Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI:10.21474/IJAR01/17083
DOI URL: http://dx.doi.org/10.21474/IJAR01/17083

RESEARCH ARTICLE

COMPARATIVE STUDY OF MACHINE TRANSLATION TECHNIQUES

Brunda B.N, Mrs. Veena Potdar, Mrs. Lavanya Santhosh, Indu N. and Brunda N.C
Dept. of Computer Science and Engineering, Dr. Ambedkar Institute of TechnologyBangalore, Karnataka, India.

Manuscript Info

Manuscript History
Received: 15 April 2023
Final Accepted: 19 May 2023
Published: June 2023

Key words: -
Machine Translation, Statistical Machine Translation, Rule-based MT, NMT, BLEU scores, Comparison, Techniques, Example-based Machine Translation

Abstract

Machine translation (MT) is an approach to give knowledge and train computer to translate the sentence from one language to another language with the same fluent and meaning. MT is evaluated by the techniques that are giving the best translating accuracy. The four major techniques are Statistical Machine Translation, Rule-based Machine Translation, Example-based Machine Translation and Neural Machine Translation. We will see comparison and evaluation between these techniques in this particular paper with the examples and BLEU scores.

Copy Right, IJAR, 2023,. All rights reserved.

Introduction:-

Translation is necessary for the spreading new information, knowledge and ideas across the world. Communication is very important to communicate between two countries or between two different languages. In the process of spreading new information, translation is something that can change history. MT's task is to translate automatically by taking the input from one language and produce the output in another language. In this translation of text from one language to another, there is no human involvement and it is the machine which performs the process of conversion. At its basic level, MT performs a replacement of atomic words in a single characteristic language for words in another. The most complex translation can be done by using corpus method. Some software cannot be able to translate accurately as humans, but it will be possible in future. There are four types of Machine Translation techniques, statistical MT (SMT), rule-based MT (RBMT), example-based MT (EBMT), Neural MT (NMT). In simple language, these techniques are software to translate text from source sentence to target sentence. We define translation closeness by counting matches of n-grams in translation, that is BLEU scores. This BLEU score will give us the translation accuracy and closeness of translation. When evaluating a technique, it is important to consider how well phonetic typology differences are handled, how express acknowledgement is handled, and how idioms are translated. The comparison is done on the BLEU scores and accuracy of the translation in each technique and how the MT got better over the time and techniques.

Machine Translation

Machine Translation involves the process of conversion of text automatically in one form of natural language to another language. The goal of machine translation (MT) is to automatically translate text from one natural language into another while maintaining the text's meaning and creating fluent content in the target language. The job of MT is to automatically translate by accepting input from one language and producing an output in another. There is no human interaction in this text translation from one language to another; the conversion is being handled by a machine. Machine translation systems use four different techniques: neural, statistical, and rule-based. A common approach is rule-based, which combines language and grammar with the use of dictionaries. The development of an entire machine translation pipeline is the main goal of this endeavour.

Corresponding Author:- Brunda B.N
Address**:-** Dept. of Computer Science and Engineering Dr Ambedkar Institute of Technology Bangalore, Karnataka, India.

A. Level 1- Meta-phrase
Meta-phrase is literal translation, i.e., “word by word and line by line” translation.
B. Level 2- Para-phrase
Para-phrase may not translate the word-to-word, but it has the original text, that has the main point for translation. The main point of original text is “dynamic equivalence”.

Machine Translation Techniques
MTuses computer software to translate from one source sentence to target sentence. There are four types of MT techniques, statistical MT (SMT), rule-based MT (RBMT), example-based MT (EBMT) Neural MT (NMT). Let’s discuss them in detail.

Statistical Machine Translation
In contrast to rule-based techniques and example-based machine translation, statistical machine translation is a tool primarily utilised for machine translation. On the basis of statistical models whose parameters are determined from the analysis of bilingual text corpora, translations are carried out. Statistical Systems are typically not a specific language pair.
Example 1: “could you recommend another hotel”
Word alignment Based Statistical Translation represents bilingual correspondence by the notion of word alignment A, allowing one-to-many generation from each source word. A is an array for target words describing the indexes to The Input words.

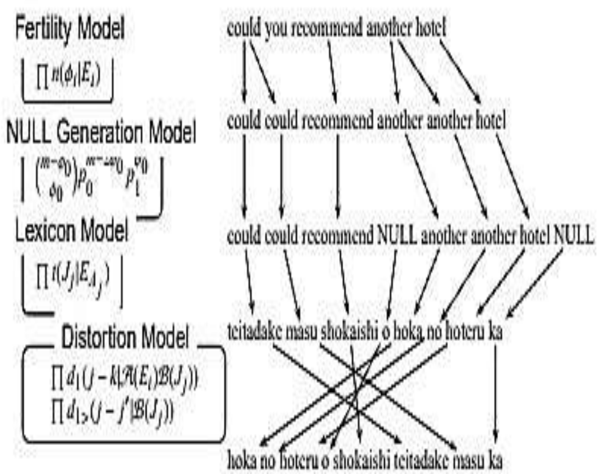

Fig 1:- Example 1.

Example 2: she gave him a watch.

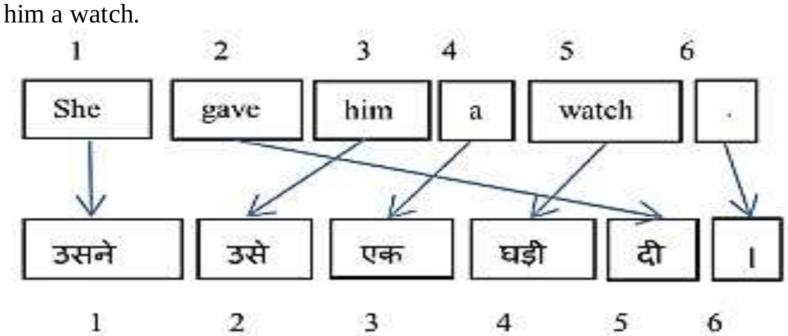

Fig 2:- Example 2.

