One Hot Encoding: Limitations

Limitations

- Consider the sentences below:

  “ the bus is running late”
  “ the train is running late”

Limitations

- Consider the sentences below:

  “ the bus is running late”
  “ the train is running late”

- One Hot Vectors (Vocab size = 21)

  bus - [0,0,0,0,0,0,0,0,0,0,0,0,0,0,**1**,0,0,0,0,0,0]

  train - [0,0,0,0,0,0,**1**,0,0,0,0,0,0,0,0,0,0,0,0,0]

Limitations

- Consider the sentences below:

  “ the bus is running late” → the ___ is running late
  “ the train is running late” → the ___ is running late

  } Context is identical

- One Hot Vectors (Vocab size = 21)

  bus - [0,0,0,0,0,0,0,0,0,0,0,0,0,0,**1**,0,0,0,0,0,0]

  train - [0,0,0,0,0,0,**1**,0,0,0,0,0,0,0,0,0,0,0,0,0,0]

Word Embeddings

- Consider the sentences below:

“ the bus is running late” → the ___ is running late
“ the train is running late” → the ___ is running late

} Context is identical

- Word Embeddings
  - Vector for a word *(does not depend on vocabulary size)*
  - Captures the context around the word

Word Embeddings

bus - [0,0,0,0,0,0,0,0,0,0,0,0,0,0,**1**,0,0,0,0,0,0] One Hot Encoded Vector

-0.01	0.02	1.1	-0.3	0.11	-0.02

Word Embeddings

train - [0,0,0,0,0,0,**1**,0,0,0,0,0,0,0,0,0,0,0,0,0,0] One Hot Encoded Vector

-0.02	0.02	0.9	-0.3	0.13	-0.01

Word Embeddings

Word Embeddings

Bus	-0.01	0.02	1.1	-0.3	0.11	-0.02
Train	-0.02	0.02	0.9	-0.3	0.13	-0.01

Similar Vectors

Thank You