Genetic Twin Support Vector Based Movie Recommendation System

Nisha Bhalse

IET, Devi Ahilya University,
IPS Academy IES,
Department of Computer Science & Engineering
Indore, India
nishabhalse@ipsacademy.org

Ramesh Thakur

Department of Computer Engineering IIPS, Devi Ahilya University, Indore, India . rthakur.iips7@gmail.com

Archana Thakur

School of Computer Science & IT Devi Ahilya University, Indore, India . archana22@gmail.com

Abstract— Movie recommendation systems have become increasingly popular in recent years, as they can help users discover new movies that they are likely to enjoy. However, existing recommendation systems often suffer from limitations such as sparsity, cold start and overfitting. In this paper, we propose a new recommendation system called GATWSVM, which combines genetic algorithm with twin support vector machine to overcome these limitations. GATWSVM works by first constructing a user-movie rating matrix from a set of user-generated ratings. Then, it uses a genetic algorithm to evolve a population of classifiers, each of which is a twin support vector machine. The classifiers are trained to predict the ratings that users would give to movies that they have not yet rated. Finally, the system recommends movies to users by selecting the movies that have the highest predicted ratings. We evaluated GATWSVM on the MovieLens dataset and compared its performance to several state-of-the-art recommendation systems. Our results show that GATWSVM outperforms all of the other systems in terms of recommendation accuracy and precision.

Keywords- Recommendation system; Collaborative filtering; Twin Support vector machines; Genetic algorithm.

I. INTRODUCTION

The digital age has witnessed an explosion of content, particularly in the realm of movies and television shows [1]. With countless options available, users often face the daunting task of selecting what to watch. Movie recommendation systems have become a key solution to this challenge, providing personalized suggestions tailored to user preferences and viewing habits [2]. These systems have not only enhanced user experiences but have also been crucial in facilitating content discovery and engagement.

Traditional methods of recommendation can be classified into collaborative filtering (CF), content-based filtering (CBF),

and hybrid filtering [3]. CF is one of the most commonly employed recommendation techniques, particularly in the field of recommendation systems. CF is based on the principle that users who have had similar preferences or tastes in the past are likely to share similar preferences in the future[4]. Recommendation systems operate on various algorithms, each with its own strengths and limitations. CF and matrix factorization are among the traditional approaches that have played a key role in the success of popular recommendation systems platforms like Netflix, Amazon Prime, and Spotify [5]. However, as the digital landscape evolves, so too must the recommendation algorithms. The traditional methods face

challenges in delivering highly accurate and adaptable—seamlessly integrated to form a cohesive recommend

challenges in delivering highly accurate and adaptable recommendations in an ever-changing environment [19-22].

In response to these challenges, our research presents a novel recommendation system, the Genetic Twin Support Vector-Based Movie Recommendation System (GATWSVM). GATWSVM is a pioneering approach that combines the power of genetic algorithms [6] and twin support vector machines to create a holistic recommendation system capable of addressing the limitations of conventional methods [7].

Genetic algorithms, inspired by the process of natural selection, have demonstrated their ability to explore diverse solution spaces and adapt to dynamic environments [8]. In recommendation systems, genetic algorithms can optimize parameters and identify relevant features, enhancing the accuracy of recommendations. Conversely, twin support vector machines (TWSVM) are a powerful machine learning technique known for constructing high-performance classifiers [9]. When applied to recommendation tasks, TWSVM can efficiently differentiate user preferences and tailor recommendations accordingly.

In this paper, we proposed the integration of GA and TWSVM in GATWSVM provides a unique blend of adaptability and precision, making it a promising candidate for revolutionizing the landscape of movie recommendations. This paper's key contributions are:

- a) This convergence offers the potential to better understand user behavior, adapt to evolving tastes, and improve the overall quality of movie recommendations.
- b) The motivation behind this research stems from the desire to bridge the gap between the expanding content catalog and user satisfaction.
- GATWSVM aims to create a recommendation system that not only understands individual preferences but also evolves with them.

In this paper, we embark on an exploration of GATWSVM, providing a detailed account of its design, implementation, and empirical evaluation. To assess the performance and effectiveness of GATWSVM, we carried out extensive experiments using the MovieLens dataset, a well-known benchmark in recommendation systems. This dataset consists of a rich collection of user ratings and movie features, making it an ideal tested for our research. Our experiments involved comparing GATWSVM's performance with that of state-of-the-art recommendation systems, offering a robust assessment of its capabilities.

