Mesopotamian journal of Big Data
Vol. **(2024)**, 2024, **pp**. 241–250
DOI: https://doi.org/10.58496/MJBD/2024/018 ISSN: 2958-6453
https://mesopotamian.press/journals/index.php/BigData

Research Article

Deep Learning-based English-Arabic Machine Translation for Sulfur Manufacture Texts

Diadeen Ali Hameed [1,*], Belal Al-Khateeb [2]

[1] *Department of Electrical Engineering, Shirqat Engineering College, Tikrit University, Tikrit, Iraq.*
[2] *College of Computer Science and Information Technology, University of Anbar, Ramadi, Anbar, Iraq.*

ARTICLE INFO

Article History
Received 21 Oct 2024
Accepted 22 Nov 2024
Published 14 Dec 2024

Keywords
Arabic machine translation
deep learning, neural MT
sulfur texts
transformers model

ABSTRACT

The field of machine translation (MT) has seen significant advancements with deep learning (DL) techniques for translating texts among different languages. Despite the wealth of studies, there exists a noticeable gap in significant research dedicated to its translate Sulfur manufacture texts, primarily hindered by resource scarcity and the intricate grammatical structures inherent to these texts. This paper explores the application of transformer-based Arabic MT for sulfur manufacture texts, including its attention mechanisms and encoder-decoder framework, focusing on the new model ability to handle the linguistic and syntactic complexities inherent in these languages, such as morphological richness and context, and how the transformer's self-attention mechanism addresses these issues. It discusses the specific challenges of our proposed translation model, the obtained results indicate that this model is effective and has an accuracy of 90.7% in comparison with Mishraq application, which has 84.9% for the same test samples.

1. INTRODUCTION

MT has become an important component of internet communication [1], facilitating the interaction among different culture backgrounds across borders. In the last ten years, DL techniques for MT [2] has achieved great success in automatically translating English-Arabic language text by transformer models, outperforming seq2-seq and encoder-decoder MT [3]. The swiftly evolving domain of artificial intelligence, MT have marked significant milestones [4], owing much of their progress to groundbreaking models like the transformer. Introduced in the seminal paper "Attention Is All You Need" [5], the transformer model has reshaped our approach to processing language computationally, moving beyond the constraints that once limited the scope of DL applications in linguistics [6]. Although the quality of the translations generated, the systems are not perfect yet, needs to increase the dataset used, and training method [7][8].

Transformer models, has made a significant advancement in the MT field, as these models demonstrate state of-the-art performance across various benchmark datasets [9]. Now, we can obtain high-quality translations for a specific task or domain by transformer models with a training data set, by fine-tuning these pretrained models [10], there are still challenges in achieving accurate and fluent translations for Arabic language in different domains [11]. The transformer-based MT requires very large corpus sizes to train and evaluate the results. This paper, explores the performance of different transformer-based English-Arabic MT model [12]. We aim to compare the effectiveness of these models that can assess state-of-the-art results in a large set of translation tasks.

2. RELATED WORK

Before the advent of DL, MT was dominated by rule-based and statistical methods. Rule-based MT, relied heavily on linguistic rules and bilingual dictionaries [13][14]. Statistical MT (SMT), SMT models used probabilistic methods to align and translate text based on large corpora of parallel texts [15]. While SMT provided improvements over rule-based approaches, it still faced limitations in handling long-range dependencies and context [16], often producing translations that lacked coherence [17][18]. The introduction of DL-based MT marked a transformative shift in MT technologies [19]. Early

*Corresponding author. Email: it@gmail.com

NMT models, sequence-to-sequence (seq2seq) architectures with recurrent neural networks (RNNs), and long short-term memory networks (LSTMs), offered better handling of context and semantic information [20][21]. Key developments in this era include:

Introducing the attention mechanism [22], which significantly improves the performance of seq2seq models by allowing the model to focus on concentrate on the most relevant parts of the input sequence dynamically [23]. Extending attention mechanisms to different forms, including global and local attention, enhancing translation accuracy and alignment. Despite these advancements, seq2seq models with RNNs and LSTMs still faced challenges with long-range dependencies and computational inefficiencies [24]. Transformer models, introduced by [25], revolutionized DL-based MT with its novel architecture. Unlike RNN-based models, transformers leverage self-attention mechanisms to process sequences in parallel [26]. Transformers have set new benchmarks in MT, significantly outperforming traditional models in terms of translation quality and efficiency [27]. Notable transformer-based models include:

- BERT ("Bidirectional Encoder Representations from Transformers"), pretrained on large corpora and fine-tuned for various tasks, demonstrating substantial improvements in language understanding [28].
- GPT ("Generative Pre-trained Transformer"), fastens on cohesive generation and contextually relevant text [29][30].
- T5 ("Text-to-Text Transfer Transformer"), treats all language tasks as text-to-text problems, enhancing model versatility [31][32].

The English-Arabic language pair poses unique challenges due to significant linguistic differences, including syntax [33], complex morphology, vocabulary, and bidirectional context [34][35]. Transformer models began addressing some of these issues, with adaptations for Arabic's morphological richness. With improvements in translation quality for this language pair [36].

