Word Embedding

A Simple Illustration

Weimao Ke
wk@drexel.edu

Target Vectors

Imagine the word embeddings in a 2D space as follows:

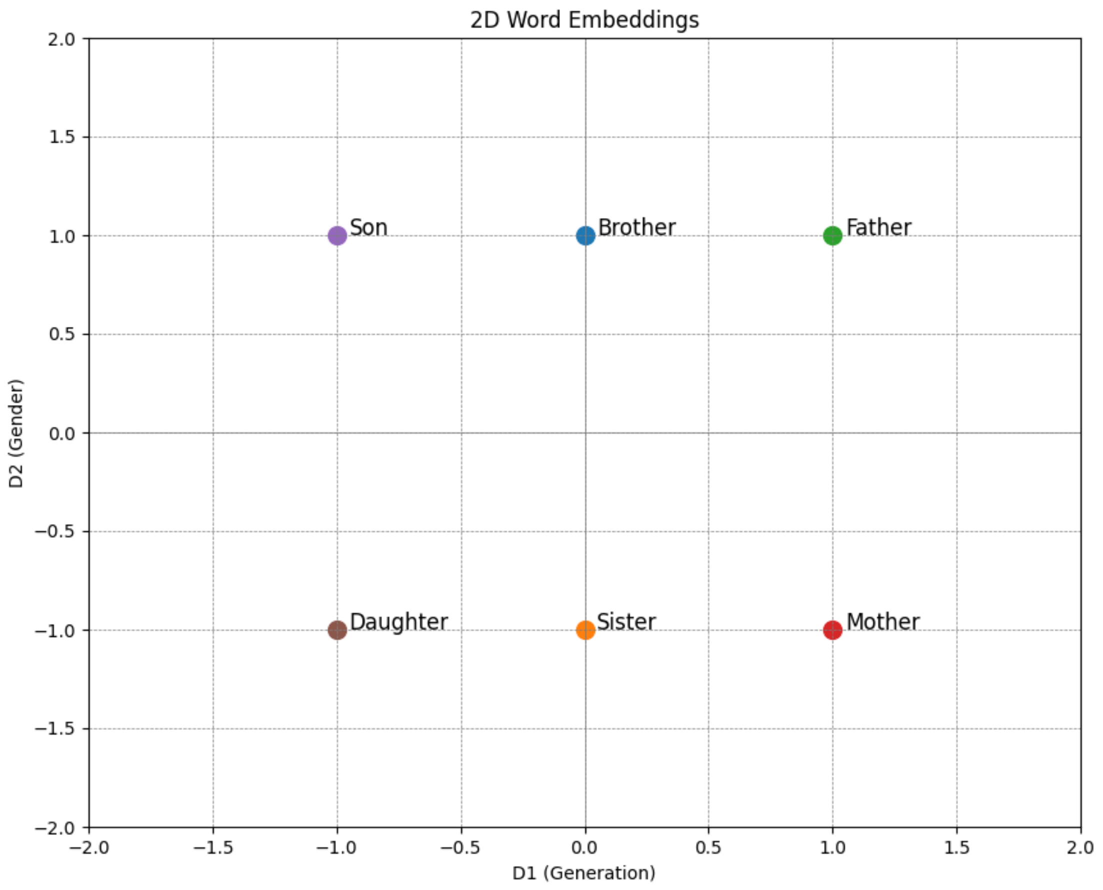
1. Dimension 1 (Generation): The further away from the origin along the x-axis, the older the generation. For simplicity, let's define 0 as the current generation (siblings), 1 as the parent's generation, and -1 as the children's generation.
2. Dimension 2 (Gender): Positive values on the y-axis represent male and negative values represent female.

Given this setup, our word embeddings could be:

```
Embedding(Brother) = [0,  1]
Embedding(Sister)  = [0, -1]
Embedding(Man)     = [0,  1]
Embedding(Woman)   = [0, -1]
Embedding(Father)  = [1,  1]
Embedding(Mother)  = [1, -1]
Embedding(Son)     = [-1, 1]
Embedding(Daughter)= [-1,-1]
```

Properties

With these embeddings:



Notes:

- The transition from Brother to Sister is essentially a flip along the y-axis.
- The arithmetic: $Brother - Man + Woman \rightarrow Sister$ translates to shifting from the male part of the current generation to the female part.
- Furthermore, the transition from Brother to Father would involve moving one unit along the x-axis.

- The arithmetic: $Son - Child$ (assuming Child maps to [-1, 0]) would translate to shifting from the child's generation to the current generation on the male side.
- It's worth noting that while this setup is quite logical and structured, real word embeddings are more intricate, with relationships not always following a direct linear pattern.

Training Process

Sample Sentences:

```
My brother is my sibling and he plays football.
My sister is my sibling and she loves dancing.
The man went to the market.
The woman went shopping.
His father works at a bank.
Her mother is a teacher.
Their son is in school.
Their daughter dances gracefully.
```

How Word2Vec learns from this data?

A. For CBOW:

1. Given the context ["My", "is", "sibling", "and", "plays", "football"], it should predict "brother".
2. Given the context ["My", "is", "sibling", "and", "loves", "dancing"], it should predict "sister".
3. ... and so on for other sentences.

B. For Skip-Gram:

1. Given the word "brother", it should predict words like "My", "is", "sibling", "and", "he", etc.
2. Given the word "sister", it should predict words like "My", "is", "sibling", "and", "she", etc.
3. ... and so on for other sentences.

Initialization

Begin with random embeddings for each word.

Training Process

In each iteration:

1. For each training sample, predict the word (for CBOW) or context (for Skip-Gram) using the current embeddings.
2. Compute the error between the prediction and the true value.
3. Adjust the embeddings slightly to reduce this error.

Training continued...

- Iterations:
  - This process is repeated many times (many epochs over the data), gradually refining the embeddings.
  - The goal is to place words in the embedding space such that words appearing in similar contexts are close to each other.

In the context of our sentences:

- "Brother" and "Man" would often appear in contexts where masculine terms or pronouns are used (like "he").
- Similarly, "Sister" and "Woman" would often appear in contexts where feminine terms or pronouns are used (like "she").
- Over time, the network adjusts the embeddings to capture these patterns.
- The spatial relationships (like `"King" - "Man" + "Woman" = "Queen"`) emerge as a by-product of the model trying to represent semantic and syntactic

2-dimensional Embedding

Hypothetical & simplified