For mapping the words present in source language Sentence into the equivalent words in target language sentence a word dictionary is used in this model. For example, if We Want to translate an English sentence ‘she gave him a watch' into Hindi, word-based translation will be as given in figure. Word Based Models are depends on lexical translation probability distribution. To put it formally, we Want to find a function,

pe: h → pe(h)

Which Represent how likely that Hindi translation h is, given a English word e. The probability of translating an English word 'e' into a Hindi word 'h' can be represented with the conditional probability function t(h|e), called as translation probability.

Example 3: She lives in the biggest house in the biggest village.
Given a corpus of target language texts, It IS Positively Possible to create a Modelof the Output Language. Based on the distribution of single Words in the corpus. Such a model is called a unigram model. By way of illustration, if we had a tiny corpus consisting of just the single 10-word sentence in (1). Here we make simple observations such as: 'she' occurs once in 10 words, while 'in', 'the' and 'biggest' each occur twice in 10 words. We move from these observations of frequency to statements of probability and say that The P of 'she' occurring is one in ten (or 0.1),The of P'in' is two in ten (or 0.2), and so on. These P can later be applied to as yet unseen strings.

Example 4: The house is small
The Language Model Typically Does Much MoreThan just enable fluent output. It Support in word Re-Ordering and word Alignment. For instance, a probabilistic language model pLM should prefer correct word order to incorrect word order: pLM(the house is small) >pLM(small the is house)
Formally, Language Model is a function that takes an English sentence and returns the probability that it was produced by an English speaker. According to the example above, it is more likely that an English speaker would utter the sentencethe house is smallthan the sentence small the is house. Hence, a good language model pLM assigns a higher probability to the first sentence.

Example 5:ಸಾಂವಿಧಾನಿಕನೈತಿಕತೆಸಹಜಭಾವನೆಯಲ್ಲ

It will take input as Kannada word and produce the output as English word. In order to match the correct answers it will check all the Probability of each sentence like

There are several sentence like
P -Constitutional morality is not a natural sentiment
-Morality constitutional is not a natural sentiment
-Is not a Constitutional Morality natural sentiment
Natural sentiment is not a morality Constitutional

These sentence are checked by Decorder and In order to Match the correct sentence or output to the input sentence it will check each sentence and check each Probabilty.

If it matches in high Probability it will take that sentence and Produce the correct answers

Example: 6

Example: source sentence:ಆದಿಯಲ್ಲಿದೇವರುಆಕಾಶವನ್ನೂಭೂಮಿಯನ್ನೂಸೃಷ್ಟಿಸಿದನು.

Target sentence: In the beginning God has created Earth and heaven

The word arrangement has a direct impact on meaning of a sentence in English. Ex: India beat Pakistan any change in the order leads to change in meaning, but Kannada sentence follow free form ordering in which inflections plays a guiding role in specifying meaning of a sentence, In Kannada inflection are attached to the root word leads to a greater number of unique words compared to English corpus.

TRANSLATION EXAMPLES GENERATED BY SMT AND NMT SYSTEMS.

SRC	去年 , 珠 三角 高 新 技术 产业带 的 产品 出口 合计 达 219.9亿 美元 , 占 全 省 高 新 技术 产品 出口额 的 98.7% 。
SMT	last year , the exports of high and new technology industrial belts in the prd reached 219.9亿 us dollars , were 98.7% percent of the province 's exports of high and new technology products.
NMT	last year , the export of high - tech and high - tech industrial belts in the prd region reached UNK billion us dollars , accounting for UNK percent of the total value of new high - tech products.
SRC	今年 前 两 月 广东 高新 技术 产品 出口 37.6亿 美元
SMT	37.6亿 us dollars during the first two months of this year guangdong high technology exports
NMT	hong kong exports UNK billion dollars in the first two months of this year.
SRC	拉斯穆森 说 : 我们 同样 认为 通过 对话 宽容 来 解决 事端 , 而 不 是 暴力 。
SMT	rasmussen said : we also believe resolved through dialogue tolerance , instead of violent incidents.
NMT	UNK said : " we also believe that we can resolve the incident through dialogue , but not violence.

Fig 3:- Example 5.

BLEU

COMPARISON WITH PREVIOUS WORK ON WMT ENGLISH-TO-GERMAN TRANSLATION. THE BLEU SCORES ARE CASE SENSITIVE.

SYSTEM	BLEU
Existing end-to-end NMT baseline [23]	16.46
WMT'14 winner system Phrase-based SMT [48]	20.67
Our work	
Moses	18.06
RNNSEARCH	16.98
+SMTrec/gate	18.22
+SMTrec/gate, UNK Replace	19.21

Fig 4:- BLEU scores.

Representing the Improvement during Translation. The Neural machine translation always take the input sentence which is source sentence and and produce the output sentence which is target sentence. There is always two source words which is single words cannot translate in NMT. As U can c in the Table , one type is not from the vocabulary, The other word is not appeared in the Corpus. Our method is to selecting proper SMT word Recommendation. We can see the most UNK words are numbers and entities. A correct translation can give good BLEU Score for every Translation as it simultaneously improves 1-gram/2gram/3-gram/4-gram BLEU scores.

a) Types of Statistical machine translation

1. Hierarchical phrase-based translation:
Example: “भारतकापर्ाानमतर्ीी”
HPBSMT, on the other hand, uses subphrases to remove issues associated with phrase based MT.
For e.g., “भारतकापर्ाानमतर्ीी ”
{bhaaratakaapradhaanamantrii}
should translate to “Prime Minister of India”. A possible grammar rule, in this case, is that the phrases on either side of the word of will be swapped when translating to Hindi

2. Syntax-based translation:
Syntax is the order or arrangement of words and phrases to form proper sentences. The most basic syntax follows a subject + verb + direct object formula.
Example: Jillian hit the ball
That is, "Jillian hit the ball." Syntax allows us to understand that we wouldn't write, "Hit Jillian the ball."