3. TRANSFORMER ARCHITECTURE

The transformer architecture [5], represents a fundamental shift in MT, moving away from recurrent and convolutional models to a fully attention-based approach [37]. The transformer model consists of an encoder-decoder framework. Each component is built from six layers that include multi-head self-attention mechanisms and feed-forward neural networks, and position-wise sublayers [38]. The model processes input sequences in parallel, which enhances computational efficiency and allows it to handle long-range dependencies effectively [39]. Fig. 1 shows the configuration of the transformer model with appropriate hyperparameters includes the number of layers, attention heads, and hidden dimensions. Typical settings might include (6) encoder layers and (6) decoder layers, with (8) attention heads and (512) hidden dimensions [40].

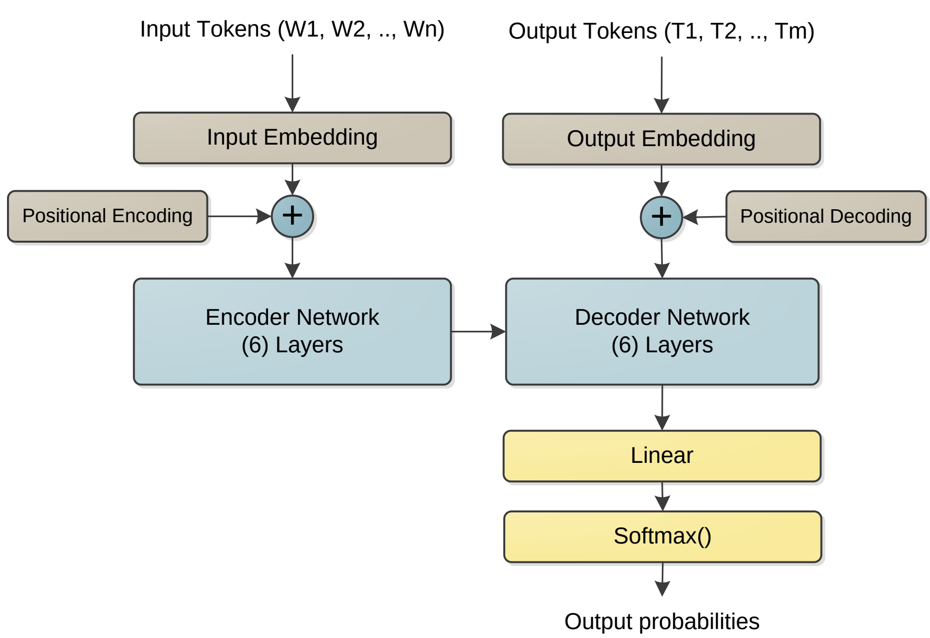

Fig. 1. The transformer model architecture [5][41].

The transformer consists of two main steps; those are Encoder (as shown in Fig. 2) and decoder (as shown in figure 3).

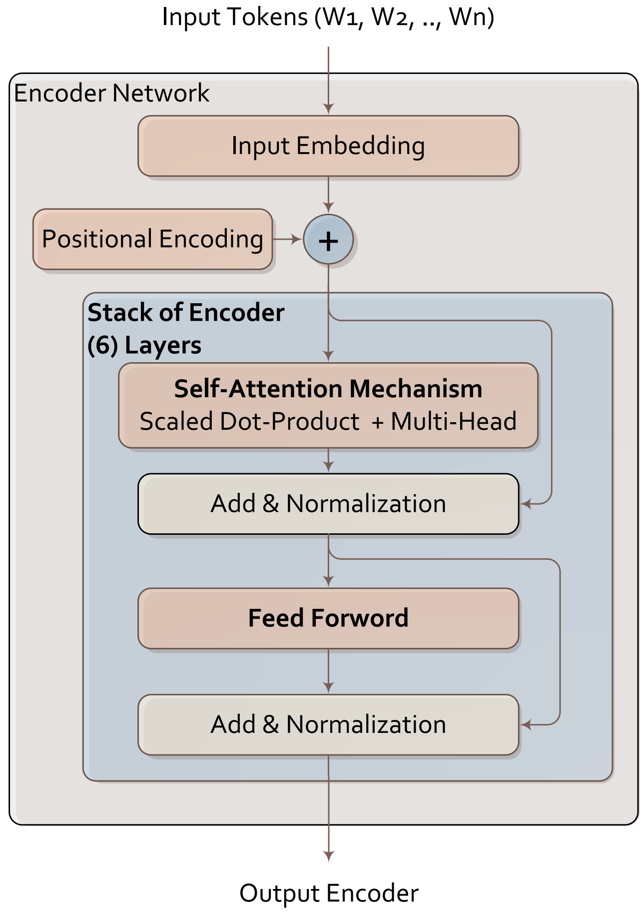

Fig. 2. The Encoder network architecture [6].

The encoder begins by converting input embedding tokens, words or sub words into vectors using embedding layers; each word is embedded into a vector of size 512 (fixed-sized). This allows them to understand the position of each word within the sentence. To do so, a combination of various sine and cosine functions are employed to create positional vectors that enable the use of this positional encoder for sentences of any length [41], as given in Eq. (1).

