International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

Improvements in Enhancing Bi-Lingual Machine Translation Approach

Manish Rana[1], Mohammad Atique[2]

[1]Research Scholar, Post Graduate Department of Computer Science & Engineering, Sant Gadge Baba Amravati University, Amravati, India

[2]Professor, Post Graduate Department of Computer Science, Sant Gadge Baba Amravati University, Amravati, India

Abstract: ***This paper shows the improvement carried in different phases of development made in "Enhancing Bi-Lingual Machine Translation Approach" . The work consists of three phases. The Initial phase, where the work was on the "Improvement of the Machine interpretation", with assistance of Example based Machine interpretation utilizing Fuzzy device, then second phase describe further improvements using Long-Short Term Memory Concept(LSTM) approach and Last phase describe utilizing Python and so forth. The proposed EBMT framework can be used for automatic translation of text by reusing the examples of previous translations through the use of Fuzzy which is proposed work.***

Keywords: NLP, Fuzzy Logic, Bi-Lingual, EMBT, LSTM etc

1. Introduction

This framework contains of 3 phases, matching, alignment and recombination.

Same type of machine translation is possible through use of soft computing tool (Fuzzy Logic). In second phase of proposed research take a shot at Enhancing Bi-lingual Machine Translation Approach utilizing LSTM is introduced. This part is composed into three areas. The Deep Neural Network (DNN) is an incredibly expressive model that can adapt profoundly complex vector-to-vector mappings. The Last phase describe the code created is a basic language interpreter which changes over English language into any client given Indian language. GOOGLE-API is utilized for this reason which contains all inclusive datasets different diverse dialects over the globe.

a) Prologue to Example based Machine Translation

Model based machine interpretation (EBMT) is a technique for machine interpretation regularly described by its utilization of a bilingual corpus with parallel messages as its fundamental learning base at run-time. It is basically an interpretation by similarity and can be seen as a usage of a case-based thinking way to deal with AI.

Example of bilingual corpus	
English	Hindi
How much is that red umbrella?	Ano akai kasa wa ikura desu ka.
How much is that small camera?	Ano chiisai kamera wa ikura desu ka.

Model based machine interpretation frameworks are prepared from bilingual parallel corpora containing sentence sets like the precedent appeared in the table above. Sentence sets contain sentences in a single language with their interpretations into another. The specific model demonstrates a case of an insignificant pair, implying that the sentences fluctuate by only one component. These sentences make it easy to learn interpretations of bits of a sentence. For instance, a model based machine interpretation framework would take in three units of interpretation from the above precedent:

How much is that X? Corresponds to Ano X wa ikura desu ka.
Red umbrella corresponds to akai kasa
Small camera corresponds to chiisai kamera

Making these units can be utilized to create novel interpretations later on. For instance, on the off chance that we have been prepared utilizing some content containing the sentences:

President Kennedy was shot dead amid the motorcade and the convict got away on July fifteenth. We could decipher the sentence. The convict was shot dead amid the procession. by substituting the suitable pieces of the sentences.

Model Based Machine Translation (EBMT), Harold Somers (1999) gives a far reaching grouping of the expansive assortment of MT look into falling inside the precedent based worldview, and makes an endeavor at catching the fundamental highlights that make a MT framework a precedent based one. The present paper accepts Somers' discourse as its beginning stage and endeavors to make further strides in addressing the inquiries presented in that [2]. We recognize now that we additionally draw vigorously from Somers' paper as far as references of past works in EBMT. In the expansive and enhanced display of MT, we trust that this definition task, a long way from being a punctilious exercise, is an imperative advance towards isolating basic contrasts among MT comes nearer from inessential ones. This exertion may prompt revealing covers between methodologies that at first sight appear to be very far separated, or on the other hand it might expose noteworthy contrasts between methodologies that are externally comparative. We trust that a superior comprehension of the relations among various methodologies gives significant knowledge that can control MT analysts in their choices about further bearings to take. [1].

In his evidently temporary decisions about a meaning of EBMT, Somers (1999) examines three progressively explicit criteria for characterizing EBMT:
1) EBMT utilizes a bilingual corpus.
2) EBMT utilizes a bilingual corpus as its primary learning base.
3) EBMT utilizes a bilingual corpus as its fundamental learning base, at run-time.