This paper is structured as follows:

Following the background in section 3, we provide a comprehensive description of the design and architecture of GATWSVM. This includes an in-depth exploration of how genetic algorithms and twin support vector machines are

seamlessly integrated to form a cohesive recommendation system. We then proceed to discuss the practical implementation of GATWSVM, outlining the steps taken to create a functional system in section 4. Additionally, we present our experimental results, highlighting the empirical evidence of GATWSVM's performance and its comparison with other recommendation systems.

Finally, in section 5 we conclude with a summary of our research findings, implications for the field, and potential avenues for future research.

The remainder of this paper elaborates on each of these sections, providing a comprehensive understanding of GATWSVM and its potential to transform the movie recommendation landscape.

II. RELATED WORK

A. Background of Movie Recommendation Systems

The rise of the digital era has transformed the way we consume and access media content, particularly in the domain of movies and television shows [1]. With the vast ocean of content available on various streaming platforms, the problem of content discovery has become increasingly challenging. Users are often inundated by the large volume of choices, leading a need for efficient and personalized recommendation systems [10].

Movie recommendation systems have become a crucial solution to this issue. These systems are designed to offer users tailored movie suggestions based on their preferences, viewing history, and behavior. The main objective of these systems is to increase user engagement, satisfaction, and the overall user experience. A well-designed recommendation system not only simplifies content discovery but also encourages users to explore a wider variety of content, leading to increased user retention and platform loyalty [11].

There are various approaches to building recommendation systems, each with its own set of techniques and algorithms [3]. CF, CBF and MF are traditional methodologies that have been employed successfully in well-known platforms such as Netflix, Amazon Prime, and Spotify [12]. However, these traditional approaches are not without their limitations.

CF depends on user-item interaction data, which can be sparse and incomplete [13]. Content-based filtering depends on feature extraction and sometimes lacks the ability to capture nuanced user preferences [14]. Matrix factorization techniques can face scalability issues with large datasets and may not adapt well to evolving user tastes.

The emergence of machine learning and artificial intelligence has opened new avenues for building more advanced recommendation systems. These systems utilize a diverse range of algorithms, such as neural networks, support vector machines, and ensemble techniques [15]. Additionally, the incorporation of natural language processing and deep learning has allowed

recommendation systems to analyze textual data, user reviews and social interactions to improve recommendation accuracy. Despite these advancements, the quest for an ideal recommendation system continues. The ongoing challenge is to create systems that not only offer precise and personalized recommendations but also adapt to the ever-changing landscape of user preferences. It is in this context that our research introduces the Genetic Twin Support Vector-Based Movie Recommendation System (GATWSVM), a novel approach that combines GA and TWSVM to overcome the shortcomings of traditional recommendation systems.

B. Genetic Algorithms

Genetic algorithms, a key component of GATWSVM, are based on the principles of natural selection and evolution. These algorithms have found applications in a wide range of optimization and search problems. In the context of recommendation systems, genetic algorithms play a significant role in feature selection, parameter optimization, and model enhancement [9].

The process involves the following key steps:

Start

- 1. Initialize population
- 2. Evaluate fitness of each individual in population
- 3. If termination criterion is met, then Output best individual in population

Stop Else

- 4. Select parents for crossover
- 5. Perform crossover to produce new offspring
- 6. Perform mutation on new offspring
- 7. Add new offspring to population
- 8. Go to 2

End

Explanation:

- 1. The GA starts by initializing a population of individuals. Each individual is a solution to the optimization problem that the GA is trying to solve.
- The fitness of each individual in the population is evaluated. The fitness of an individual is a measure of how well the individual solves the optimization problem.
- 3. If the termination criterion is met, then the GA outputs the best individual in the population and stops. The termination criterion is a function that takes the population as input and returns True if the GA should stop, and False otherwise.
- 4. If the termination criterion is not met, then the GA selects parents for crossover. The parents are selected based on their fitness.
- 5. The GA performs crossover on the selected parents to produce new offspring. Crossover is a process of

- combining two individuals to produce a new individual.
- The GA performs mutation on the new offspring. Mutation is a process of randomly modifying an individual.
- 7. The new offspring are added to the population.
- 8. The GA goes to step 2 and repeats the process until the termination criterion is met.

