The Future of Translation – Machine Translation

K. Dhanalakshmi, Lecturer (Senior Grade)
Department of Computer Engineering
Ayyanadar Janakiammal Polytechnic College, Sivakasi, Tamilnadu
kbdhana@gmail.com
==

Abstract

Being a part of fast growing digital era, we have to extend our communication to the global level. So we need Language Translators for converting a word, phrase, paragraph or whole text into many languages. Now-a-days Machines are used as Translators. Machine Translators (MT) have a large amount of data with various translating methods. MT is improve by more training data, better Computer Hardware and improved methods of creating a translation. MT Translation methods are different, and each has its own benefits and drawback. This paper explains various approaches used in machine translation process such as Rule based Translation, Statistical and Example Based Translation and Hybrid Translation methods.

Keywords: Language Translators, Machine Translators, Rule Based, Statistical, Example Based, Hybrid Translation.

Introduction

Languages are growing in importance as developing countries take their place in the Global economy [1] and more of their citizens gain Internet Access. Most of the Web Content used to be in English. And remember, even people who speak English as a second language generally prefer to read, watch and shop in their first language. That's why, Translation Services are important. It will help you to communicate globally and will ensure your message stays relatable and effective Worldwide.

Human Vs Machine Translation

The long running debate of Machine Vs Human Translation refuses to go away. This is partly because technology improves every year and the quality of translation it offers is constantly rising.

Any Translation carried out by a computer is Machine Translation (MT). It is the simplest form of translation, allowing users to choose the base and target languages. The advantage of Machine Translation generally come down to two factors. It's faster and cheaper.

==
Language in India www.languageinindia.com **ISSN 1930-2940 19:1 January 2019**
Dr. S. Jayanthi. Editor. *Translation: Exploring the Nuances and Possibilities of Mediating Intercultural Texts*
K. Dhanalakshmi
The Future of Translation – Machine Translation 108

The problems with Machine Translation begin when it actually depends on the context to know what is meant. Things like tone, idioms and other linguistic nuances won't be picked up by Machine Translation. It will only ever produce a literal word-for-word translation.

Machine Translation

Around 5000 languages are present in the world. It is not possible to know and grasp all the languages within the world by human beings. Government Agencies and Research Institutes initiates Machine controlled text translation for international communication. 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. Recently used Machine Translation Methods [6] are

1. Rule-based Machine Translation (RBMT)

RBMT builds linguistic rules based on morphological, syntactic and semantic information related to source and target language.

2. Statistical and Example based Machine Translation (SMT)

In these types, the Translations based on highest probability of existing database of examples. This method is also called Corpus-based MT. This translation requires bilingual Corpus for each target language, a language modeler and a decoder.

3. Hybrid Translation Method (HT)

Hybrid method is advanced method that combines the benefits of individual techniques to attain an overall better language translation.

Rule-based Machine Translation (RBMT)

Rule based Machine Translation (RBMT) system [2] was the first Commercial Machine Translation System. The Source text of RBMT is transferred into Target text based on dictionaries and grammars covering the main Semantic, Morphological and syntactic regularities of each language respectively. RBMT methodology applies a set of linguistic rules in three different phases.

1. Analysis -> Getting syntactic and lexical information and parsing the source sentence.
2. Transfer -> Translate source to target languages.
3. Generation -> Mapping dictionary entries into Appropriate inflected forms.
For Example: Source text in English: A GIRL EATS AN APPLE.

To get a German Translation of this English sentence one needs
1. A Dictionary that will map English word to an appropriate German word.

===
Language in India www.languageinindia.com **ISSN 1930-2940 19:1 January 2019**
Dr. S. Jayanthi. Editor. *Translation: Exploring the Nuances and Possibilities of Mediating Intercultural Texts*
K. Dhanalakshmi
The Future of Translation – Machine Translation   **109**

2. Rules representing English sentence structure.
3. Rules representing German sentence structure.

The translation can be made by the following stages.

| A | GIRL | EATS | AN | APPLE | English |

		ANALYSIS			

| A | GIRL | EATS | AN | APPLE | |
| Article | Noun | Verb | Article | Noun | |

		TRANSFER			

| A | GIRL | EATS | AN | APPLE | English |
| Ein | Madchen | essen | ein | Apfel | German |

		GENERATION			

| A | GIRL | EATS | AN | APPLE | English |
| Ein | Madchen | isst | einen | Apfel | German |
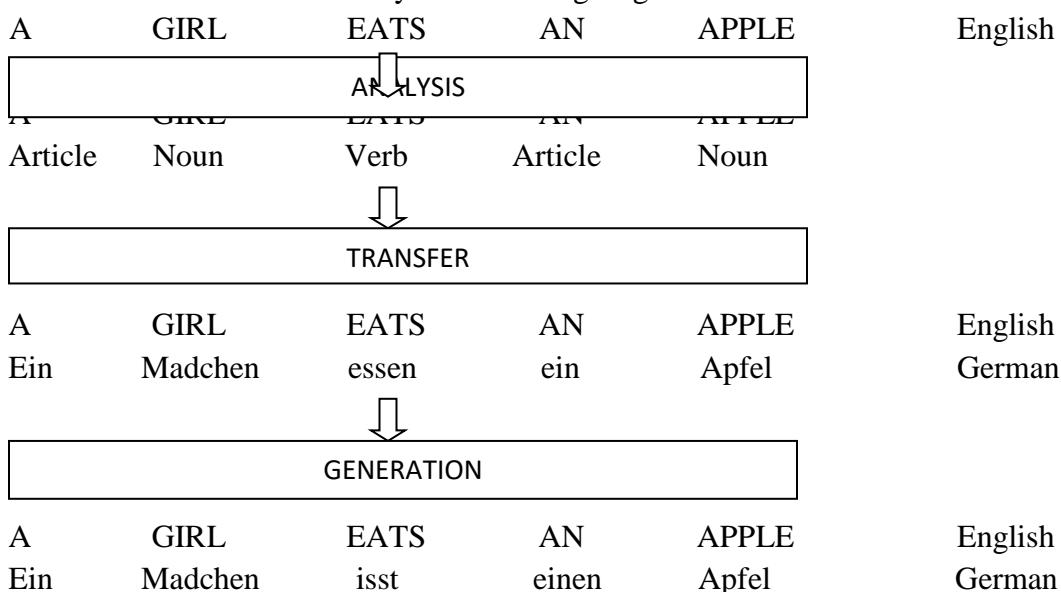
RBMT can achieve high accuracy within narrow subset of language. Many rules must be added to improve quality, lending to a very complex system

