

## Article

# An approach to integrating Sentiment Analysis into Recommender Systems

Cach N. Dang <sup>1, 2, 3, \*</sup>, María N. Moreno-García <sup>2</sup> and Fernando De la Prieta <sup>3</sup>

<sup>1</sup> Department of Information Technology, HoChiMinh City University of Transport (UT-HCMC), Ho Chi Minh 70000, Vietnam; cach@ut.edu.vn

<sup>2</sup> Data Mining (MIDA) Research Group, University of Salamanca, Salamanca 37007, Spain; mmg@usal.es

<sup>3</sup> Biotechnology, Intelligent Systems and Educational Technology (BISITE) Research Group, University of Salamanca, Salamanca 37007, Spain; fer@usal.es

\* Correspondence: cach@ut.edu.vn; Tel.: +84-919-101-086

**Abstract:** Recommender systems have been applied in a wide range of domains such as e-commerce, media, banking, and utilities. This kind of system provides personalized suggestions based on large amounts of data in order to increase user satisfaction. These suggestions help client select products, while organizations can increase the consumption of a product. In the case of social data, sentiment analysis can help gain better understanding of a user's attitudes, opinions and emotions, which is beneficial to integrate in recommender systems for achieving higher recommendation reliability. On the one hand, this information can be used to complement explicit ratings given to products by users. On the other hand, sentiment analysis of items that can be derived from online news services, blogs, social media or even from the recommender systems themselves is seen as capable of providing better recommendations to users. In this study, we present and evaluate a recommendation approach that integrates sentiment analysis into collaborative filtering methods. The recommender system proposal is based on an adaptive architecture, which includes improved techniques for feature extraction and deep learning models based on sentiment analysis. The results of the empirical study performed with two popular datasets show that sentiment-based deep learning models and collaborative filtering methods can significantly improve the recommender system's performance.

**Keywords:** sentiment analysis; deep learning; recommender system; natural language processing

## 1. Introduction

With the explosion of blogs, forums, and online social networks, differing opinions about a particular topic can be easily found from millions of users. For example, users can discuss their current experiences, share their points of view on a specific fact, or praise or complain about a product that they have just bought. With a vast amount of available online data, sentiment analysis (SA) – a method to categorize text-based opinions to determine a user's attitude – can help gain better understanding of the attitudes, opinions and emotions of the public in several domains such as business, government, and biomedicine. Several studies are summarized and discussed in [1] regarding the benefits of SA in obtaining feedbacks and determining the interests and opinions of customers.

Recommender systems, developed since mid-1990s based on users' ratings and preferences, has expanded widely in recent decades especially in e-commerce, media, banking and utilities. This type of system is used by Amazon to suggest preferred products for customers, by YouTube to suggest related videos on the auto play function, and by Facebook to recommend people and webpages to connect and follow.

Sentiment analysis can be beneficial to recommender systems. A sample of this can be found in the work of Preethi et al. [2], in which a cloud-based recommender system

uses recursive neural networks to analyze sentiments of reviews in order to improve and validate restaurant and movie recommendations. Along with behavioral analysis, sentiment analysis is also an efficient tool for commodity markets [3].

The problem of automatically extracting opinions from online user-generated texts – sentiment analysis – has been a growing research topic recently [1]. Social media data has been exploited in different ways to address some problems, especially associated with collaborative filtering approaches [4]. Methods in recommender systems are based on information filtering, and they can be classified into three categories: content-based; collaborative filtering (CF); and hybrid. Sparsity and gray-sheep problems are two of the main reasons CF methods do not provide the reliability required in some recommender systems [5]. In particular, when only sparse ratings data is available, sentiment analysis can play a key role in improving recommendation quality. This is because recommendation algorithms mostly rely on users' ratings to select the items to recommend. Such ratings are usually insufficient and very limited. On the other hand, sentiment-based ratings of items that can be derived from reviews or opinions given through online news services, blogs, social media or even the recommender systems themselves, are seen as capable of providing better recommendations to users. Sentiment-based models have been exploited in recommender systems to overcome the data-sparsity problem that exists in conventional recommender systems. Hence, integrating sentiment in recommender systems may significantly enhance the recommendation quality.

In this study, we propose a recommendation method that combines sentiment analysis and collaborative filtering. The method is implemented in an adaptive recommender system architecture in which techniques for feature extraction and deep learning based sentiment analysis are included. The results of the empirical study performed with two popular datasets show that combining deep learning based sentiment analysis and collaborative filtering methods significantly improves the recommender system's performance.

The rest of this paper is organized as follows. Section 2 provides a literature review in this research area. Section 3 presents background information. Section 4 describes the methodology for recommender systems. Section 5 outlines the results and discussion, and Section 6 offers the main conclusion.

## 2. Literature Review

Sentiment analysis is very useful in a wide range of application domains, such as business, government, and biomedicine. Application of sentiment analysis in recommender systems has also been the focus of extensive research. In this section, we start by reviewing the literature and offering an up-to-date overview of how sentiment analysis has been applied in recommender systems.

Recommender systems can be improved in a variety of ways. In [4] social tag embedding is used in a collaborative filtering approach in which user similarities based on both tag embedding and ratings are combined to generate the recommendations. Recommender systems have also benefited from sentiment analysis. An example of this can be found in the work of Preethi et al. [2], where recursive neural networks were applied to analyze sentiments in reviews. The output was used to improve and validate restaurant and movie recommendations of a cloud-based recommender system. Along with behavioral analysis, sentiment analysis is also an efficient tool for commodity markets [3]. Wang et al. [6] combined a hybrid recommender system and sentiment analysis to optimize the preliminary list and obtain the final recommendation list. Kumar et al. [7] proposed a hybrid recommender system by combining collaborative filtering and content-based filtering with the use of sentiment analysis of movie tweets to boost up the recommender system.

Rao et al. [8] designed a recommender system that contains the user list and item list with user reviews. Using the sentiment dictionaries, the researchers divided the items into three categories: brand, quality, and price. They leveraged sentiment dictionaries to calculate sentiment of a particular user on item/product. Gurini et al. [9] adopted a different approach to describe a user recommender system for Twitter. Their work emphasized the

use of implicit sentiment analysis in order to improve the performance of the recommendation process. They defined a novel weighting function that takes into account sentiment, volume, and objectivity related to the users' interests.

In yet another approach, Osman et al. [10] presented an electronic product recommender system based on contextual information from sentiment analysis. Because ratings are usually insufficient and very limited, they constructed a contextual information sentiment model for a recommender system by making use of user comments and preferences. In a similar way, Contratre et al. [11] also proposed a recommender process that includes sentiment analysis of textual data extracted from Facebook and Twitter in order to increase conversion by matching product offers and consumer preferences. We can find similar combinations in other studies [12-14].

In addition, Rosa et al. [15] used a sentiment intensity metric to build a music recommender system. Users' sentiments are extracted from sentences posted on social networks and the recommendations are made using a framework of low complexity that suggests songs based on the current user's sentiment intensity. The research by Osman, Nurul Aida, and Shahrul [16] addressed the data-sparsity problem of recommender systems by integrating a sentiment-based analysis. Their work was applied to the Internet Movie Dataset (IMDb) and Movie Lens datasets, but improvements in sentiment analysis have been made since the paper was published. Rayan et al. [17] also tried to improve recommendations by addressing the data-sparsity problem. They proposed a smart recommender system based on methods of hybrid learning that integrate the most effective and efficient learning algorithms. These methods switch among content-based and collaborative filtering, identify the user context with the integration of dynamic filtering, and finally learn the profiles.

Several research teams [6, 7, 13, 18-20] introduced the techniques for applying sentiment analysis in recommender systems. The techniques that are applicable for performing the analysis of sentiments include SVM, CNN, RNN, and DNN.

Recommender systems rely on explicit user ratings, but this is not feasible in an increasing number of domains. Moreover, when explicit ratings are available, the trust and reliability of the ratings may limit the recommender system. When we have a large number of reviews and comments on these items, analyzing sentiments in that text to obtain implicit feedback in addition to traditional ratings for items, is useful and helps to improve the recommendations to users. The above studies use sentiment analysis in recommendation methods, but most studies have used traditional sentiment techniques or a sole deep learning model.

In this study, we will apply new feature extraction techniques and hybrid deep learning methods for sentiment analysis exploiting the advantages of BERT, in order to incorporate sentiments into recommendation methods as additional feedback and thus improve the performance and the reliability of recommender systems.