b) Feature Choice

Creator further portrays that the component determination can be seen as a standout amongst the most major issues in the field of AI. The principle point of highlight determination is to decide an insignificant element subset from an issue space while holding an appropriately high precision in speaking to the first highlig hts. In genuine issues, highlight choice is an unquestionable requirement because of the wealth of boisterous, superfluous or deceiving highlights. By expelling these elements, gaining from information procedures can profit enormously. Fluffy sets and the procedure of Fuzzification give a system by which genuine esteemed highlights can be successfully overseen. By enabling qualities to have a place with more than one mark, with different degrees of enrollment, the unclearness present in information can be demonstrated. The element determination stage is performed by a fluffy deduction framework dependent on the arrangement of tenets got from the Mel recurrence coefficients. The separated 39 coefficients are utilized by the fluffy surmising framework to create Gaussian participation

c) Fluffy derivation frameworks

Fluffy derivation frameworks are otherwise called fluffy standard based frameworks. Essentially, a fluffy surmising framework is made out of four useful squares is appeared in **figure 1.1**

A Knowledge base, containing various fluffy standards and the database, which characterizes the enrollment capacities utilized in the fluffy guidelines.

An Inference motor, which plays out the surmising activities on the principles.

A Fuzzification interface, which changes the fresh contributions to degrees of match with semantic qualities [3].

A Defuzzification interface, which changes the fluffy consequences of the surmising into a fresh yield.

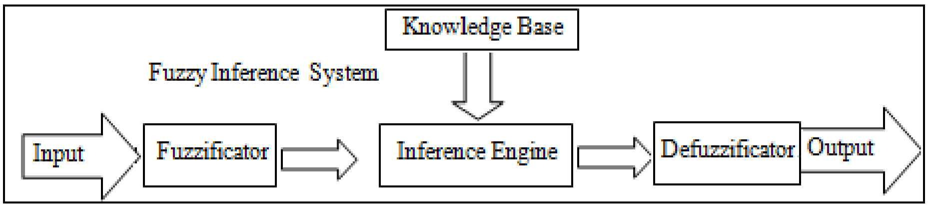

Figure 1.1: Fuzzy Inference System

The proposed EBMT [4] system can be utilized for programmed interpretation of content by reusing the instances of past interpretations. This system involves three stages, coordinating, arrangement and recombination.

d) Tokenization

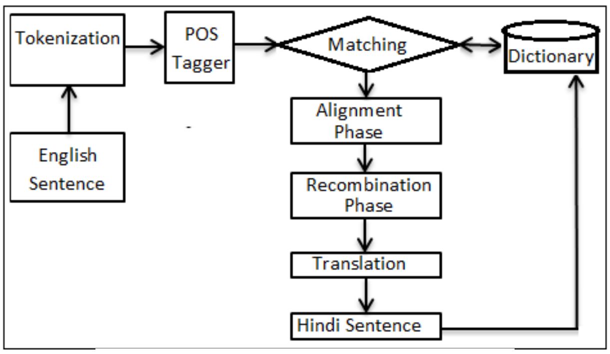

Figure 1.2: Design of system EBMT

Using NLP Tokenization [5] is an essential advance of Example based machine interpretation. In this stage, the info sentence is decayed into tokens. These tokens are given to POS stun capacity to label the tokens with their individual sort.

for example Sentence : "India has won the match by six wickets."

Tokens: "India" "has" "won" "the" "coordinate" "by" "six" "wickets."

Tokenization is an essential advance of Example based machine interpretation. In this stage, the info sentence is disintegrated into tokens. These tokens are yielded to POS amaze capacity to label the tokens with their individual sort.

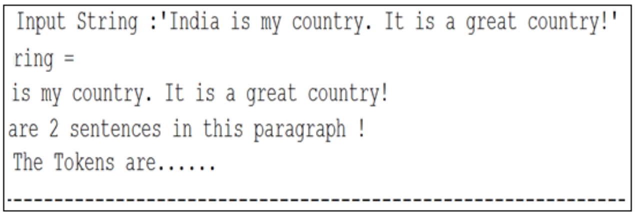

Figure 1.3 (a): Result of tokenization

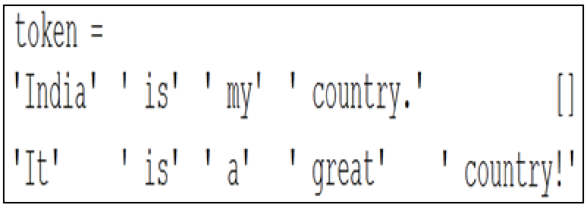

Figure 1.3 (b): Result of tokenization

Example 1
English : India won the match.
Hindi : Hkkjr us eWp ftrk |

Example 2
English : India is the best.
Hindi : Hkkjr loZJs"B gS |
Example 3
English : Sachin plays well.
Hindi : lfpu zvPNk [ksyrk gS |
Input
English : Sachin is the best.
Translation (Output)
Hindi : lfpu loZJs"B gS |

Result:
Major Expected Results
Based on this model expected **result's** as following:

