Toward Improving the Prediction Accuracy of a Product Recommendation System Based on Word Sequential Using LSTM Embedded

Jaeni Jaeni¹,², Purwanto Purwanto³, Budi Warsito⁴ and Adi Wibowo⁵

¹ Doctoral Program of Information System, School of Postgraduate Studies, Diponegoro University, Semarang 50275, Indonesia
² Department of Informatics Management, University of Amikom Yogyakarta, Jl. Ringroad Utara Condoncatur, Sleman, 55283, Indonesia
³ Department of Information Systems, School of Postgraduate Studies, Universitas Diponegoro, Semarang
⁴ Department of Statistics, Faculty of Sciences and Mathematics, Diponegoro University, Semarang
⁵ Department of Informatics, Faculty of Science and Mathematics, Diponegoro University, Semarang

Abstract. The ability to predict purchases is crucial for e-commerce decision makers when making offers and suggestions to customers. In the development of recommendation models, two common problems often encountered are a lack of personalization and irrelevant recommendations. To address these issues, it is crucial to consider user history data, such as the user's interactions with previous products. This allows the model to learn user preferences from the past and generate more personalized and relevant recommendations. In this study, word2vec is used to provide rating predictions, while document context is enhanced using LSTM capture contextual understanding of product reviews. This study makes use of an actual dataset to test our model using an Amazon Review Dress. The results of our investigation demonstrate that, as 88% of the recommendations are made in accordance with the recommendation system's criteria, it can be considered that it offers reasonably accurate and dependable recommendations with an accuracy of 0.8752

1 Introduction

Consumers' experiences when purchasing online are greatly improved by e-commerce recommender systems. In order to deliver individualized product suggestions, these systems are built to assess user behavior, preferences, and historical data [1]. E-commerce recommender systems successfully offer products that are most likely to be of interest to certain users by utilizing machine learning algorithms and data mining approaches, hence boosting customer engagement and happiness.

An e-commerce recommender system's primary goal is to solve the information overload issue that customers frequently experience when browsing through extensive online

* Corresponding author: jaeni@amikom.ac.id jaenishuri@students.undip.ac.id

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
product catalogs. Users find it difficult to locate products that match their unique needs and tastes due to the abundance of options accessible. By automatically selecting and rating items based on user profiles and previous interactions, recommender systems try to solve this problem [2].

Several crucial elements are often required for an e-commerce recommender system to operate well. The first step in the process is data collecting, which entails learning more about users' interactions with the e-commerce site. This can contain information about previous purchases, product reviews, search terms, and browsing habits. Following processing, the data is evaluated to find trends, correlations, and user preferences [3]. On Fig. 1 below, you can see an illustration of a presentation of a product recommendation.

![Fig. 1 E-commerce for women dress online shopping](image)

In the past, predicting how many stars, out of a possible 5, After watching a video, a viewer would give it an award. This is how the Netflix suggestion conundrum was explained. We did, in fact, rely heavily on such an algorithm when sending DVDs was our main line of business, in part because the only feedback we got from members at the time that they had really seen the movie was a star rating. In order to improve the accuracy of rating prediction, we even launched a competition. The winning algorithms are still employed by us in actual production [4].

This unique competition is owned by the Netflix organization. They will pay $6 million to the person who successfully increases the performance of the current Netflix engine by more than 10%. Most academics, practitioners, and researchers are exploring employing a unique collaborative filtering model based on the well-known latent factor technique known as matrix factorization. Yet, in the early 2000s, Sarwar discovered the latent component for the first time [5].

An effective way of targeted marketing, recommendation algorithms provide each customer with a unique shopping experience. Any user, regardless of how many purchases or ratings they have made, will find the recommendations produced by a good recommendation algorithm compelling. A good recommendation algorithm reacts quickly to changes in a user's data, can scale for large retailers like Amazon.com over extremely large customer bases and product catalogs, and produces recommendations that are relevant to all users. Comparatively to other approaches, collaborative filtering from item to item is more effective at solving this issue. [6].