### 3. Background

#### 3.1. Sentiment Analysis

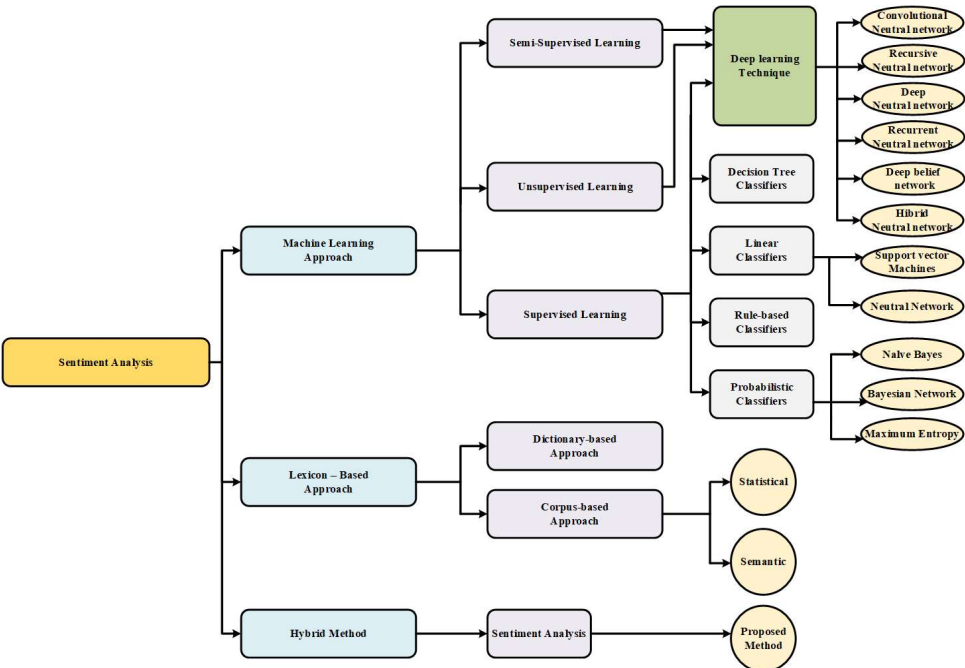
Sentiment analysis is a process of extracting information about an entity and automatically identifying any of the subjectivities of that entity. The objective is to determine whether text generated by users conveys their positive, negative, or neutral opinions. Sentiment classification can be carried out on three levels of extraction: the aspect or feature level; the sentence level; and the document level. Three approaches currently exist to address the problem of sentiment analysis[21]: (1) lexicon-based techniques; (2) machine-learning-based techniques; and (3) hybrid approaches.

**Lexicon-based** techniques were the first to be used for sentiment analysis. They are divided into two approaches: dictionary-based and corpus-based [22]. In the former, sentiment classification is performed by using a dictionary of terms, such as those found in SentiWordNet and WordNet. Nevertheless, corpus-based sentiment analysis does not rely on a predefined dictionary but on statistical analysis of the contents of a collection of

documents, using techniques based on k-nearest neighbors (k-NN) [23], conditional random field (CRF) [24], and hidden Markov models (HMM) [25], among others.

**Machine-learning-based** techniques [26] proposed for sentiment analysis problems can be divided into two groups: (1) traditional techniques and (2) deep learning. Traditional techniques refer to classical machine learning techniques, such as Naïve Bayes [27], maximum entropy [28, 29], or support vector machines (SVM) [30]. The input to those algorithms includes lexical features, sentiment lexicon-based features, parts of speech, or adjectives and adverbs. The accuracy of these systems depends on which features are chosen. Deep learning can provide better results than traditional techniques. Different kinds of deep learning models can be used for sentiment analysis, including CNN, DNN, and RNN. These models address classification problems at the document level, sentence level, or aspect level and will be discussed in the following section.

The **hybrid approaches** [31] combine lexicon-based and machine-learning-based approaches. Sentiment lexicons commonly play a key role within a majority of these strategies. Figure 1 illustrates a taxonomy of deep-learning-based methods for sentiment analysis.



**Figure 1.** Taxonomy of sentiment analysis techniques. Source: [21, 32].

In this study, we combine deep learning techniques for sentiment analysis. The resulting hybrid deep learning models for sentiment-analysis, which combine Long Short-Term Memory (LSTM) networks [33] and Convolutional Neural Networks (CNN) [34], are built and tested on two datasets containing reviews.

3.2. Recommender System

A recommender system intends to provide personalized recommendations about products or services to support decision making in the continuous increase of online information. Several systems have been developed and applied in three main domains: business, government, education, across eight categories: e-government, e-business, e-commerce/e-shopping, e-library, e-learning, e-tourism, e-resource services and e-group activities [35]. E-commerce has widely applied recommender systems to suggest additional products for customers to choose from among the multiple products available. A filtering technique has improved systems for presenting personalized choices [36].

The most common methods used for recommender systems may be grouped into three categories: content-based; collaborative filtering (CF); and hybrid recommender systems [37]. These techniques vary depending on the types of social media data that are used. Lu et al. [35] analyzed typical recommender systems and effectively identifies the specific requirements for recommendation techniques in the domain. This work also directly motivates and supports researchers and practitioners to promote the popularization and application of recommender systems in different domains.

**Content-based recommender systems:** Content-based methods make use of characteristics of items and users' profiles. User profiles are created by mining content information about items accessed over the Web by users, such as product attributes. Content-based recommender systems filter items based on the content-based similarity measures between items in the catalog and items that users have previously consumed, accessed, or rating positively. Therefore, a user receives recommendations of items similar to those that previously have been of interest. The utility of an item for a user can be a derivative done after a quantitative analysis of the metadata of the item.

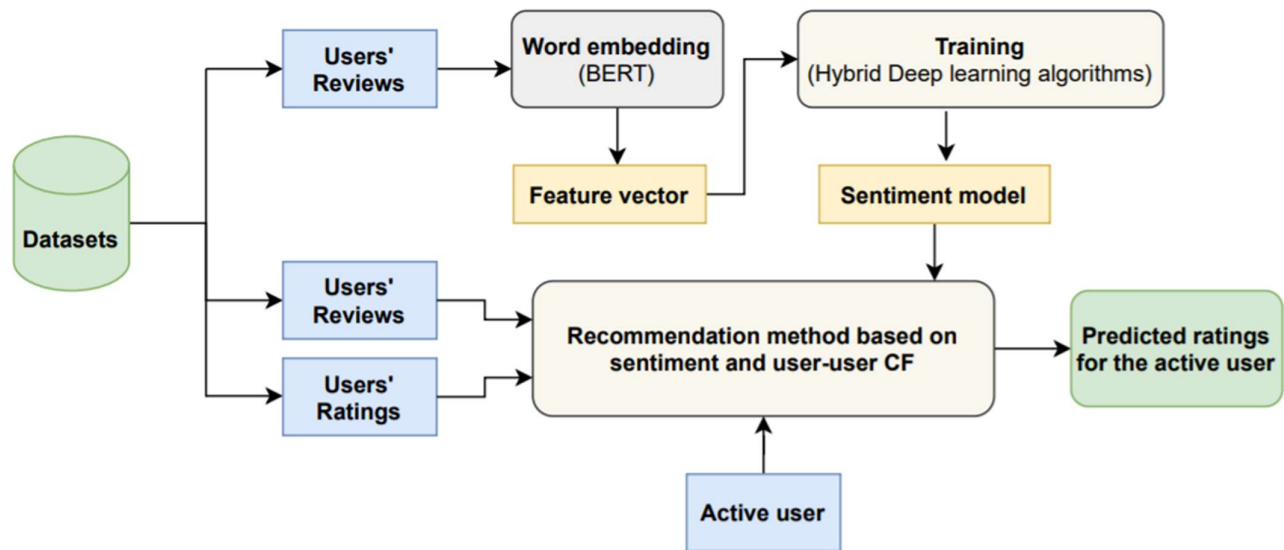
**Collaborative filtering-based recommender systems:** Collaborative filtering is a technique that can filter out items that a user might like based on reactions by similar users. It works by searching in a large group of people and finding a smaller set of users with tastes similar to those of a particular user. It looks at the items they like and combines them to create a ranked list of suggestions. We need data that contains a set of items and a set of users to perform with recommender algorithms. While working with such data, the matrix consists of the reactions given by a set of users to certain items from within a set of items. Each row would contain the ratings given by a user, and each column would contain the ratings received by an item.

