



Book Recommendation System using Machine learning

Fatima Ijaz

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

October 9, 2020

Book Recommendation System using Machine learning

Fatima Ijaz

Department of computer science
Riphah international university, Lahore, Pakistan
Fatimaijaz084@gmail.com

Abstract

Machine learning is a scientific study of statistical model and algorithms. In this research I will use the machine learning algorithms, K-NN and matrix factorization. In the books recommendations system BX books dataset is used. Suggestion method is a selection strategy which was used for collective selection and material-based sorting strategies. Pattern filtering technique is carried to suggest a consumer to an element the "rank" or "first option." Suggestion process collected information was about either the customer's first option on unusual subject relevant to films, books, travel, TV and commerce, etc. And from the other side, an effective selection of books recommendation system design utilizes prior scores or background of the customer. Cooperative sorting is a process of measuring and processing the categories across user opinions. Cooperative filtering first gathers the rankings or a preference of books provided by multiple users and then suggests books to different individuals based on various previous tastes and preferences. K-Means Multipathing together with K-Nearest Neighbor is applied on the BX dataset to achieve the greatest-optimized outcome. In prior methodology, the information is dispersed and ends in a high amount of matrices, whereas the information is collected throughout the suggested strategy and concludes in a small number of groupings. The preferred framework forecasts the customer's desire for a book based on various criteria. These consumers will affect their views on one another. It maximizes the succession and has smaller RMSE.

Keywords

Recommender system, book recommendation, collaborative filtering, KNN, matrix factorizations.

1. INTRODUCTION

Recommendation system is defined as the computer program that helps the user determine goods and content by predict the users rating of each item and presentation them the substance that they would rate highly. The recommendation system is containing three types that are: collaborative filtering, content based filtering and hybrid filtering.

1.1 Collaborative filtering

The collaborative filtering is based on customer's behaviors' and past likes. It os used to make automatic predictions, filtering collecting preferences, interest of a user and taste information of many users. .Collaborative filtering can be performed in two ways mainly, model-based and memory-based.

1.2 Memory based

In memory-based collaborative filtering, the whole dataset is used to make a recommendation.

1.3 Model based

This method does not use the complete dataset for generating recommendations. The model then makes recommendations on the test data.

There are three steps in collaborative filtering recommendation. 1) Established a user model. 2) Find the nearest neighbor user 3) finally (Gao, 2019)generates a recommendation list.

System of suggestion progressively that is used in fields, such as films, traveling, songs, books etc. Increasing social acts have amplified the use of recommending programs in persons and community recommending programs. Community suggestion structures also address the issue of cold starting which occurs within a person recommendation engine. This work provides a report on the latest technology relevant to several areas of community optimization algorithms. As for their accumulation and customer obvious sign designs, the scientist addressed prior structures. A certain organization is quite important to study each new domain complexities.

The current recommendation method does have a great performance, according to both the undertake experimentation, than numerous prior implementations, including the praised k-NN algorithm being used by suggestion especially at longer length. System of suggestion progressively that is used in fields, such as films, traveling, songs, books etc. Increasing social acts have amplified the use of recommending programs in persons and community recommending programs. Community suggestion structures also address the issue of cold starting which occurs within a person recommendation engine. This work provides a report on the latest technology relevant to several areas of community optimization algorithms. As for their accumulation and customer obvious sign designs, the scientist addressed prior

structures. A certain organization is quite important to study each new domain complexities.

We constructed a book recommender using collaborative item. For this aim, we used an open-source recommendation system package for Python. We took up the process of adding our novel approach to reduce the RMSE value of the pure ITEM-based filtering. The other frameworks we used were Pandas; for handling large datasets, SciKit learns for running the K-nearest neighbors algorithm and NumPy for its data structure capabilities. Our choice of programming language was Python.

The construction starts with the part1 with both the fundamentals of the content delivery system as Description. Part 2 addresses previous researchers' latest novel with the descriptions of optimization techniques in use by different writers. Segment 3 outlines the implementation of the proposed plan of suggestions. Segment 4 I will discuss the problems statements. The development of the current scheme is shown in Part 5. Part 6 addresses system operation and outcomes with either the aid of the program's screenshot. Part 7 of the current system shall have the summary and future products.

