CCC Publications

BioNMT: A Biomedical Neural Machine Translation System

H. Liu, Y. Liang, L. Wang, X. Feng, R. Guan

Hongtao Liu, Yanchun Liang

1. Zhuhai Sub Laboratory of Key Laboratory of Symbol Computation and Knowledge
Engineering of Ministry of Education, Zhuhai College of Jilin University
Zhuhai, 519041, China
2. Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of
Education, College of Computer Science and Technology,
Jilin University, Changchun, 130012, China
htliu17@mails.jlu.edu.cn
ycliang@jlu.edu.cn

Liupu Wang, Xiaoyue Feng

Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of
Education, College of Computer Science and Technology,
Jilin University, Changchun, 130012, China
wanglpu@jlu.edu.cn
fengxy@jlu.edu.cn

Renchu Guan*

1. Zhuhai Sub Laboratory of Key Laboratory of Symbol Computation and Knowledge
Engineering of Ministry of Education, Zhuhai College of Jilin University
Zhuhai, 519041, China
2. Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of
Education, College of Computer Science and Technology,
Jilin University, Changchun, 130012, China
*Corresponding author: guanrenchu@jlu.edu.cn

Abstract

To solve the problem of translation of professional vocabulary in the biomedical field and help
biological researchers to translate and understand foreign language documents, we proposed a
semantic disambiguation model and external dictionaries to build a novel translation model for
biomedical texts based on the transformer model. The proposed biomedical neural machine trans-
lation system (BioNMT) adopts the sequence-to-sequence translation framework, which is based
on deep neural networks. To construct the specialized vocabulary of biology and medicine, a hy-
brid corpus was obtained using a crawler system extracting from universal corpus and biomedical
corpus. The experimental results showed that BioNMT which composed by professional biological
dictionary and Transformer model increased the bilingual evaluation understudy (BLEU) value by
14.14%, and the perplexity was reduced by 40%. And compared with Google Translation System
and Baidu Translation System, BioNMT achieved better translations about paragraphs and resolve
the ambiguity of biomedical name entities to greatly improved.

Keywords: neural machine translation, Transformer, self-attention, semantic disambiguation

1   Introduction

Artificial intelligence has been used in fire monitoring[6], wireless sensor network data transmission routing protocol optimization [14], medical data classification [29] and other fields. There are many AI technologies that are constantly developing to help researchers better understand the knowledge in the field. For example, Guan *et al.*'s deep feature extraction based on text clustering can better help researchers get the deep knowledges of the text [12]. In 2014, Bengio *et al*. added the recurrent neural networks (RNN) based encoder-decoder model to statistical machine translation [9]. This is the first time that a neural network has been used in combination with machine translation. Graham first making"sequence-to-sequence" to machine translation marked a formal departure of neural networks from statistical machine translation and the independent training of machine translation models are widely used [23]. Although the work did not attract enough attention at the time, Google officially released Google neural machine translation (GNMT) [32] in 2016 and neural network-based machine translation has officially become the mainstream model for machine translation. The neural network-based machine translation encodes the entire source sentence and processes it into a sentence vector of a certain length. To complete the entire translation process, with certain related algorithms and constraints, the vector of the target language is generated, and then translated into target sentence by the decoding process of the neural network. Neural network-based machine translation not only reduces the engineering design requirement for translation, but achieves or exceeds the accuracy of phrase-based machine translation [23].

Translation systems of biomedical specialty can effectively help scientific researchers to understand the biomedical papers [5]. For example, the analysis of genetic research by meng *et al.* contains many biological terms that are not easily understood by researchers in the same field [18]. Compared with universal corpora, biomedical corpus is more specialized, requiring translation results to be more inclined to the translation habits of biomedical professionals. Therefore, biomedical corpus-based neural machine translation uses specialized corpora to train the translation model. Starting from Mitka's work on medical translation [22], medical translation researchers have been facing several challenges, for example, the untranslatable phenomenon in medical translation [19], the complexity of medical translation [11], the translation of abbreviations [15], and the translation of medical terms [26]. In 2017, Chen *et al.*[7] adopted statistical machine translation to carry out professional translation work in biomedicine and applied it to specific application. In the construction of biomedical neural machine translation, an external medical lexicon needs to be added to improve translation accuracy.

This paper mainly carried out two contributions:

1) We used our data crawler system to capture bilingual aligned corpora in the field of biomedicine, and trained a transformer framework to obtain a translation model with better professional translation results.

2) On the basis of the professional translation model, we proposed a semantic disambiguation model to the presence of the unknown words phenomenon (<unk>) [13]. <unk> are special words that cannot be translated when encountering a professional noun during the decoding process of the sequence-to-sequence model. We proposed a lexicon specifically for the biomedical vocabulary which called "MedDRA" and used semantic disambiguation model to solve <unk> problem.

The remainder of this paper is organized as follows. Section 2 introduces the related work. Section 3 describes our BioNMT system. Section 4 illustrates the experimental results and presents a discussion. Finally, Section 5 concludes our paper.