A lack of personalization and recommendations that are irrelevant are two issues that frequently arise while developing recommendation models. User history information, such as the user's interactions with past products, must be taken into account in order to address these problems.

2 Literature Review

E-commerce strategists can benefit from a superb machine learning categorization
model called purchase prediction that can help them take the appropriate actions for their future goals, like stock control. Additionally, examining user activity might make it possible to forecast whether or not a sale will occur during the next session. Purchase prediction has been connected with RS in order to anticipate whether a purchase will be made during the session and to produce suggestions based on the items accessed during the session [7].

In e-commerce recommender systems, collaborative filtering, content-based filtering, and hybrid methods are frequently employed. Collaborative filtering systems evaluate user behavior and preferences by looking for links and similarities between individuals or objects. Content-based filtering concentrates on the features and characteristics of the individual items when producing recommendations. Hybrid techniques combine collaborative and content-based tactics to maximize the advantages of both strategies. [3].

One of the most well-known and often utilized sorts of suggestions is collaborative suggestion, which is particularly well-liked in e-commerce. Without requiring explicit information on the users or the products, collaborative filtering approaches [8] result in user-specific management for products based on rating or purchasing things. To create a matrix of user goods based on the opinions of comparable consumers with the same or similar interest preferences, collaborative filtering uses user ratings. In order to create a comparable user group known as a neighborhood, the system compares user profiles to identify people with similar attitudes and interests. CF is a tried-and-true approach that is successful.

Google endorsed Word2Vec for natural language processing in 2013. It is a method for enrasing words in vector spaces to encode them as vectors. These techniques are effective in NLP and have been used to several studies, including those on recommendation systems. Word2Vec is paired with item-based CF to offer Item2Vec-based recommendation services. Without user data, Item2Vec might not be able to evaluate connections between objects. [9].

It's crucial to capture a document's contextual awareness of product reviews. Several algorithm models have been created in the previous five years to get research, including word embedding based on word2vec, quickest, glove, etc. Bidirectional word vector representation has been used in many models in recent years [10].

A deep learning model that consists of neurons is known as an RNN or recurrent neural network (RNN). Since each neuron can use its internal memory to store data about the previous input, it is most helpful when thinking about sequential data. This functions more like a loop where a neuron's output at one step is sent back to it as input at a later level. Although it looks that this is one neuron's output working as another neuron's input, this is actually just one neuron carrying out both functions [11].

Word embedding, particularly enhanced ones, are essential for capturing the semantic meaning of words. The system can get more precise and nuanced representations of words by applying advanced word-embedding approaches, such as pre-trained embedding like Word2Vec or GloVe, or by exploiting contextual embedding like BERT or ELMo. These enriched embedding can gather contextual data and enhance how well the recommendation algorithm comprehends user preferences and product descriptions [12] [13]. Before LSTM can employ a distributed word representation, each text's tokens must be transformed into a matrix of vectors. In a supervised learning setting, a specific label is also given to the training data.

Since document representation can be included into latent factor models like BOW, LDA, and CTR, the traditional NLP models were well-known for this capability early in the 2000s. Nevertheless, over the past five years, some researchers have used deep learning models including CDL, CNN, LSTM, and the Attention Model to improve document contextual awareness. According to the aforementioned literature, this work makes a number of important advances, including: Unlike earlier experiments, which did not exploit the
sequential component of the LSTM's attention mechanism LSTM and attention are used in a sequential-to-sequential contextualization method for product documentation. [14 -6].

Usually, a deep learning or machine learning model serves as the input model. The text is divided into attributes that may be entered into the associated model after being vectorized. CNN and LSTM are common components of standard deep learning models [12]. Model classification: The model classification is quite straightforward, the two categories are frequently classified using the sigmoid method, and many classifications are classified using the softmax approach.

