Al-Aqsa University Journal (Natural Sciences Series),Vol.23, No.1, Pages 1-17,Jan.2021
ISSN 2070-3155

Exploiting Sentiment Analysis to Enhance the Collaborative Filtering Recommendations

Dr. Basem H. Ahmed *

استغلال تحليل المشاعر المتواجدة في نصوص التعليقات الخاصة بالمستخدمين في تحسين جودة الاقتراح في انظمة الاقتراح الخاصة بالتصفية التشاركية

الملخص

تعد انظمة الفلترة التعاونية من أفضل أنظمة التَّوْصِيَة شيوعًا واستخدامًا، على الرغم من ذلك فان هذه الانظمة تواجه مشاكل وتحديات جمة منها البيانات المتفرقة والبداية الباردة. نظرًا لأن نهج هذه الانظمة يعتمد في الغالب على التقييمات الرقمية الصريحة للمستخدمين في التنبؤ وتقديم التوصيات وعادة ما تكون مثل هذه التقييمات الرقمية الصريحة غير كافية أو محدودة للغاية لتقديم توصيات ذات جودة عالية. لذلك، فان أنظمة الفلترة التعاونية التي تقتصر فقط على استخدام التفضيلات الرقمية الصريحة مثل الارقام تقدم توصية منخفضة الجودة. في الآونة الأخيرة، شجعت العديد من مواقع التجارة الإلكترونية، مثل موقع امازون، المستخدمين على تقديم تعليقات على المنتجات بتنسيق النص الحر، المعروف باسم مراجعات المستخدمين وذلك لوصف تجربتهم مع المنتجات. وعليه يمكن اعتبار هذه المراجعات نوعًا من تفضيلات المستخدم وذلك لأنها عادة ما تتضمن سبب إعجاب المستخدمين بالمنتج أو عدم إعجابهم به. من هنا، تهدف هذه الورقة الى اقتراح طريقة توصية من خلال استغلال قطبية المشاعر المتواجدة في تعليقات المستخدمين واعتبارها كجزء من تفضيلات المستخدمين وذلك باستخدام اساليب التعلم العميق ودمج هذه التفضيلات مع التفضيلات الصريحة مثل تقييمات المستخدمين وذلك لتحسين طُرق الفلترة التعاونيّة المقترحة في أدبيّات الموضوع. وقد أُجريت التجارب على عدة مجموعات من البيانات التابعة لأمازون، وأسفرت النتائج عن تحسُّنٍ ملحوظ حققته الطريقة المقترحة مقارنةً بالطُرق القائمة.

الكلمات المفتاحية

أنظمة التوصية، نظام التوصية بالتصفية التشاركية، تعليقات المستخدمين، التوصية بأفضل العناصر.

د. باسم احمد – * قسم علوم الحاسوب، جامعة الأقصى، غزة
Email: basem@alaqsa.edu.ps

Exploiting Sentiment Analysis to Enhance the Collaborative Filtering Recommendations

Abstract

Collaborative Filtering (CF) is considered the most popular and widely employed recommender system approach. However, the CF has exposed limitations such as cold-start and data sparsity. Since the CF approach mostly relies on users' explicit rating preferences to predict items and provided recommendations. Such explicit numerical ratings are usually insufficient or very limited to deliver a good recommendation quality. Recently, many e-commerce websites, such as Amazon, encouraged users to provide comments in free text format, well-known as user reviews to describe their experience with the products. These reviews can also be considered a type of user preference because they usually explain why users liked or disliked a product. This paper proposes to infer the sentiment polarity of user's reviews text using deep learning methods and integrate such preferences of user review sentiment with users actual rating to enhance the recommendation quality of the CF recommender system. The experimental results on different Amazon datasets demonstrate, that the proposed approach improves the performance of the CF recommender system by integrating the sentiment polarity of user reviews in the recommendation process and produces recommendations with higher quality in terms of Recall, precision and F1-measure compared to the baseline CF methods. The results show that the proposed approach achieved state-of-the-art performance, which increased the F1- measure around 3.5%. The Precision by around 3.3% and the Recall around 3.6%, compared with the baseline approaches.

Keywords—Collaborative filtering; Review text; Sentiment Analysis; Deep Learning; Top-N recommendation

