Linear Algebraic Structure of Word Meanings

Sanjeev Arora, Princeton University

Abstract

In Natural Language Processing (NLP), semantic word embeddings use vectors to capture the meanings of words. They are often constructed using nonlinear/nonconvex techniques such as deep nets and energy-based models. Recently, Mikolov et al. (2013) showed that such embeddings exhibit linear structure that can be used to solve "word analogy tasks" such as man: woman :: king: ??. Subsequently, Levy and Goldberg (2014) and Pennington et al. (2014) tried to explain why such a linear structure should arise in embeddings derived from nonlinear methods.

We provide a new generative model for language that gives a different explanation for how such linear algebraic structure arises. This new model also casts new light on older methods for generating word embeddings, such as the PMI method of Church and Hanks (1990). The model has surprising predictions, which are empirically verified. It also suggests a new linear algebraic method to detect polysemy (words having multiple meanings).

We think that our methodology and generative model may be useful for other NLP tasks and understanding the efficacy of other neural models.

Joint work with Yuanzhi Li, Yingyu Liang, Tengyu Ma, and Andrej Risteski (listed in alphabetical order)