebra which were propounded in the 1600s. Matrix operations go hand in hand with systems of equations. Thus, as systems that predicts and recommends based on quantitative representations of historical preferences are primarily based on matrices, linear algebra forms a part and parcel of the core method. A technical

article speaks at length about a tool comprising several algorithms which has been developed based on random linear algebra [12].

IV. PROPOSED CONCEPTUAL FRAMEWORK

An engine is a general piece of software capable of dealing with different datasets. The framework is proposed for a system with the following attributes:

- Recommender is created for a specific domain where suggestions can impact its current state
- The framework is able to incorporate domains where geographical location information is important
- Sufficient domain knowledge is available and domain experts are also available
- Data pertaining to that domain is also available

This paper proposes a recommender engine framework that will be useful in the following ways:

- Providing information in an easily consumable form on a domain when an analyst seeks to glean new information from evidence
- Unearthing new and hidden factors that affect the domain but cannot be not gained from evidence
- Predicting the preferences thereby completing the vital, missing information
- Providing the end user with a set of number of suggestions

The engine will thus be useful to domain analysts as well as researchers of recommenders. The components that will make up the system are three in number. They are as shown in the figure below:

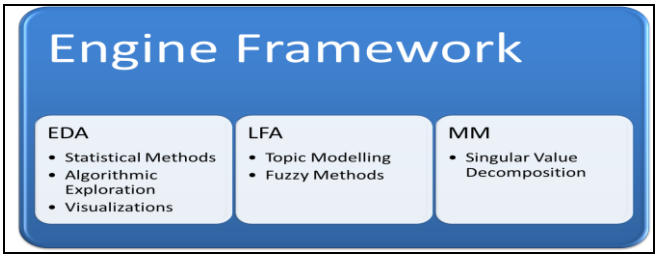

Figure 1. Proposed Framework

A. *Exploratory Data Analysis/EDA*

The first step is to gain an understanding of collected data. Prior to this step it is imperative to collect and study domain related information. Once the data is collected preparing it for its passage through the framework is essential.

Data Exploration involves studying the factors from the dataset that are most essential for the system that is being created. This is proposed to be performed in a three-fold manner: Statistical methods are classic techniques that help discover the average of the variables, the minimum and maximum values, the median, the quantiles, etc. Also, studying the measures of central tendency and dispersion is required. The general formulae are as follows:

Mean = Sum of all Values / Number of Values (3)

Median = [(Number of Values +1) /2] *th* item (4)

Standard Deviation = Square Root of Variance (5)

Following statistical analysis, visualizations are an effective way to explore and understand data. The variants of visualization techniques are:

- Word Clouds: A colourful bunch of keywords where their size depends on the frequency of occurrence
- Histograms: A visual way of understanding the binned data
- Maps: Given longitudinal and latitudinal information, plotting them on real maps can help immensely to make sense of location details

The third step is to employ an algorithm such as K means clustering to group the parameters of interest. This technique is a flat clustering method that can help to comprehend the collected data in an unsupervised manner. Given Geographical details, a two layer presentation of the clusters on a global map is also proposed.

B. *Latent Factor Exploration/LFA*

A research work proposes a stepwise methodology for implementing a recommender by using Latent Dirichlet Allocation and Fuzzy Clustering [13]. While the paper employs it in the creation of a latent feature set for prediction of rating matrix using a modified Bayesian method, the proposed framework is to make use of topic modelling and Fuzzy C means algorithm to unearth latent topics in the domain which cannot be discovered through intuition or from evidence of collected data and to gain an understanding of latent groupings of the components of the utility matrix created for the domain.

Essentially, LDA outputs a theta matrix of latent topic distributions, the top latent topics, the phi and gamma distributions. It is a generative process of modelling a document based on the subjects that occur within it and the words used to describe that subject. It is here that the latent interconnections between seemingly distant topics can be made. Fuzzy C means algorithm outputs a membership matrix, where each item under consideration has some

amount of membership in all the centres that are chosen. That is, each item belongs to several clusters with different centres with varying degrees of membership. In cases where the entities under consideration are multifaceted such a clustering is most needed.

C. Matrix Methods/MM

Finally, the Singular Value Decomposition method is proposed to be used as algorithm for predicting the missing quantitative representation of preferences and enlisting N suggestions. This algorithm is chosen as it is a tried and tested method by several experiments.

Having established the design of framework it is necessary to study the platforms of potential for the implementation of the same. The following table 1 shows the comparison of various languages in the light of creating a recommender. The complete implementation is proposed to be carried out in the R language for socially relevant and unexplored causes.