2   Relate Work

RNNs are suitable for processing time-related sequences[21]. It is the first neural machine translation mode with sequence-to-sequence translation method. Then long short-term memory networks (LSTM) [8] were proposed to solve the problems encountered by recurrent neural networks in long time series. Although the LSTM can solve the long time series problem [25], it still does not solve the problems such as the length of the fixed word vector used in encoding and decoding (see Figure 1). Bahdanau *et al.*[1] and Luong *et al.*[16] applied attention mechanisms to machine translation and

achieved significant improvements.
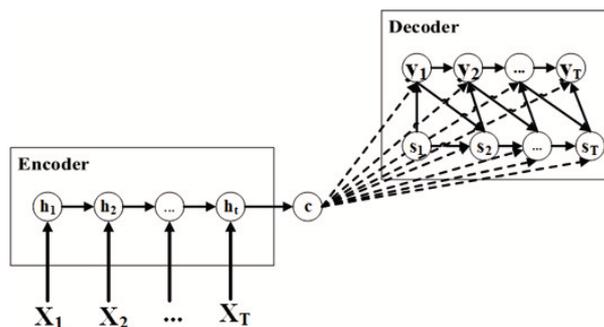
Figure 1: Encoder-Decoder model with Attention

Rocktaschel *et al.*[27] proposed the subdivision mechanism based on the attention mechanism and related visual analysis in 2015. Bahdanau *et al.*[1] improved and optimized the traditional neural network and added context semantics to the neural machine translation model. Machine translation is not only a simple translation of each word and phrase, but also refers to the meaning of the entire sentence. The supervised attention mechanism proposed by Mi *et al.*[20] improved and optimized the work based on the attention mechanism.

The combination of LSTM and attention mechanism solves the problem of fixed word vector length[1]. It solves the problem on the basis of preserving the context by retaining the intermediate output results in the input sequence in the long-term and short-term memory network. The emergence of the transformer framework has brought neural network-based machine translation results into a new era[10]. The transformer framework abandons traditional neural networks and uses a well-performing attention mechanism and feedforward network to perform the entire "encoding-decoding" process. It overcomes the bottleneck in which the recurrent neural network model cannot be calculated in parallel. In addition, the number of operations required to calculate the association between two positions will not increase as the distance between words in a sentence increases. The attention mechanism can produce a better interpretable model, and the multi-head attention mechanism can produce different results to balance and optimize the final result.

However, in the above algorithms and models, the <unk> problem caused by professional terms is not effectively solved. When using the model, there will be many professional words that do not appear during model training and need to be translated. The appearance of new words will cause problems in the translation process, and at the same time, errors will occur in the translation results. The translation of the entire sentence in the above algorithms and models focuses more on the general translation method, resulting in the translation result more like a general sentence. It would reduce and miss the professionalism and accuracy that should be included in professional translation.

3   BioNMT System

Based on transformer model, we propose a biomedical machine translation system to solve the <unk> phenomenon. As a result, a complete Chinese sentence without <unk> flags come out. And the results are more semantic smoother.

3.1   Training Model

As shown by the dotted line in Fig.1, the attention mechanism acts between the encoder and decoder. With the attention mechanism, the model can learn the alignment relationship and translation relationship between the original sentence and the target sentence at the same time. As shown in the combination of "RNN + Attention", in the encoding process, the original sentence is encoded into a set of feature vectors. When decoding, unlike the previous encoder-decoder structure of the RNN, which uses the feature vectors of the entire input sentence, the attention mechanism uses the weighted

sum of all feature vectors at time slice t; that is,

$$s_t = f(s_{(t-1)}, y_{(t-1)}, c_i) \tag{1}$$

where c_i is the weighted sum of all feature vectors h_t:

$$c_i = \sum_{t=1}^{T} \alpha_{it} h_t \tag{2}$$

$$\alpha_{it} = \frac{e^{e_{it}}}{\sum_{k=1}^{T} e^{e_{ik}}} \tag{3}$$

$$e_{it} = \tanh(s_{i-1}, h_t) \tag{4}$$

where e_{it} is the alignment model of the ith time slice of the output sequence, which indicates the correlation of the time and each time slice of the input data. Using the state s_{t-1} at the previous moment and the yth input data h_t, $\tanh(s_{i-1}, h_t)$ is calculated as the activation function.

The self-attention mechanism uses input sentences to three matrices: Q(query), K(Key), and V(value). Attention is calculated as shown in

$$self - Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{5}$$

To prevent QK^T's dot product value from being too large when the k dimension increases, we need to divide the result of the dot product by $\sqrt{d_k}$. After the $softmax$ calculation, it is multiplied by V to obtain the final attention matrix. The $softmax$ function maps multiple outputs to the interval of [0,1], as shown in

$$softmax(Z_j) = \frac{e^{Z_j}}{\sum_{k=1}^{K} e^{Z_k}} \tag{6}$$

In 2017, Google proposed the transformer framework[31]. The core part of the model consists of a self-attention mechanism and feedforward neural network (as shown in Figure 2). It uses multihead self-attention combined with a feedforward network to train the translation model in an encoder-decoder manner. The proposed transformer framework not only improves the performance of the encoder-decoder for text processing but also enables handling of long-term dependencies. Compared with the high computational complexity of the LSTM, the advantages of the fast calculation speed and high efficiency of the transformer framework make it more widely used.

