Word Vectors or word2vec

Eng. Fatma

What are word vectors?

- A word vector is a dense representation of a word.

- Word vectors are important for semantic similarity applications like similarity calculations between words, phrases, sentences, and documents, e.g. they provide information about synonymity, semantic analogies at word level.

Word Embedding

Word embeddings are a technique where individual words are transformed into a numerical representation of the word (a vector). Where each word is mapped to one vector, this vector is then learned in a way which resembles a neural network.

What are word vectors?

- Word vectors are produced by algorithms to reflect similar words appear in similar contexts. This paradigm captures target word meaning by collecting information from surrounding words which is called distributional semantics.
- Word vectors, or word2vec are important quantity units in statistical methods to
- represent text in statistical NLP algorithms. There are several ways of text vectorization to provide words semantic representation.

Example: A Simple Word Vector

Steps to assign words vectors:
• Assign an index value to each word in vocabulary and encode this value into a sparse vector.
• Consider *tennis* as vocabulary and assign an index to each word according to vocabulary order as in Fig

Vocabulary word vector will be 0, except for word corresponding index value position as in Fig

Example: A Simple Word Vector

1. a
2. go
3. I
4. tennis
5. play
6. outside
7. hot
8. swim
9. rest

Example: A Simple Word Vector

word									
a	1	0	0	0	0	0	0	0	0
go	0	1	0	0	0	0	0	0	0
I	0	0	1	0	0	0	0	0	0
tennis	0	0	0	1	0	0	0	0	0
play	0	0	0	0	1	0	0	0	0
outside	0	0	0	0	0	1	0	0	0
hot	0	0	0	0	0	0	1	0	0
swim	0	0	0	0	0	0	0	1	0
today	0	0	0	0	0	0	0	0	1

Example: A Simple Word Vector

What are word vectors?

- I play tennis today.

What are word vectors?

- I play tennis today.

word									
I	0	0	1	0	0	0	0	0	0
play	0	0	0	0	1	0	0	0	0
tennis	0	0	0	1	0	0	0	0	0
today	0	0	0	0	0	0	0	0	1

Word2vec using spacy

```
# Import spaCy and load the en_core_web_md model
import spacy
nlp = spacy.load("en_core_web_md")

# Create a sample utterance (utt1)
utt1 = nlp("I ate a banana.")
```

```
import en_core_web_md
nlp = en_core_web_md.load()
```

Use the following script to show Word Vector for *banana*:

```
utt1[3].vector
```