CHAPTER-33

MOVIE RECOMMENDATION SYSTEM MODELING USING HYBRID CLASSIFICATION APPROACH

[1]ANAND PRAKASH SRIVASTAVA, [2]Dr. SANJAY KUMAR SINHA
[1]Ph.D. Scholar, Computer Science & Engineering, Sanskriti University, Mathura
[2]Professor, Computer Science & Engineering, Sanskriti University, Mathura

Chapter ID: NSP/ICAAR-2023/A-33

ABSTRACT

A type of solution to the information overload issue experienced by users of websites that allow the rating of specific items is the recommender system. One of the most effective, practical, and well-known applications for people to view movies quickly is called movie recommendation system (MRS). There have been numerous attempts by researchers to overcome these problems, such as using MRS to watch a movie or buy a book, however the majority of these studies have been unsuccessful in addressing the cold start problem, data sparsity, and malicious attacks. In order to solve these issues, this work proposes a hybrid machine learning algorithm to indorse a suitable movie. A non-cold user went through many models with a trust filter, and a cold user produced the best possible score using their own personal preferences.

Keywords: *Hybrid Classifier, machine learning, Movie Recommendation Systems, Classification accuracy.*

INTRODUCTION

The proliferation of digital media—including websites, online stores, blogs, videos, photos, jokes, and more—has resulted in an explosion of data. It is a huge challenge to deliver accurate information from the available data. Instead of responding to specific user queries, recommender systems instead retrieve foundational pieces of data based on users' past actions. Content-enabled methods, collaborative methods, knowledge-in-built methods, utility filter techniques, and demographic recommendation filter techniques are all examples of the various flavors of recommendation strategies.

Which of the hybrid strategies (collaborative, content-based, and combination) is most popular for building a movie recommendation system (MRS)? To find products that will appeal to active users, the content enabled recommendation (Kuanr & Mohanty, 2020) method analyzes items' unique characteristics. The hypothesis that the customers who had make alike choices in the past are likely to make similar choices in the future (Guo et al., 2016), the collaborative recommendation strategy discovers the similarity for the customers by making a close check on their previous assessments. A demographic and knowledge-based recommendation system is concerned with filtering data based on geographical location and item kind, correspondingly (Kuanr et al., 2018). The Hybrid system is created by combining two or more recommendation systems. The collaborative system currently dominates the recommender system market. But it has issues like cold-start (no recommendation list for new users or new items), data sparsity, malicious attacks, and so on. Due to the limitations and difficulties of collaborative recommendation, trust filter-based classification technique between users are taken into consideration in place of similarity based user in order to provide more accurate recommendations. Without data ratings, users' personal connections to one another can serve as an alternative trust factor (Ayub et al., 2020; Massa & Bhattacharjee, 2004). There are two different kinds of trust among users. Implicit trusts (Bedi & Sharma, 2012; Gupta & Nagpal, 2015) can be calculated from the data at hand to specify the trust between users. When a user gives their trust explicitly, it means that they mean it. As there is often little or no available data, establishing explicit trust can be challenging. Thus, various metrics exist that can be used to determine the level of implicit trust. A high trust value might have an adverse effect on the business if it is shared by more than two customers. Fixing this implicit trust issue requires developing a metric for accurately estimating the level of trust amongst users with comparable profiles (Gohari et al., 2019).

Members can rate the degree to which they trust each other. In the context of recommendation engines, this evaluation has to do with how credible the user's ratings are. The trust metric serves as a node in a trust network, allowing for an easy representation of the trust calculation. Collaborative filters use trust propagation over a trust network to find reliable users. The Trust-aware Recommender System selects a group of trusted users and makes recommendations based on the items liked by those people to the current user. The ensemble model incorporates four distinct machine learning techniques, including naïve bayes (NB), support vector machine (SVM), auto encoder (AE), and auto encoder with trust (AET). The typical collaborative filtering "cold user" problem has been solved with the use of a popularity-based recommendation algorithm.

