IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 2, June 2024, pp. 2403∼2412
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i2.pp2403-2412

From recurrent neural network techniques to pre-trained models: emphasis on the use in Arabic machine translation

Nouhaila Bensalah[1], Habib Ayad[1], Abdellah Adib[1], Abdelhamid Ibn El Farouk[2]
[1]Laboratory of Mathematics, Computer Science, and Applications, Department of Mathematics, Faculty of Sciences and Technologies, University of Hassan II Casablanca, Casablanca, Morocco
[2]Laboratory of Teaching, Languages, and Cultures , Department of Linguistics, Faculty of Arts and Humanities, University of Hassan II Casablanca, Casablanca, Morocco

Article Info

Article history:

Received Feb 3, 2023
Revised May 23, 2023
Accepted Jun 12, 2023

Keywords:

AraBERT
Arabic machine translation
AraGPT
Attention mechanism
Natural language processing
Pre-trained language models
Transformers

ABSTRACT

In recent years, neural machine translation (NMT) has garnered significant attention due to its superior performance compared to traditional statistical machine translation. However, NMT's effectiveness can be limited when translating between languages with dissimilar structures, such as English and Arabic. To address this challenge, recent advances in natural language processing (NLP) have introduced unsupervised pre-training of large neural models, showing promise for enhancing various NLP tasks. This paper proposes a solution that leverages unsupervised pre-training of large neural models to enhance Arabic machine translation (MT). Specifically, we utilize pre-trained checkpoints from publicly available Arabic NLP models, like Arabic bidirectional encoder representations from transformers (AraBERT) and Arabic generative pre-trained transformer (AraGPT), to initialize and warm-start the encoder and decoder of our transformer-based sequence-to-sequence model. This approach enables us to incorporate Arabic-specific linguistic knowledge, such as word morphology and context, into the translation process. Through a comprehensive empirical study, we rigorously evaluated our models against commonly used approaches in Arabic MT. Our results demonstrate that our pre-trained models achieve new state-of-the-art performance in Arabic MT. These findings underscore the effectiveness of pre-trained checkpoints in improving Arabic MT, with potential real-world applications.

Corresponding Author:

Nouhaila Bensalah
Laboratory of Mathematics, Computer Science, and Applications, Department of Mathematics
Faculty of Sciences and Technologies, University of Hassan II Casablanca
Casablanca, Morocco
Email: nouhaila.bensalah@etu.fstm.ac.ma

1. INTRODUCTION

Over the last few years, machine translation (MT) has been extremely valuable in a wide range of applications and has made progress for almost all languages [1]-[6] . Indeed, low-resource languages' limited training corpora result in worse translation performance. Furthermore, given that utilizing an open vocabulary in MT systems yields high computational cost, such systems constrain the vocabulary to those words that occur most often in the training corpus. This degrades the performance of the system, especially for morphologically rich languages, since many words are ignored (out of vocabulary (OOV)) in the target vocabulary, and therefore remain unknown to the system. A lot of attention has been devoted to the Arabic language in the MT community in the past decade. Arabic is the official language of 25 countries, it is primarily spoken by more than 375 million people and is ranked as the fifth most spoken language in the world. It is a language that is written from right to left, using a cursive script, with 28 letters in its alphabet. These letters are consonants and vowels. However, the morphology of the Arabic language, along with other linguistic aspects, has made MT to and from Arabic much more difficult. The morphological richness of this language, which is characterized by the high presence of the agglutination phenomenon, makes that an Arabic word may represent an entire sentence in English, as illustrated by the word ”وبمدارسهم” (/wbmdArshm/) which means in English “and in their schools”. The phenomenon of agglutination in certain languages results in an increased number of OOV words in neural machine translation (NMT) systems. To address these challenges, researchers have explored alternative models that utilize smaller orthographic vocabulary units instead of complete words. One approach is to represent words as sequences of characters, which can be achieved through techniques like byte pair encoding (BPE) [7], or even by considering individual characters as the basic units. These alternatives successfully dealt with the OOV problem but involved a significant drop in semantic and syntactic information, resulting in mistranslations [8], [9].

MT can be classified into three main categories: into rule-based machine translation (RBMT), statistical machine translation (SMT), and NMT. RBMT relies on linguistic rules created by language experts, making it dependent on extensive dictionaries and significant linguistic knowledge [10]. However, building such resources can be expensive, and it is challenging to create rules that cover all languages. SMT, on the other hand, is a data-driven approach that employs probabilistic models. It consists of three primary stages: the translation model, the language model, and the decoder model. The translation model estimates the probability that a source sentence corresponds to a target sentence based on a bilingual corpus. The language model, trained on a monolingual corpus, enhances the fluency of the translation. In the decoding phase, the most probable target sentence is determined using the language and translation models. SMT can handle ambiguity by utilizing a phrase table that records phrase-based translations and their frequency of occurrence, resulting in more fluent and natural translations compared to RBMT [11].