Table 1: Overview of Main Recommender Libraries

Overview of Main Recommender Libraries			
Parameters/ Language	**Java**	**Python**	**R**
Name	*LibRec*	*SurpriseLib*	*RecommenderLab*
Functionality	Rating Prediction, Ranking	Data Handling, Prediction, Validation	Rating Matrix Prediction, Top Ranked Suggestions
Features	More than seventy algorithms; Less number of lines of code	Provides several baseline algorithms; Provides built-in datasets	Provides methods with several algorithms; Provides methods for dataset inspection and final validation
Source	https://www.librec.net/	http://surpriselib.com/	https://cran.r-project.org/web/packages/recommenderlab/vignettes/recommenderlab.pdf

V. CONCLUSION & FUTURE SCOPE

Thus, the paper has proposed a three-module recommender engine which is envisioned as a first version application and research base for less explored domains that are socially relevant and necessary such as sustainability, Corporate Social Responsibility and Skills matching. Such a framework oriented approach seeks to unify several scattered attempts at domain research that focus only on one aspect- data comprehension or prediction and so on. Firstly, domain understanding through data visualizations and statistical analysis will enable the domain actors to comprehend and assess the historical and present-day evidence. Secondly, latent factor analysis will bring to the fore topics and groupings that are hidden and often vital to bring out innovations in comparatively less explored domains of application. Thirdly, the prediction module solves the problem of completing the matrix of domain particulars that represent the defining factors. And, n number of suggestions can be presented on the basis of the computed matrix. This engine is being built in the R language.

The future scope is promising because the system is designed by taking into account the quality factors of modularity and modifiability. Thus, the three-module engine can be adapted to any similar problem with only few cosmetic changes to the implementation and formatting of the data that is to be passed through the pipeline.

.

REFERENCES

[1] Pera, M.S. and Ng, Y.K. Recommending books to be exchanged online in the absence of wish lists. *Journal of the Association for Information Science and Technology*, *69*(4), pp.541-552., 2018.

[2] Zhang, Z., Lin, H., Liu, K., Wu, D., Zhang, G. and Lu, J. A hybrid fuzzy-based personalized recommender system for telecom products/services. *Information Sciences*, *235*, pp.117-129, 2013.

[3] Yan, B. and Chen, G. June. AppJoy: personalized mobile application discovery. In Proceedings of the 9th international conference on Mobile systems, applications, and services (pp. 113-126). ACM. 2011.

[4] Meehan, Kevin, Tom Lunney, Kevin Curran, and Aiden McCaughey. "Context-aware intelligent recommendation system for tourism." In *2013 IEEE international conference on pervasive computing and communications workshops (PERCOM workshops)*, pp. 328-331. IEEE, 2013..

[5] Aishwarya Rajamani, Alpha Vijayan, "*A Survey on Realms and Applications of Social Media Data Analysis*", International Journal of Computer Sciences and Engineering, Vol.6, Issue.11, pp.844-848, 2018.

[6] Amit Palve, Rohini D.Sonawane, Amol D. Potgantwar, "*Sentiment Analysis of Twitter Streaming Data for Recommendation using, Apache Spark*", International Journal of Scientific Research in Network Security and Communication, Vol.5, Issue.3, pp.99-103, 2017

[7] A.B. Patil, S.C. Pawar, "*Keyword based Marathi Interface to the Database using Natural Language Processing*", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.1, pp.50-55, 2018

[8] Nasiri, M., Minaei, B. and Sharifi, Z. Adjusting data sparsity problem using linear algebra and machine learning algorithm. *Applied Soft Computing*, *61*, pp.1153-1159, 2017.

[9] Huang, Z., Chung, W., Ong, T.H. and Chen, H, July. A graph-based recommender system for digital library. In *Proceedings of*

the 2nd ACM/IEEE-CS joint conference on Digital libraries (pp. 65-73). ACM, 2002.

[10] Puntheeranurak, Sutheera, and Supitchaya Sanprasert. "Hybrid naive bayes classifier weighting and singular value decomposition technique for recommender system." *2011 IEEE 2nd International Conference on Software Engineering and Service Science*. IEEE, 2011.

[11] Park, Han-Saem, Ji-Oh Yoo, and Sung-Bae Cho. "A context-aware music recommendation system using fuzzy bayesian networks with utility theory." *International conference on fuzzy systems and knowledge discovery*. Springer, Berlin, Heidelberg, 2006.

[12] Drineas, Petros, and Michael W. Mahoney. "RandNLA: randomized numerical linear algebra." *Communications of the ACM* 59.6 (2016): 80-90.

[13] Liang, Qianqiao, Xiaolin Zheng, Menghan Wang, Haodong Chen, and Pin Lu. "Optimize Recommendation System with Topic Modeling and Clustering." In *2017 IEEE 14th International Conference on e-Business Engineering (ICEBE)*, pp. 15-22. IEEE, 2017.

Authors Profile

Aishwarya Rajamani pursued Bachelor of Engineering in Computer Science at M V Jayaraman College of Engineering, Bangalore, India between 2010 and 2014. She is currently pursuing her Masters in Computer Science(M. Tech) at New Horizon College of Engineering, Bangalore, India. Her technical areas of interest are Cloud Computing, Data Science and Machine Learning.

Alpha Vijayan completed Bachelor of Technology in Computer Science at College of Engineering, Chengannur, Kerala in the year 2000. She holds a Master of Engineering degree in Computer Science from Jerusalem College of Engineering, Chennai. Currently, she is pursuing PhD. in Computer Science and Engineering from Jain Deemed to be University, Bangalore. She has a total of 16.8 years of teaching experience.

