

Propaganda Detection And Challenges Managing Smart Cities Information On Social Media

Pir Noman Ahmad¹ and Khalid Khan^{2,*}

¹School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

²Computer Science and Software Engineering, University Of Stirling, UK

Abstract

Misinformation, false news, and various forms of propaganda have increased as a consequence of the rapid spread of information on social media. The Covid-19 spread deeply transformed citizens' day-to-day lives due to the overview of new methods of effort and access to facilities based on smart technologies. Social media propagandistic data and high-quality information on smart cities are the most challenging elements of this study. As a result of a natural language processing perspective, we have developed a system that automatically extracts information from bi-lingual sources. This information is either in Urdu or English (Ur or Eng), and we apply machine translation to obtain the target language. We explore different neural architectures and extract linguistic layout and relevant features in the bi-lingual corpus. Moreover, we fine-tune RoBERTa and ensemble BiLSM, CRF and BiRNN model. Our solution uses fine-tuned RoBERTa, a pretrained language model, to perform word-level classification. This paper provides insight into the model's learning abilities by analyzing its attention heads and the model's evaluation results.

Keywords: Machine translation, Span, linguistic, neural architectures, BiLSM.

Received on 10 December 2022, accepted on 25 January 2023, published on 30 March 2023

Copyright © 2023 Pir Noman Ahmad *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetsc.v7i2.2925

1. Introduction

As information dissemination has become more widespread without quality control, users can spread misinformation and target individuals with propaganda campaigns via social media to advance their ideological agendas. There are two main forms of disinformation, which are propaganda and fake news, which differ in that propaganda may be constructed based on accurate information, whereas fake news may be built on top of false information or on top of intentional distortions. Propaganda refers to actions or opinions of individuals or groups that are intentionally planned to inspire the actions or opinions of other individuals or groups toward specific goals [1,2].

Various solution utilizes BERT [3], a Transformer-based [4] model relying on multiheaded attention, its purpose of the fragment-level propaganda (FLP) classification. Social

media, mainstream media, and the online internet have become increasingly popular sources of information because of their sheer volume, which makes manual analysis impossible. Translating text is an essential task to be able to into natural language sentences to accomplish this goal. An initial method of neural machine translation based on hand-crafted translation rules [5]. Data-driven approaches have gained more attention as large-scale parallel corpora have become available.

The Covid-19 crisis redefined administrative policies, citizens' day-to-day lives, and communications among administrations and users by presenting new methods of working and providing public services. The most recent contributions in literature highlight the need to explore how humane smart cities can help manage critical issues in the administration of smart cities through entrepreneurship, governance, and citizens' inclusion [6], [7]. We are entering a new phase in automating our critical infrastructure when it

*Corresponding author. Email: Khk00014@students.stir.ac.uk

comes to smart cities. Internet-of-Things devices now allow city operators to monitor granular levels of key municipal infrastructures and adjust resources accordingly because of their low cost.

We detect bi-lingual propaganda fragments in a given text sample and perform two types of methods one is to use a direct dataset for identification. Eng/Ur or Eng-Ur using neural machine translation. Additionally, we perform deep automatic machine learning techniques to retrieve information and fact-check on detected datasets [8], [9]. We cluster each using a k-mean cluster based on the rating score of each dataset. This paper presents a computational framework for detecting and extracting knowledge conflicts in text sources and identifying misinformation/propaganda fragments. Based on deep mining algorithms, our solution translates structured information, and identify propaganda fragment and misinformation, which makes the following contributions:

- We proposed a two-step transformers model that detects propaganda fragments in a bi-lingual text sample.
- Our model predicts and gathers the text sample through information extraction architecture, which assists the dataset in the training process.
- The translation strategies applied by the subtitle in the English translation of the text, especially within the domain of the specific topics, that is, of the parts in the propaganda fragments with different denotative and connotative meanings.

Our paper is organized as section 1 represents the introduction. Section 2 overview the brief on related work. The collection of data and the proposed method is explained in section 3. This section also explains propaganda fragment detection, which uses the NMT technique to detect the information retrieved data. Section 4 explains the results and implementation of our experiments. Section 5 concludes our paper and gives us future research directions.

2. Related work

When studying the inferences of social media on smart cities, it is vital to know the phases approximately new cities have previously approved in their placement. In a framework where mobile expertise modernizes governments and local authorities, stakeholders have gradually presented an attention in numerous KPIs (Key Performance Indicators) at a city level, using social media to simplify connections with citizens. A Rahimi et al. answer numerous claims could assist in thoughtful the overall public expressive preference, a complex factor persuading decision-makers at several administration planes.

Social media stages are shown actual in identifying the granular particulars of local communities, playing a pivotal role in encouraging responsibility and transparency within a culture. A important volume of indication pointed out that US government representatives utilized social media platforms for responsibility in governing, which also aided participants

and politicians in a more creative and precise study of a sequence of events that had earlier appeared dissimilar. Health care has gone over numerous phases called Healthcare 1.0, 2.0, 3.0, and 4.0. Individually, Khairol et al. [10] proposed these phases is a logical chain of one whole progress, ensuing in answering open challenges and earlier problems. The individual phase has its benefits, shortcomings, distinguishing features, approaches to providing health care, and technical and procedural novelties, where the research is required for considering the classification as a whole, classifying trends, and evolving future directions.