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right);\ PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{1}$$

The encoder consists of a stack of identical layers, each comprising:

- Self-Attention mechanism; it computes attention scores to determine the relevance of different words in the input sequence with respect to each other. The self-attention mechanism allows the model to capture contextual relationships within the sequence. For each position in the input, it generates a weighted sum of all positions [32], considering the importance of each word in context using Eq. 2. Then, followed by add and normalization, for each sub-layer in self-attention.

$$Attention\ (Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

- Feed-forward neural network (FFNN); it applies a position-wise feed-forward network to each position is processed individually and uniformly. The network comprises two linear transformations with a ReLU activation function between them, enabling the model to capture intricate patterns in the data [42]. Then, followed by Add & Normalization; for each sub-layer in FFNN.

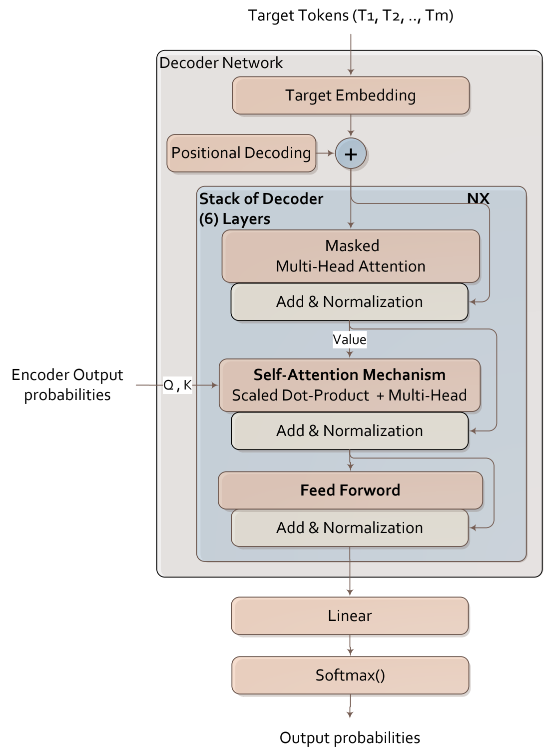

Fig. 3. Decoder network architecture [43]

The decoder is generating the target output sequence from the encoder's representation. It consists of a stack of identical layers, each comprising:

- Masked Self-Attention Mechanism; this mechanism is similar to the self-attention used in the encoder but includes masking to prevent attending to future tokens in the sequence. This ensures that predictions for a given position, maintaining the autoregressive property of the model [44]. Masked output is calculated using eq. 3. Then, followed by add and normalization, for each sub-layer.

$$Masked\ Output = \frac{QK^T - \infty .\ M}{\sqrt{d_k}} .V \tag{3}$$

- Encoder-Decoder Cross Attention [45]; this layer performs attention over the encoder's output, allowing the decoder to concentrate on related parts of the input sequence while generating each target token in the output sequence. It combines information from the encoder and the previously generated tokens. Then, followed by add and normalization, for each sub-layer. Attention weight [5]is calculated using eq. 4. The context vector represents the aggregated information from the encoder that is relevant to the current decoder position.

$$Attention\ Weights_{cross} = Softmax(\frac{Q_{dec}K_{enc}^T}{\sqrt{d_k}}) \tag{4}$$

- FFNN similar to the encoder, this applies a position-wise feed-forward network to each position in the decoder [46] using eq. 5. Then, followed by add and normalization, for each sub-layer using eq. 6.

$$FFNN(x) = ReUL(xW_1 + b_1)W_2 + b_2 \tag{5}$$

$$Loss = -\sum_{i=1}^{size}(y_i . \log \hat{y}_i) \tag{6}$$

4. METHOD

This section gives an overview of the methods and algorithms for Sulfur Manufacture Texts domain adaptation using transformer models. First, information about the corpus training dataset, vocabulary dictionary, then our approach is presented (transformer model and training model), ales presented concept word dictionary dataset. Lastly, the metric we used for evaluation are described, as shown in Figure 4.

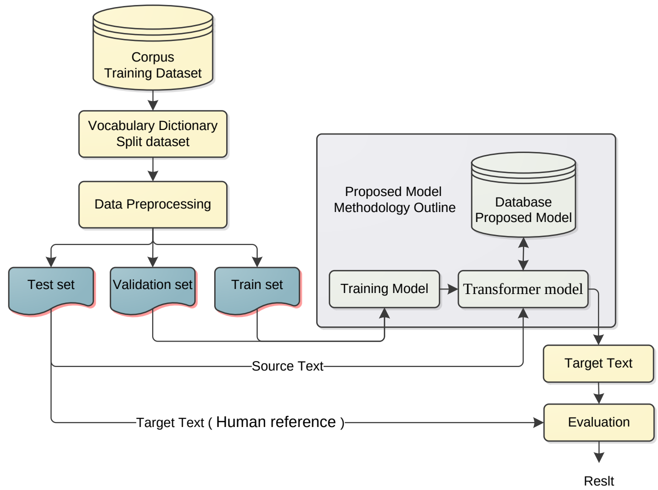

Fig. 4. Proposed methodology [41]

