Evaluation of Domain Based Hindi to Punjabi Machine Translation System

Manish Kumar
Assistant Professor, Baba Farid College, Bathinda
manish21578@gmail.com

Abstract

Author of this paper has evaluated the Hindi to Punjabi Machine Translation system which is developed for a translation of text material of a particular domain of computer science. This system is the next level of the general purpose machine Hindi to Punjabi Machine Translation (MT) system developed by Goyel and Lehel [1] in Punjabi University Patiala. Evaluation results of that system shows that accuracy level for general text is excellent, but when it translates any study material related to computer subjects the accuracy level comes down [2]. A new system has been developed to eradicates the problems related to translation of computer subjects and this paper will elaborate the evaluation of new domain based Hindi to Punjabi MT system.

1. Introduction

Hindi to Punjabi MT system described in this study is a domain based system, which translates the study material of computer subjects with higher accuracy than general purpose Hindi to Punjabi MT system. This system works on two closely related languages [3], Hindi as source language and Punjabi as target language. Efficiency of any translation system depends on the accuracy of results produced by the system in target language. In this study author has used standard evaluation measures and tools to check the accuracy level of the system.

2. Evaluation Procedure

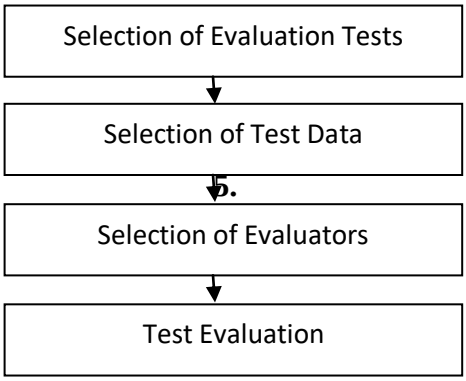

Figure 1
Evaluation Procedure

3. Evaluation Tests

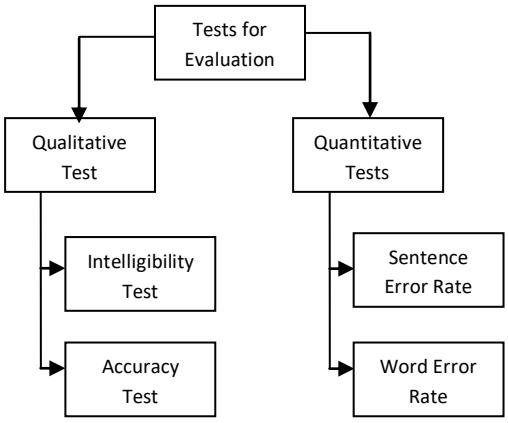

Figure 2 (Various Tests)

Intelligibility Test

Purpose of this test is to check the understandability echelon of the output generated in target language. Every evaluator has to evaluate each sentence on the scale of three. Table 1 shows the scoring sheet of intelligibility test.

Table 1 (Scoring sheet for Intelligibility Test)

Score	Meaning
0	Sentence is unintelligible
1	Sentence needs some corrections to understand
2	Sentence is understandable but needs minor corrections
3	Sentence is completely understandable.

Accuracy Test

Purpose of accuracy test is similar to intelligibility test. In this test source text is also provided along with the translated text to all the evaluators. Using source test and after comparing it with text in target language all the evaluators evaluated the system. Table 2 shows the score sheet of accuracy test.

Word Error Rate

This test evaluates the number of wrongly translated or un-translated words in comparison to the correct reference translation of the word. Word Error Rate is calculated by the following formula

$$\text{Word Error Rate} = (S+I+D)/N$$

In this formula 'S' is the number of substitutions made for words translated wrongly. Whereas 'I' stands for number of insertions done in the text and 'D' stands for number of deletions for correct translation. Sum of all these three is divided with N which is total number of translated words.

Table 2 (Score sheet for Accuracy Test)

Score	Meaning
0	Meaning of Sentence in Target language is totally different in comparison of sentence in Source language
1	Sentence hardly has the same meaning
2	Sentence is moderately accurate and gives almost same meaning
3	Sentence gives exactly same meaning

Sentence Error Rate

Sentence Error rate is also similar to Word Error rate, with the difference that in this case whole sentence is compared with the reference translation. Sentence Error rate is calculated by dividing the wrongly translated sentences with total sentences.

4. Selection of Test Data

It was very important to select the sentence set for evaluation which covers all the aspects of evaluation. It was required to take a sentence set in such a way so that evaluation regarding all the errors could be done. Following sentence set was taken for the evaluation purpose. Table 3 represents the total test data used for the evaluation of this system.

