A Survey on Hybrid Machine Translation

Anusha Anugu[1], *Gajula* Ramesh [2]

[1]PG Scholar, Department of Computer Science and Engineering, GRIET, Hyderabad, Telangana, India.
[2]Associate Professor, Department of Computer Science and Engineering, GRIET, Hyderabad, Telangana, India.

Abstract—Machine translation has gradually developed in past 1940's.It has gained more and more attention because of effective and efficient nature. As it makes the translation automatically without the involvement of human efforts. The distinct models of machine translation along with "Neural Machine Translation (NMT)" is summarized in this paper. Researchers have previously done lots of work on Machine Translation techniques and their evaluation techniques. Thus, we want to demonstrate an analysis of the existing techniques for machine translation including Neural Machine translation, their differences and the translation tools associated with them. Now-a-days the combination of two Machine Translation systems has the full advantage of using features from both the systems which attracts in the domain of natural language processing. So, the paper also includes the literature survey of the Hybrid Machine Translation (HMT).

1. Introduction

The process of retrieving and evaluating the information from the document repositories is known as "Information Retrieval (IR)" [1]. The user who needs information has to send request in the form of a query in natural language. Then the information related output will be retrieved from the IR system. The process of IR system [2] is as shown in figure1.
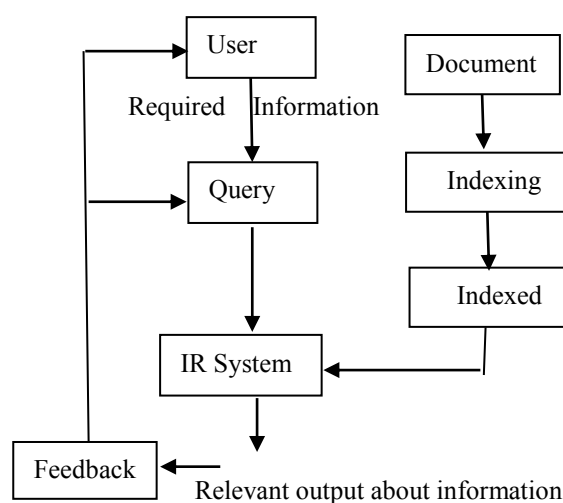

Fig. 1. Information Retrieval Process

Now-a-days, searching information in different languages has been increased rather than original language which creates a problem in IR system. Then the translation has evolved.

The domain of research work of 'Natural Language Processing (NLP)' which fulfils the interaction among the distinct classification of the nation is known as "Machine Translation". As the human made translation is expensive and time taking process, MT system is used which reduces the time and cost. MT is an automated application used by the computer to translate one language into other. In 1940's the research in MT's has been started.It is more advantageous to the industries for consumer maintenance, increasing the capacity for the ac complished translators.

2. Machine translation techniques

There are four main techniques in Machine Translation [2]. They are

2.1 Direct Machine Translation:

In initial days this type of translation is used. It translates word after word with a few word-order adjustments. It depends on dictionary look-up. Without the analysis of internal structure and grammatical correlation, the source sentence is morphologically analyzed to derive target sentence.

This process involves three steps
1. Morphological Analysis —root words are extracted from the words in source language.
2. Dictionary Lookup —searches for the matching words for target language words.

3. Syntactic Rearrangement—the order of words of source language are arranged to match the sentence in target language
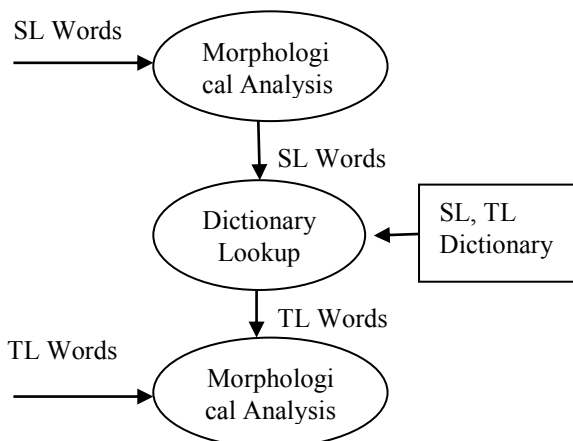

Fig. 2.1. Direct Machine Translation

.2.2 Rule Based Machine Translation (RBMT):

"Knowledge Based Machine Translation(KBMT) " is another name of RBMT, the classical approach of Machine Translation. It depends on the semantic information of both source and target languages. This information is primarily obtained from semantics and lexicons, linguistic consistencies of every language independently. In this system, the source language text is examined and an intermedial presentation is generated from which the target text is produced.

