Movie Recommendation System Using Sentiment Analysis From Microblogging Data

Sudhanshu Kumar, Kanjar De, and Partha Pratim Roy

***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.**

***Index Terms*— Collaborative filtering, content-based filtering, recommendation system (RS), sentiment analysis, Twitter.**

I. Introduction

IN TODAY'S world, the Internet has become an important part of human life. Users often face the problem of excessive available information. Recommendation systems (RSs) are deployed to help users cope up 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 [36], [51] provide personalized user experience by utilizing the user's historical data, such as previously viewed or rated history. Movie RSs [3], [25], [28], [56], [64] help us to quickly search preferred movies over online. The foremost requisite for a movie RS is that it should be trustworthy and provide the users with the recommendation of movies that are resembling their preferences. In recent times, with an exponential increase in the amount of online data, RS is beneficial for making decisions in day-to-day activities. RSs are broadly classified into two categories: collaborative filtering (CF) and content-based filtering (CBF). It is a human tendency to make decisions based on facts, predefined rules, and known information available over the Internet. The inclination of such human behavior gives rise to the concept of CF. Resnick *et al.* [43] introduced the concept of CF in netnews, to help readers find the articles they like, in a huge stream of available articles. CF helps readers make choices based on the perspective of other readers. Two users are considered like-minded when their rating for items is similar. In CBF [54], items are suggested through the similarity among the contextual information of the items. These RS algorithms need historical data to recommend the items.

Manuscript received November 3, 2019; revised April 13, 2020; accepted May 3, 2020. Date of publication May 28, 2020; date of current version August 6, 2020. *(Corresponding author: Sudhanshu Kumar.)*

Sudhanshu Kumar and Partha Pratim Roy are with the Department of Computer Science and Engineering, IIT Roorkee, Roorkee 247667, India (e-mail: skumar2@cs.iitr.ac.in).

Kanjar De is with the Department of Computer Science, Norwegian University of Science and Technology, 2815 Trondheim, Norway.

Digital Object Identifier 10.1109/TCSS.2020.2993585

To overcome this limitation, various social media platforms, such as Quora, Facebook, Instagram, and Twitter, people use to share their daily state of mind over the Internet. Twitter [1], [2], [16] is one of the most popular social media platform founded in 2006 where users can express their thoughts in limited characters. The Unique Selling Proposition of Twitter is that the existing users not only receive information according to their social links but also gain access to other user-generated information. The source of information on Twitter is called tweets. Tweets keep users updated about their favorite topics, people, and movies in limited characters.

The main contributions of this article are as follows.

1) A hybrid RS is proposed by combining CBF and CF.
2) Sentiment analysis is used to boost up this RS.

This article is organized as follows. Section II summarizes the related work. The proposed system is presented in Section III. Results obtained using the proposed framework are shown in Section IV. Finally, the conclusion is drawn in Section V.

II. Related Work

Many RSs have been developed over the past decades. These systems use different approaches, such as CF, CBF, hybrid, and sentiment analysis to recommend the preferred items. These approaches are discussed as follows.

A. Collaborative, Content-Based, and Hybrid Filtering

Various RS approaches have been proposed in the literature for recommending items [48]. The primordial use of CF was introduced in [18], which proposed a search system based on document contents and responses collected from other users. Yang *et al.* [59] inferred implicit ratings from the number of pages the users read. The more pages read by the users, the more they are assumed to like the documents. This concept is helpful to overcome the cold start problem in CF. Optimizing the RS is an ill-posed problem. Researchers have proposed several optimization algorithms, such as gray wolf optimization [26], artificial bee colony [21], particle swarm optimization [53], and genetic algorithms [6]. Katarya *et al.* and Verma [26] developed a collaborative movie RS based on gray wolf optimizer and fuzzy c-mean clustering techniques. Both techniques are applied to the Movielens data set and predicted a better RS. They improved the existing framework in [24] proposing an artificial bee colony and k-mean cluster (ABC-KM) framework for a collaborative movie RS to reduce the scalability and cold start complication. The combination of the hybrid cluster and optimization technique showed better accuracy in movie prediction compared with movie prediction by the existing frameworks. Dong *et al.* [11] proposed feature relearning with data augmentation for the Hulu Content-based Video Relevance Prediction Challenge. The result showed better improvement in TV shows and movie track in recall@100. Most approaches suffer from the sparsity problem in Social-aware Movie Recommendation systems (SMRs). Zhao *et al.* [63] developed a framework called SMR-multimodal network representation learning (MNRL) for movie recommendation to address this issue effectively. The result achieves better performance on a large-scale data set collected from the Chinese social-aware movie recommender site (Douban).

