Word embeddings & vectors

Edgar Salas Gironés,
e.girones@tudelft.nl

July 1, 2025

The Challenge of Representing Text

- Machines can‘t do much analyzing text in its raw forms:
  - Lexical similarity.
  - Count words or n-grams..
- However, machines can do way more with numbers! Some examples:
  - Arithmetic functions: Add, subtract, multiply, normalize...
  - Statistical functions: Identify variance, means, distances...
  - Probabilistic functions: KL/JS divergence, probabilities, bayesian inference....

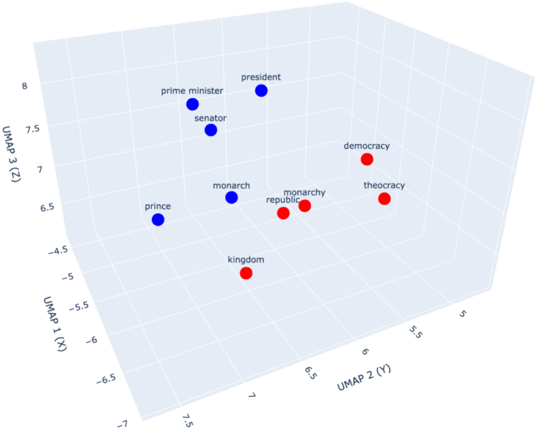
How do we represent words? Word embeddings!

What is a word embedding? It is a vector that ‘maps’ word(s) into a vector space.

$$\vec{v}_{\text{climate}} = \begin{bmatrix} 0.12 \\ -0.87 \\ \vdots \\ 0.45 \end{bmatrix}$$

- How is this value defined? Learned from large corpora.
- Words that occur in similar contexts tend to have similar meanings, “You shall know a word by the company it keeps.” (Firth, 1957)

Example: let's plot a few words



Two types of embeddings: Static vs contextual embedding

- One embedding per word!
- Why is this a problem?There is no possibility of disambiguation:
    - *party*: party leader, birthday party
    - *draft*: draft beer, football draft, policy draft
- Solution: contextual embeddings!

Contextual Embeddings (e,g, BERT)

- Transformer-based models (e.g., BERT) generate vectors *in context*.
- Word meaning varies by sentence.
- Applications in policy:
  - Argument mining.
  - Detecting changes in sentiment or position.
  - Fine-grained text classification.
- Embeddings now at sentence, paragraph, or document level.

Examples

Given this sentence: “The minister submitted a policy draft”, and these candidate sentences...

1. The pub served warm draft beer.
2. The NBA draft is taking place.
3. The bill has passed.

what would you prefer the text embedding model to do?

Go to code. . .

Extra: Dimensionality reduction!

Why dimensionality reduction? Curse of dimentionality: More dimensions, data becomes more sparse...

Solution? We reduce dimensions! Somehow transform many dimensions, (e.g. a sentence-transformers model of 768 dimensions) to a few...

- PCA.
- t-SNE, example here.
- UMAP, example here.