-Introduction

nowadays, a huge volume of information being continually created every day on the Internet. Consequently, makes challenges for Internet users to get the information they need quality and suffer from an information overload problem. In this case, it is often difficult for an Internet user to obtain useful or relevant information when it is necessary. To some extent, Search engines have partially solved this problem, but without any filtering of the information based upon internet users' favorites and interests. Hence, a Recommender System (RS) is needed to tackle the information overload problem and further filter the information to provide users with the recommendation of relevant and personalized formations. Recently, many commercial companies' sites such as Amazon.com, Netflix, YouTube, etc. employed a recommendation service. Collaborative Filtering is a promising RS approach that automating the word-of-mouth paradigm (Jeon & Ahn, 2015). The main idea behind CF, it delivers a personalized recommendation to the target users based on identifying like-minded users with similar tastes and uses their opinion to recommend items to the target user (Ayman S Ghabayen & Basem H Ahmed, 2021; Jeon & Ahn, 2015). However, the CF has exposed limitations such as cold-start and data sparsity (Natarajan, Vairavasundaram, Natarajan, & Gandomi, 2020). Since the CF approach mostly relies on users' explicit rating preferences to predict items and provided recommendations. Such explicit ratings are usually insufficient or very limited to provide a good recommendation quality. Therefore, CF recommendation limited to using only users explicit rating as users' preferences provided low recommendation quality (Jeon & Ahn, 2015). Recently, many e-commerce websites, such as Amazon, encouraged users to provide comments in free text format, well-known as user reviews to describe their experience with the products. These reviews can also be considered a type of user preference because they usually explain why users liked or disliked a (Hou, Yannou, Leroy, & Poirson, 2019). Sentiment analysis is also known as opinion mining algorithms is becoming popular in the text mining literature. Sentiment analysis algorithms identify user's sentiments and opinions and categorize user's opinions expressed in reviews text as (positive, negative, or natural) regarding different objects or topics using computational linguistic (Babu & Rawther, 2021; Kauffmann et al., 2020; Liu & Zhang, 2012; Zucco, Calabrese, Agapito, Guzzi, & Cannataro, 2020). Hence, this paper aims to contribute to user's review sentiments polarity to enhance the recombination quality of the CF recommender system.

The advantage of this system is that it will overcome a common limitation of rating-based collaborative filtering approaches: the system will be able to match users based on the similarity of their reviews, even if they rated an item differently.

The rest of the paper is organized as follows. Section 2 presents a background of a collaborative-based filtering recommender system. Section 3 describes related works. Section 4 presents the proposed approach. Section 5 presents the experiments and dataset used in addition to the state- of the art method used in comparisons. Section 6 presents the results and discussion. Finally, the last section presents the conclusions and opportunities for future work.

Background

In the last decade, Recommender System (RS) has emerged as a major research interest that aims to provide recommendations to Internet users. RS is mainly classified into two categories: collaborative filtering CF, content-based and hybrid approach based on the types of input data. A content-based approach recommends items to target users based on items they have preferred or liked in the past. A CF approach recommends items to target users based on items that other like-minded users with similar preferences preferred or liked in the past (Ricci, Rokach, & Shapira, 2015). CF approach is considered the most popular and widely employed approach for a recommendation. Hence, the main idea behind the CF method is to exploit information about the past behavior/opinions (e.g., user rating) of an existing user nearest neighbor (KNN hereafter) also known as 'like-minded' users for predicting items; the current user will most probably like or be interested in. These similarity relations are indicated from the user-item numerical rating achieved by the (Ricci et al., 2015). The most commonly used metrics to calculate the similarity between two users U and V are Cosine-based and Correlation-based similarity measures (Jannach, Zanker, Felfernig, & Friedrich, 2010). The similarity between user U and V is measured by calculating the cosine angle between users corresponding rating vector is expressed as:

$$(u,v) = \frac{\sum_{i_m \in I_{u,v}} r_{u,m} * r_{v,m}}{\sqrt{\sum_{i_m \in I_{u,v}} r^2{}_{n,m}} * \sqrt{\sum_{i_m \in I_{u,v}} r^2{}_{v,m}}} \qquad (1)$$

where $I_{u,v}$ denote the co-rated items between users U and V. In other words it denotes items rated by both users. Subsequently, to predict the rating that U would assign to an item i (unseen by user U in the past), the ratings given by user U nearest neighbors to that item are combined as follow:

$$r(U,i) = \bar{r}_U + \sum_{V \in KNN}(U,V) * \qquad (2)$$

where $\bar{r}_U$ denote the mean value of ratings entered by user U.

Related work

In recent years, big efforts have been made to exploit the rich information contained in user reviews into the RS recommendation process to enhance the quality of recommendation. According to (Aslanyan & Frasincar, 2021; L. Chen, Chen, Wang, & Interaction, 2015) different information features can be produced from review texts and invoked in the RS process. Such as review terms, review topic, and review sentiment. This study focused on related works that used review sentiment to enhance RS recommendation quality. This study aims to use the overall sentiment polarity of user review to enhance the CF recommendation quality.

Typically, the user expressed sentiment (i.e., positive or negative) on items that can be utilized to improve the rating prediction process in CF. The sentiment of user review can simply be identified by applying supervised (Ayman S Ghabayen & Basem H Ahmed, 2021) or unsupervised (C. C. Chen, Chen, Wu, & Engineering, 2011) machine learning techniques. then, the overall review sentiment is turned into a scalar rating, which can be helpful to improve CF approaches performance. For instance, user overall opinion extracted from the reviews can be associated with user rating scores to indicate the quality of user rating in CF (Raghavan, Gunasekar, & Ghosh, 2012). Also, reviews opinion can be combined with user rating to form the latent representation for users and items in a biased matrix factorization model (Pero & Horváth, 2013).

(Shen, Zhang, Yu, & Min, 2019) Proposed a sentiment-based recommendation model to enhance the recommendation accuracy. The model infers the sentiment obtained from a constructed star-based sentiment dictionary. Then, incorporate the user's feedbacks preferences from reviews and numerical rating into real values using a probabilistic matrix factorization for prediction. (Gallege & Raje, 2016) used SA to rate products for online software services. They utilized the sentiment and subjective logic of user review text as an external knowledge source to enhance the CF and content-based filtering approach.

