Deep Learning for Natural Language Processing

Transfer learning in NLP: static and contextualized representations

UNIVERSITY OF GOTHENBURG

CHALMERS

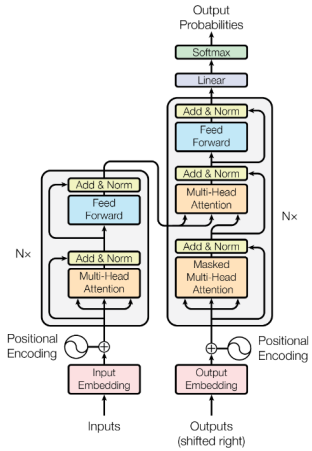
WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

Richard Johansson

richard.johansson@gu.se

plan for this lecture block

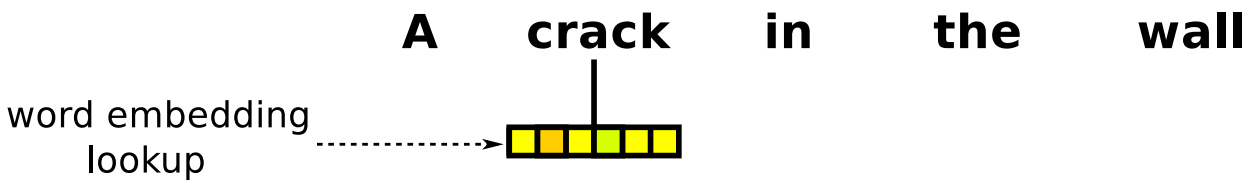
- we conclude the part of the course dedicated to representation learning
- key idea: **transfer learning** for **contextualized representations**
  - ELMo, ULMFiT, BERT, ...



