A Result Analysis of Translation Techniques of English to Hindi Online Translation Systems

Ekta Gupta
Student
Samrat Ashok Technology Institute
Vidisha (M.P.) India

Shailendra Kumar Shrivastava, PhD
Associate Prof.
Samrat Ashok Technology Institute
Vidisha (M.P.) India

ABSTRACT
If the term "translation process" makes us consider of dictionaries, grammar rules, and debates about linguistic details, we’re absolutely not alone. However, the translation process does not begin or end with transferring information from one language into another. In developing countries like Asian country and India where English is mainly half-hour recognize there automatic computational linguistics systems in education, analysis and industrial activities of very necessary role. Asian country has state a large assembly in Hindi is that the language you speak and in a very range of areas it works in all types of study and official. These days many on-line translator technologies use fully different computational linguistics approach. Like every translation approaches fully different characteristics, the result of the explanation would take issue. The purpose of this study is to create understanding about the different performance of the two online translation services due to the same actions they have. The experiment designed is meant show how the two online translation services have its have advantages and drawbacks which can affect their performance.

Keywords
Translation Quality Analysis, Translation Quality of Online Translation System for English to Hindi Translation.

1. INTRODUCTION
Machine Translation mainly deals with transformation of one language to another. Coming to the MT scenarios in India, it has huge scope due to many local languages of India. It is pertinent hat majority of the population in India are flowing in local languages such as Hindi, Punjabi etc. Given such a scenario, MT can be used to provide an interface of regional language. Machine translation is that the name for processed methods that automatize all or a section of the strategy of translating from one language to a unique. Throughout a huge multilingual society like India, there is great requirement for translation of documents from single language to a different language. As most of the higher study material, research journals and different commonplace communication tools are in English; these materials are to be translated into Hindi or particular local languages to possess an applicable higher study and communication with the individuals.

Machine Translation is a sub-field of computational linguistic that investigates the use of computer software to translate copy or speech from single natural language to another. At its basic level, MT performs simple exchange of words in single natural language for words in another. Current machine translation software often allows for customization by field or profession, improving output by limiting the scope of allowable substitutions. This technique is mainly effective in domain where formal or standard language is used it follows that machine translation of government and authorized documents more willingly produces usable output than conversation or less standardized text. Advanced study within the field of artificial intelligence and Computational Linguistics created a promising development of translation technology. This helped within the development of usable computational linguistics Systems in certain well-defined domains. Totally prime quality automatic computational linguistics system to induce could be a difficult task several organizations like Google, Microsoft, IBM, and lots of different etc. are engaged in development of MT systems. The technology is success new heights, right from idea of ideas up to the sensible implementation. It's necessary, that equal emphasis is place to remove the language divide that causes communication gap among totally different sections of societies. Natural language process (NLP) is that the field that strives to fill this gap. Machine translation (MT) principally deals with transformation of 1 language to a different. Coming back to the MT scenarios in Republic of India, it's huge scope because of several regional languages like Hindi, Punjabi etc. MT will be used to offer an interface of regional language.

2. THEORY
Machine Translation system are required to translate literary works that from any language into native languages. The literary work is fed to the MT system and translation is completed. Such MT system will break the language barriers by creating offered work rich sources of literature available to people across the globe.

MT additionally overcomes the technological barrier. Most of the data offered is in English that is known by only 3rd of the population. This has led to digital divide within which only small section of society will perceive the content presented in digital format. MT will facilitate during this regard to overcome the digital divide.

The following are four kinds of machine translation system:

MT for Watcher (MT-W)
MT for watcher is meant for readers WHO needed to achieve access to some data written in foreign language WHO also are ready to accept potential bad ‘rough’ translation instead of nothing. This was the kind of MT envisaged by the pioneers. This came in with the requirement to translate military technological documents.

MT for revisers (MT-R)
MT for revisers aims at producing raw translation mechanically with a high quality with a high quality similar to that of the primary drafts made by human. the translation output may be thought of only as brush-up so the professional translator may be free of that boring and time consuming task.
MT for translators (MT-T)
MT for translator’s aims at serving to human translators do their job by providing on-line dictionaries, synonym finder and translation memory. this kind of machine translation system is sometimes incorporated into the translation work stations and also the computer based translation tools.

