Efficiency Analysis of Collaborative Based Recommendation System

Waleed Khalid¹, Xing Xing¹*, Aikodon Julius¹, Yong Niu¹, Osama Tahir² and Imran Ihsan³

¹Department of Information Science and Technology, Bohai University, China
²IBM, Islamabad, Pakistan
³Department of Computer Science, Air University, Islamabad, Pakistan
Email: xingxing@bhu.edu.cn

Abstract. Over the recent times, there has been great enhancement towards online shopping and platforms that provide commerce. Hence, great research and work has been done and is being done in field of recommendation systems. With this great development, there has been an exponential increase in online inventory due to the great number of users excessing these online platforms for buying and selling purposes and companies are often looking for advanced recommendation systems to provide their customers with the best online experience in respect towards each individual customer. It is believed that recent advancements in Deep Learning may provide an optimal solution for better recommendation systems, but it requires validation. The main aim of this paper is to follow through different research and investigate whether modern Deep Learning algorithms live up to the expectations and demands. Different reviews have been given in support with experiments. This literature review provides an analysis of different practices, state of the industrial methodologies and current research.

1. Introduction

In today’s era of the Internet, the amount of information is growing exponentially [1]. This sharp flow of information has also resulted in the development of services to provide information to users based on their needs making it comfortable [2]. Recommendation systems (RS) that emerged in the mid-1990s [3] are the services that are typically defined to provide suggestions to users and other stakeholders [4] keeping in mind the interest for a specific user [5]. These recommendations perform various decision-making processes as what to buy, to what music to listen to, or which movies or TV series to see [1].

Conventional Recommendation Systems such as content-based [6], [7], and session-based [8] methods are normally focused on explicit user-item preferences and anonymous behaviour sessions respectively. However, these methods are primarily designed to find similar items already rated by users [9] and possibly miss out on interesting items that lie outside the range of documents already rated by users [10]. A different approach to the recommendation system is Collaborative Filtering. Collaborative Filtering filters and evaluates items based on the opinions of other people. The technique collects and establishes profiles and determines the relationships among the data according to similarity models. The profile can have various categories of the data such as user preferences, user behaviour patterns, or item properties [10]. Therefore, unlike content-based systems, collaborative filtering systems do not require to obtain and analyse content [14]. Though recommender systems are the tools to interact with large and complex information spaces and to provide a personalized view of such spaces, prioritizing items likely to be of interest to the user [15].
Today, humans undoubtedly rely on the recommendations from various sources for one purpose to another, however, the choices available to them are increasing manifold day by day [16]. Countless advancements in machine learning and artificial intelligence have significantly created an impact on recommendation systems. Similarly, promising results have been achieved by the usage of Deep Learning techniques on Recommender Systems [17] as well. In collaborative filtering recommender systems, products are the features, and users rate these products. However, there are usually a huge number of products involved, and selecting every product can make a recommender system inefficient [18]. To increase the efficiency of the recommendation, various algorithms use self-constructing clusters to reduce the dimensionality related to the number of products. In this paper, we have used four state-of-the-art recommendation systems; Matrix Factorization, Non-Negative Matrix Factorization, Neural Rating Regression, and AutoRec and analysed their respective efficiency when applied on MovieLens datasets [19] for movie recommendations.

2. Related Work
Collaborative filtering is a widely applied technique in recommender systems, in which patterns across similar users and items are leveraged to predict user preferences [20]. Collaborative filtering techniques collect and indicate profiles while determining the relationships amidst data according to similarity models. The possible categories of data in the profiles include user preferences, user behaviour patterns, or item properties [21]. The outcome that is determined through such approaches are a verdict of significance that shows an individual’s interest in a particular item. If the item is correctly recommended to the user, it provides valuable and important insights to the company of the user’s liking and disliking [22]. While Deep Learning techniques begin to be used in a recommendation system for several situations ranging from dialog systems to temporal user-item relationships through heterogeneous classification applied to recommendation environment, few efforts have been made to use collaborative filtration neural networks [23]. Matrix factorization methods have recently received greater exposure, mainly as an unsupervised learning method or latent variable decomposition and dimensionality reduction successfully applied in spectral data analysis and text mining most of the matrix factorization methods are based on the latent factor model [24]. While Non-negative Matrix Factorization (NMF), a comparatively new dimension reduction paradigm, has been in the ascendent. It incorporates the non-negativity constraint and thus obtains the parts-based representation as well as enhancing the interpretability of the issue correspondingly [25]. AutoRec partially observes vectors as an input and projects it into a low-dimensional latent space and reconstruct it in the output space to predict the missing ratings. [26]. Neural rating Regression is a multilayer perceptron network that is used to project latent user variables and latent rating variables. With a multi-task learning strategy in an end-to-end teaching paradigm, all the neural parameters in the gated recurrent neural networks and the multilayer perceptron network as well as the latent variables for users and things are learned [27].

