

Andreas van Cranenburgh¹
Mechanisms of Meaning

Distributional Semantic Models

A talk by Stefan Evert, 22 september 2010

The core of Distributional Semantic Models (DSM) is the Distributional Hypothesis by Zellig Harris. This hypothesis states roughly that:

You shall know a word by the company it keeps – Firth

In other words, the words surrounding a word say something about the meaning of a word. This is related to the idea that meaning can be derived from use. The list of words that combine with a word is called a “word sketch.”

Typically one constructs a matrix where the rows contain the target terms (words or other units) for which the meaning is to be analyzed. The columns will be features, for example different documents where the word occurs, or neighboring words with which the word occurs. The rows of this matrix will contain vectors (lists of numbers) of occurrences. These vectors can be seen as co-ordinates, so that words or points in a meaning space.

Using for example Euclidean distance measures between two points, one can find the similarity between two words. This requires normalizing the distances, so that more frequent words can be compared to less frequent words. Instead measuring distance one can also measure the angle from the origin.

Typical applications of distance measures include finding nearest neighbors (synonyms), clustering (semantic categories) and constructing semantic maps. Three famous DSMs showing these applications are Latent Semantic Analysis by Landauer & Dumais of synonymy, the word space model by Scütze for word sense disambiguation, and the Hyperspace Analogue to Language which constructed a semantic map.

There are a few important parameters for a DSM. There is the type and size of contexts (i.e., how big is the window of neighboring words to consider), the type of distance measure, and usually a form of dimensionality reduction such as Principal Components Analysis (PCA) is applied to make the data manageable and reduce noise.

¹andreas@unstable.nl

Personal view

I think Distributional Semantic Models are good in that they are quantitative and empirical. However, a severe shortcoming is that they operate only on surface forms (words). When Wittgenstein said that meaning should be understood as use (which Evert referred to in his talk), he meant not only words as used in sentences, but also words as used in situations (e.g., the famous scene of handing around slabs of concrete). I think it is only this more broad sense of usage, including daily practice, which can define meaning fully. With the current practice, DSMs can only reveal information that is encoded internal to language, not any relationship to the world.

Another problem is that current DSMs have to be optimised for each specific task independently. So for example a model trained for synonymy will not at all work well as a model for word sense disambiguation, whereas humans can do all these tasks using presumably just a single model – in other words, this detracts from their cognitive plausibility.

∞