In order to extract the most crucial essential elements of the text, learn the local information, and learn the global information, we incorporated the Word2Vec model, the RNN model, the LSTM model, and the TF-IDF algorithm in this study. A memory cell with long-term data storage is added using the LSTM technique. The TF-IDF word weighting technique is employed in this study because it distinguishes between relevant and irrelevant words in addition to counting the number of words in a document. This study's objectives and contribution are to apply the LSTM approach to contextual text, weight each word in the text using TF-IDF, evaluate the method's efficacy, and improve the efficacy of prior classification-related research.

3 Theoretical Backgrounds

A business model known as "e-commerce" includes doing transactions through digital networks, particularly the Internet. The e-commerce business model is performing transactions through electronic networks, particularly the Internet, in order to exchange products, services, and information. It involves the electronic exchange of goods, services, and information. is a type of company that involves carrying out transactions using digital networks, primarily the Internet. [17].

In order to address the aforementioned concerns, recommendation algorithms are used. Recommendation systems provide users with personalized recommendations. It is common to employ recommendation algorithms for things like music, literature, movies, and shopping. The vast majority of online retailers now have a recommendation engine. The items that the user has already chosen, the things that people who are similar to them prefer, or the item description are used to provide recommendations [18].

The primary categories of the algorithm used to develop recommender systems for e-commerce are as follows: We separated the recommender system into five categories: context awareness, collaborative filtering, content-based filtering, hybrid filtering, and knowledge-based recommender approaches. According to several literature [19], collaborative filtering is the most effective algorithm to produce product recommendation. Collaborative filtering utilizing historical data on user behavior, particularly rating information. A rating is a reflection of how happy you are with a service or a product. Collaborative filtering is therefore more effective than other algorithms, such as content-based ones that take product characteristic calculations into account. [20 - 23].

The three essential requirements for text classification-oriented models in machine learning and deep learning are text vectorization, the input model, and model classification. First, the length and consistency of the format utilized to present the material vary. Text processing is necessary. Count Vectorization, TFIDF Vectorization, Word2Vec, GloVe, and FastText are the techniques that are now in use. The first two are chosen because they retain the text's basic word frequency characteristics while eliminating the relationship between each word's context and meaning. [24].
4 Method

Get a real dataset by using Amazon Review Dress. Preprocessing is the method used to handle text data, and it involves tokenizing, stopword elimination, and stemming. Tokenizing is used to condense sentences into a single word, stopword removal to eliminate terms that are unrelated to the primary word, and stemming to isolate core words from processed words. The data is then utilized to extract each word's vector value using the TF-IDF word weighting technique. The LSTM approach was used to classify the word-weighting and preprocessing findings. The TF-IDF word weighting algorithm is then used to extract each word's vector value from the data. The results of the word-weighting and preprocessing were categorized using the LSTM method. In Fig. 2, the research technique is displayed. To determine accuracy, precision, and recall, the confusion matrix computations had to be carried out after the classification results were acquired. Performance is measured using a confusion matrix, which contains various factors.

![Fig. 2 Design of experiment process](image)

Preprocessing is done at this stage to speed up the data processing. The four preprocessing procedures used in this study are tokenizing, stopword omission, stemming, and deduplication. The first stage is tokenization. The first step in the preprocessing procedure, tokenizing, is used to separate individual words from letters or phrases in a text. Characters can be separated from one another using whitespace, such as tabulations, enter, and spaces. The next stage is to get rid of stopwords. The last phase, stemming, includes employing stopwords to eliminate less important words that have no connection to the main term. Stemming is a method for locating or changing words into their most fundamental versions that yet have the same meaning. The outcomes of the stemming procedure are shown in Table 1.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>[like, reviewers, hesitant, spend, much, pair,...</td>
</tr>
<tr>
<td>1</td>
<td>[true, bunch, fall, clothing, photos, colors, ...</td>
</tr>
<tr>
<td>2</td>
<td>[wanted, skirt, work, love, design, way, way, ...</td>
</tr>
<tr>
<td>3</td>
<td>[love, love, hesitant, buy, first, reviews, ma...</td>
</tr>
<tr>
<td>4</td>
<td>[absolutely, love, retro, look, swimsuit, firs...</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>19657</td>
<td>[feels, soft, like, good, quality, however, re...</td>
</tr>
<tr>
<td>19658</td>
<td>[oot, dress, blue, fits, greathits, knee, shor...</td>
</tr>
<tr>
<td>19659</td>
<td>[patient, dress, waiting, almost, forever, til...</td>
</tr>
<tr>
<td>19660</td>
<td>[deep, doesnt, gape, flatters, neckline, waist...</td>
</tr>
<tr>
<td>19661</td>
<td>[saw, dress, online, morning, went, store, aft...</td>
</tr>
</tbody>
</table>

