Effect of Named Entity Recognition on English-Vietnamese Neural Machine Translation

Van-Hai Vu, Quang-Phuoc Nguyen, Pum-Mo Ryu, and Cheol-Young Ock

Abstract—**Translators are becoming more and more popular and achieving reliable results since deep learning was born. English-Vietnamese machines translation (MT) still have limitations due to Vietnamese contain words with many different meanings, thus resulting in the lower accuracy of automatic MT systems. Our study applied Named Entity Recognition (NER) tool for Vietnamese sentences to determine the category of words in the English-Vietnamese parallel corpus with over 900K sentence pairs. Then, we performed experiments to assess the effect of NER on English-Vietnamese MT systems. The results showed that NER had a positive effect on MT with averagely 1.24 Bi-Lingual Evaluation Understudy (BLEU) scores and averagely 1.8 Translation Error Rate (TER) scores increased comparing to data without using NER.**

Index Terms—**English-Vietnamese machine translation, neural machine translation, named entity recognition, English-Vietnamese bilingual corpus.**

I. INTRODUCTION

In recent years, with the need to exchange large information between countries and the strong development of deep learning, automatic translation machines have been growing rapidly. Translators between popular languages like English, German, French, Chinese and Spanish have achieved remarkable results. However, translation machines related to the Vietnamese language still face many difficulties due to the lack of the Vietnamese language and research related to the Vietnamese language. A Vietnamese word may have different meanings depending on the context, thus resulting in the difficulty of MT when selecting words in context. NER classifies words in a text into predefined categories such as the name of person, organization, location, time, quantity, percent, etc. It indicates vocabulary that corresponds to different contexts, so it makes the machine easier to select words in the target language corresponding to the word in the source language.

There have been many studies on NER, especially NER for popular languages [1]–[3]. Most of the recent research about NER based on deep learning algorithms and named entities are mainly included person, location, organization etc.

NER for Vietnamese is mentioned on several papers. Le [4] described an efficient approach to improve the accuracy of a NER system for Vietnamese that combines regular

expressions over tokens and a bidirectional inference method in a sequence labelling model. Their research achieved an accuracy overall 89.66% by F1 score. Dong and Nguyen [5] proposed an attentive neural network for the task of NER in Vietnamese. Character-based language models and word embeddings were handled to encode words as vector representations while decoder layers were used to decode knowledge of input sentences and to tag label for entities. In another study, Nguyen *et al.* [6] recommended a neural architecture for Vietnamese sequence labelling tasks including NER and part-of-speech (POS). Their research paper has shown that the combination of bi-direction Long-Short Term Memory and Conditional Random Fields lead positive results in NER and POS labelling.

There are a series of papers on MT that have been published in recent years. Koehn *et al.* [7] showed an open-source toolkit for SMT named MOSES. Bahdanau *et al.* [8] established that encoding the source sentences into an auto-length vector could significantly improve the quality of the neural MT (NMT). Johnson *et al.* [9] designed a simple model that use a single NMT model to build multiple translation systems. Matusov *et al.* [10] introduced an NMT model for translation of subtitles in the domain of entertainment. Based on the NMT model, Graça *et al.* [11] have built a German-English MT system using the Back-translation method.

There are several research papers about English-Vietnamese MT. Nguyen *et al.* [12] built a bidirectional English-Vietnamese statistical MT (SMT) system using MOSES, an open-source toolkit for SMT. In this paper, they collected a parallel corpus with over 880,000 English-Vietnamese sentence pairs and over 11,000,000 Vietnamese monolingual sentences to train the statistical translation model. They used the BLEU score to evaluate the performance of their SMT system. Luong *et al.* [13] published NMT Systems for Spoken Language Domains with the MT from English to German and Vietnamese. In English-Vietnamese MT systems, they collected over 133,000 English-Vietnamese sentence pairs and the BLEU score was used to evaluate the accuracy of their MT systems. When using the test sentences available in their corpus, their MT has high accuracy, but it gives low results when using a random sentence that is not in their data. In another research, Ngo *et al.* [14] built an English-Vietnamese Bilingual Corpus for MT with over 800,000 sentence pairs and 10,000,000 English words as well as Vietnamese words has been collected and aligned at the sentence level. In term of translation experiments, they used the example-based MT models and SMT models.

