



Application of Machine Learning in Collaborative Filtering Recommender Systems

Rajesh Kumar Ojha and Dr. Bhagirathi Nayak

Sri Sri University, Odisha, India

*Corresponding author E-mail: rajesh.ojha@gift.edu.in

Abstract

Recommender systems are one of the important methodologies in machine learning technologies, which is using in current business scenario. This article proposes a book recommender system using deep learning technique and k-Nearest Neighbors (k-NN) classification. Deep learning technique is one of the most effective techniques in the field of recommender systems. Recommender systems are intelligent systems in Machine Learning that can make difference from other algorithms. This article considers application of Machine Learning Technology and we present an approach based a recommender system. We used k-Nearest Neighbors classification algorithm of deep learning technique to classify users based book recommender system. We analyze the traditional collaborative filtering with our methodology and also to compare with them. Our outcomes display the projected algorithm is more precise over the existing algorithm, it also consumes less time and reliable than the existing methods.

Keywords: Recommender Systems (RS), Collaborative filtering, k-Nearest Neighbors, Classification.

1. Introduction

The internet plays a pivotal role in the day to day life of the current generation. People are growing more internet savvy even in day to day decision making which has ignited a fierce competition among the different e-commerce companies. E-commerce websites are getting better day by day using the latest technologies for prediction of the user interest and recommend them items or services by using their past experience and behavior. Recommender Systems are very popular & extensively used for prediction about user interests from the historical data. This article suggests a model of recommender system which filters most important information from a significantly large volume of data and provides personalized services. Recommender systems are very popular applications those recommend items or services suitable for a particular user by analyzing the user's interest in that particular item or item similar to that. Recommender system also analyses the interaction between user and items, user using or preferring similar items to collect group intelligence and predict the user's preference over items or services. Deep analysis of the behavior of the user and other users associated with the similar kind of products reveals the preferences and interests of the users. This kind of analysis is done by recommender systems using various data mining and prediction algorithms, and the users are characterize on behalf of their preferences and interests on a particular item. Recommender systems have emerged with the implementation of Machine Learning but basically they are classified as content based, collaborating filtering and hybrid systems.

2. Literature review

We explore the different dimensions of Recommender Systems by

going through several articles and literatures, some related works are discussed in this section. We have reflected some Collaborative Filtering approaches by author's name and the title of the article. By Q. Wang, M. Sun, X. Yuan, [1] "Collaborative Filtering Recommendation Algorithm based on Hybrid User Model". Collaborative filtering overcomes the problems of adaptability but ineffective in case of sparsely distribution of information hence the precision of recommendation is not guaranteed so the article proposes a hybrid user model to overcome the loopholes of both Collaborative Filtering and Content Based Filtering. By M. O. Mohany, N. Hurley, N. Kushmerick and G. Silvestre [12] "Collaborative Recommendation: A Robustness Analysis" investigates the ability to make robust recommendations corresponding to "accuracy" and "stability". It also provides a frame work to examine the stability of recommendation given by different widely used collaborative filtering algorithms. By X.L. Lam, T. Vu, T.D. Le and A. D Duong [13] "Addressing cold-start problem in Recommender Systems" suggested hybrid model for Recommender Systems combining Collaborative filtering with user's information to deal with cold-start user-side problem. This paper has used the Movie Lens Group data and the metrics used for evaluation of the accuracy of the prediction are MAE and NMAE metric. By A.I. Schein, A. Alexandrin Propescul, L. H. Ungar and D. M. Pennock [14] "Methods and Metrics for Cold-start Recommendations" uses a Naïve Bayes Classifier and forms a heuristic recommender system by introducing a new performance metric, the Customer Receive Operator Characteristic (CROC) and proven it to be effective. By Hao Ma [15] "An experimental study on implicit social Recommendation" suggests how to incorporate implicit social information of a customer or a product in the absence of explicit social information and do effective prediction about user preferences and choices. By B M Riedl [16] "Application of Dimension Reduction in Recommender System – a case study" describes how the overhead of data analysis can be decreased by

reducing the number of parameters in the recommender system database by using Singular Value Decomposition (SVD) approach with collaborative filtering.

3. Machine Learning

Machine learning has drawn its root from the computational intelligence or earlier known as soft computing. It provides smart and heuristic algorithms which makes machines adaptable and develops self-learning capability. The machine learning algorithms learns by historical data and build association rules by themselves and after successful training they predict the results for unknown patterns which were not part of the historical data set. Machine Learning is otherwise known as an application of artificial intelligence (AI). One of the most important feature of machine learning is to deal with imprecise data or missing data with tremendous fault tolerance capability. So the cold-start problem in collaborative filtering may be dealt efficiently by implementing machine learning algorithms due to scarcity of data.

