Signals and Communication Technology

Asadullah Shaikh
Abdullah Alghamdi
Qing Tan
Ibrahiem M. M. El Emary *Editors*

Advances in Emerging Information and Communication Technology

Springer

Signals and Communication Technology

This series is devoted to fundamentals and applications of modern methods of signal processing and cutting-edge communication technologies. The main topics are information and signal theory, acoustical signal processing, image processing and multimedia systems, mobile and wireless communications, and computer and communication networks. Volumes in the series address researchers in academia and industrial R&D departments. The series is application-oriented. The level of presentation of each individual volume, however, depends on the subject and can range from practical to scientific.

Indexing: All books in "Signals and Communication Technology" are indexed by Scopus and zbMATH

For general information about this book series, comments or suggestions, please contact Mary James at mary.james@springer.com or Ramesh Nath Premnath at ramesh.premnath@springer.com.

Asadullah Shaikh • Abdullah Alghamdi
Qing Tan • Ibrahiem M. M. El Emary
Editors

Advances in Emerging Information and Communication Technology

Springer

Editors
Asadullah Shaikh
Department of Information Systems
College of Computer Science and
Information Systems, Najran University
Najran, Saudi Arabia

Abdullah Alghamdi
Department of Information Systems
College of Computer Science and
Information Systems, Najran University
Najran, Saudi Arabia

Qing Tan
School of Computing and Information
Systems, Faculty of Science and
Technology
Athabasca University
Athabasca, AB, Canada

Ibrahiem M. M. El Emary
Department of Information Science
King Abdulaziz University
Jeddah, Saudi Arabia

ISSN 1860-4862 ISSN 1860-4870 (electronic)
Signals and Communication Technology
ISBN 978-3-031-53236-8 ISBN 978-3-031-53237-5 (eBook)
https://doi.org/10.1007/978-3-031-53237-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

We want to thank all the authors for their participation in the International Conference on Innovation of Emerging Information and Communication Technology, ICIEICT 2023, which was held at Universidad Complutense de Madrid (Complutense University of Madrid) Escuela Profesional de Relaciones Laborales, UCM, on September 11–13, 2023, and organized by CANWEST International Research Congress, Canada. Special support was provided by the University of Almería (Universidad de Almería).

This book constitutes the refereed proceedings of the International Conference on Innovation of Emerging Information and Communication Technology, ICIEICT 2023. ICIEICT aims to provide occasions for international researchers and practitioners to present the most recent advances and future challenges in the fields of Information and Communication Technology. The conference received 128 submissions—confirming a trend of increasing interest in ICIEICT—coming from over 17 different countries including Algeria, Australia, Canada, China, Denmark, France, Japan, Kuwait, Mexico, Morocco, Poland, Romania, Saudi Arabia, Spain, Ukraine, United Kingdom, and USA.

When choosing the technical papers, the Technical Program Committee considered two factors: first, to offer a platform for more global participants by presenting and debating pertinent topics from a wide range of information and communication technologies (ICT) disciplines, reflecting the quickly advancing state of technology we are currently seeing. The second objective is to uphold a fair technical quality through the selection of about 31 technical articles, or about a 16% acceptance rate.

Four parallel-track sessions have been designed into the conference schedule to accommodate paper presentations. Deep learning, cybersecurity, blockchain technologies, health informatics, supervised and unsupervised learning, human digitization, enterprise resource planning, convolutional neural networks, sentiment analysis, endovascular intervention, robotics, and cloud computing were among the multidisciplinary topics covered in the presentation that had reasonable balance between theory and practice. Additionally, five keynote addresses from well-known ICT specialists are included in the program.

This event was only possible with the passion and diligence of several colleagues. We want to thank the members of the steering committee for their encouraging leadership and the general chairs for their help during the entire process. We also express our gratitude to the other Organizing Committee members for their productive collaboration. A special thank you to all of the referees and members of the Technical Program Committee for their essential assistance in reviewing the papers.

Najran, Saudi Arabia — Asadullah Shaikh
Najran, Saudi Arabia — Abdullah Alghamdi
Athabasca, AB, Canada — Qing Tan
Jeddah, Saudi Arabia — Ibrahiem M. M. El Emary
November 2023

Organization

Conference Chairs

Luigi Vladareanu	Institute of Solid Mechanics of Romanian Academy, Romania
Joan Lu	Hudersfield University, UK

Steering Committee

Rajesh Kumar Dhanraj	Symbiosis International, India
Jawad Rasheed	Istanbul Sabahattin Zaim University, Turkey
Sheifali Gupta	Chitkara University, India
Qurat-ul-Ain Mastoi	University of the West of England, UK
Naeem Ahmed Mahoto	MUET, Pakistan
Ahmad Taher Azar	Prince Sultan University, Saudi Arabia
Junzo Watada	Watada University, Japan
Latif Ladid	University of Luxembourg/SnT, Luxembourg

Technical Program Committee Chair

Qing Tan	Athabasca University, Canada
Asadullah Shaikh	Najran University, Saudi Arabia
Abdullah Alghamdi	Najran University, Saudi Arabia

Local Chairs

Juan E. Trinidad Segovia	Universidad de Almería, Spain

Submission and Proceedings Committee

Samar M. Alqhtani	Najran University, Saudi Arabia
Adel Sulaiman	Najran University, Saudi Arabia

Registration and Management Committee

Mohamed El Emary	Ecole Technologie Superieure, Montreal, Canada
Ibrahiem El Emary	King Abdul Aziz University, Saudi arabia

Web and Publicity Committee

Ibrahiem El Emary	King Abdul Aziz University, Saudi Arabia

Contents

Baheta: Balanced and Unbalanced Dataset in Arabic Clickbait Detection Using a Deep Learning Model (LSTM)

Batool Alharbi, Razan Alhanaya, Deem Alqarawi, and Ruwaidah Alnejaidi

1 Introduction

Over the years, the use of social networks has increased until the number of users reached 5.16 billion worldwide in 2023 [1]. As a result, about 40% of fake news [2] is on social media. One method of disseminating fake news is using Clickbait, in which the author manipulates the title head through attractive words or pictures and misleading content to grab the user's attention and persuade him to click the link.