3. Phrase-based translation:
Phrase based translation models allow lexical entries with more than one word on either the source language or target language side
Example: we might have a lexical entry (lechien, the dog) specifying that the string lechien in French can be translated as the dog in English. The option of having multi word expressions on either the source or target language side is a significant departure from IBM models1 and 2, which are essentially word-to-word translation models. i.e., they assume that each French word is generated from a single English word.

4. Word-based translation:
In word-based translation, the fundamental unit of translation is a word in some natural language.

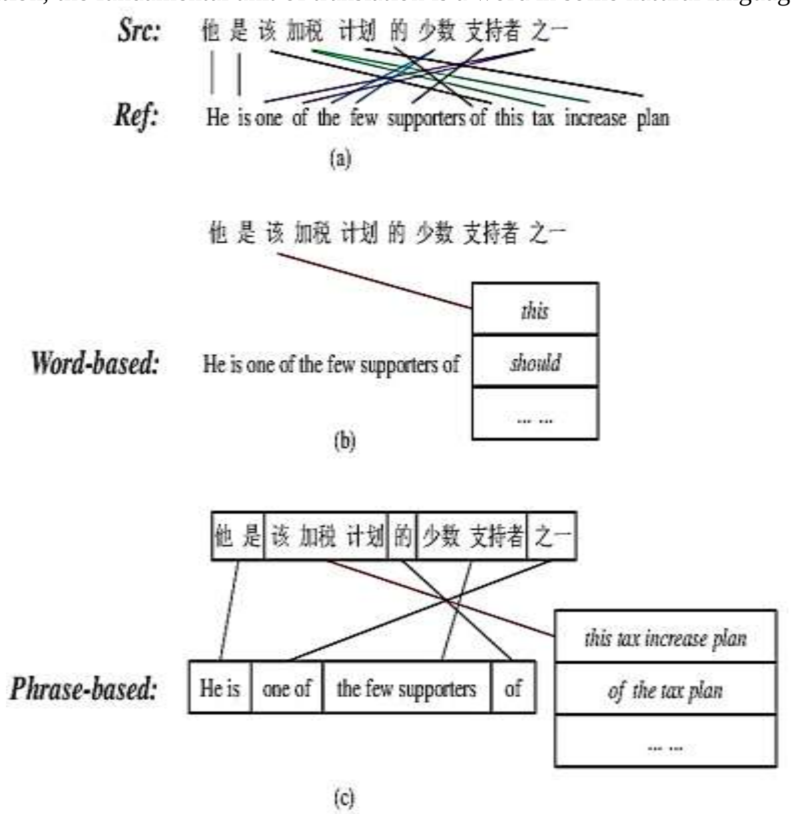

Fig 5:-Translation.

An example comparing the word-based translation and the phrase-based translation.
(a) Shows the Chinese source sentence and English reference.
(b) Illustrates a decoding phase of the word-based model translating the third Chinese word.
(c) Demonstrates a decoding stage of the phrase-based model translating the second Chinese phrase.

b) Algorithm

SMT is Based on Machine Translation. The input sentence is T in One language (that is German) to be translated on ELM(Extreme learning machine) and MERT(minimum error rate training) is Proposed Given a sentence **T** in one language (German) to be translated into another language (English), it considers **T** as the target of a communication channel, and its translation **S** as the source of the channel. Hence the machine translation task becomes to recover the source from the target. Basically, every English sentence is a possible source for a German target sentence. If we give the If a Probability P(S|T) to each Pairs of sentence (S,T), them the Problem of the translation is to find the source S for a given target T, such that P(S|T)is the maximum. According to Bayes rule.

$$P(S\,I\,T) = \frac{P(S)P(T\,I\,S)}{P(T)} \qquad (1)$$

Since the denominator is independent of S, we have

$$S = \underset{S}{\operatorname{argmax}} P(S)P(T \mid S) \qquad (2)$$

S Therefore a statistical machine translation system must deal with the following three problems:

• Modeling Problem: How to depict the process of generating a sentence in a source language, and the process used by a channel to generate a target sentence upon receiving a source sentence? The former is the problem of language modeling, and the later is the problem of trans- lation modeling. They provide a framework for calculating **P(S)** and **P(T | S)** in (2).

• Learning Problem: Given a statistical language model **P(S)** and a statistical translation model **P(T | S)**, how to estimate the parameters in these models from a bilingual corpus of sentences?

• Decoding Problem: With a fully specified (framework and parameters) language and translation model, given a target sentence T, how to efficiently search for the source sentence that satisfies

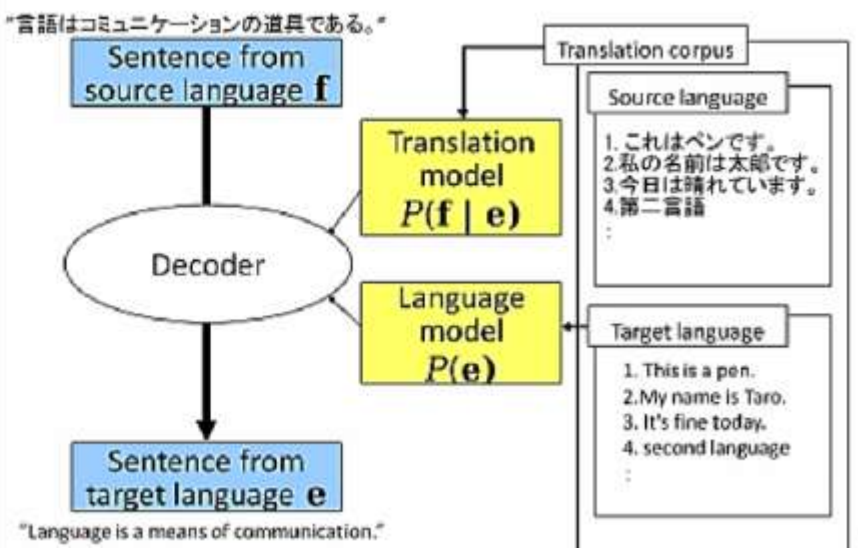

Fig 6:- SMT Framework.

c) Re-ranking in Statistical Machine Translation

Re-ranking is a significant component for effectiveness and efficiency in SMT. With consideration of cognitive computation, a novel nonlinear re-ranking model called SSCR (Scaled Sorted Classification Re-ranking) based on ELM (Extreme learning machine) and MERT (minimum error rate training) is proposed.

d) Issues in statistical machine translation

In PBSMT following two issues are extremely important for getting a high-quality translation.

a) The word alignment problem ensures the correspondence between the phrases of the two languages.
b) The Language modelling problem ensures fluency in generated output target language sentences.