The proposed model consists of the corpus training dataset, data collection and preparation and utilize large-scale parallel corpora for English-Arabic, from the special corpus of sulfur manufacturing datasets, include more than (120,000) text. These corpora provide different and aligned sentences in both English and Arabic languages. Text preprocessing is crucial for optimizing the source text before feeding it into a MT model [47] [48]. It ensures that the text is clean, consistent, and suitable for the model, thereby improving training efficiency and model robustness [49]. The objective of text cleaning is to prepare the raw text by removing unwanted elements and correcting errors, remove noise involves eliminating extraneous characters and formatting issues. Let (S) be the input text, which includes English characters, numerical, symbols and tags. The cleaning text function Remove(S), can be represented as given in Eq. 7. [47]:

$$Remove(S) = \{s_i \in S \,|s_i\ is\ Unwanted\ symbols\ and\ tags\} \tag{7}$$

While filtering involves correcting typos, grammatical errors, and inconsistencies. It also ensures that the text is in the desired language. Let Lang(S) be the detected language of text S. Define a target language as Lang target. The filtering process can be described as given in Eq. 8. [48]:

$$filter(S) = \{s_i \in S\ Lang(s_i) =\ Lang\ target\} \tag{8}$$

The objective of normalization is to convert the text into a uniform format to reduce variability and noise; therefore, all characters are converted to lowercase to ensure consistency. Let X be the original text and ϕ be the normalization function. It process can be expressed as given in Eq. 9. [47]:

$$\phi: X \rightarrow\ \phi\ (X) \tag{9}$$

Where $\phi(X)$ is the normalized text. This might include lowercasing; we can be represented as a function lowercase that maps each character c to its lowercase form. For a string S =$\{s1, s2, \ldots, sn\}$, lowercasing can be represented as given in Eq. 10. [48]:

$$Lowercase(S) = \{lowercase(s_1), lowercase(s_2), \ldots, lowercase(s_n)\} \tag{10}$$

Removing punctuation and expanding contractions. Eliminate punctuation marks from the text. Let Punc be the set of punctuation characters. The function for removing punctuation as as given in Eq. 11. [49]:

$$Remove(S) = \{s_i \in\ S\ \mid\ s_i \notin\ Punc\} \tag{11}$$

Tokenization is the process of determining the longer processing units consisting of one or more words. This task involves identifying sentence boundaries, done through break down the text into sentences, or phrases that the MT system can process, as given in Eq. 12. [19]:

$$f(text) =\ \{s_1, s_2, \ldots, s_n\} \tag{12}$$

Tokenization transforms text into a sequence of tokens Text={s_1,s_1,…,s_m}, where m ≤ n. Then, dividing each sentence into words that the MT system can process, as given in Eq. 13. [19][49]:

$$f(sentence) = \{w_1, w_2, \dots, w_m\} \tag{13}$$

Tokenization sentences into a sequence of tokens, Sentence= {w1, w2,…,wk}, where k ≤ m as shown in Figure 3.3. These prepossessing steps help ensure that the text is in the best possible shape for MT. We split dataset into 70% train set, 10% validation set, and 30% test set. Leading to more accurate and contextually appropriate translations [50]. This helps in improving the model performance. This comprehensive preprocessing strategy contributes to the model's adaptability and robustness, crucial qualities for effective MT models. The training model by a large parallel corpus get from training set and validation set with techniques such as teacher forcing. Optimize using the Adam optimizer with learning rate schedules, such as the learning rate warm-up and decay strategies [51]. Then apply techniques such as dropout and label smoothing to prevent overfitting and improve generalization. The training objective is to minimize the difference between the predicted sequence and the true sequence. This is typically done using the cross-entropy loss for each token in the target sequence [6]. Mathematically, for a given token position using eq. 14.

$$Total\ Loss = \sum_{t=1}^{T} Log(y_t \mid x, y_{<t}) \tag{14}$$

The translation model, we fed the train set, validation set to our proposed MT using transformer's model. Then training model's output is saved in a dictionary of words, which has English-Arabic words, or phrase meanings. It uses the dictionary to find the equivalent meaning of words during test step. Its innovative self-attention mechanism and parallel processing capabilities. The transformer model has indelibly transformed the landscape of MT [46]. Its development addressed the limitations of previous models and opened up new avenues for exploration and advancement in the field. The success of the transformer has inspired a plethora of subsequent models [52], such as BERT, GPT, and T5, each building on its foundational principles to enhance further our ability to process MT [53]. The transformer is a pivotal achievement in the ongoing journey of DL research, signifying a milestone in our quest to decode the complexities of Arabic language. Then, we apply techniques such as dropout and label smoothing to prevent overfitting and improve generalization. The decoder's final output is passed through a linear layer and SoftMax to produce probabilities for each token in vocabulary using eq. 15.

