Text Mining and Language Analytics

Lecture 8

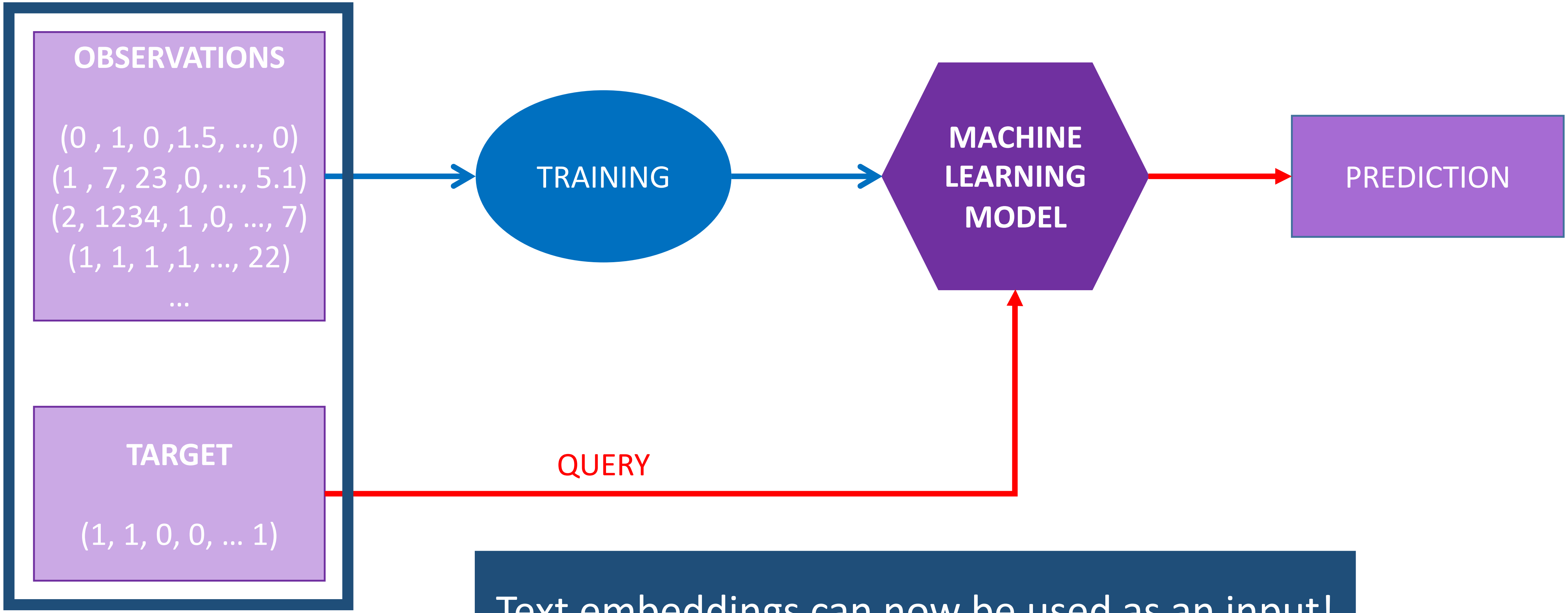
Recurrent Neural Networks & Long Short-term Memory

Dr Stamos Katsigiannis

2023-24

Durham University

Machine learning paradigm (again)



Text embeddings can now be used as an input!

Machine learning for NLP

- Sentiment analysis
- Spam detection
- Text categorisation
- Speech recognition
- Social media analysis
- Product review analysis
- And many others…