SMT has been known to struggle when translating sentences that significantly differ from the content in the training data [10]. In recent years, NMT has gained substantial attention from the research community due to its remarkable performance [12]-[15]. NMT models employ an end-to-end encoder decoder framework. In this architecture, the encoder plays a crucial role in converting a source sentence into a continuous vector, commonly known as a context vector. This vector captures the pertinent information derived from the input sentence. Once the encoder has produced the context vector, the decoder utilizes it to generate the translation in the target language, progressing word by word. Furthermore, there has been a recent surge in the use of large pre-trained transformer-driven language models (PTMs), such as the bidirectional encoder representations from transformers (BERT) [16] and generative pre-trained transformer (GPT) [17] families of models, have been storming natural language, attaining peak efficiency in many tasks. The attractive side to this overwhelming push towards using large architectures pre-trained on massive collections of text is that the pre-trained checkpoints, as well as the inference code, are freely accessible. This can result in saving hundreds of tensor processing unit (TPU)/ graphics processing unit (GPU) hours, as warm-starting a model using a pre-trained checkpoint generally required fewer fine-tuning steps, while still achieving substantial improvements in performance. More significantly, the feasibility of starting from a state-of-the-art performance model such as BERT motivates the community to significantly advance toward developing both improved and easily reusable MT systems. However, despite the success of these PTMs in tasks such as Glue and stanford question answering dataset (SQuAD), there is still a need for research to explore their potential for other applications, particularly in the area of sequence-to-sequence (Seq2Seq) models for MT. Arabic is one language that could benefit from this research, as there is a growing demand for MT systems that can accurately translate Arabic text into other languages. Hence, in this paper, we present a transformer-based Seq2Seq model for Arabic MT that leverages the publicly available AraBERT and AraGPT-2 pre-trained checkpoints. Our model is initialized using a combination of these checkpoints, and we explore various settings to find the optimal initialization method. We show that our approach outperforms randomly initialized models and achieves new state-of-the-art results in Arabic MT

The rest of the paper is organized as follows. Section 2 summarizes the research work that has been done with regard to Arabic MT. Section 3 describes the models and pre-trained checkpoints used in this work. Section 4 reports the experiments considered in this paper and discusses the findings. Lastly, a conclusion and future perspectives are set out.

2. RELATED WORKS

Over the past years, there has been a notable surge in research studies focused on the NMT paradigm. In this section, we categorize the existing research on Arabic NMT into two primary classifications:

– Pre- and post-processing: these studies aim to improve the quality of NMT systems by utilizing pre-processing and/or post-processing techniques. This includes techniques such as segmentation, normalization, tokenization, and post-processing re-scoring. The focus is on optimizing the input data and refining the output translations to improve overall performance.
– Morphology, vocabulary, and factored NMT: this category investigates the incorporation of diverse linguistic knowledge sources into baseline NMT systems. The studies investigate the impact of incorporating morphological information, exploring different vocabulary sizes and subword units, and incorporating hierarchical or factored approaches to improve translation quality. These approaches leverage linguistic factors to enhance the NMT models.

2.1. Pre- and post-processing

Several research studies have dedicated their focus to enhancing Arabic NMT baselines through pre- and post-processing techniques. Sajjad *et al.* [18] conducted a comparative analysis of language-independent segmentations, such as BPE, character-level encoding, and character convolutional neural network (CNN). Their findings indicated that BPE achieved the most favorable outcomes, surpassing even state-of-the-art morphological segmentation methods. Oudah *et al.* [19] delved into different segmentation approaches for both neural and statistical Arabic-English MT models. They observed that morphology-based segmentation, particularly the one employed in the Arabic TreeBank (ATB), proved beneficial for both NMT and SMT models. The combination of ATB with BPE yielded the most promising results for SMT models. Ameur *et al.* [20] proposed a post-processing method for n-best list re-scoring in NMT, utilizing features extracted from parallel corpora. Their approach achieved noticeable improvements in translation quality. Alrajeh *et al.* [21] explored Arabic preprocessing and found that it improved translation quality for both NMT and SMT systems. These studies highlight the significance of pre-and post-processing techniques in enhancing Arabic NMT systems. Optimal segmentation methods, appropriate preprocessing steps, and effective post-processing approaches contribute to improving translation quality. Combining linguistic knowledge with optimization algorithms can further enhance the performance of Arabic NMT systems.

2.2. Morphology, vocabulary, and factored neural machine translation

Several research studies have explored different approaches to enhance NMT models through the incorporation of linguistic knowledge. Ding *et al.* [22] determined the optimal vocabulary size for NMT models using subword units and found that smaller vocabulary sizes, containing less than 1,000 subword units, achieved the highest bilingual evaluation understudy (BLEU) scores. Ataman *et al.* [23] proposed a hierarchical latent variable approach to incorporate morphological inflection, resulting in a slight improvement in translation quality for morphologically rich languages like Arabic. Ataman *et al.* [24] introduced a hierarchical decoding method that considers both words and characters during translation generation, outperforming subword-level techniques in terms of translation quality with significantly fewer parameters. Liu *et al.* [25] shared source and target word embedding features in NMT systems, combining bilingual and monolingual characteristics, and achieved a significant performance increase over the baseline transformer model. These studies demonstrate the potential benefits of incorporating linguistic knowledge in NMT models. Optimized vocabulary sizes, modeling morphological inflection, hierarchical decoding, and shared word embedding features contribute to improved translation quality in various language pairs, including Arabic-English.

3. MODELS AND PRE-TRAINED CHECKPOINTS

Our exploration involved the analysis of different transformer encoder-decoder models, initialized in various ways. These initialization methods included random initialization and warm-starting by utilizing public checkpoints from BERT and GPT-2. The diverse initialization approaches allowed us to assess their impact on the MT' performance and capabilities; as shown in Figure 1.