Languages	Training dataset	Test dataset	BLEU Score
English-Romanian	5672	1245	18.97
English-French	14784	5890	36.14
English-Korea	4563	986	16.87
English-Russian	6000	2407	22.01
English-Gujarati	1,55,767	3149	9.8
English-Japanese	500,000	150,000	24.74
English-Hindi	28,927	10000	14.04
English-Chinese	62,878	16,683	49.50
English-Chinese	17.970	1,000	17.24

Table 1:- Comparison of BLEU scores for different examples.

1. Rule-based Machine Translation

RBMT utilizes different algorithms depending on the specific task or component of the translation process. Some commonly used algorithms in RBMT include:

The Rule-based Parsing Algorithm utilizes parsing methods, such as top-down or bottom-up parsing, to examine the syntactic structure of the given input sentence. It employs a combination of grammar rules and parsing algorithms to generate a parse tree that depicts the sentence's underlying structure

Example 1**:** Input Sentence: "The cat is on the mat"
The rule-based parsing algorithm would analyze the sentence based on predefined grammar rules and generate a parse tree that represents the sentence's syntactic structure.

The Rule-based Translation Algorithm operates by employing a collection of linguistic rules to execute the translation procedure. These rules delineate the appropriate transformations and translations of various linguistic components, including words, phrases, and grammatical structures.

Example 2: Input Expression: "I am eating an apple" Translation Rules:
"I" translates to "Je" (French)
"am" translates to "suis" (French)
"eating" translates to "mange" (French)
"an apple" translates to "une pomme" (French)
The rule-based translation algorithm applies these rules sequentially to each component of the input expression, generating the translation "Je suis en train de manger une pomme" (French). Example based machine translation

relies on a large database of bilingual translation examples. When translating an input expression, the system searches for the most similar example in the database and generates a translation based on that example. This method is beneficial because it can capture the nuances of language usage and context. However, one challenge is finding appropriate examples that closely match the input expression.
Example 3: Input Expression: "Je suis fatigué" (French)

Translation Example: "I am tired" (English)
In example-based machine translation, if the system has a translation example in its database that closely matches the input expression, it can retrieve that example and generate the translation accordingly. However, if the database lacks an appropriate example, it may struggle to produce an accurate translation.

A. Algorithm
RBMT utilizes different algorithms depending on the specific task or component of the translation process. Some commonly used algorithms in RBMT include:

Rule-based Parsing Algorithm: This algorithm applies parsing techniques, such as top-down parsing or bottom-up parsing, to analyze the syntactic structure of the input sentence. It uses grammar rules and parsing algorithms to generate a parse tree representing the structure of the sentence.
Example 1: Input Expression: "Je suis fatigué" (French)
Rule-based translation would involve applying linguistic rules to translate the input expression. For instance:

French-English

"Je" translates to "I"
"suis" translates to "am"
"fatigué" translates to "tired"

By applying the predefined rules, the rule-based system can generate the translation "I am tired" without relying on a large database of examples. However, if there are complex linguistic structures or exceptions not covered by the rules, the translation may not accurately capture the intended meaning.

Example 2: Input Expression English input: "Hi, how are you?"

English-Spanish

"Hi" can be translated as "Hola"
"how" can be translated as "cómo"
"are" can be translated as "está"
"you" can be translated as "usted"

2. Example Based Machine Translation
Example-based machine translation relies on a large database of bilingual translation examples. When translating an input expression, the system searches for the most similar example in the database and generates a translation based on that example. One advantage of this method is its ability to capture the subtleties of language usage and context. However, a challenge lies in finding suitable examples that closely align with the input expression.
Example: Input Expression: "Je suis fatigué" (French)
Translation Example: "I am tired" (English)

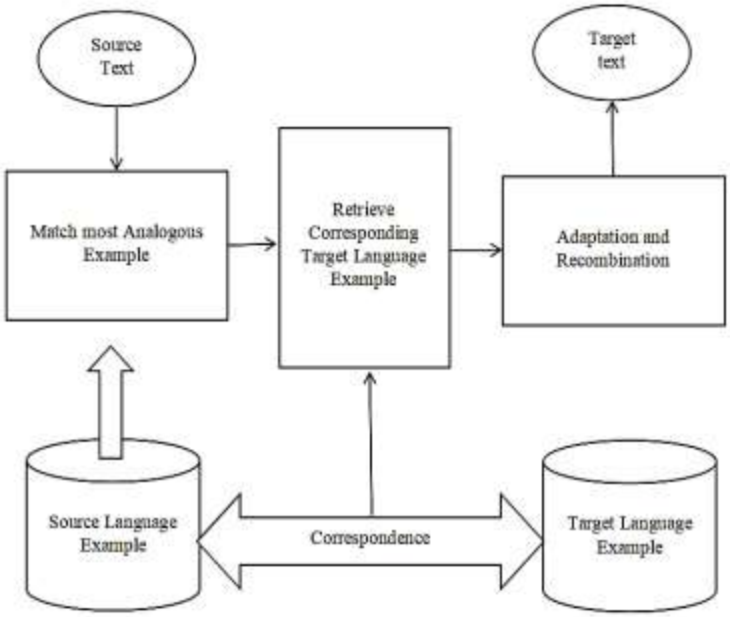

Fig 1 EBMT Architecture

The knowledge base, parallel aligned corpora consist of two sections, one for the source language examples and the other for the target language examples. Each example in the source section has one to one mapping in the target language section.

