VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wānanga o te Ūpoko o te Ika a Māui

School of Engineering and Computer Science
Te Kura Mātai Pūkaha, Pūrorohiko

PO Box 600
Wellington
New Zealand

Tel: +64 4 463 5341
Fax: +64 4 463 5045
Internet: office@ecs.vuw.ac.nz

Literature Review

Bryony Gatehouse

Submitted in partial fulfilment of the requirements
for AIML420

Abstract

Critically examining the abilities of Machine Translation, comparing the performance of the two most popular, and discussing future development.

Contents

Introduction

Machine Translation is a popular area of research as it expanded the information available to people, and reduces the workload for human translators. It is widely used on social media sites, including Facebook [1] However, there is still concern about the effectiveness of machine translation.

The two main types of Machine Translation are Statistical Machine Translation (SMT) and Neural Machine Translation (NMT). SMT uses the probability of a word to complete the translation while NMT uses Neural Networks to predict the best fitting sentence.

This review will explore and analyse the abilities of SMT and NMT. It will also focus on the evaluation of an MT system, as this is required to understand the true abilities of SMT and NMT. The two types will be considered then compared.

Background

Machine Translation is a type of Artificial Intelligence that can translate between languages. Many examples can be found online, including Google Translate, and increase the amount of information available to people.

Text Mining

To translate between languages, an MT system must know a decent amount about the two languages. This can be achieved by training the system on various texts of each language so they learn words and their position in sentences. They also need to be able to learn from bilingual texts: texts which have versions in both languages.

Text Mining is the process of extracting important information from the text. In this case, a word's meaning can be concluded from its surroundings and use. If the word "rose" is used as a verb it means the past tense of rise while if it is used as a noun it represents the red flower.

The recent growth in Machine Translation is partly due to the sudden expansion of information available on the internet. This provides a wealth of knowledge that can be used to train an MT system, particularly the websites which offer the user multiple languages to view the page in [2]. The internet also offers a variety of texts, from fantasy fiction to government documents, providing the system with many different forms of literature to learn.

Evaluation

Attempting to measure the accuracy of a Machine Translation is difficult as there are many version of a sentence which could be acceptable.

The traditional method is to use human judges who can confirm whether the translation has the same meaning as the original sentence, and how fluent the translation is in the target language. However, this process is slow and requires paying translators to judge [2].

BLEU

BLEU is an automatic machine translation evaluation that is widely used to test an MT system's efficiency [3]. Using a database of reference human translations, it compares the MT translation to one or more human translation. There can be many correct translations of a source sentence with differences in word order or word choice.

A candidate translation is compared to many different reference translations with the intent of finding the sections that match. The BLEU metric is based on the number of words that are in the matching sections vs the total number of words. Using this method, the BLEU evaluation considers the adequacy and fluency of the candidate translation. It also considers the length of the candidate translation, rewarding sentences which have the same length as a reference translation.

The resulting evaluation method isn't as accurate as a human judge but offers an automatic evaluation that produces a quick, inexpensive result.

Summary

Statistical Machine Translation

SMT was originally proposed in 1949 to translate using statistical methods and ideas from information technology [4]. They use monolingual models of the target language to find the most probable candidate translation that makes sense in the target language. SMT only started to dominate Machine Translation research in the late 1980s [5] when computer technology reached a point where it could perform the calculations required [4]. The growth of SMT was also due to the increased need for information to be available in many languages, particularly in the US government [2].

The IBM Model

The IBM Model was first described in 1990 [4] as a statistical approach to machine translation. Given the idea that a sentence in one language (T) is a possible translation of a sentence in another language (S), the probability that the translator will produce T when given S can choose the closest translation known. The IBM Model splits this probability into separate words: the probability of a word given the current sentence.

Rather than considering the probability of the next word given every single word already in the sentence, the history could be compared as a class. One method is the Trigram Model which considers two histories as equivalent if their last two words are identical. This method was tested on English sentences with less than 11 words each. The sentences were split into separate words then recovered using the Trigram Model. 63% of recovered sentences matched the original sentence, 21% had similar meanings to the original sentence, and 15% were just garbage.

As well as the probability of the word being translated into the sentence, the IBM also considers the number of words it could translate to, known as the fertility. When translating the sentence: "John does beat the dog", the word "does" is ignored in the translation, thus having a fertility of 0.

The IBM Model was tested on translating French to English. However, only 5% of sentences were decoded identically to the actual translation, but 48% were decoded to produce a legitimate translation of the French sentence.

Synchronous Context-Free Grammar Models

Context-free grammar [2] considers some of the linguistic representations of syntax which the IBM approach doesn't. Also, it can complete long-distance reordering without the exponential cost of permutations. However, there are other problems with this model, and it still an area of much active research.

SCFG uses a tree structure of non-terminal nodes representing the syntactic categories and terminal nodes representing the words. By rotating the non-terminal nodes a tree can be transformed into another, thus changing the syntactic rules of the tree/sentence. While the number of trees that can generate a string could be very large, the recursive sharing of sub-trees mean that the SCFG model can use dynamic programming algorithms to reduce the computation expense.

Neural Machine Translation

Neural Machine Translation is a new approach to Machine Translation [6]. Rather than having many small sub-components like the decoder component of the SMT, Neural Machine Translation uses a single, large neural network.

An encoder-decoder model has two neural networks for each language, an encoder neural network that encodes the sentence into a fixed-length vector and a decoder neural network that outputs a translation of the encoded vector. However, it requires all information about the sentence to be condensed into the fixed-size vector, reducing its effectiveness on longer sentences.

