Paper Title: Cross-lingual Similarity of Multilingual Representations Revisited

Paper Link: https://aclanthology.org/2022.aacl-main.15.pdf

Summary:

1.1 Motivation:

The aim of this paper evaluate the degree to which multilingual language models' cross-lingual representations are similar.

1.2 Contribution:

The author challenges the appropriateness of similarity indexes such as Canonical Correlation Analysis (CCA) and Centered Kernel Alignment (CKA) for the specific goal of measuring cross-lingual representations in multilingual language models. Points out that the assumptions of CKA/CCA methods do not align well with the objective of explaining zero-shot cross-lingual transfer.

1.3 Methodology:

The study suggests a novel technique for measuring cross-lingual representations in multilingual language models: Average NeuronWise Correlation (ANC), which can be used instead of Canonical Correlation Analysis (CCA) and Centered Kernel Alignment (CKA). To overcome the drawbacks of current techniques, ANC is predicated on the idea that neurons in representations for various languages are a priori aligned one to one. This is in line with the objective of zero-shot cross-lingual transfer learning and improves interpretability by breaking down the similarity index into correlations of individual neurons. When computing ANC, correlations between neuron pairs speaking different languages are calculated, and the average score is obtained while accounting for any possible negative correlations. Through sanity checks, the paper validates ANC, confirming established patterns from the literature and proving its usefulness in situations where CKA fails. Moreover, large-scale multilingual models are trained using ANC, which shows that the general "first align, then predict" pattern holds true for models with varying sizes and training goals. In single multilingual language models, ANC shows promise as a tool for cross-lingual similarity analysis.

1.4 Conclusion:

The study highlights the limitations of existing similarity measures like Centered Kernel Alignment and Canonical Correlation Analysis in assessing cross-lingual similarity in multilingual models. It introduces Average Neuron-Wise Correlation (ANC), a promising tool for robust analysis.

2 .limitation:

- Empirically, the study reveals that Centered Kernel Alignment (CKA) falls short in identifying connections related to similarity following architectural changes that do not adversely affect the model's performance.

- Conceptually, the inadequacy of interpretability and unsatisfactory foundational assumptions in Canonical Correlation Analysis (CCA) and CKA is highlighted.