Recurrent Neural Networks (RNNs) and Long Short Term Memory Networks (LSTMs)

Kristie Wirth

Jon Courtney

Frank Burkholder

Original content by Brandon Rohrer:

https://brohrer.github.io/how_rnns_lstm_work.html

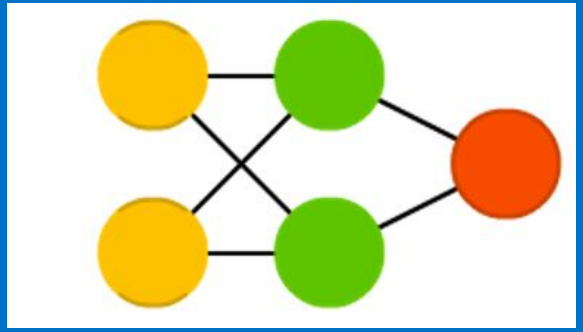
Learning Objectives

- Name some applications of RNNs & LSTMs
- Understand the structure of an RNN
- Know the problem with RNNs that LSTMs improve upon
- Describe the concept of gating & why it's used
- Name the three types of gates in LSTM

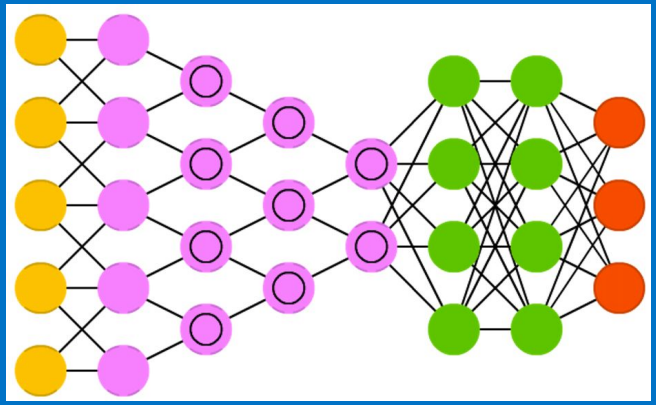
- Backfed Input Cell
- Input Cell
- Noisy Input Cell
- Hidden Cell
- Probablistic Hidden Cell
- Spiking Hidden Cell
- Output Cell
- Match Input Output Cell
- Recurrent Cell
- Memory Cell
- Different Memory Cell
- Kernel
- Convolution or Pool

Original (larger) chart comparing structures of many more neural networks: http://www.asimovinstitute.org/neural-network-zoo/

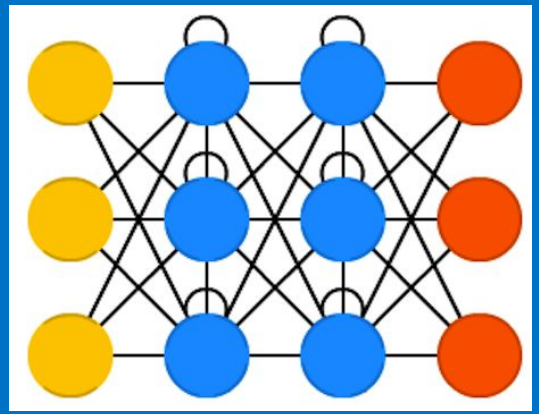
Multilayer Perceptrons (MLP)



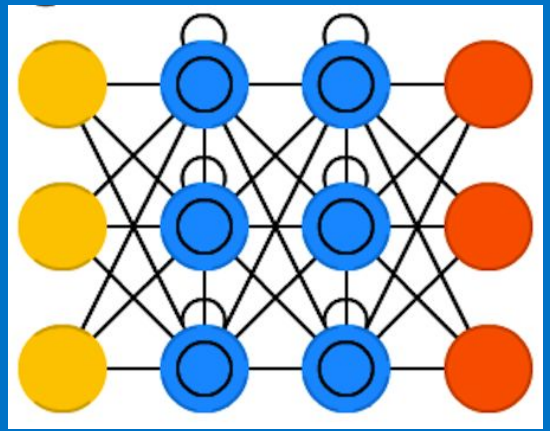
Convolutional Neural Networks (CNN)



Recurrent Neural Networks (RNN)



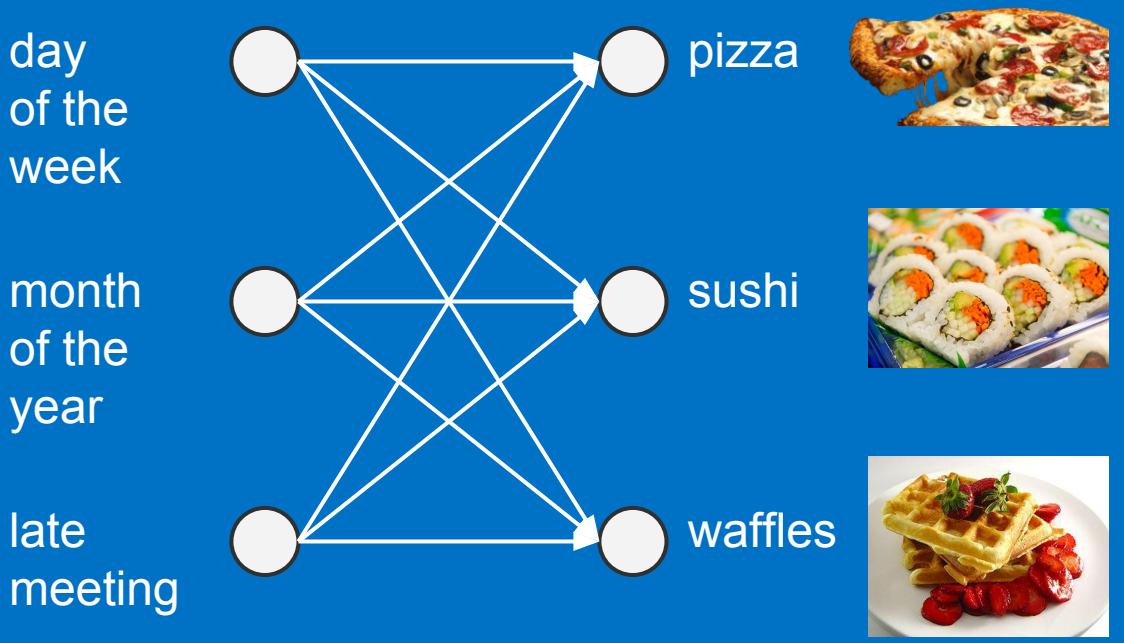
Long Short Term Memory (LSTM)