CBF [30], [39], [55], [57] is one of the most widely used and researched RS paradigm. This approach is based on the description of the item and a profile of the user's preferences. Nascimento *et al.* [35] discussed about discriminative power of the words for research articles recommendation. They deduced that title and abstract are multiple times stronger than the body text of the items and thus use the weightage scheme of the title, abstract, and body text to retrieve relevant articles. Cantador *et al.* [9] made use of user and item profiles, described in terms of weighted lists of social tags to provide music recommendations [15], [23], [32]. Meteren and Someren [54] proposed a personalized RS to suggest articles for home improvement where the similarity between the user profile vector and a document was determined by using the combination of TF-IDF and the cosine similarity. Goossen *et al.* [19] proposed a new method for recommending news items based on TF-IDF and a domain ontology, i.e., CF-IDF. The performance of this method outperformed the TF-IDF approach on several measures, such as accuracy, recall, and the F_1-measures when tested, evaluated, and implemented on the Athena framework. Ma *et al.* [31] proposed a latent genre-aware microvideo recommendation model for social media. The MovieLens and Yelp data set features (contextual and visual contents), i.e., user-item interaction and auxiliary features, were fed into a neural network for optimal recommendation scores. Du *et al.* [14] proposed a general framework that incorporates a rich content feature from the video, named as a collaborative embedding regression model, to make an effective video RS for multiple content features' scenarios. The experimental results on the MovieLens and Netflix data sets showed the effectiveness of the model. Recent research has demonstrated that the hybrid approach [5], [7], [40], [45], [50] is more effective than traditional approaches. The hybrid systems mitigate the drawback of individual technique due to the combination of multiple recommendation techniques. Melville *et al.* [34] developed a content-boosted CF system that used pure content-based features in a collaborative framework. This system further improved the prediction, first rater, and the sparsity problem. Zhang *et al.* [62] developed a framework based on user recommender interaction that takes input from the user, recommends N items to the user, and records user choice until none of the recommended items favor. Noguera *et al.* [37] developed a mobile recommender system that combines a hybrid recommendation engine and a mobile 3-D GIS architecture. For testing the proposed framework, 27 users were selected with an age range of 24–48 years. To evaluate the performance of the RS, users were instructed to find restaurants, bars, and accommodation while walking and driving along a motorway. The user feedbacks demonstrated competent performance by the 3-D map-based interface that also overcame the limited screen size of most mobile devices. Harakawa *et al.* [20] proposed a multimodal field-aware factorization machines (FFMs) algorithm to recommend the sentiment-aware personalized tweet. Users' interest is strongly influenced by sentiment factors in the tweet, and thus, this method models users' interest by deriving multimodal FFM that enables collaborative use of multiple factors in a tweet and improves performance. The experimental result of FFM evaluated through mean average precision, which showed a better result in comparison with other methods.

B. Sentiment Analysis

Sentiment analysis [8], [33], [41], [42] is a technique to computationally identifying and categorizing people's opinions expressed in the form of reviews or survey is positive, negative, or neutral. Sentiment analysis has been used TextBlob[1] library to calculate the polarity and subjectivity of the review sentences. Past research has primarily focused on analyzing the user-generated textual reviews and categorized the user reviews into positive or negative classes. In recent years, online reviews also include slang, emoticons, and some common words that help in finding the opinion of users more accurately. Hutto and Gilbert [22] proposed a valence-aware dictionary and sentiment reasoner (VADER) algorithm that is used to parse the user reviews and analyze them using a rule-based model to calculate the sentiment score of the tweets. This method is evaluated and validated in different domains, such as movie reviews, e-commerce product reviews, and news headlines. The result derived from the VADER method showed better performance than other sentiment analysis techniques. Rosa *et al.* [46] 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

[1] https://textblob.readthedocs.io/en/dev/

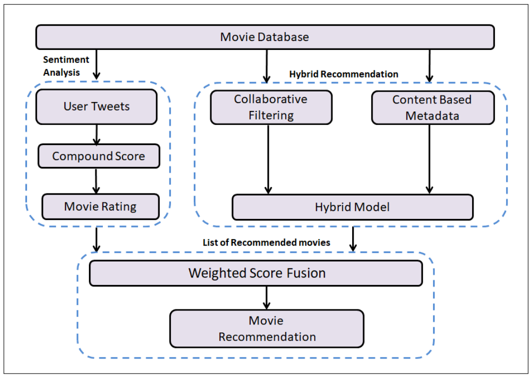

Fig. 1. Proposed movie recommendation framework.

the results showed 91% user satisfaction rating. Li *et al.* [29] 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. The polarity score of the post was assigned on a 1–5 rating scale. The result showed an enhanced RS by bridging the gap between user communication knowledge and social networking sites. Leung *et al.* [27] proposed a rating inference approach to transform textual reviews into ratings to enable easy integration of sentiment analysis and CF.

Our proposed model is a hybrid RS whose results are boosted using sentiment analysis score. Experimental evaluations, both quantitative and qualitative, demonstrate the validity and effectiveness of our method.

III. Proposed System

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 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]https://grouplens.org/datasets/movielens/100k/
[3]https://grouplens.org/datasets/movielens/20m/
[4]https://www.kaggle.com/orgesleka/imdbmovies
[5]https://www.kaggle.com/netflix-inc/netflix-prize-data/data

TABLE I
Details of the MovieTweetings Database

Metric	Value
Ratings	646410
Unique Users	51081
Unique Movies	29228
Start Year	1894
End Year	2017

TABLE II
Example of a Movie Entry in the Modified Movie Tweetings Database

Attribute	Value
MovieID	0451279
Title	Wonder Woman
Runtime	141 min
Genre	Action,Adventure,Fantasy
Director	Patty Jenkins
Writer	Allan Heinberg
Actors	Gal Gadot',Chris Pine
Rating	7.6 Massachusetts Institute of Technology in 1996.
Production Companies	DC Films,Tencent Pictures
Popularity	524.772
Language	en
Production Countries	United States of America
Budget	816303142

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

$$\text{release\_year}_{\text{movies}} \geq 2014. \tag{1}$$

The subset of the database consisted of 292 863 ratings by 51 081 users on 6209 different movies. The MovieTweetings database has three different components. The first component contains the mapping of users with their Twitter IDs. The second component contains the ratings of movies by users and their respective genres. The final component contains the information about the movies that were rated. In the proposed model, the socially filtered data, as well as the similarity of movies based on their attributes, has been used. The database had limited numbers of attributes for each movie, and thus, the Movie Database (TMDb) API was used to get more attributes of all the movies. TMDb[6] is a premier source for extensive metadata for movies that have more than 30 languages. The movie attributes of the modified MovieTweetings database are shown in Table II.

