The Educational Review, USA, 2022, 6(12), 869-874
https://www.hillpublisher.com/journals/er/
ISSN Online: 2575-7946
ISSN Print: 2575-7938

A Review of Machine Translation: Implications to Human Translators and Translation Teaching

Xiaoping Shen

Guilin Tourism University, Guilin, China, 541006.

How to cite this paper: Xiaoping Shen. (2022). A Review of Machine Translation: Implications to Human Translators and Translation Teaching. *The Educational Review, USA*, 6(12), 869-874.
DOI: 10.26855/er.2022.12.014

Received: November 22, 2022
Accepted: December 18, 2022
Published: January 5, 2023

Corresponding author: Xiaoping Shen, Guilin Tourism University, Guilin, China, 541006.

Abstract

Machine translation has witnessed great development in the recent decades and we have entered the era of neural machine translation (NMT). A review of MT is necessary for a better understanding of the relationship between MT and human translators and translation teaching in this era when MT has flourished. This paper first briefs the machine translation (MT) development in the past decades, focusing on the features, application, and drawbacks of each main paradigm of rule-based machine translation (RBMT), corpus-based translation (CBMT), and long-short term memory (LSTM), a main paradigm of NMT. It continues with a discussion of what MT means to human translators and translation teaching in universities. It concludes that MT should not and could not replace human translators which will always be vital in some fields and aspects; only a good integration between the two can ensure satisfying output with post-editing by human translators to meet the increasingly demanding market. This signifies that translation teaching in universities should embrace MT knowledge.

Keywords

Machine translation, human translator, translation teaching

1. Introduction

Machine translation has witnessed great development in the past decades and recently entered the era of neural machine translation. From rule-based machine translation to long-short term memory, a main paradigm of neural machine translation, each model of machine translation has displayed distinctive features, a mixture of cons and pros. A review of machine translation is necessary for a better understanding of the relationship between MT and human translators and translation teaching in this era when the demand for translation service is booming.

2. The development of MT: from RBMT to CBMT

The term "machine translation" (MT) generally refers to the automated translation. It means the process that uses computer software to translate a written or a spoken text from one natural language, into another, e.g., from English to Chinese, without the need or reliability of human intervention. MT is the wide application of computer and language sciences for practical needs.

MT has witnessed the pioneer researches in the 1950s and 1960s, and various approaches and systems have been put forward or/and designed to achieve automatic translation since 1970s. Along the way of MT development, two major approaches to machine translation systems have emerged: the rule-based machine translation (RBMT) and corpus-based translation (CBMT).The former was once the dominant framework of MT field. It was able to produce high quality translations with large-scale and fine-grained linguistic rules. However, constructing the system is very time-consuming and labor-intensive due to the fact that it heavily relies on the participation of translation experts. Moreover, it is of great difficulty to correct the input or add new rules to the system to generate a translation. This leads to a new dominance by the "corpus-based" methods since 1989 (Hutchins, 1995: 439).

In contrast to the RBMT, CBMT is data driven on the basis of the analysis of bilingual text corpora. To put it differently, rule-based approach falls into the domain of rationalism and the corpus-based belongs to empiricism (Okpor, 2014). Corpus based approach can be further classified into example-based machine translation approach (EBMT) and statistical machine translation (SMT), the two main approaches which emerged in the late 1980s. The essential steps of EBMT include phrase fragment matching, translation of segments, and recombination. EBMT is characterized by its use of a bilingual corpus as its main knowledge base, and can be viewed as an implementation of case-based reasoning approach of machine learning. Statistical machine translation (SMT), as its name implies, uses a mathematical model in which the process of human language translation is statistically modeled. SMT also produces translations using statistical methods on the basis of bilingual text corpora.

