

A study of attention-based Neural Machine Translation models on Indian Languages

Ayan Das, Pranay Yerra, Ken Kumar, Sudeshna Sarkar

Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur, WB, India

ayan.das@cse.iitkgp.ernet.in

ypranay.hasan@cse.iitkgp.ernet.in

ken@iitkgp.ac.in

sudeshna@cse.iitkgp.ernet.in

Abstract

Neural machine translation (NMT) models have recently been shown to be very successful in machine translation (MT). The use of LSTMs in machine translation has significantly improved the translation performance for longer sentences by being able to capture the context and long range correlations of the sentences in their hidden layers. The attention model based NMT system has become state-of-the-art, performing equal or better than other statistical MT approaches. In this paper, we studied the performance of the attention-model based NMT system on the Indian language pair, Hindi and Bengali. We analysed the types of errors that occur in morphologically rich languages when there is a scarcity of large parallel training corpus. We then carried out certain post-processing heuristic steps to improve the quality of the translated statements and suggest further measures.

1 Introduction

Deep Neural Network has been successfully applied to machine translation. The work of (Cho et al., 2014; Kalchbrenner and Blunsom, 2013; Sutskever et al., 2014) have shown that it is possible to build an end-to-end machine translation system using neural networks by introducing the *encoder-decoder* model. NMT systems have several advantages over the existing phrase-based statistical machine translation (SMT) systems (Koehn et al., 2007). The NMT systems do not assume any domain knowledge or linguistic features in source and target language sentences. Secondly, the entire encoder-decoder models are jointly trained to maximize the translation quality as opposed to the phrase-based SMT systems in which the individual components needs to be trained and tuned separately for optimal performance.

Although the NMT systems have several advantages, their performance is restricted in case of low-resource language pairs for which sufficiently large parallel corpora is not available and the language pairs whose syntaxes differ significantly. Morphological richness of language pairs poses another challenge for NMT systems that do not have any prior knowledge of the languages as it tends to increase the number of surface forms of the words due to inflectional attachments resulting in an increased vocabulary of the languages. Moreover, the inflectional forms have their semantic roles that have to be interpreted for proper translation. In order to enable the NMT systems to learn the roles of the inflectional forms automatically we need sufficiently large data. However, sufficiently large parallel data may not be available for low-resource morphologically rich language pairs. Most of the Indian languages are morphologically rich and there is lack of sufficiently large parallel corpus for Indian language pairs. Given our familiarity with Bengali and Hindi, we took up this task as a case-study and evaluated the performance of NMT models on Indian language pair-Hindi and Bengali. We then analyzed the resulting translated sentences and suggested post-processing heuristics to improve the quality of the translated sentences. We have proposed heuristics to rectify the incorrect translations of the named entities. We have also proposed a heuristic to translate and predict the position of untranslated source words.

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>

2 Related work

Neural machine translation

Neural machine translation models attempt to optimize $p(e|f)$ directly by including feature extraction using a single neural network. The entire translation process is done using an encoder-decoder framework (Cho et al., 2014; Kalchbrenner and Blunsom, 2013; Sutskever et al., 2014) where the encoder encodes f into a continuous space representation and the decoder uses the encoding of f and decoding history to generate the target language sentence e . The encoders and decoders are essentially recurrent neural networks (RNNs) (Mikolov et al., 2010; Mikolov et al., 2011) or its gated versions (Gated Recurrent Unit (GRU) (Chung et al., 2014; Chung et al., 2015) or Long-Short Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997)) capable of learning long-term dependencies.

Cho et al. (2014) proposed to use the final state of the hidden layer of the encoder as the encoding of the source sentence. Sundermeyer et al. (2014) used a bi-directional RNN in the encoder and used the concatenation of the final states of the hidden layers as the encoding of the source sentence. Sutskever et al. (2014) proposed to train the encoder using the source sentence in the reverse ordering of words and the decoder in the correct word ordering of target sentence.

