

Unsupervised Machine learning Approach for Book Recommendation System

Dr. Mrs. Radha A. Pimpale

Asst. Prof.

Department of Information Technology

PriyadarshiniBhagwati College of Engineering, Nagpur, MS

Mobile No-9423073189

Article Info

Volume 83

Page Number: 9572 - 9579

Publication Issue:

March - April 2020

Article History

Article Received: 24 July 2019

Revised: 12 September 2019

Accepted: 15 February 2020

Publication: 11 April 2020

Abstract:

In Today's Internet Era, thanks to multiple choices for purchasing product every body observe review and rating of the merchandise. Vast amount of information available on Internet for same product and more number of possibilities so continuous browsing waste lots of your time around on Internet. For helping the user on Internet we proposed a wise and Intelligent book recommendation system which is unsupervised machine learning supported document cosine similarity and association rule mining referred to as hybrid Book Recommendation System. The Proposed system is recommended the Book to the New User cold start supported their cosine similarity. More advanced techniques like unsupervised machine learning like association rule, we learn by experience and performance, given Hybrid Filtering techniques produce more accurate results which produce impact on book browsing and buying.

I. Introduction

Individuals have a tendency to believe on a book for reading which is recommended by the related individual by rating and reorganization for buying. Traditional Techniques such as popularity based book recommendation, correlation based book recommendation and content based book recommendation has more challenges and limitation than our new approaches, The task of book recommendation can be done very easily. In this paper we are making Proposal outline, information is filtered here by utilizing unsupervised machine learning algorithms. The recommended system consider keyword as information to find cosine similarity in title of the book.

In the event, if totally new user visits an e-commerce location, that location will not have any past history of that client. Our techniques find the related book by clustering using cosine similarity of book title. Its brought need and interest it create a positive impact on the user experience and lead to frequent visits and read and purchase relevant Books. If user having past history we consider this for association rule

mining, where frequent buying of item set recommended such as HaryPoter –series1, series2, series 3 etc. we also applying new rules of association rules.

As we have a monstrous measure of information over-trouble over Internet. It finds a workable pace considerable undertaking for the client to instigate the important information. A recommender system gives an answer for this by disseminating proposition for the things to be of use to a client. Recommender structure serves the "right" thing to the customers in a motorized plan to advance long haul business targets. Automation comes through calculations. Likewise, using proposition system ventures can find a good pace 'purchasing practices better' and create compelling show packaging strategies to pull in various clients.

Right now are inspecting a couple of ordinary methodologies for separating suggestion framework in detail their bit of leeway and limitation In Content-based Filtering Systems ,It Uses data of dynamic clients and information about the things. Community Filtering Systems It Uses data about a lot of clients and their relations with the thing to give suggestions to the dynamic client.

Segment Filtering Systems, Uses segment data, for example, age, sexual orientation, instruction, and so on of individuals for recognizing sorts of client. Right now are utilizing unaided AI procedures for books recommendation framework

II. Review of Literature

Straightforward recommenders offer summed up proposals to every client, bolstered ubiquity and additionally type. the fundamental thought behind this strategy is that Book which are progressively well known and having a superior likelihood of being preferred by the regular peruser are considered for suggestion. Peruser prescribed books of Popular writer.

Segment Filtering Systems utilizes segment data, for example, age, sexual orientation, training, and so on of individuals for distinguishing kinds of client. In Book suggestion framework age criteria is generally significant for book recommendation[4]. The greater part of the book are recommandated dependent on prior information on perusers age gathering and feeling about territory of intrigue. It needn't bother with history of book clients evaluations and loving. This is exceptionally simple and straightforward technique. Right now uncommonly the age quality is generally significant for instance books, for example, Harry Potter might be perused by age bunch 10 to 20 age gathering. Sexual orientation property is additionally significant Books dependent on cooking, sewing might be perused by Female. Books. For web based perusing Location savvy books having more impact. Content Based Filtering is client freedom techniques during this strategy we are suggest books which are related with each other for example Harry Potter Series or The Chronicles of Narnia like books. In content-based separating framework check the closeness between various Books which is upheld the quality of item, Distances like cosine, Eucludian separation, Pearson's Correlation is determined between at least two things, likeness will be determined. Proposal framework bolstered content moreover as clients conduct. Term recurrence is determined to gauge a watchword in any archive and relegate

the significance thereto catchphrase bolstered the measure of times it shows up inside the record also on the grounds that the IDF (reverse report recurrence) of a word is that the proportion of how critical that term is inside the entire corpus for the similitude what's more as importance figuring.