A. Algorithm

Example-Based Machine Translation (EBMT): One common algorithm used in EBMT is the nearest neighbor algorithm, which finds the most similar example in the bilingual database to the input expression. This algorithm calculates the similarity between the input expression and each example in the database, typically using measures such as edit distance, cosine similarity, or word alignment models. The example with the highest similarity score is selected as the translation reference.

Example 1: Input Expression: "I want to go home"

Translation Example 1: "Je veuxrentrer chez moi" (French)

Translation Example 2: "Quieroir a casa" (Spanish)

In this example, the nearest neighbor algorithm would compare the input expression to the translations in the database. It calculates the similarity scores for each example and determines which translation (French or Spanish) is the closest match to the input expression.

Example 2:

1. The main objective of the EBMT (Example-Based Machine Translation) system is to accurately translate an input sentence by leveraging the aligned data available in its training corpus.

Ich bin heuteglücklich ↔ I am happy today

Je vis à Paris avec ma femme ↔ I live in Paris with my wife

Il y a beaucoup à faire à Paris ↔ There's a lot to do in Paris

B. Comparison between RBMT and EBMT

Basic	RBMT	EBMT
Computational Cost	High	Low
Improvement Cost	High	Low
System Building Cost	High	Low
Context Sensitive Translation	Needs another understating device support	Architecture inbuilt

Robustness	High	Low
Measurement of reliability factor	No	Yes
Example Independency	No	Yes

C. Advantages and Disadvantages

Example: English: "I am happy." Kannada (RBMT translation): "ನಾನುಸಂತೋಷವಾಗಿದ್ದೇನೆ" (Nānusantōṣ avāgiddēne)

In this example, EBMT relies on the closest matching translation example from its database to generate the Kannada translation, while RBMT follows predefined linguistic rules to produce the translation.

MT Approach	Advantages	Disadvantages
Rule-Based	1. Easy to build an initial system 2. Based on linguistic theories 3 Effective for core phenomena 4.Better choice for domain specific translation 5.The quality of translation is good for domain specific systems	1. Rules are formulated by experts 2. Difficult to maintain and extend 3.Ineffective for managerial phenomena 4. The number of rules will grow drastically in case of general translation systems
Example Based	1. Extracts knowledge from corpus 2. Based on translation patterns in corpus 3. Reduces the human cost	1. Similarity measure is sensitive to system 2. Search cost is more 3. Knowledge acquisition problem still persists
Statistical	1. Does not consider language grammar for translation 2. Extracts knowledge from corpus 3. Reduces the human errors 4. Model is mathematically grounded	1. No linguistic background 2. Search cost is expensive 3. Hard to capture long distance phenomena 4. Require huge amount of parallel corpora 5. The translation quality will be very coarse due to lack of sufficient corpora

D. Comparison:

SMT translation: "ನಿಮ್ಮಕನಸುಗಳನ್ನುನಿಜವಾಗಿಸಿ" (Nimmakanasugalanunijavāgisi)

RBMT translation: "ನೀವುನಿಮ್ಮಕನಸುಗಳನ್ನುನಿಜವಾಗಿಯೇನೆರವೇರಿಸಿ" (Nīvunimmakanasugalanunijavāgiyeneravērisi)

The SMT translation is generated based on statistical models, while the RBMT translation is selected from the bilingual translation database. The RBMT translation captures the idiomatic expression "live your dreams" more accurately, providing a natural and meaningful translation. Please keep in mind that the quality of translations can differ depending on the specific systems and resources employed for SMT (Statistical Machine Translation) and RBMT (Rule-Based Machine Translation) BLEU Score.

BLEU is a metric for evaluating machine translation quality by comparing it to reference translations. It assigns a score between 0 and 1 based on word matching.

Languages	Training datasets	Testing Datasets	BLEU Score
Japanese - English	30000	2550	0.39
English - German	40000	3000	0.86
Hindi - English	25000	2000	0.7502
Sanskrit - English	33000	2500	0.7606
Kazakh - English	26078	1500	0.093
English - Kazakh	26078	1500	0.063
Kannada - English	20000	1000	10.68
Chinese - Spanish	14608	12080	0.0175

Table 1:- Comparison of BLEU scores for different examples.

3. Neural machine Translation

Neural machine translation (NMT) is an artificial neural network (ANN) approach to translate sentence from one language to another. Google Translate, Baidu translate are well-known example of NMT.

Technically, NMT includes all types of Machine Translation, where an ANN is used to predict a sequence of numbers when provided with a sequence of numbers. In the case of translation, each word in the input sentence is encoded as number to be translated by neural network into resulting sequence of numbers representing the translated target sentence.

Let' take a simplified example of an English to Kannada machine translation:

"It is a continuous process" is encoded into numbers 251, 252, 2, 3245, 892. The numbers 251, 252, 2, 3245, 892 are input into a neural translation model and results in output 8976, 9845, 6590. 8976, 9845, 6590 is then decoded into the Kannada translation "**ಇದುನಿರಂತರಪ್ರಕ್ರಯೆ**"

NMT model works via a complex mathematical formula. As mentioned above, this formula takes in a string of numbers as inputs and outputs a resulting string of numbers.

A. NMT improved in issues of SMT and RBMT

Attention mechanism: the issues in SMT was word alignment and Language Modelling Problem that is overcome by this mechanism.

Cross-Lingual Transfer: Issues in RBMT was depending on accurate dictionaries, grammars, and linguistic resources, difficulty with idiomatic expressions, limited adaption to new language patterns have been overcome by understanding full sentences rather than the words.

B. **Modeling**

Document, paragraph, or sentence-level are the different levels to translate the text. Here, we will discuss and translate sentence-level text in sequence. Now, the NMT model can be seen as a **sequence-to-sequence** model. Let us assume, input sentence is$X=\{X_1,\ldots, X_S\}$ and a output sentence $y=\{y_1,\ldots, y_T\}$. left-to-right(L2R) factorize can be distributed by using chain rule as:

$$P\left(\mathbf{y}|\mathbf{x} = \prod_{t=1}^{T} P\left(y_t \,|\, y_0, \ldots, y_{t-1}, \mathbf{x}\right)\right). \qquad (1)$$

the Eq. (1) is a equation for **L2R autoregressive** NMT, for the prediction at time-step **t** is taken as a input at time-step t+1.