LITERATURE REVIEW

Several approaches applied to collaborative filter can boost the quality of the recommendation system. You can look at it one of two ways. To begin, there are two main types of recommender systems: the more traditional ones and the trust-filter based ones.

Recommender Systems based on Traditional Classification Techniques

Memory-based and model-based algorithms are the two types of traditional algorithms. Memory-based algorithms provide suggestions based on a user's entire profile data set, while model-based algorithms use the user's profile history to train a model and make recommendations. Using collaborative filtering methods, researchers describe how to find comparable users in a data network (Goldberg et al., 1992). To estimate the model and compare with the statistical model of collaborative filtering, scientists depict variations of K-means clustering and Gibbs Sampling on model parameters. Another researcher use technique to automate the categorization of news articles (Konstan et al., 1997). Using a recommender system based on the Jester system, (Goldberg et al., 1992) suggest jokes to the user and books by Amazon. This system tailors its output to the individual user, surprising them with high-quality data based on their preferences. However, collaborative systems have challenges like as cold start, data sparsity, and the gray sheep problem. Due to data scarcity, the system's accuracy suffers. The data sparsity problem in feature selection for biomedical data is avoided by (Nagpal et al., 2017) by employing the Gravitational search algorithm. According to (Gupta & Nagpal, 2015), the system stops making recommendations in favor of the Gray sheep dilemma if the tastes of a single user don't align with those of the majority of its users.

Trust-filter based recommender systems

Due to the issue of sparsity and malicious attacks, there has been a lot of focus on including trust in the proposed system. Trust is an understanding between the people who have decided to associate with the user. The evaluation of trustworthy data is an attempt to fix collaborative system flaws.

Similarity between users is mined for implicit trust, whereas the user alone determines the value of explicit trust. On the basis of implicit trust at the profile and item levels, they made quality suggestions using the Resnick formula. Tidal Trust method measures the reliability of a film suggestion (Golbeck, 2006). This method uses nearest trust path in an effort to discover all users rated by the target user. The trusted k-nearest recommenders' algorithm to improve user trust in one another's recommendations and mitigate some of the problems with conventional collocation-based methods. To make up for the scarcity of ratings, (Massa & Bhattacharjee, 2004) suggested popular goods based on the explicit trust data from the users. To address the data sparsity and cold start problem in explicit recommendation, a new method based on the propagation of trust measures was suggested. Different recommendation policies were developed by (Al-Shamri & Bharadwaj, 2008), who investigated both trust and distrust. To compute more effective recommendations, researchers used Fuzzy logic and weighted trust metrics. Since the trust propagation mechanism between users is ignored by the aforementioned methods, the issues of data sparsity and the Gray sheep dilemma are not effectively addressed. We analyzed both trust and resemblance amongst users while keeping in mind the issues that have arisen in collaborative methods.

METHODOLOGY

Based on the trusted user's implicit trust value and several machine learning techniques, the suggested model will suggest a movie to the user. Collaborative filtering uses users' and products' shared characteristics to make personalized suggestions. It constantly discovers connections between users, and in turn makes suggestions. In order to estimate ratings, system can either rely on a:

(1) Memory-based approach, which uses pairs of user and object ratings, or

(2) Model-based approach, which employs some kind of machine learning technique.

Memory-based approaches struggle with data sparsity and scalability, while model-based algorithms have solved these problems by creating a model that uses user ratings to propose items (Kuanr et al., 2018). Matrix factorization is the most often used model-based technique for discovering the embeddings or features that compose a user's interest. Naïve bayes, support vector machine and a auto encoder model are only few of the machine learning approaches used to make this prediction(see Figure 1).