**Hybrid recommender systems:** Hybrid approaches take advantage of any kind of item and user information that can be extracted or inferred from web systems, social media, or other sources. Hybrid approaches are implemented by deployment individually as well as by accumulating rankings and predictions and then building a general consolidative model that resolves the common problems in recommender systems.

Each recommendation approach has advantages and limitations; for example, Collaborative Filtering has sparseness, scalability and cold-start problems [5, 38, 39]. A sparseness problem occurs when we have a vast amount of data. A scalability problem occurs when rating data is missing and when user or an item is added to the system the cold-start problem appears. Combining sentiment analysis with recommendation methods can help solve these problems.

#### 4. Methodology

In this section, the proposed recommender system is presented. It is based on a recommendation method that combines collaborative filtering and sentiment analysis. The aim is to improve reliability of the recommendations to the user by combining sentiment analysis of reviews or comments of users with traditional recommendation methods. The architecture of this system is illustrated in Figure 2. The architecture makes it easy to configure the modules and their interactions, allowing application to be composed by choosing from supported techniques and methods.



**Figure 2.** An architecture application in a recommender system.

The architecture has two separate parts, one part in charge of generating the sentiment models and other part to provide recommendations to a given user making use of the models previously generated. The reviews' data were preprocessed and used to conduct and train sentiment-based hybrid deep learning models. Then, a user-based (user-user) collaborative filtering method is combined with sentiment-based models for rating prediction.

#### 4.1. Input data and preprocessing

Sentiment analysis requires that text training data is cleaned before being used to induce the classification model. Text cleaning is a preprocessing step that removes words or other components that lack relevant information, and thus may reduce the effectiveness of sentiment analysis. After cleaning, text data can be split into individual words, which are transformed into their base form by lemmatization, then converted into numerical vectors by using methods such as word embedding or TF-IDF.

In this research, we used BERT to transform text data to word embedding. Word embedding [40] is a type of word representation that maps each word into a vector of real values in such a way that words with similar meanings have a similar representation. Value learning can be done using neural networks. BERT is a language model for natural language processing, and it was published by researchers at Google AI Language in 2018 [41]. BERT was developed after Word2vec, and includes some advances over Word2vec, such as support for out-of-vocabulary (OOV) words.

#### 4.2. Conduct and train sentiment-based hybrid deep learning models

We used the combination of several successful approaches. We start by using a pre-trained BERT model to create the feature vectors. We then vary the order of the CNN and LSTM models used in the next stages: BERT -> CNN -> LSTM or BERT -> LSTM -> CNN. The final stage of the model uses a ReLu activation function.

We labeled the reviews with one value of an ordinal scale of five classes (very negative; negative; neutral; positive; and very positive), analogous to the explicit ratings, to train and validate the result of sentiment analysis. Figure 3 visualizes the process of the hybrid methodology for sentiment analysis.



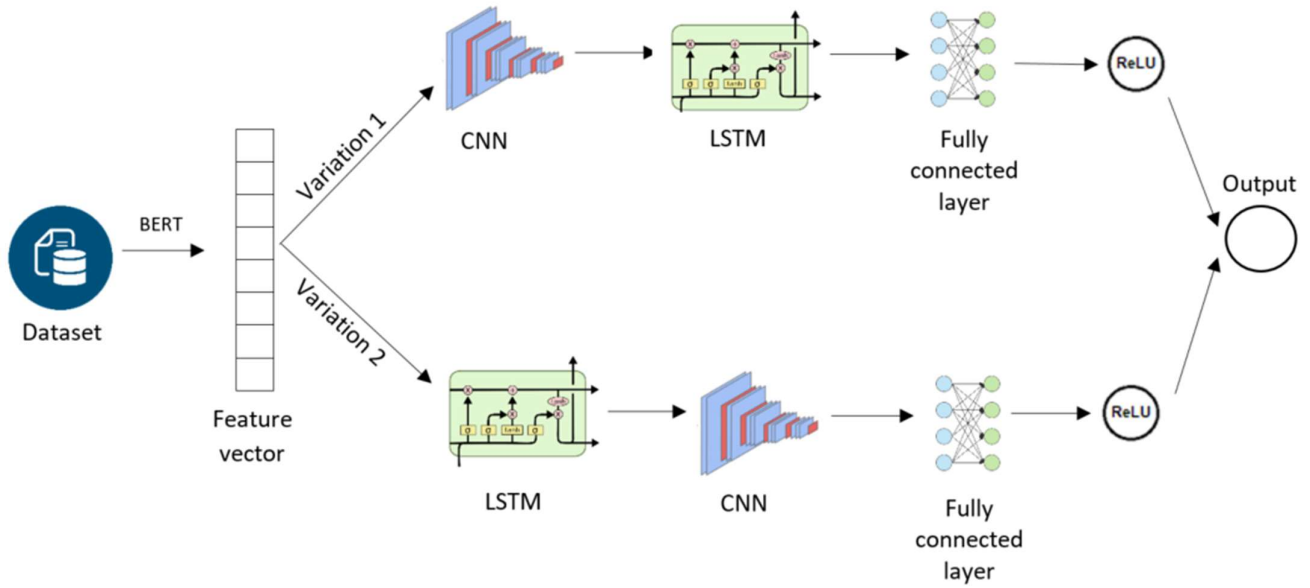


Figure 3. Process of hybrid methodology for sentiment analysis.

#### 4.3. Proposed recommendation method

The proposed recommendation method is a user-based collaborative filtering approach that takes into account explicit ratings and sentiment analysis extracted from users' reviews. We tested Singular Value Decomposition (SVD), Non-Negative Matrix Factorization (NMF), and derivative SVD++ of SVD as collaborative filtering methods. The objective is to achieve better predictive accuracy because of the addition of implicit feedback information provided by the sentiment.

Results from the CF recommendation method and sentiment analysis were combined to generate a rating used to create a list of recommendations.

Given a rating matrix  $R_{m \times n}$  ( $\mathbb{N}$ ) for training, where  $m$  is the number of users and  $n$  is the number of items,  $r_{ij} \in R_{m \times n}$  denotes the rating of user  $i$  on item  $j$ .

The rating of user  $a$  on item  $j$  in the data test set is predicted as follows:

$$pr_{aj} = \alpha * pr_{mf_{aj}} + (1 - \alpha) * pr_{sent_{aj}} \quad (1)$$

Where

- $pr_{mf_{aj}}$ : Rating for user  $a$  and item  $j$  predicted by Matrix Factorization methods (SVD, SVD++, and NMF) without using sentiments.
- $pr_{sent_{aj}}$ : Rating for user  $a$  and item  $j$  predicted by using the sentiment model.
- $\alpha$  parameter used to adjust the importance of each term of the equation.

As shown in the Algorithm 1, we used pseudocode to describe how to compute  $pr_{sent_{aj}}$ . As mentioned above, hybrid sentiment models are used for classifying each "review" in one of five possible classes. These classes are converted into sentiment scores from 1 to 5 analogous to ratings. First, for each user  $a$ , we find all items that user  $a$  already rated and the sentiment score of the corresponding review matches the explicit rating. And second, for each item  $j$ , we also find all users who already rated item  $j$  and item  $k$  (found in the first step) in the training set and their review scores also match explicit ratings.

Next, two lists of data, including items and users which are created from step 1 and step 2, are used for predicting user  $a$  rating on each item  $j$ . To do that, we compute the similarity between users by applying the cosine metric. Then, we apply the Equation 2 for rating prediction based on user similarity. The ratings of the  $k$  most similar users are used to estimate the preferences of the active user (user  $a$ ) about item (item  $j$ ) that they have not rate.

$$pr_{aj} = \bar{r}_a + \frac{\sum_{i=1}^K Sim(u_a, u_i)(r_{ij} - \bar{r}_i)}{\sum_{i=1}^K |Sim(u_a, u_i)|} \quad (2)$$

Where  $r_{ij}$  is the rating that user  $i$  give to item  $j$  respectively,  $\bar{r}_a$  and  $\bar{r}_i$  are the average ratings of user  $a$  and user  $i$ , respectively, and  $Sim(u_a, u_i)$  is the similarity between the active user  $a$  and his neighbor user  $i$ , which would be obtained by using the cosine metric.

