ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Application of Factorization machines predicting user behaviour for book Recommender Systems.

Preeti Namdeo*, Rupinder Kaur Sandh, and M Venketesan.

School of Computer Science and Engineering(SCOPE), VIT University,Vellore,TamilNadu, India-632014

ABSTRACT

With the advent of internet, where we can get anything online, uses recommender systems to track user activities for enhancing business results. In online book store, where people buy books and leave their comments, accumulates large log data set consisting user profiles, information regarding category of books which are essential to know the interest of user. Various information regarding the kind, price, author of book are made in order to predict whether the user will turn up or is just single time dealer. Hence analysis of user action logs are very important to ensure proper management and upgradation of online business enterprises. In this paper we use factorization machines to predict user activities by their ratings in order to know their interests and improve their services and products. Based on the types of ratings given by the user, analysis are drawn to prepare an efficient prediction model of user behaviour on the basis of FM's algorithm.

Keywords: Factorization machines, behaviour prediction, feature vector, recommendation.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

As internet has dominated our day to day lives, there has been huge traffic of log data which contains valuable information. Online surfing is booming and large number of organization are migrating their business on web. Online book store turns out to be more promising as accommodates large amount of books with enormous varieties. But at the same time, this variety of options can often lead to bad experience faced by user. User may need to search a lot for a single book from pool of books. Online book store may serve as a boon by providing different categories of books but can take no time to become a curse as it may lead to poor user experience. Amazon which is known to be the world's "largest online store" has been prominent enough in the collection of books. It uses recommendation model to analyse the past user action logs and forms different recommendation for different commodities to ensure good experience by the users. The choice, the user make regarding the book helps us to know users future choices making more and more user convenient. Here is where recommendations systems play a significant role in uplifting the shopping experience from user point of view and improving the product service from business point of view.

In online book store, we can enhance the business results by considering the rating given by the user. The rating is categorized as excellent, very good, good, satisfactory and poor. Different rating of books reflects the kind of books that are most admired by the users. Using such analytics we can recommend the users to buy the most rated book which can be a great help to book lovers and an added advantage for business upliftment.

This paper demonstrates the use factorization machines on "Online Book Store" which can predict the behaviour of users to estimate their interest. Section 2 describes the literature survey. Section 3 represents the scenario analysis of application. Section four shows the model of FM algorithm. Section 5 represents how FM can be applied to predict user behaviour. Section 6 describes the conclusion.

LITERATURE REVIEW

Support vector machines have been prominent enough in data mining and machine learning[1]. But it does not seem to work well where data is sparse. Other models like standard tensor factorization or standard matrix factorization like PARAFAC[2]. The drawback with these models are that they are not applied to standard prediction data and moreover specialized models are built taking specific task into consideration which requires an extra effort for modelling and designing the learning algorithm. Taking factorized parameters other specialized models are [3],[4],[5]. On the other hand, non- linear support vector machines computes prediction depending on certain parts of training data.

Factorization machines are efficient enough to predict user behaviour even though having highly sparse data. Using factorized parameterization it can nest variable interaction. One of the prominent features of factorization machines is that its equation can be computed in linear time, depending on linear number of parameters. It has been applied to various recommendation systems giving fruitful output. For e.g. FM has been used to predict student performance[6]. Further, Yuqi Wang Wenqian Shang, Zhenzhong Li used FM for E-commerce[7]. J.P. Hu used FM in predicting user behaviour in the telecommunications field[8]. P.Z. Lu used FM in the automation field[9]. In this paper we use FM for online book store to predict user behaviour[10].

SCENARIO ANALYSIS OF APPLICATION

Online book store behaviour scenarios:

Basically, there are 5 categories in rating namely Excellent, Very Good, Good, Satisfactory, Poor. Different ratings have different impacts. Taking these behaviours into consideration, gives birth to progressive relationships. The following figure 1 shows the relationship.

ISSN: 0975-8585

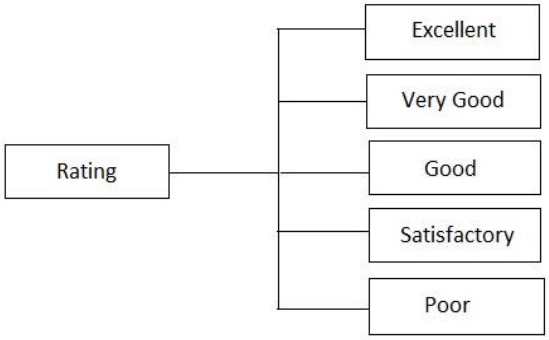

Figure 1: User behaviour type

Rating information is very crucial as it forms one of the strong basis for the decision making process of the business enterprises. If a user selects a book and rates excellent means that the score of this book is higher from the one which is rated very good. Similarly the score of the book with rating very good are higher than the books rating good, this is because the user may have interest inclination towards books of other categories, may not like the author, the content of the book did not meet the expectation of user and many more factors influence users behaviour. Giving Excellent rating means the user has been impressed by the book and may want more of the same category, same author so score of such behaviours should be more than score of other behaviours. Rating behaviour could be ambiguous for e.g. no matter what might be the experience of the user either good or bad will it will reflect users intentions so we have plan our services and products accordingly.

