iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XI **Month of publication:** November 2021

DOI: https://doi.org/10.22214/ijraset.2021.38651

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

Movie Recommendation System Mistreatment Current Trends and Sentiment Analysis from Micro Blogging Knowledge

J Manikandan[1], Abdul Kani S[2], Ch. Gopi Krishna[3], G. Sravan Rao[4], Vishnu Pantham[5]
[1]*Assistant Professor, Department of CSE, St. Martin's Engineering College, Secunderabad, Telangana, India-500014*
[2]*Sr. Assistant Professor, Department of CSE, Indur Institute of Engineering and Technology, Siddipet, India -502277*
[3, 4, 5]*Assistant Professor, Department of CSE, Indur Institute of Engineering and Technology, Siddipet, India -502277*

Abstract: ***Recommendation systems (RSs) have garnered immense interest for applications in e-commerce and digital media. Traditional approaches in RSs include such as collaborative filtering (CF) and content-based filtering (CBF) through these approaches that have certain limitations, such as the necessity of prior user history and habits for performing the task of recommendation. To minimize the effect of such limitation, this article proposes a hybrid RS for the movies that leverage the best of concepts used from CF and CBF along with sentiment analysis of tweets from microblogging sites. The purpose to use movie tweets is to understand the current trends, public sentiment, and user response of the movie. Experiments conducted on the public database have yielded promising results.***
Keywords: Collaborative filtering, Content based filtering, Recommendation System, Sentiment Analysis, Twitter

I. INTRODUCTION

Traditional approaches in RSs include such as collaborative filtering (CF) and content-based filtering (CBF) through these approaches that have certain limitations, such as the necessity of prior user history and habits for performing the task of recommendation. Users often face the problem of excessive available information. Recommendation systems (RSs) are deployed to help users cope with the information explosion. RS is mostly used in digital entertainment, such as Netflix, Prime Video, and IMDB, and e-commerce portals such as Amazon, Flipkart, and eBay. In this article, we focus on RS for movies, which is an important source of recreation and entertainment in our life. Movie suggestions for users depend on Web-based portals. Movies can be easily differentiated through their genres, such as comedy, thriller, animation, and action. Another possible way to categorize the movies based on its metadata, such as release year, language, director, or cast. Most online video-streaming services provide personalized user experience by utilizing the user's historical data, such as previously viewed or rated history.The purpose to use movie tweets is to understand the current trends, public sentiment, and user response of the movie. Experiments conducted on the public database have yielded promising results.

II. LITERATURE SURVEY

How can micro-blogging activities on Twitter be leveraged for user modeling and personalization? In this paper we investigate this question and introduce a framework for user modeling on Twitter which enriches the semantics of Twitter messages (tweets) and identifies topics and entities (e.g. persons, events, products) mentioned in tweets[5]. We analyze how strategies for constructing hashtag-based, entity-based or topic-based user profiles benefit from semantic enrichment and explore the temporal dynamics of those profiles[17]. We further measure and compare the performance of the user modeling strategies in the context of a personalized news recommendation system[10]. Our results reveal how semantic enrichment enhances the variety and quality of the generated user profiles[4]. Further, we see how the different user modeling strategies impact personalization and discover that the consideration of temporal profile patterns can improve recommendation quality[9].
How can micro-blogging activities on Twitter be leveraged for user modeling and personalization? In this paper we investigate this question and introduce a framework for user modeling on Twitter which enriches the semantics of Twitter messages (tweets) and identifies topics and entities (e.g., persons, events, products) mentioned in tweets[3].

©IJRASET: All Rights are Reserved

We analyze how strategies for constructing hashtag-based, entity-based or topic-based user profiles benefit from semantic enrichment and explore the temporal dynamics of those profiles[8]. We further measure and compare the performance of the user modeling strategies in the context of a personalized news recommendation system [23]. Our results reveal how semantic enrichment enhances the variety and quality of the generated user profiles[2]. Further, we see how the different user modeling strategies impact personalization and discover that the consideration of temporal profile patterns can improve recommendation quality[2].