In this paper, we will explain the method of detecting Clickbait using deep learning (DL). DL is the study of artificial neural networks and associated machine learning (ML) algorithms that incorporate more than one hidden layer. It is sometimes referred to as deep structured learning, deep hierarchical learning, or deep machine learning field [3]. There are many DL models used in text classification such as long short-term memory networks (LSTMs), which we will use to detect Arabic clickbait. As Arabs, it is important for us to reduce Clickbait from Arabic content as we see the effect it has on our society and culture, and because there are few papers on detecting Arabic Clickbait using DL models, we will focus on that in this paper.

In Baheta, we will use two datasets, the first is Arabic Clickbait datasets [4] that are unbalanced, as the number of Clickbait is 26% less than the non-Clickbait. The second one is a combined dataset consisting of the Arabic Clickbait datasets and an Arabic fake news dataset [5] to make the Clickbait almost equal to the non-Clickbait.

B. Alharbi (✉) · R. Alhanaya · D. Alqarawi · R. Alnejaidi
Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia

A. Shaikh et al. (eds.), *Advances in Emerging Information and Communication Technology*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-53237-5_1

Fig. 1 Logo of Baheta: balanced and unbalanced dataset in Arabic Clickbait detection using LSTM

1.1 What Baheta Means?

Since Clickbait, in short, is "lying" and "misleading" the user, we called this proposal "Baheta," which is an Arabic synonymous with lying, as it denotes Clickbait. We designed a special Baheta identity consisting of two languages because this proposal is in English and Arabic out of pride in our language, the goal of the proposal, and the logic of its work. In Fig. 1, you can see Baheta logo.

1.2 Related Work

Linguists and computer scientists use Clickbait, and it has become a crucial topic for research. In this section, a general review of relevant Clickbait detection techniques has been covered.

According to researchers in [6], LSTM is the most crucial algorithm used to detect Clickbait and rumors. This paper was based on information from Twitter and news organizations. Because of the handling of texts, the text had to be converted into a number matrix. There are two solutions, either TF-IDF or Word2Vec. They must measure the semantic sensitivity correctly, which is why they ignore TF-IDF. With the spread of global Clickbait go to select GloVe over Word2Vec. One of the LSTM benefits is that it can deal with long-range contextual information and with the exploding and vanishing gradient problems in RNNs. They achieved an accuracy of 0.859 ± 0.002 and $\text{MSE} = 0.0296 \pm 0.0002$ when over 10 different runs.

As described in [7], a technique termed the ontology-based LSTM model (OLSTM) is suggested by Vorakit et al. RNNs were employed as a classification model. By actively maintaining self-connecting layers, LSTM expands RNNs. Dataset relies on Twitter and news websites to learn about the model, and they use word embedding information to better Clickbait identification because words are the foundation for deciding whether something is Clickbait or not. According to their observations, the effectiveness of identifying Clickbait has indeed increased by word embedding approaches. The results show that OLSTM significantly outperforms approaches that did not use word embedding characteristics, with an F1-score and precision of 0.8990 and 0.8994.

In the article [8], a two-phase model has been proposed. In the proposed approach, the hybrid CNN-LSTM model is implemented in the first phase to identify Clickbait, which is further fed to the second phase where eight types of Clickbait such as reasoning, number, reaction, revealing, shocking/unbelievable, hypothesis/guess, questionable, and forward referencing are classified by using the Biterm (BTM). Three datasets were used, which collected data from different sources: Facebook, Twitter, and websites, and dataset 3 was text data and data extracted from images. The highest accuracy achieved by the hybrid CNN-LSTM model is 95.8%.

Zheng et al. [9] proposed a Clickbait convolutional neural network (CBCNN) model that takes into account not only the general features but also the specific features of different article types. The CBCNN model consists of Word2Vec models and a CNN model. CNN understands the input text from different perspectives. Each word has two meanings in the model: its general meaning and its type-related meaning, and it can be used for different natural language processing tasks. CBCNNs are used, for example, for recognizing parts of words, naming entities, extracting relations and events, and computing sentence similarities, and it shows how CNN works in Clickbait detection. One limitation of CBCNNs is that the maximum length of the headline is limited, which can lead to information loss in long headlines. CBCNNs do not include a pre-trained Word2Vec model that could allow the model to better understand the meaning of words.

In addition to the CNN [10], a CNN model is used in this study to provide a method for detecting Clickbait on online social media. The strategy being employed is concentrated on textual features. Word2vec was used to extract features from the data. The results indicate that CNN outperformed other ML methods with a high accuracy of 0.82%. Comparative analysis was carried out using the RF.