User behaviour model:

User behaviour model is primarily designed to reveal various user behaviours along with mathematical method to perform scientific analysis. As already stated above five kinds of ratings in online book store namely Excellent, very good, good, satisfactory and poor. We could find similar characteristics among various types of behaviours. Here we consider four prime elements of user namely user, book item, Rating and time. The chart below (Figure 2) presents four prime elements of different ratings.

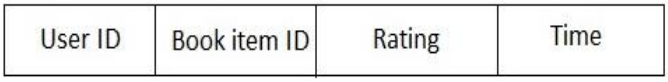

Figure 2: User behaviour model

User ID is for user identification taken as suject consideration of behaviour. Bookitem taken as object consideration behaviour. Rating contains five kinds ratings and time when the activity is being perfoemed. Following are the notations

U -> Users
I -> Bookitems
R -> Ratings
T -> Time

The formula is

$$D = <U,I,R,T> \quad (1)$$

D denotes training set. Since user behaviour data can be written as

$$d = <u,i,r,t>$$

using behaviour modelling we can perform analysis on the data obtained.

ISSN: 0975-8585

FACTORIZATION MACHINE MODEL

Factorization machine (FMs) model is introduced by Steffen Rendle in 2010 that combines the advantage of SMV (Support Vector Machine) with the factorization model [10]. FMs combines advantages of SVMs with the factorization model. FMs is able to process high dimentional data and is capable to evaluate reliable parameters under high sparse environment. FMs has a significant feature of linear complexity and a model that can perform factor decomposition.

A. Factorization machine model:

It is assumed that D is a dataset such that $D \in R^{nxp}$. $x_i \in R^p$ where, x_i represents the i-th row vector in D. Therefore, $D = \{x_1, x_2, x_3, x_4, ... x_n\}$ and n represents the total number of vectors in D.

$x = \{y_1, y_2, y_3, ...y_p\}$ where p represents the total number of features in x and y_i represents the i-th feature in x.

The model equation for 2-way factorization machine is:

$$\hat{Y}(x) = \omega_0 + \sum_{i=1}^{p} \omega_i x_i + \sum_{i=1}^{p} \sum_{j=i+1}^{p} \langle v_i, v_j \rangle x_i, x_j$$

Where, the equation parameters is defined as:

$\omega_0 \in R$, $\omega \in R^n$, $V \in R^{nxk}$

And $\langle v_i, v_j \rangle$ is the dot product of two vectors of size k:

$$\langle v_i, v_j \rangle = \sum_{f=1}^{k} v_{i,f}, v_{j,f}$$

A row v_i within V describes the i-th variable with k factors. $k \in N_0^+$ is a hyper parameter that defines the dimensionality of the factorization.

A 2-way FM (degree d = 2) captures all single and pairwise interactions between variables:

- ω_0 is the global bias.
- •ω_i models the strength of the i-th variable.
- $\langle v_i, v_j \rangle$ models the interaction between the ith and j-th variable.
- Advantages of FM:
- Under Sparcity, parameter of factorization machine can be well estimated.
- FMs have linear complexity.
- FMs can be learned in the primal whereas SVMs learned in the dual.
- FMs model equation is independent of training data whereas in SMVs prediction is depends on training data.
- FMs can work with any feature vector whereas other works with restricted input data.

BEHAVIOR PREDICTION OF USER

Feature vector:

First of all we need to create feature vector for the book recommendation according to the sample data chosen from the logs.

$U = <U_1, U_2, U_3.....>$

U represents the set of all users. To model the feature vector only three samples is taken.

ISSN: 0975-8585

I= <B_1, B_2, B_3, B_4.....>

I represents the set of all books. Four books are taken as a sample to model the feature vector.

S is the samples taken from user behaviour logs:

S= { (U_1, B_1, 2015-2, 5), (U_1, B_2, 2015-5, 4),
(U_1, B_3, 2016-3, 5), (U_2, B_1, 2015-3, 4),
(U_2, B_2, 2015-8, 5), (U_2, B_3, 2015-12, 3),
(U_2, B_4, 2016-4, 1), (U_3, B_1, 2015-9, 1),
(U_3, B_2, 2016-1, 4), (U_3, B_4, 2016-5, 3)}

Where, rating type represents by numbers like 1 represents poor, 2 represents satisfactory, 3 represents good, 4 represents very good and 5 represents excellent.