The previous research in English-Vietnamese MT mostly focuses on the SMT model or their experiments performed

with a small number of sentences in a parallel corpus. In this paper, we present the basic knowledge about a neural architecture for NER and the state-of-the-art technique in MT. Moreover, we evaluate the effectiveness of NER on English-Vietnamese MT with a large dataset to training.

II. NEURAL ARCHITECTURES FOR NAMED ENTITY RECOGNITION

In this paper, we present a up-to-date neural architecture for NER that bases on bidirectional LSTMs and conditional random field (CRF) [15].

A. Long Short-Term Memory Networks

LSTMs are a special form of Recurrent neural network (RNN) and it learns to store information over the training time by recurrent backpropagation information. LSTMs were introduced by Hochreiter and Schmidhuber [16], then it was improved and applied in many areas [17]–[19]. LSTMs are designed to avoid the problem of long-term dependency without any intervention. They take a sequence of vector as input $(x_1, x_2,…, x_n)$ and return another sequence of vector as output $(h_1, h_2,…, h_n)$. Every regression network is in the form of a series of repetitive modules of the neural network. With standard RNN networks, these modules have a very simple structure including a tanh layer, as shown in Fig. 1.
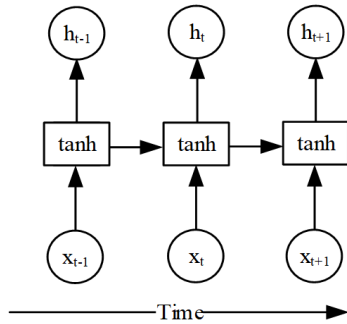

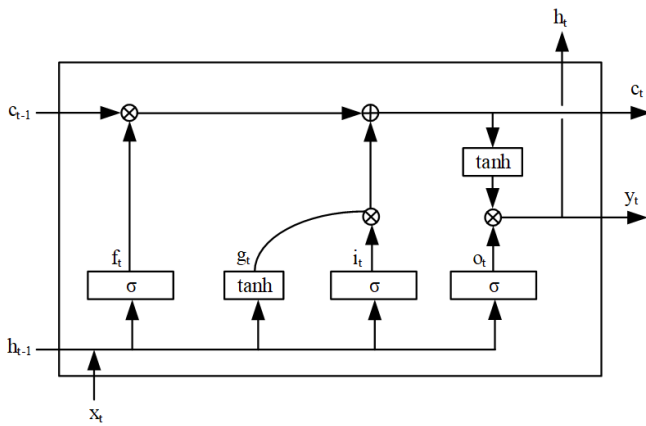Fig. 1. A standard RNN contains a tanh layer.

Fig. 2. A LSTM cell contains four interacting layers.

LSMT, shown in Fig. 2, also has a similar structure to RNN, but it has up to 4 layers interacting especially. LSTM can remove or add the needed information for the state of the cell, they are carefully adjusted by groups called gates. Gates are where filtering information passes through it, which is combined by a sigmoid layer and a multiplication. The sigmoid layer that will give the output is a number in the range [0, 1] and [0,1] describes how much information can be passed. If the output is 0, it means that no information is passed, and if that is 1, it means that all information goes through it. An LSTM consists of three such gates to maintain and regulate the state of the cell.

At step t, components in LSMT are calculated by

$$f_t = \sigma\big(W_i\,[h_{t-1}, x_t] + b_f\big) \tag{1}$$

$$g_t = tanh\big(W_g[h_{t-1}, x_t] + b_g\big) \tag{2}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{3}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{4}$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes g_t \tag{5}$$

$$h_t = o_t \otimes tanh(c_t), \tag{6}$$

where σ is the element-wise sigmoid function, $\otimes$ is the element-wise product, f_t is the forget gate, g_t is the output of the main layer, i_t is the input gate, o_t is the out gate, c_t is the cell vector, h_t is the hidden layer.

In neural architectures for NER, two LSTM networks facing each other to form a model called Bi-LSMT. In the Bi-LSTM model, the output at each step bases on the previous elements and the behind elements. For instance, to predict missing words in a sentence, it is necessary to consider both the previous parts and the next parts of the sentence.