The transformer not only used "$one\text{-}hot$"[17] to encode but also introduced positional encoding (PE) to solve the problem of position and order. Experiments have shown that the results of training and learning position encoding vectors and using formulas to perform position encoding vectors are similar on reducing the time complexity of training, as shown in

$$PE(pos, 2i) = \sin(\frac{pos}{10000^{\frac{2i}{d_{model}}}}) \tag{7}$$

and

$$PE(pos, 2i + 1) = cos(\frac{pos}{10000^{\frac{2i}{d_{model}}}}) \tag{8}$$

To calculate the position encoding vector, pos refers to the relative position of the word in the sentence, i indicates the embedding dimension, and d_{model} is the length of the pre-specified word vector. After word embedding and position embedding, each word is converted into three vectors, and each vector contains word features and location information.

In the initial stage of the transformer, the multi-head self-attention input is the feature vector obtained after embedding the sentence, and the input of each layer is the output of the previous layer, as shown in

$$MultiHead(Q, K, V) = Concat(head_1, \cdots, head_h)W \tag{9}$$
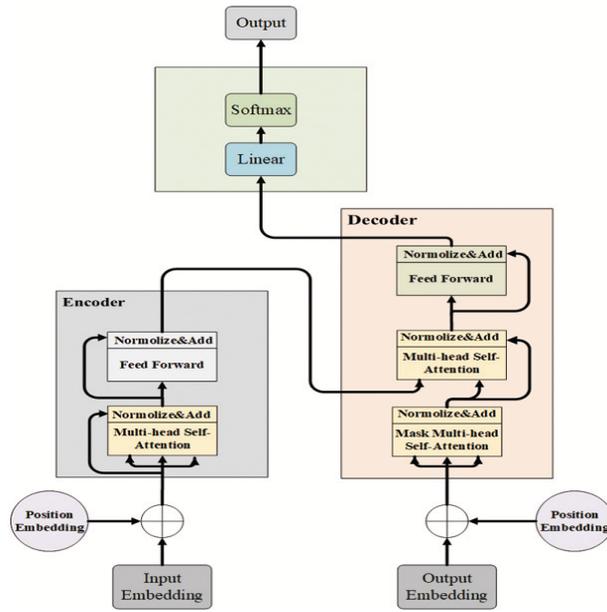
Figure 2: The Transformer framework

where

$$head_i = self - Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{10}$$

After performing the multihead self-attention calculation, layer normalization calculation and residual connection are performed. And the feature matrix obtained after the calculation is fed to the fully connected layer. For the calculation of the fully connected layer of the feature matrix y, the parameters between different rows of the same layer are the same, and the parameters of different fully connected layers are different, as shown in

$$FFN(x) = ReLU(0, xW_1 + b_1)W_2 + b_2 \tag{11}$$

and

$$ReLU(x) = \begin{cases} x, & if \quad x>0 \\ 0, & if \quad x\leq0 \end{cases} \tag{12}$$

where, ReLU is the activation function, W_1, W_2 are weight matrices, and b_1, b_2 are deviations.

The feature matrix obtained at the end of the encoder is input to the decoder for decoding. Each decoder contains three layers, which are mask multihead self-attention, encoder-decoder attention, and fully connected feedforward. In addition, in the decoder process, the K and V of the second layer of self-attention come from the encoder. To address the contextual correlation between words, the transformer uses mask multi-head self-attention (Mask-Attention) to process the contextual semantic correlation. Mask-Attention uses 0 and 1 for each dimension of the input feature matrix to choose whether to retain. It only obtains the input before the current time and perform the self-attention calculation of the input before the current time, and limit the contextual relevance. There can only be a correlation with the word before the word to be translated. The untranslated words behind do not consider the semantic relevance; that is, the information after $i + 1$ is not introduced when predicting the ith position. The specific expression is

$$Mask - Attention(Q, K, V) = Multihead(Relu(mask(Q, K)W, V)) \tag{13}$$

$$mask(x) = p * x, p \in \{0, 1\} \tag{14}$$

and Mask-Attention input is the previous output of the layer decoder.

3.2   Biomedical Dictionary and Semantic Disambiguation

<unk> (unknown words) are special words that cannot be translated when encountering a professional noun during the decoding process of the seq2seq model. Rico *et al* [28] and Issam *et al* [2]

are based on "Byte Pair Encoding" to solve <unk>. This approach solves <unk> to some extent, but in the specific field, <unk> still appears in large numbers, and there is no effection solution for the problem. Therefore, we add a lexicon specifically for vocabulary in the biomedical field and use semantic disambiguation model to solve <unk>.

