

#### Neural networks for XMTC

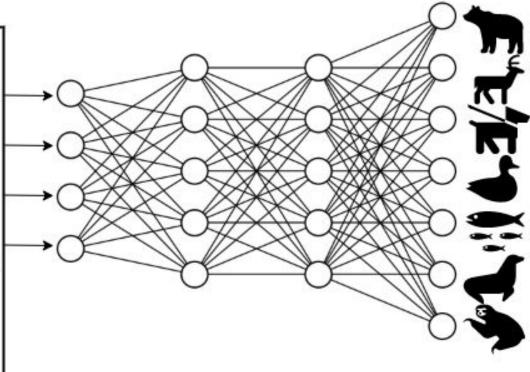
Markus Koskela



#### csc

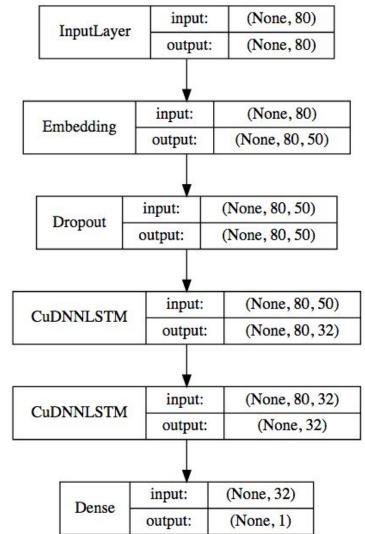
#### Neural networks for text

IT was seven o'clock of a very warm evening in the Seeonee hills when Father Wolf woke up from his day's rest, scratched himself, yawned, and spread out his paws one after the other to get rid of the sleepy feeling in the tips. Mother Wolf lay with her big gray nose dropped across her four tumbling, squealing cubs, and the moon shone into the mouth of the cave where they all lived. "Augrh!" said Father Wolf, "it is time to hunt again"; and he was going to spring downhill when a little shadow with a bushy tail crossed the



# Neural networks for text

- first layer is usually an embedding
- then there are three main approaches (that can also be combined):
  - convolutional layers
  - recurrent layers
  - attention & dense layers
- last layers are often dense



# Word embeddings

- dense vector representations
  - low-dimensional
  - learned based on context of words
- semantics
  - similar words have similar vectors
  - context-free or contextual
- suitable as input for neural networks

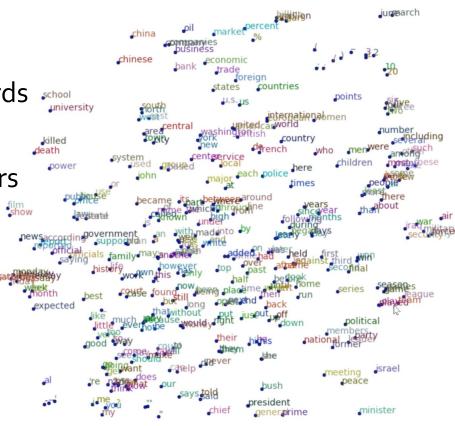
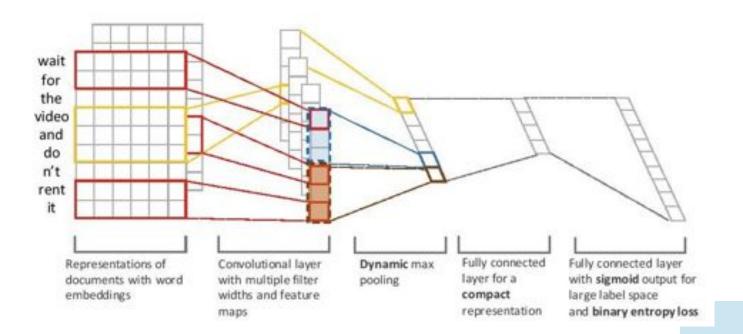