(Y. Wang, Wang, & Xu, 2018) proposed a recommender system approach by adopting a hybrid recommendation and opinion mining. Their approach gives a better performance compared to the standard recommendation models in terms of various evaluation criteria. (Li, Cui, Shen, & Ma, 2016) enlaced the recommendation quality and alleviate the cold start problem issue of the recommender system by using sentiment analysis to identify users' opinions from microblogging posts. This step attempt to enhance the recommendation results by bridging the gap between user communication knowledge and social network sites.

(H. Wang & Luo, 2014) exploited the words in the user's review text to determine the similarity between users. Particularly, two users considered more similar if there exists a significant amount of overlapping word among them on items they reviewed in the past. The proposed similarity results are then utilized in the rating prediction task in CF. In a way similar to (H. Wang & Luo, 2014), the similarity between users is measured by considering the user reviewing behavior on item (Terzi, Rowe, Ferrario, & Whittle, 2014).

(Kim, Park, Oh, Lee, & Yu, 2016) proposed a context-aware recommendation model, which exploits user's reviews as a vital source of information. The proposed model uses a convolutional neural network (CNN) and word embedding to capture the latent features of items from reviews texts on them. Subsequently, the deduced latent features are incorporated into a matrix factorization model to estimate the user's rating on items. (Zheng, Noroozi, & Yu, 2017) proposed (DeepCoNN) Deep Cooperative Neural Networks model. The model applied two parallel CNN, the first focusing on learning user behaviors' by exploiting reviews given by the user and the other one focusing on learning item features from the reviews. To accomplish the prediction task a shared layer is placed on the top to the two paralleled CNNs together. The shared layer assists latent factors to represent both users and items interaction with each other like factorization machine techniques.

Recently, items aspect in user text review utilized as a contributing source of knowledge for enhancing rating prediction accuracy of CF (Da'u, Salim, Rabiu, & Osman, 2020; HUANG, JIANG, Wu, & Wang, 2020). For instance, (Aslanyan & Frasincar, 2021; Da'u et al., 2020), intended to integrate aspects of opinion presented in users review into CF. A multichannel CNN that involves word embedding and part of speech (POS) tag embedding layers is used to extract aspects from the user's review. A Latent Dirichlet allocation (LDA) topic modeling method is used to group and fitting the importance of extracted aspects. After that, a lexicon is used for generating the aspects rating matrices, the new matrices are then weighted and combined with a classical rating matrix into a Latent Factor Model (LFM) for rating prediction. (Ayman S. Ghabayen & Basem H. Ahmed, 2021) utilized clustering technique on user review text to better grasp the best similar users, for enhancing the accuracy of CF.

A series of recent research has mentioned that the review sentiment provides a vital source of information for enhancing the rating prediction accuracy in CF. Despite this, most of these studies focusing often on prediction accuracy as the only important factor. On the other hand, few studies have tried to fully exploit it for better-personalized recommendation quality. Therefore, the main aim of this study is to enhance the recommendation quality of the CF recommender system by utilizing the overall opinion of the user's review.

proposed method

This section presents the proposed method to perform CF Top-N recommendation based on the user's review opinion and the three numerical ratings on items. Figure 1. Presents the proposed approach. The proposed approach consists of two steps. Step1: used for reviews sentiment analysis model based on deep learning GRU and Step2: combine the sentiment polarity with the traditional CF recommendation. In the following subsections, we provide an overview of the proposed approach.

1) Review Sentiment analysis

This step consists of the sentiment prediction model for user's reviews text, the model will classify the users' reviews into positive or negative sentiment. The proposed sentiment analysis model utilized the deep learning bidirectional gated recurrent unit (Bi-GRU) model.

Traditional Sentiment analysis methods are mainly based on two approaches, the first one is dictionary-based and the other is machine learning. The main idea of the dictionary-based is obtaining a large number of opinion words, it gives by weighting algorithm, then building a sentiment dictionary. So, the sentiment score of any document or sentence is computed using a specific formula depends upon the sentiment dictionary weights. On the other hand, the machine learning methods depend on extracting the text features that exist on the document or the sentence, then trains the classifier model with corpus text data to obtain sentiment classification. The most common utilized features include part of speech POS (Ghabayen & Ahmed, 2019) features, TF-IDF (Lu & Wu, 2019) , Naive Bayesian (Bouchlaghem, Elkhelifi, & Faiz, 2016) and SVM (Lu & Wu, 2019) .

Recently, sentiment analysis methods begin to gradually transit to deep learning methods such as the convolutional neural network (CNN) method and recurrent neural network