To a certain extent, translation results always include <unk>. Using the attention mechanism, we obtain the position of the generated <unk> and find the source words that produced <unk>. After that, we manually collect professional terminology to help us identify preliminary <unk> solutions, which we call the "MedDRA" dictionary [4]. If only one Chinese translation of the word is contained in the "MedDRA" dictionary, we replace the Chinese translation into the results <unk>. If the source word contains multiple Chinese translations in the dictionary (in most cases, the biomedical vocabulary contains only one Chinese translation in the dictionary), we use the semantic similarity calculation in natural language processing (NLP) to obtain the most suitable Chinese translation result [30, 33]. If the <unk> source words do not appear in the "MedDRA" dictionary. Then, we add the source word to translate the source results in replacing <unk>. When all <unk> processes end, we refine the entire sentence to make it more fluent.
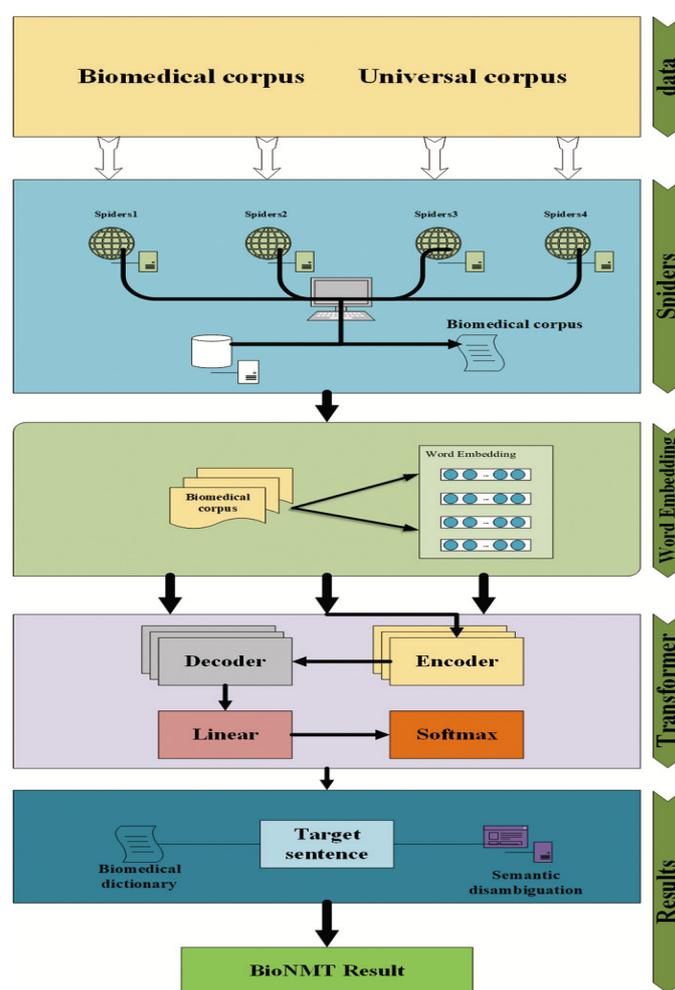
Figure 3: The BioNMT framework

3.3   Biomedical Neural Machine Translation System

The proposed biomedical neural machine translation (BioNMT) system uses its own data crawler system to collect biomedical corpora to train biomedical translation models. Based on the trained transformer framework, the translation model with the highest BLEU[24] score was obtained. Later, BioNMT added a variety of natural language processing technologies, such as semantic disambiguation, external dictionaries, regular expressions, and word similarity calculation, to improve the final

translation result.

The data used in this system were obtained from a number of websites and institutions, such as the Chinese Medical Journal Network, the Science Foundation Shared Services Network, WMT (world machine translation)[34], and the Natural Language Processing Group of Nanjing University, with a data scraping system. The English and Chinese bilingual corpus contains both of the biomedical corpus and general corpus can enrich the dictionary and help the translation model to learn the grammaticals. When training, the model looks up the word vector of each word from the dictionary and performs the position embedding. Then, the embedded word vector is inputted to the transformer to train the model. Through the encoding and decoding operations of the transformer, we obtain a translation model that conforms the grammar and semantics of the biomedical field. In order to make the sentence more accurate and fluent, we input the obtained translation result to the semantic disambiguation and "MedDRA" dictionary to handle the <unk> words. Then, we output the final results. Figure 3 is the framework of the proposed BioNMT system.

