


Applied Text Analytics & Natural Language Processing



with Dr. Mahdi Roozbahani
& Wafa Louhichi

Deep Learning
Long Short-Term Memory (LSTM) – Part 1

Some of the slides are based on Ming Li (University of Waterloo – Deep Learning Part)
with some modifications



Learning Objectives

In this lesson, you will learn a deep learning model called LSTM

- Long Short-Term Memory (LSTM) an update to RNN
- Why LSTM
- Gated Recurrent Unit (GRU)



A Long-Term Dependency Example

Fill in the blank model:

I was born in Japan and lived there for ten years, and then moved to the US. My native language is _____

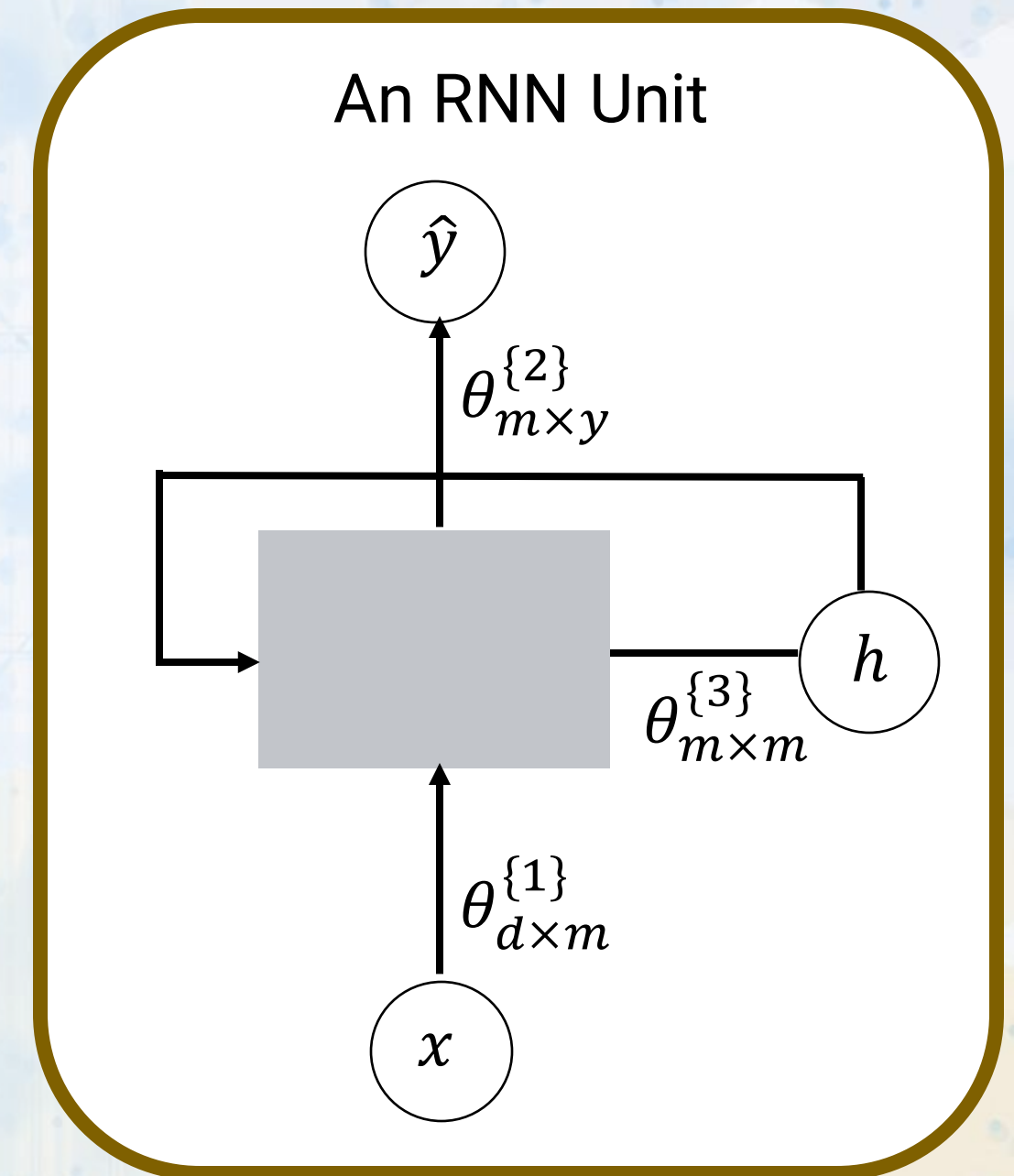
RNN could suffer from a short-memory issue in this example because of the vanishing gradient issue because of long dependencies between words to fill in the blank.