Statistical models used in SMT include word-based translation (Brown et al., 1993), phrase-based translation (Och & Ney, 2004), syntax-based translation, and hierarchical phrase-based translation (Chiang, 2007). For SMT model, translation is created by stringing together the translations of clearly identified subsegments. It takes the translation of natural language as a machine learning problem. By examining many samples of human-produced translation from the corpora, SMT algorithms automatically learn how to translate. This means building quality bilingual text corpora is crucial to the success of SMT. In other words, the translation quality relies on the number of human-translated documents available in the given languages: the more of them, the better the quality. Suraiya et al. (2018), for example, point out that, for a SMT engine to be trained, a minimum of 2 million words for a specific domain and even more for general language are required in order. Such a heavy dependence on huge amounts of parallel texts is one of the SMT's biggest downfall. Besides, the reordering problem still remains a challenge in statistical machine translations (Cui et al., 2016). In addition, the statistics-based translation model shows inferior performance in deep semantic understanding. The dominant position of this corpus-based predecessor was challenged as a breakthrough of deep learning, the Neural Machine Translation (NMT) (Sutskever et al., 2014), has emerged as a new paradigm.

3. Neural machine translation

3.1 NMT and RNN

NMT has quickly replaced SMT as the mainstream approach, although it is similar to SMT in the sense that also a corpus-based machine translation. What distinguishes NMT from SMT is that the computational approaches of NMT are artificial neural networks. The neural networks are composed of thousands of artificial units, i.e., an artificial intelligence, which resemble neurons in that they are able to learn and grow by means of data. Their output or activation (that is, the degree to which they are excited or inhibited) depends on the stimuli they receive from other neurons and the strength of the connections along which these stimuli are passed (Forcada, 2017). Another difference between SMT and NMT concerns the representations. SMT uses symbolic representations of words, while NMT uses distributed representations, also defined as word embedding, in which words are encoded as vectors in a multidimensional space where similar words are semantically close to each other.

Despite being relatively new, NMT has attracted a lot of interest among researchers (Sutskever et al., 2014; Wu et al., 2016) and already showed promising results and achieved state-of-the-art performances for various language pairs (Luong et al., 2015a; Jean et al., 2015). NMT is argued to create much more accurate translations than SMT (Doherty et al., 2010; Dorr, 1999 et al., 1999; Liu et al., 2014). The strength of NMT lies in its ability to learn directly, in an end-to-end fashion, the mapping from input text to associated output text. Its architecture typically consists of two recurrent neural networks (RNNs), one to consume the input text sequence and the other to generate translated output text (Wu et al., 2016). RNNs can be regarded as multiple copies of the same network, each passing a message to a successor. The simple RNNs consisting of standard recurrent cells are not flawless. The most noticeable downfall of such RNNs is that they are not capable of handling long-term dependencies. As claimed in Hochreiter (1991), the fundamental reason for this long-term dependency problem is that error signals flowing backward in time tend to either blow up or vanish. For instance, RNNs are able to predict the next word 'sky' for "the Sun is in the sky" but may not be able to predict the expected word 'Chinese' in a longer sentence like: "I grew up in China... I speak good Chinese" where more context is needed.

3.2 LSTM

In order to handle the exploding/vanishing gradient problem and the long-term dependencies problem that plague RNN, long short-term memory (LSTM), an RNN architecture specifically designed to address the vanishing gradient problem, is proposed (Hochreiter, 1991; Hochreiter & Schmidhuber, 1997). LSTMs are effective at capturing long-term temporal dependencies without suffering from the optimization hurdles, they thus have been used to advance the state of the art for many difficult problems.

The most remarkable development of this earlier LSTM from the even earlier RNN is that it divides the state into two states and differentiates 'cell state' c_t from the original hidden state h_t. C_t is intended to store long-terminformation, while h_t is used as representation of short-term information. The special units called memory blocks in LSTM (Hochreiter & Schmidhuber, 1997) contain memory cells with self-connections storing the temporal state of the network in addition to special multiplicative units called gates to control the flow of information. Each memory block in the original architecture contained an input gate and an output gate. The former controls the flow of input activations into the memory cell. The latter controls the output flow of cell activations into the rest of the network (Sak et al., 2014). That is why LSTM can solve numerous tasks not solvable by previous learning algorithms for RNNs (Ger et al., 2000). As a result, LSTMs have emerged as an effective and scalable model for several learning problems, as they are capable of overcoming long-term dependencies and handle the exploding/vanishing gradient problem of simple RNN (Greff et al., 2015).