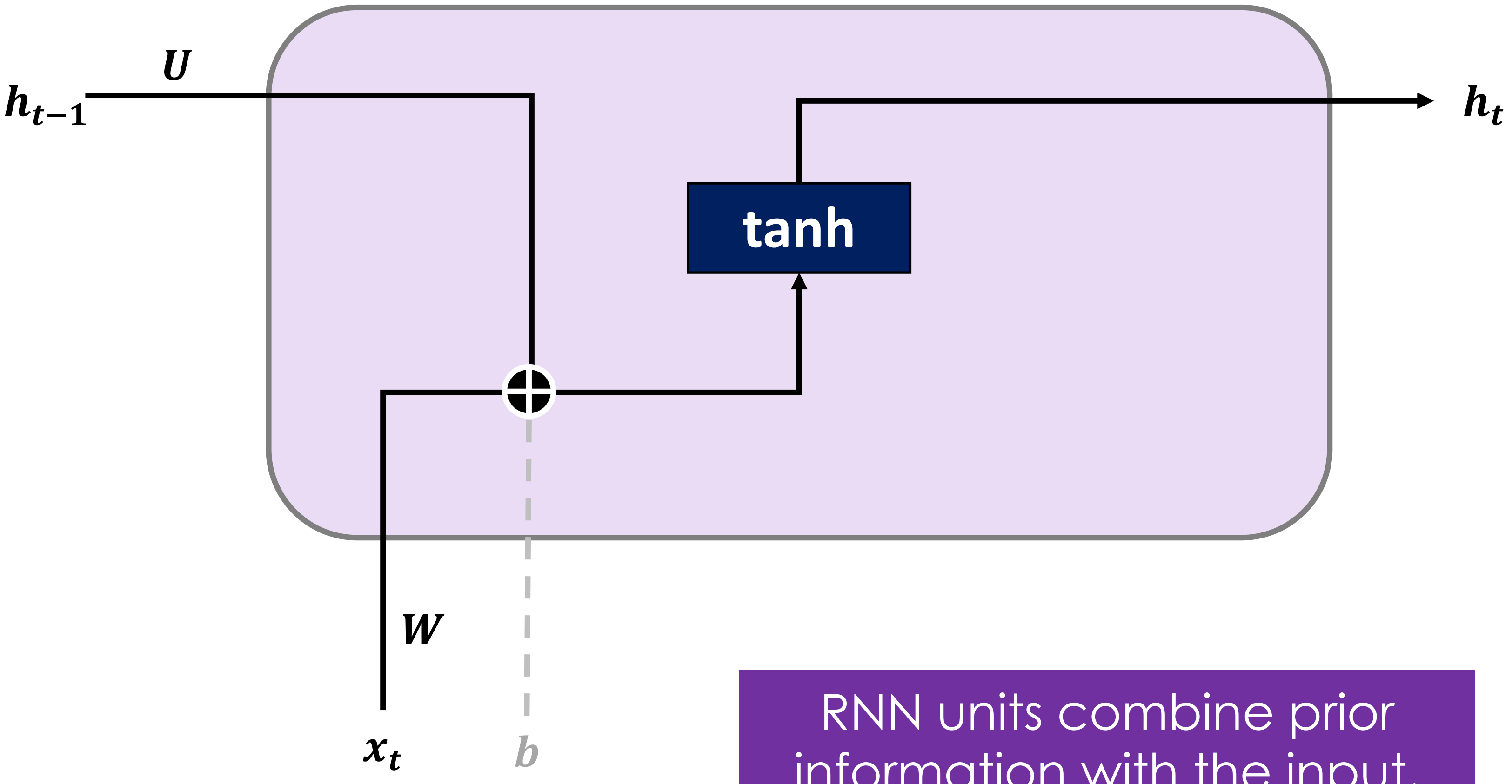
Durham University

Machine learning approaches for NLP

- Naïve Bayes
  - Simple approach
  - No embeddings needed
  - Works surprisingly well for some applications
- Convolutional Neural Networks
  - More advanced feature extraction
  - Can detect complex patterns within the data
  - Whole document must be used as input
  - Modelling of relation between words relies on embedding used
- **Words within a document have complex relations between each other**
- **More advanced models needed!**

Durham University

Recurrent Neural Network (RNN) unit



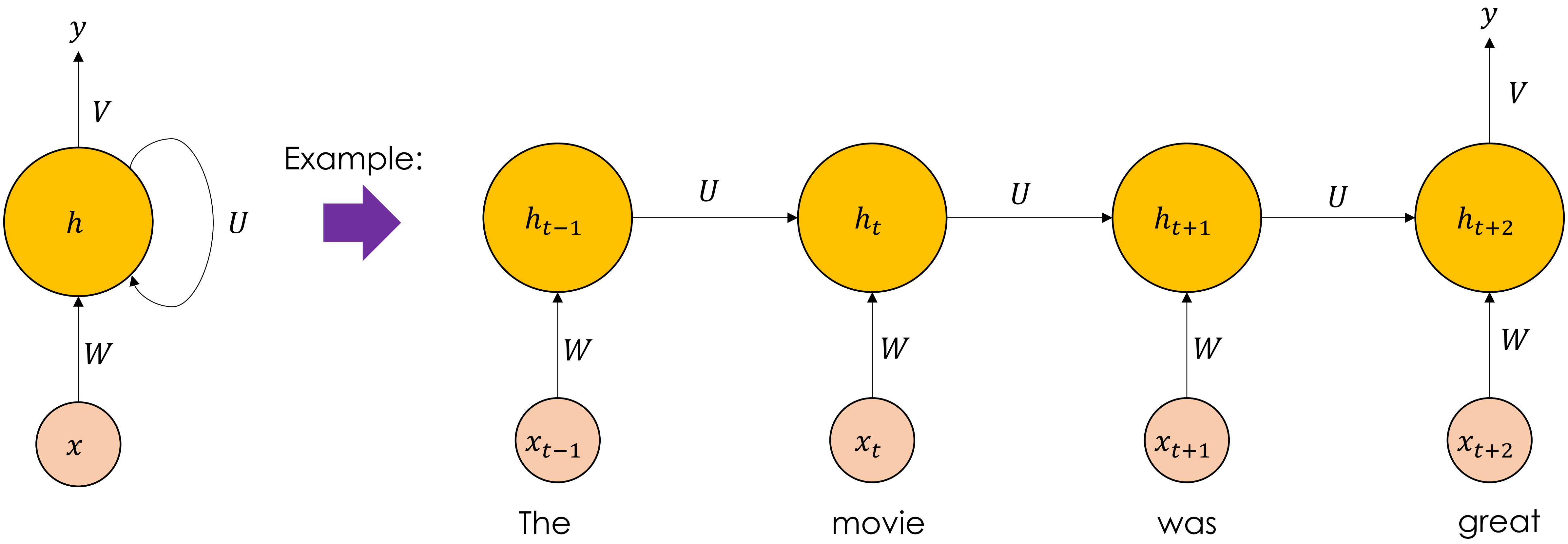
x_t	**Input vector**
h_{t-1}	**Previous state**
h_t	**New state vector**
U, W	**Weights**
b	**Bias vector (optional)**
$\oplus$	**Element-wise addition**