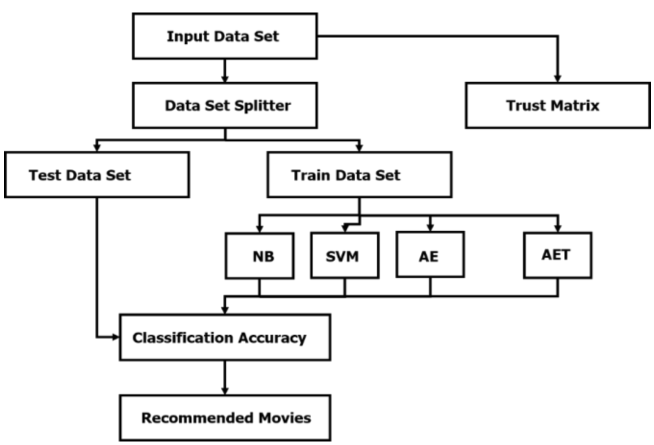

Figure 1. Methodology

Factorisation of Matrix

By multiplying two types of items together, matrix factorization can be used to produce latent features. Matrix factorization (Bedi & Sharma, 2012) is used to determine the entity-item connection. The recommendation model has high computational complexity with low accuracy since user ratings are sparse (95% missing data). It is a procedure that involves dividing the true users' rating matrix into two feature (user and item) matrices, such that the true matrix may be reconstructed by multiplying the new two matrices. The predicted matrix's output is consistent with the true values, and the projected ratings (0 in the actual matrix) are used in place of the actual ratings (see Figure 2). It is simple to construct, has the potential to be interpretable, and reduces the amount of time spent on queries. Unfortunately, the linear model has the limitation of not being able to capture complicated relations in data.

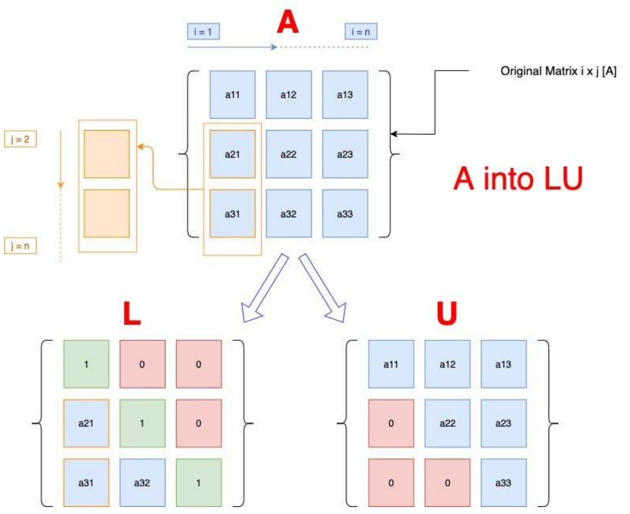

Figure 2. Matrix Factorization

Support Vector Machine

Support vector machines (SVMs) are a classic machine learning tool for dealing with general categorization issues in a large data setting. Although several different SVMs exist, a linear SVM was chosen because it is the easiest to build, has the best track record for addressing general classification issues, and is readily available in the appropriate python package. To classify the various data points, a linear SVM employs a straightforward statistical equation. Considering my issue spans multiple domains, I recognize that a linear method may not provide optimal outcomes, but one was chosen due to its seeming simplicity (Pavitha et al., 2022).

Naïve Bayes

The probability calculations at the heart of a Naive Bayes indicate that it takes a more statistical approach. It can accurately predict the results of a test instance, making it useful for supervised learning scenarios. The classifier is naive since it just uses two very simple assumptions. The first premise is that the various forecasts are fully independent of one another under all conditions. Second, it is presumed that no relevant information is missing and that all possible influences have been accounted for. The second assumption in particular makes the classifier naive since it interprets the data in a closed-world setting (Pavitha et al., 2022).