The TF-IDF method promotes giving preference to dependent terms over commonly used phrases in a text. For instance, when classifying articles, terms like "governing" or
"white house" that are frequently used in the category of "politics" impact the judgment more than words like "this" or "the," which are more frequently used. TF-IDF uses both the frequency of a term in a specific document and the frequency of the same phrase throughout the whole document collection for this purpose [25].

This approach calculates word weights by summing the frequency of a word in a text with the inverse frequency of documents containing words that appear in the text. The inverted Document Frequency (IDF) and Term Frequency (TF) values for each document will be determined by the TF-IDF technique [26]. To calculate the weight of each term \( t \) in the document \( d \), eq. (1)

\[
W_{dt} = TF_{dt} \times IDF
\]

Recurrent neural networks, often known as RNNs, are a form of artificial neural network in which the output data from one phase is also utilized as the input data for the following phase. The major problem with RNNs, on the other hand, is that they typically run into gradient vanishing and exploding issues during back propagation [27]. Hochreiter and Schmidhuber developed the Long Short-Term Memory (LSTM) in 1997 to solve this problem. Recurrent neural networks may learn information from earlier time steps and apply it to subsequent time steps, however LSTM networks are a modified version of these networks. LSTM networks use cell states to transmit data rather than the more conventional feedforward neural networks. LSTMs are able to selectively recall and exclude information in this fashion. RNNs' gating mechanism allowed them to overcome the "short-term memory" problem. An LSTM unit is made up of a cell, an input gate, an output gate, and a forget gate, which is the last gate in the chain. The three gates continuously control the flow of data into and out of the cell, which maintains data for an arbitrary amount of time. (Fig. 3) [28].

The LSTM and the categorical feed-forward neural network are examples of neural network subclasses. The LSTM method offers the benefit of accounting for the connections between the processes of earlier information sessions and the current session. This is crucial for capturing the context of a phrase in a document, according to NLP. The input layer, output layer, hidden state, and previous process are connected to a number of hidden stages in the LSTM. Several critical calculation phases in the concealed stage provide the advantage of LSTM to begin the sequential aspect.

Each of the three gates that make up a memory cell in the hidden layer—input gate, forget gate, and output gate—consists of these three components. The input gate determines how much data should be stored in the cell state. The cell can no longer store any additional data as a result. Forget gates are used to control how long a value remains in a memory cell. The output gate establishes the percentage of a memory cell's data or value that will be used to compute the output.

![Fig. 3 Basic structure of the LSTM model](image-url)
Figure 4 indicates that the first step is to input sentences that have been processed at the preprocessing stage into word weighting using the TF-IDF technique to transform words into vectors and then into TF-IDF arrays. The input layer of the LSTM technique is then given the results of the TF-IDF weighting. After that, the word will be added to the LSTM layer. The categorization of emotion classes occurs in the last stage when the words are reunited into sentences in a completely connected layer.

![Figure 4 LSTM architecture](image)

**4.1 Evaluation**

The performance of the classification method is assessed in this research using a confusion matrix. To evaluate the value of the classification performance using LSTM, the model will be put to the test using data. The accuracy, precision, and recall numbers serve as a representation of the performance outcomes. The degree to which the predicted value and the actual value are similar is a measure of accuracy. [29].

\[
\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+FN+TN)}
\]  

(2)

The accuracy of the data from prediction and accuracy is what determines precision. Precision can be determined by

\[
\text{Precision} = \frac{(TP)}{(TP + FP)}
\]  

(3)

While a model's recall rate indicates how well it recognizes a class, Recall can be calculated using (4).

\[
\text{Recall} = \frac{TP}{(TP + FN)}
\]  

(4)

According to the formula, "True Positive" (TP) denotes positive data that is accurately predicted, "True Negative" (TN), negative data that is similarly predicted, "False Positive" (FP), negative data that is predicted as positive data, and "False Negative," positive data that is similarly predicted as negative data.