Algorithm 1 Train Model&Generate Translation Results

Input: source sentence
Output: target sentence
  function TRAIN($HibridCorpus$)
    for $i = 0 \rightarrow length(HibridCorpus)$ **do**
      $SentencesList[i] \leftarrow SegWord(HibridCorpus[i])$
    end for
    for $i = 0 \rightarrow length(SentencesList) - 1$ **do**
      $WordsList \leftarrow SentencesList[i]$
      $WordVectors \leftarrow WordEmbedding(WordsList[0 : length(WordsList - 1])$
      $SentencesVector[i] \leftarrow WordVectors$
    end for
    while $k < 20$ **do**
      for $j = 0 \rightarrow length(SentencesVector) - 1$ **do**
        $TrainVectorsMatrix \leftarrow SentencesVector[i : batchsize - 1]$
        $Model \leftarrow Transformer(TrainVectorsMatrix)$
        $j \leftarrow j + batchsize$
      end for
      $k \leftarrow k + 1$
    end while
    return $Model$
  end function

  function TRANSLATE($SourceSentence$)
    $SentenceVector \leftarrow WordEmbedding(SourceSentence)$
    $FirstTarget \leftarrow Model(SentenceVector)$
    if $< unk > in\ FirstTarget$ **then**
      $SencondTarget \leftarrow SolveUnk(FirstSentence)$
      $FinalSentence \leftarrow SemanticDisambiguations(SencondTarget)$
    else
      $FinalSentence \leftarrow SemanticDisambiguations(FirstTarget)$
      return $FinalSentence$
    end if
  end function

Algorithm 1 briefly describes the model training and translation process of the BioNMT system. The detailed training process has been described in chapters 3.1 and 3.2. In the translation process, as shown in function *TRANSLATE*, the BioNMT system receives the foreign language sentence (*Source-Sentence*) that needs to be translated, and first performs regular expression matching and filtering on the sentence to be translated, and then performs word vector conversion. Here we use the same

"*position coding + one-hot*" structure as in the training process. Then, treating the resulting vector input into the model has been trained for translation. After getting the results, if there is <unk>, the aid has been tidied biological dictionary and custom rules <unk> the solution. Finally, the semantic disambiguation work is performed on the obtained sentence to obtain the final translation result.

4   Experiments

This section mainly introduces translation model training and analysis of the results. It consists of the data collection, experimental environment, evaluation indicators and complete translation system construction process. The evaluation indicators were mainly evaluated by BLEU, accuracy and perplexity, which are widely used for translation model evaluation[3].

4.1   Dataset Details

For the machine translation model training in the biomedical field, we needed to collect biomedical corpora. The Chinese Medical Journal Network[36] has collected papers from 190 medical journals. The medical corpora collected from this website are highly professional and highly aligned with bilingual corpora. Neural machine translation requires precise corpus alignment. Additionally, the Science Fund Shared Service Network[35] includes the abstracts and project briefs of all fund projects in the domestic biological and medical fields for obtaining bilingual corpora containing more specialized words in the biomedical field, and then natural language processing can be used. This technique performs bilingual alignment work. The author defined a series of regular expressions to filter and troubleshoot possible confused characters, special symbol expression errors, and repeated corpora in the corpus.

In the experimental training, a total of 5,455,344 aligned Chinese and English sentences were used, and the universal and specific corpora were combined for experiments. The number of corpora in the training set, validation set, and test set accounted for 73.4%, 12.4%, and 14.2%, respectively.

4.2   Experimental Results

We used the better-performing "LSTM + Attention" framework as a comparison framework. Then, we used four corpora to compare the results, that are universal data collections (LSTM + Attention), universal data collections (BioNMT), hybrid data collections (LSTM + Attention), and hybrid data collections (BioNMT) (hybrid data collections as shown in the last section, the universal data collections and biomedical data collections were combined). Four different training models were used to train the translation model, and the results are shown in Table1.

Table 1: The comparison the of Experimental Results of Four Different Algorithms

corpus	BLEU	Accuracy(%)	Perplexity
Universal corpus(LSTM+Attention)	24.10	36.95	50.29
Universal corpus(BioNMT)	29.06	42.16	27.85
Hybrid corpus(LSTM+Attention)	27.32	37.65	45.51
Hybrid corpus(BioNMT)	**33.17**	**50.50**	**16.71**

Table 1 shows the BLEU, accuracy, and perplexity values [3] of the four models based on different corpora or different frameworks.

From the above results, we make the following discussion:

1) With the calculation of the BLEU value, we can clearly see the advantages of the BioNMT framework. BioNMT uses the transformer for encoding and decoding. During the decoding process, mask multihead self-attention is used to select the context. Additionally, mask multihead self-attention defines the entire encoding and decoding process. The processing of the current word only has a contextual relationship with the word before the word. The words behind the word are not considered. "Mask" masks all the word information behind the word, which not only reduces the calculation cost but also ensures that the context is not lost in long sequences.