**Algorithm 1.** Rating prediction based on sentiment for user  $a$  and item  $j$ .

```

1. Function sentiment_ratingPred (user  $a$ , item  $j$ ) {
2.   // This function is used to obtain the  $pr_{sent_{aj}}$  term of Equation 1
3.   //Step 1:
4.   FOR each item  $k$  in the training set:
5.     IF user  $a$  already rated item  $k$  AND review score matches rating THEN
6.       Add  $k$  to list of items  $I$ ;
7.   // Result this step is a set of  $m$  items  $I = \{i_1, i_2, \dots, i_m\}$ 
8.   //Step 2:
9.   FOR each user  $b$  in the training set:
10.    FOR each item  $k$  in the set of items  $I$ :
11.      IF user  $b$  already rated item  $j$  AND user  $b$  already rated item  $k$ 
        AND their review scores match ratings
12.        Add user  $b$  to list of users  $U$ ;
13.   // Results this step is a set of  $n$  users  $U = \{u_1, u_2, \dots, u_n\}$ 
14.   // Step 3:
15.   IF length( $U$ )>0 THEN
16.     FOR each user  $i$  in the set of user  $U$ :
17.       Compute sim (user  $a$ , user  $i$ ) by applying cosine metric;
18.       Add value to list of similar value  $S$ ;
19.   // Results is a set of  $n$  similarity values  $S = \{s_{a,i}\}$ 
20.   Set the  $K$  value to select the  $K$  nearest neighbors using  $S$ ;
21.   Compute the predicted rating  $pr_{aj}$  by applying the Equation 2;
22.   Return  $pr_{aj}$ ;
23. ELSE
24.   Return 0;
25. }
```

## 5. Experiments and Results

In this section, we present the experiments conducted to evaluate the performance of the proposed approach to recommender systems. In particular, we used two well-known datasets, Amazon Fine Food reviews and Amazon Movie reviews, in order to validate the proposal. The results are shown and discussed in Sub-section 5.2. The metrics used to evaluate the reliability of rating predictions were Root-Mean-Square Error (RMSE), Mean Absolute Error (MAE) and Normal MAE (NMAE). In addition, Mean Reciprocal Rank



(MRR), Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG) were used for evaluating top-N recommendations.

The configuration of related parameters, hardware devices, and the necessary library facilities was carried out before performing the experiments, such as echoes = 5, and k-fold = 5. In particular, we used Google Colab Pro with GPU Tesla P100-PCIE-16GB or GPU Tesla V100-SXM2-16GB [42], Keras [43] and Tensorflow [44] libraries. We also used the implementation of the SVD, NMF and SVD++ algorithms provided by the Surprise library (<http://surpriselib.com/>).

5.1. Dataset

We chose the datasets based on availability and accessibility criteria. Moreover, we took into account that they are widely accepted by the research community. These datasets are described below:

- **Amazon Fine Foods reviews** comprise reviews of fine foods from Amazon [45]. Each review includes product and user information, as well as the rating, and the plaintext review given by each user to each product he/she rated. The data span a period of more than 10 years, including 568,454 reviews with 256,059 users and 74,258 products up to October 2012.
- **Amazon movie reviews** consists of movie reviews from Amazon [45]. Each review also includes product and user information, ratings, and plaintext reviews. It covers a period of more than 10 years as well, including 7,911,684 reviews with 889,176 users and 253,059 products up to October 2012.

Table 1. Statistics of the datasets.

#	Amazon Fine Foods reviews	Amazon movie reviews
Number of reviews	568,454	7,911,684
Number of users	256,059	889,176
Number of products	74,258	253,059
Users with > 50 reviews	260	16,341
Average no. of words per review	56	101
Timespan	Oct 1999 - Oct 2012	Aug 1997 - Oct 2012

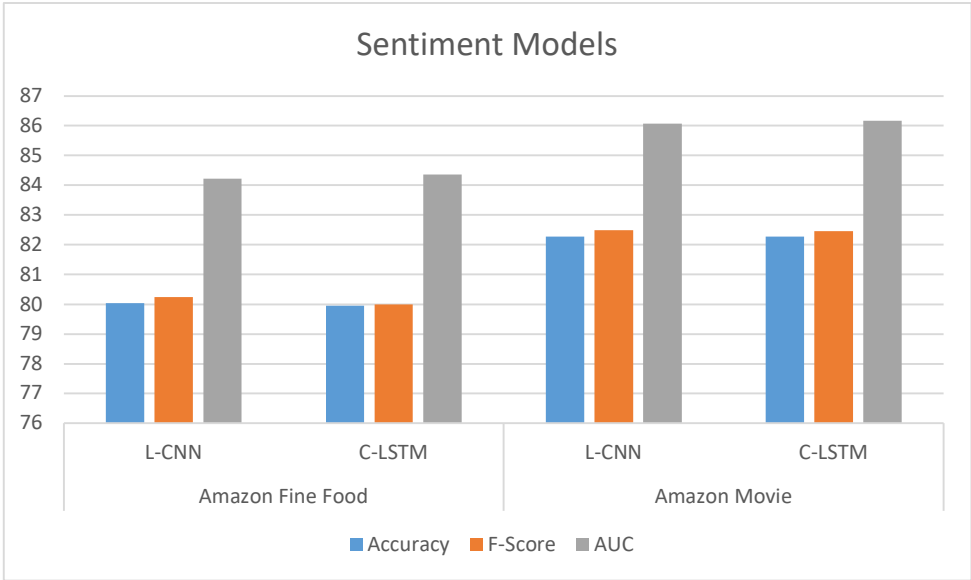
5.2. Results and Discussion

We performed experiments with two different settings without/with sentiment analysis. In the former, recommendations are based on recommender system methods without sentiment while in the second, the result of performing sentiment analysis on the reviews is incorporated to the recommendation process. We tested two hybrid deep learning models for sentiment analysis: CNN and LSTM as well as LSTM and CNN, referred to as C-LSTM, L-CNN respectively.

As presented in Figure 3, we adopt a pre-trained BERT model in order to vectorize each plaintext review. The obtained vector is then fed into C-LSTM or L-CNN followed by the fully connected layer. Finally, ReLU is stacked on the top of the classifier. The output of the sentiment classifier is exploited for recommendation. Table 2 and Figure 4 present the experimental results of sentiment classification. The results show that the performances of hybrid models are encouraging, with accuracy and F-score over 80% and AUC over 84%, respectively. These models will be applied to predict sentiment rating before being combined with recommendation methods.

Table 2. Sentiment performance of hybrid deep-learning models.

Measures	Amazon Fine Foods Reviews		Amazon Movie Reviews	
	L-CNN	C-LSTM	L-CNN	C-LSTM
Accuracy	80.04%	79.95%	82.27%	82.27%
F-Score	80.24%	80.00%	82.49%	82.46%
AUC	84.22%	84.36%	86.07%	86.17%



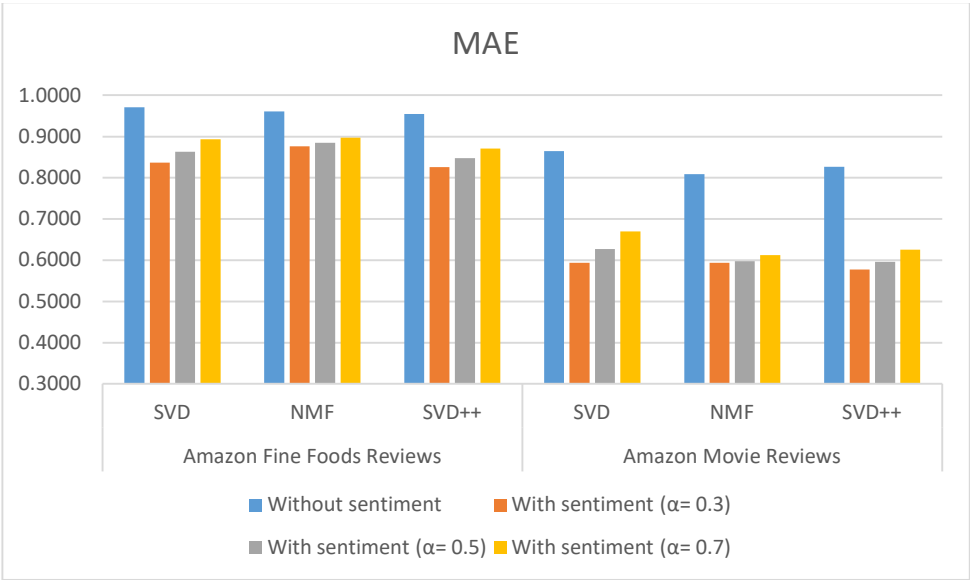
**Figure 4.** Sentiment performance of hybrid deep-learning models using BERT.

