

Practical Guide to Sentence Transformers

Sadamori Kojaku

2021/09/27

- **What is sentence BERT?**
... 5 mins
- **How to use sentence BERT**
... 30 mins
- **How to train a new model with your data** ... 15 mins

References

- **Paper**
Reimers, Nils, and Iryna Gurevych. 2019. "Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks." EMNLP. *arXiv [cs.CL]*. arXiv. <http://arxiv.org/abs/1908.10084>.
- **Library**
<https://www.sbert.net/>
- **Video**
<https://www.youtube.com/watch?v=Ey81KfQ3PQU>

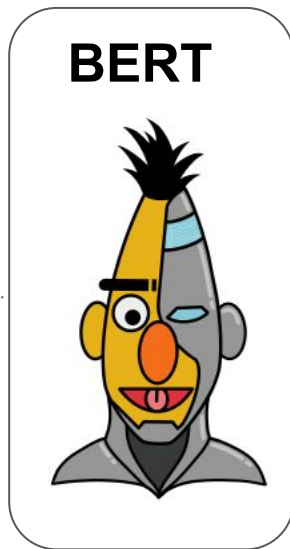
What is ~~sentence~~ BERT? (1)

- **Model for encoding text to numerical vectors**
- Trained on a **large amount of textual data**
- Able to **understand words in context**
- Achieved **the state-of-the-art performance** on various tasks

Hi, I am BERT, I'm trained on more than ten thousand of English books and Wikipedia



"I arrived at the *bank*
after crossing *the street*"



I [u_I]
arrived [u_{arrived}]
at [u_{at}]
the [u_{the}]
bank [u_{bank}]
....

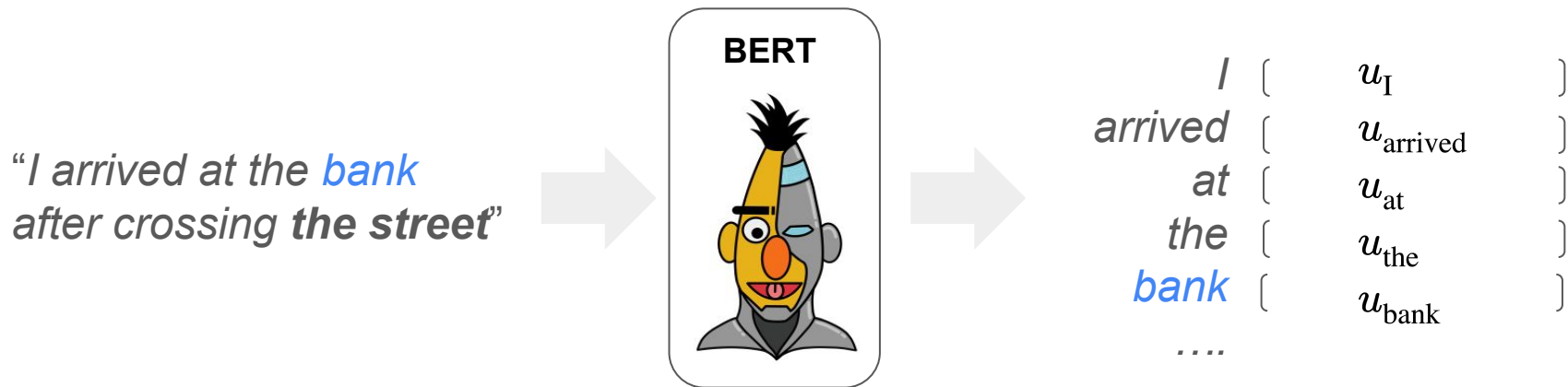
"I arrived at the *bank*
after crossing *the river*"



I [v_I]
arrived [v_{arrived}]
at [v_{at}]
the [v_{the}]
bank [v_{bank}]
....