3. Methodology
Several approaches have been taken to provide reasonable outline and see whether the latest practices cross the mark line in terms of performance. Focus would be put upon collaborative filtering with both Deep Learning and Non-Deep Learning approaches as it’s widely used in scientific society. Different algorithms have been discussed with regards to their performance, their method of usage and their shortcomings. To achieve this, following methodology has been adopted and is outlined in Figure 1.
3.1. Dataset
In education, research, and business, the MovieLens information sets are commonly used. Every year they are downloaded hundreds of thousands of times, reflecting their use in popular media programming books, internet and traditional classes, and software. There is different size of dataset available online for MovieLens available, but we have used MovieLens dataset with size of 1 million.

3.2. Preprocessing
As several features are used during the processing in later stages. Hence, first we analysed the dataset for any missing values of those features. This step is highly important because it can affect our results in later stages. After that, we normalized independent features of the dataset and brought all the variables to the same range.

3.3. Collaborative Filtering
We chose model based approach for Collaborative Filtering because it supports machine learning technique and envisages the user’s ending rating on unrated items. For machine learning techniques, we chose Matrix Factorization, Non-Negative Matrix Factorization, Neural Rating Regression and AutoRec. We selected Neural Rating Regression as it is a deep learning technique which can simultaneously predict precise ratings and generate abstractive tips with good linguistic quality simulating user experience and feelings [27]. We used AutoRec as it is a novel autoencoder framework for collaborative filtering and provides compact and efficiently trainable model [28]. We used Matrix factorization as it could incorporate implicit feedback, information that are not directly given but can be derived by analyzing user behavior [29]. Lastly, we used Non-Negative Matrix Factorization approach as it signifies speed, ease of interpretation and versatility [30].

4. Experiments and Results
To fulfill the main objective for this literature is to overview whether latest methods live upon to their performance. Several approaches have been taken to provide reasonable outline and see whether the latest practices cross the mark line in terms of performance. A high focus would be put upon collaborative filtering with both Deep Learning and Non-Deep Learning approaches as it’s widely used in scientific society. Different algorithms have been discussed with regards to their performance, their method of usage and their shortcomings. The Table 1 shows the hyperparameter used for each model on MovieLens dataset (1 Million data points) are Matrix Factorization, Non-Negative Matrix Factorization, Neural Rating Regression and AutoRec. Each model is evaluated upon Root Mean Square Error (RMSE) and its results are listed on Table 1.
Table 1. Hyper Parameters for Models.

| Parameters       | Matrix Factorization | Non-Negative Matrix Factorization | Neural Rating Regression | AutoRec |
|------------------|----------------------|-----------------------------------|--------------------------|---------|
| Learning Rate    | 0.001                | 0.001                             | 0.001                    | 0.001   |
| Regularization Rate                      | 0.01                | 0.01                              | 0.01                     | 0.00001 |
| Epochs                      | 20                  | 20                                | 20                       | 20      |
| Batch Size       | 128                  | 128                               | 128                      | 28      |
| RMSE             | 0.812                | 0.871                             | 0.826                    | 0.821   |

5. Discussion

Learning rate is the extent to which the model changes during each search step. To achieve good performance, it is the most important parameter of a model as it shows that how quickly or slowly it can learn the problem. A greater learning allows the model to learn at the cost of reaching a sub-optimal final set of weights much faster. A lower learning speed may enable the model to learn a weight set that is more ideal or even globally optimal, but it may take much longer to train. The other important inspections are regularization rate and the batch size. Regularization is a significant element in avoiding over fitting. It can be used to decrease the model volume while withholding accuracy. Whereas, batch size is relatively proportional for convergence and performance.

From the Table 1 we can observe that Matrix Factorization provided the lowest score with 0.812 while Neural Rating Regression came second. From the Figure 2, we can easily conclude from the pattern that Matrix Factorization which is a non-Deep Learning approach tends to provide better results.

![Figure 2. Graphical Overview of Different Models](image)

6. Conclusion

As mentioned before that deep learning has brought major impact on research group interest regarding recommendation systems but with reference to recent research, the classical algorithms tend to provide better results. The state of the art algorithms is more complex, and their complexity tends to provide disadvantage while delivering optimal solution. Hence, a simple approach like Matrix Factorization which is believed to be more mature in nature and not only provide optimal results but also provide a much better traditional baseline for future research to take place. The future research directs towards...
an algorithm with much accuracy and simplicity and we believe that classical algorithms fulfil this requirement.

7. Acknowledgements
This paper is partially supported by the National Natural Science Foundation of China under Grant No.61972053, by the Scientific Research Foundation of Liaoning Education Department under Grant No. LQ2019016, No. LJ2019015, and by the Natural Science Foundation of Liaoning Province, China under Grant No.2019-ZD-0505.