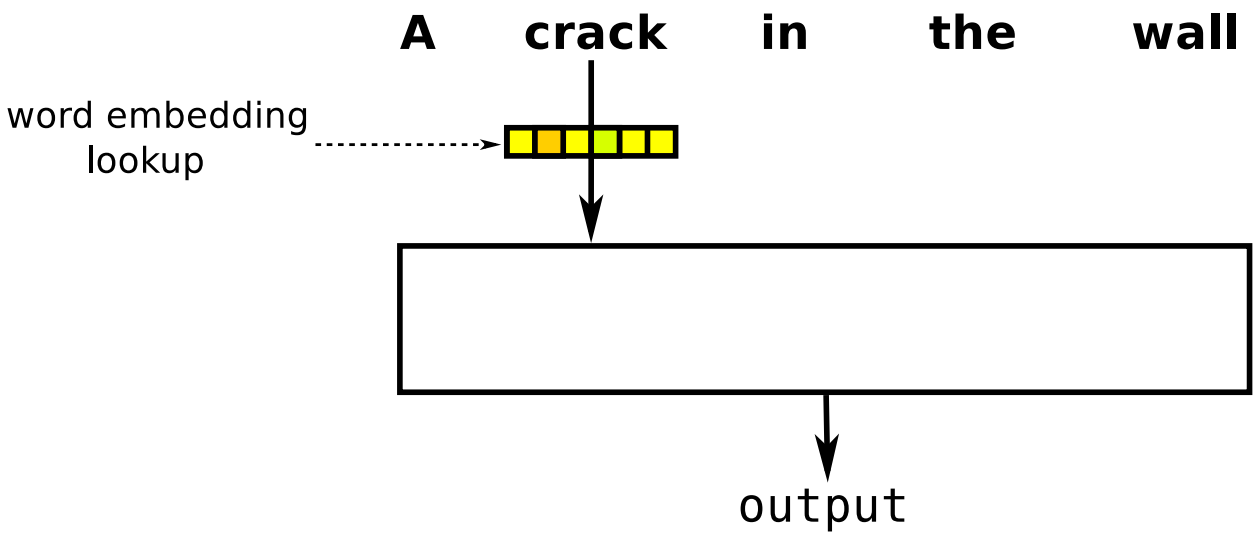
recap: representations of words

A crack in the wall

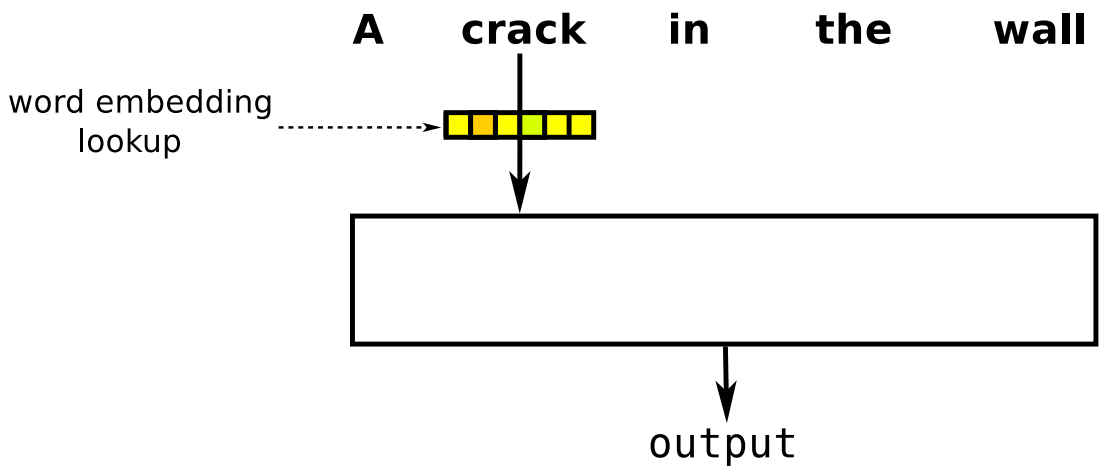
recap: representations of words



recap: representations of words

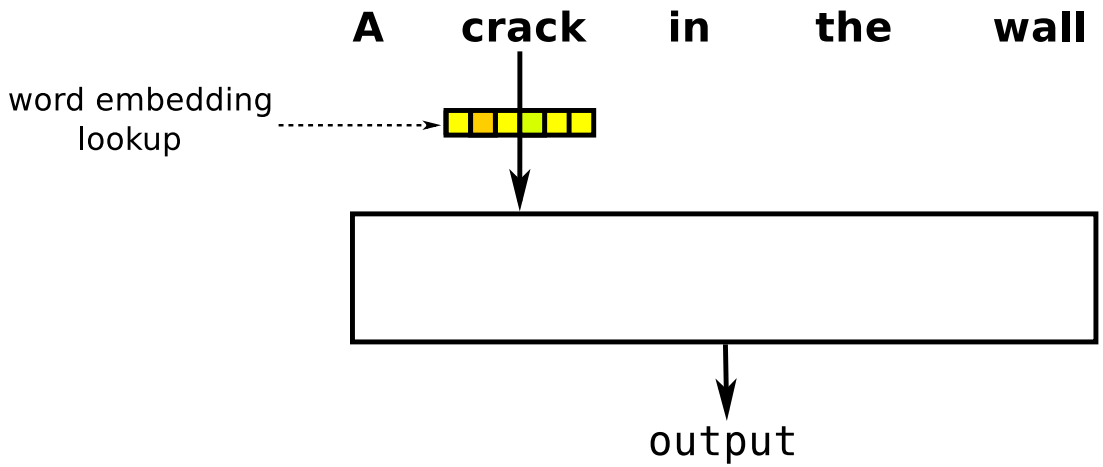


transfer learning for word representations



- in many applications, it is useful to **pre-train** the word embeddings using raw text
- this is an example of a **transfer learning** approach
  - representations are **reused** between different tasks

limitations of static word representations



- the word ***crack*** always has the **same** embedding:

  *a **crack** in the wall*
  *he was high on **crack***
  *a hard problem to **crack***