W = Number of times term w shows up in a report

N = Total number of terms in the record

$IDF(t) = \log_e(\text{Total number of records/Number of archives with term } w \text{ in it})$.

In community oriented separating need clients' verifiable inclination on a lot of things. client thing accociation network methodology considered. It additionally assessed on rating of person which might be sure or antagonistic. Communitarian separating has two sub classifications are memory based and model based methodology. Unequivocal Rating, is a rate given by a client to a thing on a sliding scale, similar to 5 stars for item. This is the most immediate input from clients to show the amount they like a thing. Certain Rating, proposes clients inclination in a roundabout way, for example, site hits, clicks, buy records, regardless of whether tune in to a music track, etc. Mooney proposed a substance based book recom-mendation procedure, right now separated and AI calculation are applied called LIBRA (Learning Intelligent Book suggestion Agent) [10]. Jomsri proposes a library book suggestion framework dependent on client profile advancing and affiliation rule [11]. Tewari A S, Priyanka K present a procedure User based synergistic sifting and affiliation rule mining[12].

III Proposed System

Proposed framework depends on Machine learning calculation. ML Algorithms are classes into four classifications into regulated, semi-managed, solo and fortification learning calculations.

Regulated Machine Learning

In regulated learning Regression is measurable based which may utilized preparing informational collection and test informational collection model

applied methods are applied strategies are choice trees, bolster vector machine, irregular timberland, k-closest neighbor, closeness learning, neural system, direct relapse, calculated relapse, and Naïve Bayes.

In regulated learning Classification, information things are classifications upheld their element are the fundamental destinations of managed learning. While arrangement recognizes the classification or class of most recent information things which is anticipated bolstered experience picked up from preparing information.

Solo Machine Learning

Solo learning (USML) strategies, bunch the comparable information things into same classification and unlabelled and different into another gathering. Grouping and affiliation rule mining are significant strategies of solo AI. In the event that the highlights of information things are increasingly heterogeneous, at that point there will be progressively number of groups and the other way around. Bunching is unaided learning strategies of AI the objective of our framework to look through book dependent on cosine then affiliation rule mining are executed to prescribe the books to the new client. Affiliation rule mining is said to discovering rules which show the pattern of some arrangement of exercises which can happen in a steady progression. Like discover the buying conduct of shoppers by observing their buy history or discover the getting example of book borrowers in libraries or predicting the following site someone will visit by seeing his internet surfing history.