B. Conditional Random Field Tagging

In the field of Natural Language Processing, most of the basic models are built on the Bag of Words method [20], [21]. But, these models cannot identify syntactic relationships between words. For example, with an emotional analysis model built on the Bag of Words method, we can not identify the difference in the sentence: "I like you", in which the verb "like" is a verb that indicates positive emotions. And similarly, with another sentence: "I am like you", the verb "like" is the similarity between two objects, not just affection. In this article, we will look at the CRF algorithm [22]–[24].

CRF is conditional probability algorithms, which are very useful for NER and POS labelling. CRF is a scalar graph model, which allows defining the probability distribution of the entire sequence of states with given conditions. In the CRF algorithm, the input is a set of attributes from the input data set according to a rule. The weight of the expression with the input attributes and the tags that have been tagged before will be used to predict the current tagged label.

The objective function in the algorithm determines the label for each word in the sentence. In CRF, the current word is formulated according to the labels of previous words. The model weights will be the most reasonable.

Fig. 3 demonstrates the architecture of Bi-LSTM-CRF for NER tagging. An input sentence is embedded into vectors, then these word embeddings are given to the forward and backward LSTM network. l_i represents the word i and its left context, r_i symbolizes the word i and its right context while c_i stands for the representation of the concatenating of two vector $\vec{l}$ and $\vec{r}$. In the next step, the CRF layer forecasts the named entity output sequence that best corresponds to the input sentence.

To tag NER to Vietnamese sentences, our research based on the system of Nguyen *et al* [6] that showed remarkable improvements as following table.
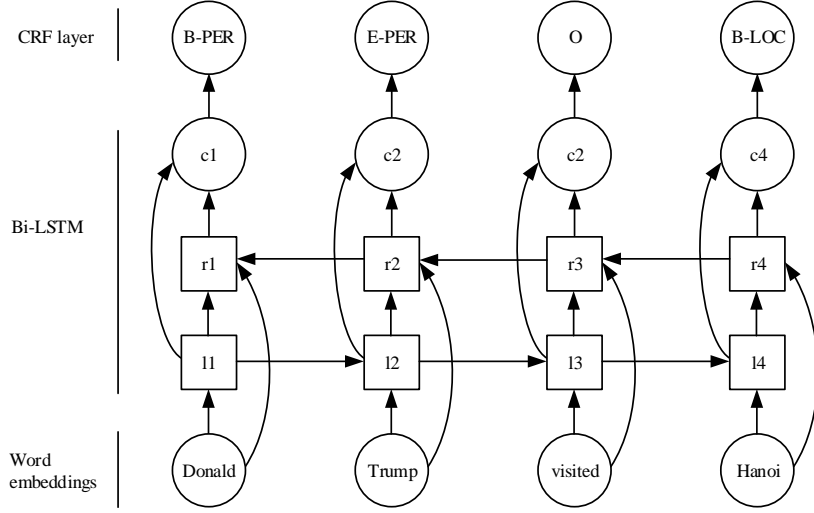
Fig. 3. Bi-LSTMs and CRF architecture.

TABLE I: NER TAGGING PERFORMANCE ON VIETNAMESE LANGUAGE AND SPEECH PROCESSING 2016 DATASET

Method	P	R	F1
Bi-LSTM-CRFs	90.97	87.52	89.21
Bi-LSTM-CRFs + POS	90.90	90.39	90.64
Bi-LSTM-CRFs + Chunk	95.24	92.16	93.67
Bi-LSTM-CRFs + POS + Chunk	**95.44**	**94.33**	**94.88**

III. NEURAL MACHINE TRANSLATION

NMT uses the neural network to train a statistical model for MT that can generate a target sentence depending on maximizing the conditional probability of a given source sentence. There are numerous studies have shown that NMT is more effective than previous MT methods [8], [13], [25], [26]. Our research-based on the encoder-decoder model with the attention that was introduced by Bahdanau [8], [27], [28].
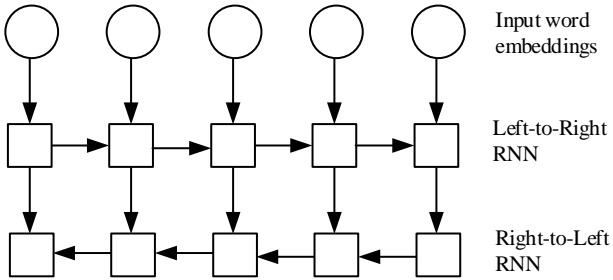

Fig. 4. The encoder of the NMT model.