To validate our recommendation approach, we compared the performance of recommendations with/without using sentiment analysis for both rating prediction and item recommendation (recommendation of top-N lists).

Tables 3 -5 show the results of MAE, RMSE and NMAE measures for rating prediction on both dataset food and movie reviews. They were calculated based on SVD, NMF and SVD++ algorithm with and without using sentiment analysis. Alpha ( $\alpha$ ) parameter is used to adjust the importance of the recommendation result without and with sentiment in the equation 1. Figures 5-7 illustrate the comparative results obtained from the recommender with sentiment analysis on different values of the  $\alpha$  parameter against those obtained from the recommender without sentiment analysis.

**Table 3.** MAE values without and with L-CNN sentiment analysis model.

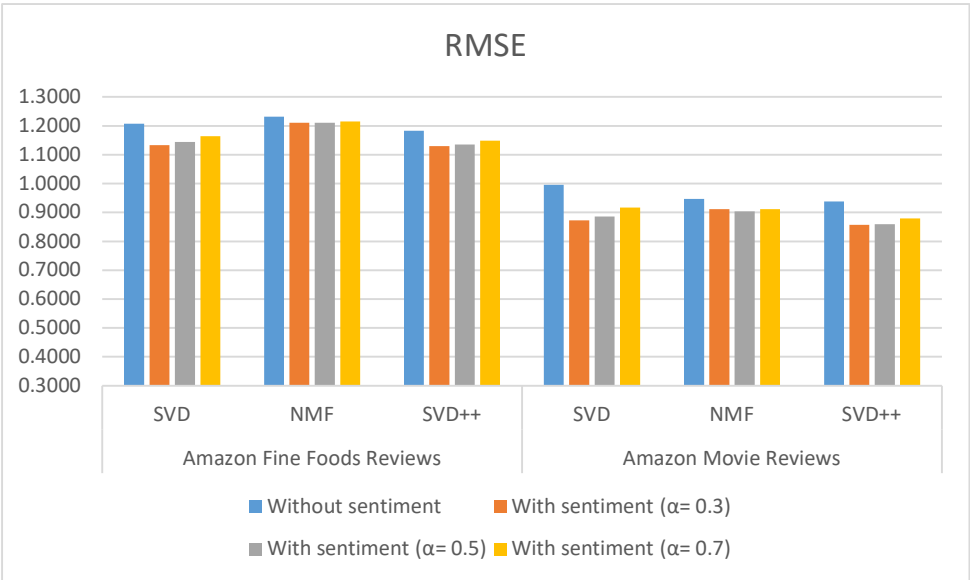
#	Amazon Fine Foods Reviews			Amazon Movie Reviews		
	SVD	NMF	SVD++	SVD	NMF	SVD++
Without sentiment	0.9706	0.9608	0.9540	0.8644	0.8087	0.8266
With sentiment ( $\alpha= 0.3$ )	0.8365	0.8762	0.8263	0.5943	0.5936	0.5770
With sentiment ( $\alpha= 0.5$ )	0.8634	0.8846	0.8470	0.6268	0.5976	0.5959
With sentiment ( $\alpha= 0.7$ )	0.8933	0.8964	0.8707	0.6701	0.6125	0.6253



**Figure 5.** MAE measures comparison for different types of method and datasets using L-CNN sentiment model.

**Table 4.** RMSE values without and with L-CNN sentiment analysis model.

#	Amazon Fine Foods Reviews			Amazon Movie Reviews		
	SVD	NMF	SVD++	SVD	NMF	SVD++
Without sentiment	1.2076	1.2312	1.1831	0.9960	0.9464	0.9376
With sentiment ( $\alpha=0.3$ )	1.1338	1.2103	1.1292	0.8732	0.9112	0.8577
With sentiment ( $\alpha=0.5$ )	1.1442	1.2102	1.1356	0.8851	0.9041	0.8598
With sentiment ( $\alpha=0.7$ )	1.1633	1.2150	1.1493	0.9166	0.9110	0.8791

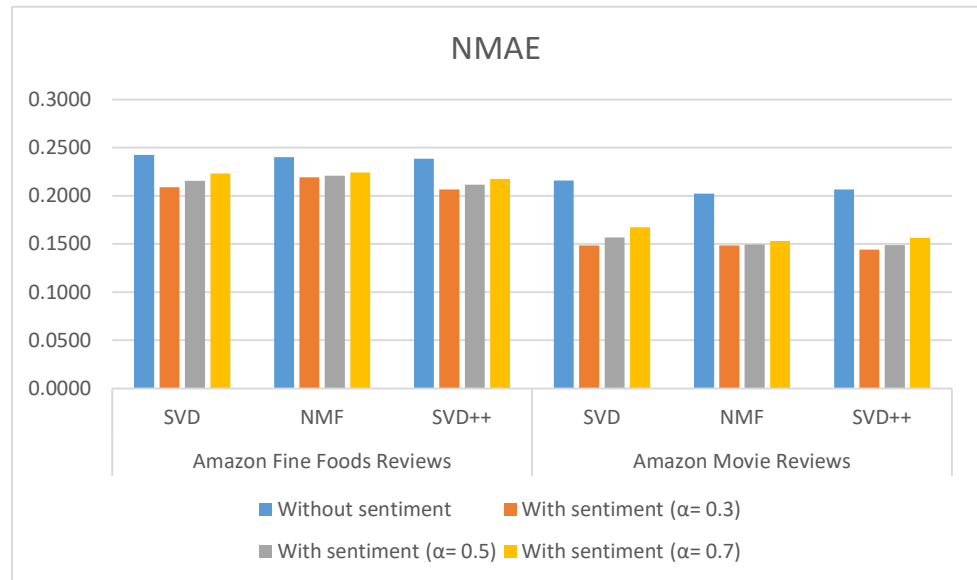


**Figure 6.** RMSE measures comparison for different types of method and datasets using L-CNN sentiment model.

**Table 5.** NMAE values without and with L-CNN sentiment analysis model.

	Amazon Fine Foods Reviews	Amazon Movie Reviews
--	---------------------------	----------------------

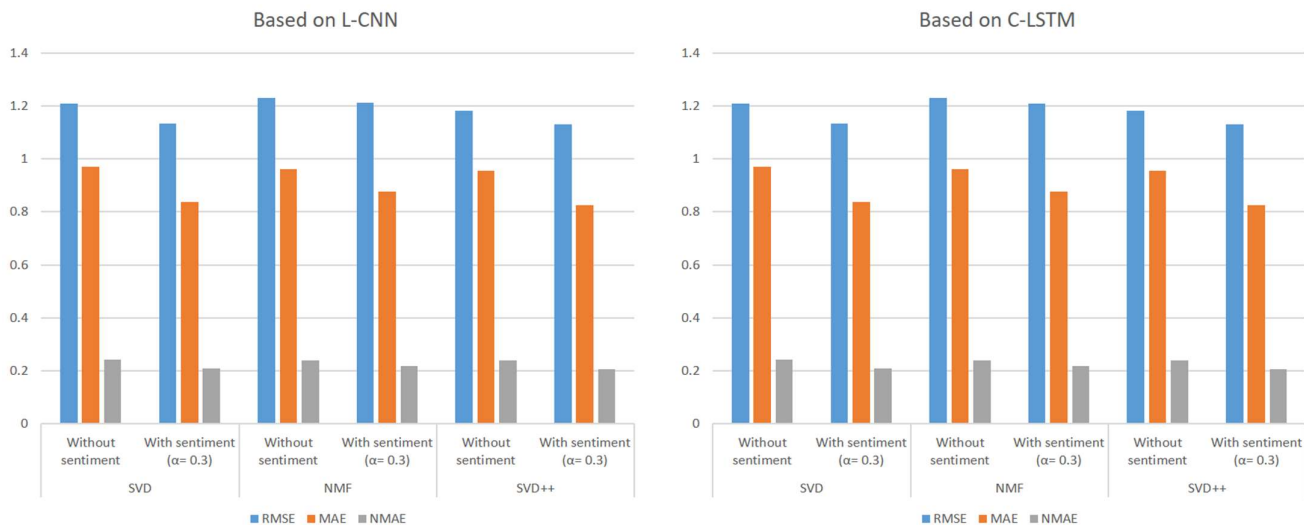
#	SVD	NMF	SVD++	SVD	NMF	SVD++
Without sentiment	0.2427	0.2402	0.2385	0.2161	0.2022	0.2066
With sentiment ( $\alpha=0.3$ )	0.2091	0.2191	0.2066	0.1486	0.1484	0.1443
With sentiment ( $\alpha=0.5$ )	0.2158	0.2211	0.2117	0.1567	0.1494	0.1490
With sentiment ( $\alpha=0.7$ )	0.2233	0.2241	0.2177	0.1675	0.1531	0.1563