Several components have been artificially separated from the information extraction problem, including entity type classification, entity mention boundary identification, and event detection and categorization. There is rarely any feedback from the downstream classifiers when upstream errors are compounded and propagated to it, such as mislabeling an entity type [11]. Bidirectional LSTMs and a bidirectional RNN (Bi-RNN) are used in the new model for joint extraction of entities and relations. Based on a bit-wise parsing operation between two entities, the Bi-RNN structure predicts their relationship [12]. We studied the likelihood of the distribution of attention heads, the importance of the attention heads, and the impact of masking out layers [13]. The direct learning of this conditional distribution has been proposed in a variety of papers [14]. The neural machine translation (NMT) approach involves encoding sentence x source and decoding sentence y to target. The variable sentence source length encodes the target sentence using RNN fix-variable vector length [5], [15]. It is essential to convert unstructured text into a structured representation prior to developing automated knowledge analysis and fusion algorithms [16]. Thus, to increase the accuracy of cross-document entities, text fusion must solve cross-document entity co-reference problems simultaneously [17].

A recent study by Rashkin et al. (2017) examined trust, satire, hoax, and propaganda (TSHP) formats of text news for the detection of deception [18]. It is also considered as a task to fill a table with different entity extractions for multi-way classifications [19]. The active research areas cover textual record analysis involving linguistic and stylistic properties [16,17,18], in contrast to [23], a bidirectional RNN to label each pair of words. Online fraud detection has extensively used authorship attribution and stylistic cues, including sockpuppet detection in Wikipedia [24], the detection of deceptive online profiles, and the detection of fraudulent online behavior on social media [25]. These methods combine statistical classifiers with rule-based feature specifications, making high-level decisions based on the outcome of these methods. Instead, our objective is to extract knowledge information in English language.

For about twenty years, the concept of a “smart city” has received increasing attention in urban planning and governance [26], [27]. In light of current worldwide events, applying new knowledge in smart cities needs a combined structure to identify and stop a community health emergency. Numerous numerical solutions have been established during the epidemic to implement an approach to comprehend the

spread of virus, which controls human anxiety, and joint comfort, and gather complex space-time procedures in a smart city associated to Covid-19 protection methods. In several smart cities, the government set up contact-tracing apps, robots, and digital thermal gantries as well as civil society involvement in managing the spread of the virus to contain the pandemic crisis after the early first wave of the pandemic [28]. This combination proved essential to contain the pandemic crisis. Technology enables people to maintain social distance while continuing their lives as a result of a pandemic, thus mitigating negative effects.

Digital skills and the willingness to adopt new technologies are not always present among citizens, governments, and organizations. It will be important to conduct further research to understand how citizens and

individuals have adjusted to the technological changes imposed by the pandemic [29]. By strengthening citizens' involvement in policymaking, creating added value in the urban context, and enhancing crisis response capabilities, multiple technological points and their real-time data collection and sharing capabilities can significantly enhance well-being and quality of life.

We are doing work on translation knowledge information extraction and bi-lingual propaganda detection is strongly related to treating conflict patterns [30], [31]. A global conflict pattern is constructed in a data corpus (text/smart cities mining) using local conflict rules as a basis for bi-lingual multi-source knowledge conflict detection, as shown in Table 1.

Table 1. The previous proposed model and limitation

Smart cities method	Limitation
Risk management misinformation and fake news [32]	smart city risks in the context of the COVID-19 pandemic.
Survey on Computing Security over Smart City [33]	It focuses only on who accesses the data developed and conceptualized for the safety and privacy
Smart Cities Architectures [34]	It does not provide Non-Repudiation that improves the quality of life
Propaganda detection system [35]	Support for numerous languages and a <i>pull</i> mode Less instances sample that can facilitate the detection of propaganda
Character-level detection [36]	Computational Text Segregation in mixed-code
Pre-trained BERT language model (LM) [31]	Identify specific propaganda span
Prosoul [32]	Generic analysis of various aspects of the propaganda detection system
Massive amounts of misinformation, Hoaxy [33]	Track spreading and fact-checking in online information
Propaganda as Neuro-linguistic [36]	Symbolic data corpora
Emerging forthcoming human-centered smart cities [41]	Safety, strength, interpretability, and principled

2.1. Processive Model

Given the best-performing systems from the last challenge, we chose the Transformer-based model as our solution, focusing primarily on RoBERTa [42]. RoBERTa improves the language masking method of BERT by eliminating the NS (next-sentence) pre-training target and training with abundant mini-batches and LR (learning rates) [43]. In recent years, many improved models based on BERT have emerged, including DistilBERT [44], XLNET [45], ALBERT [46] etc. RoBERTa has been trained for longer than BERT on massive data provided by researchers from Facebook and Washington University. We developed a RoBERTa-based architecture with language masking and byte-level byte pair encoding (BPE) as a tokenizer to classify propaganda techniques [47]. Our model used the RoBERTa transformer for text

classification. We used a pre-trained model on a given dataset and fine-tuned it with a fully-connected (FC) layer.