Every NMT model uses **encoder-decoder framework**for their model creation. The four components of Encoder-Decoder framework are: Embedding layer, encoder-decoder networks and classification layer. We use ‘‘<bos>’’ and ‘‘<eos>’’ as special symbols that mark the beginning and ending of a sentence.

We use different colours to distinguish between the embedding layer, classification layer, encoder-decoder network in NMT architecture.

The embedding layer involves the topic of **continuous representation**. It uses different symbols x_t into a continuous vector $X_t$belongs to R^d, where **d** represents the dimension of the vector. The embeddings are then fed into other layers for grater feature extraction.

The source embeddings are translated by the encoder network into concealed continuous representations. The encoder's capacity to model the complex dependencies and ordering found in the source language is a prerequisite for learning expressive representations. For modelling variable-length sequences, recurrent neural networks (RNN) are a good option. The computation required by an encoder using RNNs can be summed up as follows:

$$\mathbf{h}_t = \mathrm{RNN}_{\mathrm{ENC}}\left(\mathbf{x}_t, \mathbf{h}_{t-1}\right). \qquad (2)$$

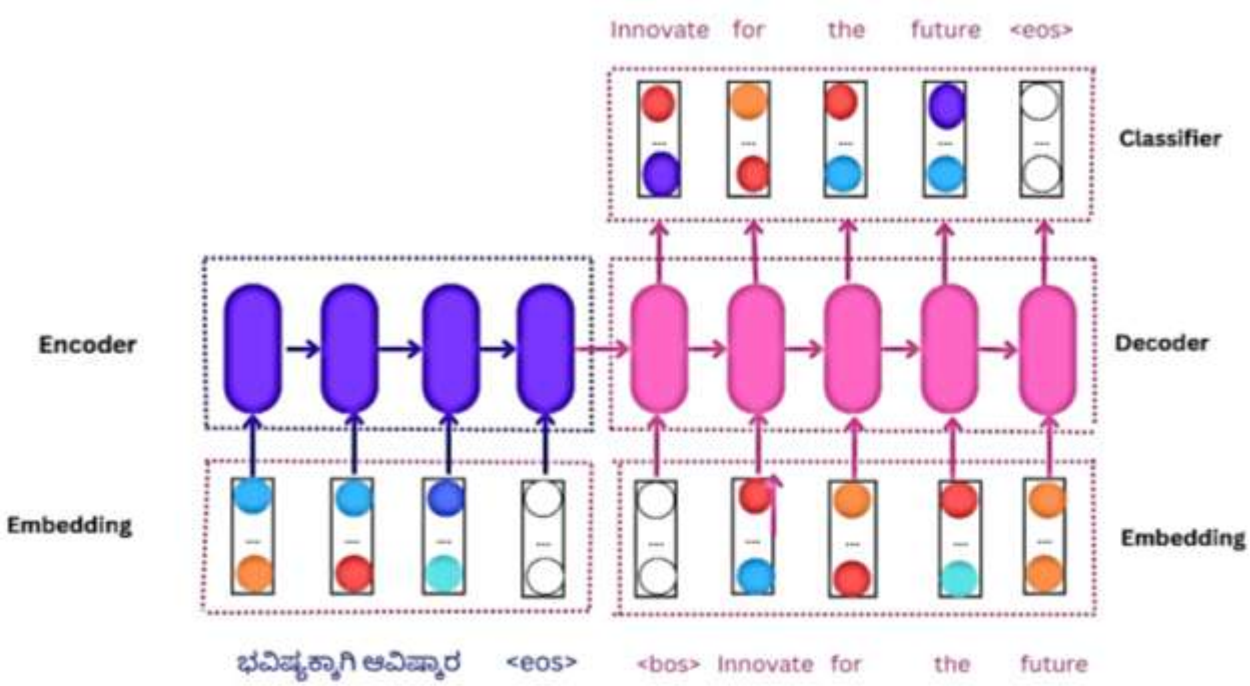

Fig 7:- Neural Machine Translation Framework.

By iteratively applying the state transition functionRNN_{ENC}over input sequence, we can use the final state h_s as the representation for the entire source sentence, and then feed it to the decoder.

The decoder network extracts the necessary information from the encoder output and models the long-distance dependencies between target words. Given the start symbol y_0=<bos> and the initial state $s_0=h_s$, the RNN decoder compresses the decoding history $y_0,...,y_t$-1 into a state vector. St belongs to R_d:

$$\mathbf{s}_t = \mathrm{RNN}_{\mathrm{DEC}}\left(\mathbf{y}_{t-1}, \mathbf{s}_{t-1}\right).$$

The target tokens' distribution is predicted by the categorization layer. It is customary for the classification layer to be a linear layer with a softmax activation function. assuming that |V| is the size of the vocabulary and V is the number of words in the target language. The classification layer initially uses the linear map to translate the decoder output St Rd from h to a vector z in the vocabulary space R|v|. The output vector is then checked to see if it is a valid

$$\mathrm{softmax}(\mathbf{z}) = \frac{\exp(\mathbf{z})}{\sum_{i-1}^{|V|} \exp(\mathbf{z}_{[i]})},$$

probability using the SoftMax function:

where we use $z_{[i]}$ to denote the **i**-th component in z.

C. Advantages of NMT

1. **Natural-sounding language**: NMT can make AI-translated language sound more human and less robotic.
2. Context: NMT isn't perfect, but it does a good job of interpreting **translation context**. Google made recent strides in training its NMT to pay greater attention to the **context of a sentence** within a body of work.
3. Generalization: NMT gets smarter over time. Like the human brain, it can generalize to make new conclusions and connections. It uses these connections to quickly learn different language pairs.
4. Good accuracy in some contexts: NMT is especially good at translating repetitive content that requires high accuracy like manuals, guides, or reference materials.

