Module 6 & 7 - Applications of NLP

# Text Summarization:

## ROUGE SCORE:

Recall- Oriented Understudy for Gisting Evaluation:

ROUGE-N

N indicates the no. of 'N' grams which can be 1 and 2.

$$\text{Recall} = \frac{\text{Overlapping number of n-grams}}{\text{Number of n-grams in the reference}}$$

$$\text{Precision} = \frac{\text{Overlapping no. of n-grams}}{\text{Number of n-grams in the candidate}}$$

F1 Score (Harmonic Mean) of Rouge

$$F1 = \frac{2 * P * R}{(P+R)}$$

Rouge-1:

Candidate 1: Summarization is cool

Reference 1: Summarization is beneficial and cool

$$\text{Recall} = 3/5 = 0.6$$

$$\text{Precision} = 3/3 = 1$$

$$\text{Rouge-1} = 2 * R * P / (R+P) = \frac{2 * 0.6 * 1}{0.6+1} = 0.75$$

Rouge-2

Candiate 1: ~~Somero~~
(Summarization is), (is cool)

Reference 1: (summarization is), (is beneficial
(beneficial and), (and cool)

Recall : 1/4 = 0.25

Precision: 1/2 = 0.5

Rouge-2 = (2 * 0.5 * 0.25) / (0.5 + 0.25)

= 0.33

Note

∴ If More candidate's are there we have to calculate Mean Value.

ROUGE-L:

ROUGE-L is the Longest Common Subsequence (LCS) oriented. LCS is the longest Sequence of words that appear in both the candidate and references Summaries, while keeping the order of the words intact. It automatically includes the longest Sequence of words.

It is important to note that LCSes are not necessarily ~~consecutive~~ consecutive but still in order.

Ex:

Model output: "A fast brown fox leaps over a sleeping dog."

Reference summary: "The quick brown fox jumps over the lazy dog" (∵ lcs → b/w two sentences)

Longest common subsequence here is "brown fox over a dog"

Precision: $\frac{\text{Number of words in LCS}}{\text{Number of words in Model output}}$

$= 4/9 = 0.444$

Recall: $\frac{\text{No. of words in LCS}}{\text{No. of words in reference summary}}$

$= 4/9 = 0.444$

F1-Score / ROUGE-L $= \frac{2 * 0.444 * 0.444}{(0.444 + 0.444)}$

$= 0.444$

Ex 2:

Consider the Reference R and the candidate

Summary C

R: The cat is on the mat

C: The cat and the dog

ROUGE-1  precision = $3/5 = 0.6$

Recall = $3/6 = 0.5$

ROUGE-1 $= \dfrac{2 * P * R}{P+R} = 0.54$

ROUGE-2

R: (The cat) (cat is) (is on) (on the) (the mat)

C: (The cat) (cat and) (and the) (the dog)

$P = \dfrac{1}{4} = 0.25$

$R = 1/5 = 0.20$

ROUGE-2 -Score = 0.22

ROUGE - L :

Remember that the words are not necessarily consecutive

$P = 3/5$ , $R = 3/6$ ,

$F = \underline{0.55}$

## Machine Translation

BLEU Score for evaulating ~~New~~ Machine Translation.

Machine Translation is a standard task in NLP that involves translating a text from a source language to target language

BLEU (Bilingual Evaluation Understudy)

It is a score used to evaluate the translations performed by a machine translator. It is calculated by comparing the n-grams of the machine-translated sentences to the n-gram of human-translated sentences.

Note: BLEU scores is b/w 0 and 1. The score Best score is 0.6 or 0.7. The score closer to '1' is unrealistic in practice it raises a flag that your model is overfitting

Precision:

$$= \frac{\text{No. of corrected predicted words}}{\text{No. of total predicted words}}$$

Ex

Target Sentence: He eats an apple
predicted Sentence: He ate an apple

$p = 3/4$

But using Precision like this is not enough. There are two cases that we still need to handle:

Repetition:

The first issue is that this formula allows us to cheat: we could predict a sentence

Target sen: He eats an apple
predicted Sen: He He He

$P = 3/3 = 1$

Multiple Target Sentences:

In the NLP Models, we might be given multiple acceptable target sentences that capture these different Variations.

# Modified Precision or clipped precision

Ex

Target S1: He eats a sweet apple

Target S2: He is eating a tasty apple

Predicted Sentence: He He He eats tasty fruit

**Two things we have to do:**

→ we compare each word from the predicted sentences with all of the target sentences. If the word matches any target sentences, it is considered as correct.