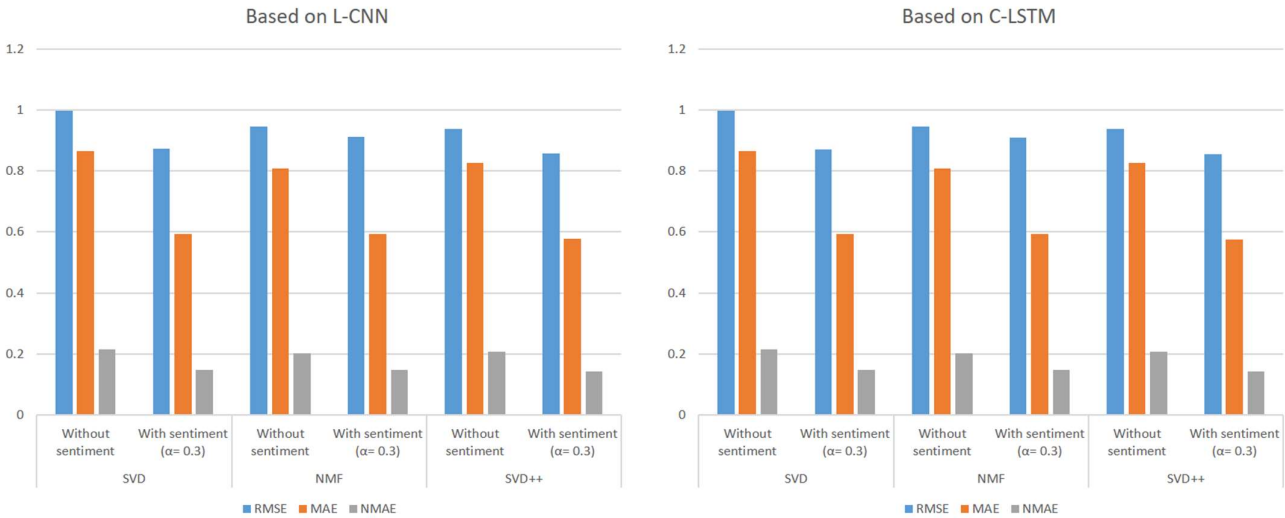
**Figure 7.** NMAE measures comparison for different types of method and datasets using L-CNN sentiment model.

The results show that RSME, MAE and NMAE yielded by the approach that combines CF with sentiment analysis are better than the error rates yielded by traditional CF methods without sentiment on all algorithm in all  $\alpha$  values. We found that the best results of the proposal are obtained with  $\alpha=0.3$ .

Regarding the type of datasets, Amazon Movie Reviews provided better results than those of Amazon Fine Foods reviews. For example, MAE measured with SVD++ is 0.577 with  $\alpha=0.3$ ; RMSE measured with SVD++ is 0.8577 with  $\alpha=0.3$ ; and NMAE measured with SVD++ is 0.1443 with  $\alpha=0.3$ .



**Figure 8.** Comparison of the sentiment-based methods with the L-CNN model (left side) and the C-LSTM (right side) and  $\alpha=0.3$  against non sentiment-based methods on Amazon Fine Foods Reviews.



**Figure 9.** Comparison of the sentiment-based methods with the L-CNN model (left side) and the C-LSTM (right side) and  $\alpha=0.3$  against non sentiment-based methods on Amazon Movie Reviews.

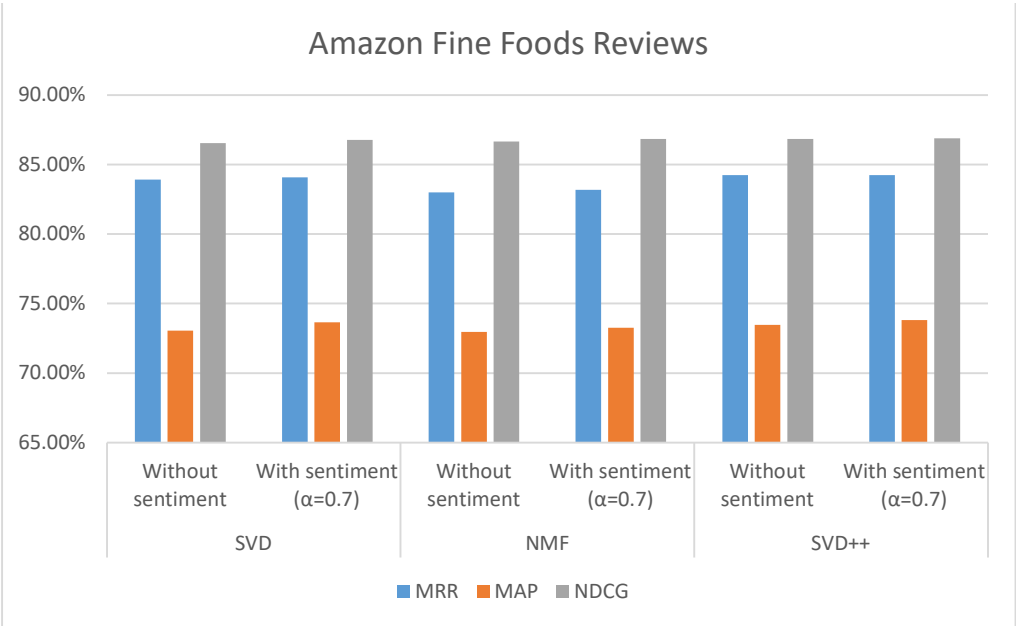
Figure 8 and Figure 9 illustrate the comparison of the sentiment-based methods with the L-CNN and the C-LCTM against non sentiment-based methods on Amazon Fine Foods Reviews and Amazon Movie Reviews. The values with sentiment are obtained with  $\alpha = 0.3$ . We found that C-LSTM and L-CNN provide similar results. In addition, the sentiment-based approach provides better results on Amazon Movie reviews.

For all algorithms applied to two datasets, the combined proposal provides lower error rates. For example, the sentiment-based approach on Amazon Movie Reviews with  $\alpha = 0.3$  when L-CNN is used for sentiment analysis provided the following percentage improvement with SVD: 12.29% in RMSE; 27.01% in MAE; and 6.75% in NMAE. With the above results, we see that the sentiment model helps to improve the predicted ratings. Instead of just using explicit rating, the predictive model now considers the aspect of analyzing reviews of related items and users. Because more information is available in the new recommendation method, we get better than usual results.

In addition to prove that the proposed method performs better in predicting ratings, we also check the performance for top-N recommendations. MRR, MAP, and NDCG rank-based metrics have been computed. The results obtained for  $N = 5$  are given in Table 6 and Table 7, and in Figure 10 and Figure 11. SVD, NMF, and SVD++ values, respectively, with L-CNN and C-LSTM sentiment models are obtained when applied on  $\alpha = 0.7$ .

**Table 6.** MRR, MAP and NDCG values without and with L-CNN sentiment model on Amazon Fine Foods Reviews with  $\alpha=0.7$ .