RNN units combine prior information with the input, having a form of *memory*

$$h_t = f(Wx_t + Uh_{t-1} + b)$$

Durham University

RNNs for NLP: Many to One

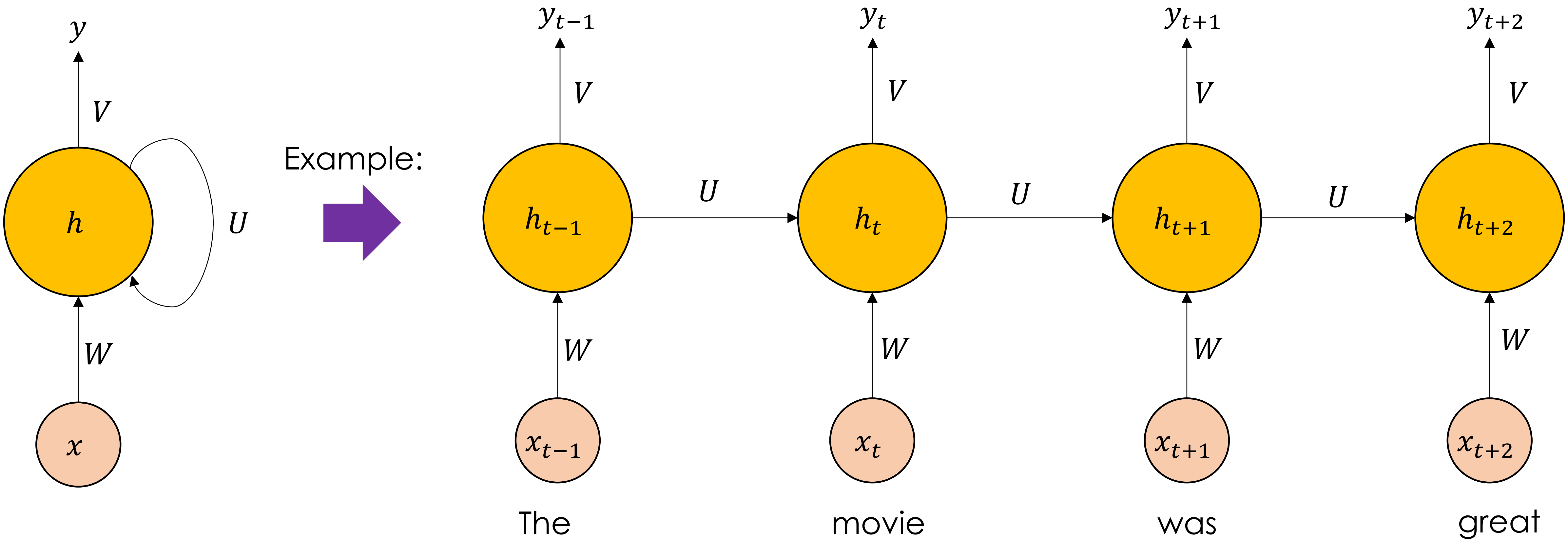


Backpropagation used for training
Softmax used for classification

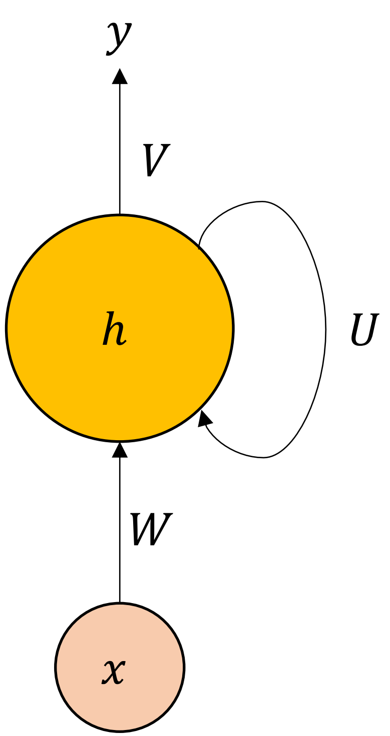
$$\boldsymbol{y_t = softmax(Vh_t)}$$

Durham University

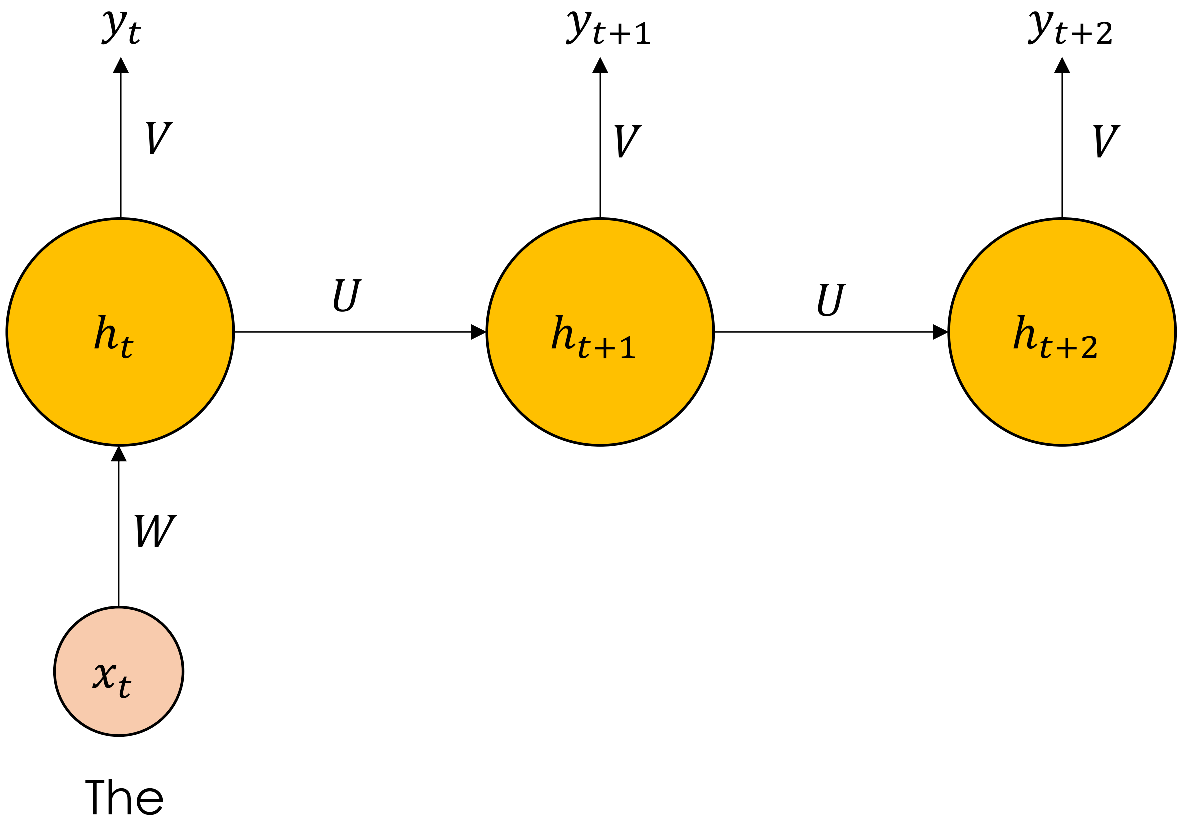
RNNs for NLP: Many to Many



RNNs for NLP: One to Many



Example:



Durham University

RNN architectures

- Many to One
  - **Multiple inputs – One output**
  - e.g. Classification
- Many to Many
  - **Multiple inputs – Multiple outputs**
  - e.g. Machine translation
- One to Many
  - **One input – Multiple outputs**
  - e.g. Data generation from a single sample

Durham University

RNN training

- Backpropagation Through Time (BPTT)
  - Present a sequence of time steps of input and output pairs to the network
  - Unroll the network then calculate and accumulate errors across each time step
  - Roll-up the network and update weights
  - Repeat

- **BPTT can be very computationally expensive!**

- Truncated Backpropagation Through Time (TBPTT)
  - Present a sequence of k_1 time steps of input and output pairs to the network
  - Unroll the network then calculate and accumulate errors across k_2 time steps.
  - Roll-up the network and update weights
  - Repeat

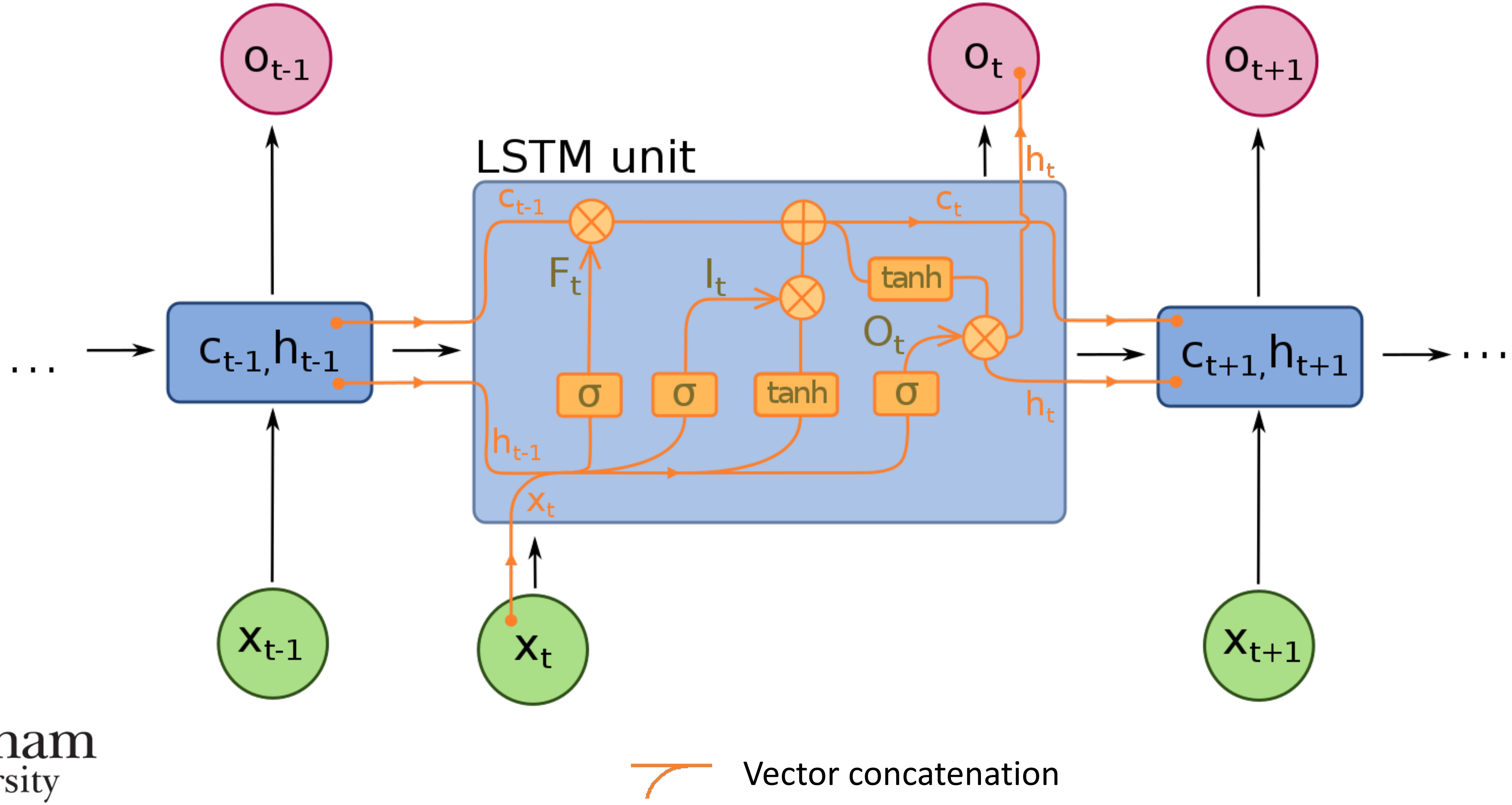
Durham University

RNNs

- **Advantages**
  - Can model sequences of data, e.g. time series, natural language
  - Flexible to work with sentences of varying size
  - Model size not increasing with size of input
  - Each sample can be assumed to be dependent of previous ones
  - Sharing features learned across different positions of the text
  - Weights are shared across time