Nevertheless, Ger et al. (2000) identify a weakness of Hochreiter & Schmidhuber (1997)'s LSTM networks processing continual input streams that are not segmented into subsequences with explicitly marked ends at which the network's internal state could be reset. Accordingly, Ger et al. (2000) put forward a remedy, i.e., adding an adaptive "forget gate" to the memory block to allow the network to reset its state. In the remainder of this paper, LSTM refers to the LSTM model with the forget gate, since this is the most popular LSTM architecture.

3.3 Classification and application of LSTM

LSTM networks can be divided into two broad categories: LSTM-dominated networks and integrated LSTM networks (Ruan & Ma, 2020; Yu et al., 2019).The former focus on optimizing the connections between inner LSTM cells, while the latter mainly pay attention to integrating the advantageous features of different components (Yu et al., 2019). After comparing the vanilla LSTM in their term, with eight different variants with experimental study, Greff et al. (2015) conclude that the former performs reasonably well on various datasets and using any of eight possible modifications does not significantly improve the LSTM performance. Moreover, the forget gate and the output transfer function are the most critical components of the LSTM block, whereas the learning rate is the most important hyperparameter in the backpropagation algorithm. Schmidhuber et al. (2007) proposed that LSTM was sometimes better trained by evolutionary algorithms combined with other techniques rather than by pure gradient descent.

There are few remarkable variants that significantly outperform standard LSTM's accuracy, even if the training would be slow due to complexity (Greff et al., 2015). LSTM has transformed both machine learning and neurocomputing fields (Houdt et al., 2020). Especially, it is well suited to handle time series predictions, but also any other problem that requires temporal memory.

The recent applications on LSTM reported in the literature empirically illustrate that this recurrent system is well capable of handling a wide variety of problems including language modeling (Zaremba et al., 2014) and translation, handwriting recognition (Graves et al., 2005; Pham et al., 2013) acoustic modeling of speech (Sak et al., 2014), video data (Donahue et al., 2014), time series forecasting, text sentiment classification (Wang et al., 2018).

LSTM is also employed in other translation related fields like morphological segmentation in (Wang et al., 2016), relation extraction (Song et al., 2018), mapping natural text to knowledge base entities (Kartsaklis et al., 2018), and translation tasks in Sutskever et al. (2014). Facebook, for instance, reaches over 4 billion LSTM-based translations perday as of 2017 (Pino et al., 2017). Google is also using LSTM. That's different from Google's previous translation method, phrase-based machine translation, which breaks sentences into individual words and phrases. The new method looks at the entire collection of words.

4. MT: implications for human translators and translation teaching

4.1 The evaluation of MT

The translation industry is steadily incorporating MT in their workflows engaging the human translator to post-edit the raw MT output in order to comply with a set of quality criteria in as few edits as possible.

The quality of MT systems is generally measured by automatic metrics, producing scores that should correlate with human evaluation. The generally well-known translation evaluation methods include Bilingual Evaluation Understudy and Translation Error Rate, generally abbreviated as BLEU and TER respectively. The former refers to an algorithm for evaluating the quality of machine translated text, helping to accelerate the MT research and development cycle. The closer a machine translation output is to a professional human translation, the better it is. Compared with human evaluation approaches, it is much more cost-effective and time-effective. Thus, it has become the standard in evaluating machine translation output. Departing from this, TER is a method used by MT specialists to determine the amount of post-editing required for machine translation jobs. TER score measures the amount of editing required for a machine translation output to be edited in line with a reference translation. Its aim is to find the minimum number of edits without generating a new reference. The score ranges from 0-1, where 1 means more post-editing effort. It's quick to use, language independent and corresponds with post-editing effort.

