

Basic Idea

The underlying idea is to independently train the embeddings in different languages using monolingual corpora, and then map them to a shared space through a linear transformation. The algorithm tries to learn the matrices for different languages which are used to map the monolingual word embeddings to the shared space. We use an unsupervised algorithm for learning the matrices, but in order to make improvements in the results we try to add minimal supervision in the algorithm.