Distributional Memory: a generalized framework for corpus-based semantics

Distributional semantic models have proved to successfully address many different tasks, from synonymy recognition to analogy learning. Their many achievements notwithstanding, the typical approach in the field has been to treat each semantic task (or set of closely related tasks) as a separate problem, that requires its own corpus-derived model and algorithms, both highly optimized to achieve the best performance in a given task, but lacking generality. Distributional Memory (DM) is a framework for corpus-based semantics that aims at overcoming this situation, moving towards a unified model that is able to generalize over different, prima facie incompatible semantic spaces. DM consists of a common distributional structure formed by word-link-word tuples representing ternary relations between linguistic units and formalized as a third-order tensor. Semantic spaces are then generated from this common structure, each acting as a different view on the same underlying distributional memory. DM has been tested on the largest and most various array of semantic tasks ever addressed by a single distributional model, achieving state of the art, sometimes even optimal results, without any task-specific tuning.