Intro to NLP Project

Team 8: NLP Rookies

Project Title: **Measure Text Fluency**

Abstract

We aim to measure the fluency of text in corpus. Fluency is commonly considered as one of the dimensions of text quality of MT. Fluency measures the quality of the generated text (e.g., the target translated sentence) i.e how much a sentence is perceived as natural by a human reader, without taking the source into account. It accounts for criteria such as grammar, spelling, choice of words, and style. To understand text fluency we first need to check how readable the text is. The readability of text depends on its content (the complexity of its vocabulary and syntax). It focuses on the words we choose, and how we put them into sentences and paragraphs for the readers to comprehend.

We will use rule based 5 readability index scores, statistical model scores and neural network model scores over the same corpus for each sentence .Then we will normalize each score.Then get a weighted mean score combining all the scores to get best score.Maximum weight is given to the NLM model.

Weights are assigned like below:

- For readability index score -> weight of 2
- For statistical model score -> weight of 3
- For NLM model score -> weight of 5

Then each weighted mean score is scaled to 10 ; low score means better in fluency and can be perceived better.

Then we randomly selected 100 sentence from corpus;and manually calculated Fluency score from scale of 1 to 10 following below rules:

v complex words , sentence length high	8+

v complex ; sentence length short	7 - 8
complex ; sentence length high	5-6
complex ; sentence length short	4-5
less complex/easy ; sentence length high	3-4
less complex/easy ; sentence length short	2-3

The manually calculated Fluency scores are then compared with the weighted means score.

Implementation

Steps

- We will use rule based 5 readability indices, statistical model and neural network model over the same corpus .
- So we will have scores from each of 3 above methods.
- Then we normalized the score obtained from each of the above 3.
- Formula used:

  $(x - x_{min}) / (x_{max} - x_{min})$

- Then using weights calculated weighted mean score for each sentence:

```
[2*(Readability score1 +Readability score2 +Readability score3
+Readability score4 +  Readability score5 )
   + 3*Statistical Model score
       + 5*NLM model score ] / (2*5 + 3+5)
```

Details

We used the *MSR Abstractive Text Compression Dataset* as corpus. Our plan is to implement the models for measuring the fluency in three different forms. First we will use rule based measures to measure the textual readability of the sentences in the corpus. Next we measure the fluency using a statistical model and in the end we measure the fluency using an LSTM NN model. We compare each of these results and weigh them against human readable scores for analysis.

1. Calculating the readability of the corpus

We have applied the below readability indices to a subset of 15,600 sentences. These are based on different rules of the given text and help determine the readability, complexity, grade level. Based on the computed readability index, we perceive it as easier to the reader and hence more natural. This might not be a direct measure of textual fluency but is an important metric to measure it. The sentences with high readability indices are more likely to score well on the fluency scale as well.

- Automated readability index
- Dale-Chall readability score
- Coleman liau index
- Linsear write formula
- Gunning fog

And calculated the readability scores for each of them.

Automated readability index :

$$4.71\left(\frac{\text{characters}}{\text{words}}\right)+0.5\left(\frac{\text{words}}{\text{sentences}}\right)-21.43$$

where *characters* is the number of letters and numbers, *words* is the number of spaces, and *sentences* is the number of sentences, which were counted manually by the typist when the above formula was developed. Non-integer scores are always rounded up to the nearest whole number, so a score of 10.1 or 10.6 would be converted to 11.

Dale-Chall readability score:

It uses a list of 3000 words that groups of fourth-grade American students could reliably understand, considering any word not on that list to be difficult.

$$0.1579\left(\frac{\text{difficult words}}{\text{words}}\times 100\right)+0.0496\left(\frac{\text{words}}{\text{sentences}}\right)$$

If the percentage of difficult words is above 5%, then add 3.6365 to the raw score to get the adjusted score.