$$P(y_t \mid x, y_{<t}) = Softmsx(Decoder\ OutputW_{Out} + b_{Out}) \tag{15}$$

Finally, after completion, we use target output evaluation

5. RESULTS AND DISCUSSIONS

This study utilizes a dataset comprising over 11,200 texts from Mishraq sulfur company catalogs and references. Table (1) presents results of the translations using adequacy. Table (2) presents results of the translations using fluency evaluation. We calculate average adequacy score via eq. 16. While we calculate average fluency score via eq. 17. Ales calculate average all of part rating score via eq. 18.

$$Adequacy\ Score\ (x) = \frac{1}{n}\sum_{i=1}^{n} s_i\ ; \quad x = \{Phrase \mid Short\ text \mid Long\ texts\} \tag{16}$$

$$Fluency\ Score(x) = \frac{1}{n}\sum_{i=1}^{n} LogP(w_i \mid w_1, w_2, \dots, w_{n-1}\} \tag{17}$$

$$\text{Average precision} = \frac{2(Phrase_{Score}) + 3(ShortText_{Score}) + 5(LongText_{Score})}{10} \tag{18}$$

Where P(wi∣w1,w2,…,wn−1) is the probability assigned by the language model to the ith word given its context.

TABLE I. AVERAGE PRECISION FOR ADEQUACY EVALUATION

Criteria translator	Phrase	Short text	Long texts	Average precision
New MT model	0.970	0.90	0.858	0.893
Mishraq translator	0.935	0.84	0.788	0.835

TABLE II. AVERAGE PRECISION FOR FLUENCY EVALUATION

Criteria translator	Phrase	Short text	Long texts	Average precision
New MT model	0.973	0.948	0.872	0.904
Mishraq translate	0.935	0.88	0.828	0.863

TABLE III. AVERAGE PRECISION TOTAL

Criteria translator	Phrase	Short text	Long texts	Average precision	Percentage method (100%)
New MT model	0.9715	0.934	0.865	0.907	90.7%
Mishraq translate	0.935	0.86	0.808	0.849	84.9%

To calculate average value for phrase, short text, and long text, see eq. 19. To calculate average percentage method for each type, see eq. 20.

$$Average\ precision\ (T) = \frac{Adequacy(x)+Adequacy(y)}{2} \tag{19}$$

$$Percentage\ method\ (T) = \frac{2(Phrase)+\ 3(ShortText)+\ 5(LongText)}{10} * 100\% \tag{20}$$

Where T is the total; x is the adequacy scores, x ={phrase, short text, long texts}; y is the fluency scores, y ={phrase, short text, long texts}.

The obtained results showed that if the system previously trained on all terms of a sentence, we get better performance. Experimental results of English to Arabic translation from testing dataset are shown in Fig. 5. The Mishraq application can be seen as like our model in most cases of translation phrase (less than 6 words), with a slight advantage to our model. Regarding the short sulfur (less than 15 words) manufacture phrases, it found that our translation model is much better than the Mishraq translate. This is primarily because of the training models of English-Arabic for our models bigger than training Mishraq translate model. Regarding the sulfur manufacture long texts, shown in figure 7. Mishraq application can be seen of inferior quality in comparison to this model in most cases. It is also noted that, while Mishraq translator cannot handle complex sentences, our new model can handle some of these sentences with accuracy. When evaluating sulfur manufacture texts, it is noted that the precision of fluency is much wider than adequacy. Finally, the obtained average results as illustrated in table (3) show that this model produces (90.7%), while Mishraq model application (84.9%).