3.1. Bidirectional encoder representations from transformers checkpoints

In this paper, we adopt AraBERT which is a pre-trained Arabic language model based on the BERT architecture developed by Google. AraBERT uses the same BERT-base configuration, consisting of 12 transformer layers, 768 hidden units, and 12 self-attention heads. We distinguish two versions of the model, AraBERTv0.1 and AraBERTv1. AraBERTv1 adopts pre-segmented text where prefixes and suffixes have been split by means of the Farasa segmenter [26]. It segments words into stems, prefixes, and suffixes, allowing the model to better handle Arabic morphology. Alternatively, a SentencePiece (an unsupervised text tokenizer and detokenizer) is trained on unsegmented text to generate the second release of ARABERT (AraBERTv0.1) that involves no segmentation. The model was trained on a large-scale dataset composed of a combination of Arabic Wikipedia, Arabic Gigaword, and OSCAR Arabic. This version of the model is particularly useful for tasks where pre-segmented text is not available, such as social media or dialectal Arabic. The final vocabulary size is also 64k tokens, but it includes fewer subword units than AraBERTv1.

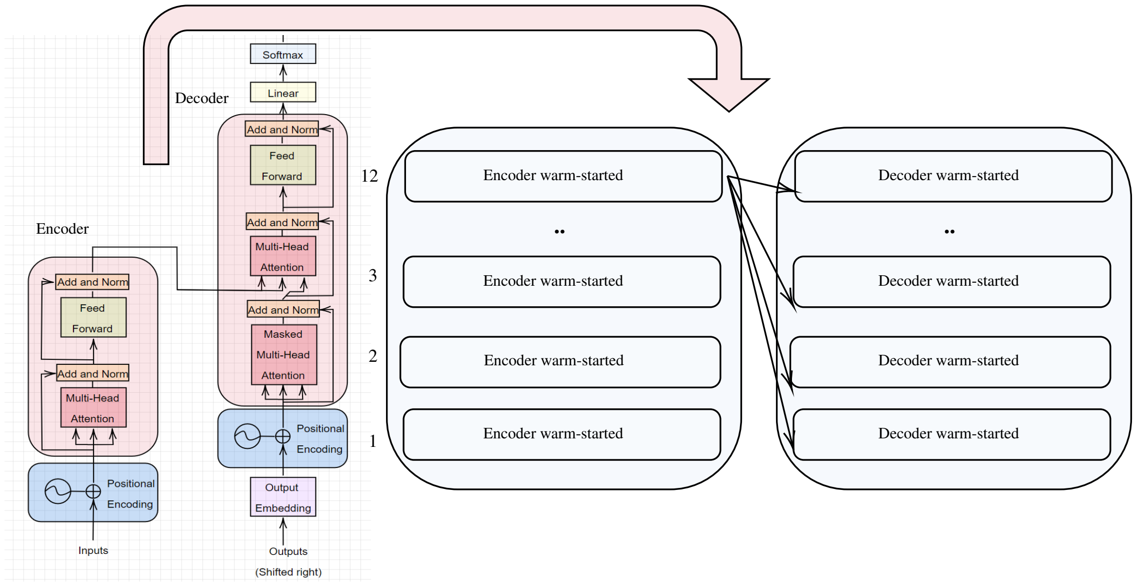

Figure 1. The proposed architecture

3.2. Generative pre-trained transformer-2 checkpoints

In this work, we use AraGPT2 which is the first advanced model for Arabic language generation that relies on the transformer architecture [27]. The model was trained on the largest publicly available collection of filtered Arabic corpus, which includes a wide range of text genres, such as news articles, web pages, and literary texts. AraGPT2 follows the variant architectures and training process of GPT-2 closely, with some modifications tailored to the Arabic language. The model consists of a varying number of transformer layers, ranging from 12 to 48, depending on the version of the model. Four versions of the model are available, AraGPT2-base, AraGPT2-medium, AraGPT2-large, and AraGPT2-mega, which differ in the number of parameters and computational resources required. Most of the training data are composed of Arabic news articles, which are mainly written in modern standard Arabic (MSA). However, the model has also been fine-tuned on various downstream tasks, such as dialect identification and named entity recognition, to improve its performance on specific applications. The total dataset size is 77 GB with 8.8 billion words, making it one of the largest publicly available Arabic language models.

4. EXPERIMENTS AND RESULTS

In this paper, we aim to investigate translation from Arabic to English. Therefore, a subset of the web inventory of transcribed and translated talks (WIT^3) corpus of technology, entertainment, and design (TED) talks [28], [29] provided for International Workshop on Spoken Language Translation (IWSLT) 2016 is used to validate our settings. A training set consisting of 108,000 sentences was utilized, while 437 sentences were allocated for validation and 524 for testing purposes. The input and output lengths were constrained to 64 tokens each. Training was conducted over 100 epochs, employing a batch size of 256. During decoding, a bundle size of 4 was employed, and the sentence length penalty was set to the default value of α=0.6. The evaluation metric employed was the BLEU score. To begin, we provide a description of the chosen combinations for model initialization:

– RND2RND: a transformer model with both the encoder and decoder initialized randomly.

