Machine Translation

Machine translation has become one of the biggest challenges for the computer from the beginning. It is a cumbersome task for machines to automatically translate the human language. Machine Translation is known as the application of computers to translate natural languages without human help. It is a process in which the machine does translation challenges by converting the input text in one language(such as English) to output text in another language (such as Hindi). The aim of this type of translation is to remove the language barrier. In earlier times, we needed a translator to be present while an important business meeting was being done. Now, with the help of computers we can do business, seminars and important meetings without a human translator. Meanwhile, with the help of this type of translation now we can study and read literature from different nations using the web services. It was found that people are able to speak, read and understand more languages than earlier and all credit goes to computer translation. Hence, machine translation has become the most vital tool for an individual to understand many languages or text and speech which needs to be decoded.

There are various different ways through which the translation is being done by the machine. These are mainly done with the help of Rule-based machine translation (RBMT), Statistical-based machine translation (SBMT) and Neural-based machine translation (NBMT). The oldest way of machine translation is rule-base MT. It is a path breaker in machine translation. RBMT has been used for a long time till other methods have not been introduced. It is mostly based on numerous in-built rules and countless dictionaries for every language. The computer analyzes text and builds a transitional image from which the word in the output language is created. This method takes a lot of vocabulary and semantic information and major types of rules[1]. The computer sets these rules and transfers the structure of the input language into the output language. This method is quite sophisticated as it is built on large dictionaries and language rules[1]. On the other hand, SBMT relies on statistical translation models. Who has parameters based from the analysis of mono and bi lingual corpora [1]. It takes less time to build a statistical model. Tech companies like Google and Bing are using the SBMT tool for translation. SBMT as defined by, when finding the sentence T, sentence T originates from the output language O, the challenge is to find the sentence S which S infer from the input language is the source used to interpret the text [2]. By picking the most credible sentence S, the probability of error of translation is less and hence the probability of having the accurate translation is high.

Neural-based machine translation(NBMT) is the recent tool introduced for machine translation. NBMT does not work on a phrase-based translation system like RBMT and SBMT, rather it builds a single neural network that can be combined to maximize the translation accuracy [3]. Also, Most of the models built for NBMT are based on the encoder-decoders system[3]. One of the important principles of this model is that it encodes the input sentence to a series of vectors and picks a subset of vectors while decoding the translation [3].

Performing a performance-based evaluation of the three methodologies utilized for machine translation;the model should be easy to access and the process should be fast and it should not take much space on the computer memory. RBMT, which employs a pattern to translate text, performs better and is more reliable because the rules utilized do not change. Owing to the lack of a corpora-based strategy, RBMT employs as much disc memory as SBMT. The system development is as crucial as its performance. As a result, designing a system for a specific pair of linguistic words needs a definite amount of effort and time as well as costly. Because RBMT relies on principles developed by a number of well-known linguistic specialists, it necessitates a lot of manual labor and, more crucially, is more error-prone. SBMT meanwhile uses different corpora of languages from transformation of text on the set of languages being translated. On the other hand, NBMT utilized regular representations instead of discrete pattern representations in SBMT. Also, it uses a single neural network for the complete translation process, which enhances its performance.

Quality of translation plays an important part in machine translation. RBMT stands on the top as it includes the grammatical rules for translation, which gives more consistent and predicted quality of translation. Meanwhile, SBMT employs probabilistic methods. When it comes to translating natural languages, the outcomes are frequently inconsistent, and the quality is typically lacking. Overall translation quality of NBMT is slightly better than phrase-based methods, the BLEU and TER clearly indicating that NBMT has outscored the rest of the machine translation tools[4]. Whereas, it suffers badly while translating the very long sentences.

Quality of translation does not really play a sole decider for better translation models. When we talk about a model, which can handle both a domain of language and the translation process, RBMT scores more points as it can clearly handle many different languages with ease, whereas SBMT does not work efficiently when language comes out of its corpora domain. Whereas, the NBMT system is built on a four layer network, which has more than one thousand dimension word embeddings, source reversing and more than fifty thousand input and output vocabularies[4]. These help to build an NBMT system more effective for machine translation.

Fluency plays a part and parcel role for translation, it helps for analyzing the better translation model and technique to use in machine translation. By employing rules that may or may not be correct for some input language, fluency is not much achievable for RBMT, whereas for SBMT, because of its use of corpora collection from input to output languages, helps make it more fluent to output text. In NBMT, the fluency of full translation depends upon the generated word, which makes a good correlation with its past and upcoming translation. Moreover, with the help of encoders the fluency is assessed. However, the slang words do not follow the rules set by regular language. Also, the metaphorical terms do not follow the rules; their significance might be deceptive.