Input: Source Language, ENGLISH (SL).
Output: Target Language, HINDI (TL).
The result obtained is with minimal human interface
The full LSTM's definition incorporates hardware for processing St and hardware for interpreting data from St.
Tragically; unique experts utilize somewhat extraordinary LSTM variations. In this work, we utilize the LSTM design that is definitely determined beneath. It is like the engineering of [6] however without peep-gap

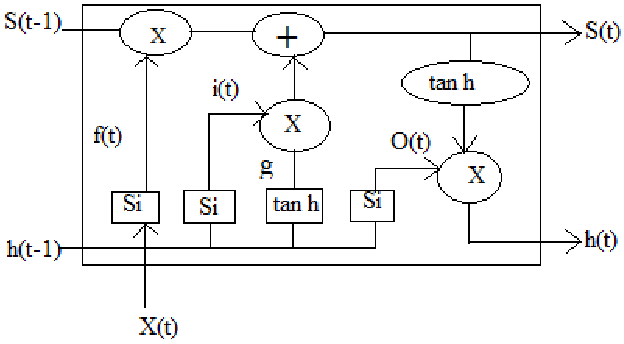

Figure 1.4: LSTM architecture

II. Long Short Term Memory Architecture

In the above LSTM engineering the images are characterized as,
S (t-1): Previous cell status
h (t-1): Previous cell concealed state
f (t): Forget door
I (t): Information door
Si: Sigmoid capacity
X (t): Current info
X: Vector duplication, in this paper it is spoken to by * documentation.
O (t): Output
S (t): Current cell status
h (t): Current cell concealed state

In this LSTM engineering the phone status store cell status. In view of current info LSTM [7] takes choice that how much past data is to erase. This activity is performed with the assistance of Forget entryway. Once past data is erased then new data is added to the cell utilizing Information entryway. The conditions are,

I (t) = Si (WxiXt +Whi ht-1 + bi)
f (t) = Si (WxfXt + Whf ht-1 + bf)
g = tanh (WxgXt + Whght-1 + bg)
(t) = Si (WxoXt + Who ht-1 + bo)
S (t) = S (t-1) * f (t) + I (t) * g
h (t) = tanh (S(t)) * O(t)

'W' is the weight vector instated haphazardly. 'b' is the predisposition esteem likewise introduced haphazardly. All the weight vectors are refreshed after every cycle.

LSTM engineering find out increasingly more via preparing and work useful for both long haul just as for momentary memory.

Standard LSTM is executed for a lot of information for 0 to 99 cycles and the outcome is appeared as the screen shot of the yield. This screen short just was indicating last piece of the yield with definite misfortune toward the finish of the yield[9].

```
cur iter: 98
y_pred[0] : -0.500087106205
y_pred[1] : 0.200498204588
y_pred[2] : 0.0994746051746
y_pred[3] : -0.499579491524
loss: 7.08662403825e-07
cur iter: 99
y_pred[0] : -0.500096072911
y_pred[1] : 0.200463300125
y_pred[2] : 0.0995059890975
y_pred[3] : -0.499595630139
loss: 6.31438766408e-07
```

Figure 1.6: Standard LSTM Output

In figure 3 after last iteration the final loss is 6.31438e-07. This loss is less than all the other loss of diffrent architecture except figure 2 modified LSTM architecture's loss.

Now, modified LSTM when execucate for the same data input the result is shown as the screen shot of the output. This screen short only was showing last part of the output.

```
cur iter: 98
y_pred[0] : -0.499712732281
y_pred[1] : 0.199925730777
y_pred[2] : 0.100122240059
y_pred[3] : -0.500299187737
loss: 1.92494594059e-07
cur iter: 99
y_pred[0] : -0.499731288362
y_pred[1] : 0.19992779384
y_pred[2] : 0.100116124116
y_pred[3] : -0.500280730381
loss: 1.69714030648e-07
```

Figure 1.5: Modified LSTM

III. Output

In Last Phase of researchIn this work it gives the promising outcomes when given the contribution to any language, think about this, whenever input given in japanese language then it would be changed over into english and hindi according to the program terms, at that point the client may choose to pick another dialect to be deciphered in.