2. LITERATURE REVIEW

[1] Suggestion framework is a very common and cold e-commerce issue. Recommendation system performs in multiple ways including faculty member based on quality, suggestion for reciprocal filtering, and suggestion for the combination technique. This article proposes a collective suggestion filtering system focused on naive Bayesian approach. The current recommendation method does have a great performance, according to both the undertake experimentation, than numerous prior implementations, including the praised k-NN algorithm being used by suggestion especially at longer length.

[2] The current recommendation method does have a great performance, according to both the undertake experimentation, than numerous prior implementations, including the praised k-NN algorithm being used by suggestion especially at longer length. System of suggestion progressively that is used in fields, such as films, traveling, songs, books etc. Increasing social acts have amplified the use of recommending programs in persons and community recommending programs. Community suggestion structures also address the issue of cold starting which occurs within a person recommendation engine. This work provides a report on the latest technology relevant to several areas of community optimization algorithms. As for their accumulation and customer obvious sign designs, the scientist addressed prior

structures. A certain organization is quite important to study each new domain complexities.

3. Problem statement

In the recommendation system the problem is trying to forecast the option the users will have on the dissimilar substance and be able to recommends the finest items to each user. Another some problems in recommendation system are data sparsity, scalability and gray sheep. Data sparsity means the data is widely spread; it has null values and missing values. Scalability means the prediction is difficult in huge amount of ratings items. Gray sheep means the time and memory requires problems.

4. Research questions

- What dataset is given for recommender system?
- How can in combine the clustering results of products and user rating matrix?
- How I can do recommendation system evaluation?
- How to find the accuracy of recommendation system?

5. Methodology

5.1 Process diagram

Fig 1 displays the Book Recommendation Model flow graph method. This chart shows the flow of the planned program operation. Method flow demonstrates how well the system works, how much the system handles the relevant data, and how ranking is expected by the system.

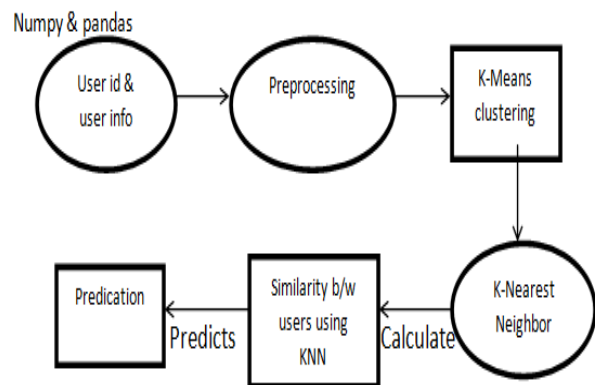


Figure 1: flow diagram

- The dataset gives the kaggle and it has three data files users, books and ratings.
- Using the numpy, pandas, sklearn and Matplotlib library the separate the raw data frames and processing.
- Squared techniques are used to find the correct number of clusters so that K-means clustering can be functional to the books.

- After applying K-means clustering a clustered matrix factorization is build which defined average rating the user gives to each cluster.
- Pearson correlation similarity between the users is calculated.
- Finally find the accuracy through root means square error.

6. Results

Python libraries import: Numpy, Pandas, MAplotlib, and sklearn. The dataset contain three excel files. The dataset of 90,000 users includes 1.1 million reviews of 270,000 books. The scores are on (1-5) scale.

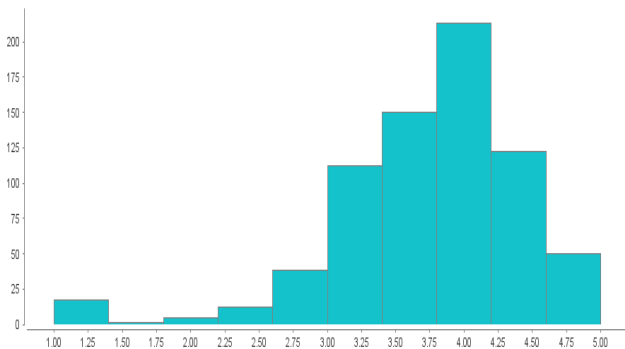


Figure 2: Rating distribution

First count the ratings of books from collaborative filtering. The total rating count is 0.990.

```
print(book_ratingCount['totalRatingCount'].quantile(np.arange(.9, 1, .03)))
0.900    5.000
0.930    7.000
0.960   11.000
0.990   31.000
Name: totalRatingCount, dtype: float64
```