Gated cell: Long Short-Term Memory (LSTM)

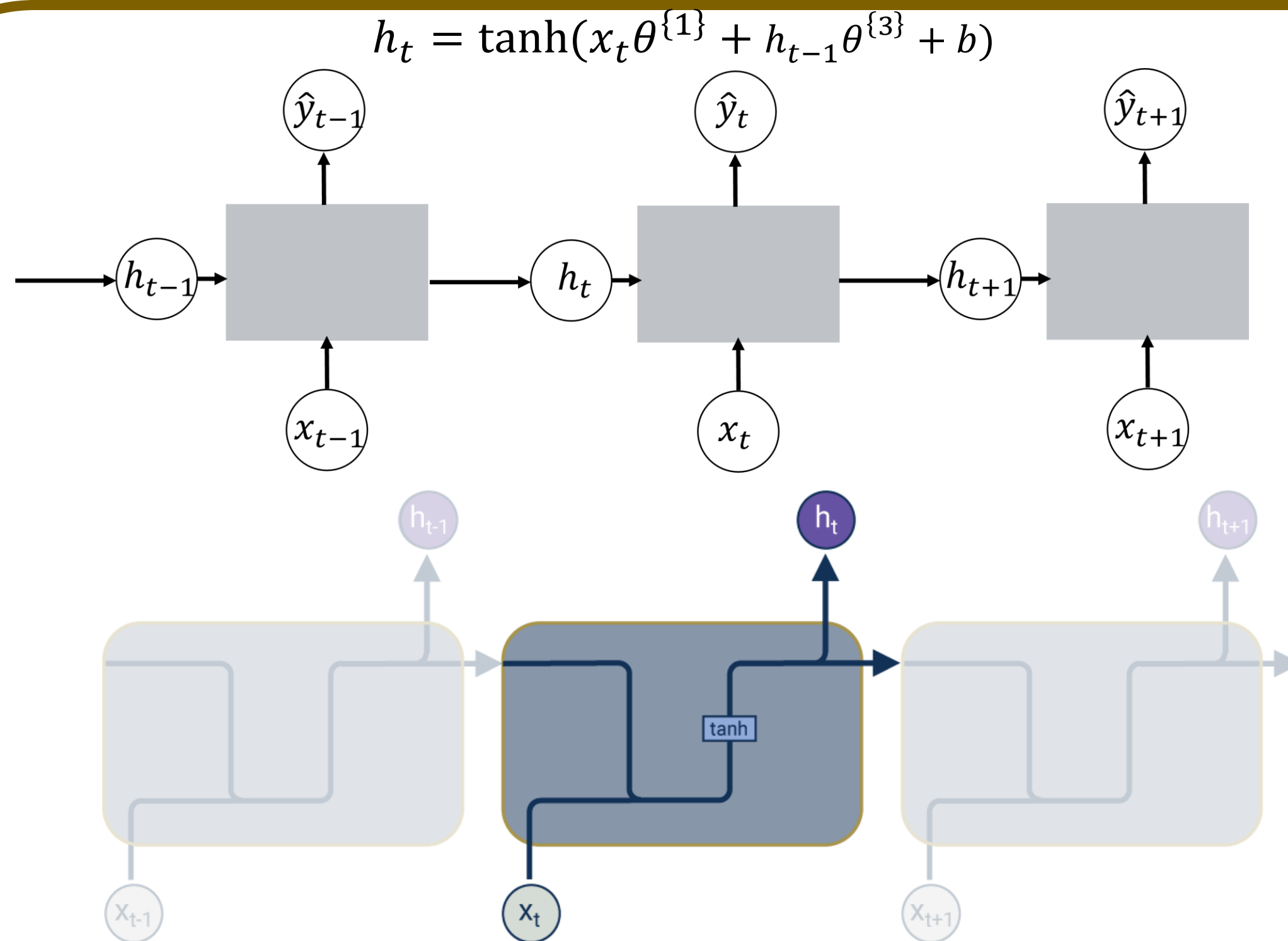
The main concept is to add a “gate” to an RNN unit to control what information is passed through the network.