Bahdanau et al. (2014) and Luong et al. (2015) have proposed the attention-based translation model. The encoder of the model is a bi-directional RNN (Schuster and Paliwal, 1997). The annotation vectors \mathbf{h}_j^T (where \mathbf{h}_j encodes the j^{th} word with respect to the other words in the source sentence) are obtained by concatenating the two sequences of hidden layers $\overrightarrow{\mathbf{h}}_j^T$ and $\overleftarrow{\mathbf{h}}_j^T$ which are obtained by training the forward RNNs on the original sequence of input sentences and the backward RNNs on the reverse sequence of input sentences, such that $\mathbf{h}_j^T = [\overrightarrow{\mathbf{h}}_j^T; \overleftarrow{\mathbf{h}}_j^T]$. The decoder consists of a single layer GRU. At time step t , the alignment layer decides the relevance of the source words for the word to be predicted. The relevance (α_{tj}) of the j^{th} annotation vector at time t is determined by a feed-forward neural network that takes the previous state of the hidden layer of the decoder (\mathbf{s}_{t-1}), embedding of the last predicted word (\mathbf{y}_{t-1}) and the j^{th} annotation vector (\mathbf{h}_j) as input. The hidden state of the decoder at time t is computed as a function f_r of the previous hidden state \mathbf{s}_t , the context vector \mathbf{c}_t and the previous predicted word \mathbf{y}_{t-1} , where f_r is a GRU and \mathbf{c}_t is the context vector for the t^{th} word is obtained as a sum of the annotation vectors weighted by the corresponding relevance scores.

$$\mathbf{s}_t = f_r(\mathbf{s}_{t-1}, \mathbf{y}_{t-1}, \mathbf{c}_t) \quad (1)$$

Finally, the conditional distribution over the words is obtained by using a deep output layer.

$$p(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{x}) \propto \exp(\mathbf{y}_t^T (W_o f_o(\mathbf{s}_t, \mathbf{y}_{t-1}, \mathbf{c}_t) + b_o)) \quad (2)$$

where, \mathbf{y}_t is the indicator vector corresponding to a word in the target vocabulary. W_o and b_o are the weights and bias of the deep layer and f_o is a single-layer feed-forward neural network with a two-way maxout layer (Goodfellow et al., 2013).

Once the model learns the conditional distribution, then given a source sentence we can find a translation that approximately maximizes the conditional probability using, for instance, a beam search algorithm.

3 Proposed Method

In this paper, we studied the performance of attention-model based NMT system (Bahdanau et al., 2014) on Bengali-Hindi language pair. The attention-based NMT models have shown near state-of-the-art performance for the language pairs, English-French and English-German. One of the advantages for these language pairs was the availability of good-quality, sentence aligned parallel corpora from WMT'14 dataset. We implemented the same attention-model based NMT system (Bahdanau et al., 2014) and studied its performance on the Indian language pair, Bengali and Hindi. Both Hindi and Bengali belong to the same family of language and share some high-level syntactic similarities such as Subject-Object-Verb (SOV) sentence structure which lead us to believe that the attention model will be useful for this language pair.

3.1 Resources used

Monolingual Hindi and Bengali corpora were used to train word2vec (Mikolov et al., 2013) to obtain the word embeddings. The monolingual Hindi corpus was obtained from the ILTP-DC (www.tdil-dc.in/) which consists of about 45 million sentences. The FIRE 2011 (<http://www.isical.ac.in/clia/2011/>) monolingual Bengali news corpus consisting of about 3.5 million sentences was used to obtain the Bengali word vectors. The Bengali-Hindi parallel corpus was obtained from ILCI (sanskrit.jnu.ac.in/ilci), comprising of 50000 sentences obtained from tourism and health domains was used for the experiments. From the 50000 Bengali-Hindi parallel sentences, 49000 sentence pairs were randomly selected for training and remaining 1000 sentence pairs were used for testing. In order to reduce the size of the vocabulary we replaced all the numeric values by the 'NUM' token.

3.2 Our implementation of the Attention-Model

The attention-model based NMT model (Bahdanau et al., 2014) was implemented in Theano (Theano Development Team, 2016). The number of hidden layer units (n) was taken as 1000, the word embedding dimensionality as 620 and the size of the maxout hidden layer in the deep output was 500. The number of hidden units in the alignment model was 1000. We used gradient-clipping with a clipping threshold of 5. The model was trained using stochastic gradient descent with a learning rate of 0.0627 and batch size of 1. The model was run on a Nvidia Tesla K40C GPU machine.