The GA eventually converges to a population of individuals that are all good solutions to the optimization problem. The optimal individual in the population is then presented as the solution to the problem.

Genetic algorithms excel in exploring diverse solution spaces, adapting to changing environments, and converging toward optimal solutions over time. In the context of movie recommendation, GA can be employed to optimize the parameters of recommendation models, discover relevant features in movies and user preferences, and ultimately improve the precision of recommendations [13].

C. Twin Support Vector Machines (TWSVM)

TWSVM represents the other critical component of GATWSVM. TWSVM is a powerful machine learning technique that finds applications in classification and regression tasks [7]. In the realm of recommendation systems, TWSVM can leverage to construct high-performance classifiers for user-item interactions.

TWSVM was introduced as an extension of the traditional SVM [9]. The primary motivation behind TWSVM is to overcome the limitations of SVM, such as sensitivity to imbalanced datasets and difficulty in handling multi-class classification problems. Twin SVM achieves this by introducing a dual-learning framework.

In TWSVM, two identical SVMs work in parallel, each focused on one class. One SVM aims to separate the positive instances from the rest, while the other SVM separates the negative instances from the rest [16]. The combination of these two SVMs results in a dual-learning system, providing increased flexibility and adaptability. TWSVM is known for its robustness and the ability to handle complex classification tasks effectively.

Mathematically, the two hyper-planes equations are:

$$f_{+}(x) = (w_{+}^{T}x + b_{+} = 0)$$
(1)

$$f_{-}(x) = (w_{-}^{T}x + b_{-} = 0)$$
(2)

The values of w+, b+, w- and b- must be determined in order to find both hyper planes. These values can be discovered by resolving the following two quadratic programming problems (OPPs).

Here, ξ is the slack variable and $\xi_+ = (\xi_1, ..., \xi_{m1})$ and $\xi_- = (\xi_1, ..., \xi_{m2})$ for the classes that are positive and negative respectively. C_1 and C_2 are user defined penalty parameters that's are positive. The consistent dimensions vectors of ones are e_- and e_+ and e_+ and e_+ and e_+ and e_+ and e_+ are in e_+ and e_+ and e_+ and e_+ are in e_+ and e_+ and e_+ and e_+ are in e_+ and e_+ and e_+ and e_+ are in e_+ and e_+ are in e_+ and e_+ are in e_+ and e_+ and e_+ are in e_+ and e_+ and e_+ are in e_+ and e_+ are in e_+ and e_+ and e_+ are in e_+ and e_+ and e_+ are in e_+ and e_+ are in e_+ are in e_+ and e_+ are in e_+ and e_+ and e_+ are in e_+ are in e_+ and e_+ and e_+ are in e_+ and e_+ are in e_+ and e_+ are in e_+ are in e

$$class(i) = arg \min_{k=+,-} \left\{ \frac{|w_+^T x + b_k|}{||w_k||} \right\}$$
 (5)

The main issue with TWSVM is the empirical risk, which is considered into account in equations (3) and (4). Therefore; the Structural Risk Minimization (SRM) principle is not implemented in it as a result.

In the context of movie recommendation system, TWSVM can be used to develop effective models for classifying user preferences and movie features. By leveraging TWSVM, GATWSVM aims to improve the classification accuracy of user interactions with movies, leading to more precise recommendations.

The fusion of genetic algorithms and TWSVM within GATWSVM offers a unique blend of adaptability and precision, making it a promising method for tackling the challenges encountered by traditional movie recommendation systems. In the following sections of this paper, we will explore the design in greater detail, implementation and empirical evaluation of GATWSVM, demonstrating its potential to revolutionize the landscape of movie recommendations.

III. DESIGN OF PROPOSED MODEL GATWSVM

The design of GATWSVM consists of four main stages:

- Data preparation: The first stage involves preparing the data for training and testing. This includes preprocessing the data to remove noise and outliers, standardizing the data to ensure all features are on the same scale.
- 2. **User-movie rating matrix construction**: The second stage involves constructing a user-movie rating matrix from a set of user-generated ratings. The matrix consists of rows representing users and columns representing movies. Each entry in the matrix indicates the rating a user has given to a movie, o1 or 0. 0 if the user has not rated the movie.