The modified database also contains some obscure movies from different countries and languages. The metadata for such movies was not available in TMDb, and therefore, those movies were discarded from the database. The final database had approximately 5000 movies.

[6]https://www.themoviedb.org/about?language=en

TABLE III
EXAMPLES OF NOISY AND UNINFORMATIVE PARTS IN THE TWEETS

Types of noise	Example
Stop words	be, of, the, being, am
Stemmer	friend, friendship and friends
Web links	www.imdb.com
Filtering of repeating words	happyyyy, heloooo
Special Characters	!, @, #, $, %, and _

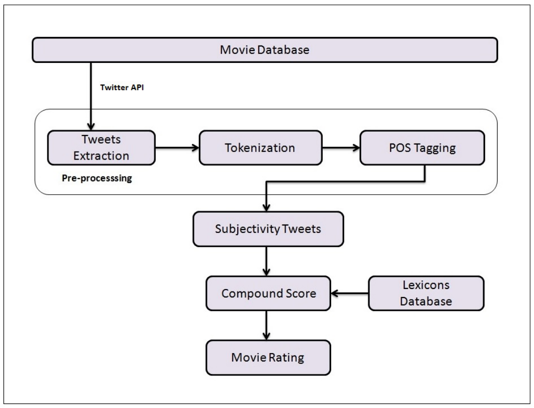

Fig. 2. Representative framework based on the VADER sentiment analysis system.

B. Analysis of User Tweets

As shown in Fig. 2, Twitter API[7] was used to fetch the tweets for the movies that were present in the MovieTweetings database. The extracted tweets consisted of tremendous amounts of noise, such as hashtags, emojis, repetitive words, and other irrelevant data that were removed using preprocessing techniques.

1) Preprocessing of Tweets: There are many short forms of words in the tweets, which converted into its original forms through gingerit[8] library. To filter unusable data and uninformative parts in tweets such as stop words, punctuations, weblinks, and repetitive words, which did not add much value to sentiment analysis, as shown in Table III. After preprocessing, the text extracted from the tweets was used for sentiment analysis.

2) Sentiment Analysis of User Tweets: VADER is a lexicon and rule-based method that is used to find the opinion expressed by the users in the form of tweets. It maps the words to sentiment by looking up the intensity of a word in the lexicon. This method produces four sentiment components for each tweet. The first three components are positive, negative, and neutral. The last component is the normalization of all the abovementioned three components of the tweet. The sum of the first three components is always 1. Compound score lies between -1 to $+1$ where -1 represents extreme negative and $+1$ denotes extreme positive sentiment rating of the movie. For calculating the rating of the movie, the compound score is scaled in the range of 1–10 using (2), where x is a compound score

$$\text{Rating} = [1 + (1 + x) \times 2] \times 2. \tag{2}$$

[7]https://developer.twitter.com/en/docs
[8]https://gingerit.readthedocs.org

VADER performance is better than the other methods, as shown in Table V.

C. 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 $f = \{f_1, f_2, \ldots, f_n\}$ and $q = \{q_1, q_2, \ldots, q_n\}$ are the content-based feature vectors and weight vectors, respectively. We construct the closeness C of two items i and j as:

$$C(i,j) = \begin{cases} \sum_{n=1}^{N} f_n(A_{n_i}, A_{n_j}), & \text{for } i \neq j \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $f_n(A_{n_i}, A_{n_j})$ corresponds to the similarity between feature values A_{n_i} and A_{n_j} corresponding to two items. In (3), the closeness of the items is determined using the metadata or the relevant information related to the items. F_{ij} is constructed by combining the closeness vector C for all the items and multiplying it with the weight vectors q. F_{ij} is a feature matrix of dimension $n \times (M(M-1)/2)$, 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 $U = \{u_1, u_2, \ldots, u_n\}$, where u_i is a user in the database. A user-item matrix is constructed for M items. An important property of the user-item matrix is that it has very high sparsity. Typical collaborative filtering [49] 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, K neighboring users. Neighboring users are recognized by similarity measures, such as cosine similarity and Pearson correlation. After selecting K neighboring users, the weighted aggregation of the ratings is as follows:

$$\text{rating}(u,i) = \frac{1}{K} \sum_{k \in K} \text{similarity}(\text{user}_u, \text{user}_{\upsilon_k}) \cdot \text{rating}_{\upsilon_{ki}} \tag{4}$$

where u and υ_k are target user and K nearest neighbors, respectively. The procedure of CF is used to overcome the sparsity of the user-item matrix instead of directly using it to predict ratings. We employ the tweaked user-item matrix to construct a social graph using items as nodes. This graph represents the user's perception of similarity between the items. The determination of feature weights complies with the social graph.

To determine the optimal feature weights q, we formulate a framework as described in the following equaion:

$$S(i,j) = q \cdot F_{ij} \tag{5}$$

which can be expanded as

$$S(i,j) = q_1 \cdot f_1(A_{1i}, A_{1j}) + q_2 \cdot f_2(A_{2i}, A_{2j}) + \cdots + q_n \cdot f_n(A_{ni}, A_{nj}). \tag{6}$$

The procedure for determining the weights for the feature vectors used for calculating the similarity scores between two items have been constructed as a linear system, $S(i,j)$. Here, $S(i,j)$ are the number of users who are interested in both items i and j. F_{ij} denotes the feature vectors, which is constructed keeping in mind the similarity in metadata between two items.

