Sentiment analysis based on deep learning in e-commerce

Ameni Chamekh[1], Mariem Mahfoudh[1,2], and Germain Forestier[3]

[1] ISIGK, University of Kairouan, 3100 Kairouan, Tunisia
[2] MIRACL Laboratory, University of Sfax, 3021 Sfax, Tunisia
amenichamekh959@gmail.com, mariemmahfoudh@gmail.com
[3] University of Hautes Alsace, 68093 MULHOUSE Cedex, France
germain.forestier@uha.fr

Abstract. Social media allow businesses to find out what customers are thinking about their products and to participate in the conversation. Companies, therefore, have an interest in using them to market their products, identify new opportunities and improve their reputation. The main objective of our study was to recognize feelings expressed in opinions, ratings, recommendations about a product using a construction based on a corpus of sentiment lexicon with different deep learning algorithms. In this work, we will then analyze an e-commerce platform in order to know the feelings of customers towards the products. This study is conducted based on a static dataset of 41,778 smartphone product reviews in french collected on Amazon.com. For the classification of reviews, we applied the Long short-term memory network (LSTM). The results showed that the LSTM deep learning algorithm yielded a good performance with an accuracy of 95%.

Keywords: Sentiment analysis · The e-commerce platform (Amazon) · Long short-term memory network (LSTM).

1 Introduction

Sentiment analysis has attracted a lot of attention in recent years. It is known as opinion mining and it is used by companies to identify what impression people have of their services and products through user reviews, tweets and comments on social media platforms [3]. These masses of data become essential and effective information to predict the consumer preferences and consumption trends of sentiment expressed in reviews, which benefits companies to improve their marketing strategies and products [7]. The fundamental purpose of sentiment analysis is to process human emotions, expressed in a text. Users freely express their opinions regarding the products they have already purchased. These reviews tend to become powerful tools that help customers analyze them for further reuse in their product purchases on e-commerce sites [9]. Sentiment analysis, is present in many areas such as politics, finance, and education. In this article, we address the field of e-commerce. The number of online purchases has been particularly increased in recent years. Therefore, e-commerce sites generate a lot of data every day in the form of customer reviews of products they have already purchased. User opinions form a kind of discussion with which they can interact in order to

This is the author's version of an article published at KSEM 2022. The final authenticated version is available online at: https://doi.org/10.1007/978-3-031-10986-7_40

have recommendations and advice on products or services. As well as, analyzing these opinions will help online retailers understand customer expectations, provide a better shopping experience and increase sales. The objective of our work is to automatically analyze the textual data extracted from the most popular e-commerce platform amazon in order to detect customer sentiment towards particular smartphone products. This, using a construction based on a corpus of feelings lexicon with the deep learning algorithm. In this study, we first extracted a corpus of static data from amazon, then we applied different natural language processing techniques such as tokenization, stopword removal and stemming, then we evaluated the performance and efficiency of the classification algorithm: Long short-term memory (LSTM). The performance result shows that LSTM performed with an accuracy of 95%. The advantage of our work over other works is that we extracted a massive amount of data from the amazon e-commerce site on smartphone products specifically using a scraping process and this helped us in our classification of reviews with LSTM since deep learning algorithms usually require a lot of data to train properly. Moreover, we opted for LSTM algorithm in our classification model, and we also achieved a very high accuracy rate compared to other previous works that used LSTM model. The rest of this paper is organized as follows: section 2 discusses related work. Section 3 describes in detail the proposed method. We report in section 4 our experimental results and give our conclusion on this work in section 5.