Coleman-Liau index:

$$CLI=0.0588L-0.296S-15.8$$

L is the average number of letters per 100 words and *S* is the average number of sentences per 100 words.

Linsear write formula:

The standard Linsear Write metric *Lw* runs on a 100-word sample:

1. For each "easy word", defined as words with 2 syllables or less, add 1 point.
2. For each "hard word", defined as words with 3 syllables or more, add 3 points.
3. Divide the points by the number of sentences in the 100-word sample.
4. Adjust the provisional result *r*:
    - If *r* > 20, *Lw* = *r* / 2.
    - If *r* ≤ 20, *Lw* = *r* / 2 - 1.

Gunning fog index:

1. Select a passage (such as one or more full paragraphs) of around 100 words. Do not omit any sentences;
2. Determine the average sentence length. (Divide the number of words by the number of sentences.);
3. Count the "complex" words consisting of three or more syllables. Do not include proper nouns, familiar jargon, or compound words.
4. Add the average sentence length and the percentage of complex words
5. Multiply result by 0.4

$$0.4\left[\left(\frac{\text{words}}{\text{sentences}}\right)+100\left(\frac{\text{complex words}}{\text{words}}\right)\right]$$

Average score : It is the average of all above indices.

Sample output:

	text	score1	score2	score3	score4	score5	avg_score
0	The KIT's production of FOREVER PLAID by direc...	12.4	14.91	13.27	11.000000	12.67	12.850000
1	FOREVER PLAID, a KIT production directed by St...	8.7	15.19	10.81	5.166667	11.01	10.175333
2	I don't have all the answers, but you can make...	9.2	7.46	5.00	14.000000	11.27	9.386000
3	You can help make positive changes in the live...	9.8	9.35	9.16	11.500000	12.76	10.514000
4	You can help make changes in lives at central ...	9.8	8.04	9.16	10.500000	10.53	9.606000

2. Usage of Statistical N-gram LM

IDEA: Use probability of a sentence obtained from N-gram model to judge fluency. This approach does not use any reference translation.

APPROACH:

1. N-GRAM MODEL: The n-gram model is used to predict the probability of a word by the left contextual words. In the n-gram model, we consider a sentence as ann order Markov chain, where a word's probability is predicted only by the n-1 words to its left. The formula to calculate the probability of sentence using n-gram model:

$$P(W_1..W_{m-1}W_m) = P(W_1)\times P(W_2 \mid W_1)\times ...\times P(W_{n-1} \mid W_1..W_{n-2}) \times \prod_{i=n}^{m} P(W_i \mid W_{i-n+1}..W_{i-1}) \quad (5)$$

2. SMOOTHING: Modified kneyser ney smoothing is done to move some probability to the unkown N-grams.

$$P_{KN}(w_n|w_1,\ldots,w_{n-1}) = \frac{C(w_1,\ldots,w_n) - D}{\sum_{w' \in L} C(w_1,\ldots,w_{n-1},w')} + \lambda(w_1,\ldots,w_{n-1}) P_{KN}(w_n|w_2,\ldots,w_{n-1})$$

3. Using N-gram conditional probability:

   Take the first term w_1 of a sentence, and lets say it is a high frequency word. A high score will be assigned to the sentence. But this is not always true, the probability that the sentence starts with w_1 might be very low. This results in loss of fluency. So we replace $P(w_1)$ with $P(w_1 \mid PSE)$ where PSE denotes the punctuations used in the end of a sentence such as full stops, question marks and so on. Using $P(w_1 \mid N-gram)$ to denote the CP of different n-grams in a sentence, the formula is written as:

$$M(W_1W_2...W_m) = [\sum_{i=1}^{m} \log(P(W_i \mid N-Gram))]/m$$

4. SENTENCE FLUENCY:

   Only few trunks of a sentence make it ill-formed. Since our aim is to select these bad sentences from good ones, we don't need to multiply all the n-grams of a sentence but only to multiply those strange ones.Thus we multiply the CP of those strange n-grams and call the result as penalty value. We multiply the CP of those good n-grams and call the result as praise value. Fluent sentences will have a high penalty value and a low praise value while