As well-known, translation is a considerably challenging task for humans. This is also true for computers, as what they deal with are highly complex natural languages. In other words, computers, like humans, cannot generate perfect translation. The prime objective of MT in an internet-dominated environment has been the rapid development of translation systems that generate high-quality translations. MT output, in practice, still have drawbacks. For instance, Koehn & Knowles (2017) examine a number of challenges to NMT and give empirical results on how well the NMT technology, compared to traditional SMT, currently holds up. They find that NMT systems have lower quality out of domain, to the point that they completely sacrifice adequacy for the sake of fluency; NMT systems have a steeper learning curve with respect to the amount of training data, resulting in worse quality in low-resource languages, but better performance in high resource languages. Besides, NMT systems that operate at the sub-word level (e.g., with byte-pair encoding) perform better than SMT systems on extremely low frequency words, but still show weakness in translating low-frequency words belonging to highly-inflected categories (e.g., verbs) (Koehn & Knowles, 2017).

The above can be attributed to the fact that a good translation for natural languages entails a solid bilingual language competence, a thorough interpretation of the source text and its intended purposes to the target readers, and good knowledge of the target language, all processes of which need to take into account not only language aspects, but also social and cultural factors. Consider, for example, natural languages are complex. To be specific, there are no clear boundaries between word senses in some languages like Chinese and English, and the overall meaning of a sentence or of a text is based on these ambiguous words. Secondly, word senses do not directly correspond across different languages. For instance, the same concept can be expressed by a single word or by a group of words, depending on the context and language concerned. At the same time, it is impossible and unrealistic to manually specify all the information that would be necessary for an automatic machine translation system. In addition, in many cases, a word in the source text doesn't have a counterpart in the target language. This also accounts for the fact that the translation task has remained highly challenging and computationally expensive up to the present time.

4.2 MT and human translators

The practical value of MT in areas of politics, economics, culture, etc., is of great significance in this era of economic globalization and internet. Compared with human translator, MT excels in data processing, recording, and translation speed. Moreover, it is expected to get mature technically. The fast development of NMT has caused worries and anxieties among some human translators to whom the MT is a challenge or threat to some extent. Wang (2016, cite from Chen, 2020), for example, claims that MT is quite expected to take the place of human translator or interpreter.

Other scholars (Chen, 2020; Zhu, 2018), however, argue that MT will never reach the quality of a professional human translator. According to Chen (2020), the lack of subjectivity is a noticeable flaw of MT. For him, what MT can do is no more than technically decoding and coding in line with the prearranged procedures by human. This means MT could only statically 'understanding' words, i.e., only get the static dictionary meaning of words, which is quite different from the comprehension by human translators, the translation subject who interpret the text in combination with the context in a deep and comprehensive way, not alone the review and pondering between lines, and selection of words by the human translator. Similarly, Zhu (2018) contends that it is a false proposition that human translators are to be substituted by MT. He holds that the purpose of MT is to serve instead of replacing human beings. He further illustrates with examples that MT cannot perform as well as human interpreters in that they could not convey emotion as human translators do as the latter can adjust their tone and intonation accordingly. MT is especially dwarfed in the literature translation area, the most-demanding and productive area where fuzziness, imagination, virtuality, metaphor, cultures and polysemy dominate. Therefore, the fear of being replaced by MT is unnecessary and the human factor will always be vital in translation (Poibeau, 2017).

The above demonstrates that MT and human translators are not a "either A or B" relation and integration of the two, i.e., technology and humanity, is necessary as demand for translation service is booming due to more and more frequent intercultural communication. Human translators should make the best use of MT for the sake of saving time and energy. For example, MT can offer many candidates for a given word or expression so that human translators can select the most suitable ones. At the same time, bear it in mind that the improvements of neural machine translation that base its strength on machine learning is far from perfect, although there is already the high-quality output of neural MT tools. MT is not different from human translation in the sense that both normally require revision by a second translator before dissemination. Thus, post-editing is crucial. In other words, MT output might serve as a rough draft or a pre-translation for human translators who are empathy, emotional, intelligent, and creative and thus naturally can be more profiled in activities where the human brain can excel to produce texts that satisfy the linguistic and culture norms of the target readers by making adaptation to the assumed knowledge of its readers.