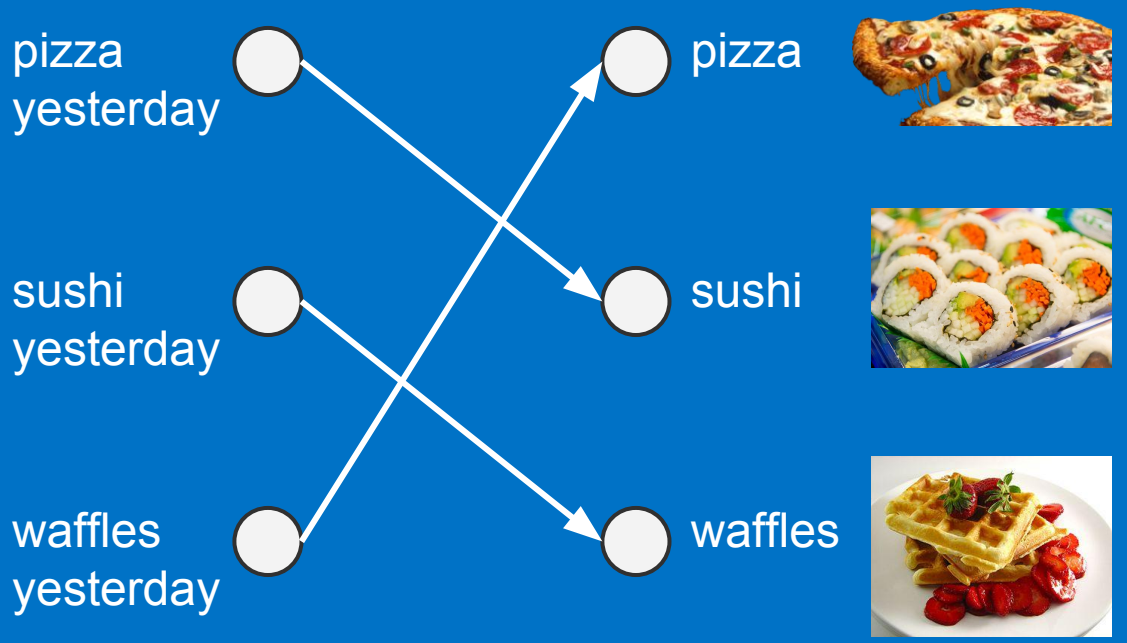
Applications for RNNs & LSTMs

- **Pattern recognition**: handwriting, captioning images
- **Sequential data**: speech recognition, stock price prediction, and generating text and news stories
- Translating between...
  - Speech
  - Text
  - Different languages
  - Audio
  - Video
- Physical processes, including robotics
- Anything embedded in time