There isn't much research focusing on Arabic Clickbait headline detection. In order to better identify Clickbait news on social networks in Arabic, this study [4] built the first Arabic Clickbait headline news dataset and presented a multiple-feature-based strategy. Data collection, data preparation, and ML model training and testing phases are the three key components of the suggested approach. A total of 54,893 Arabic news articles were included in the dataset that was gathered (after pre-processing), and 23,981 of these news stories were Clickbait headlines. The ANOVA F-test was used to choose the most significant features from this pre-processed dataset. To ensure that the best settings were discovered, a number of ML techniques were then used with hyper-parameter tweaking techniques. The evaluation of the ML models was completed, and according to the findings, the support vector machine (SVM) with the top 10% of ANOVA F-test features (user-based features (UFs) and content-based features (CFs)) had the highest accuracy of 92.16%.

Mohammad A. Bsoula et al. [11] also applied ML models to an Arabic dataset of news. The purpose of this dataset is to enable the automatic classification of news headlines as "Clickbait" or "Not Clickbait" using ML models. A total of 3235 records were gathered and 2652 news records were classified as "Not Clickbait" and 583 as "Clickbait." The class distribution within the dataset was unequal with a ratio of "Clickbait" to "Not Clickbait" of 1:4. Therefore, to extract the final dataset needed for this research, the evaluation was performed on the original dataset, oversampled dataset, and under-sampled dataset. The three train datasets were subjected to the application of different seven ML models. Model hyperparameters were fine-tuned by ten-fold cross-validation on the training dataset. The evaluation of these models using the test dataset results produced a Macro F1-Score in the range of 0.18–0.81 for the explored feature combinations, successfully demonstrating the dataset's utility. According to the results, the resampling did not improve the F1-score value.

2 Methodology

We illustrate the general methodology of recognizing the Clickbait detection using DL in Fig. 2, which describes the phases involved in our proposed detection framework. We will explain each phase in detail in this section and illustrate the expected outcomes and our reasoning.

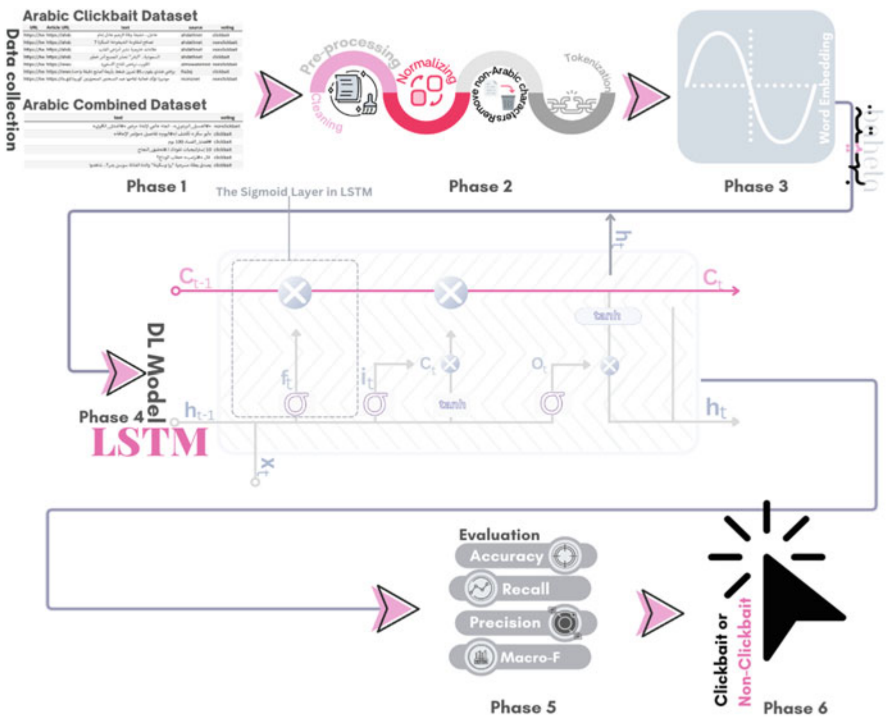

Fig. 2 Overview of basic phases of Baheta methodology approach

2.1 Dataset

The primary stage in the DL model pipeline is gathering data for training the DL model. The accuracy of the predictions provided by DL systems depends on training data. The dataset we will use is [12]; according to our knowledge, there is no other for the Arabic Clickbait. The Arabic Clickbait dataset was collected in this study [4]. This dataset contains 10,564 rows collected in Arabic from Twitter, 7735 were rated as non-Clickbait, and 2829 as Clickbait, which constitutes 26% of all datasets and makes the dataset unbalanced. As shown in Fig. 3, the dataset comprises five columns (features): URL, Article URL, text, source, and voting. Table 1 describes the dataset's features.

We combined the Arabic Clickbait dataset with the Arabic fake news dataset to enhance the Clickbait data ratio until it is equal to the non-Clickbait data because the Arabic Clickbait dataset is unbalanced. Arabic Fake News Dataset [5] contains 606,912 news classified into credible, not credible, and undecided collected from 134 different public Arabic news websites. The combined dataset consists of two columns (features): text and voting as described in Table 2. A sample of it is shown in Fig. 4. This dataset has 15,472 rows, it is an Arabic Clickbait dataset with some Arabic Fake News Datasets that are classified as not credible to become the Clickbait percentage of 50% in the combined dataset, and we published it on GitHub [12].