D. Disdvantages of NMT

1. Need for clarity in the source text: Source text needs to be very clear for NMT to generate a quality translation. Neural machine translation has difficulties with ambiguities, highly technical language, proper nouns, and rare words.
2. Poor translation of long sentences: NMT generally outperforms other machine translation methods, but there's an interesting exception. NMT can't translate long sentences very well.
3. Large data sets needed: For smaller projects, NMT isn't always a good fit. NMT systems produce comparably poor results with small data sets. Without a large amount of training data, NMT isn't going to give you accurate translations.
4. Expertise limitations: NMT can't perform well without the expertise and cognitive power of human architects and engineers. Any NMT model must be trained with a large quantity of linguistic data.
5. Human post-editing still needed: NMT text should always be checked by a human editor. Language interpretation requires a level of critical thinking and nuance that computers haven't achieved yet.

E. BLEU Scores

The BLEU scores are used to define the translation closeness and accuracy. Here we taken the Translation dataset and its BLEU scores from the different languages. In the below given table, we can observe that The BLEU score is more when the training dataset are more. Compared to other techniques SMT and RBMT, here, NMT giving greater translation closeness.

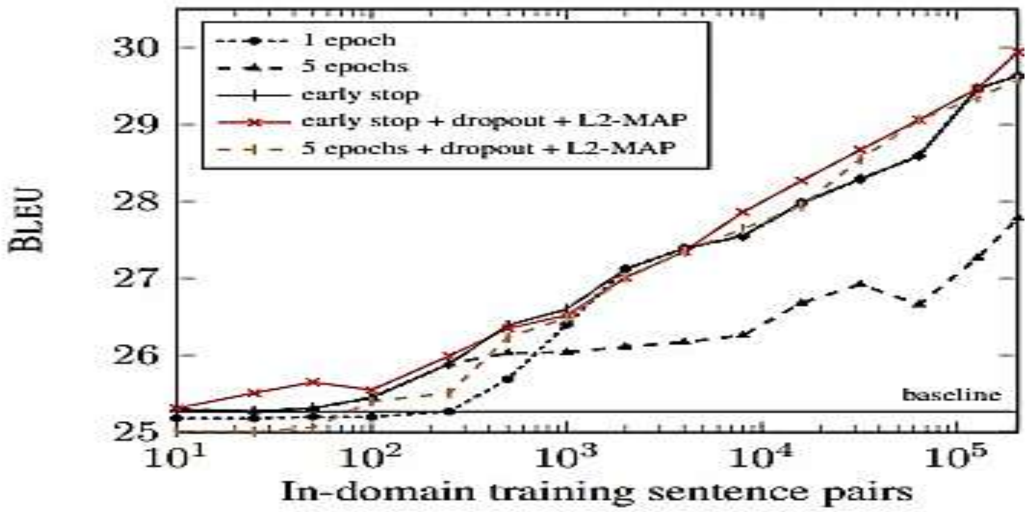

Fig 8:- BLEU scores based on the epochs and training dataset.

Languages	Training dataset	Test dataset	BLEU scores
English-Spanish	60,000	10,000	14.59
English-Spain	16.6 M	7,000	30.74
English-French	40,000	2000	38.10
English-Thai	30,781	1000	16.7
French-English	20,000	10,000	24.50
English-Persian	13,386	1000	3.40
English-Russian	45,786	2000	16.81
English-German	38,849	1750	27.76
Chinese-English	10000	2000	27.31
English-Kannada	10,000	8,000	

Table 3:-Table based on the different papers and different languages showing the BLEU scores for the training and dataset.

Conclusion:-

We observed the Comparison of SMT, RBMT and NMT by seeing the different experiment results. NMT has improved lot better than SMT and RBMT. However, the methodologies and techniques are still challenging with respect to the new words and word alignment.

We observed that SMT BLEU scores are ranges from 15-20, but with the sentence length of 3-5 and it doesn't support the large sentences. RBMT BLEU bounded between 0 and 1. NMT giving more BLEU scores with the larger sentences and complex sentences ranges from 14-30. To conclude NMT is better than other techniques with complex and difficult words and sentences.

References:-

[1] P.J. Antony, V.P. Ajith, and K.P. Soman. 2010."statistical methos for english to kannada translation" https://link.springer.com/chapter/10.1007/978-3-642-12214-9_57.V.V Das et al. (Eds.): BAIP 2010, CCIS 70,pp.356–362, 10.1007/978-3-642-12214-9_57.

[2] Arianna Bisazza,Marcello Federico. 2016,"A Survey of Word Reordering in Statistical Machine Translation: Computational Models and Language Phenomena". https://aclanthology.org/J16-2001.pdf.

[3] A. R. Babhulgaonkar, S. V. Bharad, 2017,"Statistical Machine Translation", 9781-5090-4264-7/17/$31.00 ©2017 IEEE , https://ieeexplore.ieee.org/document/812 2149.

[4] Kehai Chen , Tiejun Zhao, FEBRUARY 2018,"A Neural Approach to Source Dependence Based Context Model for Statistical Machine Translation", https://ieeexplore.ieee.org/document/8105847 IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 26.

[5] Jia Yan, Wang Chao, 2015,"English Language Statistical Machine Translation Oriented Classification Algorithm", HTTP://ieee-ims.org/publication/ieee-tim/authorshttp://dx.DOI 10.1109/ICITBS.2015.99 .

[6] Xing Wang, Zhaopeng Tu, and Min Zhang,2015,"Incorporating Statistical Machine Translation Word Knowledge into Neural Machine Translation", https://ieeexplore.ieee.org/document/8421063 ,DOI 10.1109/TASLP.2018.2860287, IEEE/ACM.

[7] Yan Liu , Chi Man Vong , Pak Kin Wong, 2017,"Extreme Learning Machine for Huge Hypotheses Re-ranking in Statistical Machine Translation", https://link.springer.com/article/10.1007/s12559-017-9452-x ,DOI 10.1007/s12559-017-9452-x.

[8] Aloka, FernandoGihan, DiasSurangika Ranathunga,2020,"Data Augmentation and List Integration for Improving Domain-Specific Sinhala-English-Tamil Statistical Machine Translation"https://arxiv.org/abs/2011.02821 .