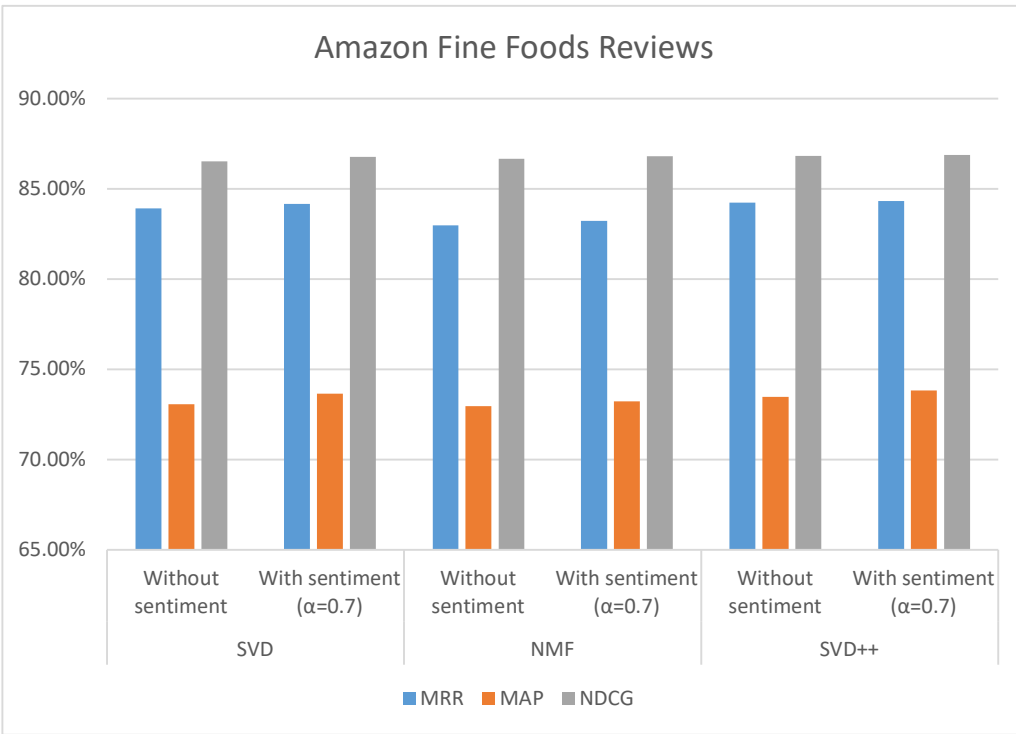
#	SVD		NMF		SVD++	
	Without sentiment	With sentiment ( $\alpha=0.7$ )	Without sentiment	With sentiment ( $\alpha=0.7$ )	Without sentiment	With sentiment ( $\alpha=0.7$ )
MRR	83.92%	84.09%	82.98%	83.19%	84.24%	84.24%
MAP	73.06%	73.67%	72.97%	73.26%	73.48%	73.82%
NDCG	86.53%	86.78%	86.67%	86.84%	86.84%	86.89%



**Figure 10.** MRR, MAP and NDCG values without and with L-CNN sentiment model on Amazon Fine Foods Reviews with  $\alpha = 0.7$ .

**Table 7.** MRR, MAP and NDCG values without and with C-LSTM sentiment model on Amazon Fine Foods Reviews with  $\alpha = 0.7$ .

#	SVD		NMF		SVD++	
	Without sentiment	With sentiment ( $\alpha=0.7$ )	Without sentiment	With sentiment ( $\alpha=0.7$ )	Without sentiment	With sentiment ( $\alpha=0.7$ )
MRR	83.92%	84.16%	82.98%	83.23%	84.24%	84.33%
MAP	73.06%	73.64%	72.97%	73.23%	73.48%	73.84%
NDCG	86.53%	86.78%	86.67%	86.82%	86.84%	86.89%





**Figure 11.** MRR, MAP and NDCG values without and with C-LSTM sentiment model on Amazon Fine Foods Reviews with  $\alpha = 0.7$ .

The values of MRR, MAP, and NDCG show that the proposed method also improve top-N recommendations. In the case of the Amazon Foods Review with  $\alpha = 0.7$  and C-LSTM sentiment model, the increase in MAP was 0.58% (SVD), 0.26% (NMF), and 0.36% (SVD++) over without sentiment. Regarding NDCG, the increase was 0.25% (SVD), 0.15% (NMF), and 0.06% (SVD++) over without sentiment. And, the value of MRR was increased on 0.24% (SVD), 0.25% (NMF), and 0.09% (SVD++).

Three algorithms (SVD, NMF, and SVD++) were tested in two ways, with explicit ratings only, and combining explicit ratings with sentiment extracted from reviews. In most cases, the combined approach with sentiments from two classification models (C-LSTM and L-CNN) on food and movie reviews datasets gave better results. However, the improvement for top-N recommendation is not as significant than the achieved for rating prediction.

In general, the sentiment-based methods proposed in this work provide better results than those based only on explicit ratings. These improvements have occurred in both data sets used in the study. General summaries of the results achieved in the experiments referenced earlier are discussed below:

- We presented and evaluated a recommendation approach that integrates sentiment analysis and collaborative filtering methods.
- Two datasets, Amazon Fine Foods Review and Amazon Movie Review, are used for evaluation. Each plaintext review is vectorized by using the pre-trained BERT model.
- Two hybrid sentiment classification models, CNN-LSTM and LSTM-CNN, are used for extracting sentiments from reviews, which are incorporated as implicit feedback into the recommender system models.
- We applied SVD, NMF, and SVD++ recommendation methods following the user-based CF approach.
- Accuracy, F-score, and AUC were computed for validating sentiment classification models.
- The evaluation of the recommendation method was performed for rating prediction and top-N recommendation. RMSE, MAE, and NMAE were the metrics used in the first case, and MRR, MAP and NDCG were the metrics used in the second case.
- The sentiment-based proposal increased the recommendation reliability in comparison to traditional, ratings-based recommendation methods on the two datasets.

## 6. Conclusion

In this paper, we have proposed an application of sentiment analysis in recommender systems that is based on hybrid deep learning models and collaborative filtering on online social networks. The system architecture presented in this work, can integrate a variety of techniques that have been proposed to perform recommendations, including the preprocessing strategy, hybrid deep learning models for sentiment analysis and methods for recommender systems. The architecture can be used to develop a recommender system in the context of social networks that takes advantage of sentiment analysis performed on user opinions and reviews in the network. We conducted experiments with reviews of food and movies. Based on such experiments, we demonstrate the utility and applicability of our approaches in producing personalized recommendations on online social networks.

The results show that the joint use of deep learning-based sentiment analysis and collaborative filtering methods significantly improves the performance last ones. This is achieved through the exploitation of additional information from user reviews / comments data. Its integration into the traditional recommendation methods makes the recommender system more reliable and capable of providing better recommendations to users.

As future work, we plan to explore other application domains to ensure that the proposed architecture can be generalized to efficiently solve similar problems. We will also consider researching new sentiment analysis techniques, such as graph convolutional networks, for a potential improvement of this aspect.

**Author Contributions:** Conceptualization, N.C.D. and M.M.-G.; methodology, M.M.-G.; software, N.C.D.; validation, M.M.-G., and F.D.I.P.; formal analysis, N.C.D., and M.M.-G.; investigation, N.C.D.; data curation, N.C.D.; writing—original draft preparation, N.C.D.; writing—review and editing, M.M.-G. and F.D.I.P.; visualization, N.C.D.; supervision, M.M.-G. and F.D.I.P.; project administration, F.D.I.P.; funding acquisition, F.D.I.P.