- AraBERT2RND: an architecture consisting of an AraBERT-initialized encoder and a randomly initialized decoder. The encoder and decoder share the same embedding matrix initialized from a pre-trained AraBERT model.
- RND2AraBERT: an architecture where the encoder is randomly initialized while the decoder is AraBERT-initialized. Autoregressive decoding is performed by masking the bidirectional self-attention mechanism of AraBERT to consider only the left context.
- AraBERT2AraBERT: an architecture with both the encoder and decoder initialized from a publicly available AraBERT checkpoint. The only randomly initialized component is the encoder-decoder attention.
- AraBERTSHARE: similar to AraBERT2AraBERT, but the parameters between the encoder and decoder are shared. This significantly reduces the memory requirements of the model.
- RND2AraGPT: an architecture featuring a randomly initialized encoder and an AraGPT-2-compatible decoder. The decoder and embedding matrix are warm-started using a publicly available AraGPT-2 checkpoint.
- AraBERT2AraGPT: an architecture combining an AraBERT-compatible encoder and an AraGPT-2-compatible decoder. Both sides of the model are warm-started separately using the publicly available AraBERT and AraGPT-2 checkpoints. The AraBERT vocabulary is used for input, while the AraGPT-2 vocabulary is used for output.

In our study, we conducted a comprehensive comparison of seven deep learning (DL) models in the Arabic-English MT context. These models include variants of recurrent neural networks (RNNs) such as long short term memory (LSTM), gated recurrent unit (GRU), bidirectional long short term memory (BiLSTM), and bidirectional gated recurrent unit (BiGRU), used as both encoder and decoder components. Additionally, we incorporated the attention mechanism in each model and experimented with different word embeddings, namely Word2Vec, GloVe, and FastText. This extensive evaluation allowed us to assess the performance and effectiveness of each model configuration in the specific task of Arabic-English MT.

The results presented in Figures 2-4 reflect that the highest performing model in terms of BLEU score is composed of BiGRU as an encoder and BiLSTM as a decoder with attention mechanism and FastText embeddings (BLEU score = 42.18%). The findings suggest that the model utilized the advantages of BiGRU, known for faster training compared to BiLSTM. Additionally, it benefited from BiLSTM, which demonstrated better performance in terms of BLEU score. The inclusion of FastText also contributed to the model's effectiveness, as it considers the internal information of subwords, enabling the model to capture word morphology and lexical similarity effectively. Furthermore, Arabic preprocessing techniques were applied, involving multiple phases such as normalization and tokenization based on the Farasa. Examining the findings in Figures 5-7, it seemed that the Arabic preprocessing was efficient while being applied to the Arabic sentences.

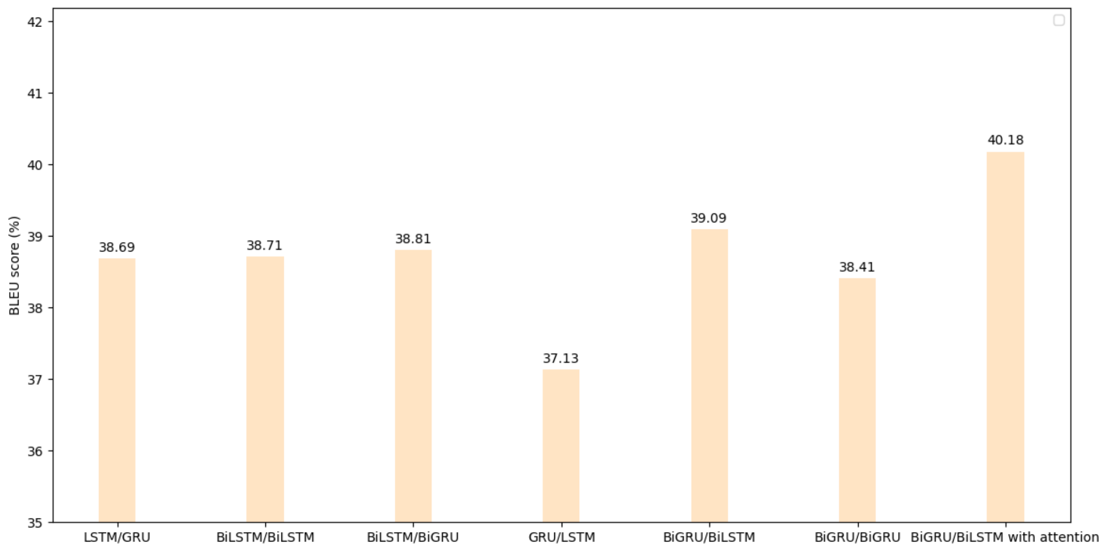

Figure 2. Performance evaluation of DL encoder-decoder models using Word2Vec embeddings without preprocessing

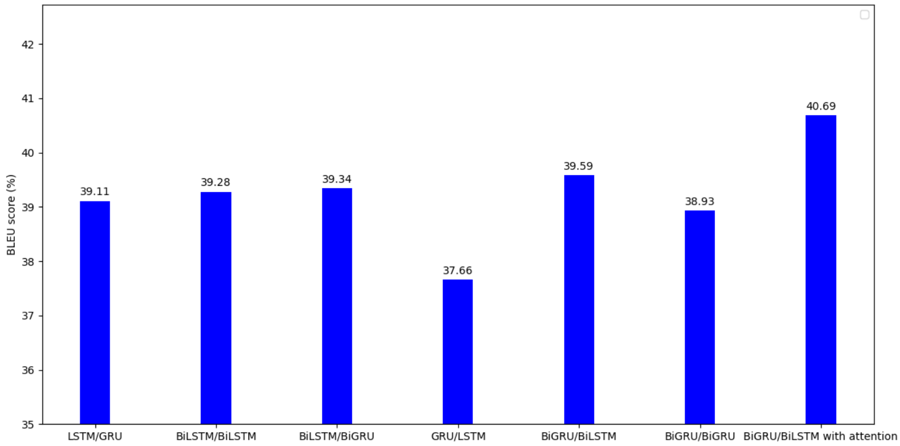

Figure 3. Performance evaluation of DL encoder-decoder models using GloVe embeddings without preprocessing