Image from: <u>http://wiki.fast.ai/index.php/Lesson\_5\_Notes</u>

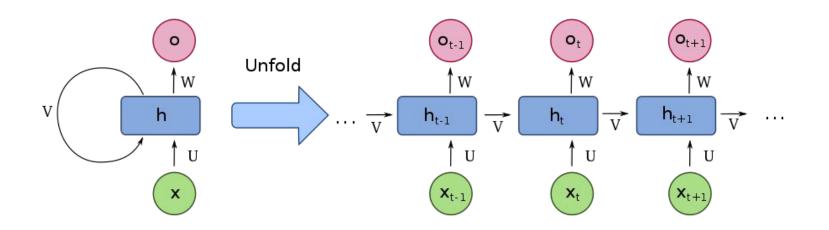
# Convolutional neural networks

- Neural networks with convolutional kernels of fixed size
- XML-CNN, Bow-CNN



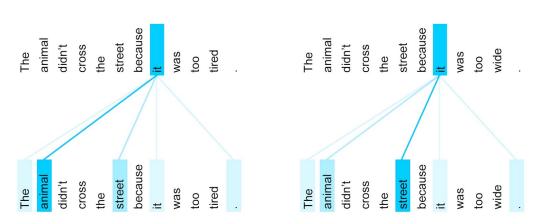
#### Recurrent neural networks

- Neural networks with memory and recurrent connections, i.e. loops
- AttentionXML



## Attention & dense layers

- Transformer models consist of layers that transform a sequence to another sequence
- Self-attention: relating different positions of a sequence in forming its representation



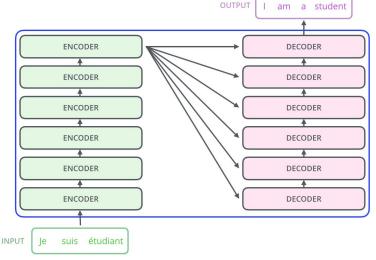
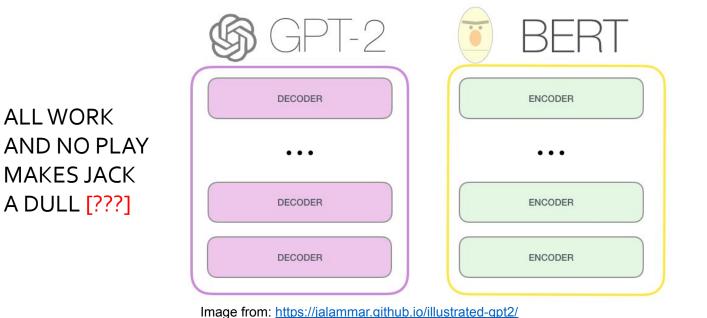


Image from: https://jalammar.github.io/illustrated-gpt2/

Image from: <a href="https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html">https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html</a>

# Models based on Transformer



ALL WORK AND NO [???] MAKES JACK A DULL BOY

CSC

- BERT-like: masked LM: predict the masked words
- GPT-like: left-to-right LM: *predict the next word*

## Models based on Transformer

#### Deep learning resources for Finnish: FinBERT

BERT for Finnish, custom Finnish vocabulary, trained from scratch for 1M steps

http://github.com/TurkuNLP/FinBERT

#### NLPL infrastructure & resources:

- Finnish Internet Parsebank: 4B + words of Finnish Internet
  - +News, +Finnish Online discussion fora
- CSC It Center for Science: Puhti supercomputer
  - CSC Grand Challenge project DeepFin