The similarity score using metadata of both items is calculated as described in (3). Therefore, the weight q here signifies the importance of a particular metadata when it is compared with the metadata of another movie. For example, the weight of the title of the movie will have more importance than the weight of the costume designer in determining the similarity between "The Dark Knight" and "The Dark Knight Rises." After having the weight matrix for the content-based metadata, we can calculate the similarity between an unknown movie A and a movie B, by using the weights present for B in the weight matrix computed from the user social graph.

For the entire database, S is a matrix of dimension $1 \times (M(M-1)/2)$ and q is a matrix of dimension $1 \times n$, where n is the number of content-based features and dimenstionality of F is $n \times (M(M-1)/2)$. We calculate the weight vectors q for all the metadata feature attributes for all the items using the Moore–Penrose pseudoinverse as in the following equation:

$$q = S^{-1} \cdot F. \tag{7}$$

D. Weighted Score Fusion

To make the system robust, we use two data sources: one from the hybrid RS and another is from sentiment analysis. The hybrid RSs gives us the similarity between two movies based on their metadata (e.g., Actor, Director, Release Year, and Producer). The weights of these metadata for computing the similarity is computed under a linear system framework, as described in Section III-C The weights q signify the importance of particular metadata when a movie is compared with another (e.g., the genre of a movie has more importance than the runtime of the movie) movie. These precomputed weights from the collaborative social graph are used for computing the similarity with another new item for which we just have metadata information but no social user rating data. 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 M movies. Let $S \in \{s_1, s_2, \ldots, s_n\}$, where s_i is the rating of movie i calculated using (2). For calculating the sentiment similarity, a function $G(i, j)$ for two movies i and j is defined based on their sentiment ratings s_i and s_j as mentioned in (8) to determine how close are the movies in terms of the polarity of the user

$$G(i, j) = D - |s_i - s_j| \tag{8}$$

where D is a constant. The constant D in (8) is taken as 10 because the ratings are on a scale of 1–10. Another function $H(i, j)$ defined as

$$H(i, j) = q \cdot f_{ij} \tag{9}$$

where f_{ij} is the feature similarity between movies i and j and q is the set of optimal weights as determined by (7). The final combined similarity $CS(i, j)$ is described in (10). It is a weighted combination of the defined functions G and H

$$CS(i, j) = \omega_1 \cdot H(i, j) + \omega_2 \cdot G(i, j) \tag{10}$$

$$\omega_1 + \omega_2 = 1, \quad \omega_1, \omega_2 \in [0, 1] \tag{11}$$

TABLE IV
CORRELATION MEASURES BETWEEN SENTIMENT AND MOVIE RATINGS

Correlation coefficient	Definition	Value
PLCC	$\frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}}$	0.76
SROCC	$1 - \frac{6}{N(N^2-1)}\sum_{i=1}^{N} d_i^2$	0.72
KRCC	$\frac{2(N_c-N_d)}{N(N-1)}$	0.51

where ω_1 corresponds to the weight of the similarity score calculated from the hybrid model and ω_2 corresponds to the weight of the sentiment similarity score. For a new movie item, we calculate this weighted similarity with all the movies present in the social graph for which we have the user rating data and then sort them by the computed similarity rating in descending order.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, quantitative, qualitative, and correlation coefficient results are discussed.

A. Correlation Between Sentiment and IMDb Movie Ratings

We conducted the statistical analysis between sentiment ratings X and movie rating Y to find the correlation coefficient. The correlation coefficient value varies from -1 to $+1$. Let D denotes a database of movies and N denote the number of total movies in the database. The statistical correlation coefficients are as follows: Spearman rank-order correlation coefficient (SROCC), Kendall rank correlation coefficient (KRCC), and Pearson linear correlation coefficient (PLCC). Table IV displays the values of different corelation coefficients utilized by us. In our experiments, we have found that sentiment and movie ratings are positively correlated. For PLCC, x_i and y_i are sentiment rating and IMDb movie rating, respectively, for the ith movie, whereas $\bar{x}$ denotes the mean sentiment score and $\bar{y}$ denotes the mean movie rating in the database. For SROCC, d_i is the difference between the sentiment rating and movie rating of the ith movie in the database. For KRCC, N_c and N_d represent the number of concordant and discordant pairs in the database, respectively.

B. Evaluation Metric

In many real-world applications, relevant recommendations are suggested by the system, instead of directly predicting rating values. This is known as Top-N recommendation [10], [47] and suggests specific items to users that are likable. The direct alternative methodologies are used for evaluation metric (e.g., precision). Precision is defined in terms of movies that are relevant (L_{rel}) and recommended (L_{rec}) by the model. In the proposed system, Precision@N is defined as follows:

$$\text{Precision@N} = \frac{L_{\text{rel}} \cap L_{\text{rec}}}{L_{\text{rec}}}. \tag{12}$$

For the proposed model, the choice of weights in the fusion in (10) is determined by evaluating the Precision@5 and Precision@10 for a different combination of weights ω_1 and ω_2 conforming with (11).