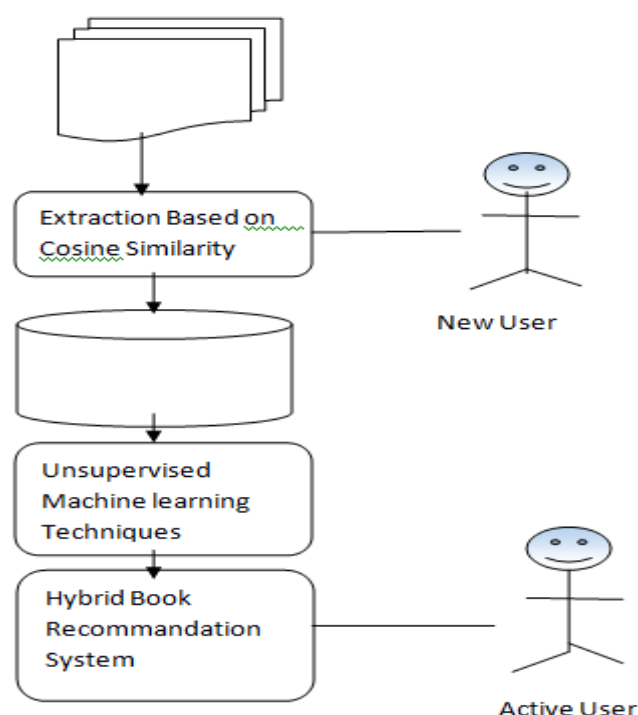


Figure Shows Overview of Proposed Work

IV Algorithmic Steps For Proposed Work

1. Data Collection
2. Data Storage
3. Data Filtering
4. Transforming and normalizing data
5. Training models
6. Selecting the optimal model
7. Evaluating model performance

1. Data Collection

The data can be collected by explicitly and implicitly. Explicit data is information that is provided purposely such as input from the users such as rating of product, movie ratings, game

rating etc. Implicit data is information that gathered from available data streams like search history, clicks, order history etc. Here engineering students books dataset is used. This dataset has been compiled and it comprises of three tables for users as Studentid, Bookid, Bookname, Author, and price.

2. Data Storage

Data can be store include a standard SQL database or some kind of object storage.

3. Data filtering

After collection and storage, to filter data to extract the related information required to make the final recommendations.

4. Transforming and Normalizing Data-

Normalization can be done on data and Data has been transform into csv file for processing

Normalize item values across users

To do this, we normalize purchase frequency of each item across users by first creating a user-item matrix as follows

```
df_matrix = pd.pivot_table(data,
values='borrower_count',
index='BookId',
columns='BookId')
```

Data acquisition and selection

The dataset is available on the UCSD website or use general Dataset.

Data cleaning

To bring the data into a consistent format, steps taken are:

Drop unnecessary columns

Drop duplicate records

Check for invalid data

Check ranges for applicable columns (such as ratings between 1 and 5)

Deal with missing values and outliers

Split train and test set

For predictive modelling we are splitting data into 75: 25, one for training data and other for testing data.

6. Assessing model execution Model Performance is assessed by applying cosine likeness and affiliation rule mining to the information, thus backing, certainty and lift is determined and afterward we determined Accuracy, Precision and Recall and F-measure..

V. Proposed Work

In our proposed work we are utilizing solo learning calculations bunching here assessment grid is utilized cosine likeness for book title , on the off chance that similitude is high, at that point the related books are prescribed to the new client. At that point we are applying affiliation rule digging for the books which are in arrangement for proposal.

Our work is to consolidate cosine closeness and affiliation rules mining to give a structure of a half and half proposal framework on two dimensional spaces

We utilize the preparation and test datasets of library. Furthermore, we use PYTHON programming to create cosine likeness of vectors or tile of books and afterward we apply affiliation runs the information mining undertakings that are required to actualize and test the proposed system

Cosine Similarity: It is the most normally utilized likeness metric in content examination. The closeness of content information is estimated by the littlest edge between two vectors. The point (Θ) is thought to be somewhere in the range of 0 and 90. A snappy boost: $\cos(\Theta = 0) = 1$ and $\cos(\Theta = 90) = 0$.

In this way, the most extreme disparity between two vectors is estimated at Cos 90 (opposite). Also, two vectors are said to be generally comparable at Cos 0 (equal). For two vectors (x,y), the cosine comparability is given by their standardized spot item demonstrated as follows:

$$\text{cossim}(x, y) \leftarrow \frac{\text{dot}(x, y)}{(\sqrt{\text{dot}(x, x) * \text{dot}(y, y)})}$$

Cosine Similarity Example

We should assume you have putting word "Programming in C" in web index. Books dependent on a title. "The C Programming Language", "Programming in ANSI C", "Commonsense C Programming", "Programming in C", "programming with c" are going first which have high cosine closeness in title and afterward books like

Think about chapter by chapter guide of the books to check Similarity Matrices

Thought about 2 words to check closeness networks

Document- Term Matrix

Word Counts	Programm ing In C	Programm ing in ANSI C	Programm ing with C
C	1	1	1
Programm ing	1	1	1