2 Related Work

[12] proposed a new sentiment analysis model SLCABG to improve the accuracy of sentiment analysis of product reviews. They combined the advantages of sentiment lexicon, CNN model, GRU and the attention mechanism for building the SLCABG model. The authors used the dataset consisting of book review data collected from Dangdang.com, it is a famous Chinese e-commerce website. After manually filtering product reviews by star rating, the authors had a data set of 100,000 reviews. To evaluate the performance of SLCABG model, [12] explored the impact of several factors such as: the length of the input text sentence, the size of the lexicon, the number of iterations of the model, the value dropout and the weighted word vector. The experimental results show that all these factors influenced the model. Thus, [12] compared the sentiment analysis effects of the SLCABG model with common sentiment analysis models (NB, SVM, CNN and BiGRU) on the dataset. The comparison results show that the classification performance of the proposed SLCABG model is effectively improved over the commonly used deep learning model. [6] presented a model based on a set of models; bi-directional LSTMs (Bi-LSTM), and convolutional neural network (CNN), one to capture temporal information from data, and the other to extract local structure to perform sentiment analysis. They used popular databases such as IMDB review and SST2 dataset to test the proposed model. The experimental results showed that this model can outperform the two individual models hence [6] observed some performance gain with 90% compared to the individual LSTM and CNN models. [11] introduced a new model, SenBERT-CNN for analyzing customer reviews. The SenBERT-CNN model combines a pre-trained bidirectional encoder representation network from transformers (BERT) with a convolutional neural network (CNN) to capture more sentiment information in sentences. So SenBERT-CNN is a pre-trained language representation model for the

e-commerce review domain. Specifically, [11] used the BERT structure to better express the semantics of the sentence as a text vector, and then they extracted the deep features of the sentence through a convolutional neural network. The authors collected reviews on the JD.com commerce platform for smartphones. [11] compared their model with 4 basic models such as TextCNN, BiGRU-Attention, Bert and LSTM. therefore, the experimental results indicate that the proposed hybrid model outperforms individual BERT and CNN. SenBERT-CNN also significantly outperforms other models, including BiGRU-Attention and LSTM. Then the proposed SenBERT-CNN model is significantly better with 95% compared to other methods. [1] proposed a BERT-CNN based sentiment analysis model for commodity product reviews. This BERT-CNN sentiment analysis model is therefore a proposal to improve the original BERT model to improve the accuracy of commodity sentiment analysis. They based the experiment on the dataset of real cell phone reviews on the JD Mall e-commerce platform. The authors then compared the BERT-CNN model to the original CNN model and to the BERT model. In order to ensure the accuracy and objectivity of the experimental results, [1] ran the three models 10 times on the same training set and the same test set. The final results revealed that the values F1 of the BERT-CNN model and the BERT model are higher than those of the CNN model in the training set and the test set. Therefore, the BERT-CNN model has a favorable capacity in general. The following table 1 summarizes the researches in the field of sentiment analysis by indicating their objectives, the classification techniques and the Corpus of data used in their work.

Table 1. Comparison table between research works

Article	Objective	Dataset	Tools
[12]	Construction of a new sentiment analysis model SLCABG whose goal is to improve classification performance	dangdang.com website	Dictionary of Sentiment, BERT Model, CNN Model, BiGRU Model and Attention Mechanism.
[6]	Creation of a model based on the set of models; bi-directional LSTMs (Bi-LSTM), and convolutional neural network (CNN), to perform sentiment analysis.	The IMDB Review dataset and the SST2 dataset.	Bidirectional LSTM (Bi-LSTM) and Convolutional Neural Network (CNN).
[11]	Creating a hybrid SenBERT-CNN model helps to accurately identify sentiment from product reviews and analysis.	Reviews of the JD.com commerce platform for smartphones.	The BERT model and the convolutional neural network (CNN).
[1]	Building a Sentiment Analysis Model Based on BERT-CNN for Product Reviews	Reviews of mobile phones in the JD Mall e-commerce platform	The convolutional neural network (CNN) and the BERT model.