LSTM essentially augments the RNN unit by creating gates that allow some information to be passed on through the network and some to be forgotten.

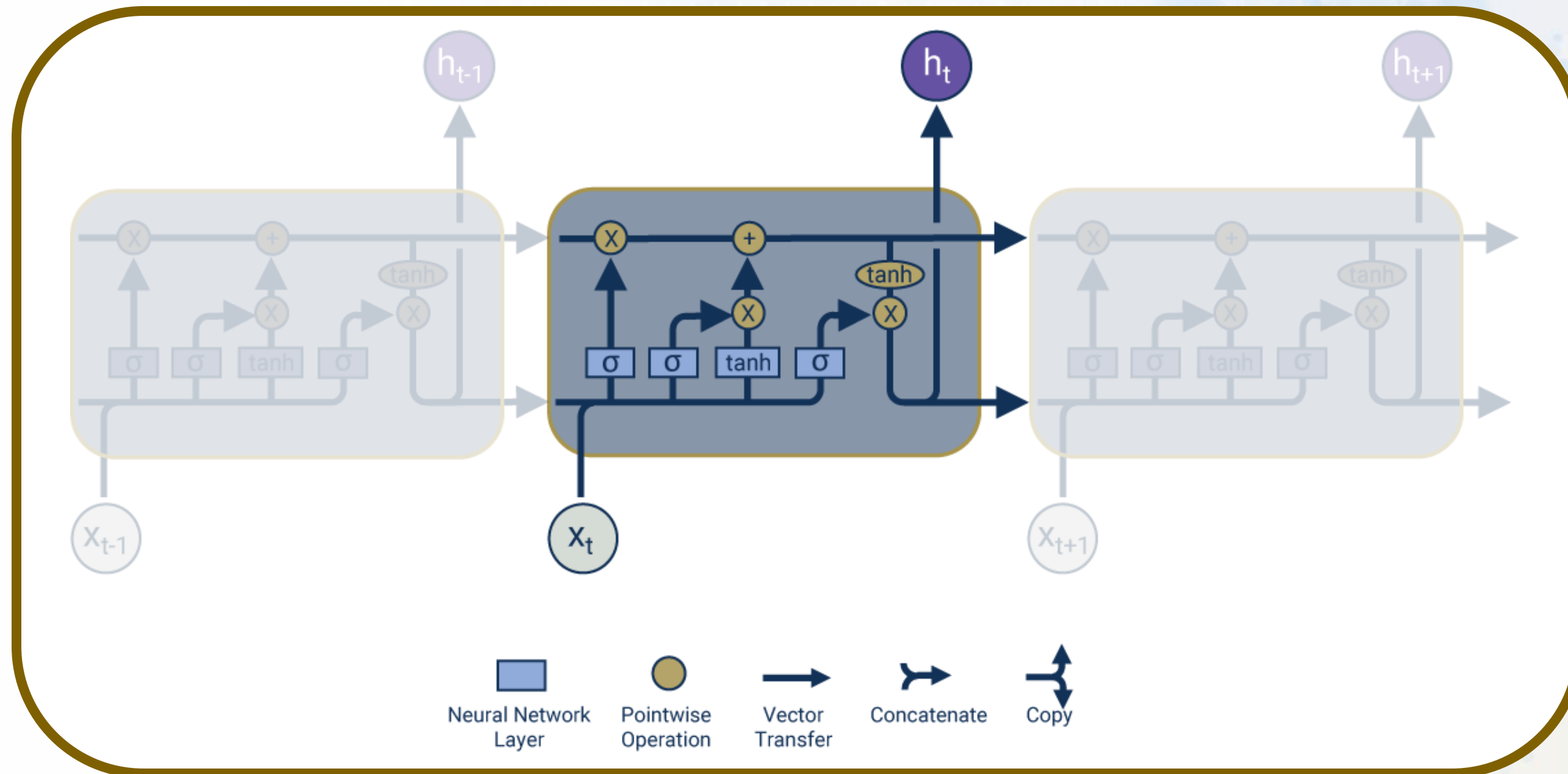
LSTM is the leading algorithm used in Sequential modeling