4.3 Implication to Translation teaching:

MT has been embraced by the professional and educational environments worldwide. The situation of MTI in mainland China is, however, not so optimistic in terms of MT teaching, when compared with universities of Hongkong SAR, Macao SAR, or some developed countries (Zhu, 2018). According to a survey by Wang (2016, cite from Zhu, 2018), 11 universities out of 43, most of which are first tier universities in Chinese mainland, have not offered any course of MT, with a high percentage of 25, let alone the second tier universities, the majority of mainland China universities. Furthermore, for those offering these courses, the teaching hours is only two per week and 34 hours altogether during their 4-year studies, which is far from enough, consider MT is involved in a wide-range of knowledge. Thus, enough teaching hours should be guaranteed so that translation students can get familiar with and acquire MT knowledge including text and picture processing, searching tools, database, etc. Post-editing is another crucial element for translation teaching as they are in increasingly great demand in translation market, which is in accordance with the growing need of integration between human-MT mentioned earlier. In order to improve students' post-editing ability, teachers should help them to develop a critical view of common merits and drawbacks in the outputs of different MT paradigms. To this end, students can be trained with case studies, identifying the MT flaws, classifying them into such different categories as semantic, lexical, cultural, to name a few. Following this, students should work on improving MT output by replacing, restructuring, omitting, reorganizing, adding, and so on. All these are inseparable from a good mastery of translation theories on their own and a good cooperation and communication with their classmates now and workmates in the future. Besides, how to incorporate discrete, complicating, and rich human knowledge into (N)MT is also an important problem that needs further exploration for translation teaching.

5. Conclusion

MT has experienced great development and exhibited promising results in language translation and related fields. A review of MT is helpful to gain a deeper and better understanding of the relationship between MT and human translators as well as translation teaching. Our discussion shows that MT will never reach the quality of and take the place of the professional human translators. Thus, the integration between the two is inevitable. This indicates that translation teaching in universities should not only equip students with a solid basis of translation theory but also make them have enough access to MT knowledge like artificial intelligence knowledge, computer-aided translation tools, word processing, database, corpus, online translation applications, etc., via translation practice and enough case studies. All these, however, pose a challenge to universities where lack of quailed faculty and fund arises.

References

Brown, P.F., Pietra, V. D. J., Pietra, S. A. D., & Mercer, R. (1993). The Mathematics of Statistical Machine Translation: Parameter Estimation. Computational Linguistics 19(2):263-311.

Chen W. (2020). The deconstruct of subjunctivity by machine translation--the foothold of machine translation. Foreign Languages Research 180(2):76-83.

Chiang, D. Hierarchical Phrase-Based Translation. 2007. Computational Linguistics (2007), 33 (2): 201-228.

Cui, Y.M., Wang, S.J. & Li, J.F. (2016). LSTM Neural Reordering Feature for Statistical Machine Translation). Conference: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: 977-982.

Doherty, S., O'Brien, S., & Carl, M. (2010). Eye tracking as an MT evaluation technique. Transl. (24): 1-13.

Dorr, B. J., Jordan, P. W., & Benoit, J. W. (1999). A survey of current paradigms in machine translation, Adv. Computer, (49):1-68.

Forcada, M. (2017). Making sense of neural machine translation. Translation Spaces 6(2): 291-309.

Graves, A. & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional lstm and other neural network architectures. Neural Networks 18(5):602-610.

Greff, K., Srivastava, R.K., Koutnik, J., Steunebrink, B.R., & Schmidhuber, J. (2015). LSTM: A Search Space Odyssey. IEEE Transactions on Neural Networks and Learning Systems 28(10): 2222-2232.

Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen Netzen. Master Thesis, Technische Universitat¨ Munchen, Munchen.

Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural computation (9):1735-1780.

Hochreiter, S., Schmidhuber, J. (1997). Lstm can solve hard long time lag problems. In: M.C. Mozer, M.I. Jordan (eds.), Advances in Neural Information Processing Systems (9): 473-479. MIT Press.

Houdt, G.V., Mosquera, C., Napoles, G. (2020). A Review on the Long Short-Term Memory Model Cognitive Science & AI (53): 5929-5955.