RF (Random Forest Classifier), providing precise baselines on classification and regression tasks. CNB (Complement Naive Bayes Classifier), using the similar value behind MNB while modifying its expectations and execution it appropriate for imbalanced data. MNB (Multinomial Naive Bayes Classifier), calculating the likelihoods of fitting to a class as a purpose of the rate of dissimilar words.

3. Method

A lot of research attention has been devoted to transformer-based networks in the recent past, such as BERT [3] TransformerXL and RoBERTa [42]. The key mechanism for

tracking mutual influence in these models is the use of a multi-head self-attention mechanism. The primary aim of the study is to signify text data based on their text contents and then assess words trend to comprehend citizens' sentiment and perception of the calculated smart cities in the pre-Covid and post-Covid times.

3.1. Dataset

The dataset provides NLP4IF [48], and the training, the development, and the test contain 16,000 and 3,400 sentences, respectively. We also collect data through the information and knowledge extraction process and apply machine translation (Google translation) to the extraction information data (IED), as shown in Table 2.

Table 2. The statistics of train/test corpus used in our experiments

Dataset		Docume nt	Senten ce	Span	NMT
TSHP	Train	430	7,184	7,184	Eng-Ur
	Test	107	1,796	1,796	
NLP4IF	Train	16,000	-	32,800	Eng-Ur
	Test	3,400	-	6,180	
ProSoul	Train	-	7400	11600	Ur
	Test	-	1100	2300	
ProText	Train	800	11,327	11,327	Ur
	Test	200	2,260	2,260	
IED	Train	256	6,528	6,528	Eng
	Test	64	2832	2832	

3.2. Smart city tweets collection

The Twitter Advanced Search allows web scrapers to collect tweets that respond to query parameters. This application meets both adoption and privacy concerns. For pre-Covid (before the Covid period) and post-Covid (after the Covid period), we searched for tweets responding to the keyword "smart city" between January to December 2019 [49]. Only the most relevant tweets from 32,334 tweets by 22,202 users are kept after a filter based on indicators related to identified constructs (e.g., virus, infrastructure, transportation, etc.). Post-Covid, there are 15,130 tweets compared to 17,204 tweets during the pre-Covid period. As shown in Table 3, cities are distributed as follows.

Table 2. An overview of the distribution of tweets among cities.

City	Tweets
London	18,086
Milan	3,896
Dublin	1,769
Berlin	4,384

Madrid	4,199
Total	32,334

3.2. Neural machine translation Encoder-Decoder

The neural machine translation (NMT) approach is emerging newly to machine translation [14], [15]. With neural machine translation, a single, large neural network is built and trained, which reads and outputs a sentence correctly. In contrast to phrase-based translation, which contains many smaller subcomponents tuned separately. Most NMT models are based on encoder-decoders in a multi-lingual task in which an encoder and a decoder for each language.

NMT involves encoding specific languages (source) onto each sentence, then comparing the outputs to the source. We now propose a model in which each conditional likelihood is defined according to Eq. 1.

$$P(y_i/y_1, y_2, \dots, y_{i-1}, x) = g(y_{i-1}, s_i, C_i) \quad (1)$$

where s_i is hidden state (h) for time T_i , computed in RNN by

$$s_i = f(s_{i-1}; y_{i-1}; C_i) \quad (2)$$

The likelihood on each target word y_i having a distinct context C_i vector, unlike the current encoder-decoder as in Eq. 2. The depends of input sentence of context C_i vector on a hidden ($h_i = h_1, h_2, \dots, h_T$) sequence annotations on an encoder. The collection of annotation sequence h_i focus on the input of i -th word which contains information of sourrounding part of input. The C_i is, computed in Eq. 3. which sum weighted of these annotations represented.

$$C_i = \sum_{k=1}^{T_x} \alpha_{ik} h_k \quad (3)$$

The weight α_{ik} of individual annotation hidden sequence h_k is calculated by

$$\alpha_{ik} = \frac{\exp(e_{ik})}{\sum_{k=1}^{T_x} \exp(e_{ik})} \quad (4)$$

Where alignment model show the position of input/output k and i match.

$$e_{ik} = a(s_{i-1}, h_k) \quad (5)$$

The BiRNN-based hidden h_i state score s_{i-1} input sentence just before emitting y_i , and the j -th annotation h_k Eq. 5. A feedforward neural network is used to parametrize the alignment model a while other components of the proposed system are used as well. By calculating an expected annotation from a weighted sum of all annotations, we can understand the approach of taking an expected alignment. The likelihood of α_{ik} the target word y_i aligning with a source word, or converting from a source word, is x_k . Then,

context C_i vector of the i -th word is expected α_{ik} likelihoods over all the annotations. The likelihood α_{ik} , or its related e_{ik} , reflects the significance of the annotation h_k with aforementioned hidden state s_{i-1} in determining the following state s_i and making y_i . This method does not consider alignment as a latent variable like traditional machine translation, as shown in Figure. 1.

