Word2vec

Representing word meanings in form of large vectors

Sahil Kakkar

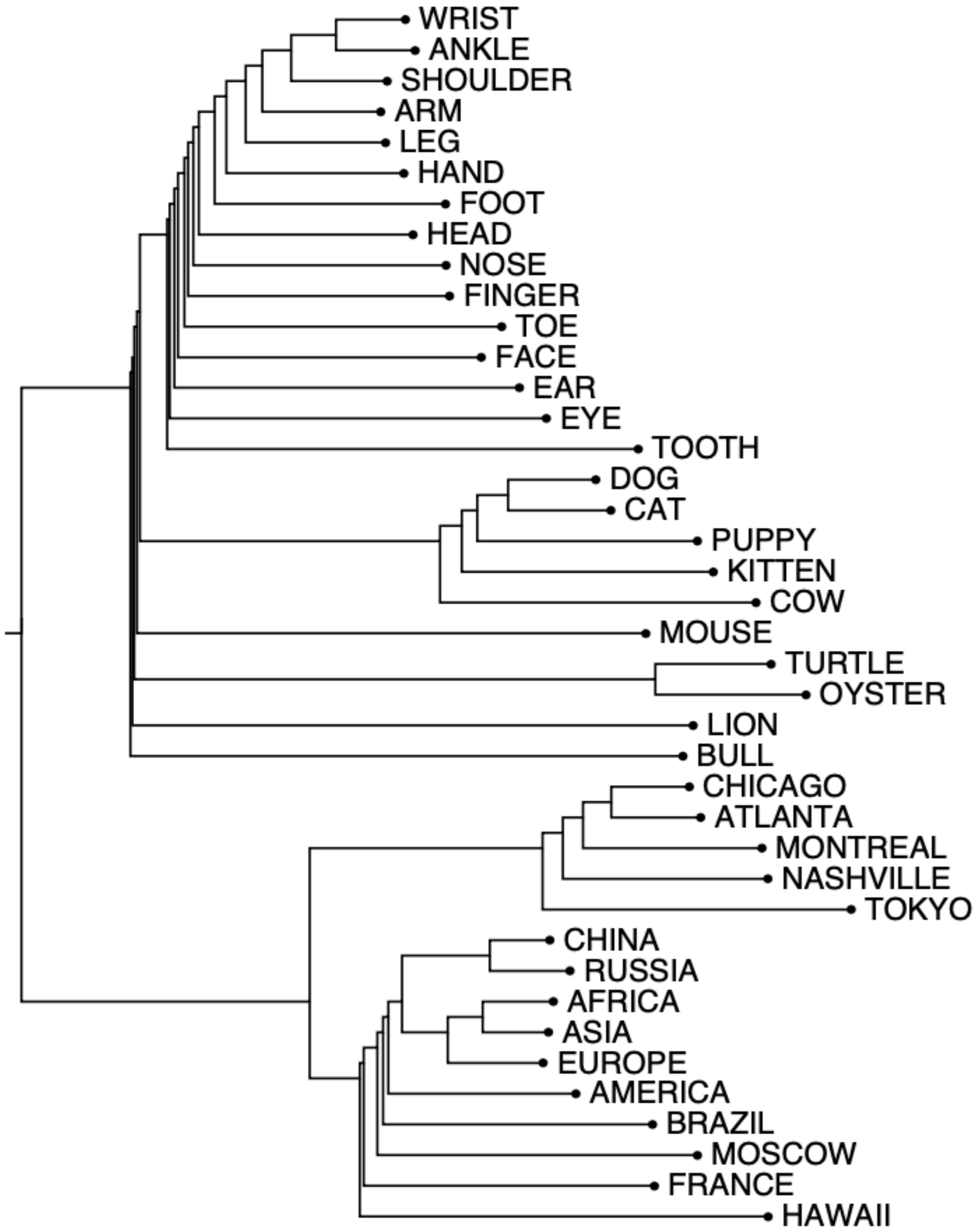
Main Idea

“You shall know a word by the company it keeps” **: J.R. Firth 1957**

Words that occurs in *similar contexts* tend to have *similar meanings*

…government debt problems turning into **banking** crisis as happened in 2009…
…saying the Europe needs unified **banking** regulation to replace the hodgepodge…
…India has just given its **banking** system a shot in the arm…

These context words represent ***banking***



Words are represented as vectors

Similar words are closer in vector space

Similarity can be computed using normalised dot product (cosine similarity) between any two vectors (words)

$$\text{similarity}(\vec{Arm}, \vec{Leg}) = \frac{\vec{Arm}.\vec{Leg}}{|\vec{Arm}||\vec{Leg}|}$$