Using a Neural Network to Predict What's for Dinner

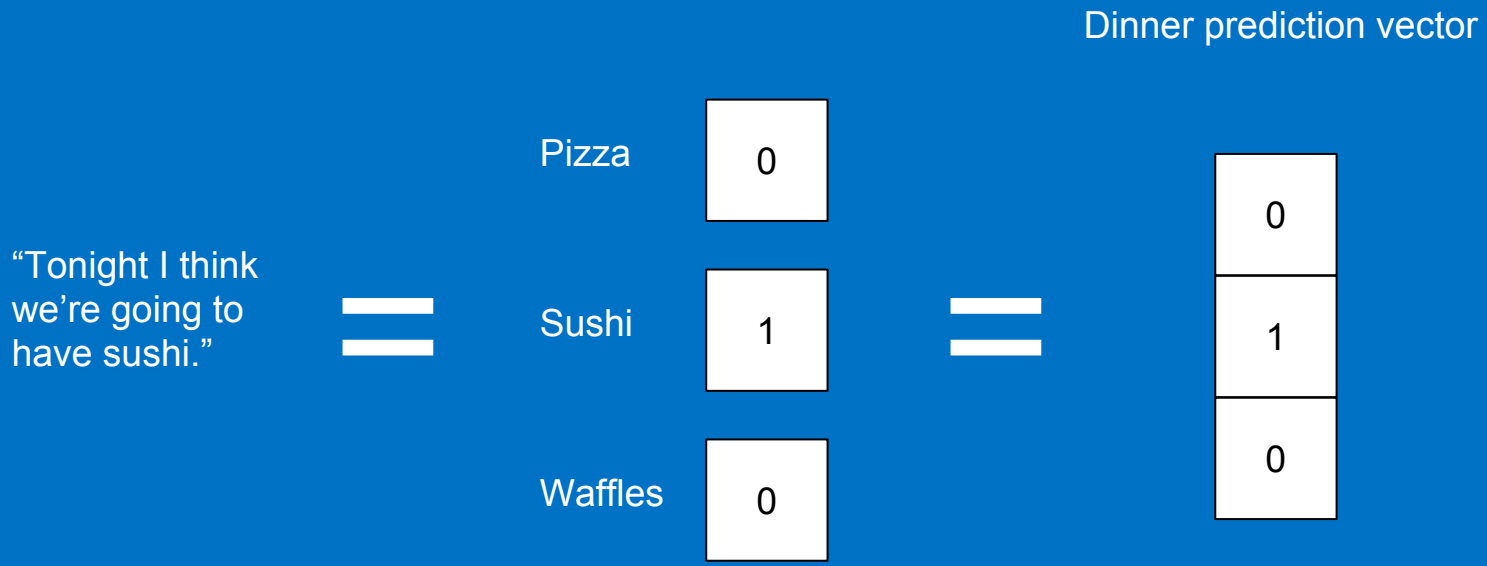


A Revised Neural Network to Predict What's for Dinner

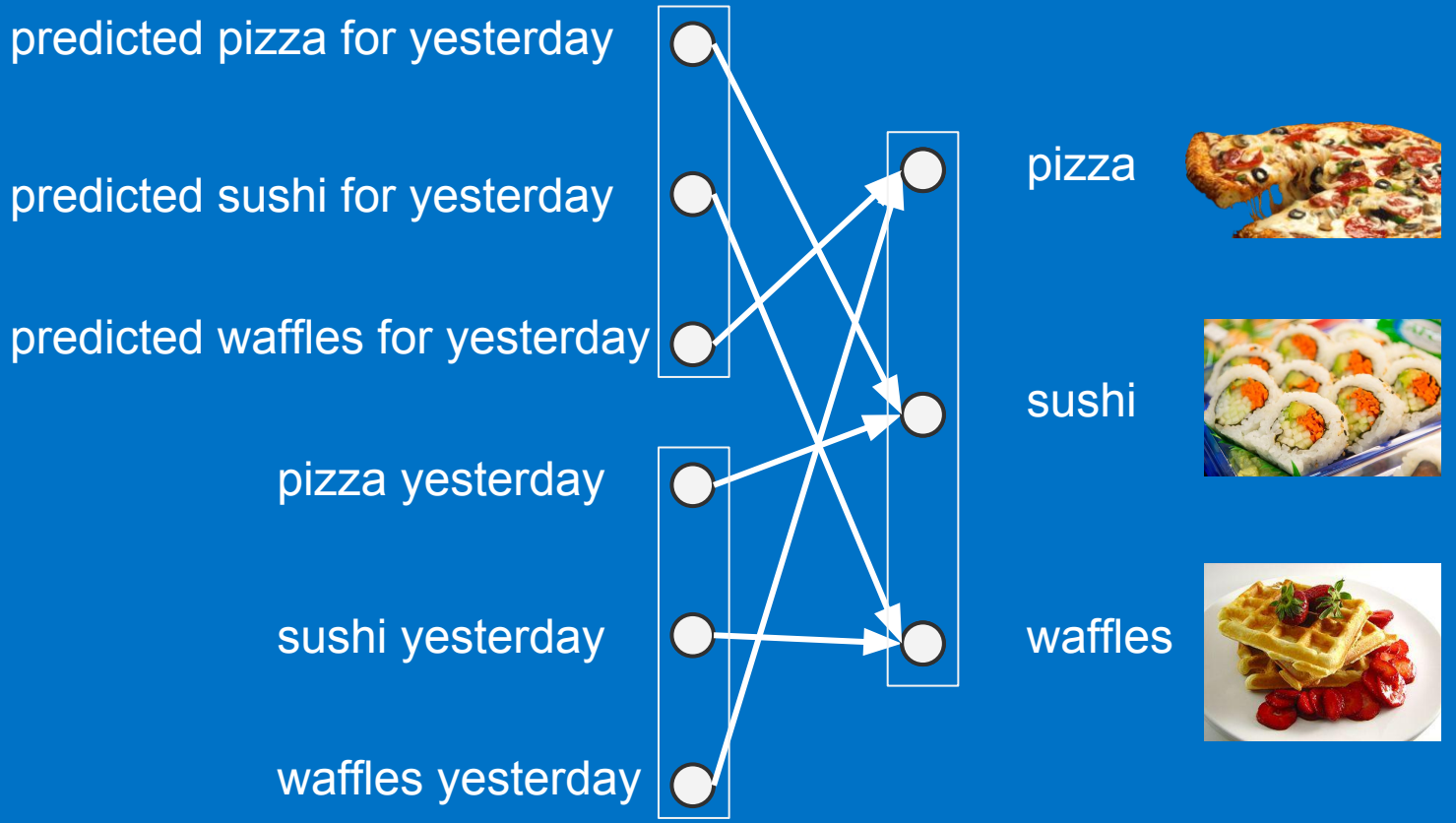


