A Study Of Machine Translation Theories And Methods

Guli I. Ergasheva
Dr, Uzbekistan State World Languages University, Uzbekistan

Zahriddin.X. Haitqulov
Teacher, Uzbekistan State World Languages University, Uzbekistan

ABSTRACT

The demand for language translation has greatly increased in recent times due to increasing cross-regional communication and the need for information exchange. Most material needs to be translated, including scientific and technical documentation, instruction manuals, legal documents, textbooks, publicity leaflets, newspaper reports etc. Some of this work is challenging and difficult but mostly it is tedious and repetitive and requires consistency and accuracy. It is becoming difficult for professional translators to meet the increasing demands of translation. In such a situation the machine translation can be used as a substitute. This paper intends to study methods and techniques of Machine Translation (MT). Through the following points: History of MT, Statistical MT, Types of MT, and evaluation of MT.

KEYWORDS

Machine translation, computational linguistics, statistical, rule based, language translation, types of MT and evaluation of MT.

INTRODUCTION

Over the years, Machine Translation has been a focus of investigations by linguists, psychologists, philosophers, computer scientists and engineers. It will not be an exaggeration to state that early work on MT contributed very significantly to the development of such fields as computational linguistics, artificial intelligence and application-oriented natural language processing. Moreover the idea of language

translation is developing currently that solves the issues of linguistic diversity. It is not possible to know and grasp all the languages within the world by human beings. Around 5000 languages present in the world that shows the need of language translation methods and its developments Researches within the field of language translation are exploring the possibilities of message transferring from one language to different. Machine translation, commonly known as MT, can be defined as "translation from one natural language (source language (SL)) to another language (target language (TL)) using computerized systems and, with or without human assistance" [1] [2]. Thus, Most recently, machine translation tools achieved translation excellence. Dictionary based machine translation was the first generation of automated language translation and it was purely based on electronic dictionaries. It translates the phrases but not sentences. Next, Rule Based Machine Translation (RBMT) systems, Corpus Based systems and Hybrid Machine Translation systems were developed. RBMT builds linguistic rules based on morphological syntactic and semantic information related to source and target language. At the same time, Corpus Based systems generate translations from bilingual text corpora. Hybrid method is advanced method that combines the benefits of individual techniques to attain an overall better language translation.

We try to give in this paper a coherent, if necessarily brief and incomplete, the development has been the field of machine translation through four points which are: first of all surveys the chronological development of machine translation, the different approaches developed (linguistic and computational), the types of machine translation and finely, we try to answer an important question which is how to evaluate a machine translation?

Machine Translation (MT) of natural human languages is not a subject about which most scholars feel neutral. This field has had a long, colorful career, and boasts no shortage of vociferous detractors and proponents alike. During its first decade in the 1950s, interest and support was fueled by visions of high-speed high-quality translation of arbitrary texts (especially those of interest to the military and intelligence communities, who funded MT projects quite heavily). During its second decade in the 1960s, disillusionment crept in as the number and difficulty of the linguistic problems became increasingly obvious, and as it was realized that the translation problem was not nearly so amenable to automated solution as had been thought. The climax came with the delivery of the National Academy of Sciences ALPAC report in 1966, condemning the field and, indirectly, its workers alike. The ALPAC report was criticized as narrow, biased, and short-sighted, but its recommendations were adopted (with the important exception of increased expenditures for long-term research in computational linguistics), and as a result MT projects were cancelled in the U.S. and elsewhere around the world. By 1973, the early part of the third decade of MT, only three government-funded projects were left in the U.S., and by late 1975 there were none. Paradoxically, MT systems were still being used by various government agencies here and abroad, because there was simply no alternative means of gathering information from foreign [Russian] sources so quickly; in addition, private companies were developing and selling MT systems

based on the mid-60s technology so roundly castigated by ALPAC. Nevertheless the general disrepute of MT resulted in a remarkably quiet third decade. We are now into the fourth decade of MT, and there is a resurgence of interest throughout the world - plus a growing number of MT and MAT (Machine-aided Translation) systems in use by governments, business and industry: in 1984 approximately half a million pages of text were translated by machine. Industrial firms are also beginning to fund M(A)T R&D projects of their own; thus it can no longer be said that only government funding keeps the field alive (indeed, in the U.S. there is no government funding, though the Japanese and European governments are heavily subsidizing MT R&D). In part this interest is due to more realistic expectations of what is possible in MT, and realization that MT can be very useful though imperfect; but it is also true that the capabilities of the newer MT systems lie well beyond what was possible just one decade ago. In light of these events, it is worth reconsidering the potential of, and prospects for, Machine Translation [3; 1-3].