URL	Article URL	text	source	voting
https://tw	https://ahda	عاجل.. حقيقة وفاة الزعيم عادل إمام	ahdathnet	clickbait
https://tw	https://ahda	نصائح لمقاومة الشيخوخة المبكرة 7	ahdathnet	nonclickbait
https://tw	https://ahda	علامات خارجية تشير أمراض القلب	ahdathnet	nonclickbait
https://tw	https://ahda	السعودية.. "أبشر" تحذر الجميع أمر خطير	ahdathnet	clickbait
https://tw	https://www	الكويت ترخص للقاح أكسفورد	almowatennet	nonclickbait
https://tw	https://www	رياضي هندي يقوم بـ 85 تمرين ضغط بأربعة أصابع دقيقة واحدة	Ra2ej	clickbait
https://tw	https://is.gd	موديرنا تؤكد فعالية لقاحها ضد النسختين المتحوّرتين كورونا	mzmznet	nonclickbait

Fig. 3 Sample of the Arabic Clickbait dataset [12]

Table 1 Description of the Arabic Clickbait dataset features

Feature	Description
URL	Tweet link
Article URL	Link to the article to which the tweet refers
text	Tweet content
source	The source who posted the tweet
voting	Tweet labeled as "Clickbait or non-Clickbait"

Table 2 Description of the combined dataset features [12]

Feature	Description
text	Tweet content and fake news
voting	Tweet labeled as "Clickbait or non-Clickbait"

text	voting
«#الغسيل_البريتوني».. اتجاه عالمي لإنقاذ مرضى «#الفشل_الكلوي»	nonclickbait
«أبو سكر» تكشف لـ«#اليوم» تفاصيل «مؤتمر الإعاقة»	clickbait
#قضايا_الفساد 100 يوم	clickbait
10 إستراتيجيات تقودك لـ #تحقيق_النجاح	clickbait
قال «#ترامب» خطاب الوداع؟	clickbait
يصدق بطلة مسرحية "ريا وسكينة" والدة الفنانة سوسن بدر؟.. شاهدوا	clickbait

Fig. 4 Sample of the combined Arabic dataset [12]

We now have two datasets: unbalance (Arabic Clickbait dataset) and balance (combined Arabic dataset).

2.2 *Pre-processing*

Raw data from the real world are frequently inconsistent, lacking in specific behaviors or trends, and incomplete. They are also probably full of mistakes. As a result, after being gathered, they undergo pre-processing to create a model-compatible format for the DL model. The data must go through a pre-processing process before classification. The proposed model will undertake data preprocessing tasks, including cleaning, normalization, tokenization, and the removal of non-Arabic characters, prior to its utilization.

- Cleaning includes:
  Convert hashtags to words and remove emails, URLs, mentions, punctuations, repeating chars, stop words, and emojis.
- Normalizing includes:
  Converting the hamza to the near letters, removing diacritics and tatweel characters, and replacing “ة” with “ه”.
- Remove non-Arabic characters:
  Remove numbers and non-Arabic characters.
- Tokenization:
  Involves breaking up the sentence into a collection of words known as tokens.

Because we care for the Arabic language, it is important that the pre-processing of Arabic data is using libraries specialized in Arabic natural language processing. So we used Ruqia [13], and pyarabic [14] libraries, as shown in Fig. 5, which we represented in our methodology.

Fig. 5 Text pre-processing phase in Baheta methodology

2.3 Word Embedding

The techniques used to map words or phrases to vectors of real numbers are referred to as word embedding. Words are represented as continuous vectors in a low-dimensional space via word embedding techniques. Word2vec technology is arguably the most popular method for using an external feed-forward neural network to learn word embeddings. We used Skip-gram word embedding for the proposed model from AraVec. AraVec [15] is an open-source word2vec project for Arabic. The process of embedding words starts with creating a co-occurrence matrix, and we take into consideration the length of texts in the dataset.

2.4 Model

The architecture of the LSTM model we created is composed of several layers. The embedding layer is the first layer followed by the SpatialDropout1D layer with 0.2 units to drop, and SpatialDropout1D will help promote independence between feature maps. The next layer is an LSTM layer with 128 neurons, employing the tanh activation function for cell state and hidden state and the Sigmoid function for LSTM gates, as shown in Fig. 6. GlobalMaxPooling and a Dropout layer with a probability of 0.2 are used. The final dense layer is the output layer, which uses a Sigmoid activation function.

2.5 Evaluation

Through the dimension reduction process of feature extraction, a large initial set of raw data is divided into smaller groups for processing, which will be done by the LSTM model. Performance metrics are used to assess the effectiveness or quality of the model. These performance measures aid in our comprehension of how well our