Figure 3: Total rating count

6.1 Accuracy

To calculate the accuracy of our predictions I available to use an ordinary statistical metric named root mean square error (RMSQ). RMSE is a measurement of the variation between the user's real books ratings and the, I predicted for the same books. If the lower the RMSE, the more acceptable the model. An RMSE of zero means our model is absolutely guess the user ratings.

rms error is: 0.3872849941150143

Figure 4: Root mean square error

7. Discussion

The structure was implemented in python programming languages utilizing K-Means cluster library and K-NN. The control system success consists of multiple post-sections that are also normal procedures to be accompanied while resolving any deep learning issue. There are fellows:

The first phase in the development method is the compilation of the information. The appropriate set of data is chosen in this phase in terms of achieving additional calculations. The database is collected from either the kaggle official website in the context of books recommendation program BX Books. The dataset contains 1048576 books from (1-10) ranking. In addition, it has 276272 consumers and 271380 books. With that kind of knowledge, additional calculations are carried out using software package Python

The second measure in the deployment phase is the method of preparing information. Preprocessing of the information is performed in this stage. It depicts the additional power vector that informs which user has classified that book. This is accomplished by dividing user information and book data in and out of multiple data blocks, first. Then, the vector factorization is generated just using the information points.

The second option in the deployment phase is the processing of the information. The grouping method of K-Means is implemented in this phase. The proper amount of cluster is selected. Using the K-Means Segmentation system, before selecting the correct no of group books are separated across groups.

RMSE is measured to determine the designer's reliability.

CONCLUSION

ML is a statistical process that automatically builds test model structures. It is primarily a AI contingent based on the idea that the system can test records, understand the style and make the choices from the smallest individual intervention. This proposed device includes a book suggestion gadget using K-Menus Clustering and the K-NN algorithm. Facts BX books are derived from facts. The device is imposing strange programming language. It is seen that when the device is implemented in Python programming language, the RMSE fee of the predicted technique goes far beyond the previous method. Likewise, it has been seen that the RMSE cost of the proposed gadget is meeting the same fees as the present approach, but without a lower number of commenters. The proposed system is find accuracy through root mean square error (RMSE) that is 0.3. The proposed career can be further expanded by using additional information units. The concept of critical analysis can be used to illustrate the performance of the book's recommendation system in destiny, so the model can be adjusted to accommodate

further conditions. In destiny, there may be an interest in a male or female feature that is hidden in consumer advice.

References

- [1] Ahuja, R. (2019). *Movie Recommender System Using K-Means Clustering AND K-Nearest Neighbor*. IEEE.
- [2] Dara, S. (2019). *A survey on group recommender systems*. Journal of Intelligent Information Systems.
- [3] Gao, Y. (2019). *Research on Book Personalized Recommendation Method Based on Collaborative Filtering Algorithm*. IEEE.
- [4] M, I. (2020). *Predicting Books' Overall Rating Using Artificial Neural Network*. International Research Journal of Engineering and Technology .
- [5] Parvatikar, S. (2015). *Online Book Recommendation System by using Collaborative filtering and Association Mining* . IEEE.
- [6] Ramakrishnan, G. (2020). *Collaborative filtering for book recommendation system*. SocProS.
- [7] Shah, K. (2019). *Book Recommendation System using Item based Collaborative Filtering* . International Research Journal of Engineering and Technology .
- [8] Sohail, S. S. (2013). *Book Recommendation System Using Opinion Mining Technique* . IEEE.
- [9] Tan, K. W. (2019). *A New Collaborative Filtering Recommendation Approach Based on Naive Bayesian Method*. Beijing: ResearchGate.