DISCUSSION

Where we are using machine translation? Language translation systems facilitate the individuals to communicate each other from different places so they can utilize the advantages of information and communication technology [4; 25-35]. Machine translation is widely employed in numerous applications and a few translation agencies including government agencies are supporting implementation of tools [5; 66-73]. Translation tools will primarily used for conducting research by reviewing foreign websites and articles. In addition,

marketing, legal purposes, software localization, email translation for customer enquiries, website translation, manuals and documents translation, customer support, personal communication like travel reservations, managing assets abroad etc, are possible with MT software.

Machine translation uses the method based on linguistic rules which convert source language to target language. Natural language understanding is the most important thing for the success of machine translation. As explained above different methods are available for automated machine translation. Type of technology chosen for machine translation is primarily depends on the source and target language pair. If customization is performed in regular basis, RBMT is better and it gives good result. But comparing with Corpus based and Hybrid method it is less efficient. Target language does not have rich morphology features it is good to use Corpus Based MT especially Statistical MT. When source and target languages are more complex, Hybrid MT is better to use because this combines the advantages of different approaches.

RESULT

Rule based machine translation (RBMT) is also called Knowledge Based Machine Translation that retrieves rules from bilingual dictionaries and grammars based on linguistic information about source and target languages. RBMT generates target sentences on the basis of syntactic, morphological and semantic regularities of each language. It converts source language structures to target language structures and it is extensible and maintainable [6;388-393]. There are three types of RBMT systems:

- Direct method (Dictionary Based Machine Translation). Source language text are translated without passing through an intermediary representation. Anusaarka is the example of system that uses direct approach. Indian Institute of Information Technology, Hyderabad, develops it.
- Transfer RBMT Systems Morphological and syntactical analysis is the fundamental approaches in Transfer based systems. Here source language text is converted into less language specific representation and same level of abstraction is generated with the help of grammar rules and bilingual dictionaries. Mantra is a transfer based tool which is a funded project of India Government.

- C. Interlingua RBMT Systems (Interlingua). This model is indented to make linguistic homogeneity across the world. In this method, source language is translated into an intermediary representation which does not depends on any languages. Target language is derived from this auxiliary form of representation [4; 20-25]. The main property of this model is single representation for different languages and much easier to multilingual machine translation. UNITRAN (UNIversal TRANslator) system is an example of Interlingua model. The edit distance and review distance of Rule Based Machine Translation system are given here.

Table - Edit/Review distance for RBMT languages

Language	Edit distance	Review distance	Volume (words)
French	46.33%	9.1%	38900
Italian	49.05%	16.94%	40149
Spanish	33.67%	6.30%	56269
Simplified Chinese	54.43%	2.69%	80367

The benefits of RBMT are easy customization and predictability. Easy customization means user dictionaries is adjusted to fix errors and predictability is the quality to understand the output you can expect with basic understanding of the tool. RBMT has some disadvantages as in [7; 217-220] and first one is unavailability of good dictionaries. New dictionary building is truly high-priced task. Another limitation is, it's necessary to set some linguistic information manually. In addition, it's very difficult to manage rule interactions and ambiguity in the large system. RBMT allows building new rules and extends it but these changes are very expensive [7; 217-220].

Evaluation of Machine Translation Systems. Evaluating Machine translation system is important not only for its potential users and buyers, also to researchers and developers. Various types of evaluation have been

developed, such as : BLEU(Bilingual Evaluation Understudy), WER (Word Error Rate), PER (Position-independent word Error Rate) and TER (Translation Error Rate).