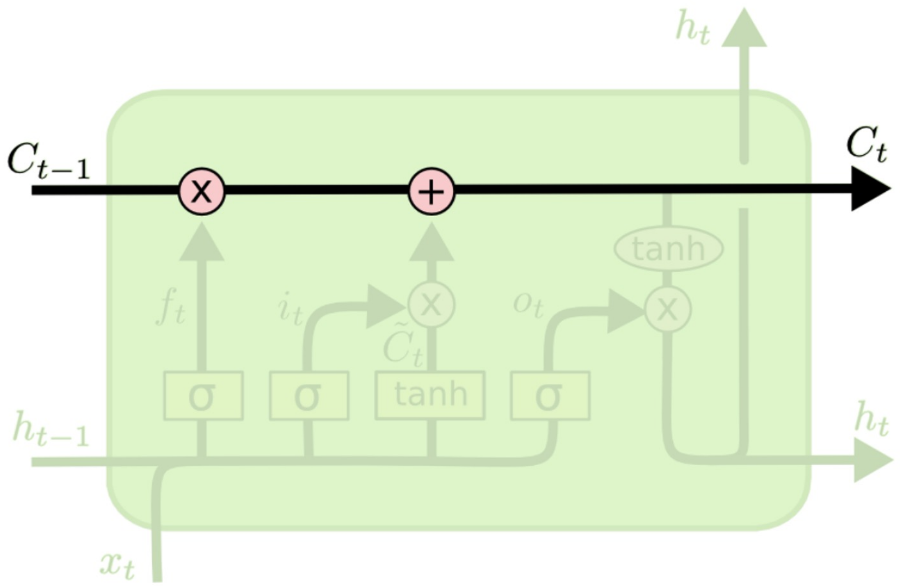
- **Limitations**
  - Computation is slow
  - Capture dependencies in only one direction of the text (past)
  - Not very good in capturing long term dependencies
  - Vanishing gradients issue during training: Gradients become too small to contribute to training for large architectures

Long Short-term Memory (LSTM)



LSTM: Cell state

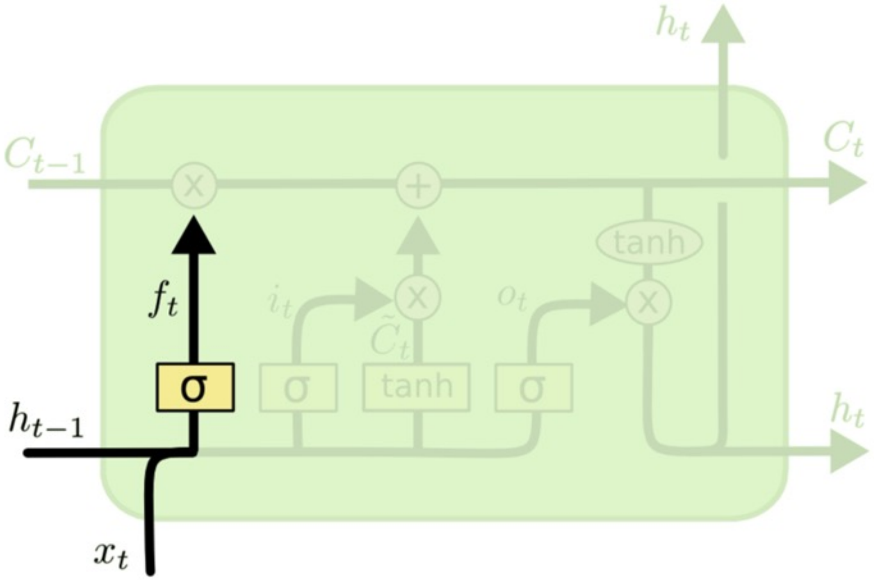
Information flows easily across it



The cell state is effectively a **memory** that propagates across units

LSTM: The Forget gate

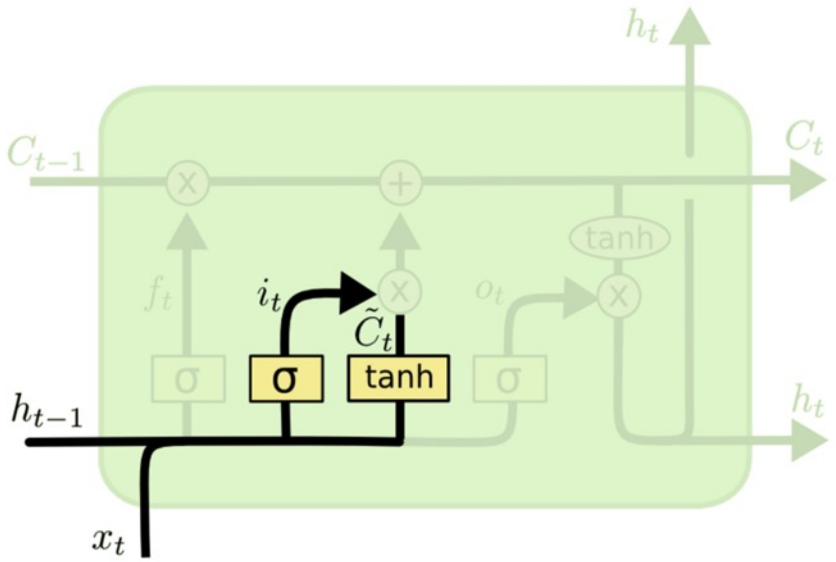
The first step is to decide what information will be thrown away from the cell state