This paper presents an overview of the field of recommender systems and describes the current generation of recommendation methods that are usually classified into the following three main categories: content-based, collaborative, and hybrid recommendation approaches [19]. This paper also describes various limitations of current recommendation methods and discusses possible extensions that can improve recommendation capabilities and make recommender systems applicable to an even broader range of applications[7]. These extensions include, among others, an improvement of understanding of users and items, incorporation of the contextual information into the recommendation process, support for multicriteria ratings, and a provision of more flexible and less intrusive types of recommendations [20].

Deep learning techniques for Sentiment Analysis have become very popular [26]. They provide automatic feature extraction and both richer representation capabilities and better performance than traditional feature-based techniques (i.e., surface methods)[1]. Traditional surface approaches are based on complex manually extracted features, and this extraction process is a fundamental question in feature driven methods [16]. These long-established approaches can yield strong baselines, and their predictive capabilities can be used in conjunction with the arising deep learning methods [27]. In this paper we seek to improve the performance of deep learning techniques integrating them with traditional surface approaches based on manually extracted features[6]. The contributions of this paper are sixfold [18]. First, we develop a deep learning-based sentiment classifier using a word embeddings model and a linear machine learning algorithm [21]. This classifier serves as a baseline to compare to subsequent results. Second, we propose two ensemble techniques which aggregate our baseline classifier with other surface classifiers widely used in Sentiment Analysis[11]. Third, we also propose two models for combining both surface and deep features to merge information from several sources [22]. Fourth, we introduce a taxonomy for classifying the different models found in the literature, as well as the ones we propose. Fifth, we conduct several experiments to compare the performance of these models with the deep learning baseline [24]. For this, we use seven public datasets that were extracted from the microblogging and movie reviews domain. Finally, as a result, a statistical study confirms that the performance of these proposed models surpasses that of our original baseline on F1-Score. [12]

Recommendation systems get ever-increasing importance due to their applications in both academia and industry. The most popular type of these systems, known as collaborative filtering algorithms, employ user-item interactions to perform the recommendation tasks. With growth of additional information sources other than the rating (or purchase) history of users on items, such as item descriptions and social media information, further extensions of these systems have been proposed, known as hybrid recommendation algorithms [14]. Hybrid recommenders use both user-item interaction data and their contextual information. In this work, we propose new hybrid recommender algorithms by considering the relationship between content features. This relationship is embedded into the hybrid recommenders to improve their accuracy. We first introduce a novel method to extract the content feature relationship matrix, and then the collaborative filtering recommender is modified such that this relationship matrix can be effectively integrated within the algorithm [13]. The proposed algorithm can better deal with the cold-start problem than the state-of-art algorithms. We also propose a novel content-based hybrid recommender system. Our experiments on a benchmark movie dataset show that the proposed approach significantly improves the accuracy of the system, while resulting in satisfactory performance in terms of novelty and diversity of the recommendation lists [15].

III. PROPOSED METHODOLOGY

The proposed sentiment-based RS is shown in Fig. 1. In this section, we describe various components of the proposed RS. A. Data Set Description The proposed system needs two types of databases. One is a user-rated movie database, where ratings for relevant movies are present, and another is the user tweets from Twitter.

1) *Public Databases:* There are many popular public databases available, which have been widely used to recommend the movies and other entertainment media. To incorporate the sentiment analysis in the proposed framework, the tweets of movies were extracted from Twitter against the movies that were available in the database. Experiments conducted using various public databases, such as the Movielens 100K,2 Movielens 20M,3 Internet Movie Database (IMDb,4) and Netflix database,5 that were not found suitable for our work due to the absence of microblogging data. After a thorough assessment of the abovementioned databases, the MovieTweetings database [12] was finally selected for the proposed system. MovieTweetings

©IJRASET: All Rights are Reserved

is widely considered as a modern version of the MovieLens database. The purpose of this database is to provide an up-to-date movie rating so that it contains more realistic data for sentiment analysis. Table I displays the relevant details of the MovieTweetings database.