8. References
[1] Lytvin V, Vysotska V, Shatskykh V, Kohut I, Petruchenko O, Dzyubyk L, Bobrivetc V, Panasyuk V, Sachenko S and Komar M 2019 Design of a recommendation system based on Collaborative Filtering and machine learning considering personal needs of the user Eastern-European J. Enterp. Technol. 4 6–28
[2] Jones M T 2013 Recommender systems , Part 1: Introduction to approaches and algorithms developerWorks 1–8
[3] Bobadilla J, Ortega F, Hernando A and Bernal J 2012 A collaborative filtering approach to mitigate the new user cold start problem Knowledge-Based Syst. 26 225–38
[4] Melville P, Mooney R J and Nagarajan R 2002 Content-Boosted Collaborative Filtering for Improved Recommendations
[5] Nambiar R, Bhardwaj R, Sethi A and Vargheese R 2013 A look at challenges and opportunities of Big Data analytics in healthcare Proc. - 2013 IEEE Int. Conf. Big Data, Big Data 2013 17–22
[6] Lops P, de Gemmis M and Semeraro G 2011 Content-based Recommender Systems: State of the Art and Trends Recommender Systems Handbook (Springer US) pp 73–105
[7] Weston J, Chopra S and Bordes A 2015 Memory networks 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. 1–15
[8] Wang M, Chen Z, Ren P, Ma J, Mei L and De Rijke M 2019 A collaborative session-based recommendation approach with parallel memory modules SIGIR 2019 - Proc. 42nd Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. 345–54
[9] Pazzani M and Billsus D 2013 Content-based recommendation Recomm. Syst in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2007,pp 325–341
[10] Khan B M, Mansha A, Khan F H and Bashir S 2018 Collaborative filtering based online recommendation systems: A survey 2017 Int. Conf. Inf. Commun. Technol. ICICT 2017 2017-Decem 125–30
[11] Linden G, Smith B and York J 2003 Amazon.com recommendations: Item-to-item collaborative filtering IEEE Internet Comput. 7 76–80
[12] Sarwar B, Karypis G, Konstan J and Riedl J 2001 Item-based collaborative filtering recommendation algorithms Proc. 10th Int. Conf. World Wide Web, WWW 2001 285–95
[13] Breese J S, Heckerman D and Kadie C 2013 Empirical Analysis of Predictive Algorithms for Collaborative Filtering 43–52
[14] Burke R, Felfernig A and Göker M 2011Recommender Systems: An Overview AI Magazine. pp. 13–18
[15] Sharma R and Singh R 2016 Evolution of recommender systems from ancient times to modern era: A survey Indian J. Sci. Technol. 9
[16] De Souza Pereira Moreira G, Ferreira F and Da Cunha A M 2018 News Session-Based Recommendations using Deep Neural Networks ACM Int. Conf. Proceeding Ser. 15–23
[17] Liao C L and Lee S J 2016 A clustering based approach to improving the efficiency of collaborative filtering recommendation Electron. Commer. Res. Appl. 18 1–9
[18] Harper F M and Konstan J A 2015 The movielens datasets: History and context ACM Trans. Interact. Intell. Syst. 5 1–19
[19] Lobel S, Li C, Gao J and Carin L 2019 Towards Amortized Ranking-Critical Training for Collaborative Filtering 1–19
[20] Mohd A, Hameed M, Jadaan O and Sirandas R 2012 A survey on collaborative filtering based recommendation system: International Journal on Computer Science and Engineering

[21] Nguyen P T, Tomeo P, Di Noia T and Di Sciascio E 2015 Content-based recommendations via DBpedia and freebase: A case study in the music domain Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9366 605–21

[22] Zhang S, Yao L, Sun A and Tay Y 2019 Deep learning based recommender system: A survey and new perspectives ACM Comput. Surv. 52 1–35

[23] Bokde D, Girase S and Mukhopadhyay D 2015 Matrix Factorization model in Collaborative Filtering algorithms: A survey Procedia Computer Science vol 49 (Elsevier B.V.) pp 136–46

[24] Ding C, Li T and Jordan M I 2008 Nonnegative matrix factorization for combinatorial optimization: Spectral clustering, graph matching, and clique finding Proc. - IEEE Int. Conf. Data Mining, ICDM 183–92

[25] Falahuddin Quadri S, Li X, Zheng D, Umar Aftab M and Huang Y 2019 Multi-Layer Graph Generative Model Using AutoEncoder for Recommendation Systems J. Big Data 1 1–7

[26] Li P, Wang Z, Ren Z, Bing L and Lam W 2017 Neural rating regression with abstractive tips generation for recommendation SIGIR 2017 - Proc. 40th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. 345–54

[27] Sedhain S, Menony A K, Sannery S and Xie L 2015 AutoRec: Autoencoders meet collaborative filtering WWW 2015 Companion - Proc. 24th Int. Conf. World Wide Web 111–2

[28] Chen S and Peng Y 2018 Matrix factorization for recommendation with explicit and implicit feedback Knowledge-Based Syst. 158 109–17

[29] Seung D and Lee L 2001 Algorithms for non-negative matrix factorization Adv. Neural Inf. Process. Syst. 13 556–62