(a) Accuracy of training

(b) Perplexity of training
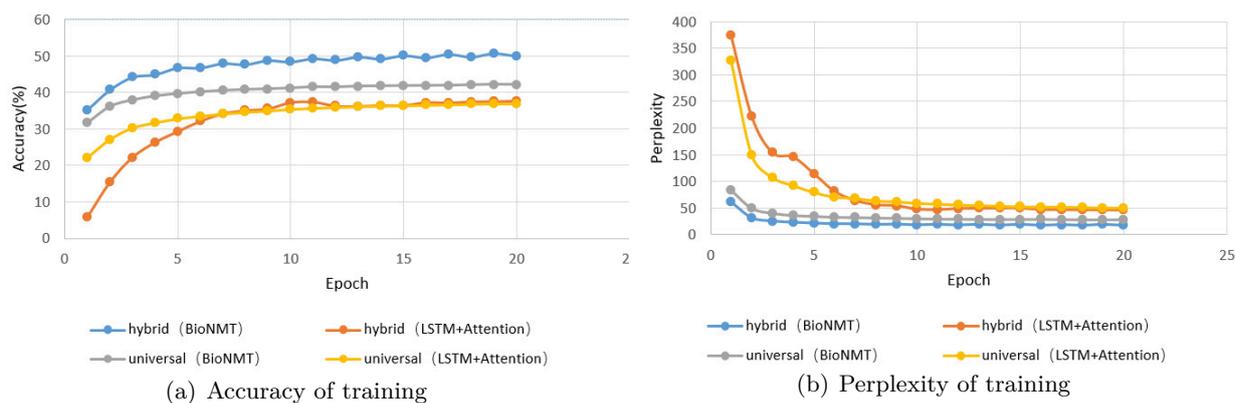
Figure 4: Results of training

2) On the universal dataset, the BioNMT framework is significantly better than the "LSTM + Attention" framework in all indicators. Compared with the translation model based on "LSTM + Attention", the accuracy of the translation model obtained by BioNMT improved by 5.21%. Additionally, the model generated by BioNMT reduced the perplexity by 44.62%. The decrease in the perplexity indicates that the probability of the BioNMT model selecting the correct translated word increased to a certain extent, which can make the translation result closer to the human translation result.

3) In comparison experiments with biomedical corpora, the translation model trained using the hybrid dataset has significantly improved on BLEU,accuracy and perplexity. Additionally, the translation model based on the hybrid dataset BioNMT is 8.34% more accurate than the translation model based on the universal dataset BioNMT, and the perplexity was reduced by 40%. BioNMT was superior to the model based on the universal dataset and the model based on the "LSTM + Attention" framework in the three evaluation indicators.

We investigated the changes in accuracy during the training process in Figure 4.a and the learning curve of the perplexity rate in Figure 4.b. Comparing the learning curves of the accuracy rate and perplexity among different algorithms during the training process, we can clearly see that with the continuous increase in the number of training epochs, the BioNMT framework tended to better convergence than the "LSTM + Attention" framework during the training process. It also ensured that the accuracy rate was higher than the latter. At the beginning of training, the perplexity of "LSTM + Attention" was significantly higher than that of the BioNMT framework. Although the perplexity of the "LSTM+Attention" framework was also significantly reduced as the number of training epochs increased, the perplexity of the transformer framework was lower than that of the "LSTM + Attention" framework when the two models converged.

The aforementioned analysis and experimental results show that the translation model obtained by the BioNMT framework based on the hybrid dataset was superior to all the other models. Additionally, we added semantic disambiguation and external dictionary assistance to improve translation accuracy. We used the optimal translation model (BioNMT model based on the hybrid dataset) to translate the biomedical corpus. To obtain more professional translation results on this basis, before returning the result, we used a lexicon for biomedical specialties called "MedDRA" and a universal English-Chinese dictionary to enrich the vocabulary of the external dictionary. After adding the semantic elimination manifold and external dictionary, as shown in Table 2, the translation results of the BioNMT system is significantly improved.

By comparing the results in Table 2, we can find that when the biomedical vocabulary has few or no biomedical nonsolitary words, there is not much difference between adding an external dictionary and semantic disambiguation and with out the strategy. However, when facing biomedical terms, the translation model and without an external semantic disambiguation dictionary generated solitary <unk>. After adding the external "MedDRA" dictionary and semantic disambiguation work, the proposed BioNMT system translates the <unk> with professional and certain extents and increases the fluency of the translation sentence.

Table 2: The effect of the MeDRA lexicon

Source sentence#1	Leukemia in infants is rare but generates tremendous interest due to its aggressive clinical presentation in a uniquely vulnerable host, and its fascinating biology.
No MedDRA and semantic disambiguation	婴儿的白血病虽然罕见，但是由于其独特的临床表现而产生巨大的兴趣，以及它迷人的生物学。
Add MedDRA and semantic disambiguation	婴儿中的白血病很少见，但由于其独特的临床表现，其强有力的临床表现以及它迷人的生物学，对其产生了巨大的兴趣。
Source sentence#2	Using representative clinical case presentations, we review the key clinical, pathologic and epidemiologic features of infant leukemia, including the high frequency of KMT2A gene rearrangements.
No MedDRA and semantic disambiguation	采用典型临床病例报告，回顾婴儿白血病的临床、病理及流行病学特点，包括\<unk\>基因重排频率的高频率。
Add MedDRA and semantic disambiguation	通过典型临床病例介绍，我们回顾了婴儿白血病的关键临床、病理和流行病学特点，包括KMT2A基因重排高频。
Source sentence#3	We highlight recent discoveries that elucidate the molecular biology of infant leukemia and suggest novel targeted therapeutic strategies, including modulation of aberrant epigenetic programs, inhibition of signaling pathways, and immunotherapeutics
No MedDRA and semantic disambiguation	我们强调最近的发现，这些发现阐明了婴儿白血病的分子生物学，并建议新的\<unk\>策略，包括异常的表观遗传程序，抑制信号通道，以及\<unk\>。
Add MedDRA and semantic disambiguation	我们着重介绍了新发现的婴儿白血病的分子生物学，提出了新的靶向治疗策略，包括表观遗传学调控的调制，抑制信号通路和免疫和免疫疗法。