**5 Result and Discussion**

The model's confidence level is determined by examining the accuracy, precision, and recall after it has generated its predictions for each class. The objective is to determine the accuracy percentage that may be relied upon for a class prediction model. Table 2 displays the results in comparison. In the context of a confusion matrix, "true labels" refers to the labels that are actually assigned to each sample in the dataset or to the labels that should be. This is the
fundamental data that is utilized to create a confusion matrix. The true label displays the sample's actual target class. The confusion matrix describes the number of true (true) and false (false) predictions made by the model.

**Table 2.** Results of comparison of model results

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>TF-IDF</th>
<th>Word2Vec</th>
<th>RNN +LSTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>accuracy</td>
<td>0.888</td>
<td>0.859</td>
<td>0.8752</td>
</tr>
<tr>
<td>precision</td>
<td>0.853</td>
<td>0.791</td>
<td></td>
</tr>
<tr>
<td>recall</td>
<td>0.751</td>
<td>0.698</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5.** Model Sequential

Decision-makers must be able to forecast future purchases when making offers and suggestions to customers through online commerce. To increase prediction accuracy in this study, we added product similarity as a component of classification models. A reminder design, forgetting the result so that it may be processed again as input, and more accurate data processing are all advantages of the LSTM model.

The preprocessing stage's processed sentences are first added to word weighting using the TF-IDF method, which first turns words into vectors and then TF-IDF arrays. The outcomes of the TF-IDF weighting are then provided to the input layer of the LSTM algorithm. The term will then be included in the LSTM layer after that. After the words have been put back together into sentences in a completely connected layer during the last stage, the categorizing of emotion classes takes place.

Words are represented as vectors in word embedding, which takes into account both the word's context and the surrounding words in the phrase. The word2Vec technique has two main word embedding strategies. Word embedding and deep learning models must be employed for better results. The goal of word embedding, an n-dimensional distributed representation of a text, is to capture the meanings of the words.
By substituting the product id for the word that makes up the phrase and the session for the sentence, we were able to calculate product similarity using the Word2Vec algorithm. Over the course of a session, we moved the products around to find the ideal configuration for the highest level of purchase prediction accuracy.

![Figure 6. Comparison Result](image)

![Figure 7. Model confusion matrix](image)

### 6 Conclusion and Future Work

Our experiment report show that model With an accuracy of 0.8752, it can be said that the recommendation system provides relatively accurate and reliable recommendations, as 88% of the recommendations are given according to user preferences.

LSTM is used in recommendation systems to model the sequence of items or user behavior from previous data. In the context of a recommendation system, LSTM can study patterns and dependencies that exist in user behavior, such as product sequences viewed or products purchased. Based on this understanding, LSTM can make relevant recommendations for the next item based on the user’s history.

Additionally, the model can gather contextual information from earlier words in sentences or messages by taking into account user history data. This makes it easier to understand user preferences and enables the model to produce recommendations that are more pertinent to the user. For instance, the model can take into account the preceding context to deliver more accurate recommendations in phrases or texts that contain several repetitions of the same term in various situations.

Recurrent neural network with long short-term memory (RNN LSTM) models provide a number of benefits in jobs involving sequence modeling, but they also have significant drawbacks. RNN LSTM models use sequential computation to process input, with each step...
depending on the preceding one. Due to the difficulty of parallelizing the training process due to this sequential processing, training times are longer than for alternative architectures, such as feed-forward neural networks.

We appreciate the support of our research from Universitas Amikom Yogyakarta and Universitas Diponegoro Semarang.

References

10. O. Barkan, ICML Microsoft 1 (2016)