

Abstract: Learning and representing semantics is one of the most important tasks that significantly contribute to some growing areas, as successful stories in the recent survey of Turney and Pantel (2010). In this thesis, we present an innovative (and first) framework for creating a multimodal distributional semantic model from state of the art text-and image-based semantic models. We evaluate this multimodal semantic model on simulating similarity judgements, concept clustering and the newly introduced BLESS benchmark. We also propose an effective algorithm, namely Parameter Estimation, to integrate text- and image-based features in order to have a robust multimodal system. By experiments, we show that our technique is very promising. Across all experiments, our best multimodal model claims the first position. By relatively comparing with other text-based models, we are justified to affirm that our model can stay in the top line with other state of the art models. We explore various types of visual features including SIFT and other color SIFT channels in order to have preliminary insights about how computer-vision techniques should be applied in the natural language processing domain. Importantly, in this thesis, we show evidences that adding visual features (as the perceptual information coming from images) is comparable (and possibly better) than adding further text features to the advanced text-based model; and more interestingly, the visual features can capture the semantic characteristics of (especially concrete) concepts and they are complementary with respect to the characteristics captured by textual features.