3 Methodology

Our work aims to analyze the data extracted from the most popular e-commerce site in the field; Amazon to detect customer sentiment towards products. The main steps of the proposed approach are illustrated in figure 1. It is based on five fundamental steps. In the following, we will describe this procedure for detecting the sentiment polarity of Amazon customers :

1. **Data Collection.** This step consits to collect reviews from amazon customers.
2. **Data preprocessing.** This is the step of normalizing raw data using different preprocessing techniques.
3. **Feature Extraction.** This the step of the Word embedding representation method aimed at representing the words of a text in the form of numerical vectors [10].
4. **Sentiment Classification.** This step include the classification using the deep learning algorithm; Long Short Term Memory (LSTM).
5. **Evaluation.** In order to evaluate the performance of the model, we used accuracy and recall metrics as evaluation measures.

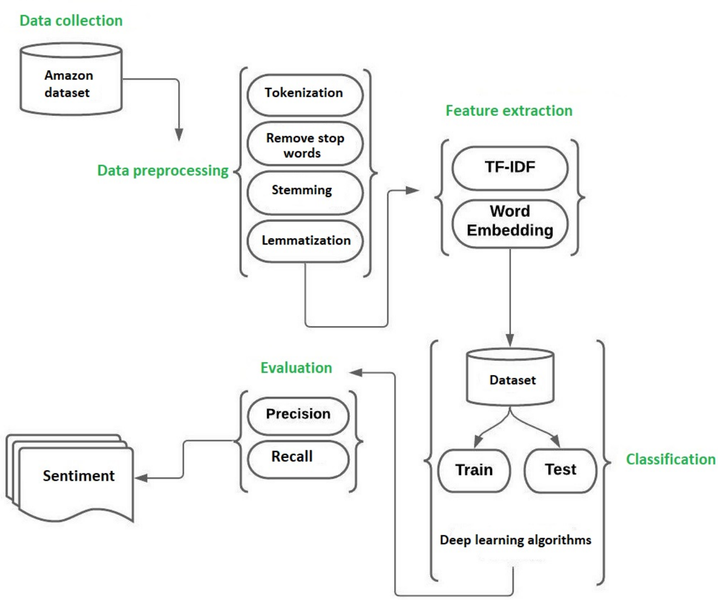

Fig. 1. Overview framework of our proposed approach.

3.1 Data Collection

Many platforms have become important sources of information such as Twitter, Facebook, Instagram and Youtube. They are powerful communication tools that allow to analyze the state of public opinion on a given subject in just a few clicks and reading comments. However, as a general rule e-commerce datasets are proprietary and therefore difficult to find among publicly available data. The extraction of data from E-commerce sites then differs from that of social networks. Hence, a process called Scraping is required. The data set consists of 41577 customer reviews of smartphones, published in the "Amazon" site.

3.2 Data preprocessing

To perform the analysis of textual data, a preprocessing step is necessary to normalize the unstructured data. Thereafter, we will detail the different preprocessing techniques used to ensure efficient sentiment classification.

Tokenization. This process involves breaking down the sequence of characters in a text by marking word boundaries, the points where one word ends and another begins [8]. The result of this segmentation is a series of elements called Tokens separated by spaces (*e.g.*, Super mobile...Samsung est très facile A manipuler. $\Longrightarrow$ ['super', 'mobile', 'samsung', 'est', 'très', 'facile', 'manipuler']).

Remove stop words. Stop words are too frequent words do not bring meaning to the text. they are useless to use in classification. In French for example, the stop words could be: the logical connectors (*e.g.*, Super mobile...Samsung est très facile A manipuler. $\Longrightarrow$ ['super', 'mobile', 'samsung', 'facile', 'manipuler']). we notice that the stop words that are going to be removed after the result are: est, très and A.

Stemming. It is the process of reducing a word to its root form [3]. This transformation aims to truncate the word of any chord by removing its prefixes and suffixes. Stemming therefore leads to forms that are not words since it consists of deleting the end of words (*e.g.*, Super mobile...Samsung est très facile A manipuler. $\Longrightarrow$ ['super', 'mobil', 'samsung', 'facil', 'manipul']).

