Learning to Abstract with NVIB

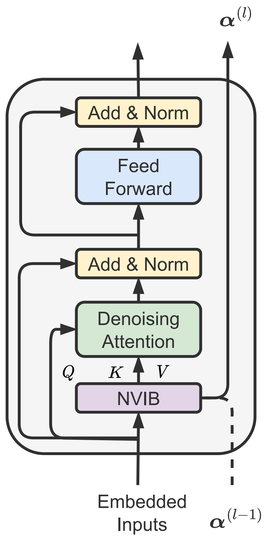
Melika Behjati*, Fabio Fehr*
James Henderson

Overview

- **Problem:** Language models' abstraction is defined and restricted by their tokenization. Training models for each tokenization is costly.
- **Why:** Different levels of abstraction are important for different linguistic tasks and applications.
- **How:** Stacked NVIB layers learn to compress to different levels of abstraction at different layers within the same model.
- **Results:** NVIB: discovers words groupings; is more robust to noise; and is more linguistically informed than standard Transformers.

Model

Nonparametric Variational Information Bottleneck (**NVIB**) is an information theoretic regulariser for attention-based latent representations. We apply NVIB to stacked self-attention layers and train it on a character-level denoising reconstruction objective.



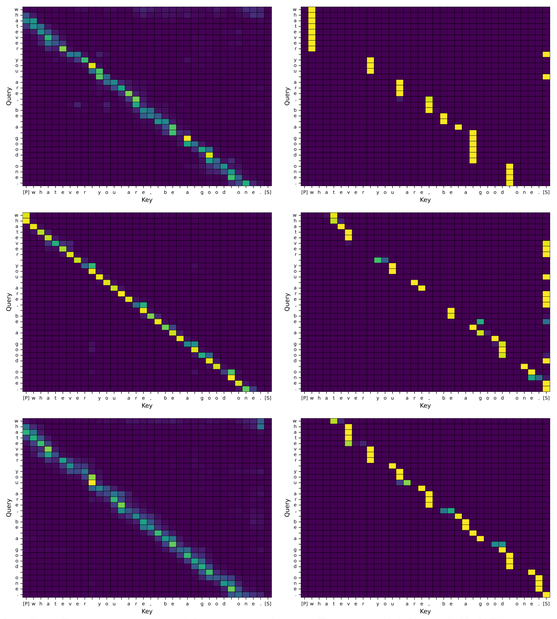
Nonparametric Variational Information Bottleneck

discovers **words** with more **linguistically informed** and **robust** representations.

Our Paper!

Our Demo!

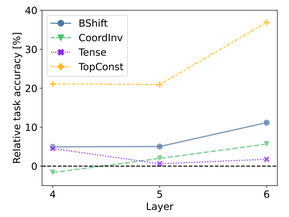
Attention Map Visualisations



Left: Standard self-attention. **Right:** With NVIB.

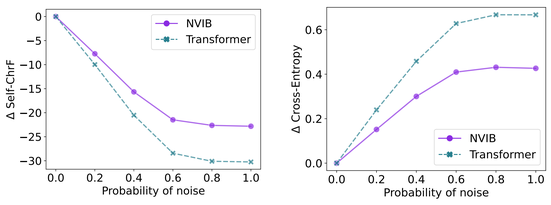
Linguistic Probing

Relative performance of NVIB over Transformers for a subset of SentEval tasks.



Robustness Analysis

Relative performance under input perturbations.



idiap RESEARCH INSTITUTE
EPFL
nccr evolving language NATIONAL CENTRE OF COMPETENCE IN RESEARCH
EMNLP 2023