MT for Authors (MT-A)
MT for authors aims at authors desperate to have their texts translated into one or many languages and acceptable to write under control of the system or to help the system clear up the vocalization so satisfactory may be obtained without any revision.

3. METHOD
3.1.Statistical Machine Translation
In this the input is considered as a distorted version of the target language sentence and the task is to find the most likely source language sentence giving the translation.

The task involves three steps:
1. Estimating the language probability P(t)
2. Estimating the translational model probability p(s/t)
3. Devising an efficient search for the target text that maximizes their product.

We have to find the sentence T for which p(s t) is maximum.

\[ P(s,t)=\text{arg max } p(s,t)=\text{arg max } p(t)p(s/t) \] (1)

In the above model „s“ is the source language sentence and „t“ is the target language sentence. The probabilities are to be calculated from the parallel corpus. Smoothing techniques are required for handling data sparsity problem that occurs in any noisy channel model.

3.2. Example based Machine translation system
An Example based Machine translation system (EBMT) system maintains a corpus consisting of translation examples between source and target languages. An EBMT system has two modules: Retrieval module and an adaptation module. The retrieval module retrieves a similar sentence and its translation from the corpus for the given source sentence. The advertisement module then adapts the retrieved translation to get the final corrected translation.

Consider the English to Hindi translation for the following sentence.

“Rama sings a song”

The retrieval module retrieves the following sentence and its translation from the corpus from a list of approximately matching sentences. It uses some similarity measures based on word similarity or syntactic and semantic similarity to identify this set of approximately matching sentences. From these the system selects the sentence with closest match with the input sentence.

If the system selects “Rohit sings a song” and its translation “Rohit geet gaati hai” as the closest one, it replaces Rohit with Rama and gaata with gaathi and finally forms the translation.

“Rama geet gaati hai”.

Here the adaptation is required to replace the word and suffix replacements. This method may not work in case of translation divergence where structurally similar sentences of the source language get translated into a different structure.

3.3. Language model
A language model gives the probability of a sentence. The probability is computed using n-gram model. Language model can be considered as computation of the probability of single word given all of the words that precede it in a sentence. The goal of statistical machine translation is to estimate the probability of a sentence. A sentence is decomposed in to the product of conditional probability.

By using chain rule, this is made possible as shown in equation 2. The probability of sentence P(S), is broken down as the probability of individual words P(W).

\[ P(S) = P(W_1,W_2,W_3,...........,W_n) \]
\[ = P(W_1)P(W_2|W_1)P(W_3|W_1W_2)P(W_4|W_1W_2W_3)........P(W_n|W_1W_2...W_{n-1}) \]

In order to calculate sentence probability, it is required to calculate the probability of a word, given the sequence of word preceding it. An n-gram model simplifies the task by approximating the probability of a word given all the previous words.

An n-gram of size 1 is referred to as a unigram; size 2 is a bigram (or, less commonly, a digram); size 3 is a trigram; size 4 is four-gram and size 5 or more is simply called a n-gram.

4. RESULT
This is algorithm which convert file to text of data

```java
//create file inputstream object to read data from file
FileInputStream fs=new FileInputStream(src);
```

```java
//create document object to wrap the file inputstream object
XWPFDocument docx=new XWPFDocument(fs);
```

```java
//create text extractor object to extract text from the document
XWPFWordExtractor extractor=new XWPFWordExtractor(docx);
```

```java
//create file writer object to write text to the output file
FileWriter fw=new FileWriter(desc);
```

```java
//write text to the output file
fw.write(extractor.getText());
```

```java
//clear data from memory
fw.flush();
```

```java
//close inputstream and file writer
fs.close();
fw.close();
```

```java
}catch(Exception e){e.printStackTrace();}
```
5. CONCLUSION
In this paper result analysis of Translation Techniques of English to Hindi Online Translation Systems which provides above results, that shows the complication of the previous ways ought to be increased. Our results show that the planned techniques considerably improve the translation quality. The performance of the two is also different and this has been proven in this study. I chose Google Translate and Bing Translator to represent English to Hindi translation. In this studies it is found that Bing Translator did a better job when translating English to Hindi and vice versa.

6. REFERENCES