Lemmatization. This technique aims to obtain the lemma of a word by reducing it to a normalized form. Lemmatization then consists in reducing a term, whatever its agreements, to its simplest form [5]. For example for French, obtaining its infinitive form for a verb and its masculine singular form for a noun, adjective ... (*e.g.*, Super mobile...Samsung est très facile A manipuler. $\Longrightarrow$ ['super', 'mobile', 'samsung', 'facile', 'manipuler']).

3.3 Feature Extraction

To ensure that we obtain a better result for our classification model, a step of transforming textual data into numerical data is necessary in order to transfer this data to the learning algorithms so that they can analyze it. We have used the Glove Embedding model [2].

3.4 Sentiment Classification

Sentiment classification is a method where algorithms are trained to learn and extract insights from data. Data is collected, pre-trained and made available for analysis. Since LSTMs showed good performance for sentiment classification, we decided to use the long-term memory network (LSTM) for our work. An LSTM network which stands for Long Short-Term Memory is a variant of a recurrent neural network (RNN). Generally, the architecture of an LSTM is composed of a memory cell, typically a layer of neurons, as well as three gates: an entry gate, an exit gate and a forgetting gate. These three gates will make it possible to manipulate the flow of information at the input, at the output and to be memorized in an analog way to a sigmoid type activation function.
The architecture of the model used is detailed as follows:

1. Embedding layer. used to generate an embedding vector for each input sequence.
2. Conv1D layer. which is used to convolve the data into smaller feature vectors.
3. LSTM. which has a memory state cell for learning the context of words that are further in the text whose purpose is to take on contextual meaning instead of neighboring words as in the case of RNN.
4. Dense. These are fully connected layers intended for classification.

The data breakdown is an optimization step that should not be left out, otherwise we risk over-evaluating our model (over-fitting) or quite simply the opposite (under-fitting). It is essential to have at least two data sets: one for training the model and the other will allow it to be tested for validation. For our work, it is up to us to define the distribution proportion of the dataset. We randomly split the dataset in two: 20% (8316 opinions) for the test data set aside and the rest 80% (33262 opinions) for training using the library of *scikit-learn "train_test_split"* which takes the desired proportion as a parameter. Since the training set must be representative of all the data to train the classifier in order to learn and provide results that is why generally the "Training set" takes the majority of the data. However, the "Testing set" is used to provide an unbiased assessment of a final fit of the model to the "training set" data set. Finally, we predicted the feelings of amazon customers on smartphone products through the LSTM classifier, and it showed us a high performance with an accuracy of 95% on the test set.

4 Experiment

Our data set is composed of **41577** customer reviews on smartphones, published in the "Amazon" site. The step of collecting reviews was done using a process called **Scraping** allowing the recovery of the content of a website in an automated way. So, we used scraping to extract data from the Amazon e-commerce site. Positive reviews represent roughly 90% (37242 positive reviews) of the collected data set, while negative reviews represent only 10% (4290 negative reviews) of the data set. In figure 2, we will show examples of positive reviews in frensh. While figure 3 shows examples of negative reviews. In the

	review_verified	review_content	polarity	subjectivity	sentiment
1	Achat vérifié	super mobile...samsung facile manipuler	0.333333	0.666667	Positive
2	Achat vérifié	téléphone niquel -1 étoile colis	0.000000	0.000000	Positive
3	Achat vérifié	bon entrée gamme.un écran top visionner videos	0.500000	0.500000	Positive
4	Achat vérifié	l'instant bon téléphone	0.000000	0.666667	Positive
5	Achat vérifié	corresponds descriptif attentes	0.000000	0.000000	Positive