Different Machine learning algorithms may be classified as:

- **Supervised (learning in presence of a moderator)**
- **Unsupervised (learning from experience)**
- **Semi-supervised (occasional involvement of a moderator)**
- **Reinforcement (learning in presence of a critic)**

A big difference between Statistical prediction and machine learning is the adaptability or self-learning capability of machine learning algorithms. Not only they learn quickly from a huge amount of data but also generate more accurate results in order to identify opportunities or risks. So a combination of machine learning and cognitive techniques may yield better performance

3.1. Machine Learning for Recommender System

Recommender systems are one of the most successful and widespread application of machine learning technologies in business. Machine learning algorithms in recommender systems are typically classified into two categories—content based and collaborative filtering methods although modern recommenders combine both approaches. Content based filtering methods are based on similarity of item attributes and collaborative methods calculate similarity from interactions.

- **Content-based Filtering**
Recommends items or services based upon items liked in the past, by user profiling and profile matching.
- **Collaborative Filtering**
Recommends items or services based upon the matches of users having similar interests, by statistical analysis of patterns and analogies of data extracted explicitly from evaluation of ratings given by different users.

Table 1: Recommendation task

	b-1	b-2	b-3	b-4
u1	5	1	3	5
u2	?	?	?	2
u3	4	?	3	?

Table shows U as set of users and B as set of items (Books) to be recommended for the user, based upon collaborative filtering and k-NN. The mapping between utility of each item $b \in B$ to each

user $u \in U$ is to be learned from past data by a function f . The methodology of study is mainly revolved around laboratory and in-situ testing. Water samples are taken at retention pond UiTMCJKPG (see Fig. 2) and kept in Environmental Laboratory at Faculty of Civil Engineering. 16 sampling points are selected for this study and each sampling point consists of 1 water sample. The water samples are taken 15 cm to 30 cm below water surface. In-situ testing which covers pH, turbidity and concentration of

Table 2: User-based k-nearest neighbors

	b-1	b-2	b-3	b-4
u1	5	1	3	5
u2	?	?	?	2
u3	4	?	3	?

Compute similarity of users, find k most similar users to user a. Recommended book not read by user a. Cosine similarity

$$\text{sim}(a,b) = \frac{a \cdot b}{\|a\| \cdot \|b\|}$$
 The simplest algorithm computes cosine or correlation similarity of rows (users) or columns (items) and recommends items that k-Nearest Neighbors.

We have used the data set given by Group Lens Research Project in which the ratings take value over a range of 1 to 5. More than 10 lakh of data are taken from the two files “books ratings.csv” and “books.csv”. Only four fields have been considered to predict user ratings using collaborative filtering. Figure -1 shows a snapshot of the file “ratings.dat” which holds the four fields i.e User, ISBN, Book-Title, Book-Ratings.

Row No.	User	ISBN	Book-Title	Book-Rating
1	8	1567407781	The Witchfind...	3
2	8	1575663937	More Cunnin...	4
3	176875	1575663937	More Cunnin...	4
4	8	1881320189	Goodbye to th...	4
5	35826	1881320189	Goodbye to th...	4
6	226745	1881320189	Goodbye to th...	4
7	10	1841721522	New Vegetari...	2
8	6366	1841721522	New Vegetari...	3
9	16654	1841721522	New Vegetari...	1
10	16673	1841721522	New Vegetari...	2
11	17183	1841721522	New Vegetari...	4
12	18384	1841721522	New Vegetari...	4
13	55217	1841721522	New Vegetari...	1
14	80908	1841721522	New Vegetari...	1
15	94985	1841721522	New Vegetari...	1
16	124868	1841721522	New Vegetari...	5
17	126644	1841721522	New Vegetari...	1

Fig. 1: Historical Dataset

```

UserKnnCosine
com.rapidminer.operator.RatingPrediction.UserKnnCosine@19741bc4
    
```

Fig. 2: User k-NN cosine

3.2. Construction of the model (k-NN as model builder)

The ratings for a predicted unrated books for a particular user is given by Apply Model(Rating Prediction) using k-NN operator for the Item, from RS→Item Rating Prediction. Figure 3 shows the work flow to measure performance of a recommendation model. The Read AML operator has been twice used for loading the input data and setting roles for item_ID and user_ID to identify the item and the user respectively. Then the Next Item k-NN operator, in the recommender model under Item Rating Prediction and also drag Apply Model under Rating Prediction

together with the Performance operator under Rating Prediction and run it. Figure-4. Shows the predicted data set by the model. Figure 5 shows the results coming in terms of terms of RMSE, MAE and Normalized Mean Absolute Error (NMAE) .It can be observed that Lesser values of RMSE and MAE gives better accuracy. MAE is usually less than or equal to RMSE. All errors have the same measure if both metrics are equal.