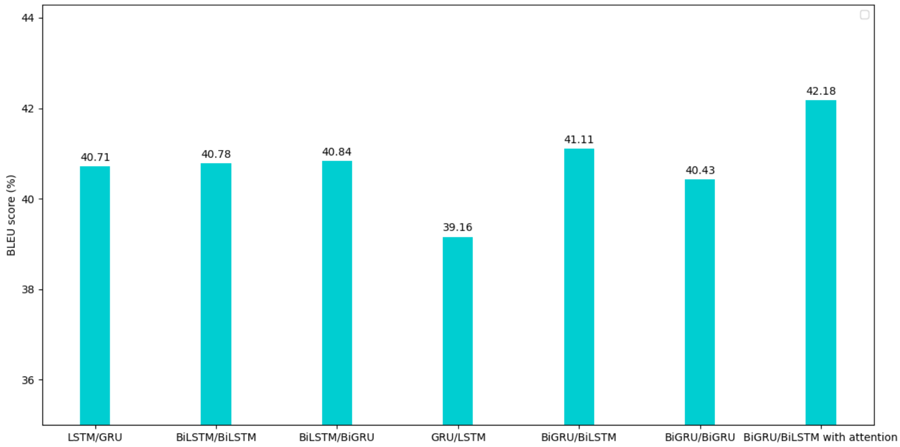

Figure 4. Performance evaluation of DL encoder-decoder models using FastText embeddings without preprocessing

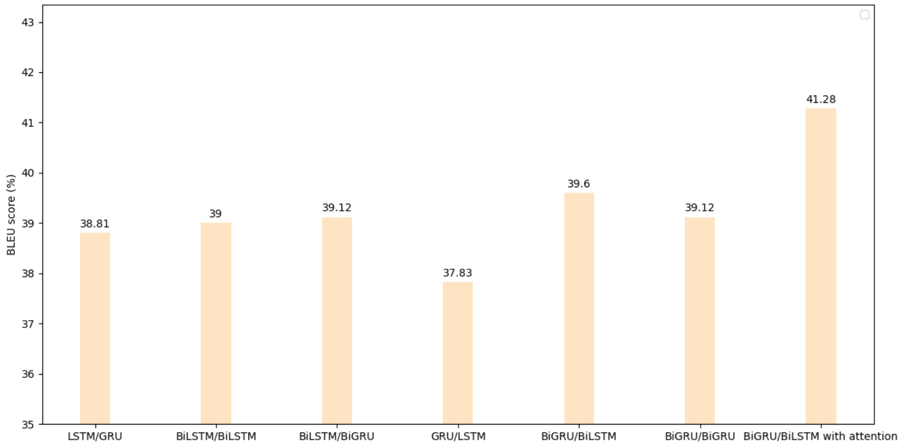

Figure 5. Performance evaluation of DL encoder-decoder models using Word2Vec embeddings with preprocessing

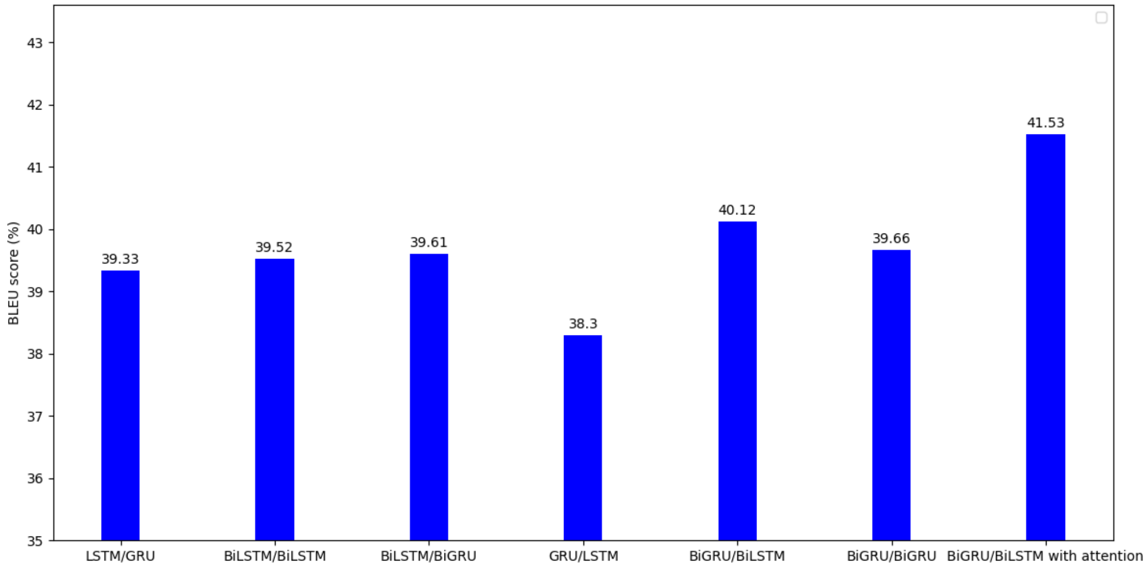

Figure 6. Performance evaluation of DL encoder-decoder models using GloVe embeddings with preprocessingg