Similarity Matrix

Similarity or distance matrix	Total Common Words	Cosine Similarity
Programming In C	2	0.86
Programming in ANSI C	2	0.86
Programming With C	2	0.84

Measure similarity between two sentences using cosine similarity

In our proposed examination we measure the similitude between two title of book utilizing cosine closeness for this we utilized the capacity nktl.tokenize which partition a major amount of content into littler parts as word and the nltk.corpus capacities gives us stop words, for example, 'a','an','the' and so on we can discover comparability cosine according to this we prescribe the book to new client as cool beginning and afterward we apply affiliation rule for additional suggestion

Capacities utilized in the given System are

nltk.tokenize: It is utilized for tokenization. Tokenization is the procedure by which enormous amount of content is partitioned into littler parts called tokens. word_tokenize(X) split the given sentence X into words and bring list back.

Cosine comparability is a proportion of closeness between two non-zero vectors of an internal item space that gauges the cosine of the edge between them.

Comparability = $(A.B)/(\|A\|.\|B\|)$ where An and B are vectors contains title of books.

Affiliation Rule Mining

Affiliation Rule Mining (ARM) is solo AI strategies which to find the connection between book titles of a huge dataset. It has two stages Minimum help and second is certainty.

In least emotionally supportive network, fixed least recurrence are sifted through. This base recurrence is known as help. It gives the portion of exchanges which contains it Book An and B. Fundamentally Support informs us concerning the every now and again purchased Books or the blend of books purchased as often as possible. We can channel here books of low recurrence

Backing: It reveals to us how regularly the things A happen in given the number occasions, by absolute exchange.

Support(A)=(Transactions containing (A))/(Total Transactions)

Certainty: It discloses to us how frequently the things An and B happen together, given the number occasions A happens.

Confidence(A→B) =((Transactions containing both (An and B)))/(Transaction containg A)

Lift: Lift demonstrates the quality of a standard over the arbitrary event of An and B. It essentially discloses to us the quality of any standard.

Lift=Support/(Supp(A)*Supp(B))

we utilized similitude and affiliation rule digging for the proposal of books. We have assessed the framework execution by posing inquiries. We have structured keen suggestion framework for books utilizing affiliation rule mining strategy. To execute the calculation they have utilized Python coding have gathered surveys about books from web sources and after pre-preparing those audits

	Antecedant s	Consequen ts	support	confiden ce	lift
0	Programmi ng In C	Programmi ng In C++	0.1709 18	0.59701 5	3.5459 07
1	Programmi ng In C++	Programmi ng In Java	0.1683 67	0.60606 1	3.5459 07
2	Programmi ng In Java	Programmi ng in C#	0.1683 67	0.53030 3	3.8496 07
3	Programmi ng in C#	Programmi ng in Python	0.1377 55	0.64814 8	3.8496 07

	Antecedant s	Consequen ts	support	confiden ce	lift
4	Programmi ng In C	Programmi ng in C#	0.1709 18	0.61194 0	4.4422 33

The point of utilizing affiliation rules is to find designs like perusers who have perused book Programming in C would likewise decide to peruse book Programming in C++. It is appeared in Table, rule 1 has certainty esteems more prominent than help an incentive to locate the intriguing guidelines and examples that could help our framework in exhibiting exact and satisfactory suggestions. The point is to find such principle that fulfills client determined least help and high certainty score.

Proposed half and half procedures

Openly accessible archives and online destinations which we chose to download books. In structuring our book prescribed framework we utilized books identified with the building understudies which would be effectively accessible. Book Title and book substance and use for cosine likeness. At that point we apply affiliation rule visit book understanding strategies and the locate the quantity of client utilizing books and the discover the measure like exactness review and F measure for additional translation. Here we considered 25% testing book information who adhere to the affiliation rule and 75% as book information considered as train information. Which creates the outcomes dependent on following recipe.

Certainty shows the occasions the in the event that/at that point explanations have been seen as obvious. On the off chance that framework propose the client that books are intrigued and helpful, at that point it is True positive. In the event that the books are not intrigued, at that point it has bogus positive. At the point when

framework can't give intrigued proposal then it is known as bogus negative.