4 Results

MOSES (a phrase-based SMT model) (Koehn et al., 2007) was used as a baseline system for comparison of the NMT model. The Bengali-Hindi parallel corpus obtained from ILCI (sanskrit.jnu.ac.in/ilci) comprising of 50000 sentences obtained from tourism and health domains was used for the experiments. From the 50000 Bengali-Hindi parallel sentences 49000 sentence pairs were randomly selected for training the model and remaining 1000 sentence pairs were used for testing. Out of the 49000 sentence-pairs in the training set, 15000 pairs (tuning set) were randomly selected for tuning the model parameter (weights) using MERT system (Minimum Error Rate Training) (Och, 2003) which searches for weights optimizing a given error measure which is BLEU score in our case. The SRILM (Stolcke, 2002) language model was trained using the entire training dataset comprising of 49000 sentence pairs.

We compared the performance of the attention-model based NMT system with that of the baseline MOSES phrase-based SMT system. We ran the NMT model for 25 epochs. Table 1 summarizes the results.

Table 1: Comparison of 1) attention-based NMT model and 2) MOSES phrase-based SMT system.

Translation model	BLEU score	Iterations
MOSES	14.35	-
Attention-based translation model	20.41	25

As the BLEU score suggests, the translation quality of the NMT system surpasses that of the MOSES (Koehn et al., 2007) by a significant margin. Out of the 1000 sentence pairs used for testing, we randomly picked up 8 sentences and present them in Appendix 1. We observe that in five of the eight examples the translation results of the attention model are clearly better than that produced by MOSES. The translation by MOSES is slightly better in two cases whereas in one example, both models have almost similar translation results. This was the general trend in all the test examples with the attention model performing relatively better than MOSES in cases of longer source sentences (Figure 1).

We also compared the BLEU score of our NMT model over 25 iterations with the MOSES system and saw that only after 5 iterations, the NMT model started performing better than MOSES (Figure 2).

5 Analysis

Our implementation of the attention based NMT model significantly outperforms MOSES in terms of BLEU scores. However on manual inspection of some random samples, we observed significant errors

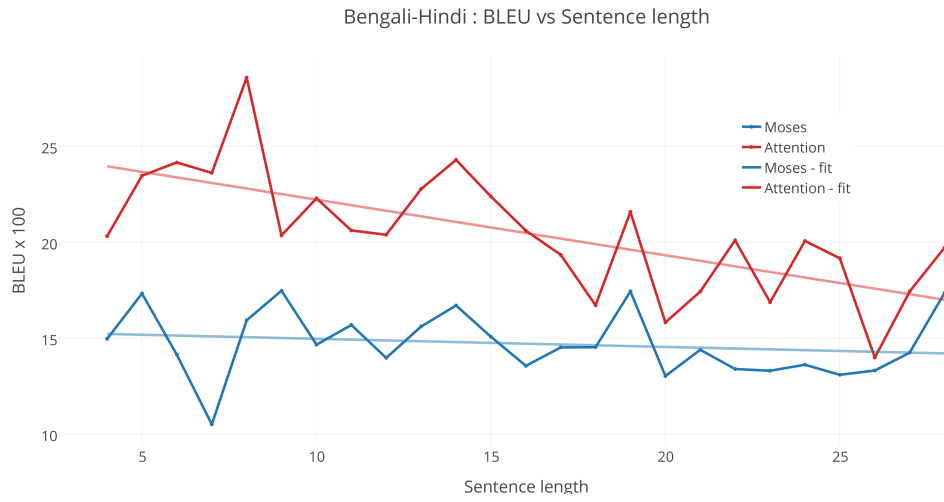


Figure 1: Variation of BLEU score with sentence length. The plot shows the BLEU score against the source sentence length.