the illformed sentences will have a lowpenalty value and a high praise value. Thus, we divide the penalty value by the praise value and use the result. The fluency of a sentence is calculated using the below algorithm:

$$
\begin{aligned}
&M(W_1W_2 \ldots W_m) = 1/m \\
&\text{for } i = 1, \ldots m \\
&\quad \text{if } (P(W_i \mid N-Gram) \geq ValForGood) \\
&\qquad M(W_1W_2 \ldots W_m) = M(W_1W_2 \ldots W_m) / P(W_i \mid N-Gram) \\
&\quad \text{elseif } (P(W_i \mid N-Gram) \leq ValForBad) \\
&\qquad M(W_1W_2 \ldots W_m) = M(W_1W_2 \ldots W_m) \times P(W_i \mid N-Gram) \\
&\quad \text{endfor}
\end{aligned}
$$

Here ValForGood and ValForBad is selected as follows:

1. Sort the conditional Probabilities obtained from the N-GRAMS.
2. ValForGood is taken as the lowest conditional probability of the first 40 percent of probabilities and ValForBad is taken as the highest conditional probability of the last 20 percent

In our implementation we have replaced probability with Perplexity to show better readable results.

SAMPLE OUTPUT:

text	N-gram LM score
The KIT's production of FOREVER PLAID by director Steven D. Kline, stage manager Alan Bolosan Campo, Technical Director Derron Peterson and costumes	39.95707347
FOREVER PLAID, a KIT production directed by Steven D. Kline. Set, lighting and technical support by Derron Peterson. Costume by Brigitte Doth, Jade Stice,	606.0699404
I don't have all the answers, but you can make a difference in a child's life by sending a gift to The Salvation Army.	74904.61989
You can help make positive changes in the lives of at-risk youth in central Indiana. Some of them are growing up in such emotional poverty, they have give	56090.28691
You can help make changes in lives at central Indiana. We hear about poverty, but what these young people are growing up with is emotional poverty-they	684.3278379
You can make positive changes for at-risk youth in Indiana. We hear about economic poverty, but these youngsters are growing up with emotional poverty	699.8647944
Help make changes in the lives of at-risk youth in Indiana! We hear about economic poverty, but what these children are growing up with is emotional pov	288.7567269
You can changes lives of at-risk youth in central Indiana who have grown up with not only economic but emotional poverty - they have often given up on h	147.886495
Each year the graduating class contributes a gift to the law school. Sometimes this is an item of remembrance, or even individual tax deductible donations	480.8738257
Each year the graduating class gives a gift to the law school. This might be an item of remembrance, a scholarship, or individual tax deductible donations.	654.0486064
regardless of the difficulties they may be facing, The Salvation Army offers grace and compassion to those in need in our community. You have been gener	70.13516323
The Salvation Army offers grace and compassion to those in need in our community. You've been so generous with your giving in the past, and we are grat	19.08334712
Like the Salvation Army offers grace and compassion to those in need, you've been so generous with your giving in the past, and we are grateful.	66.30049749
The Salvation Army offers grace and compassion to those in need in our community. You've been so generous with your giving in the past, and we are grat	11.8483819
The Salvation Army offers grace and compassion to those in need - regardless of difficulties they're facing. You've been generous with your giving and we a	330.7726172
This is possible because we are a statewide network. Join us, send in your application today!	85.59639244
This is possible because we are a statewide member network. Send in your membership to join us!	25.44940479
All achievements are possible because of our large network. Please join us by applying today!	207.2132831
We are a statewide member network. Join us, by sending in your membership-application -today!	80.13020161
Join our statewide member network by sending in a membership-application today!	31.4565954

3. Usage of LSTM neural network based language model

Training is done on 6000 sentences from corpus ; so vocab size from 6000 sentence is determined initially

Then input dimension is determined from vocab size and One hot encoding for each word in each sentence is fed to language model

Dimensions of LSTM were chosen to be 12 and 2 dropout layers were used.

