Semantic Loss in Autoencoder Tree Reconstruction Based on Different Tuple-Based Algorithms

Hiram Calvo
Ramón Rivera Camacho
Ricardo Barrón Fernández

Centro de Investigación en Computación (CIC), Instituto Politécnico Nacional (IPN). México. hcalvo@cic.ipn.mx

Abstract

Current natural language processing analysis is mainly based on two different kinds of representation: structured data or word embeddings (WE). Modern applications also develop some kind of processing after based on these latter representations. Several works choose to structure data by building WE-based semantic trees that hold the maximum amount of semantic information. Many different approaches have been explores, but only a few comparisons have been performed. In this work we developed a compatible tuple base representation for Stanford dependency trees that allows us to compared two different ways of constructing tuples. Our measures mainly comprise tree reconstruction error, mean error over batches of given trees and performance on training stage.

Keywords: *Semantic reconstruction, Parsing Structuring word embeddings*