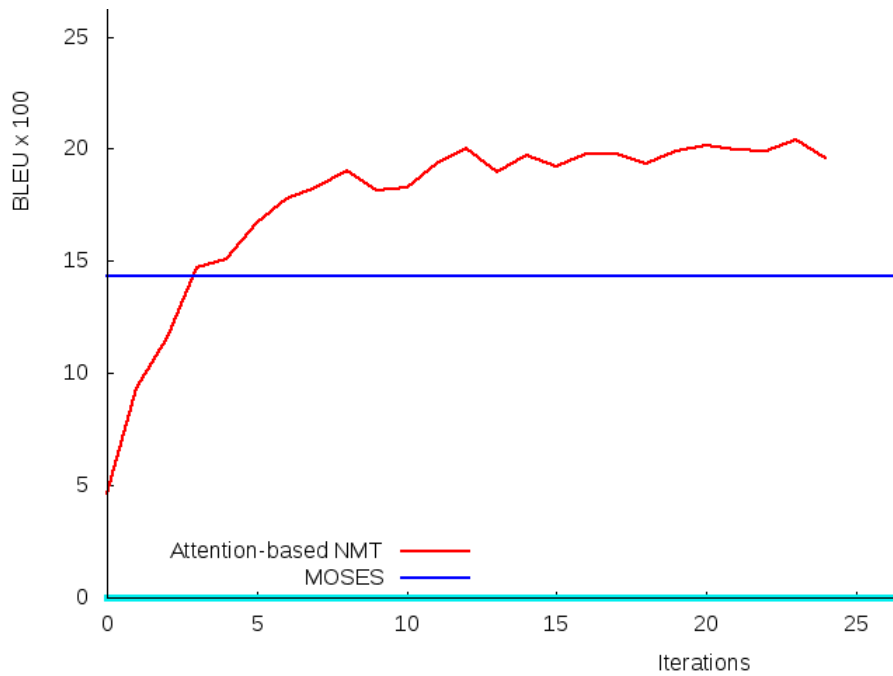


Figure 2: Comparison of BLEU score of the NMT model over 25 iterations with the baseline MOSES system.

in translation of named entities. Due to the limited size of the corpus, many named entities were absent from the vocabulary and hence the model was not able to find a suitable translation for them. Thus the quality of the translated sentences suffered. We propose the following algorithm as a post processing step in order to deal with named entities.

5.1 Dealing with named entities

Algorithm 1 summarizes the steps for correcting the errors due to wrong translation of the named entities for a test Bengali sentence.

Algorithm 1: Correction of errors due to wrong translation of the named entities

input : Bengali sentence ($B = \{b_1, b_2, \dots, b_M\}$), translated Hindi sentence, word alignment scores (α values) for the translation

output: Corrected Hindi sentence

```

1 for each word  $b_j$  in  $B$  do
2   | if  $b_j$  is named entity then
3   |   | Tag  $b_j$  as NE
4   | end
5 end
6 for each tagged  $b_j$  do
7   | Transliterate the tagged  $b_j$  into the target language (Hindi) using any open-source transliteration
   | tool.
8   | Find the index  $i$  in the translated sentence for which the value of  $\alpha_{ij}$  is maximum. /* This
   |  $h_i$  corresponds to the word in the target language sentence
   | whose translation has been most highly influenced by  $b_j$ . */
9   | Replace  $h_i$  with the transliterated word of  $b_j$ 
10 end

```

The named entities in the test sentences were identified manually. For transliterating the Bengali words to Hindi we used a Bengali-Hindi transliterator developed at our institute. We are working on developing a good quality NER system for Bengali and automating the process of identification and transliteration of the Bengali named entities. On manually observing the target sentences after performing the heuristic, it was found that the overall quality of the translated sentences had gone up and they were more relevant to the context of the source sentences. However this post-processing step resulted in slight decrease in the BLEU score. Part of this may be due to the fact that direct transliteration of the named entities from the source language to the target language without stemming or lemmatization could not take into account the inflectional differences in the source and target language. In Appendix 2 we present five examples. Words like কচ্ছের (kachchh-of) in Bengali when transliterated directly into Hindi results in कच्छेर (kachchher), which is indeed the direct Hindi transliteration of the Bengali word including the inflection *-of* but fails to capture the context in which it is used and how it should be used (with proper inflection) in the target language sentence. Similarly in the third example sentence, the Bengali word এশিয়াত (Asiyan) transliterate directly to एशियान (Asiyan) in the target language but the word एशियाई (Asia-of) was more suited to the context of the sentence. But as we mentioned earlier, it was manually observed that the relevance of the target sentences in relation to the source sentences was found to be more than those of the translated sentences before correction.

5.2 The problem of untranslated words

The lack of sufficient amount of training data meant that we had to work with a limited vocabulary size for the source as well as the target language. This resulted in many phrases in the source sentences not getting translated simply because our model was not able to find words in the target language vocabulary for that phrase. Algorithm 2 summarizes the post-processing heuristic to deal with such untranslated words.