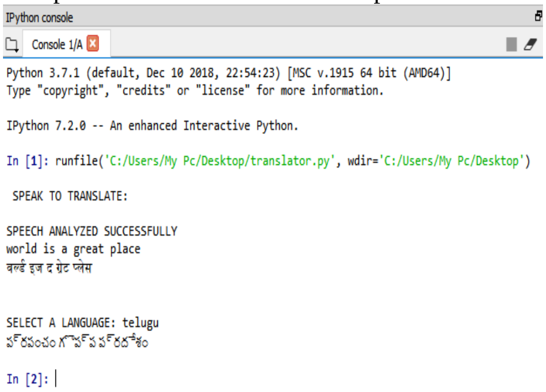

Figure 1.6: Speeches Analysed

IV. Outcomes

The result of this paper is to endeavor to accomplish 99 percent precision through the utilization of Google cloud servers and through python ML bundles. Critical, results have been accomplished and contrasted and different strategies through this paper

2. Conclusion

The general research depends on four primary commitments; Example based Machine Translation Approach, Use of Fuzzy Tools for Machine Translation, Long Term Short Memory for Machine Translation Approach and English to Indian Language Translation Using Textblob& Google-API. As referenced over the thorough experimentation is completed standard test information accumulation for these four procedures. The discoveries of the experimentation are talked about beneath.

In Example based Machine Translation Approach another framework is fabricated, which is versatile, straightforward and productive. The whole framework changes over the source language content into target language content utilizing characteristic language handling [10]. The calculation is with the end goal that, there is lexicon/corpus/vocabulary of English and Hindi. The parsing is to be done appropriate. The mapping procedure is likewise utilized. Every one of the literals are isolated utilizing parceling and stemming systems. The root words are been recognized utilizing man-made reasoning and bilingual interpretation.

Utilization of fluffy for Machine Translation gives the execution measurements utilized for assessment of interpretation are unigram accuracy, unigram review, F-measure, BLUE, NIST, mWER, SSER. The whole EBMT framework makes an interpretation of the source language into target language utilizing NLP which is roughly 0.14 percent progressively effective, than existing apparatuses RBMT and SMT in the market.

Standard LSTM engineering works better then the RNN by taking care of the disappearing inclination issue. The LSTM design isn't impeccable. To make it progressively exact we probed it by changing its engineering and thus think of another LSTM design which works superior to that of the standard engineering. In this manner, adjusting standard LSTM design by changing the overlook door structure and evacuating pointless standardization work improved the LSTM execution.

It has examined all the center functionalities of how the English language is changed over to Indian dialects utilizing TEXTBLOB and GOOGLE-API, likewise results have been contrasted and different philosophies gave in various papers and a normal of +10 percent increments in productivity and exactness was taken note. The result of this examination is to attempt to accomplish 99 percent precision through the use of Google cloud servers and through python ML bundles. Huge, results have been accomplished and contrasted and different strategies.

References

[1] Manish Rana & Mohammad Atique “Enhancing Bi-Lingual Example Based Machine Translation Approach” International Journal of Advanced Engineering Research and Science (IJAERS), Vol-3, Issue-10, Oct 2016, ISSN: 2349-6495(P)/ 2456-1908(O) Pp 8-12.

[2] Manish Rana & Mohammad Atique “Example Based Machine Translation Using Fuzzy Logic from English to Hindi” *Int'l Conf. Artificial Intelligence | ICAI'15 | Pp -354-359*

[3] Manish Rana & Mohammad Atique“A Review: Machine Translation Using Various Soft-Computing Tools” Volume-ll ISBN : 978-0-9884925-8-5 International Conference on Communication Computing & Virtualization 2015 Vol 2, Pp : 813-816

[4] Manish Rana & Mohammad Atique “ Example Based Machine Translation Using Various Soft-Computing Techniques Review” International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015 ISSN 2229-5518 , Pp-1100- 1106.

[5] Manish Rana & Mohammad Atique “English to Indian languages Using TEXTBLOBand GOOGLE-API in Python” Multiconw-2019, Mc-Graw Hill Publication, Pp215-219.

[6] Manish Rana & Mohammad Atique “Use of Long Short-Term Memory for Enhancing Bi-Lingual Machine Translation Approach” Proceeding of conference on Recent Trends in Computer Engineering 2018, ISBN 978-0-9994483-0-4. Pp 255-258.

[7] Manish Rana & Mohammad Atique “ Enhancing Example Based Machine Translation Using Fuzzy Tool” 8th Internation onference on communication , computing and Virtualization 2017, MulticonW 2017, Pp 211-216.

[8] Manish Rana & Mohammad Atique “Use of Fuzzy tool for example based machine translation” 7th International Conference on Communication, Computing and Virtualization 2016, Elsevier Pp 119-206.

[9] M. Rana "Example Based Machine Translation Using Fuzzy Logic", Procedia Computer Science 00 (2016) 000–000, ISBN: 978-0988-492-56-1, .ICCCV 2016.
[10] M. Rana," A Review: Machine Translation Using Various Soft-Computing Tools-27th & 28th Feb 2015 Multiconw-2015 Proceedings ISBN: 978-0-9884925-8-5, ISSN 0975- 8887.