A Different Simple Representation of RNN

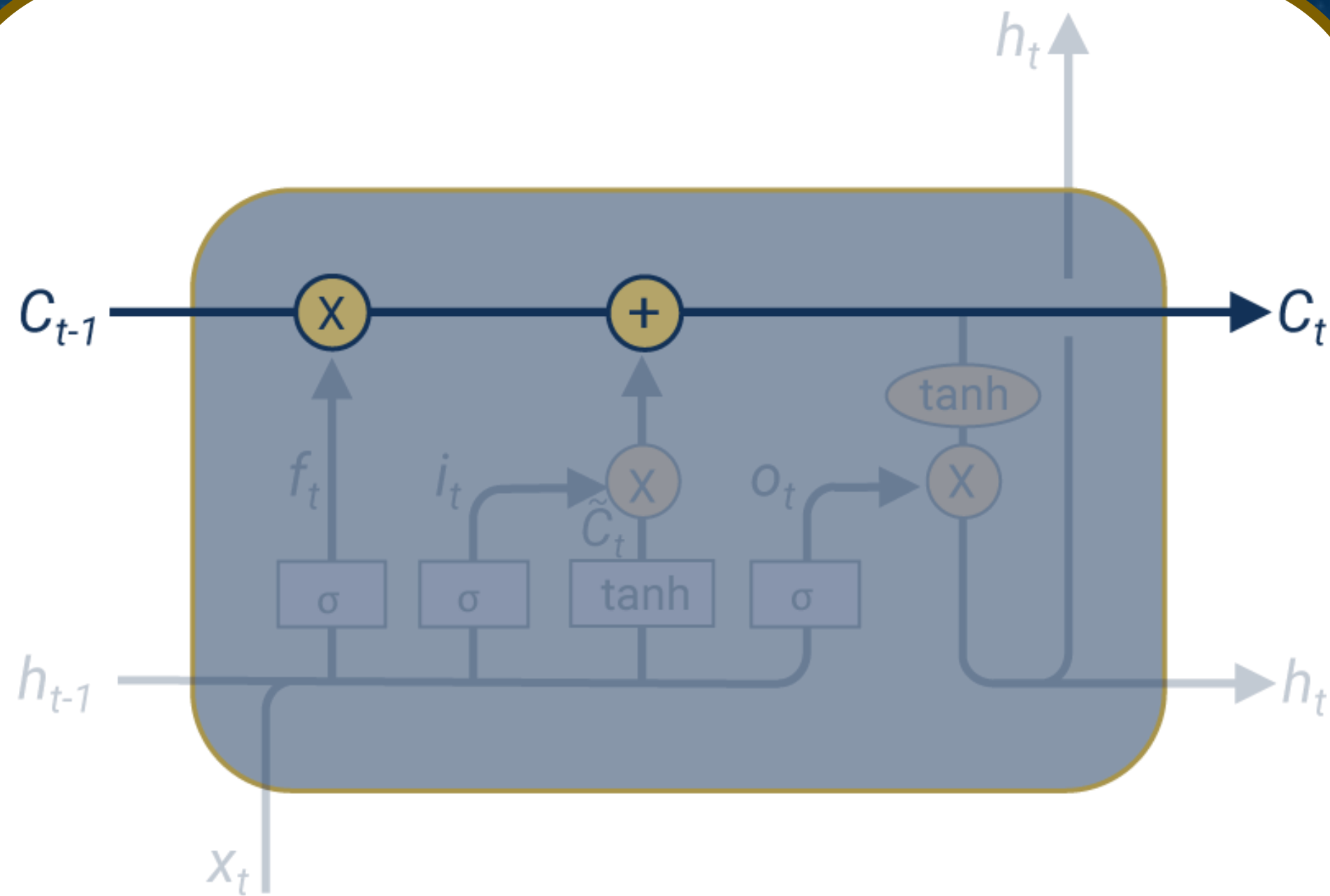


A Simple Representation of LSTM



 `tf.keras.layers.LSTM(num_units)`

The *cell state* in LSTM



C_t : cell state vector

How LSTM Can Control Information Removal and Addition for Cell State?

- This is called “gate” (the sigmoid operator) and controls how much information goes through.
- The sigmoid layer scales numbers between zero and one.
- A value of zero means do not let anything through (removal) and one means let all the information through (addition)