Algorithm 2: Prediction of translations for untranslated words

input : Bengali sentence ($B = \{b_1, b_2, \dots, b_M\}$), translated Hindi sentence ($H = \{h_1, h_2, \dots, h_N\}$), word alignment scores (α values) for the translation

output: Corrected Hindi sentence

```
1 for each word in  $b_j$  in  $B$  do
2   if  $b_j$  is NOUN then
3      $untranslated = false$  for all  $h_i$  in  $H$  do
4       if  $\alpha_{ij} > threshold$  then
5          $untranslated = true$ 
6         break
7       end
8     end
9     if  $untranslated$  then
10      Find the index  $i$  in the target sentence for which the value of  $\alpha_{ij}$  is maximum.
11      Insert the transliteration of  $b_j$  in Hindi into the target sentence at the  $i^{th}$  position.
12    end
13  end
14 end
```

The intuition behind this heuristic is very simple. The index which is most highly influenced by the untranslated word in the source sentence is the probable position for the translation of that word to occur. We simply transliterated those words and put them at position that they influence the most in the target sentence. We show five randomly selected examples in Appendix 3 ($\alpha_{ij} = 0.2$). Out of the five examples, we find that the quality of 4 sentences improved, while for one sentence, it did not improve much.

We observed that the NMT system is better at translating the postpositions than the SMT system. We need to further investigate this observation. The reason is not yet clear to us and we are working to find the explanation for this observation.

6 Conclusion

In this paper we showed that the performance of the attention-model based NMT system for the Indian language pair, Bengali and Hindi is better than the existing SMT model of MOSES. We then analysed the output translated sentences and observed that there were significant translation errors in case of named entities and rare words. In order to improve the results, we implemented certain post-processing heuristic steps and manually observed that we were able to make the translated sentences more relevant in context to the source sentences.

References

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. In *ICLR 2015*.
- Kyunghyun Cho, Bart van Merriënboer, Çağlar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. 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)*, pages 1724–1734, Doha, Qatar, October. Association for Computational Linguistics.
- Junyoung Chung, Çağlar Gülçehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, abs/1412.3555.
- Junyoung Chung, Çağlar Gülçehre, KyungHyun Cho, and Yoshua Bengio. 2015. Gated feedback recurrent neural networks. *CoRR*, abs/1502.02367.

- Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron C. Courville, and Yoshua Bengio. 2013. Maxout networks. In *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*, pages 1319–1327.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November.
- Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent continuous translation models. Seattle, October. Association for Computational Linguistics.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, ACL '07*, pages 177–180, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. *CoRR*, abs/1508.04025.
- Tomas Mikolov, Martin Karafiát, Lukás Burget, Jan Cernocký, and Sanjeev Khudanpur. 2010. Recurrent neural network based language model. In *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010*, pages 1045–1048.
- Tomas Mikolov, Stefan Kombrink, Lukás Burget, Jan Cernocký, and Sanjeev Khudanpur. 2011. Extensions of recurrent neural network language model. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2011, May 22-27, 2011, Prague Congress Center, Prague, Czech Republic*, pages 5528–5531.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.
- Franz Josef Och. 2003. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1, ACL '03*, pages 160–167, Stroudsburg, PA, USA. Association for Computational Linguistics.
- M. Schuster and K.K. Paliwal. 1997. Bidirectional recurrent neural networks. *Trans. Sig. Proc.*, 45(11):2673–2681, November.
- Andreas Stolcke. 2002. Srilm – an extensible language modeling toolkit. In *IN PROCEEDINGS OF THE 7TH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (ICSLP 2002*, pages 901–904.
- Martin Sundermeyer, Tamer Alkhouli, Joern Wuebker, and Hermann Ney. 2014. Translation modeling with bidirectional recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 14–25, Doha, Qatar, October. Association for Computational Linguistics.
- Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.
- Theano Development Team. 2016. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688, May.

Appendix A. Comparison of output of attention-based NMT model and MOSES.