A. Encoder

The encoder is a bidirectional RNN, including a forwarding RNN and a backward RNN. The input sentence is encoded into a matrix, then the basic language model will process this matrix with an RNN network as Fig. 4. The encoder states are the concatenation of the hidden state left-to-right RNN $\vec{h}$ and the hidden state right-to-left $\overleftarrow{h}$. Mathematically, the input sentence $x = (x_1, x_2, \dots, x_n)$, where n is sentence length, will be encoded by the following equation:

$$\vec{h_i} = f\left(\overrightarrow{h_{i-1}}, \overline{E}x_i\right) \tag{7}$$

$$\overleftarrow{h_i} = f\left(\overleftarrow{h_{i+1}}, \overline{E}x_i\right) \tag{8}$$

$$h_i = \left(\vec{h_i}, \overleftarrow{h_i}\right) \tag{9}$$

where f is a typical feed-forward neural network function such as tanh; $\overline{E}$ is the word embedding matrix for the source language.

B. Decoder

The decoder is also an RNN, it takes the representation of the input context, the previous hidden state and the output word prediction to generate a new hidden decoder state and a new output word prediction $y = (y_1, y_2, \dots, y_m)$, where m is the sentence length. The hidden state of the decoder is computed as:

$$s_i = (1 - z_i)°s_{i-1} + z_i°\tilde{s_i} \tag{10}$$

where

$$\tilde{s_i} = tanh(WEy_{i-1} + U[r_{i-1}] + Cc_i) \tag{11}$$

$$z_i = \sigma\left(W_z E_{y-1} + U_z s_{i-1} + C_z c_i\right) \tag{12}$$

$$r_i = \sigma\left(W_r E_{y-1} + U_r s_{i-1} + C_r c_i\right) \tag{13}$$

E_*, U_* and C_* are weight matrixes; E is the word embedding for the target language; c is the input context vector.
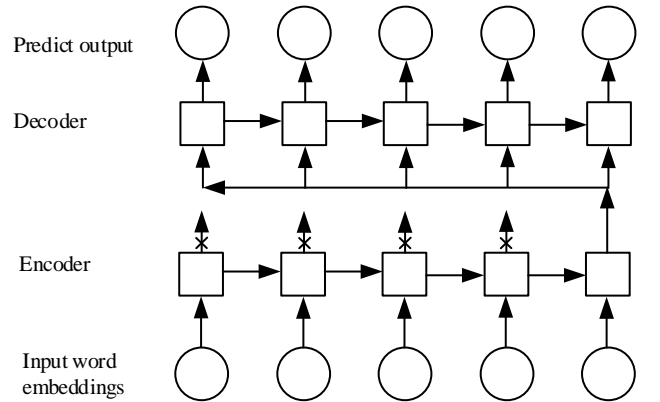

Fig. 5. The common sequence to sequence mechanism (without attention).

C. Attention

The source sentence is encoded into a fixed vector, as Fig. 5. Despite using LSTM to overcome the weaknesses of traditional RNN networks with Vanishing Gradient

phenomenon [29], [30], it makes the model difficult especially with long sentences. Hayashi *et al*. [31] introduced the attention mechanism applied in sequence-to-sequence model and they proved that attention is effective for MT. Bahdanau *et al*. [8] proposed a mechanism that allowed the model to focus on the important parts (associated words from source to target). Instead of just using the context layer created from the last layer of the encoder, they used all the outputs of each cell through each timestep and the hidden state of each cell to synthesize a context vector (attention vector), as Fig. 6.
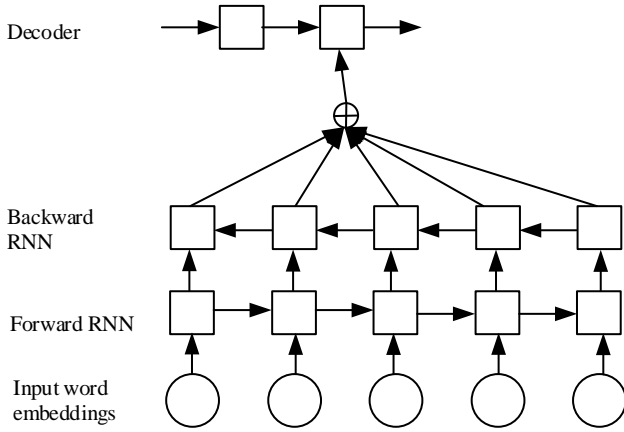

Fig. 6. The NMT model with attention.

