Distributional semantics and compositionality

Summary

The main question is how we can represent the meaning of a sentence or phrase using distributional, vector-based models? One way to answer this question is to extend the hypothesis of the distributional meaning from the word level to the sentence level. Baroni & Zamparelli developed a vector-based framework where the adjective-noun combinations are represented by multiplying the vectors' representation of the noun and the adjective. A neural network is then used to learn this matrix representation of the noun-adjective combination from a corpus. In other words, the context representation of noun-adjective combinations can be learned. It is unclear if such methodology can be extended to predict the context of more complex syntactic combinations. One argument against the distributional framework described above is that such a framework —that represents a sentence\complex structure from its context— is inconsistent with the compositionality nature of the natural language. The principle of compositionality states that the meaning of a complex linguistic expression is based on its constituents' meanings and the rules used to combine these constituents. In that sense, we need to examine different methods to combine word vector representation to represent a complex linguistic expression.

In compositional semantic, Montague semantics tries to give a meaning representation of a sentence based on its syntax structure. In compositional distributional semantic, the distributional models are extended by using the syntax rules that combine the semantics of the sentence constituents to represent the semantics of the entire sentence. This syntax-semantics interface is type-driven. For example, the adjective can be seen as a function that takes a noun and returns another noun. By combining formal and distributional semantics we are able to differentiate between "a man is killed by a dog" and "a man killed a dog", these two sentences are completely different and integrating formal semantics enabling us to capture such differences.

Different methods could be used to represent the adjective-noun compound, the tensor product is one of them. One of the advantages of using a tensor product that it is non-commutative so it reserves the order of the words. The adjectives, in this case, is a matrix (a linear map) that is when it is multiplied by a noun vector, it returns a modified version of the noun vector. Another motivation for using the tensor product is the ability to represent the multilinear map; a function that takes multiple arguments like a transitive verb. Building relation function for complex syntactic types is an open question.

Based on the above, the distributional semantics model can be combined with compositionality such that for a given sentence, there exists a syntax-driven linear map from the context of the constituents' vector representation to a single vector space for the whole sentence. Such a transition can be implemented by combinatory categorial grammar (CCG).

The compositional distributional semantics requires extensive knowledge in both mathematics and linguistics. It also needs a tagged corpus to be able to integrate the formal semantics with the distributional. The compositional distributional semantics cannot cover

the instance of events and situations in the sentences, also it cannot represent the meaning variations liked to speaker intent (pragmatics).

Questions

Maybe could formal methods be used to get contexts for words which then are used in distributional models?

How to represent quantifiers ⇒ **A** dog bites man