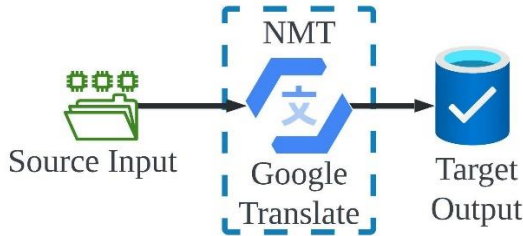


Figure. 1 Neural machine translation source-target pipeline

The input sequence \mathbf{x} in Eq. 1 described RNN starting symbol x_i to the last one x_{T_k} . The proposed method summarizes the annotation of each word, which lead us to use multi-layer bidirectional RNN and LSTM (BiRNN|BiLSTM) [50], which has been used successfully in text classification, sentiment analysis, and speech recognition [51]. A BiRNN and BiLSTM consists of forward and backward layers. The forward and backward layers RNN/LSTM \vec{f} reads the sequence input/output as forward and reverse order. The RNN/LSTM read $(x_i - x_{T_k})$ and calculates hidden states

$(\vec{h}_i = h_1; h_2, \dots, x_{T_k})$ of the \vec{f} forward layer. While the backward RNN/LSTM read $(x_i - x_{T_k})$ and calculates hidden states $(\overleftarrow{h}_i = h_1; h_2, \dots, x_{T_k})$ of the \overleftarrow{f} backward layer. The forward \vec{f} and backward hidden state $[\overleftarrow{h}_i, \vec{h}_i]$, i.e., $h_k = [\vec{h}_i^t, \overleftarrow{h}_i^t]$ concatenating an annotation for each word x_j . To summaries preceding and the following words contains the annotation h_j using forward backward RNN/LSTM approach.

3.3. Propaganda Fragment

Given a text sample, detect all fragments/spans of propaganda present in the given sample text. It is also necessary to identify, for each span, the propaganda technique that was employed. The sentence-level propaganda (SLP) and Fragment-level Propaganda (FLP) approach is followed by two components: the information extraction feature and information classification and ensemble component. Information extraction features indicate the combination of linguistic investigation, layout, and topical features of RNN and logistic regression. While our RNN/BiLSTM model, we concentrate the pruning feature in the last hidden layer before classification using RoBERTa. Our final model fine-tuned RoBERTa and achieved the SOTA results with an ensemble of Bi-RNN, CRF and BiLSTM. In the final component, we gather the likelihoods propaganda label for each sentence and thus, obtain M number of prediction classifiers for each word token, as shown in Figure. 2.

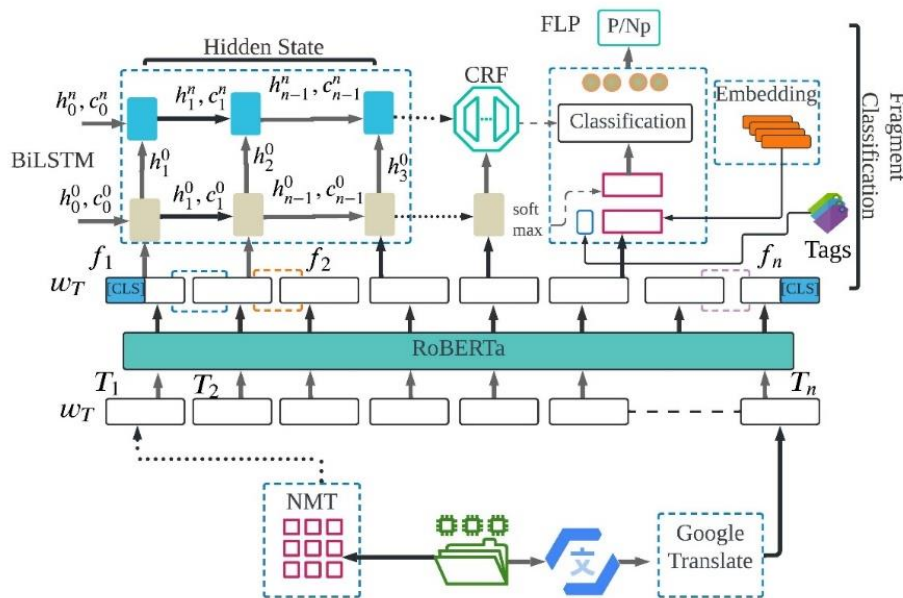


Figure 2. The proposed fragment level propaganda classification ensemble model flowchart

In multi-models, we designed fragment-level propaganda (FLP) as sequence taggers [19]. BiLSTM-CRF with word embeddings (E_w) and character embeddings E_c , tokenlevel features (E_T) I.e, POS, NER, etc. FLP and SLP performs jointly with LSTM-CRF+fine-tune that RoBERTa E_{sent} , and E_w (sentence and word embedding) respectively.

We fine-tune RoBERTa and Ensemble Multi-task BiLSTM-CRF, in each sequence tagger considering propagandista span/fragments. When the span overlaps exactly, our model performs at fragment level majority voting for span/fragment, while non-overlap span/fragment our model considers all and truncates it. But when span/fragment overlap with the identical label our model selects the larger span, as shown in Figure. 3.