Fig. 2. Examples of positive reviews

	review_verified	review_content	polarity	subjectivity	sentiment
0	Achat vérifié	dommage manque écouteurs j'ai regardé commenta...	0.100000	0.4	Négative
16	Version: FranceCouleur: NoirTaille: 32 GoAchat...	téléphone lent u démarrage	0.000000	0.0	Négative
35	Achat vérifié	vois écouteurs.c scandaleux qu vendus principe...	0.000000	0.0	Négative
41	Version: FranceCouleur: NoirTaille: 32 GoAchat...	j'ai acheté téléphone problème : grisaillement...	0.000000	0.0	Négative
44	Achat vérifié	impossible paramétrer mms vraiment déçu	-0.666667	1.0	Négative

Fig. 3. Examples of negative reviews

experiment, the preprocessing step is necessary to normalize unstructured data using different preprocessing techniques to ensure efficient sentiment classification. The result is shown in the table 2. Then, We have explained our LSTM model in detail as follows:

1. An input layer for variable-length integer sequences.
2. The model uses an embedding layer which is used to generate an embedding vector for each input sequence. Knowing that, the entries have a maximum number of words 30, and the dimension of the vector representation equals 300. Then, the embedding layer will associate each word of a sentence with a feature vector of 300 dimensions.

Table 2. Examples of normalized review

Id	Original review	Normalized review
0	Dommage il manque les écouteurs j'ai regardé les commentaires précédent apparemment ça arrive souvent	['dommage', 'manqu', 'écouteur', 'commentaire', 'précédent', 'apparemment', 'arrive']
1	Super mobile...Samsung est très facile A manipuler	['super', 'mobile', 'samsung', 'facile', 'manipuler']
3	Très bon entrée de gamme.¡br¿Un écran au top pour visionner les videos.	['bon', 'entrée', 'écran', 'visionner']

3. The number of embedding parameters is equal to 7879800.
4. The output of this layer is then passed to a convolution layer with 64 different filters, in order to capture the local information needed to classify sentiment.
5. Next, a Bidirectional LSTM layer takes the output from the convolutional layer and it will run the memory state cell to learn the context of words that are further in the text to convey contextual meaning.
6. After passing through the Bidirectional LSTM layer, the fully connected layer with 512 hidden neurons connects all input and output neurons. A vector that passes through this layer forms an output that is classified as positive or negative by applying the activation function Relu.

The model configuration is shown in the figure 4 below. The LSTM network is

```
Model: "model_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_3 (InputLayer)         [(None, 30)]              0
_________________________________________________________________
embedding_1 (Embedding)      (None, 30, 300)           7879800
_________________________________________________________________
spatial_dropout1d_2 (Spatial (None, 30, 300)           0
_________________________________________________________________
conv1d_2 (Conv1D)            (None, 26, 64)            96064
_________________________________________________________________
bidirectional_2 (Bidirection (None, 128)               66048
_________________________________________________________________
dense_6 (Dense)              (None, 512)               66048
_________________________________________________________________
dropout_2 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_7 (Dense)              (None, 512)               262656
_________________________________________________________________
dense_8 (Dense)              (None, 1)                 513
=================================================================
Total params: 8,371,129
Trainable params: 491,329
Non-trainable params: 7,879,800
_________________________________________________________________
None
```

Fig. 4. The LSTM model configuration.

a variant of recurrent neural network (RNN) allowing to capture the long-term dependence between words thanks to their door systems. The sentiment of each review can then be effectively categorized as positive or negative. We used the LSTM for sentiment classification and it showed great performance with 95% accuracy.

5 Evaluation and discussion

The performance results table shows that the deep learning algorithm LSTM obtained a high accuracy: 95%. We believe this is primarily because the LSTM has the ability to learn the context of words that are later in the text for the purpose of taking on contextual meaning. Moreover, by comparing our results with those of other works which have worked on Amazon product reviews as well, we notice that we obtained better results. For example in [4], the authors have worked on Amazon product reviews and they have obtained the following precision values: 55%, 64% and 65% with three different algorithms.