Statistical and Example Based Machine Translation (SMT)

Statistical Machine Translation (SMT) based upon analyzing existing translations. Machine translator can use a database as the source for all the information it needs for translation. Parallel Corpora [3] (bitexts) are collections of texts available in two languages. Some Examples are Multilingual legal text translated sections of Wikipedia.

Some challenges in translation [4] are

❖ Word order in language differs and thus word translation is difficult.

For Eg : English :	MY	NAME	IS	RAM	(SVO)
		Subject	verb	Object	
Hindi :	MY	NAME	RAM	IS	(SOV)
	Mera	nam	Ram	Hai	

❖ A word has more than one meaning.
   For Example: "He went to Bank, but it was closed for Lunch"

   Here, "Bank" refers to Financial Institution and not a river bank.

==
Language in India www.languageinindia.com **ISSN 1930-2940 19:1 January 2019**
Dr. S. Jayanthi. Editor. *Translation: Exploring the Nuances and Possibilities of Mediating Intercultural Texts*
K. Dhanalakshmi
The Future of Translation – Machine Translation   **110**

In SMT, the probability for translating Source (S) to Target (T) is done by Bayse' Theorem[5].
$P(S/T) = P(S). P(T/S) / P(T)$

The three major aspects are
1. <u>Translation Model</u> -> P(T/S), search for the most probable translation (T) for a given Source (S)
2. <u>Language Model</u> -> P(S), which models the fluency of the proposed target sentence. This mode use N-gram models, conditioning on two or more previous words for predicting the probability of the current word.
3. <u>Decoding</u> -> A search process which is concerned with navigating through the possible target translations.

Hybrid Machine Translation: (HMT)

HMT is a method of Machine Translation that is characterized by the use of multiple Machine Translation approaches within a single Machine Translation system. The motivation for developing hybrid machine translation [6] systems stems from the failure of any single technique to achieve a satisfactory level of accuracy. There are several forms of HMT such as

1. <u>Multi-Engine</u>: In this method, input sentence is divided into some parts and translates them by switching RBMT and each part. The splitting is based on its syntactic structure.

2. <u>Multi-Pass</u>: This method uses RBMT as baseline and refines through Statistical Model [7]
Rule Post Processed by Statistics: Translations are performed using a Rules Based Engine.

The output from the Rules Engine are then adjusted or corrected by the Statistics Engine. This is also known as Statistical smoothing and Automatic post editing. This attempt improves the lower quality output from an RBMT Engine.

Statistics Guided by Rules: Rules are used to Post-process the Statistical output to perform functions such as normalization. This approach has a lot more power, flexibility and control when translating.

3. <u>Multi-System</u>: Multiple Machine Translation systems run in parallel. Then the system selects the best translation as output.

Conclusion

Machine Translations from companies like Google, Bing and SYSTRAN have come a long way in the past five years. And advancement in Artificial Intelligence and Neural Networks

==
Language in India www.languageinindia.com **ISSN 1930-2940 19:1 January 2019**
Dr. S. Jayanthi. Editor. *Translation: Exploring the Nuances and Possibilities of Mediating Intercultural Texts*
K. Dhanalakshmi
The Future of Translation – Machine Translation   **111**

improve the Translation methods. Each method has its own pros and cons. By understanding a suitable method for our use, one can improve the cost and efficiency of Translations.

===

References

[1] Translations.com, GMS Technology Solutions, Globalization Management Systems. Available at: http://www.translations.com/quality/solutions/gmsTechnology-solutions.html (Accessed on 19th June, 2015)

[2] Yang V S-C, Electronic dictionaries in machine translation Encyclopedia of Library and Information Science,48 (1991) 74-92.

[3] Hutchins J, Example based machine translation- a review and commentary. In Recent advances in example-based machine translation .(Ed. Micheal Carl and AndyWay) (2003).Available at: http://www.hutchinsweb.me.uk/MTJ-2005.pdf (Accessed on 21- June, 2015)

[4] Macherey K, Bender O and Ney H, Applications of statistical machine translation approaches to spoken languages understanding, IEEE Transactions On Audio Speech and Language Processing, 17(4) (2009) pp.803- 818

[5] B Dorr, E Hovy, L Levin, Machine Translation, Interlingual Methods, Elsevier (2008). [22] Available at: http://www.cut-theknot.org/Probability/BayesTheorem.shtml (Accessed on 09- July 2015)

[6] en.wikipedia.org/wiki/Machine translation.

[7] S. Corston-Oliver, M. Gamon, and C. Brockett. A machine learning approach to the automatic   evaluation of machine translation. In Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, pages 148–155. Association for Computational Linguistics, 2001.

===
Language in India www.languageinindia.com **ISSN 1930-2940 19:1 January 2019**
Dr. S. Jayanthi. Editor. *Translation: Exploring the Nuances and Possibilities of Mediating Intercultural Texts*
K. Dhanalakshmi
The Future of Translation – Machine Translation   **112**