Based on intermedial presentation RBMT is categoriz ed as "Interlingua Machine Translation" and "Transfer-Based Machine Translation".

2.2.1 Interlingua Machine Translation:

Interlingua MT is one of the major advanced system in which the source text might be translated beyond one language. The system involves in creating a universal language text called Inter language text from source text and the target text is produced from that Inter language text as shown in Figure 2.2.1.
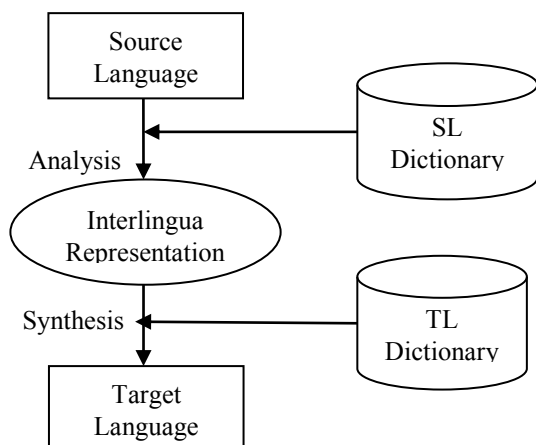

Fig. 2.2.1 Interlingua Machine Translation

It follows two steps [8][9]

1. Analysis-- This step translates the source text to an inter language text which doesn't depends on both the source and target languages.
2. Synthesis-- This step translates inter language text of source into its corresponding target text.

2.2.2 Transfer Based Machine Translation:

TBMT has the same idea of translating the source text into an intermedial text from which the target text is produced [10]. Unlike Interlingua MT, the intermediate language partially depends on the pair of languages involved.

The model comprises of three phases [8]. They are

1. Analysis—the sentence structure and constituents of source sentence are indentified by analyzing the source sentence.
2. Transfer—the source sentence is translated into its corresponding target sentence.
3. Generator—depending on the syntax of the target language, translation is done.
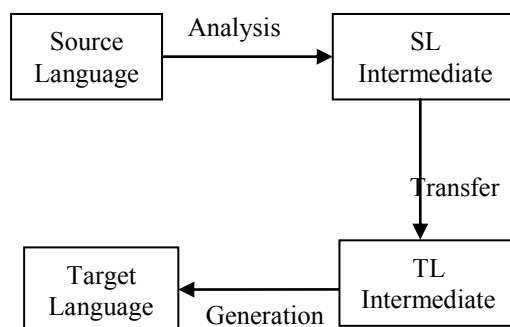

Fig. 2.2.2. Transfer Mased Machine Translation

2.3 Corpus Based Machine Translation (CBMT)

The group of spoken or written content that are used in the research is known as Corpus. The source language and its translations are stored in the parallel corpus. CBMT [8] is the approach that carries a lot of work in present days and it doesn't require any linguistic knowledge for translating source to target language.

This approach is categorized into two approaches. They are "Statistical MT" and "Example Based MT".

2.3.1 Statistical Machine Translation (SMT):

This system translates source text to target text depends on the analytical methods uprooted from the huge volume of alliance bilinguals collection. These statistical representations contain particulars such as interaction between sources, target texts and are build by using supervised and unsupervised algorithms. The statistical

models [11] help the system to obtain the perfect translation for the text while translation is performed.

SMT comprises of decoder, translation model and the decoders and is shown in the Figure2.3.1.
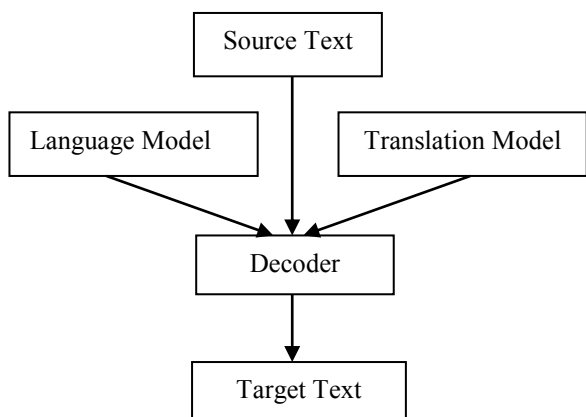
Fig. 2.3.1. Statistical Machine Translation