IV. EXPERIMENTS AND RESULTS

A. Datasets

In this paper, we combined the dataset of Nguyen *et al*. [12] and the dataset of Luong *et al*. [13] into a parallel corpus. Then, we removed the sentences that are too short or the sentences that are extremely long in this dataset. The English-Vietnamese parallel corpus be using in our system is demonstrated as Table II. After combining two above datasets and removing some sentences, the dataset has over 900k sentence pairs. While the average length in English sentences is 7.9 words, that in Vietnamese sentences is 10.7 words. The token in Vietnamese sentence is also outstanding that of English sentence with approximately 9.8 million token and overly 7.2 million tokens, respectively.

TABLE II: ENGLISH-VIETNAMESE PARALLEL CORPUS

	Sentences	Average length	Token
English	917,395	7.9	7,234,159
Vietnamese		10.7	9,786,958

B. Implementations

We perform many preprocessing steps before putting them into the NMT system [32] including tagging NER for Vietnamese sentences. Our bi-direction English-Vietnamese NMT systems include:

Baseline: uses the original English sentences and the Vietnamese sentences with Word Segmentation (Word Seg.)

NER: uses the English texts and Vietnamese sentences after applying both Word Seg. and NER.

In Vietnamese texts, we used the research of Nguyen *et al*. [33] to address the problem of the unclear word boundaries in Vietnamese, then, we applied the research of Nguyen *et al*. [6]

to tag NER to the sentences. The sentence in Table III (Nhiên đang chơi ở công viên Cầu Giấy - *Nhien is playing in Cau Giay park*), in which the word "Nhien", and the word "Cau Giay" are transformed into "Nhien|B-PER" and "Cau_Giay|B-LOC", respectively.

TABLE III: FORM OF VIETNAMESE SENTENCE AFTER APPLYING NER

Form	Sentences
Original	Nhiên đang chơi ở công viên Cầu Giấy
Word Seg.	Nhiên đang chơi ở công_viên Cầu_Giấy
Word Seg.+NER	Nhiên|B-PER đang chơi ở công_viên Cầu_Giấy|B-LOC

In the above NMT systems, we used deep LSTMs with the 4 layers and with 1000 cells at each layer and 1000 dimensional word embedding. We used BLEU [34] and TER [35] scores to evaluate the performance of each NMT system.

C. Results

The results of our Korean-Vietnamese NMT systems have shown in Table IV. Generally, the NMT systems using NER for English in Vietnamese-to-English systems have better results than the NMT system in the reverse dimension. After using NER for Vietnamese sentence, English-to-Vietnamese MT system improved 1.34 BLEU points and 2.28 TER points and while the reverse dimension systems improved BLEU 1.13 points and 1.32 TER points.

TABLE IV: THE BLEU AND TER SCORES OF ENGLISH-VIETNAMESE NMT SYSTEMS

Systems		BLEU	TER
English-to-Vietnamese	Baseline	11.76	69.84
	NER	**13.10**	**67.56**
Vietnamese-to-English	Baseline	12.31	68.15
	NER	**13.44**	**66.83**

Depending on the context, a word may have a different meaning. In Vietnamese, various words are having different meanings based on their context. For instance: the word "*Huế*" can be the name of a person or it can be the name of a street in Hanoi. It can be tagged into "*Huế/B-PER*" or it also is transformed into "*Huế/B-LOC*". The toolkit of Nguyen *et al*.[6] was used for Vietnamese NER labelling, it may divide the word in text into Location, Person, Organization. After using NER, the Vietnamese sentence become clearer, thus increasing the quality of NMT systems.

In another example, the word "*nhà*" can mean the wife or the husband or the house depending on the context. When NER tags the word "nhà" to "*nhà/PER*" or "*nhà/LOC*", NMT will understand that these are two different words. After using NER, the translation sentence has higher accuracy than the translation of the systems without using NER. Therefore, NER has a positive effect on English-Vietnamese MT systems.

