FRAMEWORK FOR MULTILINGUAL UD PROBING

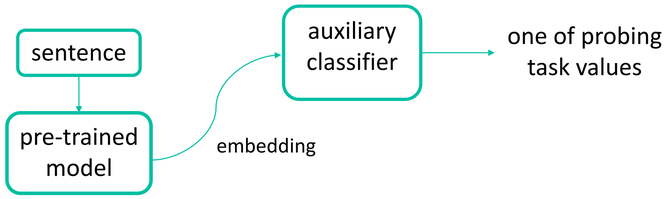
by Kniazkova Viktoria

BACKGROUND

Modern language models show great results in generating coherent text with correct grammar and style. That led to a question: does it mean that model knows something about morphosyntactic properties of the language?

PROBING

Probing tasks (Conneau et al., 2018) help in finding out what linguistic properties are possibly encoded in neural models.



If the classifier succeeds, it means pre-trained model encodes readable tense information in sentence embeddings.

MY GOAL

Create a framework that will allow you to conduct (with minimum extra work) large-scale studies based on probing tasks. And more precisely – come up with a way of constructing sentence datasets for custom probing task for any language. For that I need:

- Select some corpora with many languages presented and with the consistent annotation.
- Create a tool that transforms a set of sentences suitable for probing task into a file used by a classifier.
- Create a tool that filters sentences by probing task.

HOW ABOUT UD?

> 100 languages

can help you make probing of many different languages

universal annotation

fixed tagsets make cross-linguistical comparison more easy

grammatical relations annotation

so you can explore not only morphology but also syntax

CONLL-U & PROBING TASK FORMAT

From the set of annotated sentences to dataset for classifier

ID	FORM	LEMMA	UPOSTAG	XPOSTAG	FEATS	HEAD	DEPREL	DEPS
1	They	they	PRON	PRP	Case=Nom|Number=Plur	2	nsubj	2:nsubj|4:nsubj
2	buy	buy	VERB	VBP	Number=Plur|Person=3|Tense=Pres	0	root	0:root
3	and	and	CONJ	CC	_	4	cc	4:cc
4	sell	sell	VERB	VBP	Number=Plur|Person=3|Tense=Pres	2	conj	0:root|2:conj
5	books	book	NOUN	NNS	Number=Plur	2	obj	2:obj|4:obj
6	.	.	PUNCT	.	_	2	punct	2:punct

↓

PART	TAG	SENTENCE
tr	PRES	" Okay now , okay , settle down .
tr	PRES	" Ugh .... " he mumbled .
tr	PAST	Now it frames her slender face like the perfect picture frame , causing her eyes to
tr	PAST	The movement of his lips tickled the shell of her ear .
tr	PAST	I perch on the edge , clinging to my purse like a lifeline .
tr	PRES	He touches my elbow .
va	PRES	High above a mass of rock forms the roof .
va	PAST	They still bustled about , mindless of the weather .
va	PRES	He forgets himself and smiles .
te	PAST	Why deny them both ?
te	PAST	He heaved once , violently .
te	PRES	' None of the other staff there today recognise Thomas from the photo . '

MORPHOLOGY PROBING TASKS

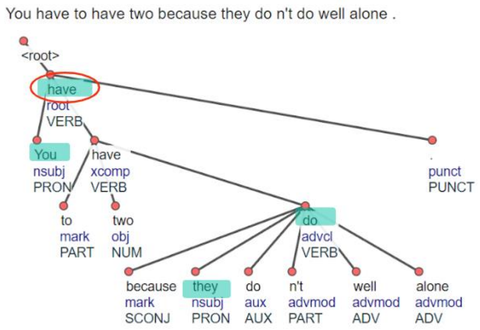
Does the model contain information about <grammar category>?

Sentence filtration based on the values of selected grammar category.

```
def generate_probing_file(category,
                          conllu_path,
                          result_path,
                          partition=(0.8, 0.1, 0.1),
                          shuffle=True):
```

TREE DISAMBIGUATION

classifying sentences by the token which is the closest to the *root*.



CUSTOMIZABLE PROBING TASKS

Tool that would allow you to construct datasets for more complex and customizable probing tasks. It should be able to filter conllu file by syntax and morphology at the same time.

UD QUERY

Based on Grew-match my implementation of query language made in the form of python dictionaries

```
{
        'V': {},
        'S': {},
        'BY': {'lemma': '^by$'},
        'N': {},
}
{
        ('V', 'S'): {'deprels': '^aux:pass$'},
        ('V', 'N'): {'deprels': '^obl$'},
        ('N', 'BY'): {'deprels': '^case$'},
}
```

An example of a query for a passive construction with a nominal by-agent

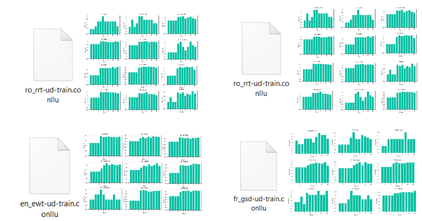
**This algorithm is still in development and need to be tested.*

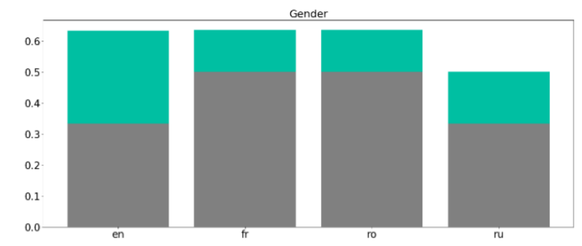
EXAMPLES OF PROBING RESULTS

This illustrates how many probing experiments can be conducted only by running my function for 4 conllu files in a loop:

Each bar chart shows accuracy score of classification on every layer of M-BERT model.

(this visualization is just to show how large-scale experiments can be conducted with minimal effort)





The best accuracy scores for Gender probing in English, French, Romanian and Russian. Grey columns on the chart represent the level of random guessing for every language.