Bengali	এরা শ্বেত সাধারণ বাজেটের পর্যটক।
MOSES	ये चंदोला आम बजट वाला पर्यटक है।
Attention-based TM	ये आम बजट के पर्यटक है।
Ref. translation	ये सामान्य बजट के पर्यटक है।
Bengali	পোর্ট ব্লেয়ারে পর্যটকদের থাকার জন্য ছোটো ছোটো হোটেল আর গেস্ট হাউসের ব্যবস্থা আছে।
MOSES	पोर्ट ब्लेयर में पर्यटकों के ठहरने की के लिए छोटे छोटे होटल और गेस्ट हाउस के की व्यवस्था है है।
Attention-based TM	पोर्ट ब्लेयर में पर्यटकों का ठहरने के लिए छोटे होटल होटल और गेस्ट हाउस की सुविधा है।
Ref. translation	पोर्ट ब्लेयर में पर्यटकों के ठहरने के लिए छोटेछोटे होटल व गेस्ट हाउसों की व्यवस्था है।
Bengali	পেটে চোট লাগলে ক্ষত হতে পারে যা মাংসপেশী ও হাড় পর্যন্ত পৌঁছে যায় আর পায়ে পুঁজ হতে পারে আঙুল গলে যেতে পারে বা কালো হয়ে যেতে পারে।
MOSES	पेट में चोट लगने पर जख्म हो सकती है हो सकता है है जो की मांसपेशियों तथा की हड्डी तक पहुँच जाती है और पैरों में पीप हो सकती है हो सकता है गल के की अंगुलियों के जा सकता है या फरि काले हो जा सकता है।
Attention-based TM	पेट में चोट लगने पर जख्म हो सकते हैं जो मांसपेशियों तथा हड्डियों तक पहुँच हो और पैरों में मवाद हो पाती है या काले हो सकते हैं।
Ref. translation	पेट में चोट लगने से नासूर बन सकता है जो मांसपेशियों व हड्डियों तक पहुँच जाता है और पैर में मवाद पड़ सकती है अंगुलियाँ गल सकती हैं या काली पड़ सकती हैं।
Bengali	কম্পিউটার যতই উন্নত বা বুদ্ধিমান হোক তা কেত ও শুধু সেই কাজই করতে পারে যার বিধিটা আমরা স্বয়ং জানি ও ওকে দিয়েছি।
MOSES	कंप्यूटर चाहे कतिना भी या कुशाग्र के उन्नत से ही क्यों न हो तथा सरिफ उस के काम ही कर सकता है जिससे मैं स्वयं [] [] [] [] [] हम जानते हैं तथा उसे दी और है।
Attention-based TM	कंप्यूटर चाहे भी बेहतर या फायदेमंद साबित हो न हो या सरिफ वही काम कर सकते हैं जसि हम हम स्वयं जानते हैं।
Ref. translation	कंप्यूटर चाहे कतिना ही उन्नत या चतुर क्यों न हो वह केवल वही काम कर सकता है जिसकी वधि हिमें स्वयं ज्ञात है और जसि हमने कंप्यूटर को सखिा दिया है।
Bengali	তৈল চিকিৎসায় খনিজ তেলের কোনও ব্যবহার নেই কারণ তা স্বকে প্রবেশ করতে পারে না।
MOSES	तैलीय के इलाज में खनजि तेल के कसिी का इस्तेमाल नहीं है क्योंकि उसे त्वचा में के प्रवेश के लिए भी नहीं कर सकता।
Attention-based TM	तैलीय चिकित्सा में खनजि तेल का उपयोग न है क्योंकि यह त्वचा में प्रवेश नहीं है।
Ref. translation	तैलीय चिकित्सा में खनजि तेलों का कोई उपयोग नहीं क्योंकि वे त्वचा में प्रवेश नहीं कर सकते।
Bengali	আমাকে বাইরে দাঁড়িয়ে থাকতে দেখে এক জন মহিলা এলেন আর মসজিদের ভেতর ডেকে নিয়ে গেলেন।
MOSES	मुझे के बाहर खड़े हो महिला से नरिणय लथिा था रह को देखकर के एक जाँन और मसजदि के के भीतर ले पड़े।
Attention-based TM	मुझे बाहर खड़े होकर देखकर एक सौ आए थे और मसजदि को भीतर ले गए।
Ref. translation	मुझे बाहर खड़ा देख कर एक महिला आई और बुलाकर मसजदि के अंदर ले गई।
Bengali	হেমাৰঘাট ঝিলে নৌকায় বসে ঝিল ভ্রমণ করা আর আশেপাশের প্রাকৃতিক দৃশ্য দেখতে ভালো লাগে।
MOSES	हेमारघाट झील में नाव में बैठकर झील का भ्रमण कथिा और के आसपास के के प्राकृतिक दृश्य देखने को अच्छा लगता है है।
Attention-based TM	बरैला झील में नौका में बैठकर झील भ्रमण करने और आसपास प्राकृतिक दृश्यों को देखना अच्छा लगता है।
Ref. translation	हेमारघाट झील में नाव में बैठकर झील का भ्रमण करना और आसपास के प्राकृतिक दृश्यों को देखना बहुत अच्छा लगता है।
Bengali	ভৈরবনাথ মন্দির সোত প্রয়াগ থেকে NUM কি সি পরে আর কেদারনাথ থেকে NUM কি সি আগে পায়ে হেঁটে হল এক অত্যন্ত গুরুত্বপূর্ণ তীর্থ এবং বিশ্রাম স্থল।
MOSES	भैरवनाथ मंदिर सोन प्रयाग से NUM कमी की मी के बाद और केदारनाथ से NUM कमी की मी से पहले पैदल चलकर एक अत्यंत महत्त्वपूर्ण तीर्थ एवं वशिराम स्थल है।
Attention-based TM	भैरवनाथ मंदिर सोन महादेव से NUM कमी आगे और केदारनाथ से NUM कमी आगे पैदल यात्रा अत्यंत महत्त्वपूर्ण तीर्थ है।
Ref. translation	भैरवनाथ मंदिर सोनप्रयाग से NUM कमी आगे और केदारनाथ में NUM कमी पहले पैदल पड़ने वाला एक अत्यंत महत्त्वपूर्ण तीर्थ एवं वशिराम स्थल है।