6 Conclusion

Nowadays, online business and shopping has become more popular. Therefore analysis of huge user feedback becomes essential to judge what people think about the product, which benefits companies to improve the products. In this study, we applied sentiment analysis with the amazon dataset. In order to achieve our goal, we spent a lot of time reading and reviewing publications and articles to see and understand the concepts and how to apply a deep learning model to our problem. Then, a data collection step is performed by automatically extracting 41,577 customer reviews in french on smartphones, published on the "Amazon" site using the process called scraping. Then, to perform an analysis of textual data, we performed a preprocessing step to normalize unstructured data to structured data. After the normalization, a step of transforming the textual data into numerical data is necessary to transfer this data to the learning algorithms so that they can analyze them in order to ensure the obtaining of a better result for our classification model. We have applied the GLOVE model. Finally, the data is collected, pre-trained and made available for classification. As a result, LSTM classifier shows good performance with 95% accuracy. The advantage of our work over other works is that we extracted massive amount of data from amazon e-commerce site on smartphone products specifically using a scraping process and this helped us in our classification of reviews with LSTM since deep learning algorithms generally require a lot of data to train properly. Moreover, the choice of the LSTM algorithm in our classification model allows us to achieve a very high accuracy. So, these results show that our model is efficient. For future preferences, we think to go a little further in the analysis with the following ideas: in the beginning, we will test our model on other data sets: we can pull data from other e-commerce platforms and do the comparison. In addition, we can collect a larger amount of data sets and with other categories of products not only smartphones. Moreover, it is not enough to use notices in French only. But, it is necessary to test the classification with other different languages. This allows us to study the influence of the challenges of various languages on the classification.

References

1. Dong, J., He, F., Guo, Y., Zhang, H.: A Commodity Review Sentiment Analysis Based on BERT-CNN Model. No. 14, International conference on Computer and Communication Systems (2020)
2. Ghannay, S.: Etude sur les representations continues de mots appliquees à la detection automatique des erreurs de reconnaissance de la parole. No. 29, Mémoire présenté en vue de l'obtention du grade de Docteur de Le Mans Université sous le sceau de l'Université Bretagne Loire (2018)
3. Jabbar, J., Urooj, I., JunSheng, W., Azeem, N.: Real-time Sentiment Analysis On E-Commerce Application. No. 17, IEEE xplore (2019)
4. Kumar, K.S., Desai, J., Majumdar, J.: Opinion Mining and Sentiment Analysis on Online Customer Review. No. 10, IEEE Xplore (2016)
5. Mhatre, M., Phondekar, D., Kadam, P., Chawathe, A., Ghag, K.: Dimensionality Reduction for Sentiment Analysis using Pre-processing Techniques. No. 28, International Conference on Computing Methodologies and Communication (ICCMC) (2017)
6. Minaee, S., Azimi, E., Abdolrashidi, A.: Deep-Sentiment: Sentiment Analysis Using Ensemble of CNN and Bi-LSTM Models. No. 11, arXiv.org (2019)
7. Mowlaei, M.E., Abadeh, M.S., Keshavarz, H.: Aspect-based sentiment analysis using adaptive aspect-based lexicons. No. 21, ELSEVIER (2020)
8. Ong, J.Y., Mun'im Ahmad Zabidi, M., Ramli, N., Sheikh, U.U.: Sentiment analysis of informal Malay tweets with deep learning. No. 23, IAES International Journal of Artificial Intelligence (IJ-AI) (2020)
9. Smetanin, S., Komarov, M.: Sentiment Analysis of Product Reviews in Russian using Convolutional Neural Networks. No. 19, IEEE 21st Conference on Business Informatics (CBI) (2019))
10. ThomasB: Word2vec : NLP & Word Embedding. No. 16, DataScientes (2020)
11. Wu, F., Shi, Z., Dong, Z., Pang, C., Zhang, B.: Sentiment analysis of online product reviews based on SenBERT-CNN. No. 12, The International Conference on Machine Learning and Cybernetics (ICMLC) (2020)
12. Yang, L., Li, Y., Wang, J., Sherratt, R.S.: Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. No. 8, IEEE Xplore (2020)