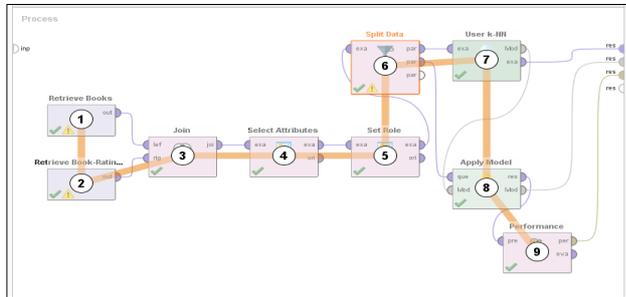


Fig 3: Prediction of Ratings to be put on multiple models using Multiply Operator

Row No.	User	ISBN	Book Title	Book-Rating	prediction
1	8	1552041778	Jane Doe	5	3.109
2	213915	1575683937	More Cummin...	5	3.383
3	3675	1841721522	New Vegetari...	1	3.743
4	11676	1841721522	New Vegetari...	2	2.922
5	13872	1841721522	New Vegetari...	2	2.831
6	105476	1841721522	New Vegetari...	4	3.778
7	134556	1841721522	New Vegetari...	4	3.056
8	136633	1841721522	New Vegetari...	4	2.828
9	152107	1841721522	New Vegetari...	5	2.995
10	156502	1841721522	New Vegetari...	3	3.022
11	158627	1841721522	New Vegetari...	5	3.140
12	187810	1841721522	New Vegetari...	2	3.802
13	229501	1841721522	New Vegetari...	3	2.972
14	241268	1841721522	New Vegetari...	3	2.099
15	248325	1841721522	New Vegetari...	3	2.388
16	22	3404921038	Wie Barney e...	2	3.166
17	141609	3442410665	Sturmzeit Ro...	5	2.800

Fig. 4: Prediction of the Ratings of the unrated books

PerformanceVector

PerformanceVector:
 RMSE: 1.556
 MAE: 1.312
 NMAE: 0.262

Fig. 5: The performance of the System for the recommending BookInternational

4. Conclusion

The paper has proposed a recommender system for prediction of user preferences of books using an integration of collaborative filtering and machine learning technique as an extension to the traditional recommender systems. This model can be used to test various recommender algorithms for performance evaluation. This paper has used only a single model (i.e. k-NN) but more heuristic algorithms belonging to machine learning may be implemented on the same model for more précised recommendations. The knowledge gained from this paper can be used to solve other recommendation problems and to improve quality of prediction and to take knowledgeable decisions in future.

5. References

- [1] Q. Wang, M. Sun, X. Yuan, "Collaborative Filtering Recommendation Algorithm based on Hybrid User Model", FSKD, 2010.
- [2] D J.Bobadilla, F. Ortega, A. Hernando, A. Gutierrez, Recommender systems survey, Knowledge-Based Systems,46(2013) 109-132.
- [3] Ricci F, Rokach L, Shapira B, Kantor PB (eds) (2011) Recommender Systems Handbook. Springer

- [4] P. Resnick, H.R. Varian, Recommender systems, Communications of the ACM, 40(1997) 56-58.
- [5] G. Zhuo, Jingyu Sun and Xueli Yu "A Framework for Multi- Type Recommendations", Eighth International Conference on Fuzzy Systems and Knowledge Discovery, 2007.
- [6] Deshpande M. and Karypis, G. Item-based top-n recommendation algorithms. ACM Trans. Inf. Syst. 22, 1 (2004), 143-177.
- [7] B. J., Heckerman D., and Kadie C., Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence, page 4352, 1998.
- [8] B.N. Miller, J.A. Konstan, J. Riedl, PocketLens: Toward a personal recommender system, ACM Transactions on Information Systems (TOIS), 22 (2004) 437-476. [8] B.N. Miller, J.A. Konstan, J. Riedl, PocketLens: Toward a personal recommender system, ACM Transactions on Information Systems (TOIS), 22 (2004) 437-476.
- [9] C. Huang and J. Yin "Effective Association Clusters Filtering to Cold-Start
- [10] Dr. B. Nayak, C A. V. Bath, "Association Rules Mining using Apriori algorithm for work-related beliefs of Generation X and Generation Y, 2015, 3324-3330.
- [11] Y. Jiang, J. Liu, M. Tang and X. (Frank) Liu "An Effective Web Service Recommendation Method based on Personalized Collaborative Filtering", 2011 IEEE International Conference on Web Services.
- [12] M. O Mohany, n. Hurley, n. Kushmerick and G. Silverstre, " Collaborative Recommendation: A Robustness Analysis", ACM Transacions on Internet Technology Vol. 4, November 2004, 344-377.
- [13] X N Lam, T Vu, T. Duc Le, A. Duc Duong, "Addressing Cold-Start in Recommendation Systems", proceedings of the 2nd international conference on Ubiquitous information management and communication,2008,208-211
- [14] I. Schein, A. Popescul, L. H. Ungar and D. M. Pennock, " Methods and Metrics for Cold-Start Recommendations", ACM1-58113—561-0/02/0008,2002.
- [15] Hao Ma,"An Experimental Study on Implicit Social Recommendation",2013, ACM 978-1-4503-2034-4/13/07.
- [16] B. M. Sarwar, G. Karypis, J. A. Konstan and J. T. Reidl, " Application of Dimensionality Reduction in Recommender System – A Case Study", Defence Technical Information Center, ADA439541.