(RNN) method. This study utilizes the RNN methods. RNN is a type of recurrent neural network, that uses sequence data as input and recursively links all nodes in the direction of sequence progression, which forms a directed cycle of feedback loop. It commonly comprises three layers input, hidden, and output layer. In this way, RNN can remember the previous calculation of information and can reuse it by utilizing it to the next variable in the sequence of inputs.(Hochreiter & Schmidhuber, 1997) proposed the long short-term memory (LSTM), which is a special type of RNN. LSTM uses the long memory as an input of activation function located in the hidden layer of the RNN. The work of (Abdi, Shamsuddin, Hasan, & Piran, 2019) proposed a sentiment analysis method using the LSTM deep learning method to get over the drawback of the classical RNN method. GRUs (Dey & Salem, 2017) is an alternative version of RNN. It strengthens the short-term memory with long-term memory through delicate gate control, and |handles the problem of gradient disappearance to a certain level. Following the previous works (Ahmed & Ghabayen, 2020; L.-x. Luo, 2019) the GRU model achieves a higher accurate text classification result compared with the state of art Long short-term memory network (LSTM), also, GRU is easier to train than LSTM. Therefore, this study utilized the GRU model for the sentiment analysis step. The architecture of the sentiment analysis model based on Bi-GRU is depicted in Fig. 2. Different from previous studies by (Ahmed & Ghabayen, 2020). A new attention layer is added after the hidden layers of Bi-GRU. Hence, the attention mechanism is utilized as in (Zhou et al., 2016). The attention layer is applied to extract the most informative words that reflect the sentiment of the sentence. For more details about the usage of other layers presented in this model, see (Ahmed & Ghabayen, 2020).

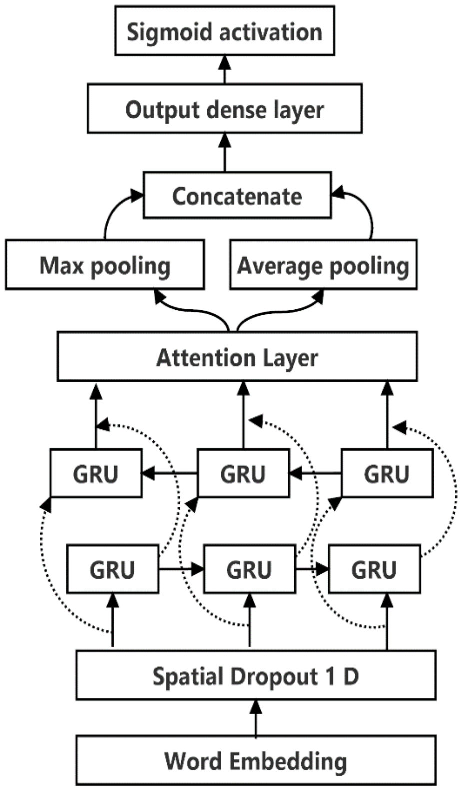

Fig. 2 Sentiment analysis model based on GRU.

2) Integration of Reviews polarity with Actual Rating

User overall polarity (negative/positive) of users' reviews obtained from the previous step (bi-GRU sentiment analysis model) usually contains the opinion of the user towards different items.

The next figure shows the steps of the Integration of Reviews polarity with Actual Rating. First, the number of items to be recommended to the target user, the User to recommend items to u, the List of all Items, and the User -Item matrix of rating R are passed to the algorithm. In the next step, the algorithm finds the nearest neighbor user k to the target user based on the user's previous numerical rating. In the next step, the algorithm re-rank the recommendation list-based user review polarity on an item Polarity Bias (incentive or penalty). Finally, a list of top N items is recommended to the target user.

Proposed approach Algorithm

Input: Number of items to be recommended $N \in$ N,

Number of items to be recommended $k \in$N

User to recommend items to u

Output: N items to be recommended

foreach *item* ∈ ***Items*** do
{
If *item* ∉ u, *rated-items* **then**
    Item rank-set ← rank -according -to-nearest-neighbors (k, u, item)
descending-rank-sort(*items*)
}
foreach *item* ∈ ***Item rank-set*** do
    {
    check user review polarity on an item.
    // Polarity Bias (incentive or penalty)
if positive **then**
 rating = rating + 1
 else
 rating =rating- 1
}
return top (*N*, Items)
Afterward, the review's polarity is then fed into the CF framework. Based on the polarity of the review of users on items in the ranked items resulted -

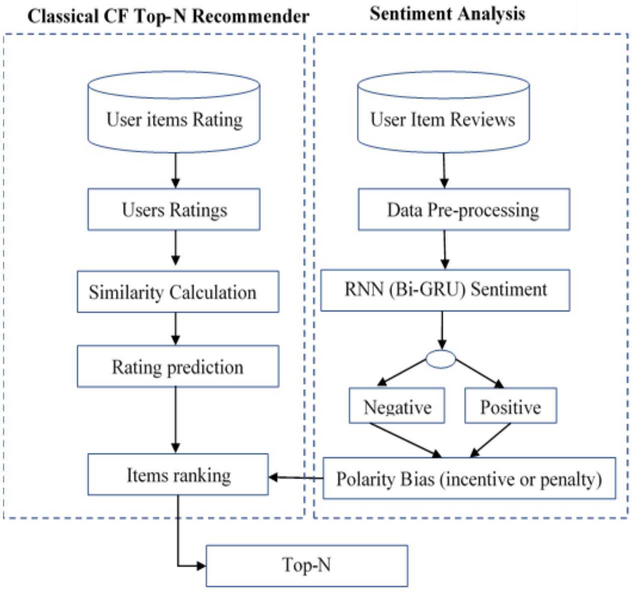

Fig 1. Proposed Recommendation Approach

from the traditional CF task, each item on the ranked list is affected by a Polarity bias according to its cross-ponding review polarity results from step1. Hence, an incentive bias is given to items with positive polarity on the other hand penalty is given to items with a negative polarity.
Experiments

The purpose of this set of experiments is to examine that the integration of collaborative filtering and user's review sentiment can generate more accurate recommendations compared to only traditional collaborative filtering systems based on users' numerical ratings. All the programs were written in Python 3.7, running on a computer with Intel(R) Core(TW) i7-4790K CPU, 4.00 GHz, 32GB of RAM.
This section presents the methodology and metrics used in the experiments and different experiment details on the Amazon dataset and the comparisons with baseline approaches. The finally the experimental results.