TABLE V

COMPARATIVE ANALYSIS OF RATING AMONG VADER, NAIVEBAYES, TEXTBLOB, PRETRAINED WORLD EMBEDDING, AND ATTENTION MODEL ON MOVIE LISTS

S.No.	Movie Lists	Vader rating	Naivebayes rating	Textblob rating	PTWE rating	Attention model rating	IMDB rating
1	Baby Driver	7.61	7.37	7.25	6.35	6.0	7.6
2	Snowden	7.72	6.06	7.45	5.86	5.0	7.3
3	Arrival	8.45	7.7	7.68	7.87	8.0	7.9
4	A Dog Purpose	7.54	6.37	7.38	6.67	6.0	7.0
5	Alien Covenant	6.77	6.19	6.03	3.45	7.0	6.4
6	Captain America	6.04	7.29	7.15	4.67	6.0	7.8
7	Storks	7.54	6.39	7.24	6.39	7.24	6.8
8	Mother	6.63	6.15	6.32	6.09	6.0	6.6
9	Neerja	8.21	7.48	7.38	7.08	8.0	7.7
10	The Legend of Tarzan	6.53	6.41	6.6	6.14	7.0	6.3

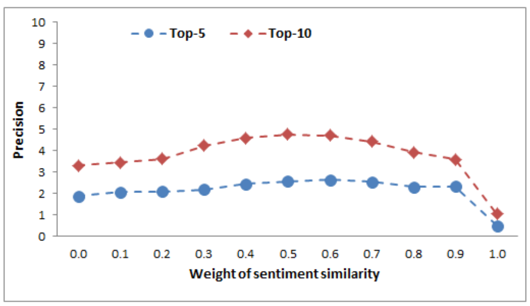

Fig. 3. Precision of Top-5 and Top-10 movies with varying sentiment similarity weights.

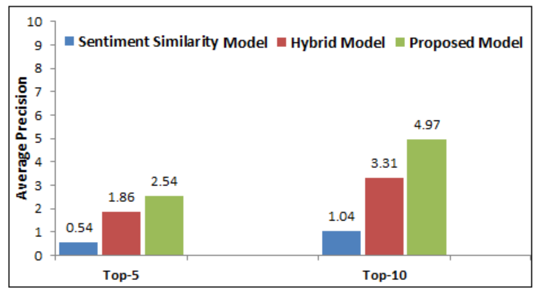

Fig. 4. Comparison of the proposed model with baselines models.

C. Weight Selection for Weighted Fusion

For every movie, the Top-N recommendation list is evaluated using (10). The choice of the weights ω_1 and ω_2 in (10) is decided by experiments conducted on the metric mentioned in Section IV-B. The Precision@N is evaluated as in (12). The recommendations of all movies are collected from public databases, such as IMDb and TMDb. These recommended movies are considered as the ground truth. We compare the results of the Precision@5 and Precision@10 for different values of ω_1 and ω_2. We choose the values of ω_1 and ω_2 for which the precision values are the best.

From Fig. 3, an observation can be made that the maximum precision for weight values is between 0.5 and 0.6. Hence, ω_1 and ω_2 values are selected as 0.5 in the proposed system.

D. Comparative Analysis

In this section, we present a comparative analysis of our proposed system with the pure hybrid model (PH Model) and sentiment similarity models (SS Models). The PH Model is a combination of CBF and CF. The recommended movies are based on the similarity of attributes, such as genre, director, and cast. The similarities are evaluated using weights obtained by a social graph, as described in Section III-C. SS Model recommends movies based solely on the similarity of the movie tweets of the corresponding tuple of movies. We evaluate our proposed method using Precision@5 and Precision@10. Fig. 4 shows the quantitative comparative results of our proposed system with the baseline models. For Precision@5, the average precision values of the SS Model and PH Model are 0.54 and 1.86, respectively. Similarly, for Precision@10, the average precision values of SS Models and PH Model are 1.04 and 3.31, respectively. Our proposed model achieves a better precision value in both cases with 2.54 for Top-5 and 4.97 for Top-10 in comparison with the PH and SS Models. Thus, we can infer that our method will suggest at least two recommended movies out of five and five recommended movies out of ten.

In addition, we have studied the FFM algorithm [20], [38] that uses personalized tweet recommendations for comprehending quick and accurate access to the desired information in the area of effective advertisements or election campaigns. This model is primarily effective when a fine-grained analysis is needed on the user's tweet along with its retweet to analyze multiple factors in a tweet, i.e., publisher, topic, and sentiment factors. Since this article is to propose a movie RS using approaches, such as hybrid, CF, and CBF against the modified MovieTweeting database and sentiment analysis on the user's tweets, respectively, therefore, the FFM algorithm is not suitable in this article.

1) Comparison With Pretrained Word Embedding and Attention Mechanism: Deep learning-based models are mostly used in the natural language processing and vision domain. The pretrained word embedding (e.g., GloVe algorithm) and attention mechanism models are used to compare our proposed system. Both models have used the IMDb database for training purposes. We have used the GloVe algorithm to initialize the pretrained vectors. Bidirectional LSTM, adam optimizer, and dropout layer are the parameters used to train this model.