→ we limit the count for each correct word the maximum no. of times that the word occurs in Target Sentence. It helps to avoid Repetition Problem.

| Word | Matching Sentence | Matched Predicted count | Clipped count |
|---|---|---|---|
| He | Both | 3 | 1 |
| eats | Target 1 | 1 | 1 |
| tasty | Target 2 | 1 | 1 |
| fruit | None | 0 | 0 |
| Total | | 5 | 3 |

Clipped Precision =

$$\frac{\text{clipped no. of correct predicted words}}{\text{~~Total~~ No. of total predicted words}}$$

$$CP = \frac{3}{6}$$

Compute BLEU Score for the given example.

Machine Translated Text: the picture the picture by me

Reference Text-1: This picture is clicked by me

Reference Text-2: the picture was clicked by me.

Step 1: Count no. of unique words in the predicted sentence.

Step 2: Unigram ~~clipped~~ Modified Precision

| Unigram | Count in Machine Translation | Max Count in Ref | Clipped count |
|---|---|---|---|
| the | 2 | 1 | 1 |
| Picture | 2 | 1 | 1 |
| by | 1 | 1 | 1 |
| Me | 1 | 1 | 1 |
| | Total: 6 | | Total: 4 |

$$P_1 = \frac{\text{clipped count}}{\text{Count in MT}} = \frac{4}{6} = \frac{2}{3}$$

Bigram Modified Precision. $n=2$.

clipped count
min(coont im MT, Max count in Ref)

| Bigram | Count in MT | Max Count in Ref | Clipped count |
|---|---|---|---|
| the picture | 2 | 1 | 1 |
| picture the | 1 | 0 | 0 |
| picture by | 1 | 0 | 0 |
| by me | 1 | 1 | 1 |
| | Total 5 | | Total 2 |

$P_2 = 2/5$

Trigram Modified Precision

$n = 3$

| Trigram | Count in MT | Max Count in Ref | Clipped count |
|---|---|---|---|
| the picture the | 1 | 0 | 0 |
| picture the picture | 1 | 0 | 0 |
| the picture by | 1 | 0 | 0 |
| picture by me | 1 | 0 | 0 |

4-gram Modified Precision

| 4 gram | Count in MT | Max Count Ref | Clipped Count |
| --- | --- | --- | --- |
| the picture the picture | 1 | 0 | 0 |
| picture the picture by | 1 | 0 | 0 |
| the picture by me | 1 | 0 | 0 |

$$P_4 = \frac{0}{3} = 0.0$$

Third step Estimate Brevity Penality

BP penalizes translations that are shorter than the reference translations.

$$BP = \min\left(1, \frac{\text{Machine Translation Output length}}{\text{Maximum Reference Output length}}\right)$$

$BP = \min\left(1, \frac{6}{6}\right) = 1$

4th Step

$$\boxed{BLEU\ Score = BP * \exp \sum_{n=1}^{4} w_i * \log(P_i)}$$

Substitute the values

BLEU Score = ~~1 * exp (0.25~~

→ $w_i$ = weight of n-gram precision of order i → take maximum value in Ref.

→ $p_i$ is the n-gram modified precession score of order i

→ N is the maximum n-gram order to consider

Substitute the values

BLEU Score =

$$1 * \exp\left(0.25 * \ln(2/3) + 0.25 * \ln(2/5) + 0 * \ln(0) + 0 * \ln(0)\right)$$

BLEU Score = 0.718

ROUGE vs BLEU:

→ BLEU focuses on precision: how much the words in the candidate model outputs appear in the human references.

→ ROUGE focuses on the recall: how much the words in the human references appear in the candidate model outputs.

Cosine Similarity:

Cosine Similarity is a measure of similarity between two non-zero vectors of an inner product space. It calculate the cosine of the angle between the two vectors and returns a value between -1 and 1, where '1' means the vectors are perfectly similar and -1 means they are perfectly dissimilar.

Suppose you have a following two sentences represented as 4-dimensional word vectors, $Y_1$ and $Y_2$ respectively.

$Y_1$: "The sun is bright", with vector values (0.6, 0.1, 0.7, 0.3)

$Y_2$: "The moon is ~~gleaming~~ shining" with vector values (0.4, 0.3, 0.5, 0.6)

Evaluate the cosine similarity between $Y_1$ and $Y_2$. Based on the similarity score, ~~do~~ would you consider $Y_1$ and $Y_2$ similar or dissimilar?