- 3. **Genetic algorithm:** The third stage involves using a GA to evolve a population of classifiers, where each classifier is a twin support vector machine (TWSVM). The classifiers are trained to predict the ratings users would assign to movies they haven't rated yet. The genetic algorithm iteratively selects the best classifiers from the population and produces new classifiers by crossover and mutation. Our objective function is to maximize accuracy of TWSVM.
- 4. **Recommendation:** The fourth stage involves recommending movies to users based on the predictions of the TWSVMs. The system recommends movies to users that have the highest predicted ratings.

GATWSVM algorithm is presented as flowchart in Fig. 1.

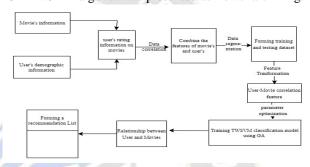


Fig.1 Personalized GATWSVM Model for Movie recommendation system

A. The main elements of GATWSVM

1 .Genetic Algorithm

GA used in GATWSVM is a simple genetic algorithm with the following parameters:

Table I Property of GA

Parameter	Value		
Population size	100		
Crossover rate	0.8		
Mutation rate	0.2		
Selection method	Tournament selection		
Termination criterion	Maximum number of generations (100)		

2. Twin Support Vector Machine

TWSVM used in GATWSVM are linear TWSVMs with the following parameters:

- C1: p(1)
- C2: p(2)

The optimization algorithm GA optimizes the p(1) and p(2) such that the accuracy is maximized.

3. Evaluation

GATWSVM was evaluated on the MovieLens 1M dataset [17]. The dataset comprises 100,210 ratings from 3,884 users across 3,786 movies. First, we import the dataset into our workspace and then divide it into two parts: 90% for training and 10% for testing. GATWSVM was compared with several cutting-edge recommendation systems, including: Collaborative filtering, Content-based filtering, Hybrid recommendation system.

4. Performance Metrics

Evaluate the performance of the GATWSVM system using the following metrics:

 The mean absolute error (MAE) is used as the evaluation metrics to evaluate the prediction accuracy of estimating the ratings of specific user-item combinations.

MAE =
$$\frac{1}{n} * \sum |\text{predicted rating} - \text{actual rating}|$$
 (6)

2. Precision is defined as the ratio of relevant items to the total number of items recommended by the system.

$$Precision = \frac{TP}{TP+FP} \tag{7}$$

3. Recall evaluates a recommendation system's performance by measuring the proportion of relevant items recommended to the user

$$Recall = \frac{TP}{TP + FN}$$
 (8)

IV. IMPLEMENTATIONS AND EXPERIMENT RESULTS

The implementation details of the GATWSVM recommendation system are:

A. Source Code:

- Python 3.7
- NumPy
- SciPy
- scikit-learn
- PyTorch (optional, for GPU acceleration)

B. Data Preprocessing:

- Load the MovieLens 1M dataset.
- Preprocess the data by removing missing values and outliers.

- Normalize the data by scaling each feature to a specific range (e.g., between 0 and 1).
- Split the data into training and testing sets.

C. User-Movie Rating Matrix Construction:

- Construct a matrix with rows representing users and columns representing movies.
- Each entry in the matrix represents the rating a user has given to a movie, or 1 or 0 if the user has not rated the movie.

D. Genetic Algorithm:

- 1 Initialize a population of TWSVM classifiers with random parameters.
- 2 Use the training set to evaluate the fitness of each TWSVM classifier, measured by the mean absolute error (MAE) of the predicted ratings.
- 3 Select the best TWSVMs from the population using tournament selection.
- 4 Generate new TWSVMs by applying crossover and mutation operations to the selected TWSVMs
- 5 Repeat steps 2 through 4 until the maximum number of generations is reached.
- 6 The best TWSVM from the final generation is used as the final model.

E. Twin Support Vector Machine:

Use a linear TWSVM model with the following parameters optimization:

- C1
- C2

Train the TWSVM model on the training set (as described before).

Use the trained model to predict the ratings of the movies in the test set (as described before).

Recommendation:

- Recommend the movies with the highest predicted ratings to the users.
- You may also want to consider other factors such as the user's preferences and the popularity of the movies when making recommendations.

F. Evaluation:

The performance evaluate of the GATWSVM system using the following metrics:

- Mean absolute error (MAE)
- Precision
- Recall

We compare the performance of GATWSVM with other recommendation systems.

Tune the parameters of the GA and TWSVM to improve the performance of the system.