Based on the above sample data, the feature vector is obtained shown in Table 1.

Table 1: Feature Vector of some User Behaviour logs in book recommendation systems

X_1	1	0	0	...	1	0	0	0	...	0.3	0.2	0.3	0	...	2	0	0	0	0	...	5
X_2	1	0	0	...	0	1	0	0	...	0.3	0.2	0.3	0	...	5	1	0	0	0	...	4
X_3	1	0	0	...	0	0	1	0	...	0.3	0.2	0.3	0	...	15	0	1	0	0	...	5
X_4	0	1	0	...	1	0	0	0	...	0.2	0.3	0.2	0.06	...	3	0	0	0	0	...	4
X_5	0	1	0	...	0	1	0	0	...	0.2	0.3	0.2	0.06	...	8	1	0	0	0	...	5
X_6	0	1	0	...	0	0	1	0	...	0.2	0.3	0.2	0.06	...	12	0	1	0	0	...	3
X_7	0	1	0	...	0	0	0	1	...	0.2	0.3	0.2	0.06	...	16	0	0	1	0	...	1
X_8	0	0	1	...	1	0	0	0	...	0.06	0.2	0	0.2	...	9	0	0	0	0	...	1
X_9	0	0	1	...	0	1	0	0	...	0.06	0.2	0	0.2	...	13	1	0	0	0	...	4
X_{10}	0	0	1	...	0	0	0	1	...	0.06	0.2	0	0.2	...	18	0	1	0	0	...	3
X_{11}	U_1	U_2	U_3	...	B_1	B_2	B_3	B_4	...	0.06	0.2	0	0.2	...	time	B_1	B_2	B_3	B_4	...	Rating
	user				Books					Other book rated						Last book rated					

The table 1 shows the feature vector based on the sample data taken for the book prediction. It contains six parts. The first part indicates the active users. The second part indicates the active items i.e. books. The third part indicates the other books rated by the active user. This part is calculated based on the ratings by the user. There are five types of ratings user U_1 can give to the particular book. So, we can find out the value of B_1 by the following formula:

$$V_1 = \frac{rating_{type}(i_1)}{\sum_{k=1}^{i} rating_{type}(i_k)}$$

Where rating_type(i) = {1,2,3,4,5} and vales for U_1 is [0.3,0.2, 0.3,0].

The fourth part indicates the total time in months starting from the January, 2015. The fifth part indicates the last book rated by the active user. And the sixth part indicates the ratings of the active item i.e. book.

CONCLUSION

Factorization machines proved to be reliable in predicting user behaviour. Taking into consideration user name, bookitem, rating and time alongwith the feature vector provides a deep insight of user interaction and characteristics. All other models had certain drawbacks and had not been efficient enough to answer product management questions. Hence Factorization machines proved to be capable enough the extract the

real crux from the interactions between the user and the data. Further we can add more features in the feature vector table to enhance the analysis and perform higher level prediction.

REFERENCES

[1] H.L. Wang, "User behavior classification and feature analysis in search advertising," Application Research of Computers, vol. 30, No.5, pp. 1413-1418, May 2013.

[2] R. A. Harshman, "Foundations of the parafac procedure: models and conditions for 'exploratory' multimodal factor analysis." UCLA Working Papers in Phonetics, pp. 1–84, 1970.

[3] Y. Koren, "Factorization meets the neighborhood: a multifaceted collabo-rative filtering model," in KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: ACM, 2008, pp. 426–434.

[4] S. Rendle and L. Schmidt-Thieme, "Pairwise interaction tensor factoriza-tion for personalized tag recommendation," in WSDM '10: Proceedings of the third ACM international conference on Web search and data mining. New York, NY, USA: ACM, 2010, pp. 81–90.

[5] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, "Factorizing per-sonalized markov chains for next-basket recommendation," in WWW '10: Proceedings of the 19th international conference on World wide web. New York, NY, USA: ACM, 2010, pp. 811–820.

[6] Nguyen Thai-Nghe ∗,† , Lucas Drumond † , Tom´a˘s Horv´ath † , and Lars Schmidt-Thieme †, " Using Factorization Machines for student modelling"

[7] Yuqi Wang Wenqian Shang, Zhenzhong Li," The Application of Factorization Machines in User Behavior Prediction".

[8] J.P. Hu, "prediction algorithm of user behavior based on community mining in telecom networking," Journal of Information Engineering University, vol. 11, No. 5, pp. 540-543, October 2010.

[9] P.Z. Lu, "Research on user behavior prediction in home automation,"China New Technologies and Products, No. 02, pp. 19-20, 2016.

[10] S. Rendle, "Factorization Machines," 2010 IEEE International Conference of Data Mining, pp. 995-1000, 2010.