To further evaluate the proposed BioNMT model, we also released a web service (http://39.98.161.93:8000/index) and compared it with the widely used translation systems, Google Translate and Baidu Translate. The compared results are shown in Table 3.

Table 3: Comparison of translation results between BioNMT system and other translation systems

Source sentence	With the ever-growing volume of online information, recommend systems have been an effective strategy to overcome such information overload. The utility of recommendation systems cannot be overstated, given their widespread adoption in many web applications, along with its potential impact to ameliorate many problems related to overchoice.
BioNMT System	随着在线信息信息量的不断增加，推荐系统已成为解决这些信息过载的有效策略。由于在许多网络应用中广泛采用，推荐系统的效用不可小觑，而且它对解决相关问题有潜在的影响。
Baidu Translation System	随着在线信息量的不断增加，推荐系统已成为克服此类信息过载的有效策略。由于推荐系统在许多Web应用程序中的广泛采用，以及它对改善许多与过度选择相关的问题的潜在影响，因此不能夸大推荐系统的实用性。
Google Translation System	随着在线信息量的不断增长，推荐系统已成为克服此类信息过载的有效策略。推荐系统的实用性不能高估，因为它已在许多Web应用程序中得到广泛采用，并且它可能会缓解与选择过多相关的许多问题。

In Table 3, we can clearly see that during the paragraph translation, the word "web" was not successfully translated in Google Translate and Baidu Translate, and the translated meaning of the sentence was misinterpreted. More importantly, for the translation of "the utility of recommendation systems cannot be overstated" in the source sentence, neither the Google Translation System nor the Baidu Translation System gave the correct translation, and the semantics of their translation results were quite different from the actual meaning of the sentence. Experiments show that after adding semantic disambiguation, BioNMT has a better ability to understand paragraphs and resolve ambiguity.

5 Conclusion

In this paper, two types of machine translation models that have become popular in the past decade were introduced: statistical-based machine translation and neural network-based machine translation. Based on transformer, a highly professional machine translation model was proposed. This model improved the accuracy of the original transformer training model by integrating related technologies in natural language processing and the use of auxiliary dictionaries.

For the future work, we will focus on two aspects. (1) Optimizing the current BioNMT system, increasing the accuracy by adding professional training data, and attempting to use the BERT[10]

model for training. (2) Most of the translation work in the biomedical field are paragraph translations. However, the current stage of BioNMT's translation of paragraphs is sentence-to-sentence translation, and then overall splicing and semantic disambiguation are performed. The next step is to conduct comprehensive research on paragraph translation. It is expected that paragraph translation in the professional field can be achieved instead of only maintaining the current level of sentence translation.

Acknowledgment

References

[1] Bahdanau, D.; Cho, K.; Bengio, Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate. *3rd International Conference on Learning Representations*, San Diego, 2015.

[2] Bazzi, I.; Glass, J.R. (2000). Modeling Out-of-vocabulary Words for Robust Speech Recognition. *Proc. of ISCA ASR2000*, 401-404, 2000.

[3] Blei, D.M.; Ng, A.Y.; Jordan, M.I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022, 2003.

[4] Bo, Q.; Xiong, N.; Zou J. et al. (2007). Internationally agreed medical terminology: Medical Dictionary for Regulatory Activities. *Chinese Journal of Clinical Pharmacology and Therapeutics*, 2007.

[5] Brazill, S. (2016). *Chinese to English Translation: Identifying Problems and Providing Solutions*. Graduate Theses & Non-Theses. 71, 2016.

[6] Bu, F.; Gharajeh, M.S.(2019). Intelligent and Vision-based Fire Detection Systems: a Survey. *Image and Vision Computing*, 91, 2019.

[7] Chen, H.B.; Hsen, H.H.; Chang, H.A. (2017). A Simplification Translation Restoration Framework for Domain Adaptation in Statistical Machine Translation: A Case Study in Medical Record Translation. *Computer Speech & Language*, 42:59-80, 2017.

[8] Cheng, J.; Dong, L.; Lapata, M. (2016). Long Short-Term Memory-Networks for Machine Reading. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 551-561, 2016.

[9] Cho, K.; van Merrienboer, B.; Gulcehre, C. et al. (2014). Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724-1734, 2014.

[10] Devlin, J.; Chang, M.W.; Lee, K. et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT 2019*, 1, 4171-4186, 2019.

