AGH University of Science
and Technology

Seminar in *Artificial Intelligence*
Word embedding

Marcin Trebunia, Dominik Rygiel, Konrad Adamczyk

Department of Telecommunications

27.05.2019

Agenda

1. **Introduction**
   - Necessity for encoding text
   - Simpler types of encoding
2. **Word embedding details**
3. **Applications**
4. **Problems and limitations**

Introduction

What is word embedding?

Introduction

What a **lovely** day.
What a **nice** day.

Encoding text

- Machine learning models take vectors (arrays of numbers) as input.
- ....

One hot encoding

$$
\begin{aligned}
\text{What} &= [1\ 0\ 0\ 0\ 0] \\
\text{a} &= [0\ 1\ 0\ 0\ 0] \\
\text{lovely} &= [0\ 0\ 1\ 0\ 0] \\
\text{nice} &= [0\ 0\ 0\ 1\ 0] \\
\text{day} &= [0\ 0\ 0\ 0\ 1]
\end{aligned}
$$

One hot encoding (cont.)

- Words completely independent of each other
- Inefficient approach: vector is sparse

One hot encoding (cont.)

Example:

- Dictionary of 10,000 words
- One hot encode each word
- Each vector's elements are 99.99% zeros!

Unique number encoding

$$
\begin{aligned}
\text{What} &= [1] \\
\text{a} &= [2] \\
\text{lovely} &= [3] \\
\text{nice} &= [4] \\
\text{day} &= [5]
\end{aligned}
$$

Unique number encoding (cont.)

+ Efficient - dense vector
– Encoding arbitrary - does not catch relationships between words
– Can be challenging for a model to interpret

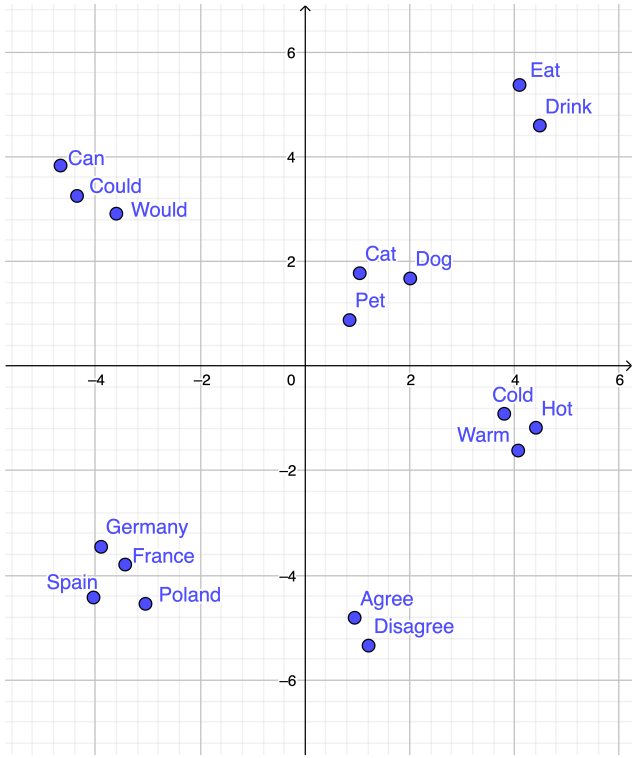
Word embedding

$$
\begin{aligned}
\text{What} &= [1.2 \quad -0.1 \quad 4.3 \quad 3.2] \\
\text{a} &= [0.4 \quad 2.5 \quad -0.9 \quad 0.5] \\
\text{lovely} &= [2.1 \quad 0.3 \quad 0.1 \quad 0.4] \\
\text{nice} &= [2.0 \quad 0.4 \quad 0.3 \quad 0.5] \\
\text{day} &= [3.0 \quad -0.6 \quad 3.5 \quad -0.8]
\end{aligned}
$$