Table 3 (Total Test Data for Evaluation)

	Number of Sentences	Number of Words
Total Test Data for Evaluation	176	3222

5. Selection of Evaluators

Table 4 (Evaluators for Evaluation)

Category	Number
Computer Science Teachers	25
Students of Computer Science	50
Teachers of Social Science	25
Teachers of Punjabi Language	25
Students of Social Science	50

Around 175 evaluators were selected for various evaluation tests. Those evaluators were selected on the basis of their profession and according to their understanding of target language. Table 4 describes the detail of different kind of evaluators selected for this evaluation.

6. Evaluation of Tests

Analysis of Intelligibility Evaluation

All the evaluators did an intelligibility test for the complete sentence set of evaluation. Following were the results for intelligibility test.

Table 5 (Evaluation results of Intelligibility Test)

Category	Percentage
Percentage number of sentences for which Score 0 was given	0.99%
Percentage number of sentences for which Score 1 was given	2.36%
Percentage number of sentences for which Score 2 was given	25.33%
Percentage number of sentences for which Score 3 was given	71.32%

From the above figures it was analysed that most of the translated sentences were understandable.

Analysis of Accuracy Evaluation

Table 6 (Evaluation results of Accuracy Test)

Category	Percentage
Percentage number of sentences for which Score 0 was given	4.20%
Percentage number of sentences for which Score 1 was given	6.35%
Percentage number of sentences for which Score 2 was given	21.22%
Percentage number of sentences for which Score 3 was given	68.23%

There were more than 89% sentences which were evaluated as understandable and matched with the meaning inference by evaluators from source language text. There were around 7% sentences which required some correction to understand completely.

Analysis of Word Error Rate

Word Error Rate evaluation is word based evaluation system. In the sentence set we took sentences which were causing different kind of problems in the baseline MT system. While evaluating the system of Word Error

Rate we have evaluated it separately for different types of sentences.

Analysis of Word Error Rate shows that there is 2.86% Word Error Rate in the system.

Table 7 (Results of Word Error Rate)

	Total Percentage of Error in Complete Test Data
Wrongly Translated Words	1.44%
Un Translated Words	1.01%
Insertion or Deletion required	0.41%
Total	2.86%

Analysis of Sentence Error Rate

Sentence Error evaluation was done by checking the correctness of sentences. Sentences were compared with the correct reference sentences.

Table 8 (Result of Sentence Error Rate)

	Percentage Sentence Error
Sentences having technical ambiguous words	10%
Sentences having Non technical ambiguous words	13.50%
Sentences having word with inflectional errors	15.45%
Sentences having words which were out of vocabulary of baseline MT system	9.36%
Average	**12.00%**

Sentence error evaluation strictly checks the output generated by the system with the reference text in target language. It was analyzed that sentences having inflectional error issues were having high Sentence Error Rate. On the average 12% Sentence Error Rate was evaluated for updated Hindi to Punjabi MT System.

Table 9 Comparison of Evaluation results with general purpose MT system

Test	Accuracy Level of New domain based Hindi to Punjabi MT System	Accuracy Level of general purpose Hindi to Punjabi MT system
Intelligibility Test	96.55%	90.01%
Accuracy Test	89.45%	82.11%
Word Error Rate	97.14%	86.76%
Sentence Error Rate	88%	78.1%

Above table clearly shows that refurbished domain specific Hindi to Punjabi Machine Translation is giving better results as compare to general purpose MT system with the same data set.

Conclusion

In this paper author has evaluated the new domain based Hindi to Punjabi Machine Translation system. On the basis of various tests conducted it is evaluated that accuracy level of the system has increased to very prominent level.

References

1. *Goyal, V, Lehal, G.S., (2009) "Evaluation of Hindi to Punjabi Machine Translation System," International Journal of Computer Science, 4(1)*
2. *Goyal, V, Lehal, G.S., (2011) "Hindi to Punjabi Machine Translation System,"Proceedings of the ACL-HLT 2011, pp. 1-6.*
3. *Altintas K., "A Machine Translation System between a pair of closely related languages ".Internet:* http://www.cs.bilkent.edu.tr/ ~ilyas / PDF / iscis2002.pdf.
4. *Marrafa A. "Quantitative Evaluation of Machine Translation Systems : Sentence Level ", Proceedings of the MT Summit VIII Fourth ISLE workshop 2001, spain, pp 39*