2) *Modified MovieTweetings Database:* In the proposed work, the MovieTweetings database is modified to implement the RS. The primary objective to modify the database was to use sentiment analysis of tweets by the users, in the prediction of the movie RS. The MovieTweetings database contains the movies with published years from 1894 to 2017. Due to the scarcity of tweets for old movies, we only considered the movies that were released in or after the year 2014 and extracted a subset of the database which complied with our objective.

A. *Content-based Filtering*

Content-based filtering uses item features to recommend other items similar to what the user likes, based on their previous actions or explicit feedback.To demonstrate content-based filtering, let's hand-engineer some features for the Google Play store. The following figure shows a feature matrix where each row represents an app and each column represents a feature. Features could include categories (such as Education, Casual, Health), the publisher of the app, and many others. To simplify, assume this feature matrix is binary: a non-zero value means the app has that feature.

You also represent the user in the same feature space. Some of the user-related features could be explicitly provided by the user. For example, a user selects "Entertainment apps" in their profile. Other features can be implicit, based on the apps they have previously installed. For example, the user installed another app published by Science R Us.

The model should recommend items relevant to this user. To do so, you must first pick a similarity metric (for example, dot product). Then, you must set up the system to score each candidate item according to this similarity metric. Note that the recommendations are specific to this user, as the model did not use any information about other users.

B. *Collaborative Filtering*

To address some of the limitations of content-based filtering, collaborative filtering uses similarities between users and items simultaneously to provide recommendations. This allows for serendipitous recommendations; that is, collaborative filtering models can recommend an item to user A based on the interests of a similar user B. Furthermore, the embeddings can be learned automatically, without relying on hand-engineering of features.

C. *Sentiment Analysis*

Sentiment analysis is a widely used technique by researchers to acquire people's opinions. Sentiment analysis has been used for rating products for online software services. Their research enhances both CBF and CF algorithms, using external reviews such as sentiment analysis and subjective logic. Sentiment analysis technique has been used to calculate the polarity and confidence of review sentences. The authors proposed Valence Aware Dictionary and Sentiment Reasoner (VADER) model for sentiment analysis. Lexical features were combined for five general rules that embody grammatical and syntactical conventions for expressing and emphasizing sentiment intensity. An F1 score of 0.96 was recorded for classifying the tweets into positive, neutral, and negative classes. In, the author proposed an automatic feedback technique on the basis of data collected from Twitter. Different classifiers such as Support Vector Machine, Naive Bayes, and Maximum Entropy were used on twitter comments. The authors proposed a music recommendation framework for mobile devices where recommendations of songs for a user were based on the mood of the user's sentiment intensity. The studies were performed on 200 participants (100 men and 100 women) to fill out their musical preference choice in his or her profile. Later, the participant's profile was analyzed and the results showed 91% user satisfaction rating. In, the author proposed the KBridge framework to solve the cold start problem in the CF system. Sentiment analysis was also used for microblogging posts in this framework and the polarity score of the post was assigned in 1 to 5 rating. The result showed an enhanced recommendation system by bridging the gap between user communication knowledge between social networking sites and programs watching over TV. The author proposed a rating inference approach to transform textual reviews and ratings to enable easy integration of sentiment analysis and CF.

D. *Hybrid Recommendation*

In this section, we describe the combination of content-based similarity features with collaborative social filtering to generate a hybrid recommendation model. Let f = {f1,f2,· · · ,fn} and q = {q1,q2,· · · ,qn} are the content-based feature vectors and weight vectors, respectively.