$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

LSTM: The Input gate

The next step is to decide what new information will be stored in the cell state



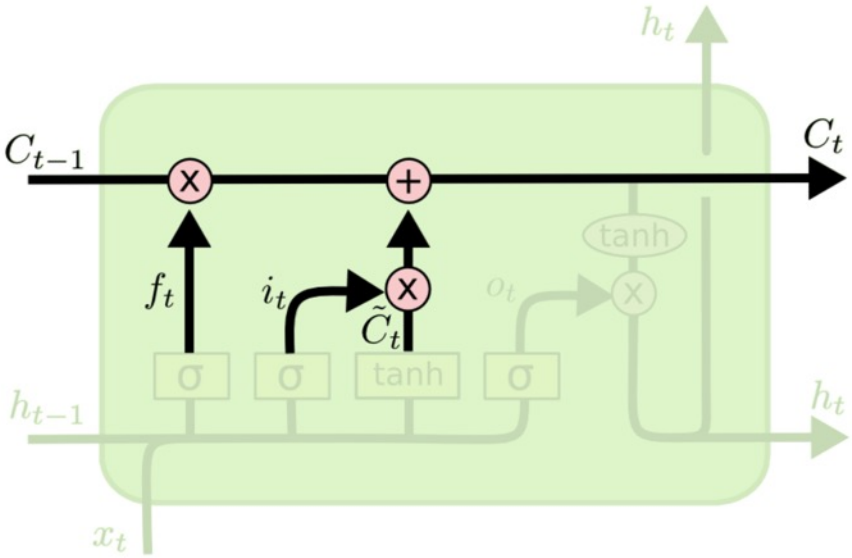
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Durham University

LSTM: Cell state update

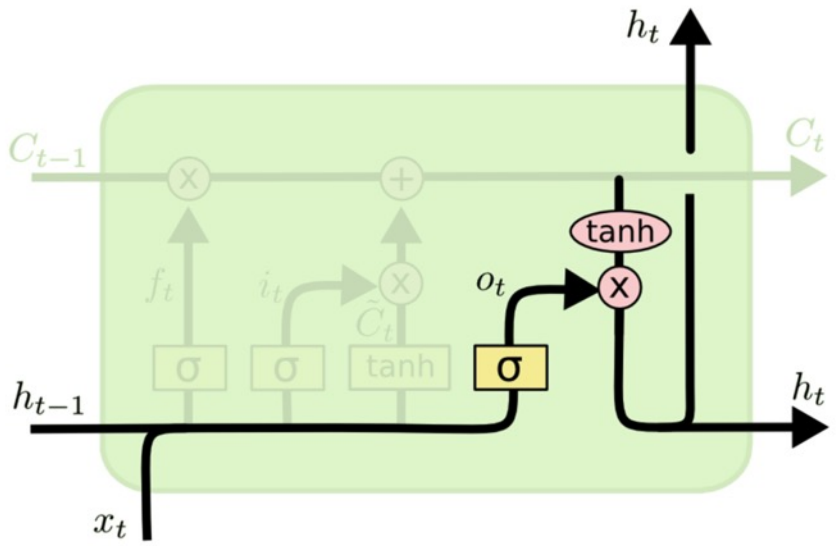
Next cell state is computed from the Forget and Input gates, along with the candidate for the next cell state, and the previous cell state