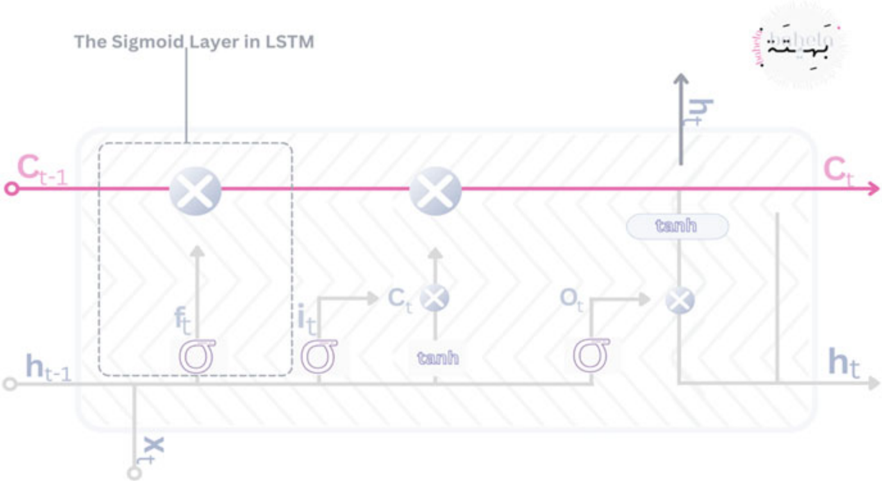

Fig. 6 LSTM layer architecture of the Baheta proposed model

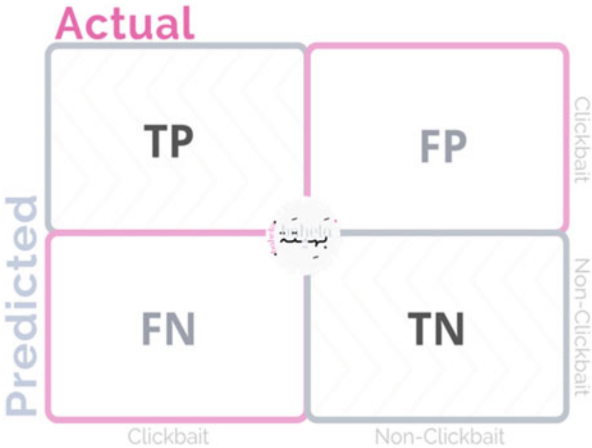

Fig. 7 Confusion matrix for Clickbait detection

model handled the supplied data. Performance metrics assist measure how well a DL model generalizes on new or previously unexplored data. In our proposed work, we will analyze our chosen set of approaches using performance metrics like accuracy, precision, recall, and Macro-F score. To understand the necessity of using the performance measurements chosen for this study, it is important to take into account the confusion matrix notion. For a prediction task, a confusion matrix summarizes examples of accurate and incorrect predictions. A confusion matrix is graphically represented in Fig. 7, based on [16].

- *True Positives (TP)*: The cases in which the model predicted Clickbait and the actual output were also Clickbait.
- *True Negatives (TN)*: The cases in which the model predicted non-Clickbait and the actual output was non-Clickbait.
- *False Positives (FP)*: The cases in which the model predicted Clickbait and the actual output was non-Clickbait.
- *False Negatives (FN)*: The cases in which the model predicted non-Clickbait and the actual output was Clickbait.

We found several metrics for evaluation during our research, which are explained below:

The accuracy formula is used to determine the percentage of all accurate predictions, and it can be represented by the following equation:

$$\mathrm{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

The recall formula is used to determine the percentage of true positive predictions among all possible positive predictions a model may have made, and it can be represented by the following equation:

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{2}$$

The precision formula is used to determine the proportion of accurately predicted actual cases to all positively predicted instances, and it can be represented by the following equation:

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{3}$$

The macro-F formula is used to calculate the arithmetic mean of the individual class-related f1 score, and it can be represented by the following equation:

$$\mathrm{Macro} - F = \frac{\mathrm{sum}(F1 \quad \mathrm{scores})}{\mathrm{number\ of\ classes}} \tag{4}$$

2.6 Clickbait or Non-clickbait

After applying the LSTM model to the processed data, which involves feature extraction, the results are evaluated using our model. In this phase, the model determines if the headline is categorized as Clickbait or non-Clickbait based on the final result.

2.7 Hyperparameters and Training

For our model, we chose 500 epochs with a batch size of 512. The dimension of the word embedding is kept at 300. Also, we will use binary-cross entropy as a loss function for binary classification and Adam as an optimizer. The parameters are kept the same for all the experiments for a fair comparison.

3 Results and Discussion

Our model focuses on the text feature to detect Clickbait. Because the dataset is unbalanced, we will adopt an F-macro measurement to compare the results as a fair measure. The model was trained using Word2vec and also without Word2vec using both the balanced and unbalanced datasets. Using random splitting, the datasets will be divided into 20% for test data and 80% for training data. The model obtained the best performance using Word2vec, where the Macro-F value reached 0.79 using the unbalanced (raw) dataset, and the Macro-F value reached 0.76 using the balanced dataset. As we can see the results of the model in Table 3, the results showed that the process of merging to make the dataset balanced did not give better. We can see that the LSTM model performed better with the unbalanced dataset because it achieved a higher Macro-F value of 0.02 than the LSTM did with the balanced dataset. The difference in the performance of LSTM is shown in Fig. 8.

We can see the performance of the LSTM model during training on detecting Clickbait in an unbalanced dataset in Fig. 9, and using the balanced dataset in Fig. 10.

Figure 9 is divided into three subfigures. In (a), the confusion matrix of the model shows the number of correct and false classifications, as we can see the model predicted 1775 of data correctly out of 2113. (b) shows the performance of the model during training, and the model reached a validation accuracy ranging from 0.81 to 0.84. In (c), the amount of loss of the model during training is shown, and we see that the loss value fell to less than 0.4 in the first Epochs and began to rise in Epoch 10.

Figure 10 is divided into three subfigures. As can be seen in (a), the model's confusion matrix displays the amount of true and wrong classifications, and the model correctly predicted 2390 data out of 3095. The model's performance during training is shown in (b), and it achieved a validation accuracy between 0.7 and 0.79. The model's training loss is displayed in (c), and it dropped to less than 0.5 in the early epochs before starting to grow.