The BLEU metric, proposed by Papineni in 2001 was the first automatic measurement accepted as a reference for the evaluation of translations. The principle of this method is to calculate the degree of similarity between candidate (machine) translation and one or more reference translations based on the particular n-gram precision [8; 311-318].

The WER metric, Proposed by Popovic and Ney in 2007. Originally used in Automatic Speech Recognition, compares a sentence hypothesis refers to a sentence based on the Levenshtein distance. It is also used in machine translation to evaluate the quality of a translation hypothesis in relation to a reference translation. For this, the idea is to calculate the minimum number of edits (insertion, deletion or substitution of the word) to be performed on hypothesis translation to make it identical to the reference translation [9].

The PER metric, proposed by Tillman in1997. Compare the words of machine translation with those of the reference regardless of their sequence in the sentence [10; 2667-2670].

The TER metric, proposed by Snover in 2006. Is defined as the minimum number of edits needed to change a hypothesis so that it exactly matches one of the references. The possible edits in TER include insertion, deletion, and substitution of single words, and an edit which moves sequences of contiguous words. Normalized by the average length of the references. Since we are concerned with the minimum number of edits needed to

modify the hypothesis, we only measure the number of edits to the closest reference [11].

CONCLUSION

Machine Translation is an automated process within which computer software is used to convert text from one natural language to another. Translator ought to interpret the contents within the source text and build sentence structure of target language for translation. This process demands wide knowledge in grammar, structure of sentence and its meanings in the source and target languages. Machine Translation has an important role today in various applications such as customer management, documents translation, communications, software localization website translation etc. Dictionary Based, Rule Based, Corpus Based and Hybrid approaches are the main methods for machine translation. Each of these has its own advantages and limitations as explained above. It's a proven fact that no two translation system can produce identical translations of same text in the same language pair. Also it is necessary to perform post editing for quality translations. Moreover, the development of translation as an application of Computational Linguistics will require substantial research in its own right in addition to the work necessary in order to provide the basic multilingual analysis and synthesis tools. Translators must be consulted, for they are the experts in translation. None of this will happen by accident; it must result from design.

REFERENCES

1. Hutchins, W. J. and Somers, H. L. (1992). An introduction to machine translation, Academic Press, London.

2. Baumgartner-Bovier, (2003). "La traduction automatique, quel avenir ? Un exemple basé sur les mots composés", Cahiers de Linguistique Française N°25

3. Jonathan Sloculm. (1985). A survey of machine translation: its history, current status, and future prospects. Microelectronics and Computer Technology Corporation .Austin, Texas. Computational Linguistics.

4. T. V. Prasad, G.M. Muthukumaran, (2013)"Telugu to English Translation using Direct Machine Translation Approach, International Journal of Science and Engineering Investigations, 25-35.

5. M. N. Al-Kabi, T. M. Hailat, E.M Al-Shawakfa, I. M Alsmadi, (2013)."Evaluating English to Arabic Machine Translation Using BLEU", IJACSA, 66-73.

6. S. Tripathi, J. K. Sarjgek, (2010). "Approaches to machine translation", Annals of Library and Information Studies,388-393.

7. A -L Lagarda, V. Alabau, Casacuberta, R. Silva, E. Díaz-de-Liaño, (2009)"Statistical Post-Editing of a Rule-Based Machine Translation System". Proceedings of NAACL HLT. 217–220,

8. K. Papineni, S. Roukos, T. Ward, and W. Zhu, (2002). "Bleu: a Method for Automatic Evaluation of Machine Translation", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, 311-318

9. M. Popovic and H. Ney (2007)." Word error rates: Decomposition over POS classes and applications for error analysis". In Proceedings of ACL Workshop on Machine Translation.

10. C. Tillman, S. Vogel, H. Ney, H. Sawaf, and A. Zubiaga. (1997)."Accelerated DP-based search for statistical translation". In Proceedings of the 5th European Conference on Speech Communication and Technology, 2667.2670. Rhodes, Greece.

11. M. Snover, B. Dorr, R. Schwartz, L. Micciulla, J. Makhoul. (2006). "A Study of Translation Edit Rate with Targeted Human Annotation". In Proceedings of AMTA, Boston.