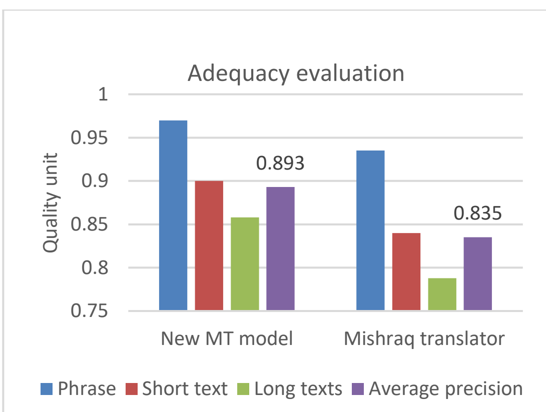

Fig. 5. Average precision for adequacy evaluation

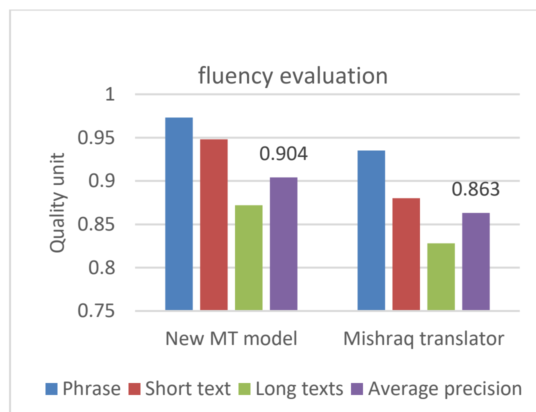

Fig. 6. Average precision for fluency evaluation

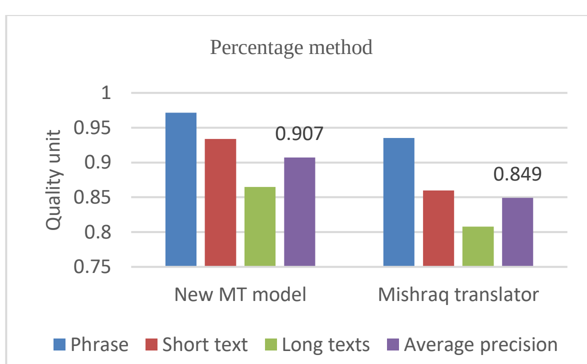

Fig 7. Summary of average precision

6. CONCLUSION

This study employs transformer model, in DL network, in translating sulfur manufacture texts from English into Arabic. The obtained results show that transformer model MT outperforms in comparison with other NN based models in processing sequential data. Its capability to capture long-term dependencies and utilize parallel computation has established its superiority in this domain. The achieved results indicate that the transformer model MT software accuracy is approximately 90.7%, while the Mishraq software is approximately 84.9% in the sulfur manufacture texts. Finally, from the obtained results of each test, it could be concluded that the proposed transformer model MT has a high accuracy and superiority over many translation application systems for sulfur manufacture filed. In the future, exploring various fine-tuning strategies, increasing the training dataset, and enhancing generalization capabilities could be employed.

Conflicts Of Interest

The author's disclosure statement confirms the absence of any conflicts of interest.

Funding

The author's paper does not provide any information on grants, sponsorships, or funding applications related to the research.

Acknowledgment

The authors expresses appreciation to the institution for their continuous support and access to relevant research materials.

References

[1] N. Alsohybe, N. Dahan, and F. Ba-Alwi, “Machine-Translation History and Evolution: Survey for Arabic-English Translations,” Curr. J. Appl. Sci. Technol., vol. 23, no. 4, pp. 1–19, 2017, doi: 10.9734/cjast/2017/36124.

[2] R. Torjmen and K. Haddar, “Translation from Tunisian Dialect to Modern Standard Arabic : Exploring Finite-State Transducers and Sequence-to-Sequence Transformer Approaches,” 2024, doi: 10.1145/3681788.

[3] X. Ho et al., “A Survey of Pre-trained Language Models for Processing Scientific Text,” 2024, [Online]. Available: http://arxiv.org/abs/2401.17824

[4] D. A. Hameed, T. A. Faisal, A. K. Abbas, H. A. Ali, and G. T. Hasan, “DIA-English-Arabic neural machine translation domain: sulfur industry,” Indones. J. Electr. Eng. Comput. Sci., vol. 27, no. 3, pp. 1619–1624, 2022, doi: 10.11591/ijeecs.v27.i3.pp1619-1624.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, “Attention Is All You Need”, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

[6] X. Han et al., “Pre-trained models: Past, present and future,” AI Open, vol. 2, no. August 2021, pp. 225–250, 2021, doi: 10.1016/j.aiopen.2021.08.002.

[7] J. Xu, X. Liu, Y. Wu, Y. Tong, Q. Li, and M. Ding, “ImageReward : Learning and Evaluating Human Preferences for Text-to-Image Generation,” no. NeurIPS, 2023.

[8] O. Abend, L. Choshen, L. Fox, and Z. Aizenbud, “On the Weaknesses of Reinforcement Learning for Neural Machine Translation”.

[9] F. Casacuberta, “Segment-Based Interactive Machine Translation for Pre-trained Models”, arXiv:2407.06990v1 [cs.CL] 9 Jul 2024

[10] P. M. Reader, “From Cloze to Comprehension : Retrofitting Pre-trained Masked Language Models to Vanilla Discriminative Fine-tuning Generative Fine-tuning Fine-tuning with PMR g n i n i a r t - e r P,” no. NeurIPS, 2023.

[11] Y. Ye and A. Toral, “Fine-grained Human Evaluation of Transformer and Recurrent Approaches to Neural Machine Translation for English-to-Chinese”, arXiv:2006.08297v1 [cs.CL] 15 Jun 2020.

[12] B. Zhang, D. Xiong, and J. Su, “Neural Machine Translation with Deep Attention,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 1, pp. 154–163, 2020, doi: 10.1109/TPAMI.2018.2876404.

[13] A. Alqudsi, N. Omar, and K. Shaker, “A Hybrid Rules and Statistical Method for Arabic to English Machine Translation,” 2nd Int. Conf. Comput. Appl. Inf. Secur. ICCAIS 2019, 2019, doi: 10.1109/CAIS.2019.8769545.

[14] K. S. Alubaidi, “Hybrid Arabic-English Machine Translation to Solve Reordering and Ambiguity Problems,” J. Univ. Hum. Dev., vol. 1, no. 4, p. 413, 2015, doi: 10.21928/juhd.v1n4y2015.pp413-416.

[15] A. Alqudsi, N. Omar, and K. Shaker, “Arabic machine translation: a survey,” Artif. Intell. Rev., vol. 42, no. 4, pp. 549–572, 2014, doi: 10.1007/s10462-012-9351-1.