$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

LSTM: The Output gate

The output will be a filtered version of the cell state



$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \cdot \tanh(C_t)$$

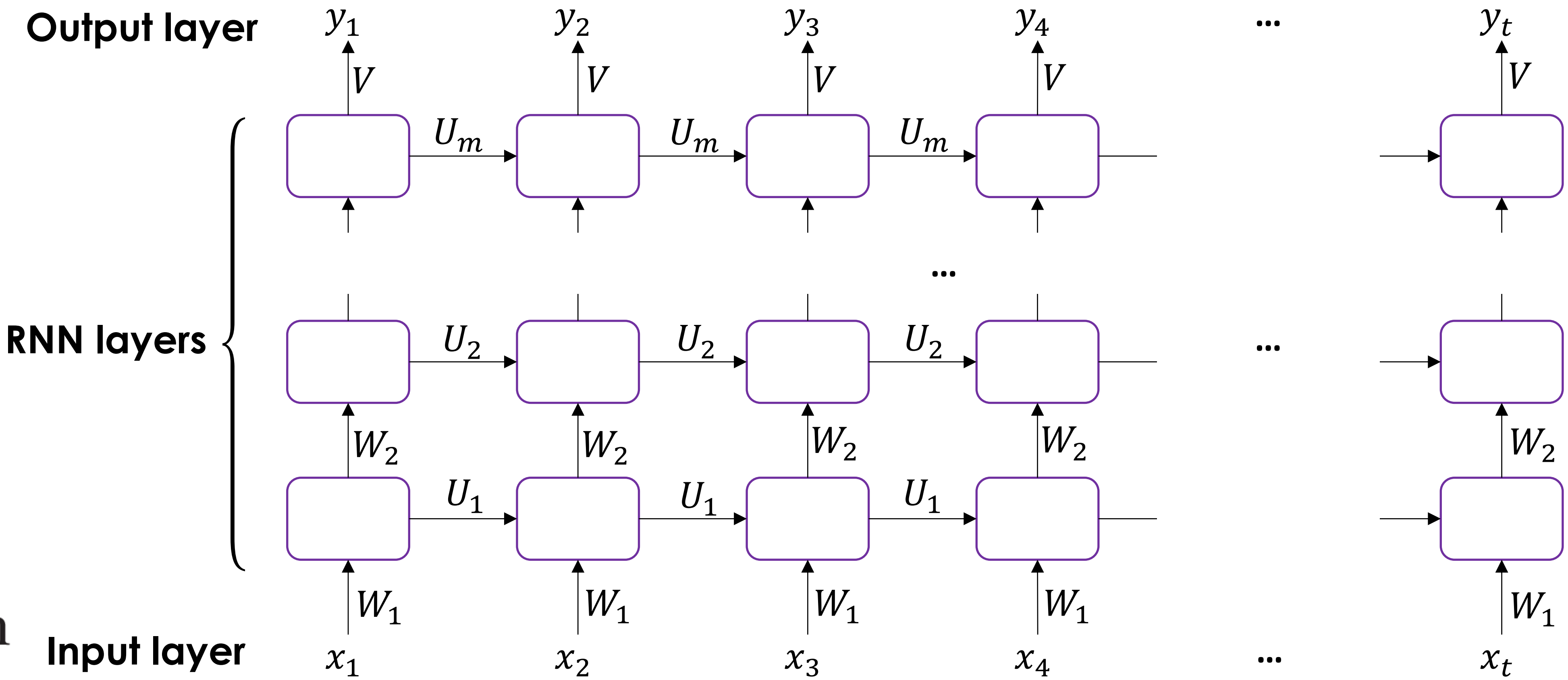
LSTMs vs. Basic RNNs

- **Advantages**
  - Partially address vanishing gradients problem but do not eliminate completely
  - Longer memory → Ability to bridge very long time lags
  - Generalise well