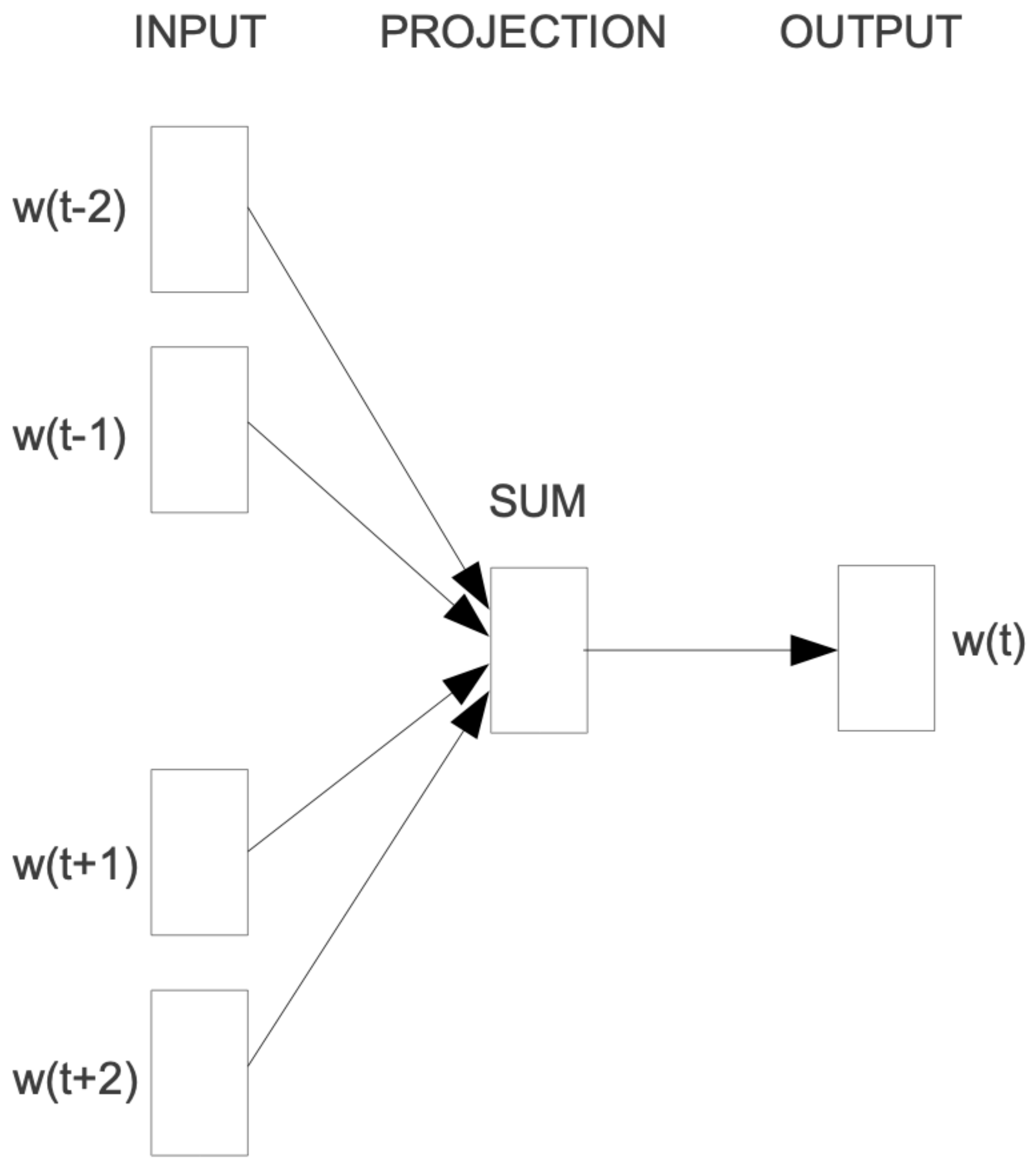
Two methods

Continuous bag of words

Use the context and predict the word in the middle

More intuitive

One prediction

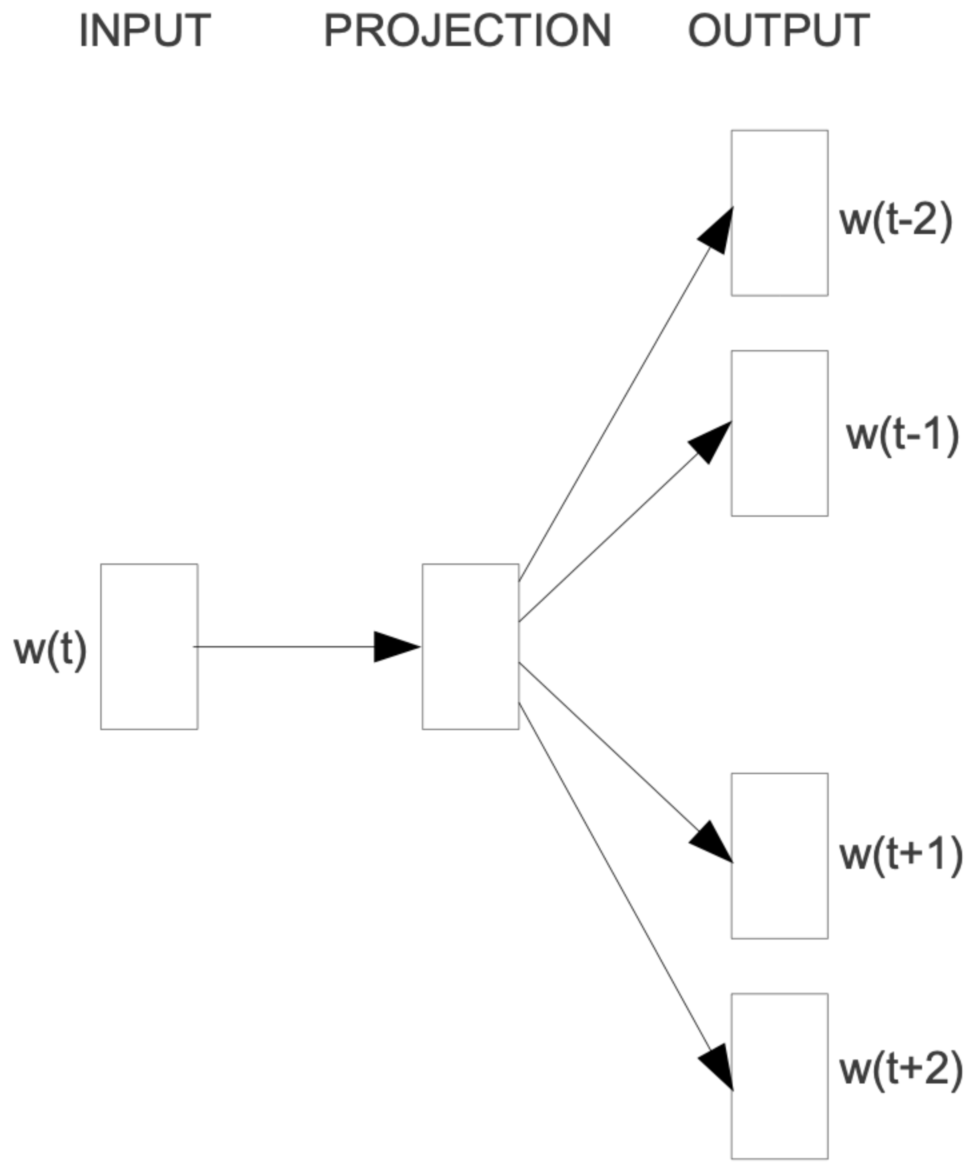


Skip N-gram

Use the word in the middle to predict the context words

Commonly used because it works better for rarer words

Multiple predictions



Methodology

Represent each word as a d *(around 300)* dimensional vector

Represent each context word as a d dimensional vector

Initialise all vectors to random weights

Arrange them into two matrices W (word) and C (context word)

Cost Function

$$\log\ p\ (\mathrm{c}|\mathrm{w};\ \theta) = \frac{\exp v_c . v_w}{\sum_{c' \in C} \exp v_{c'} . v_w}$$

Predict context word(s) from focus word

Objective Function

$$\arg\max_{\theta} \sum_{(w,c)\in D} logp(c|w) = \sum_{(w,c)\in D} \left[log \exp v_c . v_w - log \sum_{c'} \exp v_{c'} . v_w \right]$$

Negative Sampling Distribution

$$\mathrm{p}^{NS}(w) = \frac{f(w)^{\frac{3}{4}}}{\sum_{w'} f(w)^{\frac{3}{4}}}$$

Some Insights

Why two different representations of the same word (word and context)?

What about the words that have multiple meanings?

Why d=300?

GitHub Repo Link

https://github.com/Sahil1776/NLP-TeachBack-Assignment

References

- YouTube. (2021a, October 29). *Stanford CS224N: NLP with Deep Learning | Winter 2021 | lecture 1 - Intro & Word Vectors*. YouTube. https://www.youtube.com/watch?v=rmVRLeJRkl4&list=PLoROMvodv4rMFqRtEuo6SGjY4XbRIVRd4&index=1

- YouTube. (2021c, October 29). *Stanford CS224N NLP with Deep Learning | Winter 2021 | lecture 2 - neural classifiers*. YouTube. https://www.youtube.com/watch?v=gqaHkPEZAew&list=PLoROMvodv4rMFqRtEuo6SGjY4XbRIVRd4&index=2

- YouTube. (2019, February 18). *Understanding word2vec*. YouTube. https://www.youtube.com/watch?v=QyrUentbkvw

- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013, September 7). *Efficient estimation of word representations in vector space*. arXiv.org. https://arxiv.org/abs/1301.3781

- Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing*. Pearson Education.