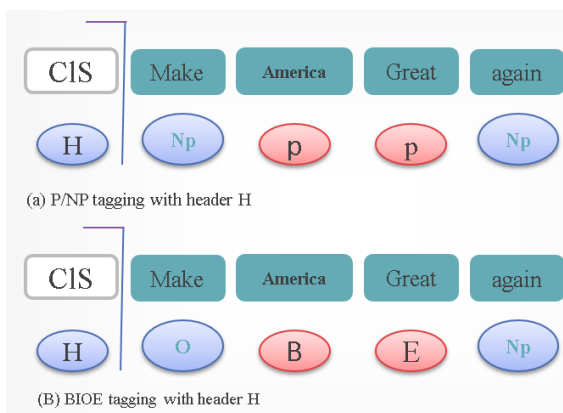


Figure. 3. Prop/NProp and BIOES tags samples with the given instance

4. Method

The model trained Nvidia RTX 3090 Ti graphic cards. A Python-Torch-transformers package is used to implement our implementation. A mixed precision model was used to train all the models in the aim of accelerating training time.

4.1. Evaluation metrics

The F1-score can be classified as a primary metric in the context of our study, whereas P and R can be classified as a secondary metrics, as shown in Eq. 6.

$$F_1 = \left(\frac{P * R}{P + R} \right) * 2 \tag{6}$$

$$P = \text{Precision} = \frac{TP}{TP + FP}$$

$$R = \text{Recall} = \frac{TP}{TP + FN}$$

Where TP is true positive, FP, FN is a false positive and false negative, respectively, while P = precision =

$\frac{TP}{TP+FP}$, $R = \text{Recall} = \frac{TP}{TP+FN}$ are statistics for the binary classification. Evaluation metrics for classification were used the following metrics evaluating the rule-based classification results.

As RoBERTa-based is case-insensitive, we chose a version that uses lowercase letters throughout the text, whereas BioBERT only uses case-insensitive letters. According to Table 3, all models were based on the same hyperparameter values. Using Tensorflow, the RoBERTa layer's hyper parameters were the same as those used in the RoBERTa layer.

Table 3 List of all hyperparameter values used in experiments.

Parameter	Values
Learning rates	3×10^{-5}
Epochs	15
Bach-size	16-64
Sequence length	80-150
Embedding size	8-16
Token size	4-50

4.2. Implementation process

The results obtained on bi-lingual corpus, which identify propaganda fragment, with fine-tuned RoBERTa, and ensemble our model with Bi-RNN, CRF and BiLSTM model. Our model with twelve layers and 64 batches, 180 sequences, 0.1 weight decay, and 15 epochs were trained on all models, on ProText Ur, as shown in Table 4.

Table 4. fragment identification on Bi-lingual corpus results ProText Ur.

Model	F1	P	RI
Baseline	0.4134	0.4029	0.4188
RoBERTa-RNN	0.4657	0.4780	0.4622
RoBERTa-CRF	0.5969	0.5803	0.5811
RoBERTa-BiLSTM	0.6615	0.6628	0.6529
RoBERTa-BiLSTM-CRF	0.7031	0.7002	0.6984
Our TC	0.8871	0.8840	0.8798

Our model use the RoBERTa baseline show low results with F1-score 0.4134, precision 0.402, and recall 0.418. RoBERTa-RNN tokenized with CRF with F1-score of 0.4657, precision 0.478, and recall 0.462. RoBERTa-CRF tokenized with CRF with F1-score of 0.5969, precision 0.5803, and recall 0.5811. RoBERTa-BiLSTM tokenized with CRF with F1-score of 0.661, precision 0.662, and recall 0.652. RoBERTa-BiLSTM-CRF tokenized with CRF with F1-score of 0.703, precision 0.7002, and recall

EAI Endorsed Transactions on

0.698. Our tokenized with CRF with F1-score of 0.887, precision 0.884, and recall 0.879. The corpus on fragment identification on ProText Eng (English), as shown in Table 5.

Our model use the RoBERTa baseline show low results with F1-score 0.452, precision 0.461, and recall 0.451. RoBERTa-RNN tokenized with CRF with F1-score of 0.312, precision 0.328, and recall 0.320. RoBERTa-CRF tokenized with CRF with F1-score of 0.490, precision 0.485, and recall 0.489. RoBERTa-BiLSTM tokenized with CRF with F1-score of 0.545, precision 0.539, and recall 0.541. RoBERTa-BiLSTM-CRF tokenized with CRF with F1-score of 0.516, precision 0.509, and recall 0.519. Our tokenized with CRF with F1-score of 0.628, precision 0.624, and recall 0.628. The corpus on fragment identification on ProText Eng (English), as shown in Table 6.

Table 5. Fragment identification on Bi-lingual corpus results on ProText Eng.