[9] Rajitha M.D.C, Piyarathna L.L.C, 04-07 November 2020," Sinhala and English Document Alignment using Statistical Machine Translation", https://ieeexplore.ieee.org/abstract/document/9325462,https://doi.org/10.1109/ICTer51097.2020.9325462 .

[10] Alberto Poncelas, Gideon Maillette de Buy Wenniger and Andy Way, Volume 28 , Issue 1 , January 2022 , pp. 71 - 91," Improved feature decay algorithms for statistical machine translation", https://doi.org/10.1017/S1351324920000467 .

[11] Sitender, and Seema Bawa. "A Sanskrit-to- English machine translation using hybridization of direct and rule-based approach." Neural Computing and Applications 33 (2021): 2819-2838. DIO:https://doi.org/10.1007/s00521-02005156-3

[12] Chen, An-Pin, and Mu-Yen Chen. "An implementation of learning classifier systems for rule-based machine learning." In Knowledge-Based Intelligent Information and Engineering Systems: 9th International Conference, KES 2005, Melbourne, Australia, September 14-16, 2005, Proceedings, Part II 9, pp. 45-54. Springer Berlin Heidelberg, 2005.DOI https://doi.org/10.1007/11552451_7

[13] Evans, Richard. "A comparison of rule-based and machine learning methods for identifying non-nominal it." In Natural Language Processing—NLP 2000: Second International Conference Patras, Greece, June 2–4, 2000 Proceedings 2, pp. 233-240. Springer Berlin Heidelberg, 2000. https://doi.org/10.1007/3-540-45154-4_22

[14] Rios, Annette, and Anne Göhring. "Machine learning applied to rule-based machine translation." Hybrid approaches to machine translation (2016): 111-.https://doi.org/10.1007/978-3-31921311-8_5

[15] Sumita, Eiichiro, and Hitoshi Iida. "Experiments and prospects of examplebased machine translation." In 29th Annual Meeting of the Association for Computational Linguistics, pp. 185-192. 1991.DIO: 10.3115/981344.981368

[16] Carl, Michael. "Inducing translation templates for example-based machine translation." In Proceedings of Machine Translation Summit VII, pp. 250-258. 1999. https://aclanthology.org/1999.mtsummit1.37.pdf

[17] Costa-Jussà, Marta R., and Jordi Centelles. "Description of the Chinese-to-Spanish rulebased machine translation system developed using a hybrid combination of human annotation and statistical techniques." ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 15, no. 1 (2015): 113 https://doi.org/10.1145/3312575

[18] Islam, Md Adnanul, Md Saidul Hoque Anik, and ABM Alim Al Islam. "Towards achieving a delicate blending between rulebased translator and neural machine translator." Neural Computing and Applications 33 (2021): 1214112167. https://doi.org/10.1007/s00521-021-05895-x

[19] Sitender, & Bawa, S. (2020). A Sanskrit-toEnglish machine translation using hybridization of direct and rule-based approach. Neural Computing and Applications. doi:10.1007/s00521-020-05156-3

[20] Chen, A.-P., & Chen, M.-Y. (2005). An Implementation of Learning Classifier Systems for Rule-Based Machine Learning. Knowledge-Based Intelligent Information and Engineering Systems, 45–54. doi:10.1007/11552451_7

[21] Joint Training for Neural Machine Translation 2019 ISBN: 978-981-32-9747-0 https://doi.org/10.1007/978-981-32-9748-7

[22] Apaza H,Aruhuanca B,Nina M,Flores A,Silva C andTito E (2023). Neural Machine Translation for Native Language Aymara to English. Proceedings of the Future Technologies Conference (FTC) 2022, Volume 3. 10.1007/978-3-031-18344-7_40. (565576) https://link.springer.com/10.1007/978-3-031-18344-7_40

[23] Philipp Koehn and Rebecca Knowles. 2017.Six Challenges for Neural Machine Translation.In Proceedings of the First Workshop on Neural Machine Translation, pages 28–39, Vancouver. Association for Computational Linguistics http://dx.doi.org/10.18653/v1/W17-3204

[24] Joel Escudé Font and Marta R. Costa-jussà. 2019.Equalizing Gender Bias in Neural Machine Translation with Word Embeddings Techniques.In Proceedings of the First Workshop on Gender Bias in Natural Language Processing, pages 147–154, Florence, Italy. Association for Computational Linguistics. http://dx.doi.org/10.18653/v1/W19-3821

[25] Fei Z. Efficient Modeling of Future Context for Image Captioning. Proceedings of the 30th ACM International Conference on Multimedia. (5026-5035). https://doi.org/10.1145/3503161.3547840

[26] Roman Grundkiewicz and Kenneth Heafield. 2018.Neural Machine TranslationTechniques for Named Entity Transliteration. In Proceedings of the Seventh Named Entities Workshop, pages 89–94, Melbourne, Australia. Association for Computational Linguistics.http://dx.doi.org/10.18653/v1/W18-2413

[27] Antonio Valerio Miceli Barone, Barry Haddow, Ulrich Germann, and Rico Sennrich. 2017.Regularization techniques for finetuning in neural machine translation. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1489–1494, Copenhagen, Denmark. Association for Computational Linguistics. http://dx.doi.org/10.18653/v1/D17-1156

[28] Guo ZandHan D. (2023). Sparse co attention visual question answering networks based on thresholds. Applied Intelligence. **53**:1. (586-600). Online publication date: 1-Jan-2023. https://doi.org/10.1007/s10489-022-03559-4

[29] Marco Turchi, Tijl De Bie, Cyril Goutte, and Nello Cristianini. 2012." Learning to transalation: A statistical and computational analysis". https://www.hindawi.com/journal s/aai/2012/484580/. Hindawi Publishing Corporation Advances in Artificial Intelligence Volume 2012, Article ID 484580, 15 pages, https://doi.org/10.1155/2012/484580

[30] Stahlberg, Felix. "Neural machine translation: A review." Journal of Artificial Intelligence Research 69 (2020): 343-418. https://doi.org/10.1613/jair.1.12007.