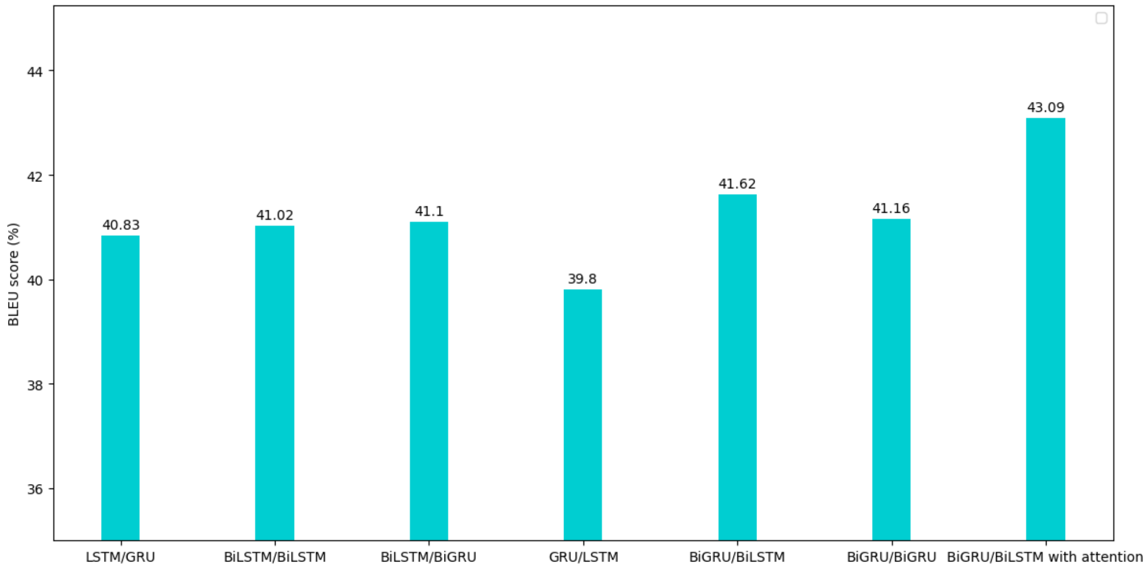

Figure 7. Performance evaluation of DL encoder-decoder models using FastText embeddings without preprocessing

The best model (BiGRU as an encoder and BiLSTM as a decoder with an attention mechanism and FastText embeddings) got in this case a BLEU score of 42.18% compared to 43.09% achieved without preprocessing. The observed results can be attributed to the effectiveness of Arabic preprocessing in addressing data sparsity and managing tokens that may not be present in the training corpus. Considering this analysis, the optimal combination for achieving desirable outcomes would involve utilizing BiGRU as the encoder, BiLSTM as the decoder, employing the attention mechanism, and incorporating Arabic preprocessing techniques.

In Table 1, the baseline scores for, the best model (BiGRU as an encoder, BiLSTM as a decoder, the attention mechanism and FastText embeddings), the original transformer model and our transformer implementation with the same hyper-parameters are presented. Our implementation achieves significantly higher BLEU points than the best model. The middle section of Table 1 presents the findings for different initialization schemes using AraBERT and AraGPT-2 pre-trained checkpoints. For AraBERT, we choose the AraBERTv0.1-base checkpoint for initializing the encoder or the decoder, or both. First, we note that is more beneficial to initialize the model, on the encoder side, with the AraBERT checkpoint. In addition, models initialized with the AraBERT checkpoint (AraBERT2RND, RND2AraBERT, AraBERT2AraBERT, and AraBERTSHARE) receive a significant boost.

Table 1. BLEU scores on a subset of the WIT corpus

	AR $\longrightarrow$ EN	EN $\longrightarrow$ AR
BiGRU/BiLSTM/attention/FastText	43.09	45.94
transformer	46.4	49.3
RND2RND	44.2	47.4
AraBERT2RND	48.3	30.5
RND2AraBERT	45.1	48.2
AraBERT2AraBERT	48.4	50.8
AraBERTSHARE	47.6	50.5
RND2AraGPT	37.4	41.6
AraBERT2AraGPT	41.1	49.7

For AraGPT, to initialize the decoder, we adopt the AraGPT2-base checkpoint. The AraGPT-based models (RND2AraGPT and AraBERT2AraGPT) are not as efficient, mainly when using AraGPT as a decoder and the target language is English. The reason behind this is the fact that the AraGPT model has been pre-trained primarily on Arabic text.

5. CONCLUSION

MT is a complex task, and different languages may require different approaches to achieve the best results. Arabic is a Semitic language with a complex structure that differs from that of European languages. Therefore, the same MT approach may not work as well for Arabic as for European languages. Recently, neural network-based MT has emerged as an alternative approach to traditional SMT. In this study, we compare the performance of seven DL models based on LSTM, GRU, BiLSTM, and BiGRU as simple encoders/decoders with attention mechanisms and different word embeddings, including Word2Vec, GloVe, and FastText. We also investigate the effect of Arabic text preprocessing on the MT models' performance. We explored different transformer encoder-decoder models and initialized them in different ways, including random initialization and warm-starting with public checkpoints of AraBERT and AraGPT-2. Our findings suggest that pre-trained encoder checkpoints are crucial for Arabic MT as they enable shared weights between the encoder and decoder, which minimizes the memory footprint. Our model is initialized using a combination of these checkpoints, and we explore various settings to find the optimal initialization method. We also found that the combination of AraBERT and AraGPT-2 in a single model does not improve efficiency compared to a randomly initialized base model. However, we noted that is more beneficial to initialize the model, on the encoder side, with the AraBERT checkpoint. Our findings provide insights into the selection and use of pre-trained checkpoints in neural network-based MT models, which can facilitate the development of more accurate and efficient MT systems for Arabic. As part of future work, we believe that there is still a lot of potential in combining different pre-trained models for MT, and we plan to investigate the impact of BERT and GPT checkpoints for multilingual NMT. Additionally, we aim to evaluate different language-specific BERT model checkpoints and assess the performance of the transformer when using the multilingual version. These investigations will help us to better understand the strengths and limitations of different MT models and inform the development of more effective and efficient MT systems.