**Funding:** This work was supported by the Spanish government and European (Fondo Europeo de Desarrollo Regional) FEDER funds, project InEDGEMobility: Movilidad inteligente y sostenible soportada por Sistemas Multi-agentes y Edge Computing (RTI2018-095390-B-C32).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Dang, N. C.; Moreno-García, M. N.; De la Prieta, F., Sentiment analysis based on deep learning: A comparative study. *Electronics* **2020**, *9*, (3), 483.
2. Preethi, G.; Krishna, P. V.; Obaidat, M. S.; Saritha, V.; Yenduri, S. In *Application of Deep Learning to Sentiment Analysis for recommender system on cloud*, 2017 International Conference on Computer, Information and Telecommunication Systems (CITS), 2017; IEEE: 2017; pp 93-97.
3. Keenan, M. J. S., *Advanced Positioning, Flow, and Sentiment Analysis in Commodity Markets: Bridging Fundamental and Technical Analysis*. 2nd ed.; 2018.
4. Sánchez-Moreno, D.; Moreno-García, M. N.; Mobasher, B.; Sonboli, N.; Burke, R., Using Social Tag Embedding in a Collaborative Filtering Approach for Recommender Systems. In *The 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*.
5. Sánchez-Moreno, D.; López Batista, V. F.; Muñoz Vicente, M. D.; Gil González, A. B.; Moreno-García, M. N., A session-based song recommendation approach involving user characterization along the play power-law distribution. *Complexity* **2020**.
6. Wang, Y.; Wang, M.; Xu, W., A sentiment-enhanced hybrid recommender system for movie recommendation: a big data analytics framework. *Wireless Communications and Mobile Computing* **2018**.
7. Kumar, S.; De, K.; Roy, P. P., Movie recommendation system using sentiment analysis from microblogging data. *IEEE Transactions on Computational Social Systems* **2020**, *7*, (4), 915-923.
8. Rao, K. Y.; Murthy, G.; Adinarayana, S., Product recommendation system from users reviews using sentiment analysis. *International Journal of Computer Applications* **2017**, 975, 8887.
9. Gurini, D. F.; Gasparetti, F.; Micarelli, A.; Sansonetti, G., A Sentiment-Based Approach to Twitter User Recommendation. *RSWeb@ RecSys* **2013**, 1066.
10. Osman, N.; Noah, S.; Darwich, M., Contextual sentiment based recommender system to provide recommendation in the electronic products domain. *International Journal of Machine Learning and Computing* **2019**, *9*, (4), 425-431.
11. Contrates, F. G.; Alves-Souza, S. N.; Filgueiras, L. V. L.; DeSouza, L. S. In *Sentiment analysis of social network data for cold-start relief in recommender systems*, World Conference on Information Systems and Technologies, 2018; Springer: 2018; pp 122-132.
12. Nabil, S.; Elbouhdidi, J.; Yassin, M. In *Recommendation system based on data analysis-application on tweets sentiment analysis*, 2018 IEEE 5th International Congress on Information Science and Technology (CiSt), 2018; IEEE: 2018; pp 155-160.
13. Ziani, A.; Azizi, N.; Schwab, D.; Aldwairi, M.; Chekkai, N.; Zenakhra, D.; Cheriguene, S. In *Recommender system through sentiment analysis*, 2nd International Conference on Automatic Control, Telecommunications and Signals, 2017; 2017.

14. Abbasi, F.; Khadivar, A.; Yazdinejad, M., A Grouping Hotel Recommender System Based on Deep Learning and Sentiment Analysis. *Journal of Information Technology Management* **2019**, 11, (2), 59-78.
15. Rosa, R. L.; Rodriguez, D. Z.; Bressan, G., Music recommendation system based on user's sentiments extracted from social networks. *IEEE Transactions on Consumer Electronics* **2015**, 61, (3), 359-367.
16. Osman, N. A.; Noah, S. A. M. In *Sentiment-based model for recommender systems*, 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), 2018; IEEE: 2018; pp 1-6.
17. Nouh, R. M.; Lee, H.-H.; Lee, W.-J.; Lee, J.-D., A smart recommender based on hybrid learning methods for personal well-being services. *Sensors* **2019**, 19, (2), 431.
18. Devipriya, K.; Prabha, D.; Pirya, V.; Sudhakar, S., Deep learning sentiment analysis for recommendations in social applications. *Int J Sci Technol Res* **2020**, 9, (1), 3812-3815.
19. Singh, V. K.; Mukherjee, M.; Mehta, G. K. In *Combining collaborative filtering and sentiment classification for improved movie recommendations*, International Workshop on Multi-disciplinary Trends in Artificial Intelligence, 2011; Springer: 2011; pp 38-50.
20. Nimirthi, P.; Krishna, P. V.; Obaidat, M. S.; Saritha, V., A framework for sentiment analysis based recommender system for agriculture using deep learning approach. In *Social Network Forensics, Cyber Security, and Machine Learning*, Springer: 2019; pp 59-66.
21. Bhavitha, B.; Rodrigues, A. P.; Chiplunkar, N. N. In *Comparative study of machine learning techniques in sentimental analysis*, 2017 International conference on inventive communication and computational technologies (ICICCT), 2017; IEEE: 2017; pp 216-221.
22. Salas-Zárate, M. d. P.; Medina-Moreira, J.; Lagos-Ortiz, K.; Luna-Aveiga, H.; Rodriguez-Garcia, M. A.; Valencia-García, R., Sentiment analysis on tweets about diabetes: an aspect-level approach. *Computational mathematical methods in medicine* **2017**, 2017.
23. Huq, M. R.; Ali, A.; Rahman, A., Sentiment analysis on Twitter data using KNN and SVM. *Int J Adv Comput Sci Appl* **2017**, 8, (6), 19-25.
24. Pinto, D.; McCallum, A.; Wei, X.; Croft, W. B. In *Table extraction using conditional random fields*, Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, 2003; 2003; pp 235-242.
25. Soni, S.; Sharaff, A. In *Sentiment analysis of customer reviews based on hidden markov model*, Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015), 2015; 2015; pp 1-5.
26. Zhang, X.; Zheng, X. In *Comparison of text sentiment analysis based on machine learning*, 2016 15th International Symposium on Parallel and Distributed Computing (ISPDC), 2016; IEEE: 2016; pp 230-233.
27. Malik, V.; Kumar, A., Sentiment Analysis of Twitter Data Using Naive Bayes Algorithm. *International Journal on Recent and Innovation Trends in Computing and Communication* **2018**, 6, (4), 120-125.
28. Mehra, N.; Khandelwal, S.; Patel, P., Sentiment identification using maximum entropy analysis of movie reviews. *Stanford University* **2002**.
29. Wu, H.; Li, J.; Xie, J., Maximum Entropy-based Sentiment Analysis of Online Product Reviews in Chinese. In *Automotive, Mechanical and Electrical Engineering*.
30. Firmino Alves, A. L.; Baptista, C. d. S.; Firmino, A. A.; Oliveira, M. G. d.; Paiva, A. C. d. In *A Comparison of SVM versus naive-bayes techniques for sentiment analysis in tweets: a case study with the 2013 FIFA confederations cup*, Proceedings of the 20th Brazilian Symposium on Multimedia and the Web, 2014; ACM: 2014; pp 123-130.
31. Pandey, A. C.; Rajpoot, D. S.; Saraswat, M., Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing & Management* **2017**, 53, (4), 764-779.
32. Medhat, W.; Hassan, A.; Korashy, H., Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal* **2014**, 5, (4), 1093-1113.

33. Hochreiter, S.; Schmidhuber, J. In *LSTM can solve hard long time lag problems*, Advances in neural information processing systems, 1997; 1997; pp 473-479.
34. Yamashita, R.; Nishio, M.; Do, R. K. G.; Togashi, K., Convolutional neural networks: an overview and application in radiology. *Insights into imaging* **2018**, 9, (4), 611-629.
35. Lu, J.; Wu, D.; Mao, M.; Wang, W.; Zhang, G., Recommender system application developments: a survey. *Decision Support Systems* **2015**, 74, 12-32.
36. Betru, B. T.; Onana, C. A.; Batchakui, B., A Survey of State-of-the-art: Deep Learning Methods on Recommender System. *International Journal of Computer Applications* **2017**, 162, (10).
37. Kardan, A. A.; Ebrahimi, M., A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups. *Information Sciences* **2013**, 219, 93-110.
38. Adomavicius, G.; Tuzhilin, A.; engineering, d., Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge* **2005**, 17, (6), 734-749.
39. Schafer, J. B.; Frankowski, D.; Herlocker, J.; Sen, S., Collaborative filtering recommender systems. In *The adaptive web*, Springer: 2007; pp 291-324.
40. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; Dean, J. In *Distributed representations of words and phrases and their compositionality*, Advances in neural information processing systems, 2013; 2013; pp 3111-3119.
41. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K., Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* **2018**.
42. Making the Most of your Colab Subscription. <https://colab.research.google.com/notebooks/pro.ipynb> (accessed on 22 January 2021)
43. Keras: the Python deep learning API. <https://keras.io/> (accessed on 10 December 2020)
44. TensorFlow. <https://www.tensorflow.org/> (accessed on 10 December 2020)
45. McAuley, J. J.; Leskovec, J. In *From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews*, Proceedings of the 22nd international conference on World Wide Web, 2013; 2013; pp 897-908.