A Revised Neural Network to Predict What's for Dinner

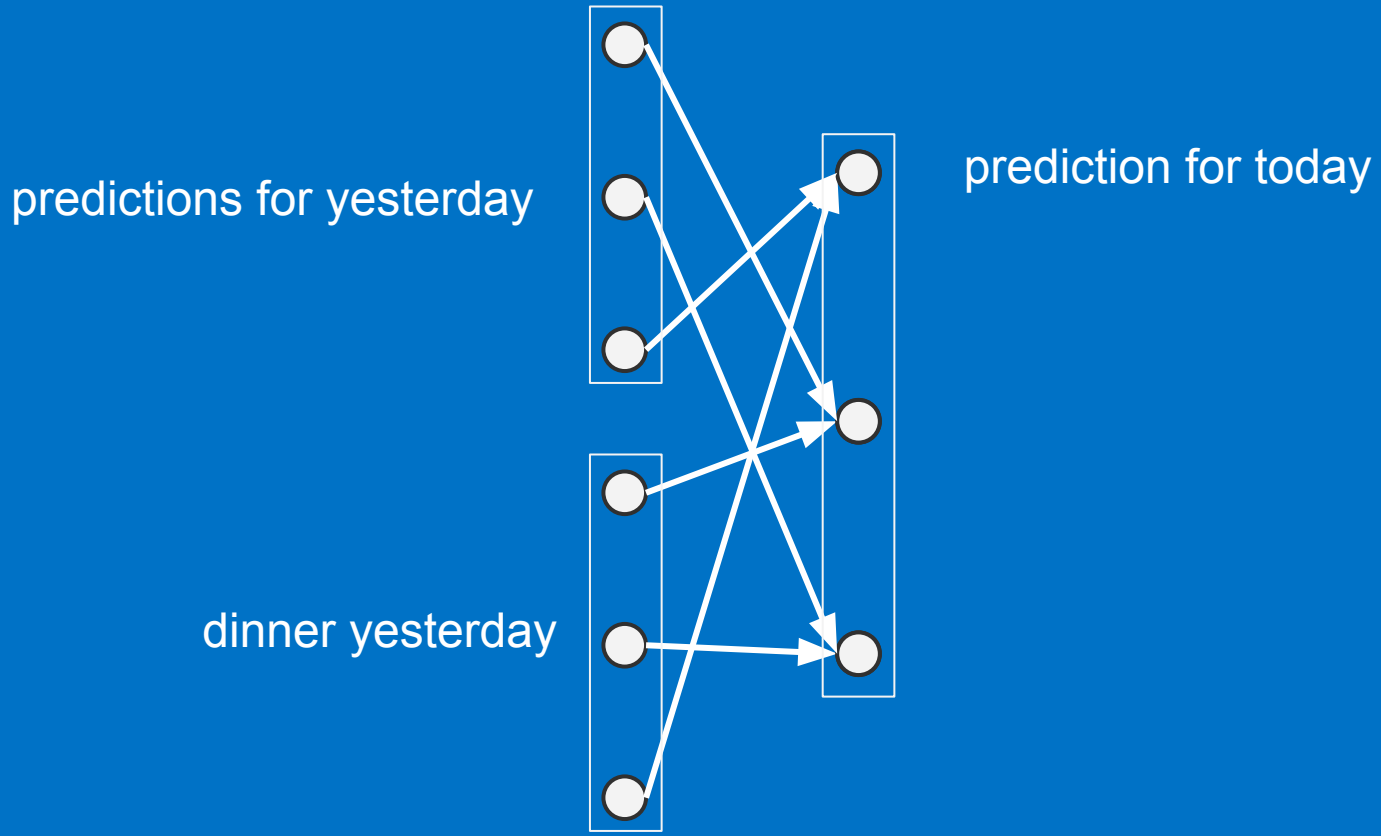
Dinner Prediction Vector



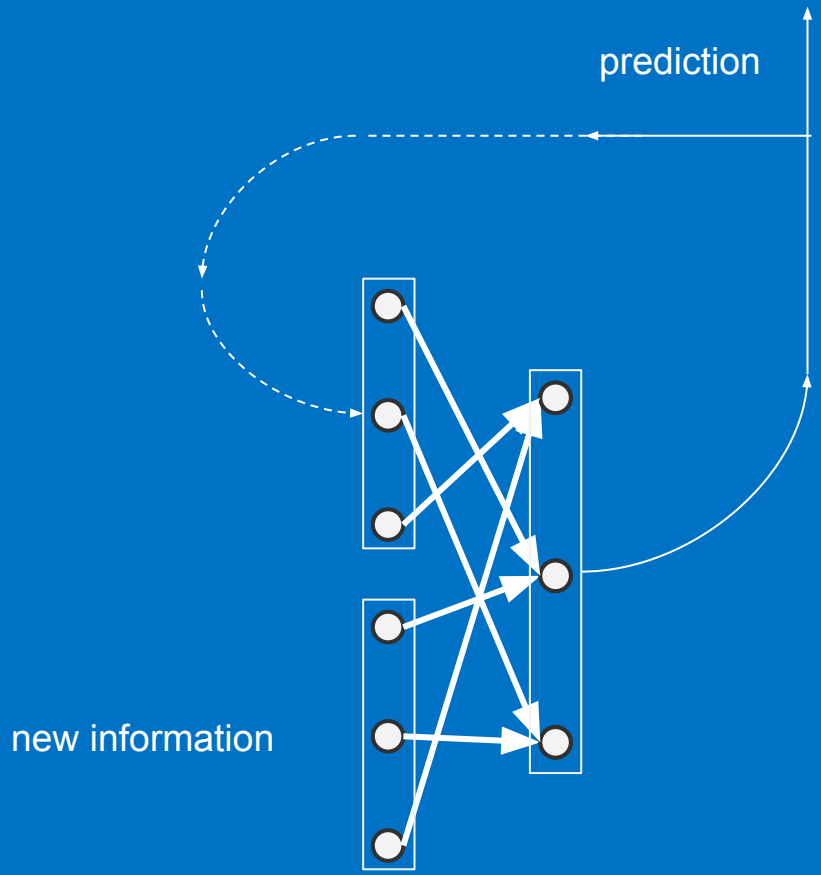
A Revised Neural Network to Predict What's for Dinner