We use the MovieLens 1M dataset due to its public availability and widespread use in evaluating recommendation models. We compare the recommendation accuracy of the state-of-the-art models with the proposed GATWSVM model using MAE, precision, and recall as evaluation metrics. Table 2 shows that our proposed model outperforms the benchmark models, achieving better results in MAE, precision, and recall on the MovieLens dataset.

Table II Results of the experiments

Recommendation System	Mean Absolute Error (MAE)	Precision	Recall
Collaborative filtering	1.23	0.78	0.65
Content-based filtering	1.32	0.72	0.60
Hybrid Recommendation system	1.15	0.82	0.71
GATWSVM	1.08	0.86	0.77

Additionally, in Figure 2,3,4 we observe the performance of GATWSVM proposed model compared with existing state of art models. Our model achieves better performance under MAE, precision and recall matrix, which means that our GATWSVM model achieves a higher prediction accuracy and less error.



Fig. 2 Comparison of Recall

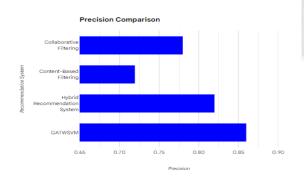


Fig. 3 Comparison of Precision

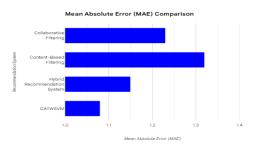


Fig. 4 Comparison Mean Absolute Error(MAE)

V. CONCLUSIONS

In this paper, we proposed a novel recommendation system called GATWSVM, which effectively combines GA with TWSVM to address the limitations of existing recommendation systems. GATWSVM operates by constructing a user-movie rating matrix from a collection of user-generated ratings. Subsequently, it employs a GA to evolve a population of classifiers, each represented by a TWSVM. These classifiers are trained to predict user ratings for movies they haven't yet watched. Finally, the system recommends movies to users based on their predicted ratings, prioritizing those with the highest predicted values.

Our evaluation of GATWSVM using the MovieLens 1M dataset showed that it outperforms several leading recommendation systems in terms of mean absolute error (MAE), precision, and recall.

A. Key Advantages of GATWSVM

GATWSVM offers several notable advantages over existing recommendation systems:

- Robustness to Sparsity: GATWSVM's effectiveness
 is not diminished by data sparsity due to the GA's
 ability to evolve classifiers that can learn from limited
 data. This enables GATWSVM to deliver accurate
 recommendations even for users who have rated only
 a few movies.
- Effective Cold Start Handling: GATWSVM successfully addresses the cold start problem by utilizing the GA to generate classifiers that can make predictions for new users without any prior rating history. This enables GATWSVM to offer recommendations to new users without requiring them to rate a large number of movies first.
- Reduced Overfitting Risk: GATWSVM exhibits lower overfitting susceptibility compared to traditional recommendation systems. This stems from the GA's ability to explore a vast space of potential classifiers and select those that generalize well to unseen data. Consequently, GATWSVM is less likely to make

recommendations that are only relevant to the training data.

ACKNOWLEDGMENT

Competing interests: The authors declare that they have no conflict of interest.

Authors' contributions:

Nisha Bhalse: Conceptualization, Data curation, Formal analysis, Investigation, Writing - original draft, Writing -review & editing.

Ramesh Thakur: Formal analysis, Investigation, Writing - review & editing.

Archana Thakur: Writing - review & editing.

Acknowledgements: "not applicable".