Appendix B. Example of sentences containing named entities before and after post-processing.

Table 2: Target sentences after transliterating the named entities in the source sentences

Bengali	কচ্ছের ছোটো মরুভূমি হল জাদুনগরী।
Attention-based TM	कच्छ का छोटा रेगसितान भी है।
After transliteration	कच्छेर का छोटा रेगसितान जादुनगरी है।
Ref. translation	कच्छ का छोटा रेगसितान “ जादुईनगरी “ है।
Bengali	এগুলির নাম হল কাফনি হিমবাহ পিণ্ডারী হিমবাহ লাবন হিমবাহ ও শলাঙ্গ হিমবাহ।
Attention-based TM	इनके नाम है कि हमिनद ग्लेशियर हमिनद हमिनद है ग्लेशियर और हमिनदों ग्लेशियर भी है।
After transliteration	इनके नाम है काफनी हमिनद ग्लेशियर पण्डारी हमिनद लाबन ग्लेशियर और शलाङ्ग ग्लेशियर भी है।
Ref. translation	इनके नाम हैं कफनी ग्लेशियर पंडारी ग्लेशियर लावन ग्लेशियर और शलांग ग्लेशियर।
Bengali	তিতটি এশিয়াত গণ্ডারের মধ্যে পাওয়া সবথেকে বড় গণ্ডারটিকে দেখার জন্য পর্যটকদের বেশ ভিড় জমে।
Attention-based TM	तीन चार गैंडे के पाए जाने के सबसे बड़े गाँव को देखने के लिए पर्यटकों की काफी भीड़ है।
After transliteration	तीन एशियान गण्डारेर के पाए जाने के सबसे बड़े गाँव को देखने के लिए पर्यटकों की काफी भीड़ है।
Ref. translation	तीनों एशियाई गैंडों में यहाँ पाया जाने वाला सब से बड़ा गैंडा देखने के लिए पर्यटकों की अच्चीखासी भीड़ जुटती है।
Bengali	পোর্ট ব্লেয়ারে পর্যটকদের থাকার জন্য ছোটো ছোটো হোটেল আর গেস্ট হাউসের ব্যবস্থা আছে।
Attention-based TM	पोर्ट ब्लेयर में पर्यटकों का ठहरने के लिए छोटे होटल होटल और गेस्ट हाउस की सुविधा है।
After transliteration	पोर्ट ब्लेयारे में पर्यटकों का ठहरने के लिए छोटे होटल होटल और गेस्ट हाउस की सुविधा है।
Ref. translation	पोर्ट ब्लेयर में पर्यटकों के ठहरने के लिए छोटेछोटे होटल व गेस्ट हाउसों की व्यवस्था है।
Bengali	সম্পূর্ণ গ্যাঙ্গটক পারম্পরিক মুখোশ পরে লামারা নৃত্য করে।
Attention-based TM	संपूर्ण पर्यटन पारम्परिक जीवन की पश्चात बाद लामाओं के नृत्य करते हैं।
After transliteration	संपूर्ण ग्याङ्गटक पारम्परिक जीवन की पश्चात बाद लामारा के नृत्य करते हैं।
Ref. translation	सम्पूर्ण गंगटोक में पारम्परिक मुखोटे पहने लामाओं द्वारा नृत्य करि जाते हैं।

Appendix C.Example of sentences with untranslated words before and after post-processing.