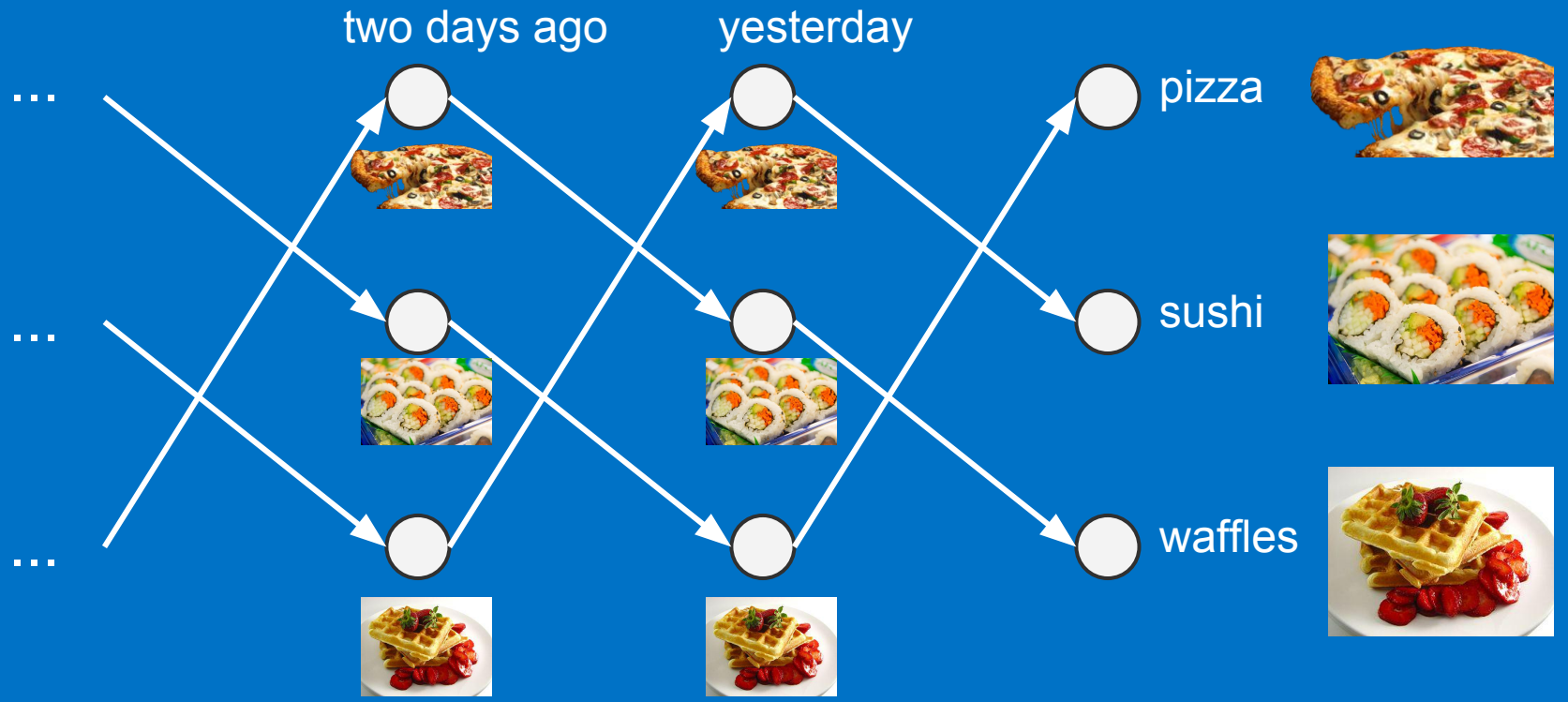
A Revised Neural Network to Predict What's for Dinner