- **Limitations**
  - More complex → Slow computation
  - Prone to overfitting

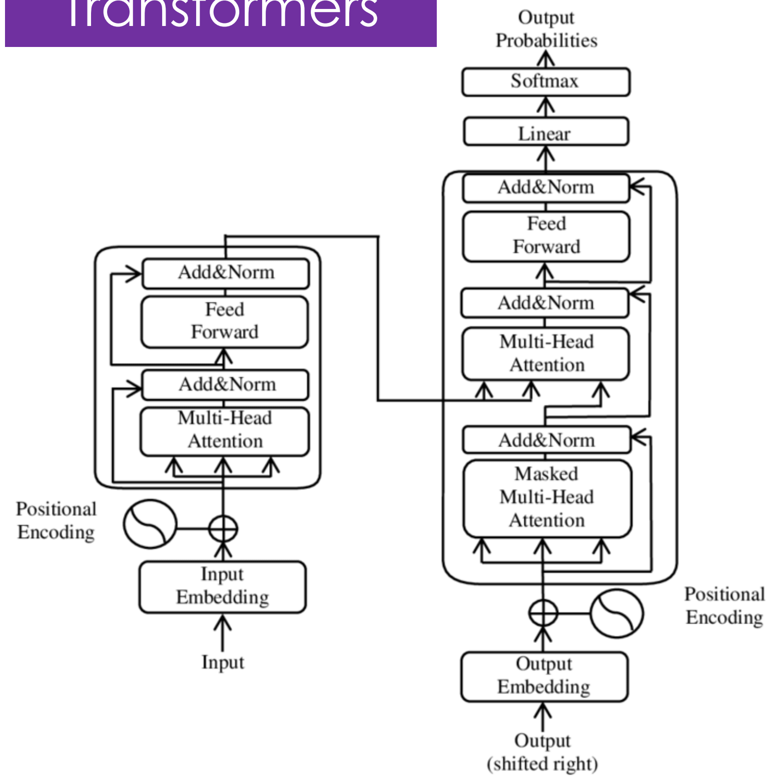
Deep RNNs

Basic RNN or LSTM units can be stacked to form deep complex RNN architectures

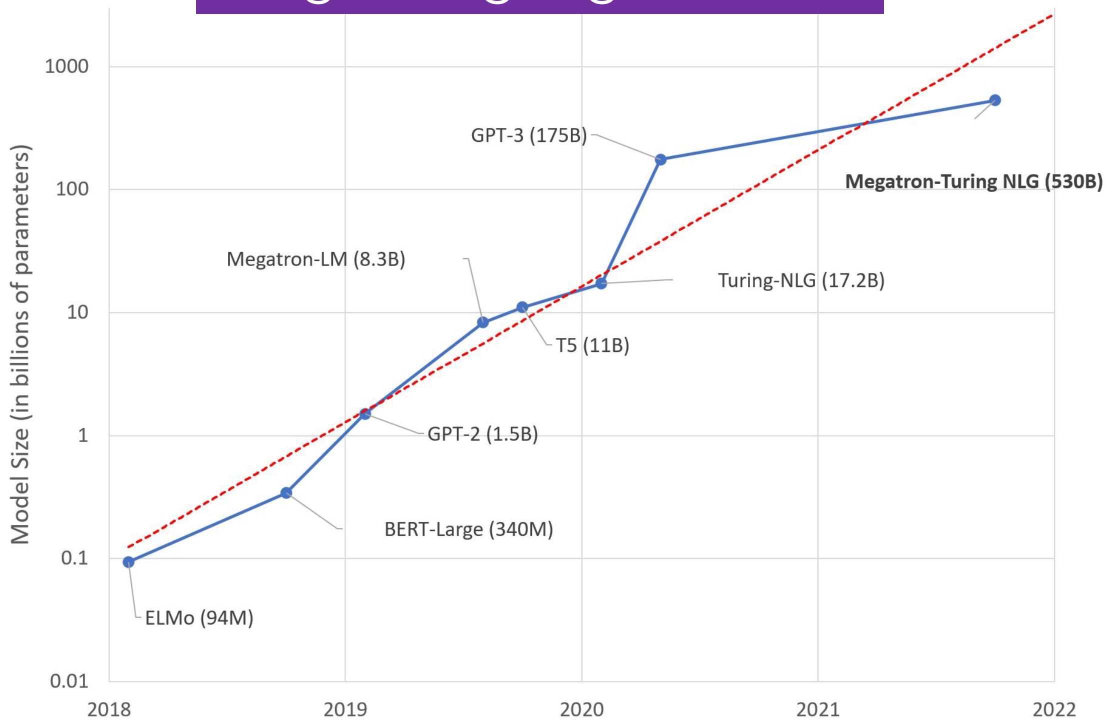


Recent advances in NLP

Transformers



Large Language Models



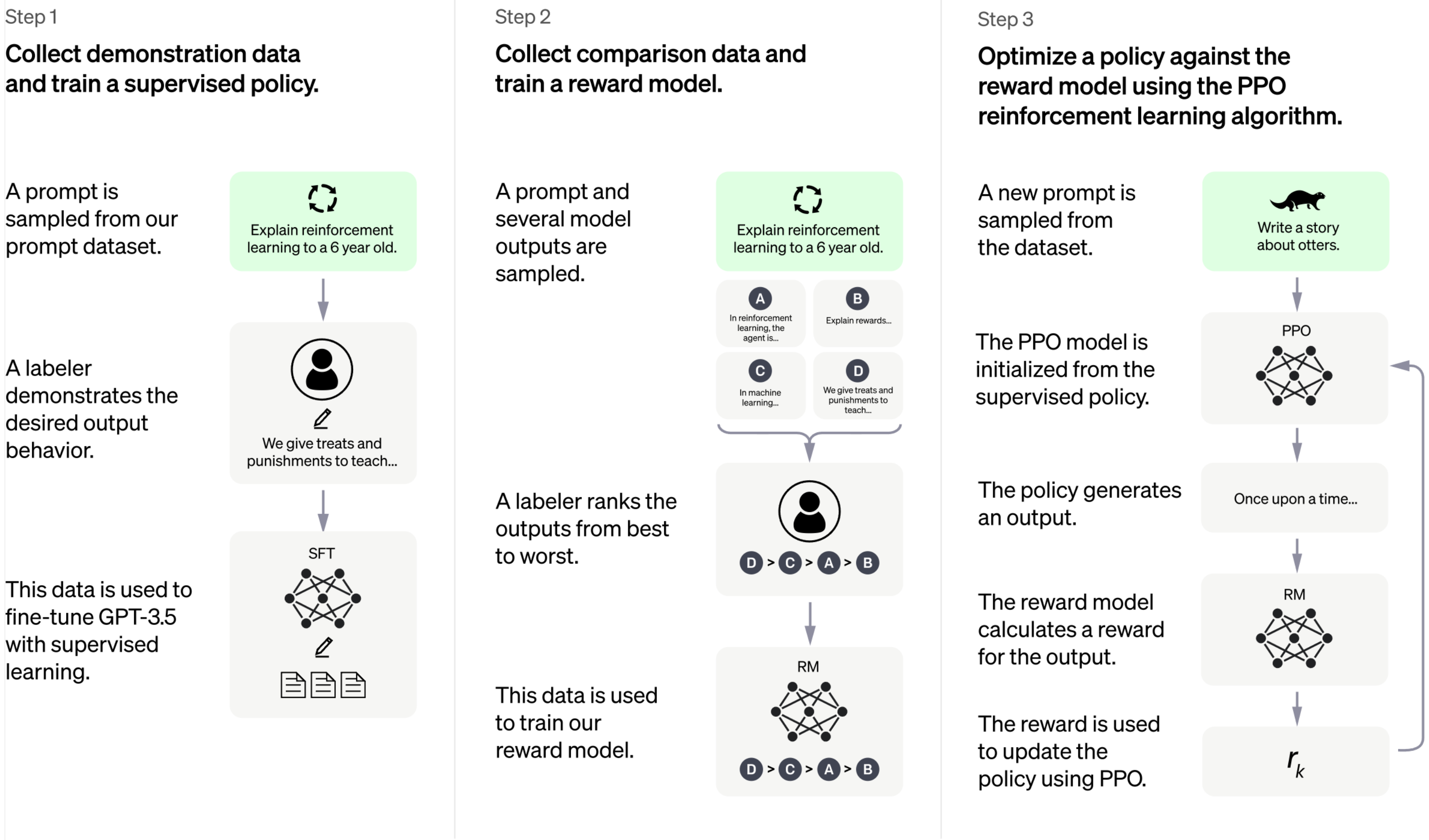
Durham University

OpenAI’s ChatGPT

Model follows an instruction in a prompt and provides a detailed response

Based on GPT-3.5

Trained using Reinforcement Learning from Human Feedback (RLHF)



Released: November 2022

SFT: Supervised Fine-Tuning
PPO: Proximal Policy Optimisation
RM: Reward Model

Durham University

OpenAI's ChatGPT examples

S What is a large language model?

A large language model is a type of artificial intelligence system that uses machine learning algorithms to understand and generate human language. It is trained on massive amounts of text data, often consisting of billions of words, to develop a deep understanding of language structure, grammar, and meaning.

Large language models can perform a variety of language-related tasks, such as language translation, question-answering, summarization, sentiment analysis, and more. They are able to generate human-like responses to input text and can be used in chatbots, virtual assistants, and other natural language processing applications.

Examples of large language models include OpenAI's GPT-3, Google's BERT, and Facebook's RoBERTa. These models have been shown to achieve state-of-the-art performance on a variety of language tasks and are constantly improving as they are trained on more data and fine-tuned for specific applications.

S Write a poem about Natural Language Processing

In the world of words and speech we dwell,
A vast and varied linguistic well,
From tongues and pens, the language flows,
Endless stories, thoughts, and woes.

But as we strive to understand,
The meanings hidden in each command,
We turn to NLP, a wondrous art,
To teach machines the human heart.

Through algorithms and data streams,
We build a bridge between two extremes,
Of human thought and machine might,
A union of logic and insight.

We seek to parse each phrase we hear,
To bring meaning to the words so dear,
And teach the machines to comprehend,
The subtleties of language without end.

With chatbots, voice assistants, and more,
We explore the bounds of what's in store,
And with each passing day we find,
New ways to teach, new paths to bind.

So let us celebrate this field of wonder,
Where language and tech become a thunder,
Of innovation and endless scope,
The power of NLP, a field of hope.

Durham University

*Queries submitted to ChatGPT on 09/03/2023

Questions?

Durham University