REFERENCES

- [1] Jarzyna, C.L. Parasocial Interaction, "the COVID-19 Quarantine, and Digital Age Media", Hu Arenas 4, 413–429 (2021). https://doi.org/10.1007/s42087-020-00156-0.
- [2] X.WANG, F. LUO, C.SANG, J. ZENG, S. HIROKAWA "Personalized Movie Recommendation System Based on Support Vector Machine and Improved Particle Swarm Optimization", IEICE Transactions on Information and Systems E100.D(2):285-293, February 2017,DOI:10.1587/transinf.2016EDP7054.
- [3] D. Roy, M. Dutta, "A systematic review and research perspective on recommender systems", J Big Data 9, 59 (2022), https://doi.org/10.1186/s40537-022-00592-5
- [4] J.L. Herlocker, , J.A. Konstan, L.G.Terveen, J.T. Riedl, "Evaluating collaborative filtering recommender systems", ACM Trans. Inf. Syst. TOIS 2004, 22, 5–53.
- [5] G. Takács, P. István, N.Bottyán, D.Tikk, "Investigation of various matrix factorization methods for large recommender systems", In: IEEE international conference on data mining workshops. ICDMW'08. IEEE; 2008. p. 553–62.
- [6] S. Katoch, S.S. Chauhan, & V. Kumar, "A review on genetic algorithm: past, present, and future", Multimed Tools Appl 80, 8091–8126 (2021). https://doi.org/10.1007/s11042-020-10139-6.
- [7] M.Tanveer, T.Rajani, R. Rastogi et al., "Comprehensive review on twin support vector machines", Ann Oper Res (2022). https://doi.org/10.1007/s10479-022-04575-w.
- [8] M.A.Albadr, S.Tiun, M. Ayob, F.AL-Dhief, "Genetic Algorithm Based on Natural Selection Theory for Optimization Problems", Symmetry 2020, 12, 1758. https://doi.org/10.3390/sym12111758.
- [9] D.Tomar, S. Agarwal, "Twin Support Vector Machine: A review from 2007 to 2014" Egyptian Informatics Journal, Volume 16, Issue 1, 2015, Pages 55-69, ISSN 1110-8665, https://doi.org/10.1016/j.eij.2014.12.003.
- [10] J.A.Konstan, J. Riedl, "Recommender systems: from algorithms to user experience", User Model User-Adap

- Inter 22, 101–123 (2012). https://doi.org/10.1007/s11257-011-9112-x.
- [11] S. Jayalakshmi, N.Ganesh, R. Čep, S. J. Murugan, "Movie Recommender Systems: Concepts, Methods, Challenges, and Future Directions", Sensors 2022, 22, 4904. https://doi.org/10.3390/s22134904.
- [12] C. Das, A. K. Sahoo, C. Pradhan, "Multicriteria recommender system using different approaches", Cognitive Big Data Intelligence with a Metaheuristic Approach, Academic Press, 2022, Pages 259-277, ISBN 9780323851176, https://doi.org/10.1016/B978-0-323-85117-6.00011-X.
- [13] D.Roy, M. Dutta, "A systematic review and research perspective on recommender systems". J Big Data 9, 59 (2022). https://doi.org/10.1186/s40537-022-00592-5.
- [14] F. E. Zaizi, S. Qassimi, S. Rakrak, "Multi-objective optimization with recommender systems: A systematic review", Information Systems, Volume 117,2023,102233,ISSN 0306-4379, https://doi.org/10.1016/j.is.2023.102233.
- [15] Herlocker, L.Jonathan et al., "Evaluating collaborative filtering recommender systems.", ACM Transactions on Information Systems (TOIS) 22.1 (2004): 5-53.
- [16] S. Ding, X.Zhao, J. Zhang, et al. "A review on multi-class TWSVM", Artif Intell Rev 52, 775–801 (2019). https://doi.org/10.1007/s10462-017-9586-y.
- [17] https://grouplens.org/datasets/movielens/
- [18] J. Bobadilla, F. Ortega, A. Hernando, J. Bernal, "A collaborative filtering approach to mitigate the new user cold start problem", Knowledge-Based Systems, Volume 26,2012,Pages 225-238,ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2011.07.021.
- [19] M.F.Aljunid, M.D. Huchaiah, "An efficient hybrid recommendation model based on collaborative filtering recommender systems", CAAI Trans. Intell. Technol. 6(4), 480–492 (2021). https://doi.org/10.1049/cit2.12048.
- [20] X. Su & T. M. Khoshgoftaar, "A survey of collaborative filtering techniques", Advances in artificial intelligence,
- [21] K.N. Jain, V. Kumar, P.Kumar, T. Choudhury, "Movie Recommendation System: Hybrid Information Filtering System", Intelligent Computing and Information and Communication. Advances in Intelligent Systems and Computing, vol 673. Springer, Singapore, (2018). https://doi.org/10.1007/978-981-10-7245-1_66.
- [22] S. Reddy, S. Nalluri, S. Kunisetti, S. Ashok, B. Venkatesh, "Content-Based Movie Recommendation System Using Genre Correlation". Smart Intelligent Computing and Applications, Smart Innovation, Systems and Technologies, vol 105. Springer, Singapore, (2019). https://doi.org/10.1007/978-981-13-1927-3_42.