TABLE VI

QUALITATIVE ANALYSIS OF WONDER WOMAN HOLLYWOOD MOVIE: LANGUAGE (ENGLISH). MOVIES IN BOLD ARE INTERSECTING WITH EITHER IMDB OR TMDB

IMDb	**TMDb**	**Recommendations from the proposed system**
Justice League	**Guardians of the Galaxy Vol. 2**	**Batman v Superman: Dawn of Justice**
Batman v Superman: Dawn of Justice	Spider-Man: Homecoming	**Suicide Squad**
Suicide Squad	Logan	**Thor: Ragnarok**
Thor: Ragnarok	**Thor: Ragnarok**	**Justice League**
Spider-Man: Homecoming	**Justice League**	Warcraft
Deadpool	Pirates of the Caribbean: Dead Men Tell No Tales	**Doctor Strange**
Logan	**Doctor Strange**	**Guardians of the Galaxy Vol. 2**
Captain America: Civil War	Baby Driver	**Kong: Skull Island**
Doctor Strange	**Kong: Skull Island**	The LEGO Batman Movie
Guardians of the Galaxy Vol. 2	Life	Batman and Harley Quinn

TABLE VII

QUALITATIVE ANALYSIS OF NEERJA BOLLYWOOD MOVIE: LANGUAGE (HINDI). MOVIES IN BOLD ARE INTERSECTING WITH EITHER IMDB OR TMDB

IMDb	**TMDb**	**Recommendations from the proposed system**
Airlift	**Airlift**	**Simran**
Pink	Pink	**Fan**
Kapoor & Sons	Rustom	**Raabta**
Udta Punjab	Ghayal Once Again	**Udta Punjab**
Drishyam	Mary Kom	**Rocky Handsome**
Rustom	**Udta Punjab**	**Rangoon**
M.S. Dhoni: The Untold Story	**Force 2**	**Raabta**
Raabta	**Fan**	**Force 2**
Dear Zindagi	**Rocky Handsome**	Te3n
Rangoon	**Simran**	**Airlift**

After training this model, our database is used to calculate the polarity score that is eventually converted into a rating using (2). As shown in Table V, the movie's rating is the average rating of the movie's tweets. The pretrained rating results are inferior to Vader rating due to ignorance of the tweet's context [4], [17], [44], [52].

Attention models are used bidirectional LSTM with attention layer [13], [58], [60], [61]. The model has been trained on 10 epochs with 32 batch size. The result shows the inferior performance of this model than the performance of the VADER method. Deep learning requires a huge amount of relevant data to give an accurate result. In this article, the performance is inferior due to not having a large amount of data.

E. Qualitative Analysis

In this section, we show the qualitative results for some of the movies recommended by the proposed system. The results also include movies from both Hollywood as well as Bollywood, as shown in Tables VI and VII, respectively. It is interpreted from these tables that the recommendations from the proposed system have many intersecting movies, with the recommendations from both IMDb and TMDb.

V. CONCLUSION AND FUTURE WORKS

RSs are an important medium of information filtering systems in the modern age, where the 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 is respond 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] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Analyzing user modeling on Twitter for personalized news recommendations," in *Proc. 19th Int. Conf. Modeling, Adaption, Pers. (UMAP)*. Berlin, Germany: Springer-Verlag, 2011, pp. 1–12.

[2] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Twitter-based user modeling for news recommendations," in *Proc. Int. Joint Conf. Artif. Intell.*, vol. 13, 2013, pp. 2962–2966.

[3] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005.

[4] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," *Expert Syst. Appl.*, vol. 77, pp. 236–246, Jul. 2017.

[5] E. Aslanian, M. Radmanesh, and M. Jalili, "Hybrid recommender systems based on content feature relationship," *IEEE Trans. Ind. Informat.*, early access, Nov. 21, 2016, doi: 10.1109/TII.2016.2631138.

[6] J. Bobadilla, F. Ortega, A. Hernando, and J. Alcalá, "Improving collaborative filtering recommender system results and performance using genetic algorithms," *Knowl.-Based Syst.*, vol. 24, no. 8, pp. 1310–1316, Dec. 2011.

[7] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Model. User-Adapted Interact.*, vol. 12, no. 4, pp. 331–370, 2002.

[8] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, Mar./Apr. 2016.

[9] I. Cantador, A. Bellogín, and D. Vallet, "Content-based recommendation in social tagging systems," in *Proc. 4th ACM Conf. Rec. Syst. (RecSys)*, 2010, pp. 237–240.

[10] P. Cremonesi, Y. Koren, and R. Turrin, "Performance of recommender algorithms on top-N recommendation tasks," in *Proc. 4th ACM Conf. Rec. Syst. (RecSys)*, 2010, pp. 39–46.

[11] J. Dong, X. Li, C. Xu, G. Yang, and X. Wang, "Feature re-learning with data augmentation for content-based video recommendation," in *Proc. ACM Multimedia Conf. (MM)*, 2018, pp. 2058–2062.

[12] S. Dooms, T. D. Pessemier, and L. Martens, "MovieTweetings: A movie rating dataset collected from Twitter," in *Proc. Workshop Crowdsourcing Hum. Comput. Rec. Syst.*, 2013, pp. 1–2.

[13] C. Du and L. Huang, "Text classification research with attention-based recurrent neural networks," *Int. J. Comput. Commun. Control*, vol. 13, no. 1, pp. 50–61, 2018.

[14] X. Du, H. Yin, L. Chen, Y. Wang, Y. Yang, and X. Zhou, "Personalized video recommendation using rich contents from videos," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 3, pp. 492–505, Mar. 2020.

[15] A. Flexer and J. Stevens, "Mutual proximity graphs for improved reachability in music recommendation," *J. New Music Res.*, vol. 47, no. 1, pp. 17–28, Jan. 2018.

[16] X. Fu and Y. Shen, "Study of collective user behaviour in Twitter: A fuzzy approach," *Neural Comput. Appl.*, vol. 25, nos. 7–8, pp. 1603–1614, Dec. 2014.

[17] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," *Expert Syst. Appl.*, vol. 69, pp. 214–224, Mar. 2017.

[18] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Commun. ACM*, vol. 35, no. 12, pp. 61–70, Dec. 1992.