1) Datasets

The well-known Amazon dataset is used to study the efficiency of the proposed approach. due, to the huge size of the Amazon dataset five real-life datasets (Books, Mobile applications, Music, Office products, and PC) are selected from the Amazon dataset to train and evaluate the proposed approach. The datasets include user id, user numerical ratings with a scale from 1 to 5, and user reviews text. The statistics of Amazon datasets are presented in Table 1

Dataset	**#Users**	**#Items**	**#Reviews**
Books	4608044	2264749	9292094
Mobile Apps	2348575	126758	4607913
Music	1937864	781329	4494733
Office Products	1789171	313144	2421350
PC	4055621	441912	6385034

Table 1. Description of Amazon Dataset.

2) Methodology and Metrics

At first, to train the proposed model on review texts and user's star rating from the same user, the data need to be pre-processed and cleaned. The review text was cleaned by removing format, punctuation, and extra whitespace. Besides, stop word with no information value but appears common in a text, were also removed. Finally, all review text in the dataset is converted to lowercase, so there is no need to pre-process uppercase letters. As in previous work, each dataset was randomly sampled into five-folds, a portion of 80% considered as training set and the remaining 20 % considered as test data. This procedure is repeated five times and finally, reports the average results to guarantees that recommendation results are not biassed toward a certain test/ training set.

3) Classification Metrics

The well-known standard evaluation metrics Precision and Recall and F-measure are used on the test set to report the performance of the proposed approaches. The detailed calculation of the used metrices are given as follows:

$$precision = \frac{tp}{tp+fp} \quad (8)$$

$$Recall = \frac{tp}{tp+fn} \quad (9)$$

$$F1 - meaure = 2 \times \frac{Rcall \times Precision}{Rcall + Precision} \quad (10)$$

4) Comparisons

To study the performance of the proposed approach compared to the state of the art CF approach, we compare the proposed method with the Traditional K-Nearest Neighborhood CF (KNN hereinafter) (Deshpande & Karypis, 2004), Singular Value Decomposition (SVD hereinafter) (Sarwar, Karypis, Konstan, & Riedl, 2000) and Non-negative matrix factorization (NMF hereinafter) (X. Luo, Zhou, Xia, & Zhu, 2014).

- **KNN** is the classic rating-based CF recommendation technique that utilizes user rating to recognize the group of neighborhood users. This model utilized the cosine or Pearson's correlation coefficient to measure the similarities between users. Hence, the items are recommended to the target user based on Top-ranked items rated by target uses neighbors.
- **SVD** this approach considered user ratings of both similar users and other users who considered not similar. In such cases, Numerous users getting to be predictors for other user preference events without any similarities between them and the target users. Finally, the missing rating is prefilling with the rating data statistics. This model has been considered as a baseline approach in several studies in the literature (Barathy & Chitra, 2020; V. X. Chen & Tang, 2019; Xian, Li, Li, & Li, 2017).
- **NMF** a model based on rating matrix manipulating. The idea of this model is based on the dimensionality of the user-item rating matrix to a low-dimensional space. Then the similarities between users are calculated in the new space. The difference between the NMF model and the SVS is investigating a non-negative update process based on each disturbed feature parameter instead of the whole feature matrices.

Results and discussion

This section presents the experimental results concerning the Top-N recommendation quality using the evolution metrics Recall, Precision, and F1-Measure on various values of Top-N recommendation. Consequently, the values of N are (5,10,15,20,25,30) for each user in the test set. Concerning the value of K similar neighbor; depending on subsequent findings most baseline approaches achieve higher performance at K equal to 20, therefore this value has been considered in experiments.

The following tables show the performance comparisons of the proposed approach refer as Prop with all the baseline methods in terms of the recommendation quality metrics precision, recall, and F1-measure on the Amazon dataset. The result of Books, Mobile Apps, Music, Office Products, and PC are presented in Tables 2,3,4,5, and 6 respectively. Hence, the best prediction result is shown in bold.

The experimental results in Tables 2 demonstrate that the proposed approach shows the best performance compared to the baseline approaches. The result shows that the recommendation quality is improved with the increment of the number of Top-N recommended items in terms of recall in all experiments, however, the value of precision is gradually decreasing. The reason for this result is that with the higher value of Top-N items more false positives recommendation will arise, and thus causing lower precision performance, but true positive recommendation resulting in increment of the recall performance. This finding is common in recommender systems. Regarding the F1-Score metric, which combines precision and recall, it also shows the best results, this means that the proposed approach reflects the behavior expressed in the corresponding users rating, whether that means the user is interested in the items or not. For instance, in Amazon dataset is a huge dataset compared with the other dataset.