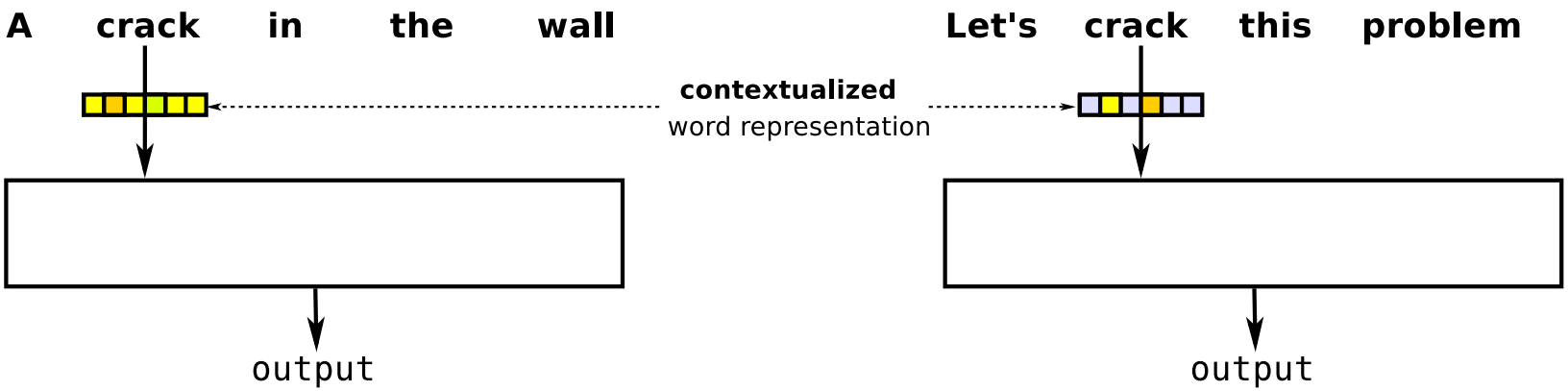
- the pre-trained representation is **static**

contextualized word representations

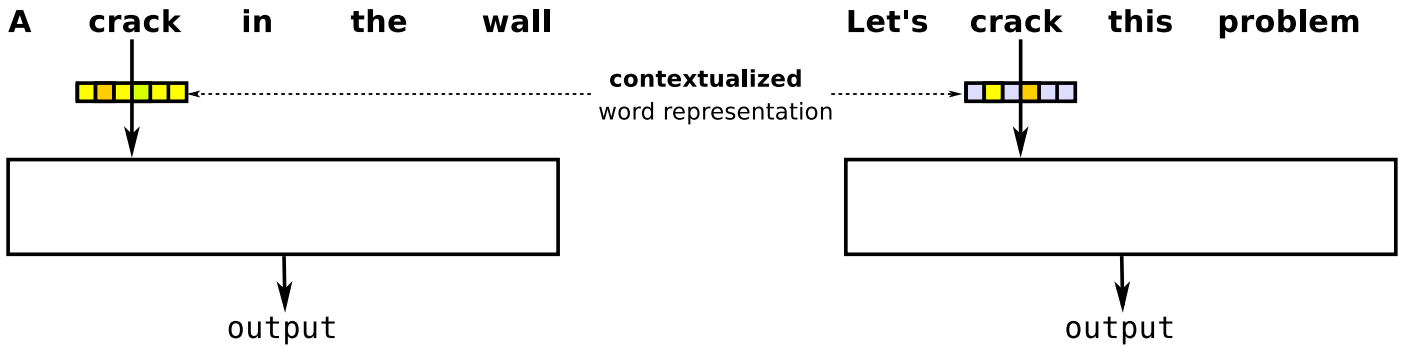
A crack in the wall **Let's crack this problem**

contextualized word representations

contextualized word representations



pre-training contextualized word representations



- we already have the building blocks to make contextualized representations
  - embeddings, CNNs, LSTMs, attention, ...
- the key difference is that we will use a transfer learning approach: the representation will be **pre-trained**

key challenges for transfer learning

- learning **generally useful** representations
  - so we need fairly general training tasks
- finding **training data**
  - ideally, an unlimited supply!

key challenges for transfer learning

- learning **generally useful** representations
  - so we need fairly general training tasks
- finding **training data**
  - ideally, an unlimited supply!

- we will consider **language modeling** as a pre-training task

I was sad because my football team had ___

the remainder of this lecture block

- transfer learning based on language models
- the Transformer model
- pre-trained Transformer models