Given

$Y_1 = (0.6, 0.1, 0.7, 0.3)$

$Y_2 = (0.4, 0.3, 0.5, 0.6)$

Step 1: Calculate the dot product of $Y_1$ and $Y_2$

$Y_1 \cdot Y_2 = (0.6 \times 0.4) + (0.1 \times 0.3) + (0.7 \times 0.5) + (0.3 \times 0.6)$

$= 0.24 + 0.03 + 0.35 + 0.18 = 0.8$

Step 2: Calculate the Magnitudes of $Y_1$ and $Y_2$

Magnitude of $Y_1$ is

$||Y_1|| = \sqrt{(0.6)^2 + (0.1)^2 + (0.7)^2 + (0.3)^2}$

$= \sqrt{0.36 + 0.01 + 0.49 + 0.09}$

$= \sqrt{0.95} = 0.975$

Magnitude of $Y_2$ is

$|Y_2|| = \sqrt{(0.4)^2 + (0.3)^2 + (0.5)^2 + (0.6)^2}$

$= \sqrt{0.16 + 0.09 + 0.25 + 0.36}$

$= \sqrt{0.86} \approx 0.927$

Step 3: Calculate the cosine similarity

$$\text{Cosine Similarity} = \frac{Y_1 \cdot Y_2}{||Y_1|| \, ||Y_2||}$$

$$= \frac{0.8}{0.975 \times 0.927} = 0.885$$

The cosine similarity between $Y_1$ and $Y_2$ is approximately 0.885. Since this value is close to '1', both $Y_1$ and $Y_2$ have same meaning and context.

So "The Sun is bright" and "the moon is shining" have a high degree of similarity based on their vector representations

Ex 2: The sky is blue (0.7, 0.2, 0.4, 0.6)
The ocean is vast (0.5, 0.3, 0.6, 0.8)

Jaccard

$$\text{Sim}(D, Q) = \frac{\Sigma(a_i * b_i)}{\Sigma a_i^2 + \Sigma b_i^2 - \Sigma(a_i * b_i)}$$

## Named Entity Recognition:

It is a part of NLP process. The primary objective of NER is to process structured and constructured data and classify these named entities into predefined categories.

. NER deals with
- Named entity recognition / detection
→ identifying a word & series of words in a document
- Named entity classification - classifying every detected entity into predefined categorie

NLP has 3 categories ← Syntax, Semantics, Speech

NER helps in the Semantic part of NLP, extracting the meaning of words.

Types of NER Entity:

Named Entity Recognition Models, Named categorized entities into various predefined types.

Person(PER): Identifies individuals names including first, middle, LN, titles etc

Ex Lakshmi Harika

**Organization (ORG):**

Recognizes companies, institute, government agencies, and other organized group.

Ex: GOOGLE,

**Location (LOC)**: Detects geographical locations including countries, cities, state, address and landmarks.

Ex: LONDON

**Date (DATE)**: Extracts date in various format

Ex. 2024-01-01

**Time (TIME)**: Identifies time expressions

Ex: 3:00PM , 15:00

**Quantity (QUANTITY)**: Recognizes numerical quantities and units of measurement.

Ex: 10 kilogram

**Percentage (PERCENT)**: Detect percentage

Ex 50% & 0.5

**Money (MONEY)**: Extracts money values and currencies $100, €50

**Other (MISC)**

A catch-all category for entities that don't fit into the other types

Ex: iphone 15

Ex:
Apple ORG today DATE announced the second QUANTITY generation iPhone(SE) COMM a powerfull new iphone E COMM fealuring a 4.7-inch QUANTITY Retina HD display

dialogue state architecture:

AI assisted chat-bots utilize dialogue state architecture to structure and manage interactions b/w the user and AI system.

It has five component

1. Natural Language Understanding
2. Dialogue State Traker
3. Dialogue Policy Manager
4. Natural Language Generation
5. knowledge Base and External API Integration

Ex: Travel Booking chat Bot

Imagine a user wants to book a hotel room using a chatbot.

1. NLU: The user types, "I'd like to book a hotel room in the new York".
The NLU identifies the intent ("book hotel") and extracts entities ("location: New York").

2. Dialogue State Tracker:

The tracker records that the user wants to book a hotel room in New York. The state is kept as context, so future queries like "Can it close to Central Park"? refer to hotel booking in New York.

3. Policy Manager:

Based on the state, the policy manager decides that the next action should be asking for additional details, like check-in date and number of guests

4. NLG: Generate the responses, "Got it when will you be checking in, and how many guests will be staying?

5. Knowledge Base Interface: As the conversation progresses, if the user asks about available amenities, the chatbot can retrieve specific details from the hotel amenities database.

6. Dialogue Manager:

Manages each of these steps, ensuring that the chatbot maintains the conversation flow by following up appropriately or redirecting the user's input is unclear.