Geometric Models of Meaning
Lecture 3

Daniel Edmiston

October 7, 2019

Similarity Metric

- We ended last time talking about distance metrics, and the most common distance metric, Euclidean distance.

Similarity Metric

- We ended last time talking about distance metrics, and the most common distance metric, Euclidean distance.

- Recall, a distance metric over a set $\mathcal{A}$ is a function $d : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$, which meets four criteria (non-negativity, symmetry, identity of indiscernibles, triangle inequality)

Similarity Metric

- We ended last time talking about distance metrics, and the most common distance metric, Euclidean distance.

- Recall, a distance metric over a set $\mathcal{A}$ is a function $d : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$, which meets four criteria (non-negativity, symmetry, identity of indiscernibles, triangle inequality)

- We can also define similarity functions, which are in some sense the inverse of distance metrics. They are less constrained than distance metrics, and of the same type $s : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$.

- Perhaps the simplest similarity metric is the negative Euclidean distance.

- Perhaps the simplest similarity metric is the negative Euclidean distance.

- More common, especially in information retrieval and related fields, is the cosine similarity.

- Perhaps the simplest similarity metric is the negative Euclidean distance.

- More common, especially in information retrieval and related fields, is the cosine similarity.

- The cosine of two non-zero vectors can be calculated as follows:

$$sim_{cos}(x, y) := \frac{x \cdot y}{\|x\| \|y\|}$$

- How do we read this? We need to define two functions:

- How do we read this? We need to define two functions:

- dot product, written as $\cdot : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, calculated as

$$\Sigma_{i=1}^{n} x_i y_i$$

- How do we read this? We need to define two functions:

- dot product, written as $\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, calculated as

$$\Sigma_{i=1}^{n} x_i y_i$$

- The geometric interpretation is the magnitude of x multiplied by the magnitude of the projection of y onto x.

- We also need to define what is called the *norm*, written $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$. This is calculated as

$$\sqrt{\sum_{i=1}^{n} x_i^2}$$

- We also need to define what is called the *norm*, written $\| \cdot \| : \mathbb{R}^n \rightarrow \mathbb{R}$. This is calculated as

$$\sqrt{\Sigma_{i=1}^{n} x_i^2}$$

- We can think of the norm intuitively as the length of the vector. Note its relation to the Euclidean distance.

- We also need to define what is called the *norm*, written $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$. This is calculated as

$$\sqrt{\Sigma_{i=1}^{n} x_i^2}$$

- We can think of the norm intuitively as the length of the vector. Note its relation to the Euclidean distance.

- Now, back to cosine similarity!

- Recall the definition:

$$sim_{cos}(x, y) := \frac{x \cdot y}{\|x\| \|y\|}$$

- Recall the definition:

$$sim_{cos}(x, y) := \frac{x \cdot y}{\|x\| \|y\|}$$

- The numerator measures similarity, as the more similar two vectors are the higher it will be, and the denominator normalizes the score, abstracting away magnitudes and leaving only "directional information."

- Recall the definition:

$$sim_{cos}(x, y) := \frac{x \cdot y}{\|x\| \|y\|}$$

- The numerator measures similarity, as the more similar two vectors are the higher it will be, and the denominator normalizes the score, abstracting away magnitudes and leaving only "directional information."

- Let's do some examples on the board to try to build some intution, then look at code.

Distributional Hypothesis

- We have seen some of the tools and power the geometric metaphor for meaning provides us

Distributional Hypothesis

- We have seen some of the tools and power the geometric metaphor for meaning provides us

- The question is now how to assign vectors to our linguistic entities

Distributional Hypothesis

- We have seen some of the tools and power the geometric metaphor for meaning provides us

- The question is now how to assign vectors to our linguistic entities

- The answer is based in the distributional hypothesis

- We will automatically extract representations from language data, providing a purely empirical approach

- We will automatically extract representations from language data, providing a purely empirical approach

- **The distributional hypothesis**: Words with similar distributional properties have similar meanings

- We will automatically extract representations from language data, providing a purely empirical approach

- **The distributional hypothesis**: Words with similar distributional properties have similar meanings

- Words with similar distributions in the data will arrive at similar geometric representations; similar geometric representations are relatively close in the semantic space, which we interpret as similar in meaning.

- Distributional hypothesis motivated by the distributional methodology of Zellig Harris.

- Distributional hypothesis motivated by the distributional methodology of Zellig Harris.

- Harris' idea was that classes of linguistic entities could be grouped automatically according to distributional behavior

- Distributional hypothesis motivated by the distributional methodology of Zellig Harris.

- Harris' idea was that classes of linguistic entities could be grouped automatically according to distributional behavior

- This applied to phonemes, morphemes, and syntactic units. Meaning (i.e. semantics) is notably absent, as Harris believed it beyond the reach of linguistic theory (with all its social manifestations, etc.)

- That said, Harris says the following:

“[A linguistic phenomenon] may be ‘due to meaning’ in one sense, but it accords with a distributional regularity.”

- That said, Harris says the following:
"[A linguistic phenomenon] may be 'due to meaning' in one sense, but it accords with a distributional regularity."

- That is to say, while some linguistic phenomena may influenced by extralinguistic factors, they have distributional correlates.

- That said, Harris says the following:
"[A linguistic phenomenon] may be 'due to meaning' in one sense, but it accords with a distributional regularity."

- That is to say, while some linguistic phenomena may influenced by extralinguistic factors, they have distributional correlates.

- We can interpret this as saying the distributional method is able to reflect meaning. A vector is certainly not what any word means, but correlations between meanings can be captured between correlations between distributions.

- A longish quote from Harris:

- A longish quote from Harris:

- "...if we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution."

- A longish quote from Harris:

- "...if we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution."

- This hypothesis has been validated by comparisons of distributions with native speakers' judgments; this forms the basis of tasks like intrinsic evaluation, which we'll discuss later.