V. CONCLUSION

In this paper, we applied NER for Vietnamese sentences. Then, we built bi-directional English-Vietnamese NMT systems and evaluated the effectiveness of NER for NMT systems. Our results showed that NER has a positive effect on

NMT systems. In the future, we intend to build word sense disambiguation system for English and Vietnamese sentences to improve the quality of our NMT systems.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Van-Hai Vu, Quang-Phuoc Nguyen researched NER for Vietnamese language and collected data; Van-Hai Vu, Pum-Mo Ryu, Cheol-Young Ock performed the research on NMT; Van-Hai Vu, Cheol-Young Ock executed the experiments and evaluated the accuracy of the MT model; Van-Hai Vu, Quang-Phuoc Nguyen, Pum-Mo Ryu, Cheol-Young Ock wrote the paper; All authors had approved the final version.

ACKNOWLEDGEMENT

REFERENCES

[1] J. R. Finkel and C. D. Manning, "Joint parsing and named entity recognition," in *Proc. Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Boulder, Colorado, 2009, pp. 326–334.

[2] J. Li, A. Sun, J. Han, and C. Li, "A survey on deep learning for named entity recognition," *arXiv:1812.09449 [cs]*, Dec. 2018.

[3] V. Yadav and S. Bethard, "A survey on recent advances in named entity recognition from deep learning models," in *Proc. the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA, 2018, pp. 2145–2158.

[4] H.-P. Le, "Vietnamese named entity recognition using token regular expressions and bidirectional inference," *arXiv:1610.05652 [cs]*, Oct. 2016.

[5] N. Dong and K. A. Nguyen, *Attentive Neural Network for Named Entity Recognition in Vietnamese*, 2018.

[6] D. N. Anh, H. N. Kiem, and V. N. Van, "Neural sequence labeling for Vietnamese POS Tagging and NER," *arXiv:1811.03754 [cs]*, 08-Nov-2018.

[7] P. Koehn *et al.*, "Moses: Open source toolkit for statistical machine translation," in *Proc. the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, Prague, Czech Republic, 2007, pp. 177–180.

[8] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv:1409.0473 [cs, stat]*, Sep. 2014.

[9] M. Johnson *et al.*, "Google's multilingual neural machine translation system: Enabling zero-shot translation," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 339–351, 2017.

[10] E. Matusov, P. Wilken, and Y. Georgakopoulou, "Customizing neural machine translation for subtitling," in *Proc. the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, Florence, Italy, 2019, pp. 82–93.

[11] M. Graça, Y. Kim, J. Schamper, S. Khadivi, and H. Ney, "Generalizing back-translation in neural machine translation," in *Proc. the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, Florence, Italy, 2019, pp. 45–52.

[12] N. Q. Phuoc and C.-Y. Ock, "Building a bidirectional english-vietnamese statistical machine translation system by using moses," *IJCEE*, vol. 8, no. 2, pp. 161–168, 2016.

[13] M.-T. Luong and C. D. Manning, "Stanford neural machine translation systems for spoken language domains," in *Proc. the 12th International Workshop on Spoken Language Translation*, Da Nang, Vietnam, 2015, pp. 76–79.

[14] Q. H. Ngo and W. Winiwarter, "Building an english-vietnamese bilingual corpus for machine translation," in *Proc. 2012 International Conference on Asian Language Processing*, Hanoi, Vietnam, 2012, pp. 157–160.

[15] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," in *Proc. 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, California, 2016, pp. 260–270.

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[17] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," *arXiv:1303.5778 [cs]*, Mar. 2013.

[18] H. Mayer, F. Gomez, D. Wierstra, I. Nagy, A. Knoll, and J. Schmidhuber, "A system for robotic heart surgery that learns to tie knots using recurrent neural networks," in *Proc. 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, 2006, pp. 543–548.

[19] R. Jia and P. Liang, "Data recombination for neural semantic parsing," *arXiv:1606.03622 [cs]*, Jun. 2016.

[20] F.-F. Li and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 2005, vol. 2, pp. 524–531.

[21] G. Qiu, "Indexing chromatic and achromatic patterns for content-based colour image retrieval," *Pattern Recognition*, vol. 35, no. 8, pp. 1675–1686, Aug. 2002.

