



Workshop EPFL-Inria January 9 and 10, 2020, Montbonnot

Alberto Duran Garcia, EPFL

Title: « Crosswalk: Learning Aligned Word and Entity Embeddings »

Abstract:

Cross-lingual word embeddings are vector representations for the words of multiple languages such that similar words have close-by vectors both within and across languages. They have recently been at the core of major advances on machine translation in the last years.

Word embeddings are often learned from text corpora wherein span of words are linked to entity nodes in a knowledge graph. For instance, the mention "l'hexagone" in Wikipedia is often linked to the entity node France in the Wikidata graph. Different to previous works, the project we propose here, Crosswalk, will also capitalize on the alignment between the multilingual corpora and the underlying knowledge graph.

Typically, applications, such as entity linking, align independently learned entity and word embeddings in a common space by applying some transformation. This is an advantage of Crosswalk embeddings, as both word and node latent representations are learned to naturally live in the same space.

The talk will provide a description of the problem, the technical details of our approach and the current progress of the project. We will also elaborate on other applications of interest to Crosswalk.