Activation functions relu and softmax are used

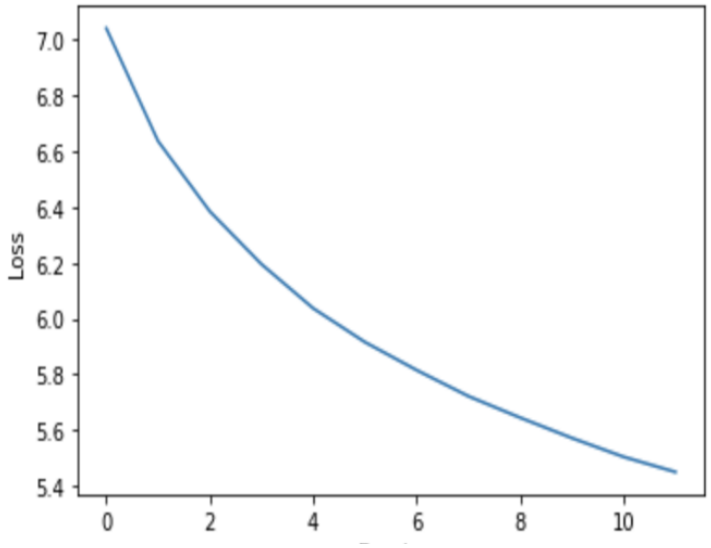
LSTM Neural network model is trained on 6000 sentences for 15 epochs to predict perplexity of sentences using a probability matrix for each word for each sentence.After training it was tested on 4000 sentences.

```
def create_model():
    model = Sequential()
    model.add(Embedding(input_dim=len(vocab) + 1,  # vocabulary size. Adding an
                                                   # extra element for <PAD> word
              output_dim=6,  # size of embeddings
              input_length=maxlen - 1))  # length of the padded sequences
    model.add(LSTM(12))
    model.add(Dropout(0.3))
    model.add(Dense(len(vocab), activation='relu'))
    model.add(Dropout(0.5))
    model.add(Dense(len(vocab), activation='softmax'))
    model.compile(loss='categorical_crossentropy',
              optimizer='adam',
              metrics=['accuracy'])
    return model
```

Training output

```
Epoch 1/12
1550/1550 [==============================] - 2363s 2s/step - loss: 7.0401 - accuracy: 0.0626
Epoch 2/12
1550/1550 [==============================] - 3141s 2s/step - loss: 6.6360 - accuracy: 0.0856
Epoch 3/12
1550/1550 [==============================] - 3333s 2s/step - loss: 6.3844 - accuracy: 0.1051
Epoch 4/12
1550/1550 [==============================] - 3261s 2s/step - loss: 6.1949 - accuracy: 0.1197
Epoch 5/12
1550/1550 [==============================] - 3207s 2s/step - loss: 6.0369 - accuracy: 0.1291
Epoch 6/12
1550/1550 [==============================] - 3115s 2s/step - loss: 5.9158 - accuracy: 0.1333
Epoch 7/12
1550/1550 [==============================] - 3081s 2s/step - loss: 5.8154 - accuracy: 0.1385
Epoch 8/12
1550/1550 [==============================] - 2696s 2s/step - loss: 5.7221 - accuracy: 0.1430
Epoch 9/12
1550/1550 [==============================] - 2975s 2s/step - loss: 5.6452 - accuracy: 0.1470
Epoch 10/12
1550/1550 [==============================] - 2852s 2s/step - loss: 5.5723 - accuracy: 0.1495
Epoch 11/12
1550/1550 [==============================] - 2887s 2s/step - loss: 5.5048 - accuracy: 0.1535
Epoch 12/12
1550/1550 [==============================] - 3296s 2s/step - loss: 5.4509 - accuracy: 0.1559
Text(0, 0.5, 'Loss')
```