		Precision				Recall				F1			
	Top N	**KNN**	**NMF**	**SVD**	**Prop**	**KNN**	**NMF**	**SVD**	**Prop**	**KNN**	**NMF**	**SVD**	**Prop**
Amazon Books	**5**	93.8	94.3	94.2	**97.6**	51	50.7	50.7	**53.1**	66.1	66	65.9	**68.7**
	10	93.7	94.1	94	**95.8**	78	76.6	77	**79.6**	85.2	84.4	84.6	**86.9**
	15	93.7	94	93.8	**94.8**	89.4	87	87.9	**90.3**	91.5	90.3	90.8	**92.5**
	20	93.7	93.9	93.8	**94.4**	94	91.3	92.4	**94.7**	93.9	92.6	93.1	**94.5**
	25	93.7	93.9	93.8	**94.2**	96.2	93.2	94.5	**96.8**	94.9	93.6	94.1	**95.5**
	30	93.7	93.9	93.7	**94.1**	97.3	94.2	95.6	**97.8**	95.5	94	94.7	**95.9**
Amazon Mobile Apps	**5**	84.3	85.2	84.3	**88.1**	79.8	79.7	79.8	**83.5**	82	82.4	82	**85.7**
	10	84.1	84.9	84.1	**87.7**	83	82.9	83	**86.6**	83.5	83.9	83.5	**87.2**
	15	84.1	84.9	84.1	**87.6**	83.4	83.4	83.4	**87.1**	83.8	84.1	83.8	**87.3**
	20	84.1	84.9	84.1	**87.6**	83.5	83.5	83.5	**87.2**	83.8	84.2	83.8	**87.4**
	25	84.1	84.9	84.1	**87.6**	83.6	83.6	83.6	**87.2**	83.8	84.2	83.8	**87.4**
	30	84.1	84.9	84.1	**87.6**	83.6	83.6	83.6	**87.3**	83.8	84.2	83.8	**87.4**
Amazon Music	**5**	93.8	93.5	93.5	**94.8**	82.2	83.8	82.8	**85.2**	87.6	88.4	87.8	**89.8**
	10	93.6	93.3	93.3	**94.4**	89.3	91.3	90.8	**92.6**	91.4	92.3	92	**93.5**
	15	93.6	93.2	93.3	**94.3**	91.1	93.3	92.9	**94.5**	92.3	93.2	93.1	**94.4**
	20	93.5	93.2	93.	**94.**	92	94.2	93.	**95.**	92.8	93.7	93.	**94.**

				2	**2**			8	**4**			5	**8**
	25	93.5	93.2	93.2	**94.2**	92.4	94.6	94.3	**95.9**	93	93.9	93.7	**95**
	30	93.5	93.2	93.2	**94.1**	92.7	94.9	94.5	**96.1**	93.1	94	93.9	**95.1**
Amazon Office Products	**5**	92.2	92.8	92.5	**93.4**	93.4	90.2	91.3	**94.1**	92.8	91.5	91.9	**93.8**
	10	92.3	92.8	92.4	**93.1**	98.1	94.5	95.6	**98.4**	95.1	93.6	94	**95.7**
	15	92.2	92.8	92.4	**93.1**	98.7	95	96.2	**98.9**	95.4	93.9	94.2	**95.9**
	20	92.3	92.8	92.4	**93**	98.8	95.1	96.3	**99**	95.4	93.9	94.3	**95.9**
	25	92.3	92.8	92.4	**93**	98.9	95.1	96.3	**99.1**	95.5	93.9	94.3	**96**
	30	92.3	92.8	92.4	**93**	98.9	95.2	96.4	**99.1**	95.5	93.9	94.3	**96**
Amazon PC	5	89.5	89.5	89.8	**91.1**	86.5	89.3	86.1	**91.4**	88	89.4	87.9	**91.3**
	10	89.4	89.4	89.7	**90.8**	89.5	92.5	89.9	**94.5**	89.4	90.9	89.8	**92.6**
	15	89.4	89.4	89.7	**90.7**	89.9	93	90.4	**94.9**	89.6	91.1	90	**92.8**
	20	89.4	89.4	89.7	**90.7**	90	93.1	90.6	**95.1**	89.7	91.2	90.1	**92.8**
	25	89.4	89.4	89.7	**90.7**	90.1	93.2	90.6	**95.1**	89.7	91.2	90.1	**92.9**
	30	89.4	89.4	89.7	**90.7**	90.1	93.2	90.6	**95.2**	89.7	91.2	90.2	**92.9**

Table 2. Amazon dataset results.

The proposed approach shows the best results in terms of recall and improvement 2.08% to 2.4 % also 3.22 % to 3.76 % in terms of Precision, regarding the F1-score it shows an improvement from 2.68% to 2.84. compared to the baseline approaches when the value of Top-N=5. The result shows that the progress of performance improvement has an inverse relation to the value of the Top-N values for the baseline and the proposed approach.

For example, the improvement with smaller values of N is larger than bigger values of N. One possible explanation for this result is that relevant items related to the target user will be retrieved and involved in the Top-N recommendation in case of high values of N. Furthermore, it’s interesting to see that the proposed approach is able to achieve better precision values compared to the other baseline approaches. This means that the purposed approach is able to rank and recommend relevant items to the users. Also, the result shows that recall values increased as the value of Top-N increased, which means that more relevant items are retrieved among the retrieved items.

