Title:
Cross-lingual Similarity of Multilingual Representations Revisited

Authors:

Maksym Del and Mark Fishel

Institute of Computer Science

University of Tartu, Estonia

Presented by:

Sohrab hossain,17101321

Introduction

- Use CCA/CKA to find a similarity between parallel sets of different representations of the same data
- Discuss the general problems of CKA/CCA
- Averaged Neuron-Wise Correlation (ANC) as an alternative
- Dataset: English, French, Estonian, and Bulgarian
- Experiment 1
- Experiment 2
- Experiment 3

Related Work

- CCA as a method for measuring canonical correlations between two sets of random variables
- Multilingual neural machine translation system.
- Measuring canonical correlations between two sets of random variables.

SIMILARITY INDEXES BACKGROUND:

- Neuron Definition: Neurons are defined as vectors representing the values a neuron takes over a dataset
- Layer Definition: Layers in neural networks are defined as lists of vectors comprising neurons at a specific depth.
- CA/CKA Indexes and Subspaces: Canonical Correlation Analysis (CCA) and Centered Kernel Alignment (CKA) leverage the idea of subspaces spanned by neurons.
- Dominant Correlations and Eigendecomposition: The initial step for these methods involves centering each neuron in layer representations.

Method:

- Proposing an alternative method that neurons in representations for different languages are aligned one-to-one a priori.
- Computing individual correlations between pairs of English and, for example, French neurons.
- Calculating an average score to gauge overall cross-lingual similarity.
- Taking absolute values of correlations to account for potential polarity changes in subsequent layers.

Dataset

- 50M sentences uniformly sampled from four languages: English, French, Estonian, and Bulgarian
- Trained the model for 1M batches of 512 sentences from the CC100 dataset using two Nvidia A100 GPUs

Result:

The study reveals the limitations of Centered Kernel Alignment (CKA) in revealing cross-lingual similarity in multilingual models. It introduces Average Neuron-Wise Correlation (ANC), which surpasses CKA's sanity check and generalizes the "first align, then translate" pattern to Causal Language Models and Masked Language Models.