ACKNOWLEDGEMENTS

We acknowledge the financial support for this research from the Centre National pour la Recherche Scientifique et Technique (CNRST) Morocco and Khawarizmi Project.

REFERENCES

[1] N. Alsohybe, N. Dahan, and F. B. -Alwi, "Machine-translation history and evolution: survey for arabic-english translations," *Current Journal of Applied Science and Technology*, vol. 23, no. 4, pp. 1–19, 2017, doi: 10.9734/cjast/2017/36124.

[2] A. Al-Janabi, E. A. Al-Zubaidi, and B. M. Merzah, "Detecting translation borrowings in huge text collections using various methods," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 3, pp. 1609–1616, 2023, doi: 10.11591/ijeecs.v30.i3.pp1609-1616.

[3] R. Chingamtotattil and R. Gopikakumari, "Neural machine translation for Sanskrit to Malayalam using morphology and evolutionary word sense disambiguation," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 3, pp. 1709–1719, 2022, doi: 10.11591/ijeecs.v28.i3.pp1709-1719.

[4] M. K. Nyein and K. M. Soe, "Source side pre-ordering using recurrent neural networks for English-Myanmar machine translation," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 5, pp. 4513–4521, 2021, doi: 10.11591/ijece.v11i5.pp4513-4521.

[5] P. Wijonarko and A. Zahra, "Spoken language identification on 4 Indonesian local languages using deep learning," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 6, pp. 3288–3293, 2022, doi: 10.11591/eei.v11i6.4166.

[6] T. M. Angona et al., "Automated bangla sign language translation system for alphabets by means of MobileNet," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 3, pp. 1292–1301, 2020, doi: 10.12928/TELKOMNIKA.V18I3.15311.

[7] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 2016, pp. 1715–1725, doi: 10.18653/v1/P16-1162.

[8] D. Ataman, M. Negri, M. Turchi, and M. Federico, "Linguistically motivated vocabulary reduction for neural machine translation from Turkish to English," *The Prague Bulletin of Mathematical Linguistics*, vol. 108, no. 1, pp. 331–342, 2017, doi: 10.1515/pralin-2017-0031.

[9] A. Tamchyna, M. W. -D. Marco, and A. Fraser, "Modeling target-side inflection in neural machine translation," in *Proceedings of the Second Conference on Machine Translation*, 2017, pp. 32–42, doi: 10.18653/v1/W17-4704.

[10] L. S. Hadla, T. M. Hailat, and M. N. Al-Kabi, "Evaluating Arabic to English machine translation," *International Journal of Advanced Computer Science and Applications*, vol. 5, no. 11, pp. 68–73, 2014.

[11] M. Alkhatib and K. Shaalan, "The key challenges for Arabic machine translation," *Studies in Computational Intelligence*, vol. 740, pp. 139–156, 2018, doi: 10.1007/978-3-319-67056-0_8.

[12] K. Cho et al., "Learning Phrase Representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724–1734.

[13] A. Vasvani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, 2017, pp. 5998–6008.

[14] G. Manias, A. Mavrogiorgou, A. Kiourtis, and D. Kyriazis, "An evaluation of neural machine translation and pre-trained word embeddings in multilingual neural sentiment analysis," in *2020 IEEE International Conference on Progress in Informatics and Computing (PIC)*, 2020, pp. 274–283, doi: 10.1109/PIC50277.2020.9350849.

[15] B. Klimova, M. Pikhart, A. D. Benites, C. Lehr, and C. S. -Stockhammer, "Neural machine translation in foreign language teaching and learning: a systematic review," *Education and Information Technologies*, vol. 28, no. 1, pp. 663–682, 2023, doi: 10.1007/s10639-022-11194-2.

[16] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, vol. 1, pp. 4171–4186.

[17] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," *OpenAI*, pp. 1–12, 2018.

[18] H. Sajjad, F. Dalvi, N. Durrani, A. Abdelali, Y. Belinkov, and S. Vogel, "Challenging language-dependent segmentation for Arabic: an application to machine translation and part-of-speech tagging," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 2017, vol. 2, pp. 601–607, doi: 10.18653/v1/P17-2095.

[19] M. Oudah, A. Almahairi, and N. Habash, "The impact of preprocessing on Arabic-English statistical and neural machine translation," in *Proceedings of Machine Translation Summit XVII Volume 1: Research Track*, 2019, pp. 214–221.

[20] M. S. H. Ameur, A. Guessoum, and F. Meziane, "Improving Arabic neural machine translation via n-best list re-ranking," *Machine Translation*, vol. 33, no. 4, pp. 279–314, 2019, doi: 10.1007/s10590-019-09237-6.

[21] A. Alrajeh, "A recipe for Arabic-English neural machine translation," *Arxiv-Computer Science*, vol. 1, pp. 1–5, 2018.