Word embedding

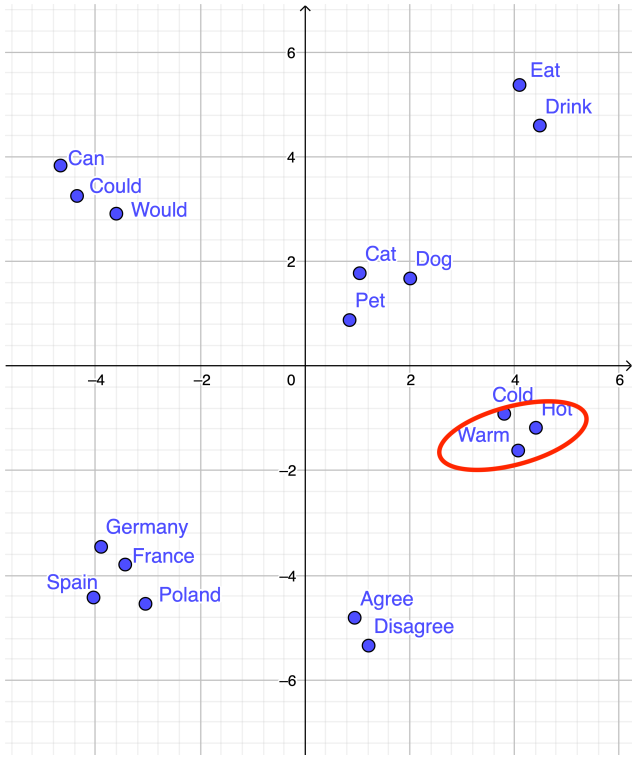
- Words with similar context occupy close spatial positions
- The cosine of the angle between words' vectors should be close to 1 (angle close to 0)