Model	F1	P	R
Baseline	0.4527	0.4613	0.4518
RoBERTa-RNN	0.3120	0.3283	0.3204
RoBERTa-CRF	0.4909	0.4856	0.4891
RoBERTa-BiLSTM	0.5425	0.5399	0.5419
RoBERTa-BiLSTM-CRF	0.5164	0.5098	0.5197
Our TC	0.6285	0.6243	0.6280

Table 6. Fragment identification on Bi-lingual corpus results on ProText NMT.

Model	F1	P	R
Baseline	0.6582	0.6677	0.6648
RoBERTa-RNN	0.5378	0.5289	0.5274
RoBERTa-CRF	0.6909	0.6896	0.6886
RoBERTa-BiLSTM	0.7815	0.7628	0.7649
RoBERTa-BiLSTM-CRF	0.8137	0.8095	0.8177
Our TC	0.9071	0.8940	0.8898

Compared to the Protex with Ur, and Eng, the NMT achieved F1-score of 0.658, precision of 0.667, and recall of 0.664 on the baseline. RoBERTa-RNN tokenized with CRF with F1-score of 0.537, precision 0.528, and recall 0.527. RoBERTa-CRF tokenized with CRF with an F1-score of 0.690, precision 0.689, and recall 0.688. RoBERTa-BiLSTM tokenized with CRF with F1-score of 0.781, precision 0.762, and recall 0.764. RoBERTa-BiLSTM-CRF tokenized with CRF with F1-score of 0.813, precision 0.809, and recall 0.817. Our tokenized with CRF

with an F1-score of 0.907, a precision of 0.894, and recall of 0.889. The ProText Ur uses Google translation to achieve the target language from ProText Eng. In the above performance comparison, the ProText NMT corpus shows a higher F1-score of 0.9071 among all other corpora. Finally, the evaluation performance accuracy is compared with 100 examples for each dataset, as shown in Figure 4.

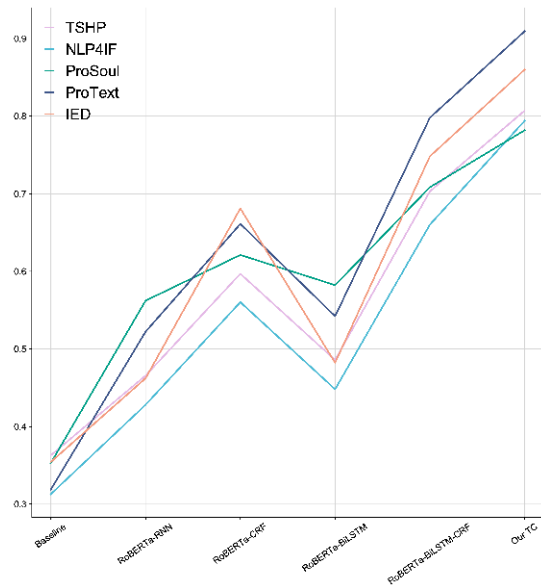


Figure 4. Prop model evaluation performance comparison with the given instance (100) of various datasets

In Covid-19, the widespread propaganda sentiment analysis that appeared from civilians' tweets from various cities (Milan, Madrid, London, and Dublin) is propagandist. The analysis reveals that in Madrid, the propaganda is created from the awareness of the failure to use identity and technology, the requirement to raise information and study how digital tools would be employed. Furthermore, p-values are re-counted to estimate the coefficients' degree of awareness (importance). The dissimilar significance values are $p < 0.26$, as shown in Figure 5.

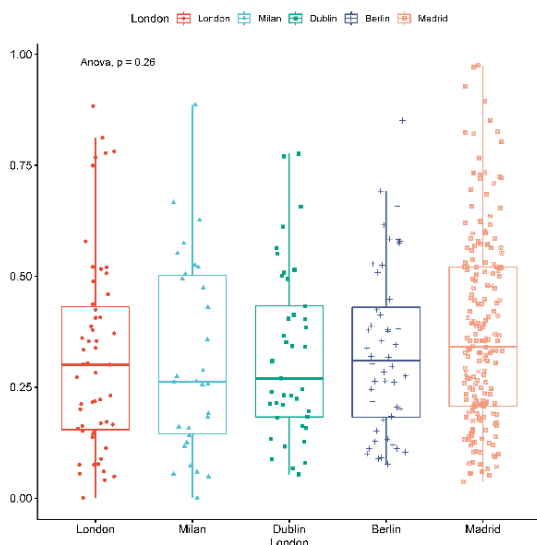


Figure 5. Covid-19, the widespread propaganda sentiment analysis in cities-based

4. Conclusion and future direction

The rapid spread of information on social media has resulted in an increase in misinformation, false news, and propaganda, which is most challenging aspects of this study, securing high-quality bi-lingual propaganda data from news sources. Our system analyzes bi-lingual sources and extracts information automatically based on a natural language processing perspective. Using neural machine translation or Google translation, to get Urdu or English bi-lingual corpus instances is explore neural architectures, as well as extract linguistic features. Furthermore, we use fine-tuned RoBERTa, a pretrained language model, to perform word-level classification as well as experiments with different ensemble schemes such as majority votes and relax votes. Moreover, the misinformation and propaganda on social media are detected in the framework, and it can be comprehensive to smart cities contexts. Additionally, we describe the model's learning abilities and aspects for improvement in our analysis of the attention heads in the model. We will also discuss the context of smart cities and how they could enhance a city's smartness through NLP.

