Natural Language

Practical Classes

Luísa Coheur

2025

P3

From Lexical Distances to Evaluation Metrics

Image generated by ChatGPT

- **Summary**:
    - Lexical Distances/Similarities
    - Evaluation Metrics in NLP
    - Evaluation of evaluation metrics

- **Operational objectives**:
    - Apply similarity/distance metrics to check how "close" words/sentences are
    - Understand that different tasks benefit from different evaluation metrics
    - Practice the application of some evaluation metrics

- **This class needs**: paper, a pen/pencil, computer, and a gmail account for colab

- **Class material**: this guide and a notebook

Your first client

Surprise: the message is from your NL prof., Luísa Coheur:

> I've seen your post in LinkedIn. I have a NLP problem and need your help. I've prepared an Escape Game for one of the labs about metrics and I am unable to say if it is too easy or not, or even if it is satisfactory. Can the "NLP detective" help on this, please? If so, I will send you the Escape Game link. Thanks,
>
> Luísa Coheur

Wow! An NLP escape game.
While you wait for the link, you decide to do some pen-and-paper work before moving on.

Towards expertise

1. Fill in the following matrix and find the Minimum Edit Distance (MED) between the words "BIO" and "BIFE".

   Explicitly indicate in the table:

   a) the MED value;

   b) the path(s) that led to that value.

   Then, briefly explain the results in terms of insertions/removals/substitutions.

		B	I	O
	0	1	2	3
B	1			
I	2			
F	3			
E	4			

2. Consider the following part of a table used to calculate the MED between two words, where:

   a) one of the words contains the character sequence C_1 C_2 and the other word contains the character C_3;

   b) the calculated values are k, m, n, o.

	...	C_1	C_2
...	k	m	n
C_3	o	p	??

   (a) Assuming $k = p$, $o \geq k$ and $m \geq k$, what can you conclude about characters C_1 and C_3?

   (b) Assuming $C_3 \neq C_2$ and $m > n > p$, what value should be assigned to the cell with ?? ?

3. Consider sentences S_1 = "*the zombie in their room*" and S_2 = "*the flower in the room*". Apply the Jaccard measure — $J(s,t) = |s \cap t|/|s \cup t|$ — between these two sentences. Always consider sets (no repetition of elements). Whitespaces should be used to split sentences into words. Consider as comparing units:

   (a) words

   (b) characters

   (c) bigrams (example, bigrams("the cat") = {th, he, ca, at})

4. Consider systems A and B, designed to extract named entities of type Location from text. Considering a given Golden Collection and system A and B results in the table below, compute the Precision, Recall and F-measure (F1) of both systems.

Golden Collection	System A	System B
Igreja de São Domingos, Rua Augusta, Praça do Comércio, Rua do Arsenal, Santa Engrácia	Igreja de São Domingos, Praça do Comércio, Rua do Arsenal, Santa Engrácia, S. Julião, KnitPro	Rua Augusta, Praça do Comércio, Rua do Arsenal

5. Consider a *Sentiment Analysis* task with three categories: Positive, Neutral, and Negative. The model is evaluated on 12 reviews, and the resulting confusion matrix is:

Reference (true)\Predicted	Positive	Neutral	Negative
Positive	2	2	1
Neutral	1	6	3
Negative	1	0	4

(a) Calculate Precision, Recall and F1 for each category;

(b) Calculate Macro-Precision, Recall and F1 (*macro averaging* treats all classes equally, useful when each category is equally important; to compute the metrics, apply the arithmetic mean across categories);

(c) Calculate Micro-Precision, Recall and F1 (*micro averaging* treats all samples equally, giving more weight to frequent classes; to compute these metrics, sum the true positives, false positives, and false negatives across classes).

6. You also decide to give a look at the given jupyter notebook, P3_metrics.ipynb, and this was a good decision, as you will use it to solve one of the Escape Game quizzes.
You run it (use Google Colab[1]; otherwise, install pytorch[2]).

Let's go!

You consider that you are now able to evaluate the escape game. You go to:

`https://forms.gle/11np8XfaZ1vcvy1s8`

and play it.

Are we done?

You found the escape game a misery, quite meh, yet you don't have the courage to tell your professor that she needs to put more effort. You reply to the email saying it's great. She thanks you very much and says she will send you a snowball candy. You think that maybe it would be a good idea to explain that you don't pay bills with a snowball candy, but right after, you receive another email from a potential client. A real wow! Next mission will be great!

---

[1] `https://colab.research.google.com`
[2] `https://anaconda.org/pytorch/pytorch`