CS689: COMPUTATIONAL LINGUISTICS FOR INDIAN LANGUAGES

ASSIGNMENT 3

April 2024

Neelu Lalchandani
231110031

1 Introduction

Machine translation is the process of automatically converting text or speech from one language into another. It's a crucial application of computational linguistics and artificial intelligence, aiming to bridge language barriers and facilitate communication between people who speak different languages.

2 Scores on different models

2.1 NLLB-200 with 600M parameters

- English to Hindi

ROUGE Scores:
rouge-1:
r: 0.5673244464391796
p: 0.6138173950443074
f: 0.5848161609314676
rouge-2:
r: 0.3379879945430296
p: 0.3671030800977633
f: 0.34880633440256265
rouge-l:
r: 0.5268853560781062
p: 0.5695201785034644
f: 0.5429572079803355
BLEU Score:
0.6258782765551912

- Hindi to English

ROUGE Scores:
rouge-1:
r: 0.6109130131344749
p: 0.6265071113633824
f: 0.6136309769878373
rouge-2:
r: 0.39097728857825575
p: 0.39794851846258766
f: 0.39061256117595217
rouge-l:
r: 0.5810031725845388
p: 0.5953442800345884
f: 0.58335953711382
BLEU Score:
0.676214216695444

- Hindi to Marathi

ROUGE Scores:
rouge-1:
r: 0.3883062874628678
p: 0.4177619649227626
f: 0.39782587920011836
rouge-2:
r: 0.17215114305334045
p: 0.1841238058209423
f: 0.1757473977623874
rouge-l:
r: 0.36241911174839403
p: 0.3894393365191745
f: 0.37112840951840786
BLEU Score:
0.547942429361091

- Marathi to Hindi

ROUGE Scores:
rouge-1:
r: 0.5080351124308534
p: 0.5528065916298678
f: 0.5241000342501483
rouge-2:
r: 0.2785240852566996
p: 0.3022761561799483
f: 0.28692531515869024
rouge-l:
r: 0.47426944035394625
p: 0.5154070180988227
f: 0.489094621494122
BLEU Score:
0.5802169830982862

2.2 indicTRANS

- English to Hindi

ROUGE Scores:
rouge-1:
r: 0.6270386143636398
p: 0.6344910701438935
f: 0.627473479134096
rouge-2:
r: 0.39380706318151115
p: 0.399636179290827
f: 0.39458509897405
rouge-l:
r: 0.5888695393414554
p: 0.5962371067409259
f: 0.5894927752055814
BLEU Score:
0.6995483147835915

- Hindi to English

ROUGE Scores:
rouge-1:
r: 0.6667621128652064
p: 0.662975116834313
f: 0.6612807966510005
rouge-2:
r: 0.4576863711711367
p: 0.451573672589976
f: 0.4517390846536463
rouge-l:
r: 0.6336902759953175
p: 0.6295916692638548
f: 0.6282762248654941
BLEU Score:
0.7516008909688211

- Hindi to Marathi

ROUGE Scores:
rouge-1:
r: 0.4235107105756269
p: 0.4199960895720604
f: 0.4178299267877743
rouge-2:
r: 0.19286763302103752
p: 0.18900911053580094
f: 0.1891159785980255
rouge-l:
r: 0.39528685803641705
p: 0.3918145260136416
f: 0.3898598573554864
BLEU Score:
0.6091614587017943

- Marathi to Hindi

ROUGE Scores:
rouge-1:
r: 0.5315656827163657
p: 0.5445237554258757
f: 0.5332282783011526
rouge-2:
r: 0.28752005153928917
p: 0.2949085828895486
f: 0.2884293405003241
rouge-l:
r: 0.4930805188256202
p: 0.5052356285819584
f: 0.4947890659945677
BLEU Score:
0.6139302618015272

2.3 chatGPT

- English to Hindi

ROUGE Scores:
rouge-1:
r: 0.5233651144494275
p: 0.5549458207422198
f: 0.534411759911741
rouge-2:
r: 0.2906069734460578
p: 0.3040811738489409
f: 0.29471983800942403
rouge-l:
r: 0.48377628770025966
p: 0.5086612214854404
f: 0.4921473783509021
BLEU Score:
0.6021075411885106

- Hindi to English

ROUGE Scores:
rouge-1:
r: 0.7756314051226327
p: 0.7946025182795945
f: 0.7828875654188898
rouge-2:
r: 0.6438798137110392
p: 0.6554244359756742
f: 0.6479524435635742
rouge-l:
r: 0.7481099609653086
p: 0.7663721743578132
f: 0.7553112601161088
BLEU Score:
0.8318640473554315

- Hindi to Marathi

ROUGE Scores:
rouge-1:
r: 0.31117307074851813
p: 0.3232613435183094
f: 0.313708864011215
rouge-2:
r: 0.1176444859021417
p: 0.12180796796083176
f: 0.11844703944322145
rouge-l:
r: 0.2840205010959485
p: 0.2950764328407517
f: 0.28646455956076006
BLEU Score:
0.5126514588699214

- Marathi to Hindi

ROUGE Scores:
rouge-1:
r: 0.602929845589998
p: 0.6222393317385311
f: 0.6073591689708325
rouge-2:
r: 0.4021685124464836
p: 0.41155851118349707
f: 0.4033469580315313
rouge-l:
r: 0.557264301897987
p: 0.5723772448637542
f: 0.56035257085593
BLEU Score:
0.6591325750493927

3 Observations/Learnings

The ROUGE scores are a set of metrics used to evaluate the quality of machine-generated text by comparing it to reference text. Higher ROUGE scores indicate better performance, with the maximum score being 1.0.

The BLEU score is another metric used to evaluate the quality of machine-generated text by comparing it to reference text. It measures the similarity between the generated text and the reference text, looking at the overlap of n-grams (sequences of n words).

Let's look at the observations:-

- The indicTRANS model outperforms the NLLB-200 model in all four language translation cases, as evidenced by the higher ROUGE and BLEU scores.
- The chatGPT model outperforms both the NLLB-200 and indicTRANS models in the Hindi to English translation case, with significantly higher ROUGE and BLEU scores.
- For the English to Hindi translation, the indicTRANS model has the highest ROUGE and BLEU scores, followed by the NLLB-200 model and then the chatGPT model.

- In the Hindi to Marathi and Marathi to Hindi translation cases, the indicTRANS model has the best performance, followed by the NLLB-200 model and then the chatGPT model.
- The ROUGE-2 scores, which measure the overlap of bigrams between the translated text and the reference, are consistently lower than the ROUGE-1 and ROUGE-L scores across all models and language pairs, indicating room for improvement in capturing longer-term dependencies.
- The BLEU scores, which measure the similarity of the translated text to reference translations, are generally lower than the ROUGE scores, suggesting that the models may be producing fluent translations that do not exactly match the reference translations.
- The performance of the models varies significantly across the different language pairs, with the Hindi to English and Marathi to Hindi translations generally having higher scores than the Hindi to Marathi and English to Hindi translations.
- The indicTRANS model appears to be the most consistent performer, with relatively high scores across all language pairs, while the NLLB-200 and chatGPT models show more variability in their performance.
- The lower scores for the Hindi to Marathi and Marathi to Hindi translations compared to the other language pairs may be due to the linguistic and cultural differences between these languages, which could pose additional challenges for the models.
- Overall, the results suggest that while the models have made significant progress in machine translation, there is still room for improvement, particularly in capturing longer-term dependencies and producing translations that are more faithful to the reference.