A Revised Neural Network to Predict What's for Dinner



Unrolled Dinner Predictions

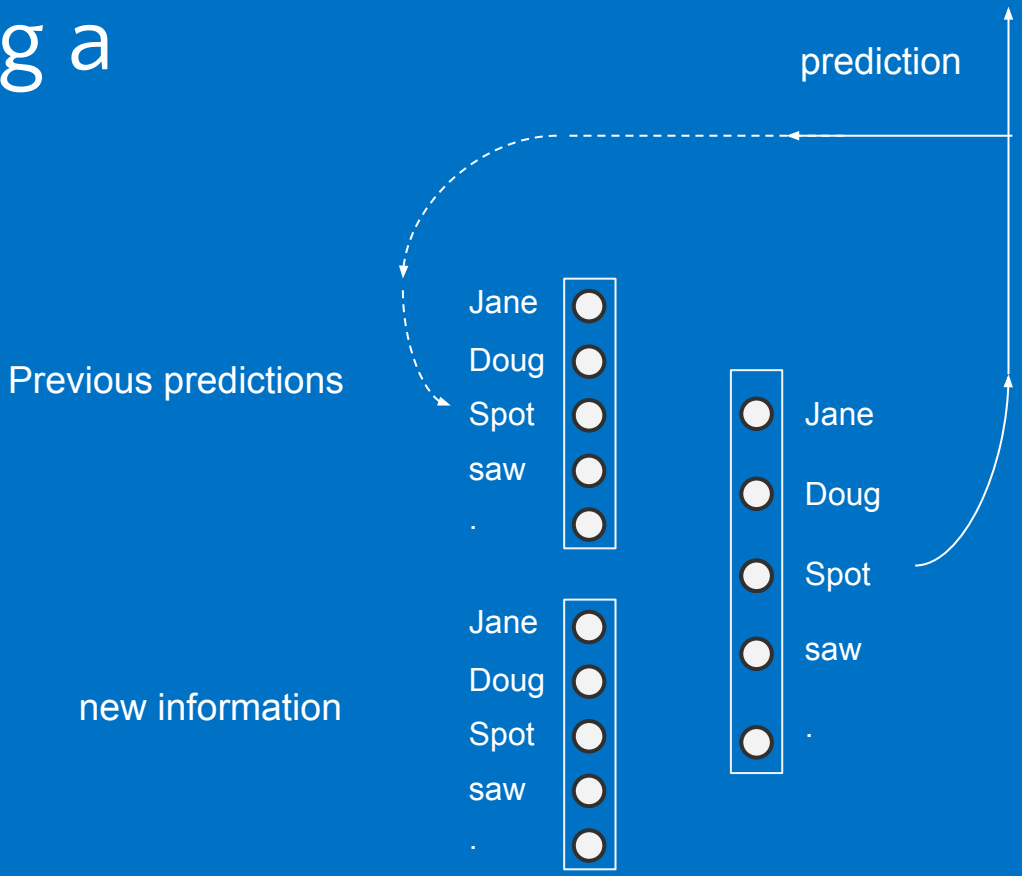


Another RNN Example: Let's Write a Children's Book

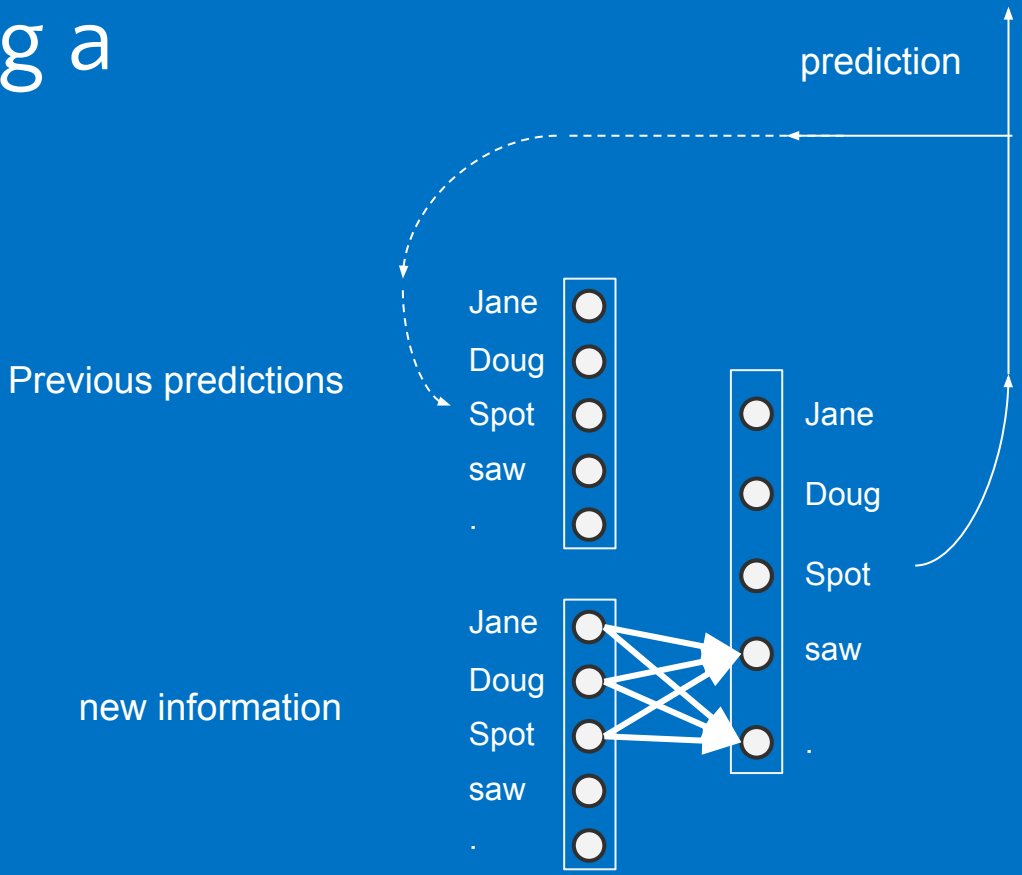
Doug saw Jane.
Jane saw Spot.
Spot saw Doug.
...

Your dictionary is small: {Doug, Jane, Spot, saw, .}

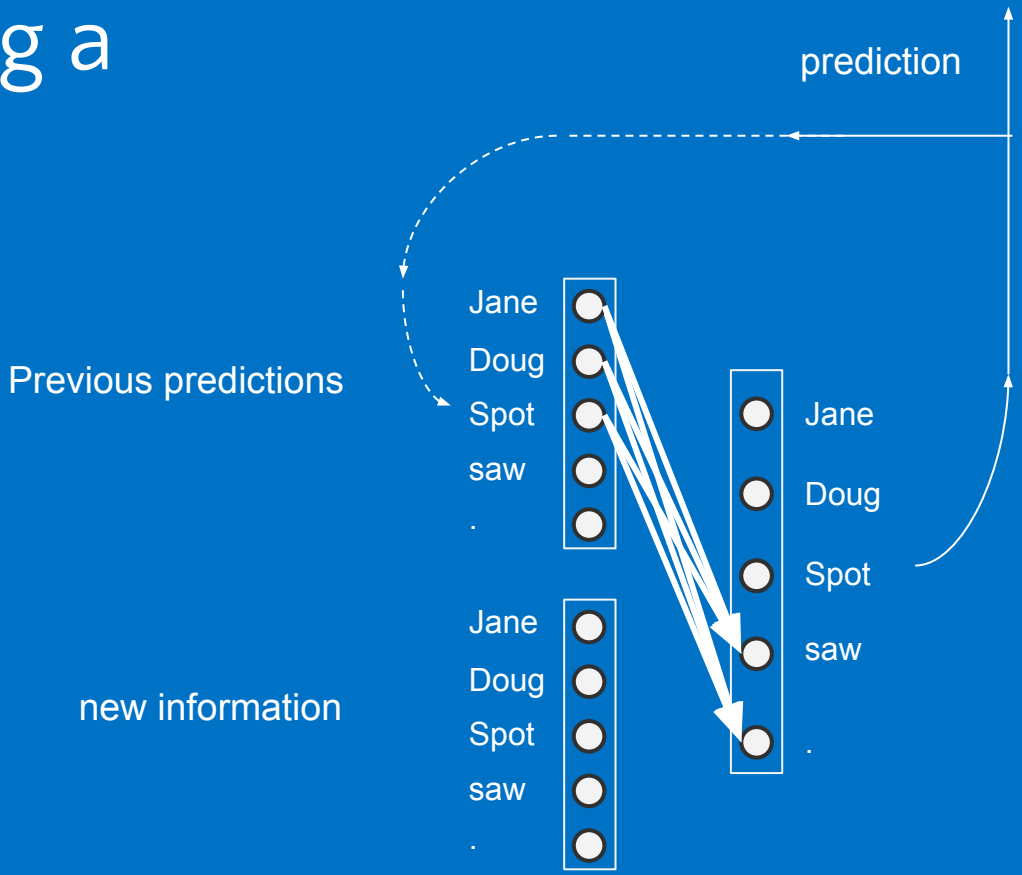
Using a RNN


prediction
Previous predictions
Jane
Doug
Spot
saw
.
Jane
Doug
Spot
saw
.
new information
Jane
Doug
Spot
saw
.