[16] Z. Tan, J. Su, B. Wang, Y. Chen, and X. Shi, “Lattice-to-sequence attentional Neural Machine Translation models,” Neurocomputing, vol. 284, pp. 138–147, 2018, doi: 10.1016/j.neucom.2018.01.010.

[17] F. Stahlberg, “Neural machine translation: A review,” J. Artif. Intell. Res., vol. 69, pp. 343–418, 2020, doi: 10.1613/JAIR.1.12007.

[18] H. M. Lateef, A. M. Awaad, D. A. Hameed, G. T. Hasa, and T. A. Faisal, “Evaluation of domain sulfur industry for DIA translator using bilingual evaluation understudy method,” Bull. Electr. Eng. Informatics, vol. 13, no. 1, pp. 370–376, 2024, doi: 10.11591/eei.v13i1.4489.

[19] D. A. Hameed, T. A. Faisal, A. M. Alshaykha, G. T. Hasan, and H. A. Ali, “Automatic evaluating of Russian-Arabic machine translation quality using METEOR method,” AIP Conf. Proc., vol. 2386, no. January, 2022, doi: 10.1063/5.0067018.

[20] W. Ma, B. Yan, and L. Sun, “Generative Adversarial Network-Based Short Sequence Machine Translation from Chinese to English,” Sci. Program., vol. 2022, 2022, doi: 10.1155/2022/7700467.

[21] L. Jian, H. Xiang, and G. Le, “LSTM-Based Attentional Embedding for English Machine Translation,” Sci. Program., vol. 2022, 2022, doi: 10.1155/2022/3909726.

[22] D. Bahdanau, K. H. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–15, 2015.

[23] A. V. M. Barone, J. Helcl, R. Sennrich, B. Haddow, and A. Birch, “Deep architectures for neural machine translation,” WMT 2017 - 2nd Conf. Mach. Transl. Proc., pp. 99–107, 2017, doi: 10.18653/v1/w17-4710.

[24] V. Karyukin, D. Rakhimova, A. Karibayeva, A. Turganbayeva, and A. Turarbek, “The neural machine translation models for the low-resource Kazakh–English language pair,” PeerJ Comput. Sci., vol. 9, pp. 1–20, 2023, doi: 10.7717/peerj-cs.1224.

[25] M. Ott, S. Edunov, D. Grangier, and M. Auli, “Scaling Neural Machine Translation,” WMT 2018 - 3rd Conf. Mach. Transl. Proc. Conf., vol. 1, pp. 1–9, 2018, doi: 10.18653/v1/w18-6301.

[26] W. Xie, Y. Feng, S. Gu, and D. Yu, “Importance-based neuron allocation for multilingual neural machine translation,” ACL-IJCNLP 2021 - 59th Annu. Meet. Assoc. Comput. Linguist. 11th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf., pp. 5725–5737, 2021, doi: 10.18653/v1/2021.acl-long.445.

[27] J. Brown, “Enhancing Translation Accuracy with Transformer Models in Neural Machine Translation”, IJST−Computer Science , Vol 1, Issue, pp 1-7. 7, July2024.

[28] G. Miao, H. Di, J. Xu, Z. Yang, Y. Chen, and K. Ouchi, “Improved Quality Estimation of Machine Translation with Pre-trained Language Representation,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11838 LNAI, pp. 406–417, 2019, doi: 10.1007/978-3-030-32233-5_32.

[29] N. Moghe et al., “Machine Translation Meta Evaluation through Translation Accuracy Challenge Sets,” pp. 1–56, 2024, [Online]. Available: http://arxiv.org/abs/2401.16313

[30] P. Nayak, R. Haque, J. D. Kelleher, and A. Way, “Instance-Based Domain Adaptation for Improving Terminology Translation,” MT Summit 2023 - Proc. 19th Mach. Transl. Summit, vol. 1, pp. 222–234, 2023.

[31] J. Li, T. Tang, W. X. Zhao, and J. R. Wen, “Pretrained Language Models for Text Generation: A Survey,” IJCAI Int. Jt. Conf. Artif. Intell., vol. 1, no. 1, pp. 4492–4499, 2021, doi: 10.24963/ijcai.2021/612.

[32] M. Popel et al., “Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals,” Nat. Commun., vol. 11, no. 1, pp. 1–15, 2020, doi: 10.1038/s41467-020-18073.

[33] H. M. Elsherif and T. R. Soomro, “Perspectives of arabic machine translation,” J. Eng. Sci. Technol., vol. 12, no. 9, pp. 2315–2332, 2017.

[34] S. H. Ahmed , Tran., “An Analytical Study on Improving Target Tracking Techniques in Wireless Sensor Networks Using Deep Learning and Energy Efficiency Models”, BJN, vol. 2023, pp. 100–104, Nov. 2023, doi: 10.58496/BJN/2023/013.

[35] Y. K. Hussein, D. A. Hameed, L. I. Kalaf, B. Rahmatullah, and A. T. Al-Taani, “Automatic Evaluating Russian-Arabic Machine Translation Quality Using BLEU Method,” Rev. AUS 25, pp. 155–162, 2019, doi: 10.4206/aus.2019.n25-9/.