[19] F. Goossen, W. IJntema, F. Frasincar, F. Hogenboom, and U. Kaymak, "News personalization using the CF-IDF semantic recommender," in *Proc. Int. Conf. Web Intell., Mining Semantics (WIMS)*, 2011, pp. 10:1–10:12.

[20] R. Harakawa, D. Takehara, T. Ogawa, and M. Haseyama, "Sentiment-aware personalized tweet recommendation through multimodal FFM," *Multimedia Tools Appl.*, vol. 77, no. 14, pp. 18741–18759, Jul. 2018.

[21] C.-C. Hsu, H.-C. Chen, K.-K. Huang, and Y.-M. Huang, "A personalized auxiliary material recommendation system based on learning style on Facebook applying an artificial bee colony algorithm," *Comput. Math. Appl.*, vol. 64, no. 5, pp. 1506–1513, Sep. 2012.

[22] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. 8th Int. Conf. Weblogs Social Media*, 2014, pp. 216–225.

[23] L. Jin, D. Yuan, and H. Zhang, "Music recommendation based on embedding model with user preference and context," in *Proc. IEEE 2nd Int. Conf. Big Data Anal. (ICBDA)*, Mar. 2017, pp. 688–692.

[24] R. Katarya, "Movie recommender system with Metaheuristic artificial bee," *Neural Comput. Appl.*, vol. 30, no. 6, pp. 1983–1990, Sep. 2018.

[25] R. Katarya and O. P. Verma, "An effective collaborative movie recommender system with cuckoo search," *Egyptian Informat. J.*, vol. 18, no. 2, pp. 105–112, Jul. 2017.

[26] R. Katarya and O. P. Verma, "Recommender system with grey wolf optimizer and FCM," *Neural Comput. Appl.*, vol. 30, no. 5, pp. 1679–1687, Sep. 2018.

[27] C. W. Leung, S. C. Chan, and F.-L. Chung, "Integrating collaborative filtering and sentiment analysis: A rating inference approach," in *Proc. ECAI Workshop Rec. Syst.*, 2006, pp. 62–66.

[28] F. Li, G. Xu, and L. Cao, "Two-level matrix factorization for recommender systems," *Neural Comput. Appl.*, vol. 27, no. 8, pp. 2267–2278, Nov. 2016.

[29] H. Li, J. Cui, B. Shen, and J. Ma, "An intelligent movie recommendation system through group-level sentiment analysis in microblogs," *Neurocomputing*, vol. 210, pp. 164–173, Oct. 2016.

[30] P. Lops, M. de Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. Kantor, Eds. Boston, MA, USA: Springer, 2011, pp. 73–105.

[31] J. Ma, G. Li, M. Zhong, X. Zhao, L. Zhu, and X. Li, "LGA: Latent genre aware micro-video recommendation on social media," *Multimedia Tools Appl.*, vol. 77, no. 3, pp. 2991–3008, Feb. 2018.

[32] Y. Matsumoto, R. Harakawa, T. Ogawa, and M. Haseyama, "Music video recommendation based on link prediction considering local and global structures of a network," *IEEE Access*, vol. 7, pp. 104155–104167, 2019.

[33] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014.

[34] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-boosted collaborative filtering for improved recommendations," in *Proc. AAAI*, vol. 23, 2002, pp. 187–192.

[35] C. Nascimento, A. H. F. Laender, A. S. da Silva, and M. A. Gonçalves, "A source independent framework for research paper recommendation," in *Proc. 11th Annu. Int. ACM/IEEE Joint Conf. Digit. Libraries (JCDL)*, 2011, pp. 297–306.

[36] J. Niu, S. Wang, Y. Su, and S. Guo, "Temporal factor-aware video affective analysis and recommendation for cyber-based social media," *IEEE Trans. Emerg. Topics Comput.*, vol. 5, no. 3, pp. 412–424, Jul. 2017.

[37] J. M. Noguera, M. J. Barranco, R. J. Segura, and L. Martínez, "A mobile 3D-GIS hybrid recommender system for tourism," *Inf. Sci.*, vol. 215, pp. 37–52, Dec. 2012.

[38] J. Pan *et al.*, "Field-weighted factorization machines for click-through rate prediction in display advertising," in *Proc. World Wide Web Conf. (WWW)*, 2018, pp. 1349–1357.

[39] S. Philip, P. B. Shola, and A. Ovye, "Application of content-based approach in research paper recommendation system for a digital library," *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, no. 10, pp. 37–40, 2014.

[40] C. Porcel, A. Tejeda-Lorente, M. A. Martínez, and E. Herrera-Viedma, "A hybrid recommender system for the selective dissemination of research resources in a technology transfer office," *Inf. Sci.*, vol. 184, no. 1, pp. 1–19, Feb. 2012.

[41] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach," *J. Informetrics*, vol. 3, no. 2, pp. 143–157, Apr. 2009.

[42] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.

[43] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An open architecture for collaborative filtering of netnews," in *Proc. Conf. Comput. Supported Cooperat. Work*, 1994, pp. 175–186.

[44] S. M. Rezaeinia, R. Rahmani, A. Ghodsi, and H. Veisi, "Sentiment analysis based on improved pre-trained word embeddings," *Expert Syst. Appl.*, vol. 117, pp. 139–147, Mar. 2019.

[45] A. Romadhony, S. Al Faraby, and B. Pudjoatmodjo, "Online shopping recommender system using hybrid method," in *Proc. Int. Conf. Inf. Commun. Technol. (ICoICT)*, Mar. 2013, pp. 166–169.