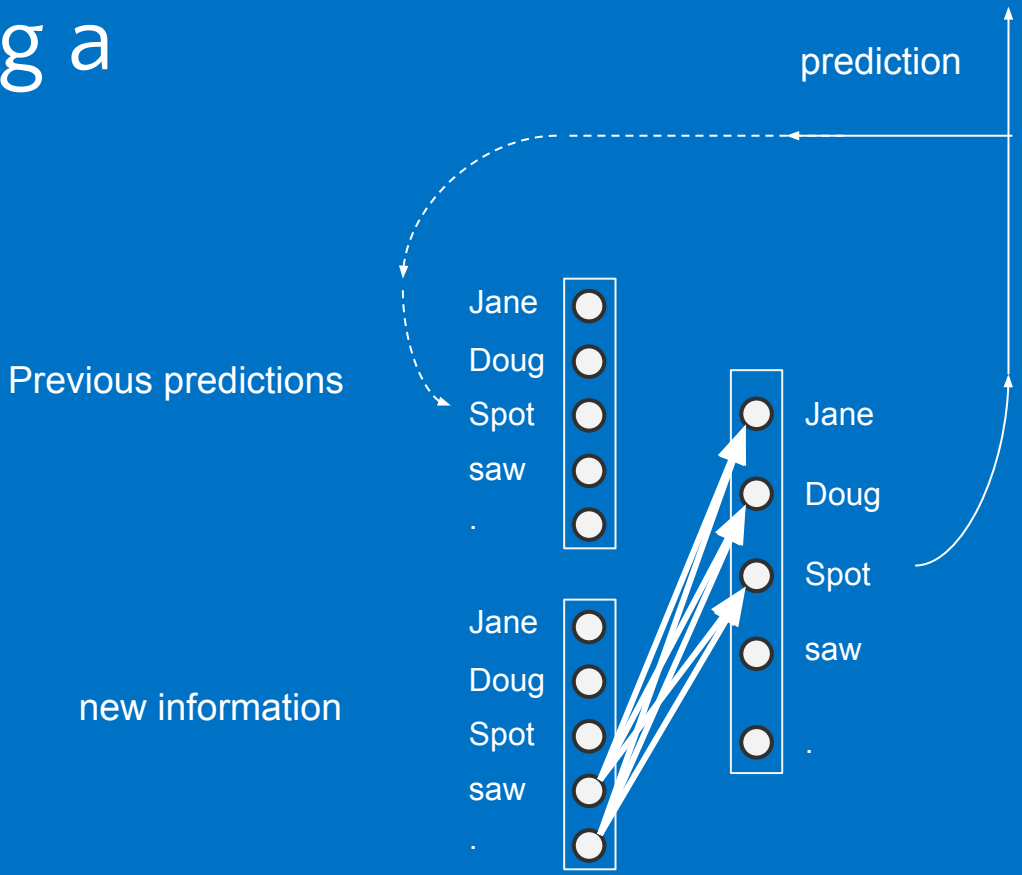
Using a RNN


prediction
Previous predictions
Jane
Doug
Spot
saw
.
Jane
Doug
Spot
saw
.
new information
Jane
Doug
Spot
saw
.

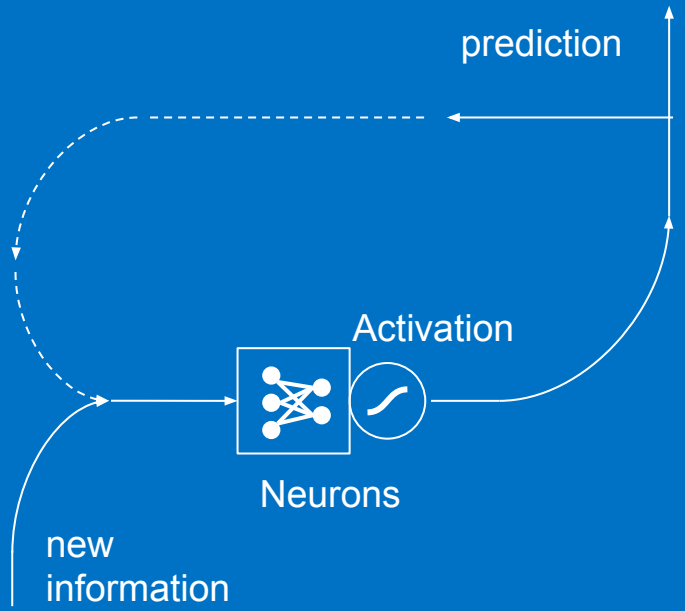
Using a RNN


prediction
Previous predictions
Jane
Doug
Spot
saw
.
new information
Jane
Doug
Spot
saw
.
Jane
Doug
Spot
saw
.

Using a RNN


prediction
Previous predictions
Jane
Doug
Spot
saw
.
new information
Jane
Doug
Spot
saw
.
Jane
Doug
Spot
saw
.

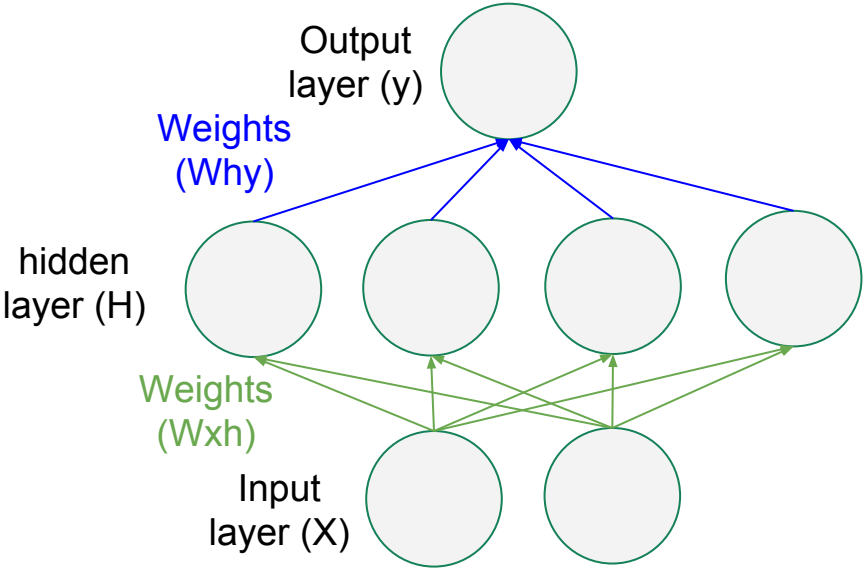
Using a RNN



In an RNN, what you predicted previously influences your current prediction.