A different approach uses a bidirectional Recurrent Neural Network as the encoder which obtains an annotation for each word that contains summaries of both the preceding and following words. This model was compared against the original model in a test where each model was trained twice: once with sentences up to 30 words and then with sentences up to 50 words. The new approach was able to outperform the old approach and maintained its effectiveness with longer sentences or 50 words or more.

Statistical vs Neural Machine Translation

Neural Machine Translation is considered better than Statistical Machine Translation now and is widely used, including for the majority of languages for Google Translate [7] [8].

Translating a Novel

A study compared the results from translating part of a novel from English to Catalan of a Statistical Machine Translator and a Neural Machine Translator [9]. It has been found that novels tend to express greater cohesion than the news, referring to knowledge beyond the sentence level. The study was based on post-editing: how long/many keystrokes it took to fix the translated text to a readable format. To remove translator bias or differences, they were chosen randomly and received random sections of text from a random model to process.

As expected the translators processed the translation from scratch slower, with the NMT providing text that was quickest to edit. However, with longer sentences, the advantage of NMT compared to from scratch reduced to almost nothing. This did not happen with the text from the SMT. The NMT required the least keystrokes to produce the final translation while, for half of the translators, the SMT text required more keystrokes than writing the translation from scratch.

This study concluded that the NMT produced more coherent sentences, reducing the amount of editing required afterwards, but its longer sentences tended to reduce in quality. It also proved that SMT was able to produce average quality sentences regardless of length, shown by the fact that the ratio of time-to-sentence-length was similar to translating from scratch, though with lower values.

Translation Errors

Considering the difficulty of Machine Translators to choose the best word for the sentence, a study compared the performance of an NMT and an SMT system [10]. The results from translating from both English to Hindi and Hindi to English were evaluated using three different metrics: BLEU, METEOR, and TER. Going from English, a morphologically poor language, to Hindi, a morphologically rich language, proved more difficult than the other way round as there were more translation choices for each English word. However, in both language orientations and all metric types, the NMT received higher marks than the SMT.

Thus, in terms of correctness, the NMT produces more sentences that keep the original meanings considering the format of the entire text while SMT tended to choose more literal translations.

Conclusion

Challenges and Future Directions

Neural Machine Translation is a new technology and is very popular. However, it is still lacking compared to human translation. The study comparing performance when translating a novel [9] proved that the NMT performance dropped with longer sentences. Thus, the focus of future work in this area should be increasing the capabilities for longer sentences before improving the overall performance.

The Statistical Machine Translation produced lower results in every category - language combination and evaluation metric - in the study on translation errors [10] than the NMT. However, post-processing of longer sentences translated by the SMT took less time than the NMT, proving it is still more effective in some areas. Future work on SMT should be either focused on improving the performance of the SMT algorithms or combining the SMT with an NMT to increase performance on long sentences.

Finally, the largest challenge in Machine Translation is still the evaluation of a system's performance. The BLEU method requires a database of reference human translations and is unable to keep up with the accuracy of human translators. However, it is faster and cheaper than human translators so more widely used. Future research should focus on increasing the capabilities of an evaluation system.

Conclusion

Though the abilities of Machine Translation are still lacking compared to human translators, they can produce readable translations with typically little meaning lost. This provides people with the ability to read anything on the internet, regardless of language.

However, this can come with consequences. Facebook's translator wrongly translated ″good morning″ in Arabic to ″hurt them″ in Hebrew, resulting in a man being arrested and almost prosecuted [11]. Thus, the use of translators should be still regarded with suspicion.

Bibliography

[1] A. Sidorov, "Transitioning entirely to neural machine translation," *AI RESEARCH, ML APPLICATIONS*, Aug. 2017. [Online]. Available: https://www.facebook.com/help/364760868183100/app-translate-facebook

[2] A. Lopez, "Statistical machine translation," *ACM Computing Surveys*, vol. 40, no. 3, pp. 1–49, aug 2008. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/1380584.1380586

[3] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002, pp. 311–318. [Online]. Available: https://www.aclweb.org/anthology/P02-1040.pdf

[4] P. F. Brown, J. Cocke, S. A. Della Pietra, V. J. Della Pietra, F. Jelinek, J. Lafferty, R. L. Mercer, and P. S. Roossin, "A statistical approach to machine translation," *Computational linguistics*, vol. 16, no. 2, pp. 79–85, 1990. [Online]. Available: https://www.ling.upenn.edu/courses/Spring_2011/cogs502/Brown1990.pdf

[5] D. Kenny, "The ethics of machine translation," *Proceedings of the XI NZSTI National Conference*, 2011. [Online]. Available: https://core.ac.uk/download/pdf/11311284.pdf

[6] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *3rd International Conference on Learning Representations, ICLR 2015*, 2014. [Online]. Available: https://arxiv.org/pdf/1409.0473.pdf

[7] B. Turovsky, "Found in translation: more accurate, fluent sentences in google translate," *The Keyword, Google*, Nov. 2016. [Online]. Available: https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/

[8] "See which features work with each language." *Google Translate Languages*, 2021. [Online]. Available: https://translate.google.com/intl/en/about/languages/

[9] A. Toral, M. Wieling, and A. Way, "Post-editing effort of a novel with statistical and neural machine translation," *Frontiers in Digital Humanities*, vol. 5, p. 9, 2018. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fdigh.2018.00009/full

[10] R. Haque, M. Hasanuzzaman, and A. Way, "Analysing terminology translation errors in statistical and neural machine translation," *Machine Translation*, vol. 34, no. 2-3, pp. 149–195, aug 2020. [Online]. Available: https://link.springer.com/article/10.1007/s10590-020-09251-z

[11] "Israel facebook: Mistranslated post creates security alert," *BBC News*, Oct 2017. [Online]. Available: https://www.bbc.com/news/world-middle-east-41714152