[46] R. L. Rosa, D. Z. Rodriguez, and G. Bressan, "Music recommendation system based on user's sentiments extracted from social networks," *IEEE Trans. Consum. Electron.*, vol. 61, no. 3, pp. 359–367, Aug. 2015.

[47] A. Said, A. Bellogín, and A. D. Vries, "A top-N recommender system evaluation protocol inspired by deployed systems," in *Proc. LSRS Workshop ACM RecSys*, 2013, pp. 1–7.

[48] V. Sembium, R. Rastogi, A. Saroop, and S. Merugu, "Recommending product sizes to customers," in *Proc. 11th ACM Conf. Rec. Syst.*, Aug. 2017, pp. 243–250.

[49] Y. Shi, M. Larson, and A. Hanjalic, "Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges," *ACM Comput. Surv.*, vol. 47, no. 1, pp. 3:1–3:45, Jul. 2014.

[50] S. Spiegel, "A hybrid approach to recommender systems based on matrix factorization," Dept. Agent Technol. Telecommun., Tech. Univ. Berlin, Berlin, Germany, 2009.

[51] L. Sun, X. Wang, Z. Wang, H. Zhao, and W. Zhu, "Social-aware video recommendation for online social groups," *IEEE Trans. Multimedia*, vol. 19, no. 3, pp. 609–618, Mar. 2017.

[52] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for Twitter sentiment classification," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2014, pp. 1555–1565.

[53] S. Ujjin and P. J. Bentley, "Particle swarm optimization recommender system," in *Proc. IEEE Swarm Intell. Symp. (SIS)*, 2003, pp. 124–131.

[54] R. V. Meteren and M. V. Someren, "Using content-based filtering for recommendation," in *Proc. Mach. Learn. New Inf. Age, Workshop*, 2000, pp. 47–56.

[55] T. Viard and R. Fournier-S'niehotta, "Movie rating prediction using content-based and link stream features," 2018, *arXiv:1805.02893*. [Online]. Available: http://arxiv.org/abs/1805.02893

[56] X. Wan and T. Okamoto, "Utilizing learning process to improve recommender system for group learning support," *Neural Comput. Appl.*, vol. 20, no. 5, pp. 611–621, Jul. 2011.

[57] M. Wasid and V. Kant, "A particle swarm approach to collaborative filtering based recommender systems through fuzzy features," *Procedia Comput. Sci.*, vol. 54, pp. 440–448, Jan. 2015.

[58] Z. Wu, X.-Y. Dai, C. Yin, S. Huang, and J. Chen, "Improving review representations with user attention and product attention for sentiment classification," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 5989–5996.

[59] C. Yang, B. Wei, J. Wu, Y. Zhang, and L. Zhang, "CARES: A ranking-oriented CADAL recommender system," in *Proc. 9th Joint Conf. Digit. Libraries*, 2009, pp. 203–212.

[60] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2016, pp. 1480–1489.

[61] A. Zadeh, P. P. Liang, S. Poria, P. Vij, E. Cambria, and L.-P. Morency, "Multi-attention recurrent network for human communication comprehension," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 5642–5649.

[62] H.-R. Zhang, F. Min, X. He, and Y.-Y. Xu, "A hybrid recommender system based on user-recommender interaction," *Math. Problems Eng.*, vol. 2015, pp. 1–11, Feb. 2015.

[63] Z. Zhao *et al.*, "Social-aware movie recommendation via multimodal network learning," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 430–440, Feb. 2018.

[64] Y. Zhou, L. Zhang, and Z. Yi, "Predicting movie box-office revenues using deep neural networks," *Neural Comput. Appl.*, vol. 31, no. 6, pp. 1855–1865, Jun. 2019.

Sudhanshu Kumar received the bachelor's degree from Uttar Pradesh Technical University (UPTU), Lucknow, India, in 2006, and the master's degree in computer science from Teerthanker Mahaveer University (TMU), Moradabad, India, in 2012. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, IIT Roorkee, Roorkee, India.

His research interests include sentiment analysis and machine learning.

Kanjar De received the bachelor's degree from the Sikkim Manipal Institute of Technology (SMIT), Sikkim, India, in 2007, and the master's degree in information technology from the International Institute of Information Technology Bangalore (IIITB), Bengaluru, India, in 2009, and the Ph.D. degree from the Department of Computer Science and Engineering, Indian Institute of Information Technology Design and Manufacturing (IIITD&M), Kancheepuram, India, in 2017.

After that, he was a Post-Doctoral Scholar with IIT Roorkee, Roorkee, India, from 2017 to 2019 and the Norwegian University of Science and Technology, Trondheim, Norway. His research interests include the applications of machine learning in computer vision, natural language, and other domains.

Partha Pratim Roy received the master's and Ph.D. degrees from the Universitat Autonoma de Barcelona, Barcelona, Spain, in 2006 and 2010, respectively.

He has gathered industrial experience while working in TCS, Kolkata, India, and Samsung Noida, India. In Samsung, he was a part-leader of the Computer Vision Group. He and his team developed several prototypes for smart-phone users. His work on automatic text transcription from scene image was highly acclaimed. He is working as an Associate Professor with the Department of Computer Science and Engineering, IIT Roorkee, Roorkee, India. He has published more than 180 research papers in various international journals and conference proceedings. His main research area is pattern recognition.

Dr. Roy has coorganized several international conferences and workshops, has been a member of the Program Committee for a number of international conferences, and acts as a reviewer for many journals in the field.