2.3.2 Example Based Machine Translation:

EBMT [5] contains intelligence for translation from bivocal or multivocal corpus which contains numerous data. Sample of sentences from source language and target language are stored as examples in corpus which are considered for further translation. Whenever content is taken for translation, the approach searches in the corpus whether it is available or else looks for any related text which can support in its translation. If the text that is not yet translated or a superior translation is found, then the corpus is updated by the administrator for future translations.
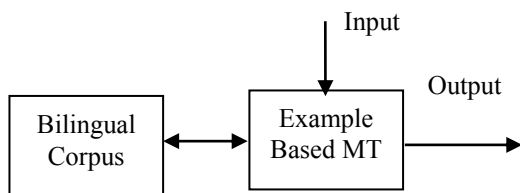

Fig. 2.3.2. Example Based Machine Translation

Recently "Neural Machine Translation (NMT)" is added to the "Corpus Based Machine Translation".

### 2.3.3 Neural Machine Translation (NMT):

It is a recently proposed method utilizing special neural grid framework called Encoder-Decoder architecture [12]. It doesn't require any predefined features (features which are designed, not learned from the data).The goal of NMT is to build a model that maximizes the translation performance. The representation of source sentence is a sequence of words and it uses vector representation to store the encoded meaning of the source.
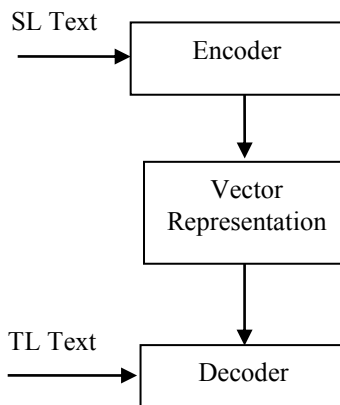
Fig. 2.3.3 Neural Machine Translation

2.4 Hybrid Machine Translation:

The aggregation of two or more translation systems is known as Hybrid Machine Translation. This approach seeks to extract the benefits of both the systems by utilising accessible assets.
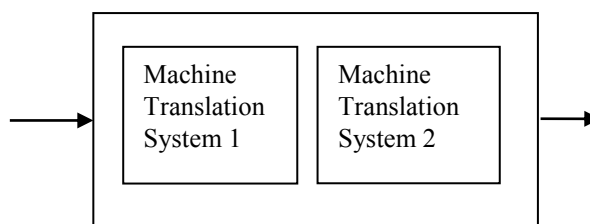
Fig. 2.4 Hybrid Machine Translation

3. Evaluation techniques of machine translation

There are various ways for the evaluation of the results of Machine Translation. They are

- *Round Trip Evaluation:*

In Round Trip Evaluation method, "the target language text which is translated from source language is further translated back to its original source language using same translation system". However the method is convenient to use, it results in the quality deficiency.

- *Human Evaluation System:*

This method depends on the knowledge of the language professionals affiliated with source and target languages.

- *Automated Evaluation System:*

This method needs a source translation in accurate to correlate a human translated text and machine translated text.
  - BLEU (Bi-lingual Evaluation under Study)--It is performed by detecting the translation BLUE n-

grams and correlating it with n-grams of its source translation and computing the scores which is the harmonic mean of changing rigor. It's range is 0-1, if >0.30 then the translation is understandable and If >0.5, a better translation.

- NIST-- NIST is based on BLEU metric with slight modification. In augmentation to n-gram precision, it adds some mass to n-gram.

- METEOR—this metric is from the "harmonic mean of unigram recall and rigor. The researchers proved that recall based metric accomplishes correct results than rigor based".

- LEPOR—A composite metric which depends on the numerous estimation components along with the current one and the revised one

4. Results and discussions

The different machine translation approaches that are used in translating source language texts to target language texts are reviewed individually. From the review the different machine translation systems are compared and differences among them are shown in the below Table 5.1.