References

- [1] G. S. Jowett and V. O'donnell, *Propaganda & persuasion*. Sage publications, 2018.
- [2] G. D. S. Martino, S. Yu, A. Barrón-Cedeño, R. Petrov, and P. Nakov, "Fine-grained analysis of propaganda in news articles," *arXiv preprint arXiv:1910.02517*, 2019.
- [3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [4] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [5] J. Devlin, R. Zbib, Z. Huang, T. Lamar, R. Schwartz, and J. Makhoul, "Fast and robust neural network joint models for statistical machine translation," in *proceedings of the 52nd annual meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2014, pp. 1370–1380.
- [6] E. Costales, "Identifying sources of innovation: Building a conceptual framework of the Smart City through a social innovation perspective," *Cities*, vol. 120, p. 103459, 2022.
- [7] S. Zeng, Y. Hu, and C. Llopis-Albert, "Stakeholder-inclusive multi-criteria development of smart cities," *Journal of Business Research*, vol. 154, p. 113281, 2023.
- [8] S. Ahmed, K. Hinkelmann, and F. Corradini, "Fact Checking: An Automatic End to End Fact Checking System," in *Combating Fake News with Computational Intelligence Techniques*. Springer, 2022, pp. 345–366.
- [9] A. Ali, M. F. Pasha, O. H. Fang, R. Khan, M. A. Almaiah, and A. K. Al Hwaitat, "Big Data Based Smart Blockchain for Information Retrieval in Privacy-Preserving Healthcare System," in *Big Data Intelligence for Smart Applications*, Y. Baddi, Y. Gahi, Y. Maleh, M. Alazab, and L. Tawalbeh, Eds. Cham: Springer International Publishing, 2022, pp. 279–296. doi: 10.1007/978-3-030-87954-9_13.
- [10] K. A. B. Ahmad, H. Khujamatov, N. Akhmedov, M. Y. Bajuri, M. N. Ahmad, and A. Ahmadian, "Emerging trends and evolutions for smart city healthcare systems," *Sustainable Cities and Society*, vol. 80, p. 103695, 2022.
- [11] Q. Li, H. Ji, and L. Huang, "Joint event extraction via structured prediction with global features," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2013, pp. 73–82.
- [12] M. Miwa and M. Bansal, "End-to-end relation extraction using lstms on sequences and tree structures," *arXiv preprint arXiv:1601.00770*, 2016.
- [13] K. Clark, U. Khandelwal, O. Levy, and C. D. Manning, "What does bert look at? an analysis of bert's attention," *arXiv preprint arXiv:1906.04341*, 2019.
- [14] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, vol. 27, 2014.
- [15] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.
- [16] Q. Li and H. Ji, "Incremental Joint Extraction of Entity Mentions and Relations.," in *ACL (1)*, 2014, pp. 402–412.
- [17] S. Singh, S. Riedel, B. Martin, J. Zheng, and A. McCallum, "Joint inference of entities, relations, and coreference," in *Proceedings of the 2013 workshop on Automated knowledge base construction*, 2013, pp. 1–6.
- [18] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking," in *Proceedings of the 2017 conference on empirical methods in natural language processing*, 2017, pp. 2931–2937.
- [19] P. Gupta, H. Schütze, and B. Andrassy, "Table filling multi-task recurrent neural network for joint entity and relation extraction," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 2537–2547.
- [20] S. Afroz, M. Brennan, and R. Greenstadt, "Detecting hoaxes, frauds, and deception in writing style online," in