[22] S. Ding, A. Renduchintala, and K. Duh, "A call for prudent choice of subword merge operations in neural machine translation," in *Proceedings of Machine Translation Summit XVII: Research Track*, 2019, pp. 204–213.

[23] D. Ataman, W. Aziz, and A. Birch, "A latent morphology model for open-vocabulary neural machine translation," in *8th International Conference on Learning Representations*, ICLR 2020, 2020, pp. 1–15.

[24] D. Ataman, O. Firat, M. A. D. Gangi, M. Federico, and A. Birch, "On the importance of word boundaries in character-level neural machine translation," in *Proceedings of the 3rd Workshop on Neural Generation and Translation*, 2019, pp. 187–193, doi: 10.18653/v1/D19-5619.

[25] X. Liu, D. F. Wong, Y. Liu, L. S. Chao, T. Xiao, and J. Zhu, "Shared-private bilingual word embeddings for neural machine translation," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 3613–3622, doi: 10.18653/v1/P19-1352.

[26] A. Abdelali, K. Darwish, N. Durrani, and H. Mubarak, "A fast and furious segmenter for Arabic," in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Demonstrations*, 2016, pp. 11–16.

[27] W. Antoun, F. Baly, and H. Hajj, "ARAGPT2: pre-trained transformer for Arabic language generation," in *WANLP 2021 - 6th Arabic Natural Language Processing Workshop, Proceedings of the Workshop*, 2021, pp. 196–207.

[28] M. Cettolo, C. Girardi, and M. Federico, "WIT3: web inventory of transcribed and translated talks," in *Proceedings of the 16th Annual Conference of the European Association for Machine Translation, EAMT 2012*, 2012, pp. 261–268.

[29] M. Cettolo, J. Niehues, S. Stüker, L. Bentivogli, R. Cattoni, and M. Federico, "The IWSLT 2016 evaluation campaign," in *Proceedings of the 13th International Conference on Spoken Language Translation*, 2016, pp. 1–14.

BIOGRAPHIES OF AUTHORS

Nouhaila Bensalah received her M.Sc. degree in 2018 in Informatics and Telecommunications from the Department of Physics, Faculty of Sciences, Mohammed V University, Rabat, Morocco. Currently, she is actively pursuing a Ph.D. degree at the esteemed LIM Laboratory of Informatics, Faculty of Sciences and Techniques Mohammedia. With a strong academic background and a keen interest in cutting-edge technologies, Nouhaila is actively engaged in research activities. She has made significant contributions to her field through her participation in international and national conferences, where she has presented her work in the form of eight publications. Her research interests revolve around machine learning and natural language processing (NLP), with a particular focus on Arabic machine translation. She is deeply committed to advancing the understanding and application of these fields, aiming to contribute to the development of improved techniques and methodologies in the domain of Arabic machine translation. She can be contacted at email: nouhaila.bensalah@etu.fstm.ac.m.

Habib Ayad is a highly accomplished individual in the field of Computer Science. He completed his Ph.D. in Computer Science from the esteemed National School of Applied Sciences of Marrakech, Cadi Ayyad University in 2013. Currently, he serves as an associate professor at Hassan II University in Casablanca, where he actively contributes to the academic community. His research primarily revolves around cutting-edge areas such as data science, AI, ML, DL, and NLP. Through his extensive studies, he explores the frontiers of these fields, seeking innovative solutions and advancements that drive technological progress. In addition to his academic role, Dr. Ayad is a valued member of the ACM Casablanca Chapter. His active involvement in this professional community demonstrates his commitment to knowledge sharing and fostering collaboration among peers. He can be contacted at email: ayad.habib@gmail.com.

Abdellah Adib received the Doctorat de 3rd Cycle and the Doctorat d'Etat-es-Sciences degrees in Statistical Signal Processing from the Mohammed V University, Rabat, Morocco, in 1996 and 2004, respectively. Since 1997, he has been an assistant professor at the Scientific Institute of Rabat and a professor of higher education at the Faculty of Science and Technology of Mohammedia since 2008. He was head of the Department between 2012 and 2015. He was a member of the Scientific Committee of FSTM for two terms, 2015-2018 and 2018-2021. He was also a member of the CNRST scientific committees as well as an expert evaluator for information technologies for two consecutive terms, 2013-2016 and 2016-2020. Since 1993, his research has focused on automatic information processing, source separation, and applications (seismic, biomedical, and speech signals). He is also the author or co-author of more than 30 papers in international journals and more than 80 papers in international conferences. He has been a member of several technical committees of IEEE, EURASIP, and Springer. He has supervised more than 20 theses in different fields related to his favorite areas. He can be contacted at email: abdellah.adib@fstm.ac.ma.

Abdelhamid Ibn El Farouk is a full professor at Hassan II University in Casablanca. Currently, he serves as the dean of the Faculty of Letters and Humanities in Mohammedia. He is a linguist, sociolinguist, and translator, holding a doctorate thesis from the University of René Descartes in 1994 on the verbal system of Written Arabic and a Doctorat d'Etat in 1996 in the functional grammar of written Arabic from the Hassan II of Casablanca. He has several publications in the fields of his specialization, linguistics, sociolinguistics, and translation. Since March 2018, he has held the position of dean of the Faculty of Letters and Human Sciences in Mohammedia. Additionally, he serves as the Director of the Research Center for Translation in Humanities since December 2020, where he oversees research projects related to translation in the field of human sciences. He can be contacted at email: farouklettres@gmail.com.