Auto Encoder

Auto encoder is a collection of neural building blocks that may be combined to form a trained network that is consistent from beginning to conclusion (Choudhury et al., 2021). Inductive bias that is tailored to the input data and end-to-end differentiability are the most intriguing qualities. Reasonable performance gains can be achieved in response to data complexity and extensive training samples. This deep learning approach attempted to propose items to users based on their predicted ratings from active users. The feature matrix between the user and the item is fed into the embedding layer. Two feature matrices are multiplied together using the dot product operation, and then fed into a dense hidden layer with as many nodes as features. Mean square error is used to evaluate the model's accuracy**.**

Auto Encoder Trust

With collaborative method, the consumer does the research and the producer provides the results. Trust properties allow for the complexities of collaborative systems to be overcome. While evaluating a client's trustworthiness, experts and past results are taken into account. We can think of a trust system as a set of interconnected peer groups. Prescribed representations of trust view the exchange of information between two domestic entities as a by-product of establishing a trust relationship with a third party. There is a significant difference between local and global trust. A separate score, called Global Trust, can be calculated independent of peer influence(Choudhury et al., 2021).. The local trust metric provides individual scores, which indicates which peers are trustworthy.

RESULT AND DISCUSSION

In this section all the outputs and results of the abovementioned models, processes and comparisons were given.

Naïve Bayes Algorithm

The dataset used in this model is a user-item rating matrix. There are a total of 300 users and 50,000 ratings over 6000 different things. The model's projected rating value was substituted for every non-zero value.

SVM Algorithm

The mean square error (MSE) approach is utilized to compute both training error and validation error. When a backpropagation model is overfit to its training data, it produces subpar results on test data (Table 1).

Table 1. MSE of SVM algorithm

No. of Epoch	Training MSE	Testing MSE
1	1.35	1.51
2	0.98	1.31
3	0.81	1.20
4	0.60	1.15
5	0.45	1.11
6	0.38	1.08
7	0.30	1.09
8	0.25	1.08
9	0.21	1.09
10	0.21	1.11

11	0.21	1.11
12	0.21	1.12
13	0.21	1.14
14	0.21	1.15
15	0.21	1.15

Auto Encoder Algorithm

For movie suggestions, we take into account the preferences of users most like our chosen one using the auto encoder algorithm. In a feature map, you can see this clearly (Table 2).

Table 2. MSE of AE algorithm

No. of Epoch	Training MSE	Testing MSE
1	3.00	1.15
2	1.85	1.05
3	1.44	0.89
4	1.11	0.81
5	0.85	0.78
6	0.75	0.75
7	0.75	0.75
8	0.75	0.75
9	0.75	0.75
10	0.75	0.75
11	0.75	0.75
12	0.75	0.75
13	0.75	0.75
14	0.75	0.75
15	0.75	0.75

Auto Encoder Trust Algorithm

With the consumer-rating matrix and the user-feature matrix, we may derive the producer-consumer trust matrix using Donovan's trust approach. As a buyer, you should not put your faith in the rating produced by the diagonal cells in your computer's trust computation. Hence, the value "Nan" has been entered into these diagonal cells. NB, SVM, and AE models have all been found to result in reducing Mean Square Error levels. The MSE value has increased when trust was integrated with AE (see Figure 3).

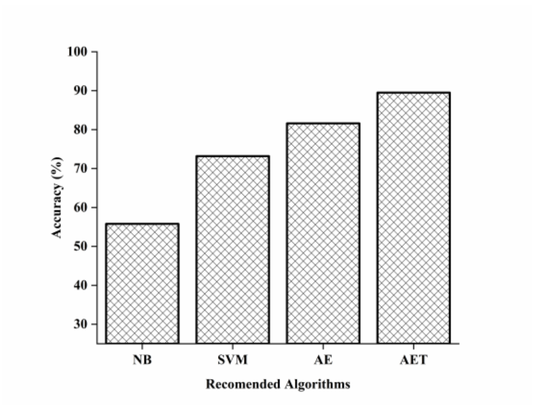

Figure 3. Accuracy of recommended algorithms

CONCLUSION AND FUTURE WORK

Several machine learning methods are being incorporated into the model's development in order to provide more accurate MRS, and then trust-based filtering is being utilized to round things out. Percentage of correct predictions using back propagation method (55.8%), super value decomposition method (73.2%), and deap neural network (81.6%). When the validation accuracy of deap neural network with trust (89.5%) was compared to that of Backpropagation method, the latter was found to be superior. High values of the optimum score are preferred since Cold users have always represented a risk for colloborative filtering. Further improvements in suggestion quality can be achieved by implementing a variety of bio-inspired optimization strategies into collaborative method, Content-based, and Demographic-based filtering and comparing their results.