[22] L. Getoor and B. Taskar, "An introduction to conditional random fields for relational learning," *Introduction to Statistical Relational Learning*, The MIT Press, 2007.

[23] J. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," p. 10.

[24] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," *arXiv:1210.5644 [cs]*, Oct. 2012.

[25] Q.-P. Nguyen, A.-D. Vo, J.-C. Shin, P. Tran, and C.-Y. Ock, "Building a korean-vietnamese neural machine translation system with Korean morphological analysis and word sense disambiguation," *IEEE Access*, pp. 1–13, 2019.

[26] T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proc. the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, 2015, pp. 1412–1421.

[27] A. Vaswani *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.

[28] H. Bahuleyan, L. Mou, O. Vechtomova, and P. Poupart, "Variational attention for sequence-to-sequence models," in *Proc. the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA, 2018, pp. 1672–1682.

[29] M. Massaoudi, I. Chihi, L. Sidhom, M. Trabelsi, S. S. Refaat, and F. S. Oueslati, "A novel approach based deep RNN using hybrid NARX-LSTM model for solar power forecasting," p. 9.

[30] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, pp. 107–116, Apr. 1998.

[31] H. Kamigaito, K. Hayashi, T. Hirao, H. Takamura, M. Okumura, and M. Nagata, "Supervised attention for sequence-to-sequence constituency parsing," in *Proc. the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Taipei, Taiwan, 2017, pp. 7–12.

[32] G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. Rush, "OpenNMT: Open-source toolkit for neural machine translation," in *Proc. 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, 2017, pp. 67–72.

[33] D. Q. Nguyen, D. Q. Nguyen, T. Vu, M. Dras, and M. Johnson, "A fast and accurate vietnamese word segmenter," in *Proc. 7th International Conference on Language Resources and Evaluation*, Miyazaki, Japan, 2018, p. 6.

[34] K. Papineni, S. Roukos, T. Ward, and W.J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proc. 40th Annual Meeting on Association for Computational Linguistics*, Philadelphia, Pennsylvania, 2001, pp. 311–318.

[35] M. Snover, B. J. Dorr, R. I. Schwartz, and L. Micciulla, "A study of translation edit rate with targeted human annotation," in *Proc. Association for Machine Translation in the Americas*, pp. 223-231, 2006.

Van-Hai Vu received his B.S. from Hanoi University of Industry, Hanoi, Vietnam, in 2011 and his M.S. from VNU University of Engineering and Technology as a member of Vietnam National University, Hanoi, Vietnam, in 2014, both in information technology. His research interests include natural language processing, machine learning, and machine translation.

Currently, he is a Ph.D. candidate at the University of Ulsan, Ulsan, Republic of Korea.

Quang-Phuoc Nguyen received his B.S. from the University of Sciences, Vietnam National University, Ho Chi Minh City, Vietnam, in 2005, his M.S. from Konkuk University, Seoul, Republic of Korea, in 2010, and his Ph.D. from the University of Ulsan, Ulsan, Republic of Korea both in information technology. His research interests include natural language processing, machine learning, and machine translation.

Currently, he is an expert of FPT Korea, Seoul, Republic of Korea.

Pum-Mo Ryu received his B.S. from KyungPook National University, Republic of Korea, in 1995. Then, he received his M.S from POSTECH, Gyeongsangbuk, Republic of Korea, in 1997. And, he received his Ph.D. from KAIST, Daejeon, Republic of Korea, in 2009. His research interests include language processing & creative convergence, cyber security & police.

Currently, he is an associate professor at the Busan University of Foreign Studies, Busan, Republic of Korea.

Cheol-Young Ock received his B.S., M.Sc., and Ph.D. in information technology from Seoul National University, Seoul, Republic of Korea, in 1982, 1984, and 1993, respectively. He was a visiting professor at TOMSK Institute, Russia, in 1994 and at Glasgow University, Glasgow, UK, in 1996. His research interests include natural language processing, ontology, information retrieval, and machine

Currently, he is a professor at the University of Ulsan, Republic of Korea.

Prof. Ock is honorary doctor of the Department of IT Studies at Mongolia National University, Ulaanbaatar, Mongolia. He was the winner of the Hangul Development Merit Pavement, 2016.