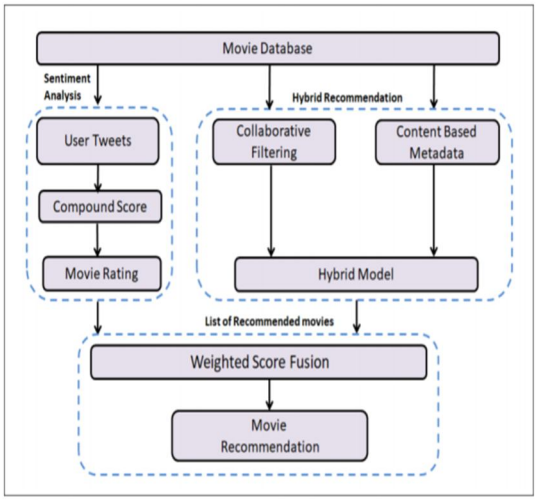
©IJRASET: All Rights are Reserved

We construct the closeness C of two items i and j as: C(i, j) = (PN n=1 fn(Ani , Anj), for i 6= j 0, otherwise (3) where fn(Ani , Anj), corresponds to the similarity between feature values Ani and Anj corresponding to two items. In Eq. (3) the closeness of the items is determined using the metadata or the relevant information related to the items. Fij are constructed by combining the closeness vector C for all the items and multiplying it with the weight vectors q. Fij is a feature matrix of dimension n × M(M−1) 2 , where n and M are the number of feature attributes and number of items, respectively. The weight vectors q are evaluated using a social graph of items that indicate the user likeness of items. Let U = {u1, u2, · · · , un} where ui is a user in the database. A User-Item matrix is constructed for M items. An important property of the UserItem matrix is that it has very high sparsity. Typical collaborative Filtering [42] uses this User-Item matrix to predict a user's rating of a particular item i by analyzing the ratings of other users in the user's neighborhood, normally a K neighboring users. Neighboring users are recognized by similarity measures like Cosine Similarity, Pearson Correlation etc. After selecting K neighbouring users, the weighted aggregation of the ratings are as follows: rating(u, i) = 1 K X k∈K similarity(useru, uservk) · rating vki (4) where u and vk are target user and K nearest neighbors, respectively. The procedure of collaborative filtering is used to overcome the sparsity of the User-Item matrix instead of directly using it to predict ratings. We employ the tweaked UserItem matrix to construct a social graph using items as nodes. This graph represents user perception of similarity between items. The determination of feature weights complies with the social graph. In order to determine the optimal feature weights q, we formulate a framework as described in Eq. (5): S(i, j) = q · Fig (5) which can be expanded as: S(i, j) = q1 · f1(A1i , A1j) + q2 · f2(A2i , A2j) + · · · + qn · fn(Ani, Anj) (6) Now for two items i and j, S(i, j) is evaluated from the User-Item matrix, where S(i, j) are the number of users who are interested in both items i, j . For the entire database, S is a matrix of dimension 1 × M(M − 1) 2 and q is a matrix of dimension 1 × n, where n is the number of content-based features and dimensionality of F isMovie Recommendation System using Sentiment Analysis from Microblogging Data 11 n × M(M−1) 2 . We calculate the weight vector q for all the metadata feature attributes for the complete items using the Moore-Penrose Pseudoinverse as in Eq. (7) : q = S −1 · F

E. Weighted Score Fusion

We derived the weights q of the feature vectors using both movies similarity and user-similarity paradigms. These weights q are normalized between [0, 1] and the concept of sentiment-fusion is utilized in the proposed system. Through the retrieved user tweets, a sentiment rating is fabricated for all movies M. Let S ∈ {s1, s2, · · · ,sn} where si is the rating of movies i calculated using Eq. (2). A function G(i, j) for two movies i, j are defined based on their sentiment ratings si and sj as mentioned in Eq. (8): G(i, j) = D − |si − sj | (8) where D is a constant. The constant D in Eq. (8) is taken as 10 because the ratings are in a scale of 1 − 10. Another function H(i, j) defined as: H(i, j) = q · fij (9) where fij is the feature similarity between movies i, j and q are the set of optimal weights as determined by Eq. (7). The final combined similarity CS(i, j) is described in Eq. (10). It is a weighted combination of the defined functions G and H. CS(i, j) = ω1 · H(i, j) + ω2 · G(i, j) (10) ω1 + ω2 = 1, ω1, ω2 ∈ [0, 1] (11) where ω1 corresponds to the weight of the similarity score calculated from the hybrid model and ω2 corresponds to the weight of the sentiment similarity score.

