

Q3) What do you learn from this assignment?

Ans) We were given 3 models for evaluating how they perform translation task. These 3 models are as follows:

- 1) Distilled model of NLLB-200 with 600M parameters
- 2) IndicTrans
- 3) ChatGPT

4 tasks were given for evaluating these 3 models. These tasks were as follows:

- 1) English to X. As my mother tongue is Marathi, X = Marathi
- 2) X to English
- 3) One Indian language (other than X), say Y, to X. I selected Y = Hindi.
- 4) Y to X

Each model was evaluated on the basis of evaluation metrics that contain BLEU and ROUGE scores.

For the ChatGPT model, I tried translation using api calls but due to exhausted credit limit, I was not able to perform translation using api calls. Therefore, I performed task using the ChatGPT prompt.

After obtaining the BLEU and ROUGE scores for each of the task performed on every model, here are some observations:

- 1) Looking at the scores obtained while translating the 4 tasks, the sentence bleu and corpus bleu scores depicted a moderate similarity between the predicted translation and ground truth translation.
- 2) With respect to ROUGE score variants(rouge1, rouge2, rougeL), the results for NLLB and IndicTrans were moderate, indicating that there was significant overlap between the unigrams and bigrams of predicted and ground truth sentences.
- 3) For ChatGPT, the rouge scores(all variants) were not that good. From this we can conclude that the overlap between unigrams, bigrams is less between the predicted and ground truth sentences.
- 4) Results are as follows:

Model	BLEU Scores		Rouge Scores			
NLLB-200	EN-MR	0.6291820934924778	Tasks	Rouge1	Rouge2	RougeL
			EN-MR	0.440	0.187	0.41527
	MR-EN	0.6660380432754996	MR-EN	0.5829	0.310	0.5447
	HI-MR	0.631046032549561	HI-MR	0.4553	0.1964	0.4265
	MR-HI	0.7965474262811929	MR-HI	0.546	0.283	0.50799

IndicTrans	EN-MR	0.6168745608330305	Tasks	Rouge1	Rouge2	RougeL
			EN-MR	0.4869	0.2382	0.4611
	MR-EN	0.647562271776767	MR-EN	0.6071	0.3616	0.5779
	HI-MR	0.6177154817226991	HI-MR	0.4613	0.2175	0.4399
	MR-HI	0.8020251302159406	MR-HI	0.5302	0.2991	0.5044
ChatGPT	EN-MR	0.6123738933621928	Tasks	Rouge1	Rouge2	RougeL
			EN-MR	0.2443	0.0528	0.2186
	MR-EN	0.7310365171555685	MR-EN	0.4562	0.2388	0.4031
	HI-MR	0.5595757524151437	HI-MR	0.2452	0.0552	0.2191
	MR-HI	0.7453722040405232	MR-HI	0.4334	0.1735	0.3729

Note: The BLEU scores are calculated using sentence\_bleu.