Loss graph



Sample perplexity score output for sentences

	text	LSTM PERPLIXITY SCORES
0	The KIT's production of FOREVER PLAID by director Steven D. Kline, stage manager Alan Bolosan Campo, Tech	1287.487645
1	FOREVER PLAID, a KIT production directed by Steven D. Kline. Set, lighting and technical support by Derron F	1208.361921
2	I don't have all the answers, but you can make a difference in a child's life by sending a gift to The Salvation Ar	71.35250931
3	You can help make positive changes in the lives of at-risk youth in central Indiana. Some of them are growing	197.9581537
4	You can help make changes in lives at central Indiana. We hear about poverty, but what these young people a	156.8697479
5	You can make positive changes for at-risk youth in Indiana. We hear about economic poverty, but these youn	192.6421418
6	Help make changes in the lives of at-risk youth in Indiana! We hear about economic poverty, but what these	223.7802422
7	You can changes lives of at-risk youth in central Indiana who have grown up with not only economic but emc	282.7450008
8	Each year the graduating class contributes a gift to the law school. Sometimes this is an item of remembrance	194.2167521
9	Each year the graduating class gives a gift to the law school. This might be an item of remembrance, a scholars	141.3241913
10	regardless of the difficulties they may be facing, The Salvation Army offers grace and compassion to those in n	98.26851694
11	The Salvation Army offers grace and compassion to those in need in our community. You've been so generous	83.69398536
12	Like the Salvation Army offers grace and compassion to those in need, you've been so generous with your givi	94.0919769
13	The Salvation Army offers grace and compassion to those in need in our community. You've been so generou	83.69398536
14	The Salvation Army offers grace and compassion to those in need - regardless of difficulties they're facing. You	132.0462658
15	This is possible because we are a statewide network. Join us, send in your application today!	215.4916993
16	This is possible because we are a statewide member network. Send in your membership to join us!	115.8457872
17	All achievements are possible because of our large network. Please join us by applying today!	256.3986946
18	We are a statewide member network. Join us, by sending in your membership-application -today!	159.4522679
19	Join our statewide member network by sending in a membership-application today!	331.8865979
20	She wondered what would change for a child from the city if s/he was out there in the natural beauty and ser	295.5316648
21	She wondered what would happen if she could take one child from the city and place that youngster where s	194.0658188
22	She wondered what would happen if a city child was in her place in the midst of this beauty and serenity. Wh	164.3869212
23	What would happen if she could take one kid from the city and put that youngster where she was now,in the	196.2156644
24	She wondered what would change for a child if she could take them from the city and place that youngster in	211.8402145
25	My love affair with IRT is long. I joined the company in 1974 with my show, "Harvey" and I just completed my	182.180966
26	I joined the company of the IRT in 1974 with my first show, "Harvey" and I just completed my 45th producti	75.45334554
27	I joined the company of the IRT in 1974 with my first show and I just completed my 45th production at the I	78.54217511
28	I joined the company of the IRT, whom I love, in 1974 with my first show, "Harvey." I completed my 45th pro	88.88086571
29	I just completed my 45th production at the Indiana Repertory Theatre. I joined IRT in 1974 with my first sho	150.4459495
30	Scientific investigation is a part of the School of Medicine's mission. We take responsibility to push the boun	83.52973748
31	Scientific investigation is an integral part of the School of Medicine's mission. We take responsibility to furth	96.21247786
32	The School of Medicine aims to benefit society. We are aware of the weight we carry and are up for the task of	153.2249266