Finally, the proposed approach result is compared with a recent work related to the filed (Ayman S. Ghabayen & Basem H. Ahmed, 2021) (Rev-clust hereafter) , which unitize user review clustering to enhance the CF accuracy. Table 3. present the experimental results on Amazon Book dataset.

	Precision		Recall		F1	
Top N	**Rev-clust**	**Prop**	**Rev-clust**	**Prop**	**Rev-clust**	**Prop**
5	98.1	**97.6**	52.9	**53.1**	68.7	**68.7**
10	96	**95.8**	78.9	**79.6**	86.6	**86.9**
15	94.9	**94.8**	98.7	**90.3**	92.2	**92.5**
20	94.4	**94.4**	94.2	**94.7**	94.3	**94.5**
25	94.1	**94.2**	96.4	**96.8**	95.2	**95.5**
30	94	**94.1**	97.4	**97.8**	95.7	**95.9**

Table 3. Amazon Book dataset results.

The result illustrates that the proposed approach archives better accuracy in terms of well-known evaluation metrics F1-measure, Recall, and Precision compared to the Rev-cluster approach. On the other hand, the performance complexity of Rev-clust is very high, due to the utilizing of text clustering opposite to the proposed approach.

In concluding the result shows that the proposed approach achieved the best result in different values of Top-N even with smaller values of N=5. This situation is completely appropriate in a real scenario since online users are normally attracted to the top few numbers of ranked items.

Conclusion and future work

This paper proposed to incorporate user reviews in developing a CF recommendation approach. Since, user reviews usually contain the opinion of the user towards different items, at first, the proposed approach employs Bi-GRU based on the attention layer to infer the sentiment polarity of user reviews text. The inferred overall sentiment polarity is then integrated into the CF recommendation process. Having the review polarity on items, the resulted ranked Top-N recombination list of the traditional CF approach is then affected by a Polarity bias according to its cross-ponding polarity. Hence, an incentive bias is given to items with positive polarity on the other hand penalty is given to items with a negative polarity. The experimental results on different Amazon datasets demonstrate, that the proposed approach improves the performance of the CF recommender system by integrating the sentiment polarity of user reviews in the recommendation process and produces recommendations with higher quality in terms of Recall, precision and F1-measure compared to the baseline CF methods. Also. The proposed approach allows the CF approach, which operates mostly on user numerical rating, to utilize user reviews as an additional source of user preferences. For future research directions, examining alternative review features to recognize the correlation between users reviews and users rating. Furthermore, including more types of item information, such as category, brand, and price to improve the recommendation quality.

References

Abdi, A., Shamsuddin, S. M., Hasan, S., & Piran, J. (2019). Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion. *Information Processing & Management, 56*(4), 1245-1259. doi:https://doi.org/10.1016/j.ipm.2019.02.018

Ahmed, B. H., & Ghabayen, A. S. (2020). Review rating prediction framework using deep learning. *Journal of Ambient Intelligence Humanized Computing*, 1-10.

Aslanyan, T. K., & Frasincar, F. J. A. S. A. C. R. (2021). LDA-LFM: a joint exploitation of review text and ratings in recommender systems. *21*(2), 33-47.

Babu, N. V., & Rawther, F. A. (2021). *Multiclass Sentiment Analysis in Text and Emoticons of Twitter Data: A Review.* Paper presented at the Second International Conference on Networks and Advances in Computational Technologies.

Barathy, R., & Chitra, P. (2020). *Applying Matrix Factorization In Collaborative Filtering Recommender Systems.* Paper presented at the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS).

Bouchlaghem, R., Elkhelifi, A., & Faiz, R. (2016). *A machine learning approach for classifying sentiments in Arabic tweets.* Paper presented at the Proceedings of the 6th international conference on web intelligence, mining and semantics.

Chen, C. C., Chen, Z.-Y., Wu, C.-Y. J. I. T. o. K., & Engineering, D. (2011). An unsupervised approach for person name bipolarization using principal component analysis. *24*(11), 1963-1976.

Chen, L., Chen, G., Wang, F. J. U. M., & Interaction, U.-A. (2015). Recommender systems based on user reviews: the state of the art. *25*(2), 99-154.

Chen, V. X., & Tang, T. Y. (2019). *Incorporating singular value decomposition in user-based collaborative filtering technique for a movie recommendation system: A comparative study.* Paper presented at the Proceedings of the 2019 the International Conference on Pattern Recognition and Artificial Intelligence.

Da'u, A., Salim, N., Rabiu, I., & Osman, A. J. E. S. w. A. (2020). Weighted aspect-based opinion mining using deep learning for recommender system. *140*, 112871.

Deshpande, M., & Karypis, G. J. A. T. o. I. S. (2004). Item-based top-n recommendation algorithms. *22*(1), 143-177.

Dey, R., & Salem, F. M. (2017, 6-9 Aug. 2017). *Gate-variants of Gated Recurrent Unit (GRU) neural networks.* Paper presented at the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS).

Gallege, L. S., & Raje, R. R. (2016). *Towards Selecting and Recommending Online Software Services by Evaluating External Attributes*. Paper presented at the Proceedings of the 11th Annual Cyber and Information Security Research Conference, Oak Ridge, TN, USA. https://doi.org/10.1145/2897795.2897797

Ghabayen, A. S., & Ahmed, B. H. (2019). Polarity Analysis of Customer Reviews Based on Part-of-Speech Subcategory. *29 %J Journal of Intelligent Systems*(1), 1535-1544.

