Título: Studying the Dependence of Word Embeddings Dimensions on the Target of NLP Tasks

Data: 03/05/2022

Horário: 14h00

Local: GREAT

Resumo:

In many human languages, most information about the structure of texts can be represented in

the form of linguistic units, particularly words. The most common approach to representing word meaning is through vector semantics, which uses word embeddings, indicating each word as points in a multidimensional semantic space. The existing techniques that generate word embeddings can be divided into count-based, static/classic, and contextual (including the popular Language Models, such as BERT and GPT). These groups differ in how they represent words as vectors, and their technical evolution in high quantity has significantly impacted not only the Natural Language Processing (NLP) community but also the entire Artificial Intelligence community. Evaluating the learned representations helps to identify the critical differences between word embedding models to choose the best one for a specific task. The evaluation of word embeddings is complex, and two approaches exist: intrinsic and extrinsic. The former evaluates vectors' structure in space, while the latter measures their performance on specific tasks. The evaluation process remains complex, and the different methods lead to varying structures with pros and cons for solving particular tasks. Considering this scenario and the rapid technical progress in NLP textual representations, this work investigates the dependencies and correlations between word embeddings on the target of NLP tasks. The study aims to efficiently analyze, before using certain word embeddings to represent a corpus or corpora of an NLP task, how to initially verify if the vectors' dimensions (features) depend on the final task and, therefore, provide quality representation. Based on this, this work explores two main research questions and presents several interesting findings across an extensive set of experiments.

Banca examinadora:

- Prof. Dr. José Antônio Fernandes de Macedo (MDCC/UFC - Orientador)
- Profa. Dra. Ticiana Linhares Coelho da Silva (UFC - Coorientadora)
- Prof. Dr. César Lincoln Cavalcante Mattos (UFC)
- Prof. Dr. João Bosco Ferreira Filho (UFC)