IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A Review on Various Techniques of Machine Translation

Richu Mittal[*1], Naresh Kumar Garg[2]
[*1,2] Department of Computer Science & Engineering, GZS PTU Campus, Bathinda, India
richumittal28@gmail.com

Abstract

Machine translation is a very crucial field of (NLP).It plays a very important role in language processing. It can translate one natural language into another natural language . Due to Globalization Language translation is becoming a necessary part of human being's life . Various machine translation techniques are there such as direct translation, example based, statistical machine translation, rule based translation. They are used according to the problem domain. Objective of this paper is to provide a brief overview of different approaches of Machine Translation.

Keywords: Machine, Translation, Rule based Translation.

Introduction

Language is a very effective communication medium. With the help of a language we can communicate with another person either verbally or in a written form. It represents the ideas and expressions of human mind. Thousands of the languages available in the world, which represent the language diversity. Everyone can not understand all these languages so their must be a translation system which can translate from one system to another system. Developments in information and communication technology have brought a revolution in the machine translation. Now a days there are a wide variety of translation tools. Which is used by language professionals. These tools dose not generate the exact translated meaning,but it gave a central meaning with the help of which language professional can understand the meaning of contained information.Over internet online translators are available such as Google translate from Google. Major amount of web literature is available in English so translation mechanism must be there, machine translation is a research area under computational linguistic. it can also be defined as a system which take one natural language as a input and gave another natural language as a output. There are various approaches such as a rule based translation, transfer based, knowledge based translation, corprus based translation. The language which is given as a input is called source language and output language is known as target language.

Existing Methodologies

In general there are two levels in the translation process. Metaphrase is word to word translation and paraphrase is related to dynamic equivalence. Different approaches are as follows.

Direct approach

In this approach source language text is directly translated in to target text without translated into intermediate representation. Anusaarka is a machine translation system based on direct approach. **Hindi to Punjabi MT system (2009)** et al. [2], Goyal and Lehal of Punjabi University, Patiala, developed a Hindi to Punjabi Machine translation system based on direct approach word-to-word translation approach (Goyal *et al*., 2009; Dwivedi *et al*., 2010). The system consists of the various modules: pre-processing, a word-to-word Hindi-Punjabi lexicon, morphological analysis, word sense disambiguation, transliteration, and post-processing. They also have developed a system with a 95% accuracy.

Rule based translation

In this approach linguistic rules are build considering both the languages source language and target language . RBMT systems is based on linking the structure of the given input sentence with the structure of the target output sentence, preserving their unique meaning. It use large collection of manually developed rules used for mapping source language into target language text. These can be

edited to improve translations. Anglabharati and Anubharti is rule based machine translation system from English to Hindi and another Indian Languages.

Types of RBMT
 Direct System : Map input to output with basic rules. Directly word to word translation is there.
 Transfer RBMT System : it is a 2nd generation based translation.Employ morphological and syntactical analysis.
 Interlingual RBMT System : it is considered into 3rd generation translation.Use an abstract meaning.it creates a interlingual homogeneity .UNITRAN is a translation system which is based on interlingual rule based system.

K.K. Batra , G.S. Lehal , Automatic Translation System from Punjabi toEnglish for Simple Sentences in Legal Domain et al. [3], The system has been developed to translate simple sentences in legal domain from Punjabi to English. Since the structure of both the languages is different, so direct approach of translation in which word by word is translated,is not possible. So, another approach indirect approach i.e. rule based approach of translation is used. The system has basic three components analysis, translation and synthesis component. The steps involved are pre processing, tagging, ambiguity resolution, phrase chunking, translation and synthesis of words in target language.

Example Based Machine Translation
Example based machine translation is based on the idea of analogy. Concept of Analogy was firstly introduced by Makato Nagao in 1981.Basic idea is to reuse the existing examples as the basis for the new translations. Example-based machine translation systems are trained from bilingual parallel corpora , which contain sentence pairs. Sentence pairs contain sentences in one language with their translations into another language.An Example based machine translation system a set of sentences is given in source language and corresponding translation of each sentence in target language is done through the point to point mapping. Example based machine translation is used to translate similar type of sentences.
EBMT is done in to three stages
Matching
Alignment
Recombination
R.M.K. Sinha and A. Jain, AnglaHindi: et al. [4], An English to Hindi Machine-Aided Translation System called AnglaHindi which is an English to Hindi version of the ANGLABHARTI translation

methodology with a mixture of some example based translation methodology.

Statistical Machine Translation (SMT)
In 1949 warren weaver introduced the concept of SMT.Statistical machine translation statistical methods are applied to generate results using bilingual corpora. statistical word based, statistical phrase based and statistical syntax based models are three different models of statistical machine translation.
Advantages of EBMT over SMT
- EBMT works with a small set of data.
- Decodes more quickly as compared with another approaches

Dictionary Based Machine Translation
In dictionary based machine translation, dictionary entries are considered. This method is used for the translation of phrases not for the sentences.

Context Based Machine Translation
Context based method is corpus based method .It does not require a parallel corpora or rules for performing translation.It just requires a extensive monolingual target based corpus and a bilingual dictionary.

Conclusion
In this paper author represent the various machine translation approaches. Source language can be converted into another language or in target language with the use of machine translation are discussed. But none of the mentioned approaches gave a complete dynamic equivalent result of the source text in to a target text. Language is progressive in nature so it would be very difficult to conclude that a single approach can be sufficient. So development is still going on to create a complete system which would generate a sufficient results

Reference
[1] K. Deep, Dr. V. Goyal "Hybrid Approach for Punjabi to English Transliteration System," International Journal of Computer Applications (0975 – 8887) Volume 28–No.1, August 2011.
[2] Goyal, V., & Lehal, G. S. (2009). Evaluation of Hindi to Punjabi Machine Translation System. IJCSI International Journal of Computer Science, 4(1), 36-39.

[3] *K.K. Batra , G.S. Lehal , Automatic Translation System from Punjabi toEnglish for Simple Sentences in Legal Domain*

[4] *R.M.K. Sinha and Ajay Jain, AnglaHindi: An English to Hindi Machine Translation System, MT Summit IX, New Orleans, USA, Sept.23-27, 2003.*

[5] *Latha R. Nair and David Peter S., Machine Translation systems for Indian Languages, IJCA, Feb 2012.*

[6] *Cheragui M.A., "Theoretical Overview of Machine Translation", Proceedings ICWIT 2012.*

[7] *Kamaljeet Kaur Batra, and G S Lehal ,Rule Based Machine Translation of Noun Phrases from Punjabi to English.*

[8] *R.M.K. Sinha and Anil Thakur, Divergence Patterns in Machine Translation between Hindi and English, 10th Machine Translation summit (MT Summit X), Phuket,Thailand, September 13-15, (2005), 346-353.*