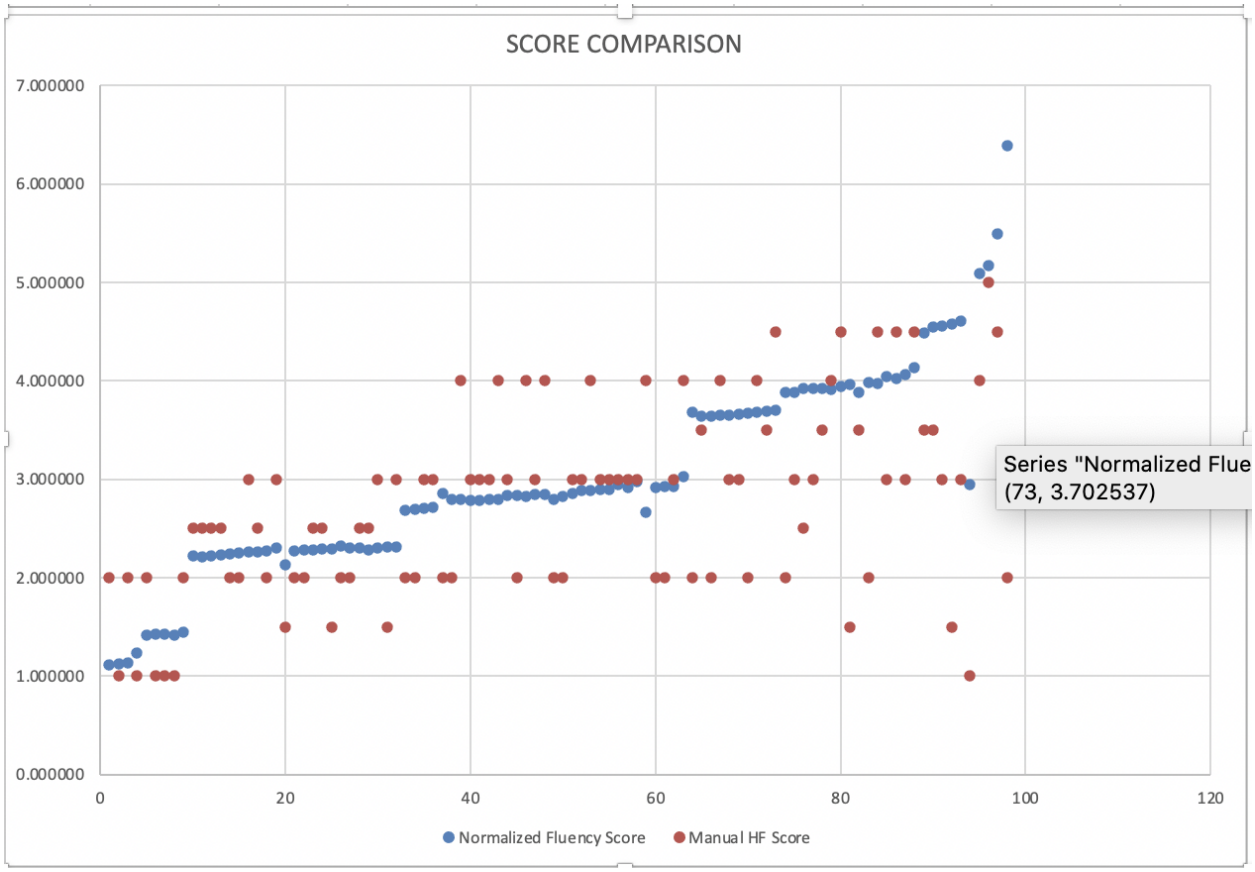
SUMMARY

We have used rule based readability indices, statistical model and neural network model over the same corpus to understand the complexity and fluency of sentences.

Then we compared the weighted mean score with manual human fluency scores for 100.

OBSERVATION

Manual human fluency scores were in range of those detected by experiment.



Output excel sheet:

Github link : https://github.com/anan-123/NLP-PROJECT/blob/main/Experiment_Human_fluency.xlsx