IV. SYSTEM ARCHITECTURE



©IJRASET: All Rights are Reserved

A simple flowchart representing a process for dealing with a non-functioning lamp.A flowchart is a type of diagram that represents a workflow or process. A flowchart can also be defined as a diagrammatic representation of an algorithm, a step-by-step approach to solving a task. The flowchart shows the steps as boxes of various kinds, and their order by connecting the boxes with arrows. This diagrammatic representation illustrates a solution model to a given problem. Flowcharts are used in analyzing, designing, documenting or managing a process or program in various fields

V. RESULT AND DISCUSSION

In this article, we have proposed a movie RS that uses sentiment analysis data from Twitter, along with movie metadata and a social graph to recommend movies. Sentiment analysis provides information about how the audience responds to a particular movie and how this information is observed to be useful. The proposed system used weighted score fusion to improve the recommendations. Based on our experiments, the average precision in Top-5 and Top-10 for sentiment similarity, hybrid, and proposed model are 0.54 and 1.04, 1.86 and 3.31, and 2.54 and 4.97, respectively. We found that the proposed model recommends more precisely than the other models. In the future, we plan to consider more information about the emotional tone of the user from different social media platforms and non-English languages to further improve the RS.

VI. CONCLUSION

RSs are an important medium of information filtering systems in the modern age, where an enormous amount of data is readily available. In this article, we have proposed a movie RS that uses sentiment analysis data from Twitter, along with movie metadata and a social graph to recommend movies. Sentiment analysis provides information about how the audience responds to a particular movie and how this information is observed to be useful. The proposed system used weighted score fusion to improve the recommendations. Based on our experiments, the average precision in Top-5 and Top-10 for sentiment similarity, hybrid, and proposed model are 0.54 and 1.04, 1.86 and 3.31, and 2.54 and 4.97, respectively. We found that the proposed model recommends more precisely than the other models. In the future, we plan to consider more information about the emotional tone of the user from different social media platforms and non-English languages to further improve the RS.