Table. 5.1. Comparison of Machine Translation Techniques

S.NO	Machine Translation Approach	Advantages	Disadvantages
1	Direct Approach	• It's implementation is easy. • For languages with similar structure and grammar rules, the translation is effective.	• For multilingual scenarios, it can be expensive. • Involves only in lexical analysis.
2	Interlingua Approach	• Well suited for domain approach. • Produces content stationed portrayal and can be applied in IR.	• Time efficiency is less than Direct Approach. • Major problem in defining Interlingua.
3	Transfer Based Approach	• It defines the modular structure. • It can easily handle the ambiguities in different languages.	• A part of the origin text content may be vanished at the foot of the translation.
4	Statistical Approach	• SMT systems are not designed for any particular language pair. • It is less expensive compared to RBMT. • Translations used to be natural as it is trained by the real time texts.	• The Creation of corpus is expensive. • Errors are hard to find and fix • It requires extensive hardware configuration.
5	Example Based Approach	• This approach avoids the use of manually driven rules • It is adaptable to many languages. • To make the system attractive, minimum prior knowledge is required.	• It requires large database. • The system is not efficient as it consists of noisy corpora.
6	Neural Approach	• NMT outputs are more fluent. • It performs better in terms of inflection and re-ordering. • Occupies lesser memory than SMT.	• NMT performs rather poorly for long sentences. • The vocabulary is limited and source coverage problem. • Emprises translation problem.

The Machine Translation approaches and their corresponding tools are reviewed from several researches in the Natural Language Processing. The results from the review are formatted in the following Table 5.2.

Table. 5.2. Machine Translation Tools

S.No	Machine Translation Approach	Machine Translation Tools	Tools Description
1	Direct Approach	• ANUSAARKA	• It converts different languages like Kannada, Marathi, Punjab, Bengali, Marathi and Telugu to Hindi.
2	Interlingua Approach	• ANGLABARTI	• It translates English to Indian Languages.
3	Transfer Based Approach	• MANTRA • MATRA	• "Machine Assisted Translation Tool", it can be adapted in different fields like legislation, service system and advertisement. • Human Aided Machine, it can be adapted in the fields of broadcast, industrial remarks.
4	Statistical Approach	• Google Translator • Bing Translator	• It supports 90 languages and has limit to the number of paragraphs to be translated. • It supports 50 languages and is a Microsoft tool.
5	Example Based Approach	• ANUBHARTI • ANUBAAD	• It translates from Hindi to English. • It translates English into Bengali.

In order to overcome the translation problems in different Machine Translation(MT) systems and to receive the accurate and better results of the translation process, Hybrid Machine Translation(HMT) is used. HMT incorporates the full advantages of the MT systems that are used in the creation of HMT systems. Out of the several HMT systems, some of them are taken into consideration which is developed in recent years to get overview of those Machine Translation Systems. The overview of some HMT systems is shown in the Table 5.3.

Table. 5.3. Comparison of Hybrid Machine Translation Systems

S.No	Hybrid Machine Approach	Author	Year	Language Pair	Description
1	RBMT+SMT	Diluk Kayahan Tungg Gungor	2019	Turkish spoken to Sign	Initially, the Turkish input sentence is evaluated grammatically and then the translation rules are applied. The output of RBMT is boosted to the SMT to increase the quality of translation.
2	NBMT+RBMT	Muskaan Singh, Ravinder Kumar and Inderveer Chana	2019	Sanskrit - Hindi	It adopts deep learning feature to defeat the dis-advantage of RBMT and SMT approach. This system has figured out that it has lower feedback time and higher speed.
3	NMT+SMT	Xing Wang, Zhaopeng Thu and Min Zhang.	2018	Chinese-English and English-German	It uses the SMT word recommendations to predict the NMT word predictions and those are combined together to get the accurate translation Performance without any UNK signs(which are not in the original vocabulary)
4	EBMT+SMT+RBMT	Omkar Dhariya, Shrikant Malviya and Umashankar Tiwary	2017	Hindi-English	HMT approach is able to outperform many baseline translation systems based on EBMT, RBMT and EBMT individually. The system works only on four types of sentences. They are "Simple Present and Past Tense", "Present Continuous and Past Continuous Tense".
5	RBMT+EBMT	Paween Unlee and Pusadee Sere Sangatakul	2016	Thai- Isam	If a source word that is to be translated has several meanings, we use this system to select the most suitable meaning that is equivalent to its original source word.

5. Conclusion

This paper represents a summary of machine translation techniques along with their respective research work in a classified way. These techniques, improve the quality of translation, after the emergence of hybrid machine translation. The factors that improve the quality of translation are taking out anomalies to an extent possible, improving volubility and durability and also by contracting the dependence on both the source and target language. But "due to the demand of bulk vocabulary, domain specific nature of structure, lexicon and linguistic irregularities, etc of the automated system, its translation quality is not superior to manual translation etc".