- 2012 *IEEE Symposium on Security and Privacy*, 2012, pp. 461–475.
- [21] P. Juola, “Detecting stylistic deception,” in *Proceedings of the Workshop on Computational Approaches to Deception Detection*, 2012, pp. 91–96.
- [22] L. Zhou, D. P. Twitchell, T. Qin, J. K. Burgoon, and J. F. Nunamaker, “An exploratory study into deception detection in text-based computer-mediated communication,” in *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*, 2003, pp. 10–pp.
- [23] M. Miwa and Y. Sasaki, “Modeling joint entity and relation extraction with table representation,” in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1858–1869.
- [24] L. Song, W. Zhang, S. SY Liao, and R. CW Kwok, “A critical analysis of the state-of-the-art on automated detection of deceptive behavior in social media,” 2012.
- [25] T. Solorio, R. Hasan, and M. Mizan, “A case study of sockpuppet detection in wikipedia,” in *Proceedings of the Workshop on Language Analysis in Social Media*, 2013, pp. 59–68.
- [26] F. Loia, “A Co-evolutionary Perspective on Data-driven Organization: Highlights from Smart Cities in the Covid-19 Era,” in *Big Data and Decision-Making: Applications and Uses in the Public and Private Sector*, Emerald Publishing Limited, 2023, pp. 181–201.
- [27] A. van Twist, E. Ruijter, and A. Meijer, “Smart cities & citizen discontent: A systematic review of the literature,” *Government Information Quarterly*, p. 101799, 2023.
- [28] W. Basmi, A. Boulmakoul, L. Karim, and A. Lbath, “Distributed and scalable platform architecture for smart cities complex events data collection: Covid19 pandemic use case,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 75–83, 2021.
- [29] H. Abusaada and A. Elshater, “COVID-19 challenge, information technologies, and smart cities: considerations for well-being,” *International Journal of Community well-being*, vol. 3, no. 3, pp. 417–424, 2020.
- [30] H. Ji, H. Deng, and J. Han, “Uncertainty reduction for knowledge discovery and information extraction on the world wide web,” *Proceedings of the IEEE*, vol. 100, no. 9, pp. 2658–2674, 2012.
- [31] G. Levchuk, M. Jackobsen, and B. Riordan, “Detecting misinformation and knowledge conflicts in relational data,” in *Signal Processing, Sensor/Information Fusion, and Target Recognition XXIII*, 2014, vol. 9091, pp. 235–248.
- [32] M. Petrova and I. Tairov, “Solutions to Manage Smart Cities’ Risks in Times of Pandemic Crisis,” *Risks*, vol. 10, no. 12, p. 240, 2022.
- [33] A. I. Tahirkheli *et al.*, “A survey on modern cloud computing security over smart city networks: Threats, vulnerabilities, consequences, countermeasures, and challenges,” *Electronics*, vol. 10, no. 15, p. 1811, 2021.
- [34] A. Sajid, S. W. Shah, and T. Magsi, “Comprehensive Survey on Smart Cities Architectures and Protocols,” *EAI Endorsed Transactions on Smart Cities*, vol. 6, no. 18, 2022.
- [35] A. Barrón-Cedeno, I. Jaradat, G. Da San Martino, and P. Nakov, “Proppy: Organizing the news based on their propagandistic content,” *Information Processing & Management*, vol. 56, no. 5, pp. 1849–1864, 2019.
- [36] A. Tundis, G. Mukherjee, and M. Mühlhäuser, “An Algorithm for the Detection of Hidden Propaganda in Mixed-Code Text over the Internet,” *Applied Sciences*, vol. 11, no. 5, p. 2196, 2021.
- [37] O. Altit, M. Abdullah, and R. Obiedat, “JUST at SemEval-2020 task 11: Detecting propaganda techniques using BERT pre-trained model,” in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, 2020, pp. 1749–1755.
- [38] S. Kausar, B. Tahir, and M. A. Mehmood, “ProSOUL: a framework to identify propaganda from online Urdu content,” *IEEE access*, vol. 8, pp. 186039–186054, 2020.
- [39] C. Shao, G. L. Ciampaglia, A. Flammini, and F. Menczer, “Hoaxy: A platform for tracking online misinformation,” in *Proceedings of the 25th international conference companion on world wide web*, 2016, pp. 745–750.
- [40] R. Torok, “Symbiotic radicalisation strategies: Propaganda tools and neuro linguistic programming,” 2015.
- [41] K. Ahmad, M. Maabreh, M. Ghaly, K. Khan, J. Qadir, and A. Al-Fuqaha, “Developing future human-centered smart cities: Critical analysis of smart city security, Data management, and Ethical challenges,” *Computer Science Review*, vol. 43, p. 100452, 2022.
- [42] Y. Liu *et al.*, “Roberta: A robustly optimized bert pretraining approach,” *arXiv preprint arXiv:1907.11692*, 2019.
- [43] V. Balakrishnan, Z. Shi, C. L. Law, R. Lim, L. L. Teh, and Y. Fan, “A deep learning approach in predicting products’ sentiment ratings: a comparative analysis,” *The Journal of Supercomputing*, vol. 78, no. 5, pp. 7206–7226, 2022.
- [44] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” *arXiv preprint arXiv:1910.01108*, 2019.
- [45] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, “Xlnet: Generalized autoregressive pretraining for language understanding,” *Advances in neural information processing systems*, vol. 32, 2019.
- [46] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, “Albert: A lite bert for self-supervised learning of language representations,” *arXiv preprint arXiv:1909.11942*, 2019.
- [47] Z. Abbasiataeb and S. Momtazi, “Text-based question answering from information retrieval and deep neural network perspectives: A survey,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 11, no. 6, p. e1412, 2021.
- [48] G. D. S. Martino, A. Barrón-Cedeño, and P. Nakov, “Findings of the NLP4IF-2019 Shared Task on Fine-Grained Propaganda Detection,” *arXiv preprint arXiv:1910.09982*, 2019.
- [49] O. Troisi, G. Fenza, M. Grimaldi, and F. Loia, “Covid-19 sentiments in smart cities: The role of technology anxiety before and during the pandemic,” *Computers in Human Behavior*, vol. 126, p. 106986, 2022.
- [50] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
- [51] A. Graves, N. Jaitly, and A. Mohamed, “Hybrid speech recognition with deep bidirectional LSTM,” in *2013 IEEE workshop on automatic speech recognition and understanding*, 2013, pp. 273–278.