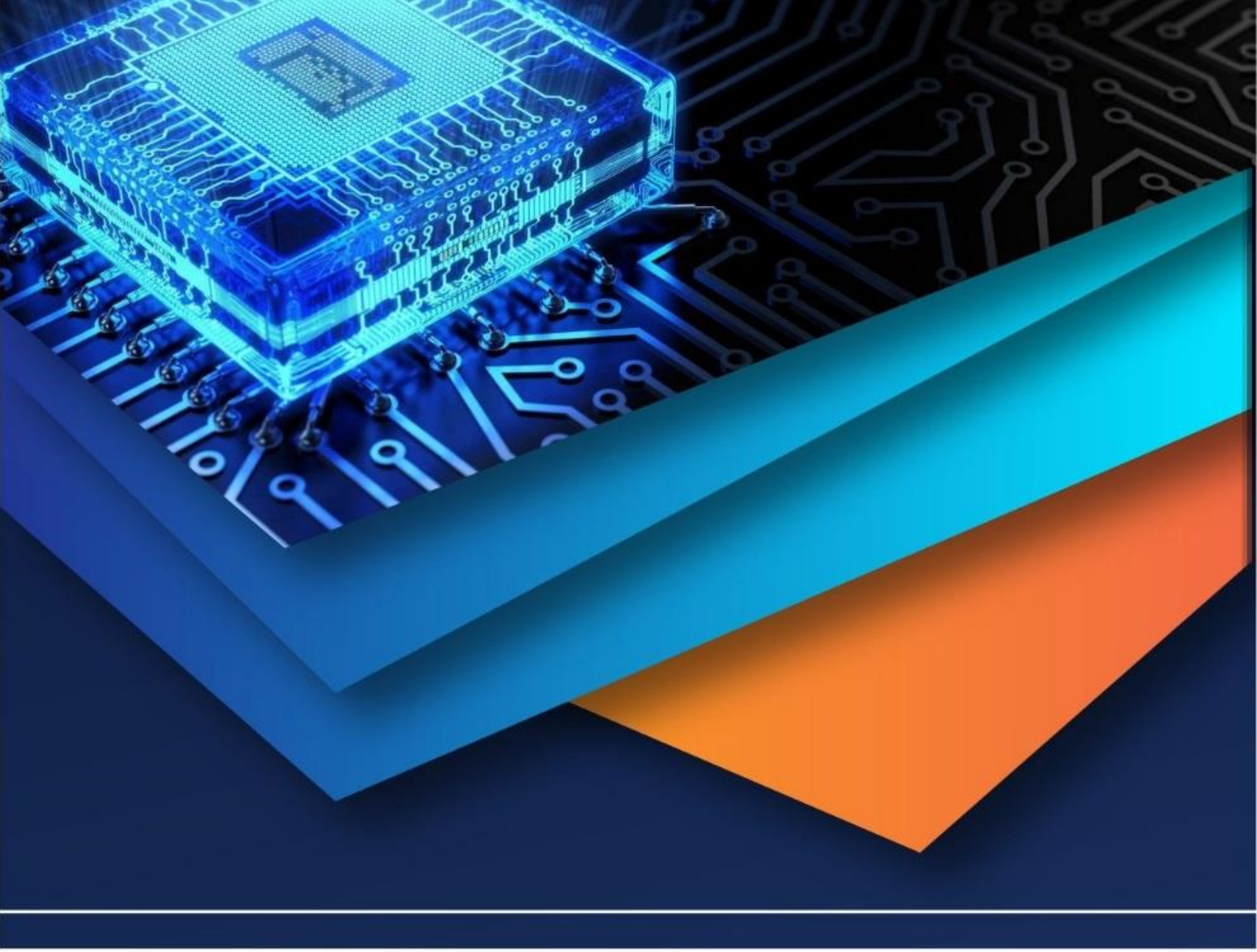
REFERENCES

[1] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Analyzing user modeling on twitter for personalized news recommendations. In International Conference on User Modeling, Adaptation, and Personalization, pages 1–12. Springer, 2011.
[2] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Twitter-based user modeling for news recommendations. In International Joint Conference on Artificial Intelligence, volume 13, pages 2962–2966, 2013.
[3] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6):734–749, 2005.
[4] Minara P Anto, Mejo Antony, KM Muhsina, NivyaJohny, Vinay James, and Aswathy Wilson. Product rating using sentiment analysis. In International Conference on Electrical, Electronics, and Optimization Techniques, pages 3458– 3462. IEEE, 2016.
[5] Ehsan Aslanian, Mohammadreza Radmanesh, and Mahdi Jalili. Hybrid recommender systems based on content feature relationship. IEEE Transactions on Industrial Informatics, 2016.
[6] Jesus Bobadilla, Fernando Ortega, Antonio Hernando, and Javier Alcala. Im- ´ proving collaborative filtering recommender system results and performance using genetic algorithms. Knowledge-based systems, 24(8):1310–1316, 2011.
[7] Robin Burke. Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction, 12(4):331–370, 2002.
[8] Erik Cambria. Effective computing and sentiment analysis. IEEE Intelligent Systems, 31(2):102–107, 2016.
[9] Ivan Cantador, Alejandro Bellog ´ ´ın, and David Vallet. Content-based recommendation in social tagging systems. In Proceedings of the Fourth Conference on Recommender systems, pages 237–240. ACM, 2010.
[10] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. Performance of recommender algorithms on top-n recommendation tasks. In Proceedings of the Fourth Conference on Recommender Systems, pages 39–46. ACM, 2010. ISBN 978-1- 60558-906-0.
[11] Mukund Deshpande and George Karypis. Item-based top-n recommendation algorithms. ACM Transactions on Information Systems, 22(1):143–177, 2004. ISSN 1046-8188.
[12] Simon Dooms, Toon De Pessemier, and Luc Martens. Movietweetings: a movie rating dataset collected from twitter. In Workshop on Crowdsourcing and Human Computation for Recommender Systems, 2013.
[13] Xin Fu and Yun Shen. Study of collective user behaviour in twitter: a fuzzy approach. Neural Computing and Applications, 25(7-8):1603–1614, 2014.Movie Recommendation System using Sentiment Analysis from Microblogging Data 17
[14] Lahiru S Gallege and Rajeev R Raje. Towards selecting and recommending online software services by evaluating external attributes. In the Eleventh Annual Cyber and Information Security Research Conference, page 23, 2016.