[36] H. Atwany, N. Rabih, I. Mohammed, A. Waheed, and B. Raj, “OSACT 2024 Task 2: Arabic Dialect to MSA Translation,” 6th Work. Open-Source Arab. Corpora Process. Tools, OSACT 2024 with Shar. Tasks Arab. LLMs Hallucination Dialect to MSA Mach. Transl. Lr. 2024 - Work. Proc., pp. 98–103, 2024.

[37] J. Moorkens, “What to expect from Neural Machine Translation: a practical in-class translation evaluation exercise,” Interpret. Transl. Train., vol. 12, no. 4, pp. 375–387, 2018, doi: 10.1080/1750399X.2018.1501639.

[38] X. Liu, J. He, M. Liu, Z. Yin, L. Yin, and W. Zheng, “A Scenario-Generic Neural Machine Translation Data Augmentation Method,” Electron., vol. 12, no. 10, pp. 1–15, 2023, doi: 10.3390/electronics12102320.

[39] S. Hajbi, O. Amezian, N. El Moukhi, R. Korchiyne, and Y. Chihab, “Moroccan Arabizi-to-Arabic conversion using rule-based transliteration and weighted Levenshtein algorithm,” Sci. African, vol. 23, no. October 2023, p. e02073, 2024, doi: 10.1016/j.sciaf.2024.e02073.

[40] H. Lai and M. Nissim, “A Survey on Automatic Generation of Figurative Language: From Rule-based Systems to Large Language Models,” ACM Comput. Surv., vol. 56, no. 10, pp. 1–34, 2024, doi: 10.1145/3654795.

[41] S. Badawi, “UHD JOURNAL OF SCIENCE AND TECHNOLOGY A Transformer-based Neural Network Machine Translation Model for the Kurdish Sorani Dialect,” vol. 7, no. 1, pp. 15–21, 2023, doi: 10.21928/uhdjst.v7n1y2023.pp15-21.

[42] A. V Hujon, T. Doren, and K. Amitab, “ScienceDirect Transfer Learning Based Neural Machine Translation of Transfer Learning Based Machine Translation of on Neural Settings English-Khasi on Low-Resource Settings,” Procedia Comput. Sci., vol. 218, pp. 1–8, 2023, doi: 10.1016/j.procs.2022.12.396.

[43] T. Kavitha, G. Amirthayogam, J. J. Hephzipah, R. Suganthi, V. A. Kumar G, and T. Chelladurai , Trans., “Healthcare Analysis Based on Diabetes Prediction Using a Cuckoo-Based Deep Convolutional Long-Term Memory Algorithm”, Babylonian Journal of Artificial Intelligence, vol. 2024, pp. 64–72, Jun. 2024, doi: 10.58496/BJAI/2024/009.

[44] H. Al-Khalifa, K. Al-Khalefah, and H. Haroon, “Error Analysis of Pretrained Language Models (PLMs) in English-to-Arabic Machine Translation,” Human-Centric Intell. Syst., vol. 4, no. 2, pp. 206–219, 2024, doi: 10.1007/s44230-024-00061-7.

[45] T. Yang, S. Zhao, H. Chen, and B. Chen, "Machine translation system based on deep learning," J. Phys. Conf. Ser., vol. 2030, no. 1, 2021, doi: 10.1088/1742-6596/2030/1/012098.

[46] W. He, Y. Wu, and X. Li, "Attention Mechanism for Neural Machine Translation: A survey," IEEE Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2021, vol. 5, pp. 1485–1489, 2021, doi: 10.1109/ITNEC52019.2021.9586824.

[47] S. Dalmia et al., "LegoNN : Building Modular Encoder-Decoder Models," pp. 1–15.

[48] B. N. V. Narasimha Raju, M. S. V. S. Bhadri Raju, and K. V. V. Satyanarayana, "Effective preprocessing based neural machine translation for english to telugu cross-language information retrieval," IAES Int. J. Artif. Intell., vol. 10, no. 2, pp. 306–315, 2021, doi: 10.11591/ijai.v10.i2.pp306-315.

[49] Worood Esam Noori and A. S. Albahri , Trans., "Towards Trustworthy Myopia Detection: Integration Methodology of Deep Learning Approach, XAI Visualization, and User Interface System", Applied Data Science and Analysis, vol. 2023, pp. 1–15, Feb. 2023, doi: 10.58496/ADSA/2023/001.

[50] Prepro, "Dictionary-based Phrase-level Prompting of Large Language Models", arXiv:2302.07856v1 [cs.CL] 15 Feb 2023 no. Section 2, 2021.

[51] M. Kale, "mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer", arXiv:2010.11934v3 [cs.CL] 11 Mar 2021.

[52] A. Hendy, M. Abdelrehim, A. Sharaf, and V. Raunak, "How Good Are GPT Models at Machine Translation ?", arXiv:2302.09210v1 [cs.CL] 18 Feb 2023.

[53] M. Mars, "From Word Embeddings to Pre-Trained Language Models: A State-of-the-Art Walkthrough," Appl. Sci., vol. 12, no. 17, 2022, doi: 10.3390/app12178805.