$$u_{\text{bank}} \neq v_{\text{bank}}$$

Embedding is *context dependent!*



“I arrived at the *bank* after crossing *the street*”

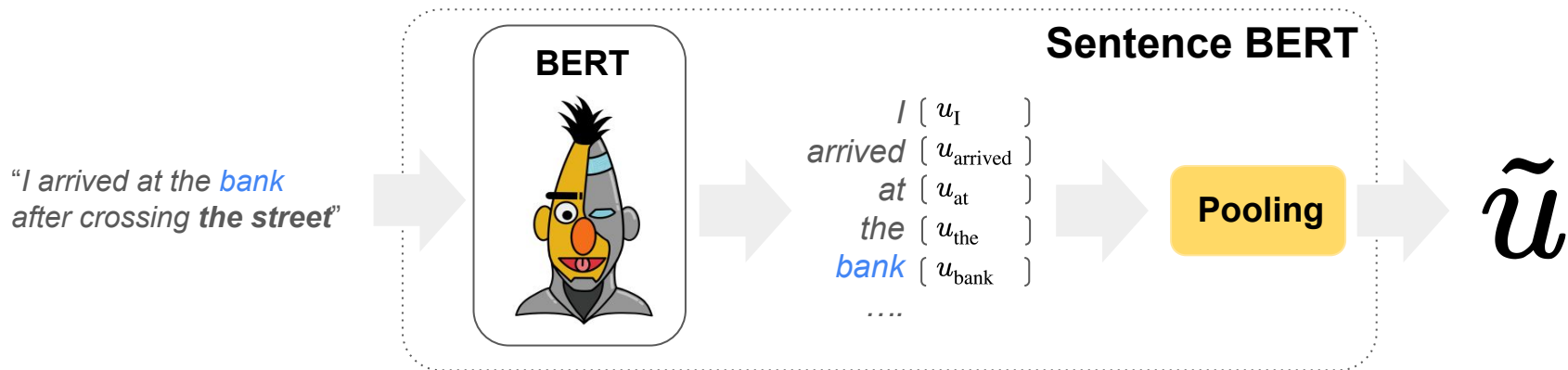
→ ?

u_I arrived at the bank after crossing the street

A number of heuristics to aggregate word embeddings,
e.g., averaging, and the [cls]-token

Sentence BERT aggregates word embeddings
in a way **to capture the semantics of the given sentence!**

Architecture

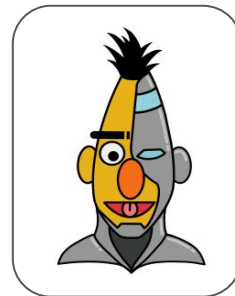


Fine-tuning

Sentence 1	Sentence 2	1 = similar 0 = dissimilar
x	x'	y
“A man is eating a piece of bread”	“A man is eating food”	1
“A woman is playing a violin”	“A man is carrying a baby”	0
\vdots	\vdots	\vdots

$$\cos(\tilde{u}_{x'} \tilde{u}_{x'}) \simeq y$$

Sentence BERT



Hands on Tutorial

https://github.com/skojaku/Practical-Guide-to-Sentence-Transformers/blob/main/notebook/Practical_Guide_to_Sentence_Transformers.ipynb



The **Internet** (or **internet**) is the system (a)___ interconnected com(b)___ networks that uses the Internet protocol suite (TCP/IP) to communicate between networks and devices. The origins of the Internet date ba(c)___ to the dev(d)___ of packet switching and research commissioned by the Un(e)___ States Department of Defense in the 1960s to enable time-(f)___ of computers.^[1] The Internet car(g)___ a vast range of information resources and services, such as the inter-linked hyper-(h)___ documents and applications of the World (i)___ Web (WWW), electronic mail, telephony, and file sharing.

Source: <https://en.wikipedia.org/wiki/Internet>

(d) devel**opment**

devel**opment** ... 70%

devel**op**ping ... 10%

dev**il** ... 5%