[15] HimaanshuGauba, Pradeep Kumar, ParthaPratim Roy, Priyanka Singh, Debi Prosad Dogra, and Balasubramanian Raman. Prediction of advertisement preference by fusing eeg response and sentiment analysis. Neural Networks, 92: 77–88, 2017.
[16] David Goldberg, David Nichols, Brian M Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12):61–70, 1992.

©IJRASET: All Rights are Reserved

[17] Frank Goossen, WouterIJntema, Flavius Frasincar, Frederik Hogenboom, and UzayKaymak. News personalization using the cf-idf semantic recommender. In Proceedings of the International Conference on Web Intelligence, Mining and Semantics, page 10. ACM, 2011.
[18] Guibing Guo and Mohamed Elgendi. A new recommender system for 3d ecommerce: An EEG based approach. Journal of Advanced Management Science, 1(1):61–65, 2013.
[19] Chia-Cheng Hsu, Hsin-Chin Chen, Kuo-Kuang Huang, and Yueh-Min Huang. A personalized auxiliary material recommendation system based on learning style on facebook applying an artificial bee colony algorithm. Computers & Mathematics with Applications, 64(5):1506–1513, 2012.
[20] Clayton J Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth International Conference on Weblogs and Social Media, 2014.
[21] Rahul Katarya. Movie recommender system with metaheuristic artificial bees. Neural Computing and Applications, pages 1–8, 2018.
[22] Rahul Katarya and Om Prakash Verma. Recommender system with grey wolf optimizer and fcm. Neural Computing and Applications, pages 1–9, 2016.
[23] Rahul Katarya and Om Prakash Verma. An effective collaborative movie recommender system with cuckoo search. Egyptian Informatics Journal, 18(2): 105–112, 2017.
[24] Cane WK Leung, Stephen CF Chan, and Fu-lai Chung. Integrating collaborative filtering and sentiment analysis: A rating inference approach. In Proceedings of the ECAI workshop on recommender systems, pages 62–66, 2006.
[25] Fangfang Li, Guandong Xu, and Longbing Cao. Two-level matrix factorization for recommender systems. Neural Computing and Applications, 27(8):2267–2278, 2016.
[26] Hui Li, Jiangtao Cui, Bingqing Shen, and Jianfeng Ma. An intelligent movie recommendation system through group-level sentiment analysis in microblogs. Neurocomputing, 210:164–173, 2016.
[27] Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. Content-based recommender systems: State of the art and trends. Recommender Systems Handbook, pages 73–105. Springer, 2011.
[28] Steven Loria. Textblob: simplified text processing. Secondary TextBlob: Simplified Text Processing, 2014.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)