Table 3: Target sentences after transliterating and inserting the untranslated words

Bengali	মুবারক মন্ডী প্যালেস মহল সবথেকে প্রাচীন ইমারত হল NUM সতের ।
Attention-based TM	मुबारक मंडी महल महल में सबसे प्राचीन इमारत NUM वी सदी में है ।
After transliteration	मुबारक मंडी प्यालेस महल महल में सबसे प्राचीन हल इमारत NUM वी सदी में है ।
Ref. translation	मुबारक मंडी पैलेस महल परसिर में सबसे प्राचीन इमारत NUM की है ।
Bengali	ওজন কমানো বা স্বকের দেখাশোনা প্রসঙ্গ ঘাই হোক তা কেত খাবারে তাজা ফল খাওয়ার পরামর্শ সব ব্যাপারেই দেওয়া যায় ।
Attention-based TM	वजन घटाना या त्वचा की देखभाल का चाहे चाहे क्यों न ताजे आहार को खाने की सलाह हमेशा ही दी जा है ।
After transliteration	वजन घटाना या त्वचा की देखभाल का चाहे चाहे क्यों न ताजे फल आहार को खाने की सलाह हमेशा ही दी जा है ।
Ref. translation	बात वजन घटाने की हो या त्वचा की देखभाल की खाने में ताजे फल खाने की सलाह सभी मामलों में दी जाती है ।
Bengali	পূজারী সাদা ধুতি সাদা পাঞ্জাবী ও মাথায় রাজস্বাস্ত্রী ডঙ্গীমায় পাগড়ী ধারণ করে খালি পায়ে মন্দির আসেন ।
Attention-based TM	पुजारी सफेद धोती सफेद तथा सरि में राजस्थानी रूप में धारण करके ऊपर से पैदल मंदिर हैं ।
After transliteration	पुजारी सफेद धोती सफेद पाञ्जाबी तथा सरि में राजस्थानी रूप में धारण करके ऊपर से पैदल मंदिर हैं ।
Ref. translation	पुजारी सफेद धोती सफेद कुरता व सरि पर राजस्थानी स्टाइल की पगड़ी धारण कर नंगे पैर मंदिर परसिर में आए ।
Bengali	হৃদ রোগ খুব ঝুলাস্ত হয়ে পড়া উত্তেজনা মেজাজ অস্থির বোধ করা প্রচণ্ড নিরাশা ঘুম কম হওয়া উচ্চ রক্তচাপ ইত্যাদি উত্তেজনাজনিত সাধারণ রোগ ।
Attention-based TM	दलि रोग रोग बहुत थक जाना उत्तेजना उत्तेजना का भय अत्यधिक गर्मी से कम नींद कम नींद उच्च रक्तचाप आदि आम तौर पर आम बीमारी है ।
After transliteration	दलि रोग रोग बहुत थक जाना उत्तेजना उत्तेजना का भय अत्यधिक गर्मी निराशा से कम नींद कम नींद उच्च रक्तचाप आदि आम तौर पर आम बीमारी है ।
Ref. translation	दलि की बीमारी बेहद थकान उत्तेजना के दौरे बारबार मूड बदलना बेचैनी की अवस्था घोर निराशा नींद कम आना उच्च रक्त चाप आदि तनाव की आम समस्याएँ है ।
Bengali	রক্তে অধিক শর্করা দীর্ঘকালীন জটিলতা ।
Attention-based TM	रक्त अधिक मात्रा अधिक जटिलताओं उत्पन्न होती है ।
After transliteration	रक्त अधिक मात्रा शर्करा अधिक जटिलताओं उत्पन्न होती है ।
Ref. translation	रक्त में शर्करा की अधिकता दीर्घकालीन जटिलताएँ ।