REFERENCES

1. *Al-Shamri, M. Y. H., & Bharadwaj, K. K. (2008). Fuzzy-genetic approach to recommender systems based on a novel hybrid user model. Expert Systems with Applications, 35(3), 1386–1399. https://doi.org/10.1016/j.eswa.2007.08.016*
2. *Ayub, M., Ghazanfar, M. A., Mehmood, Z., Alyoubi, K. H., & Alfakeeh, A. S. (2020). Unifying user similarity and social trust to generate powerful recommendations for smart cities using collaborating filtering-based recommender systems. Soft Computing, 24(15), 11071–11094. https://doi.org/10.1007/s00500-019-04588-x*
3. *Bedi, P., & Sharma, R. (2012). Trust based recommender system using ant colony for trust computation. Expert Systems with Applications, 39(1), 1183–1190. https://doi.org/10.1016/j.eswa.2011.07.124*
4. *Choudhury, S. S., Mohanty, S. N., & Jagadev, A. K. (2021). Multimodal trust based recommender system with machine learning approaches for movie recommendation. International Journal of Information Technology (Singapore), 13(2), 475–482. https://doi.org/10.1007/s41870-020-00553-2*
5. *Gohari, F. S., Aliee, F. S., & Haghighi, H. (2019). A Dynamic Local–Global Trust-aware Recommendation approach. Electronic Commerce Research and Applications, 34(March 2018), 100838. https://doi.org/10.1016/j.elerap.2019.100838*
6. *Golbeck, J. (2006). Generating predictive movie recommendations from trust in social networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 3986 LNCS, 93–104. https://doi.org/10.1007/11755593_8*
7. *Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to Weave an Information tapestry. Communications of the ACM, 35(12), 61–70. https://doi.org/10.1145/138859.138867*
8. *Guo, G., Zhang, J., & Yorke-Smith, N. (2016). A Novel Recommendation Model Regularized with User Trust and Item Ratings. IEEE Transactions on Knowledge and Data Engineering, 28(7), 1607–1620. https://doi.org/10.1109/TKDE.2016.2528249*
9. *Gupta, S., & Nagpal, S. (2015). An empirical analysis of implicit trust metrics in recommender systems. 2015 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2015, 636–639. https://doi.org/10.1109/ICACCI.2015.7275681*
10. *Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). Applying Collaborative Filtering to Usenet News. Communications of the ACM, 40(3), 77–87. https://doi.org/10.1145/245108.245126*
11. *Kuanr, M., Kesari Rath, B., & Nandan Mohanty, S. (2018). Crop Recommender System for the Farmers using Mamdani Fuzzy Inference Model. International Journal of Engineering & Technology, 7(4.15), 277. https://doi.org/10.14419/ijet.v7i4.15.23006*
12. *Kuanr, M., & Mohanty, S. N. (2020). Location-based personalised recommendation systems for the tourists in India. International Journal of Business Intelligence and Data Mining, 17(3), 377–392. https://doi.org/10.1504/IJBIDM.2020.109294*
13. *Massa, P., & Bhattacharjee, B. (2004). Using trust in recommender systems: an experimental analysis. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2995, 221–235. https://doi.org/10.1007/978-3-540-24747-0_17*
14. *Nagpal, S., Arora, S., Dey, S., & Shreya, S. (2017). Feature Selection using Gravitational Search Algorithm for Biomedical Data. Procedia Computer Science, 115, 258–265. https://doi.org/10.1016/j.procs.2017.09.133*
15. *Pavitha, N., Pungliya, V., Raut, A., Bhonsle, R., Purohit, A., Patel, A., & Shashidhar, R. (2022). Movie recommendation and sentiment analysis using machine learning. Global Transitions Proceedings, 3(1), 279–284. https://doi.org/10.1016/j.gltp.2022.03.012*