Ghabayen, A. S., & Ahmed, B. H. (2021). ENHANCING COLLABORATIVE FILTERING RECOMMENDATION USING REVIEW TEXT CLUSTERING. *Jordanian Journal of Computers and Information Technology (JJCIT), 7*(02).

Ghabayen, A. S., & Ahmed, B. H. (2021). Enhancing collaborative filtering recommendation using review text clustering. *Jordanian Journal of Computers and Information Technology (JJCIT), 07*(02), 152-165. doi:10.5455/jjcit.71-1609969782

Hochreiter, S., & Schmidhuber, J. J. N. c. (1997). Long short-term memory. *9*(8), 1735-1780.

Hou, T., Yannou, B., Leroy, Y., & Poirson, E. (2019). Mining customer product reviews for product development: A summarization process. *Expert Systems with Applications, 132*, 141-150.

HUANG, C., JIANG, W., Wu, J., & Wang, G. J. A. T. o. I. T. (2020). Personalized Review Recommendation based on Users' Aspect Sentiment.

Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). *Recommender Systems: An Introduction*: Cambridge University Press.

Jeon, B., & Ahn, H. J. J. o. I. (2015). A collaborative filtering system combined with users' review mining: application to the recommendation of smartphone apps. *Journal of Intelligence, 21*(2), 1-18.

Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management, 90*, 523-537.

Kim, D., Park, C., Oh, J., Lee, S., & Yu, H. (2016). *Convolutional matrix factorization for document context-aware recommendation.* Paper presented at the Proceedings of the 10th ACM conference on recommender systems.

Li, H., Cui, J., Shen, B., & Ma, J. (2016). An intelligent movie recommendation system through group-level sentiment analysis in microblogs. *210*(C %J Neurocomput.), 164–173. doi:10.1016/j.neucom.2015.09.134

Liu, B., & Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 415-463). Boston, MA: Springer US.

Lu, K., & Wu, J. (2019). *Sentiment Analysis of Film Review Texts Based on Sentiment Dictionary and SVM*. Paper presented at the Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence, Suzhou, China. https://doi.org/10.1145/3319921.3319966

Luo, L.-x. (2019). Network text sentiment analysis method combining LDA text representation and GRU-CNN. *Personal and Ubiquitous Computing, 23*(3), 405-412. doi:10.1007/s00779-018-1183-9

Luo, X., Zhou, M., Xia, Y., & Zhu, Q. J. I. T. o. I. I. (2014). An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. *10*(2), 1273-1284.

Natarajan, S., Vairavasundaram, S., Natarajan, S., & Gandomi, A. H. (2020). Resolving data sparsity and cold start problem in collaborative filtering recommender system using linked open data. *Expert Systems with Applications, 149*, 113248.

Pero, Š., & Horváth, T. (2013). *Opinion-Driven Matrix Factorization for Rating Prediction*, Berlin, Heidelberg.

Raghavan, S., Gunasekar, S., & Ghosh, J. (2012). *Review quality aware collaborative filtering.* Paper presented at the Proceedings of the sixth ACM conference on Recommender systems.
Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender Systems: Introduction and Challenges. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 1-34). Boston, MA: Springer US.
Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). *Application of dimensionality reduction in recommender system-a case study*. Retrieved from In Proceedings of the ACM WebKDD 2000 Web Mining for E-Commerce Workshop:
Shen, R.-P., Zhang, H.-R., Yu, H., & Min, F. J. E. S. w. A. (2019). Sentiment based matrix factorization with reliability for recommendation. *135*, 249-258.
Terzi, M., Rowe, M., Ferrario, M.-A., & Whittle, J. (2014). *Text-Based User-kNN: Measuring User Similarity Based on Text Reviews.* Paper presented at the User Modeling, Adaptation and Personalization: 22nd International Conference, UMAP 2014, Aalborg, Denmark, July 7-11, 2014. Proceedings.
Wang, H., & Luo, N. (2014). Collaborative filtering enhanced by user free-text reviews topic modelling.
Wang, Y., Wang, M., & Xu, W. (2018). A Sentiment-Enhanced Hybrid Recommender System for Movie Recommendation: A Big Data Analytics Framework. *Wireless Communications and Mobile Computing, 2018*, 8263704. doi:10.1155/2018/8263704
Xian, Z., Li, Q., Li, G., & Li, L. (2017). New Collaborative Filtering Algorithms Based on SVD++ and Differential Privacy. *Mathematical Problems in Engineering, 2017*, 1975719. doi:10.1155/2017/1975719
Zheng, L., Noroozi, V., & Yu, P. S. (2017). *Joint deep modeling of users and items using reviews for recommendation.* Paper presented at the Proceedings of the Tenth ACM International Conference on Web Search and Data Mining.
Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., & Xu, B. (2016). *Attention-based bidirectional long short-term memory networks for relation classification.* Paper presented at the Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Berlin, Germany.
Zucco, C., Calabrese, B., Agapito, G., Guzzi, P. H., & Cannataro, M. (2020). Sentiment analysis for mining texts and social networks data: Methods and tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(1), e1333.