Here we considered book information of designing understudies, 25 percent information considered as testing information which is originating from affiliation rule and 75% as preparing information. Result are determined by

by Formula

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{P} + \text{N})$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F-measure} = 2\text{TP} / (2\text{TP} + \text{FP} + \text{FN})$$

We get results from proposed half breed framework where test data are considered from affiliation rule and train information, we got result for book suggestion framework

Exactness: 0.8920 Precision: 0.826576 Recall: 0.853175 F1 score: 0.844794

VI. Conclusion

The point of this paper is to introduce a book suggestion framework dependent on AI where cold beginning with new client effectively make do with of vector put together cosine comparability with respect to books proposal. The primary accentuation of this methodology is the recognizable proof of cosine likeness based books and connected based books with the framework. We structured and built up the framework for book suggestion framework in Python. The idea of suggestion is additionally acquainted right now make trust based proposal calculation cosine comparability is vigorous to deal with the recommender framework. There are numerous information investigation instruments accessible to the python examiner A helpful system is called affiliation examination which endeavors to discover basic examples of things in enormous informational indexes called advertise crate investigation. It all the more testing to know

which ones to use in a specific circumstance. We reason that our proposed half and half methodology beats customary methodologies for book suggestions.

References

- [1] <https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-recommendation-engine-python/>
- [2] <https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-recommendation-engine-python/>
- [3] F.O.Isinkaye^a, Y.O.Folajimi^b, B.A.Ojokoh^c, "Recommendation systems: Principles, methods and evaluation", <https://doi.org/10.1016/j.eij.2015.06.005> access on 26/11/2019.
- [4] Dr. Sarika Jain, Anjali Grover, Praveen Singh Thakur, Sourabh Kumar Choudhary, "Trends, Problems And Solutions of Recommender System" ISBN:978-1-4799-8890-7/15/\$31.00 ©2015 IEEE, 955-958 International Conference on Computing, Communication and Automation (ICCCA2015)
- [5] Silvana Aciar, Girona, Spain, Debbie Zhang, "Recommender System Based on Consumer Product Reviews", Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06), 0-7695-2747-7/06 \$20.00 © 2006
- [6] Weiwei Yuan¹, Donghai Guan^{2*}, Linshan Shen¹ and Haiwei Pan¹, "Finding Suitable Number of Recommenders for Trust-Aware Recommender Systems: An Experimental Study", 978-1-4799-4808-6 /14/\$31.00 ©2014 IEEE.
- [7] Nicholas Jing Yuan, Member, IEEE, Yu Zheng, Senior Member, IEEE, Lihang Zhang, and Xing Xie, Senior Member, IEEE, "T-Finder: A Recommender System for Finding Passengers and Vacant Taxis", 1041-4347/13/\$31.00 _ 2013 IEEE Published by the IEEE Computer Society
- [8] Zafar Ali, Shah Khusro, Irfan Ullah and Association Rule Mining", *ACM* 978-1-4503-4062-5/16/05\$15.00 DOI: <http://dx.doi.org/10.1145/2908446.2908481>, 18 JUNE 2018
- [9] Silvana Aciar¹, Debbie Zhang², Simeon Simoff³, John Debenham⁴, "Recommender System Based on Consumer Product

Reviews” Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06) ,0-7695-2747-7/06 \$20.00 © 2006

- [10] Mooney RJ, Roy L. Content-based Book Recommending using learning for text categorization. Proceeding 5th ACM Conference on Digital Libraries; San Antonio, USA. 2000. p. 195–204. CrossRef.
- [11] Jomsri P. Book recommendation system for digital library based on user profiles by using association rule. Proceeding 2014 4th International Conference on Innovative Computing Technology (INTECH); Luton. 2014. p. 130–4. CrossRef.
- [12] Tewari AS, Priyanka K. Book recommendation system based on collaborative filtering and association rule mining for college students. Proceedings of 2014 International Conference on Contemporary Computing and Informatics (IC3I); Mysore. 2014. p. 135–8.