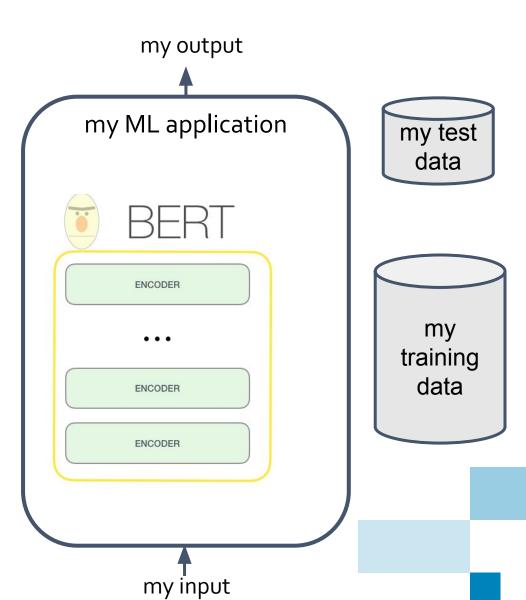
NLPL

CSC

#### From: https://twitter.com/NoDaLiDa/status/1178660680477630464

# Fine-tuning

- Pre-trained general model can be fine-tuned to a specific task
- Some amount of further training with task-specific data is needed
- Lots of different kinds of applications exist



# X-Transformer

• First "working" application of Transformers to XMTC

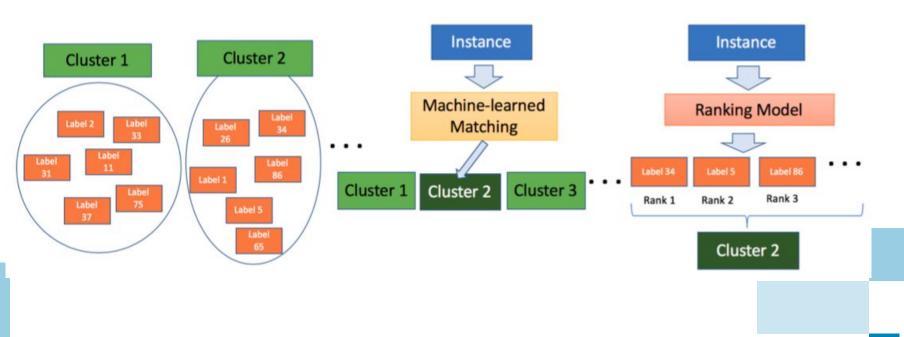
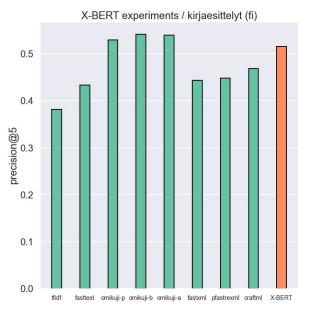
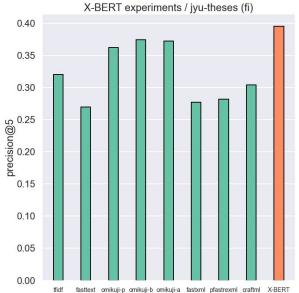
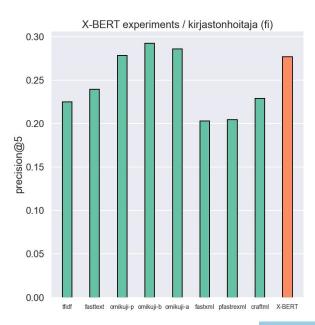


Image from: https://arxiv.org/abs/1905.02331

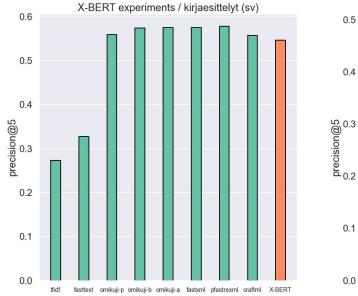
#### Some results: Finnish

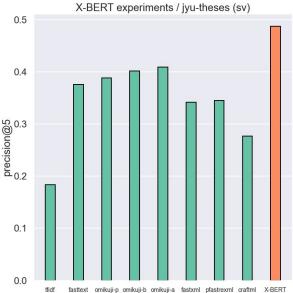


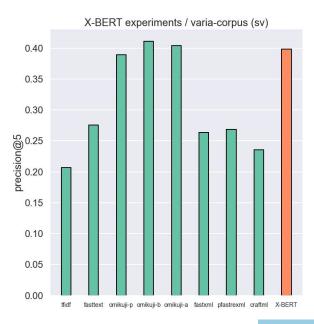




#### Some results: Swedish







# Some results: English