Hutchins, W. J. (1995). Machine Translation: A Brief History. Computer Science: 431-445.

Jean, S., Cho, K., Roland Memisevic, & Bengio, Y. (2015). On using very large target vocabulary for neural machine translation. In ACL.[1412.2007] On Using Very Large Target Vocabulary for Neural Machine Translation (arxiv.org)

Kartsaklis, D., Pilehvar, M.T., & Collier, N. (2018). Mapping text to knowledge graphentities using multi-sense lstms. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: 1959-1970.

Koehn P. K. & KNowles, R. (2017). Six Challenges for Neural Machine Translation. Proceedings of the First Workshop on Neural Machine Translation: 28-39, Vancouver, BC, Canada.

Liu, S., Yang, N., Li, M., & Zhou, M. (2014). A recursive recurrent neural network for statistical machine translation. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics:1491-1500. Association for Computational Linguistics, Baltimore, MD, USA.

Luong, C. T., Sutskever, I., Le, Q. V., Vinyals, O. V., & Zaremba, W. 2015. Addressing the Rare Word Problem in Neural Machine Translation. arXiv preprint arXiv:1410. 8206, URL http://arxiv.org/abs/1410. 8206.

Och, F. J. & Ney, Hermann. (2004). The Alignment Template Approach to Statistical Machine Translation. Computational Linguistics (4): 417-449.

Okpor, M. D. (2014). Machine Translation Approaches: Issues and Challenges. Computer science: 159-165.

Pham, V., Bluche, T., Kermorvant, C., & Louradour, J. (2013). Dropout improves Recurrent Neural Networks for Handwriting Recognition. IEEE Transactions on Signal Processing 45(11): 2673 -2681.

Poibeau, T. (2017). Machine Translation. London: MIT press.

Ruan, S-m. & Ma, Y. (2020). Real-Time Energy Management Strategy Based on Driver-Action-Impact MPC for Series Hybrid Electric Vehicles. Complexity. https://doi.org/10.1155/2020/8843168.

Sak, H., Senior, A. & Beaufays, F. (2014). Long Short-Term Memory Recurrent Neural Network Architectures for Large ScaleAcoustic Modeling. Proceedings of the Annual Conference of International Speech Communication Association (INTERSPEECH): 338-342.

Schmidhuber, J., Wierstra, H., Gagliolo, J., & Gomez, D. (2007). A new class of learning algorithms for supervised recurrent neural networks (RNNs). Neural Computation 19(3):757-79.

Song, L., Zhang, Y., Wang, Z., & Gildea, D. (2018). Nary relation extraction usinggraph-state lstm. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: 2226-2235.

Suraiya, J., Samak, S., & Sokphyrum, Kim. (2018). How to Translate from English to Khmer using Moses. International Journal of Engineer Inventions (3): 71-81.

Sutskever, I., Vinyals, O., & Le, Q. (2014). Sequence to sequence learning with neural networks. In: Advances in neural information processing systems.

Wang, L., Cao, Z., Xia, Y., & De Melo, G. (2016). Morphological segmentation with window lstm neural networks. In: Thirtieth AAAI Conference on Artificial Intelligence.

Wang,T. W., Liu, X. L., & Long, W. (2018). An LSTM Approach to Short Text Sentiment Classification with Word Embeddings Jenq-Haur. The 2018 Conference on Computational Linguistics and Speech Processing: 214-223. The Association for Computational Linguistics and Chinese Language Processing.

Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. Computation and Language, 1-23.

Yu, Y., Si, X., Hu, C., Zhang, J. (2019). A review of recurrent neural networks: Lstm cells and network architectures. Neural computation 31(7): 1235-1270.

Zaremba, W., Sutskever, I. & Vinyals, O. (2014). Recurrent Neural Network Regularization. arXiv:1409.2329 [cs], September URL http://arxiv.org/abs/1409. 2329.

Zhu. (2018). The Machine to Replace Human Beings as a Translaor? A Discussion of the Relationship Between Science and Humanity in the Cultivation of Translators and Interpreters. *Foreign Language and Literature* (bimonthly) 34(3):101-109.