[11] Garcia-Castillo, D.;Fetters, M.D. (2007). Quality in Medical Translations: A Review. *Journal of Health Care for the Poor and Underserved*, 18(1): 2007.

[12] Guan, R.;Zhang, H.; Liang, Y. et al. (2020). Deep feature-based text clustering and its explanation. *IEEE Transaction on Knowledge and Data Engineering*, 2020.

[13] Gulcehre, C.; Ahn, S.; Nallapati, R. et al. (2016). Pointing the Unknown Words. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016.

[14] Khanmohammadi, S.; Gharajeh, M.S. (2017). A Routing Protocol for Data Transferring in Wireless Sensor Networks Using Predictive Fuzzy Inference System and Neural Node, *Ad Hoc & Sensor Wireless Networks*, 38(1-4), 103–124, 2017.

[15] Kuzmina, O. D., Fominykh, A.D; Abrosimova, N.A. (2015). Problems of the English Abbreviations in Medical Translation. *Procedia-Social and Behavioral Sciences*, 199, 548-554, 2015.

[16] Luong, M.T.; Pham, H.; Manning, C.D. (2015). Effective Approaches to Attention-based Neural Machine Translation. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 1412-1421, 2015.

[17] Manning, C.D.; Raghavan, P.; Schutze, H. (2008). *Introduction to information retrieval*, vol. 1, Cambridge University Press, Cambridge, 2008.

[18] Meng, X.; Liu, X.Z. Li, Y.Y. et al. (2020). Correlation between Genotype and Phenotype in 69 Chinese Patients with USH2A Mutations: A comparative study of the patients with Usher Syndrome and Nonsyndromic Retinitis Pigmentosa. *Acta Ophthalmologica*, 2020.

[19] Mercy, O.E. (2006). English-Edo Medical Translation. *Perspectives: Studies in Translatology*, 13(4), 268-277, 2006.

[20] Mi, H.; Wang, Z.; Ittycheriah, A. (2016). Supervised Attentions for Neural Machine Translation. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2283-2288, 2016.

[21] Mikolov, T.; Karafiát, M.; Burget, L. et al. (2010). Recurrent Neural Network based Language Model. *INTERSPEECH*, 1045-1048, 2010.

[22] Mitka, M. (2001). Tearing down the Tower of Babel: Medical Translation in today's world. *Journal of the American Medical Association*, 285(6), 722-3, 2001.

[23] Neubig, G. (2017). Neural Machine Translation and Sequence-to-sequence Models: A Tutorial. arXiv:1703.01619 [cs.CL], 2017.

[24] Papineni, K.; Roukos, S.; Ward, T. et al. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. *Computational Linguistics*, 40, 311-318, 2002.

[25] Pascanu, R.; Mikolov, T.; Bengio, Y. (2013). On the Difficulty of Training Recurrent Neural Networks. *Proceedings of the 30 th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013*, (3), 1310-1318, 2013.

[26] Peters, P., Qian, Y. Ding, J. (2018). Translating medical terminology and bilingual terminography. *Lexicography ASIALEX*, 3, 99–113, 2018.

[27] Rocktäschel, T.; Grefenstette, E.; Hermann, K.M. et al. (2015). Reasoning about Entailment with Neural Attention. *Proceedings of ICLR2015*, 2015.

[28] Sennrich, R.; Haddow, B.; Birch, A. (2016). Neural Machine Translation of Rate Words with Subword Units. *The 54th Annual Meeting of the Association for Computational Linguistics*, 1715-1725, 2016.

[29] Shen, L.; Chen, H.; Yu, Z. et al. (2016). Evolving support vector machines using fruit fly optimization for medical data classification. *Knowledge-Based Systems*,96, 61-75. 2016.

[30] Slimani, T.(2013). Description and Evaluation of Semantic Similarity Measures Approaches *International Journal of Computer Applications*, 80(10), 25-33, 2013.

[31] Vaswani, A.; Shazeer, N.; Parmar, N. et al. (2017). Attention is All You Need. *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA*, 5998-6008, 2017.

[32] Wu, Y.; Schuster, M.; Chen, Z. et al. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. arXiv:1609.08144, 2016.

[33] Yandell, M.D.; Majoros, W.H. (2002). Genomics and Natural Language Processing. *Nature Reviews Genetics*, 601-610, 2002.

[34] Ziemski, M.; Junczys-Dowmunt, M.; Pouliquen, B. (2016). The United Nations Parallel Corpus. *Language Resources and Evaluation (LREC'16)*, 2016.

[35] [Online]. Science Fund Shared Service Network. Available: https://output.nsfc.gov.cn/

[36] [Online]. Chinese Medical Journal Network. Available: https://medjournals.cn/index.do

C | O | P | E

Member since 2012
JM08090

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:
H. Liu, Y. Liang, L. Wang, X. Feng, R. Guan (2020). BioNMT: A Biomedical Neural Machine Translation System, *International Journal of Computers Communications & Control*, 15(6), 3988, 2020.
    https://doi.org/10.15837/ijccc.2020.6.3988.