Table 3 LSTM model results when trained on unbalanced and balanced datasets

	Without Word2vec				With Word2vec			
	Acc	Recall	Precision	Macro-F	Acc	Recall	Precision	Macro-F
Unbalanced	0.78	0.84	0.85	0.72	0.84	0.92	0.87	**0.79**
Balanced	0.68	0.7	0.67	0.68	0.76	0.71	0.78	**0.76**

Fig. 8 LSTM model results when it's trained on balanced and unbalanced datasets

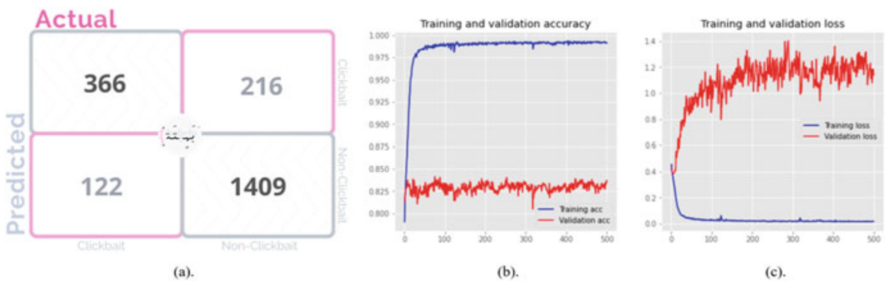

Fig. 9 LSTM performance during training on an unbalanced dataset with 500 Epochs

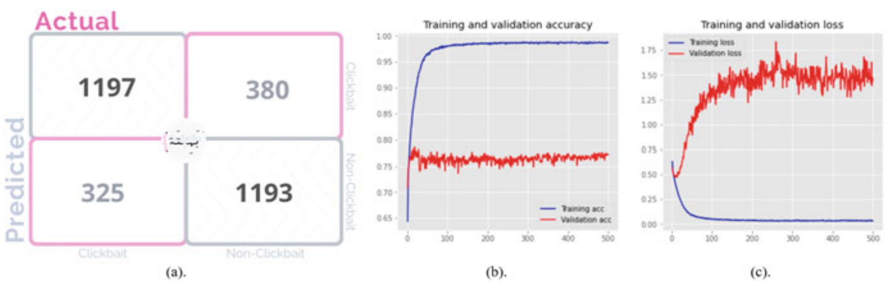

Fig. 10 LSTM performance during training on a balanced dataset with 500 Epochs

4 Conclusion and Future Work

For the first time ever, Arabic Clickbait is detected using DL in this paper, and we offer an LSTM model for its detection. The text is the key clickbait feature, and Word2vec was used to extract the features from the text. The LSTM model was trained on two Arabic datasets: the original, which was unbalanced; and the combined, balanced version of the original and a dataset of Arabic fake news. We conclude that employing Word2vec with LSTM for the Arabic content is preferable for both datasets. The LSTM model showed better performance with the unbalanced dataset, as it obtained a higher Macro-F value of 0.02 than that obtained by the LSTM with the balanced dataset. There are multiple ideas for future development, the most significant of which is to experiment with various deep DL models to find Clickbait in Arabic content.

References

1. Satista. *Number of Internet and Social Media Users Worldwide as of January 2023*. [January 2023 May 2023]; Available from: https://www.statista.com/statistics/617136/digital-population-worldwide/
2. D. Ruby, *69+ Fake News Statistics Revealed For 2023 (Updated)*. [April 17, 2023 May 2023]; Available from: https://www.demandsage.com/fake-news-statistics/
3. P. Ongsulee, Artificial intelligence, machine learning and deep learning, in *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*, (IEEE, 2017)
4. M. Al-Sarem et al., An improved multiple features and machine learning-based approach for detecting Clickbait news on social networks. Appl. Sci. **11**(20), 9487 (2021)
5. M. Yaseen, *Arabic Fake News Dataset (AFND)*. 2022. Available from: https://www.kaggle.com/datasets/murtadhayaseen/arabic-fake-news-dataset-afnd?resource=download
6. C. Oliva et al., Rumor and clickbait detection by combining information divergence measures and deep learning techniques, in *Proceedings of the 17th International Conference on Availability, Reliability and Security*, (2022)
7. V. Vorakitphan, F.-Y. Leu, Y.-C. Fan, Clickbait detection based on word embedding models, in *Innovative Mobile and Internet Services in Ubiquitous Computing: Proceedings of the 12th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2018)*, (Springer, 2019)
8. S. Kaur, P. Kumar, P. Kumaraguru, Detecting clickbaits using two-phase hybrid CNN-LSTM biterm model. Expert Syst. Appl. **151**, 113350 (2020)
9. H. Zheng et al., *Clickbait Convolutional Neural Network. Symmetry* (Multidisciplinary Digital Publishing Institute (MDPI), 2018)
10. M.A. Shaikh, S. Annappanavar, A comparative approach for clickbait detection using deep learning, in *2020 IEEE Bombay Section Signature Conference (IBSSC)*, (IEEE, 2020)
11. M.A. Bsoul, A. Qusef, S. Abu-Soud, Building an optimal dataset for arabic fake news detection. Procedia Comput. Sci. **201**, 665–672 (2022)
12. T. Wals, *Clickbait-Fake-News-Dataset*. 2023. Available from: https://github.com/RazanALhanaya/Clickbait-Fake-News-Dataset
13. Ruqiya. *Ruqia Library*. Available from: https://github.com/Ruqyai/Ruqia-Library
14. T. Zerrouki, *Pyarabic Library*. 2023. Available from: https://github.com/linuxscout/pyarabic
15. A.B. Soliman, *AraVec 3.0*. Available from: https://github.com/bakrianoo/aravec
16. A. Luque et al., The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recogn. **91**, 216–231 (2019)

Introducing a Vision of Regulation More Complex Than the Traditional One

Olivier Lefebvre

1 Introduction

There is a doxa of regulation, in particular on the crucial topic of the "good" number of competitors in an economic sector. Examples of this doxa are: (1) any entry is good for the consumers; (2) when competition exists, it is easy to maintain it; (3) a monopoly is always bad for the consumers, etc. This doxa is a simplification. There are breaches of this doxa. To show it, a model already presented by the author is used. It is Bertrand competition [1]. In general, when Bertrand competition is concerned, the demand functions are given [2]. One prefers to deduce the demand functions from the utilities of the consumers [3]. Conditions to observe breaches of the doxa are to consider multiproduct firms, demand functions that are not "regular" (twice continuously derivable), and adopt a dynamic view of the equilibrium of the sector (which can evolve through stages).