The importance of our work is that the translation quality is enhanced due to the legitimate study of source and target languages that includes the analysis of syntaxes and linguistic formation. Furthermore, by thoroughly ignoring the anomaly of the sophisticated languages and the quality of translation can be improved. But still there are complexities in translating phrases which are considered for future research work.

References

1. Xing Wang, Zhaopeng Tu, and Min Zhang, "*Incorporating Statistical Machine Translation Word Knowledge into Neural Machine Translation*" IEEE /AC M Transactions on Audio, Speech, and Language Processing,**14**) (2018)
2. O.B Abiola, A.O Adetunmbi and A Oguntimilehin, IJCSMC, **4,**308-313(2015)
3. Paweena Unlee and Pusadee Seresangtakul, "*Thai to Isarn dialect Machine Translation using Rule-based and Example- based*", (JCSSE, 2016).
4. Shantanoo Dubey, "*Survey of Machine Translation Techniques*", IJARCSMS, **5** (2017)
5. Muskaan Singh, Ravinder Kumar and Inderveer Chana, "*Improving Neural Machine Translation Using Rule-Based Machine Translation*", ICSCC, (2007)
6. Pramod Salunkhe, Aniket .D. Kadam, Prof., "*Hybrid Machine Translation For English to Marathi:A Reasearch Evaluation In Machine*", ICEEOT (2016).
7. Sandeep Saini and Vineet Sahuta, "*A Survey of Machine Translation Techniques and Systems for Indian Languages*", ICICT (2017).
8. B. N. V Narasimha Raju and M S V S Bhadri Raju, "*Statistical Machine Translation System for Indian*
9. Suresh Kumar Tummala, Dhasharatha G, *E3S Web of Conferences* 87, 01030 (2019).
10. Lakshman Pandian S and Kumanan Kadhirvelu, "*Machine Translation from English to Tamil using Hybrid Technique*", *IJCS,* **46** (2012).
11. B.N.V Narasimha Raju , Dr. M S V S Bhadri Raju and Dr. K V V Satyanarayana, "*Translation Approaches in Cross Language Information Retrieval*".
12. Jayashree Nair, Amrutha Krishnan K and Deetha R1, "*An Efficient English to Hindi Machine Translation System Using Hybrid .Mechanism*", (ICACCI, 2016)
13. Afsana Parveen Mukta, Al-Amin Mamun, Chaity Basak, "*A Phrase Based Machine Translation from English to Bangla Using Rule-Based Approac*",ECCE (2019).
14. Dilek Kayahan and Tunga Gungor, "*A Hybrid Translation System from Turkish Spoken Language to Turkish Sign Language*".
15. Aditya Kaustav Das, Manabhanjan Pradhan, Amiya Kumar Dash, Chittaranjan Pradhan and Himansu Das, "*A Constructive Machine Translation System for English to Odia Translation*", ICCSP(2018)
16. Sunita Chand*, "Empirical survey of Machine Translation Tools",* ICRCICN(2016).
17. S.Anbukkarasi and Dr. S.Varadhaganapathy, "*Machine Translation (MT) Techniques for Indian Languages",* IJRTE, **8** ,Issue-2S4 (2019).
18. Rajagiri A, MN Sandhya, Nawaz S, Suresh Kumar T, *E3S Web of Conferences* 87 01004 (2019).
19. Mahjabeen Akter, M. Shahidur Rahman, Muhammed Zafar Iqbal and Mohammad Reza Selim*, "A Review of Statistical and Neural Network Based Hybrid Machine Translators*", ICBSLP (2018).
20. J. Sangeetha, S. Jothilakshmi and R.N.Devendra Kumar, "*An Efficient Machine Translation System for English to Indian Languages Using Hybrid Mechanism*", in J. Sangeetha et al. IJET.
21. Ravali Boorugu, Dr. Gajula Ramesh, Dr. Karanam MadhaviSummarizing "*Product Reviews Using NLP Based Text Summarization*", **8**, issue 10, (2019).
22. Ch. Mallikarjuna Rao, G. Ramesh, Madhavi, K., "*Feature Selection Based Supervised Learning Method for Network Intrusion Detection*", IJRTE, 8(2019).