RBMT is not cost-effective, it just requires a linguistic professional for fine tuning. Whereas, SBMT is more cost-efficient as publicly accessible literature has increased as a result of the technology. Meanwhile, in NBMT, active learning techniques have helped to be more cost-effective and reduce the cost constraints while providing the best quality translation result. While having a lot of corpora of text in a domain, SBMT does not require any alteration for building the new set of words to

translate. Nowadays creating a corpus of languages is quite easier than before. But still bilingual and multilingual corpus for comparing language while translating is not that easy, as it needs a domain related translation otherwise it can produce various meanings to some words.

While comparing the first two types of machine learning technique. SBMT has gained popularity in recent years due to the tech revolution and also by using modern supercomputers [5]. Moreover, the data used for building the SBMT model is more widely available today. The adaptation of machine translation has become a more possible alternative. One more article, which compares the RBMT and SBMT, has discovered that SBMT has better translation quality than SBMT, while comparing the text translation from the same and different languages[6]. This article also pinpoints that the first MT techniques need a lot of human labor and assets. Adding to one more point about this article, it reflects that the oldest techniques are more viable to inflect morphemes and suffixes, while the SBMT has got many incorrect words.

There is a comparison done among all the three MT techniques on English to Finnish language mixing up product and process related method[7]. Almost more than thirty students have participated in such experiments by providing the post and process edit information. It was discovered that RBMT has tackled this challenge quite well and corrected most of the verb form type edit, whereas the rest two has found errors while predicting the accurate form of the verb. While doing the process based experiment, it was found that there was a sharp decline in mean pause length for NBMT as compared to RBMT and SBMT. Also, statistical parity was checked in per segment, NBMT has less statistical difference as compared to other machine translation tools.

The most popular and used machine translation system that is Google translate, was using the statistical based technique from april 2006. It has used SBMT to train the hefty data warehouse for most of the types of domains, and has provided excellent results. Later in 2016, Google switched to a neural based translation engine (GNMT); which is able to translate whole sentences rather than pieces of words. GNMT is now being used in more than 130 languages, which is a great feat for machine translation.

To evaluate the essay, I believe the most probable option to be used would be NBMT. It is faster, dynamic and more accurate for machine translation. After NBMT, I would pick SBMT, as it is training with larger data than rule based, predicting more correct and accurate translation than rule based. There are few stumbling blocks on the way of RBMT, since it being the oldest method, many internet revolutions like slang words have triggered its techniques and also it is outdated as compared to the modern usage of translation. While I would use SBMT when it comes to use a language which has the bigger corpus of data.

Since the rise of popularity of technology, there is further work to be done to make machine translation more efficient. Hybrid machine translation techniques have been introduced with the help of SBMT and NBMT. In this technique, translation is done by using multiple machine translation methods within a single system. According to the rules of proper spelling and sound structure RBMT provides a better translation method whereas SBMT is well with the high fluency. We should focus more on creating a system which can use both the RBMT and SBMT too. Since the hybrid

based model uses a parallel method, which is far better with dealing with error and inefficiency of other translation methods. This method is getting a lot of praise for supporting the speech especially during the international summit or business meetings. There are further improvements needed in machine learning; scoring the edit score higher and building algorithms to work better for minority languages. On the whole, the language has been introduced and created by humans so it is not a child play for computers to perform in a similar way as humans do.

Bibliography

[1] What is Machine Translation? Rule Based Machine Translation vs. Statistical Machine Translation.https://www.systransoft.com/systran/translation-technology/what-is-machine-translation/ [Accessed 1 May 2022].

[2]Brown, P.F., Cocke, J., Della Pietra, S.A., Della Pietra, V.J., Jelinek, F., Lafferty, J., Mercer, R.L. and Roossin, P.S., 1990. A statistical approach to machine translation. Computational linguistics, 16(2), pp.79-85.[Accessed 2 May 2022].

[3]Bahdanau, D., Cho, K. and Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.[Accessed 10 May 2022].

[4]Bentivogli, L., Bisazza, A., Cettolo, M. and Federico, M., 2016. Neural versus phrase-based machine translation quality: a case study. arXiv preprint arXiv:1608.04631.[Accessed 6 May 2022].

[5] Kantan MT (2014). RBMT vs SMT. [Blog] Kantan MT Blog. Available at: https://kantanmtblog.com/2014/02/13/rbmt-vs-smt/ [Accessed 23 April 2022].

[6]Sreelekha, S. (2019). Statistical Vs Rule Based Machine Translation; A Case Study on Indian Language Perspective. [online] Available at: https://arxiv.org/ftp/arxiv/papers/1708/1708.04559.pdf [Accessed 6 May 2022].

[7]Koponen, M., Salmi, L. and Nikulin, M., 2019. A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output. Machine Translation, 33(1), pp.61-90.[Accessed 10 May 2022].