Examples of these breaches are:

- "Any entry is good for the consumers." Indeed, if the demands are not "regular," an entry can trigger a decrease in the consumers' surplus.
- "When competition exists, it is easy to maintain it." In fact, it is possible that the number of products that are sold decreases, and the consumers' surplus diminishes. We have called this situation a profitable "close down" [4, 5].
- "A monopoly is always bad for consumers." A monopoly is neither good nor bad for the consumers. Sometimes, it is good from the point of view of the diversity of the products sold and bad from the point of view of the prices. In the case of a profitable "close down," the diversity of the products sold is not maintained if there is a competitive equilibrium and is maintained if there is a monopoly. But the prices are high.

O. Lefebvre (✉)
Olivier Lefebvre Consultant, Paris, France

A. Shaikh et al. (eds.), *Advances in Emerging Information and Communication Technology*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-53237-5_2

In this paper, one uses two methods:

- Building up tractable examples, the demand functions being not "regular."
- Making demonstrations, but only supposing that the demands are "regular" functions.

Before setting out the plan of the paper, let us make two remarks:

- One uses the symbol (i, j) to describe a firm selling the products i and j ($1 \leq i \leq 3$, $1 \leq j \leq 3$). And one uses brackets [.] to describe the sector. Thus, [(1, 2), 3] means that there is a multiproduct firm selling products 1 and 2, and a firm selling product 3.
- Several tractable examples are shown in the paper. Thus, the model (Bertrand competition, the demands being deduced from the utilities of the consumers) is set out. The utilities are represented in axes O u_1 u_2 u_3, the demands are deduced, and one calculates the Nash equilibrium.

Now we set out the plan of the paper:

1.1 Methodology

We show an example of entry triggering a decrease of the consumers' surplus. This is because the demand functions are not regular (in a sense that one defines). In the following, the demand functions are supposed "regular."

1.2 Results and Discussion

The following topics are studied:

- *Decrease of the number of the products sold.* Is studied the stability of the equilibrium, in that sense that the number of products sold can decrease (mainly because of profitable "close down"). Also, one demonstrates that when the demand functions are "regular," the profitable close down triggers an increase in the prices (therefore a decrease of the consumers' surplus).
- *One studies this question: in the case of entry [(1, 2)] → [(*1, 2), 3*], do the prices increase or decrease?* Here the demand functions are supposed "regular." One states that if the demands are linear, the prices decrease. But there is the possibility that the prices increase.
- *One studies the particular case of start-ups, when the question of the diversity of the products sold is concerned.*

1.3 Conclusion

One sets out the hypothesis that the regulators are strategists. There is not a certain doctrine showing what the regulator has to do. No wonder there are two "styles of regulation." The watchdog style of regulation relies on the criteria of welfare of the consumers. The permissive style of regulation relies on the criteria of value. For a regulator of the watchdog style, the stake is competition as intense as possible. For a regulator of the permissive style, some remedy to a saturated market is tolerated. The affair Illumina (a merger approved by some regulators and refused by others) shows that there are two styles of regulation. They are very different and incommensurable. The same phenomenon, the profitable "close down," can be interpreted in two different ways.

2 Methodology

Now one studies the consequences of an entry [1] → [1, 2], or [1, 2] → [1, 2, 3] on the consumers' surplus. If the demand functions are "regular,"[1] the reaction functions exist, and an easy reasoning on these reaction functions allows us to demonstrate that the prices decrease, and the consumers' surplus increases.

But it is easy to build up an example, with demand functions not regular, where the entry triggers a decrease of the consumers' surplus (example 1).

In example 1, the density is on O u_1, linear, homogeneous equal to 2/3, and the weight 1/3 is in A (1/3, 1) (Fig. 1).

If the sophisticated product 2 is not sold, $p_1 = 1/3$. If it is sold, $p_1 = 1/2$ and $p_2 \approx 0, 87$. There is a qualitative explanation: E_2 decreases the price p_2 just enough to discourage E_1 from attracting the consumers in A. E_1 "capitulates" and chooses a price p_1 which is more than 1/3. As one price increases (p_1 from 1/3 to 1/2) and the other decreases (p_2 from more than 1 to 0, 87), one has to calculate the consumers' surplus, in the two situations. If this surplus is S before the entry and S' after, S' is less than S since $S \approx 0, 148$ and $S' \approx 0, 125$.

This example seems artificial because of the concentration of consumers in A. But there is joint purchasing. A batch of products is bought, at some price. Therefore, a unit is bought at the same price as any, in the batch. An example is the online platform Groupon.

Suppose the product is a book, with two variants. One is a standard product (product 1). The other is a luxurious book, sold by a club of readers to its members. This market niche represents 1/3 of the potential consumers. The welfare of these consumers increases when the sophisticated product is sold. But the welfare of the

[1] "Regular" means two times derivable, the reaction functions exist, and are continuous, increasing in the plane O $p_i\, p_j$ (p_k being fixed), and there is a unique stable Nash equilibrium.

