Brain - computer interface

① BIC

Motivation (why?)

Patients with condition like brainstem stroke or ALS often lose the ability to speak

But their language cognition remains intact

Aim

make them be able to communicate naturally again

② Basic Idea

Step1: record brain activity using implanted electrodes (ECoG or Utah arrays)

Step2: Target language-related regions (motor area / Broca's area)

Step3: Feed these signals into NLP models (Transformer / LSTM / GRU)

often use GRU, Cause it's faster than Transformer / LSTM

Step4: Convert neural signals -> phonemes (音素) -> words -> sentences

Essentially: Neural signals -> sequence modeling -> text

③ Technical Framework

△1 Feature Extraction from Neural Signals

ECoG / cortical recordings provide high-resolution temporal signals

↓

Preprocessing (include Filtering -> Band-power extraction -> spatio-temporal feature encoding)

↓

Resulting feature from a time series suitable for sequence models

△2 To map neural activity to linguistic units, researches use:

RNNs (LSTM / GRU) - historically used

Temporal convolution networks

Transformer

But, why a sequence model?

└ because neural signals and phonemes are both time-aligned sequences often a model predicts probilities over phoneme classes at each time step

△3 Use of CTC (Connectionist Temporal Classification)

why use it?

neural input and phoneme output have different lengths and we don't know their precise alignment

how does CTC work?

※ Introducing a special "blank" sign

1. Instead of outputing character directly, output layers output a probability distribution which cover all the possible characters (eg. 26 English letters + 1 blank sign)

↓

2. Aligned multiple paths

we have these rules:

rule1: allow repeated characters

rule2: allow the insertion of blank sign

example:
- c - aa - tt -
- cca - t -
- ca - tt
- c - - at -

all of them ⟹ "cat"

3. loss function: compute the sum of all the valid paths' probabilities

goal of the loss function:

maximize the sum of probabilities for all paths that can be mapped to a real label sequence (such as 'cat')

4. reasoning / decoding: from output to final texts

- best path decoding
- Beam search decoding

---

(4) Language models in the loop

why?

decoding only with CTC will result in missing context.

no grammer, no correction of ambignous phonemes

so,

- lightweight LM (n-gram) for real-time decoding

Large Transformer LM for re-scoring

↓

this improve: syntax, semantics, word choices, naturalness

Ultimately:

NLP LMs help turn noisy neural predictions into fluent language

(5) Summary

Neural signals

↓

Feature extraction

↓

Seq2Seq model

↓

CTC

↓

Beam search

↓

LM Re-scoring

↓

Final text output