How flexible are categorical models of meaning?

March 1, 2015

1 Introduction

The distributional compositional model of meaning of Clark et al. (2008) is motivated by category theory, and enjoys hence many useful mathematical properties. For instance, the spurious ambiguity in CCG has no impact on the final semantic representation of a sentence. In practice however, implementations often depart from the original framework: after all, if stacking two vectors yields better results than taking their outer product, why should we bother abiding by this mathematical orthodoxy?

In this poster, we argue that we can get both the theoretical guarantees of category theory and the flexibility of alternate models of meaning using the notion of *free compact closed category*. This notion can be used to recast alternate models of meaning, such as neural networks for instance, in the type-driven framework. This observation calls for a generic tool easing the implementation of categorical models of meaning, a linguistic counterpart to the Quantomatic software used in quantum physics.

2 The traditional model: linear maps and the tensor product

The most popular semantic category for distributional models is the autonomous category of finite-dimensional vector spaces and linear maps between them, denoted by Vect, with the tensor product $\otimes$ as monoidal operation. This tradition has been initiated in the early works of Clark and Pulman (2007); Clark et al. (2008); Coecke et al. (2011).

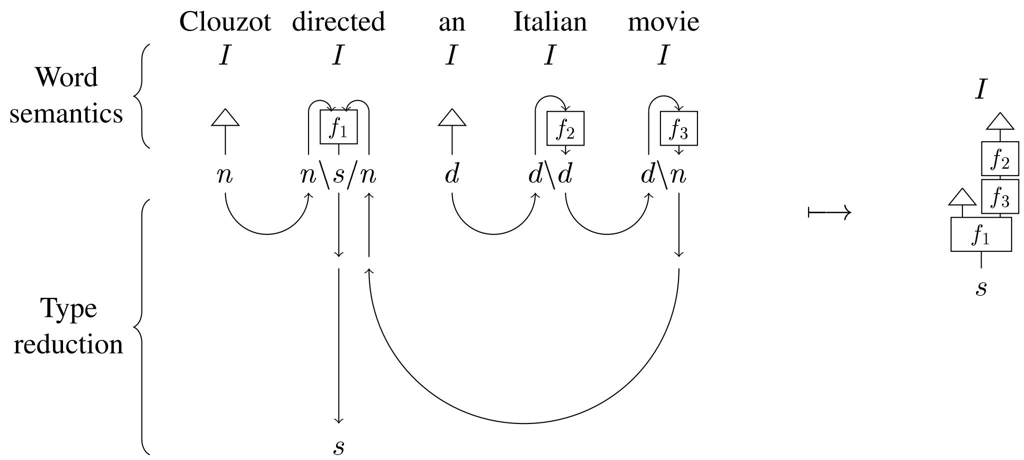
The main problem with this category is that the dimensions of the vector spaces associated with complex syntactic types are prohibitively large. Dimensionality reduction techniques have been devised, but the nature of the category restricts the range of possible algebraic operations. For instance, a very useful baseline for the composition of word vectors consists in taking the sum of the vectors for each word in the sentence. This baseline cannot be recast in Vect because $u + v$ is not a linear function of $u \otimes v$.

Another argument against linearity is that it ignores some kind of structure in the vectors, such as the famous $\overrightarrow{King} - \overrightarrow{Man} + \overrightarrow{Woman} \simeq \overrightarrow{Queen}$ of Mikolov et al. (2013). When this kind of relation holds, one might want to represent adjectives such as *female* by a function adding a vector to their argument, but it is impossible if our arrows are simply linear maps: we would need affine maps.

3 Alternate models of meaning

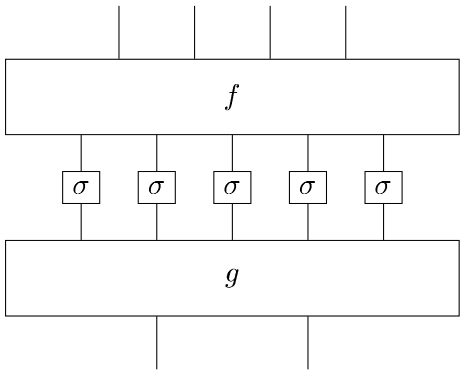
Replacing the tensor product by the direct sum $\oplus$ solves the problems mentionned above: $u + v$ is a linear function of (u, v). Moreover, the dimension of $U \oplus V$ is $\dim U + \dim V$, which reduces the dimension of compound spaces.

However, there is no equivalent of tensor contraction (also called *counit*) for the direct sum. This is where free compact closed categories come in handy: we can embed our candidate model of meaning in a larger category where units and counits exist. These formal units and counits are then eliminated using the yanking equations:

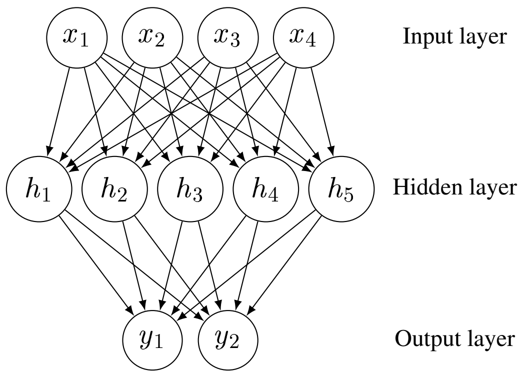


We can go further by allowing nonlinearities, which leads us to neural models of meaning.

The following figure compares the classical representation of a neural network to its string diagram in Net, the category of vector spaces and continuous maps between them.



(a) An arrow in Net.
f and g are affine maps and σ is a nonlinearity.



(b) Its traditional representation as a neural network

References

Clark, S., B. Coecke, and M. Sadrzadeh (2008). A compositional distributional model of meaning. In *Proceedings of the Second Quantum Interaction Symposium (QI-2008)*, pp. 133–140.

Clark, S. and S. Pulman (2007). Combining symbolic and distributional models of meaning. In *AAAI Spring Symposium: Quantum Interaction*, pp. 52–55.

Coecke, B., M. Sadrzadeh, and S. Clark (2011). Mathematical foundations for a compositional distributional model of meaning. *Linguistic Analysis 36*(1-4), 345–384.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean (2013). Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pp. 3111–3119.