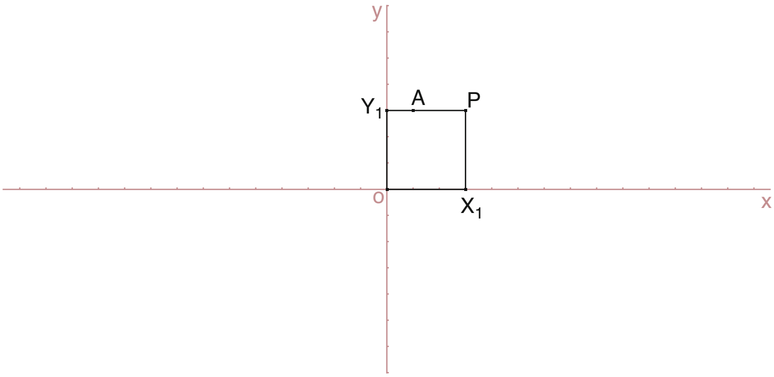

Fig. 1 *Example 1*. The density is on the segment O X_1 (1, 0) and equal to 2/3 and there is the weight 1/3 in A (1/3, 1)

consumers who can buy the standard product only decreases. And the surplus of all the consumers decreases.

To have an example of entry [1, 2] → [1, 2, 3], which triggers a decrease of the consumers' surplus, one can use example 1. The density is: 1/3 on the segment (0, 0), (0, 1) and on the segment (0, 0), (0, 1). And there is a weight of 1/6 in A (1/3, 0, 1) and in B (0, 1/3, 1). The calculations are exactly the same as in example 1.

3 Results and Discussion

One studies the following examples:

3.1 Study of the Stability of the Equilibrium

Here we suppose the demand functions "regular." There is an entry [1, 2] → [1, 2, 3]. Is it stable? Can product 3 be withdrawn from the market? Yes, it is possible. But there are degrees of "seriousness." What is possible is this: First, there is a merger [1, 2, 3] → [(1, 3), 2]. It is always profitable: the profit of the merged firm increases and the prices increase. The consumers' surplus decreases. This is easily demonstrated, using the reaction functions. Then, possibly, [(1, 3), 2] → [1, 2]. The asset E_3 is closed down if it is profitable. Finally, product 3 has been withdrawn from the market. The "close down" that is profitable (probably) triggers an increase in the prices (the demonstration is in an Appendix). Therefore, the withdrawal of the product from the market triggers an increase in the prices and should be prevented by the regulator. But it is awkward [5].

Suppose that at the start the profits are P_1, P_2, P_3. After the merger, they are P''_1 and P''_2. And after the close down, they are P'_1 and P'_2. One notes $\Delta P = P_1 + P_2 + P_3$ and $\Delta P' = P'_1 + P'_2$. There are four cases:

Stable equilibrium. It is if the joint profit increases when there is the entry. Δ P > Δ P′. An easy example is when the products are perfectly differentiated. The density is linear, homogeneous, and equal to 1/3, on the axes O u_i between 0 and 1.

Possible withdrawal of the product thanks to lateral payments. If $\Delta P < \Delta P'$, and if the "buy and close down" is not profitable ($P'_1 < P_1 + P_3$), what is possible is a withdrawal of the product thanks to lateral payments. There is a transfer t from E_2 to E_1 such that: $P'_2 – t > P_2$ and $P'_1 + t > P_1 + P_3$, and each firm makes a gain. But it is highly improbable for two reasons:

- A complex "mercato" is needed, with complex exchanges of assets between the firms E_1 and E_2 since the transfer t has to be hidden from the regulator.
- There is a problem of moral hazard. The firm E_1 can accept the transfer, then … not to close down asset E_3. Its gain is more. Let us imagine an example:

  At the start: $P_1 = P_2 = P_3 = 100$

  E_2 gives 60 to E_1

  E_1 uses 60 to buy E_3 and closes down it.

  After, $P'_1 = P'_2 = 180$. Each firm makes a gain thanks to the lateral payment: 180–60 = 120 > 100.

  But E_1 can betray E_2, not closing down E_3 … after having cashed out 60.

  The profit of E_1 is 220–60 = 160 > 120 (the profit of E_1 after the merger is 220, which is more than 100 + 100 since the merger is profitable).

  The firm E_2 has lost 60, even if its profit after the merger is more than 100.

It is example 4 (Fig. 2).

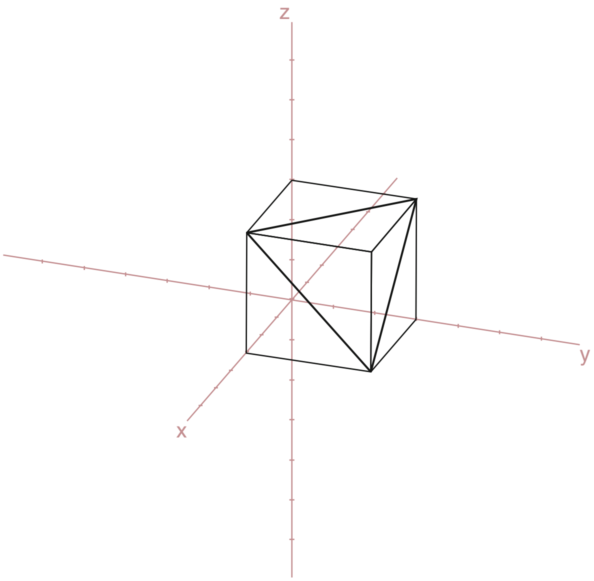

Fig. 2 *Example 4.* The density is linear equal to 1/3√2, on the second bisector of the external faces $u_i = 1$