Word Embedding



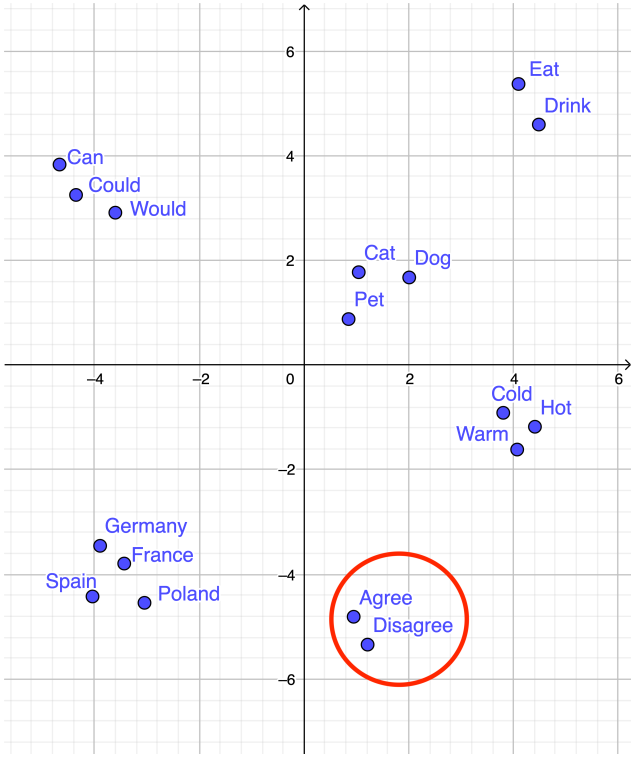
Caption of the figure

Word Embedding



Words are synonyms

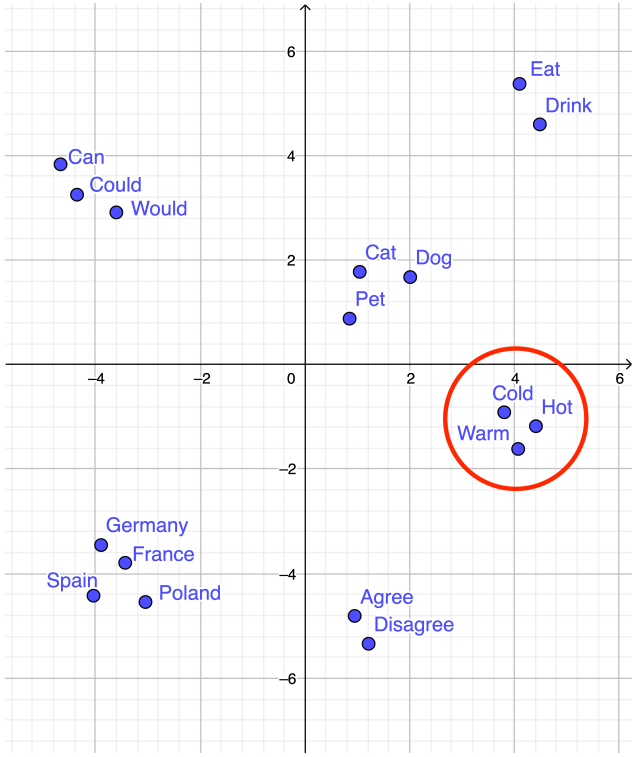
Word Embedding



Words are antonyms

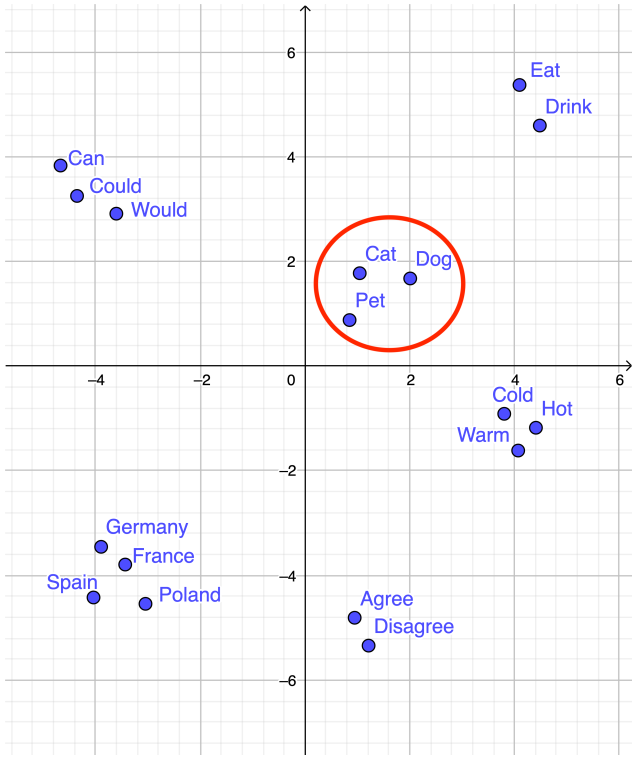
Word Embedding

Slide with a Figure from a File



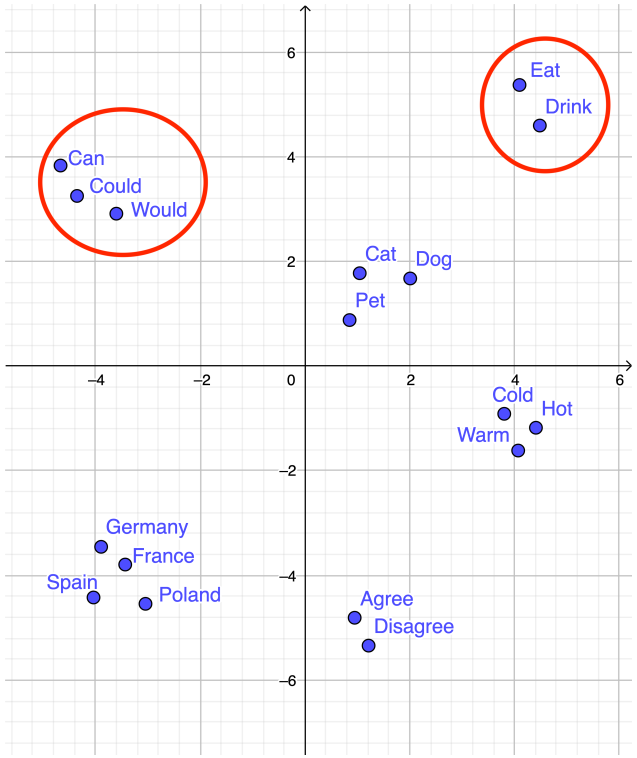
Words are value on a scale

Word Embedding



Words are hyponym - hypernym

Word Embedding



Words appear in similar context

How can we use it?

- If user search for "Dell notebook battery size" we would like to match it also with "Dell laptop battery capacity"
- If user search for "Cracow Motel" we would like to match it also with "Krakow Hotel"

Problems and limitations

- Multiple meanings of a word: solution - *Sense* embeddings
- Inability to handle unknown or out-of-vocabulary (OOV) words
- Scaling to new languages
- No shared representations at sub-word levels

AGH

Thank you for your attention!

AGH

Q & A

References

- Intro to word embeddings.
  https://www.tensorflow.org/alpha/tutorials/text/word_embeddings.
  Accessed: 2019-05-11.
- Introduction to word embedding and word2vec.
  https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060f
  fbclid=IwAR3c2RpZOmbWC84_mKFtRI6PwTD7vJRxiquKPp2Y3en3_OfDpBsWjjSinv8.
  Accessed: 2019-05-11.

References (cont.)

Word embeddings and their challenges.
http://blog.aylien.com/word-embeddings-and-their-challenges/.
Accessed: 2019-05-12.