How many weights are in each network?

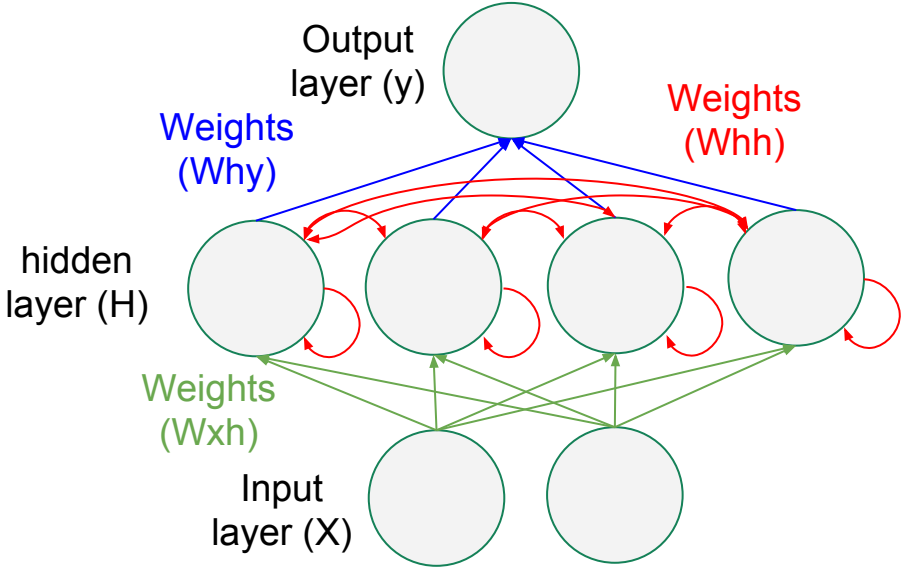
Vanilla MLP



Why =

Wxh =

Vanilla RNN



Note: A double arrow indicates a weight in each direction (2 weights).

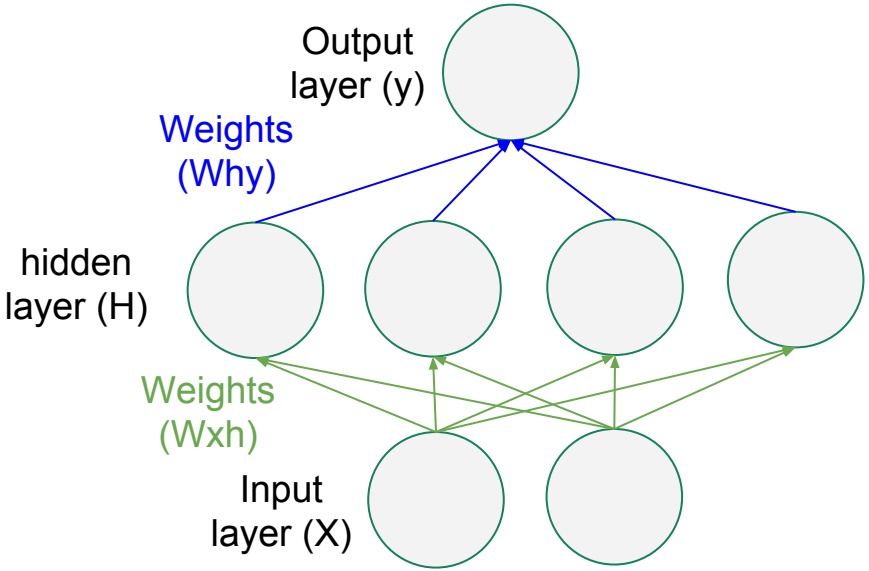
Why =

Whh =

Wxh =

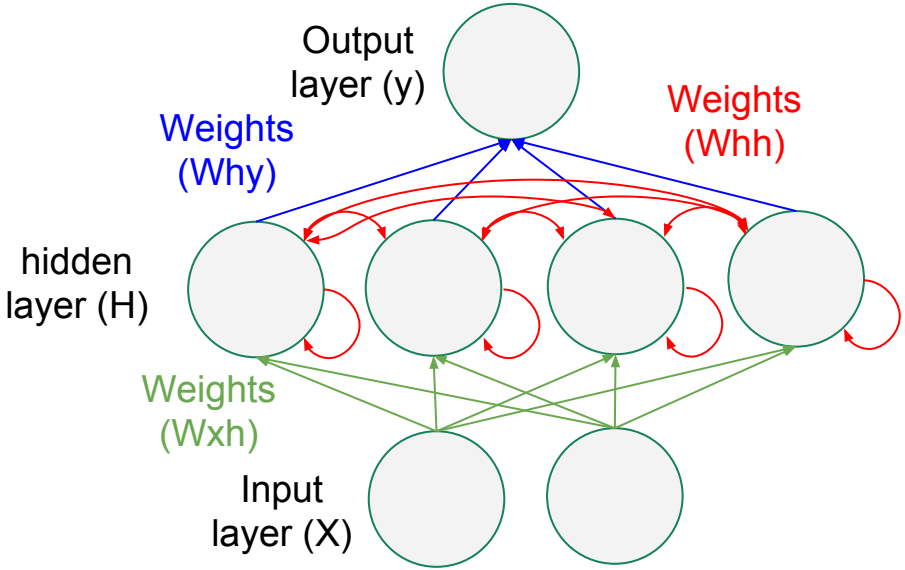
How many weights are in each network?

Vanilla MLP



Why = 4
Wxh = 8

Vanilla RNN



Why = 4
Whh = 16
Wxh = 8

Breakout: Examine an RNN

Run the min-char-rnn.py code, a RNN that is learning how to write like Dr. Seuss. As the model trains it will eventually write some new Dr. Seuss inspired books!

Look over the code and answer the following questions:

- How many hidden layers are there?
- What are the inputs?
- What are the targets?
- Why is the model training using these inputs & targets?
- What activation function is used?
- Can you make any other observations on how the model works?

Mistakes an RNN can make

Doug saw Doug.

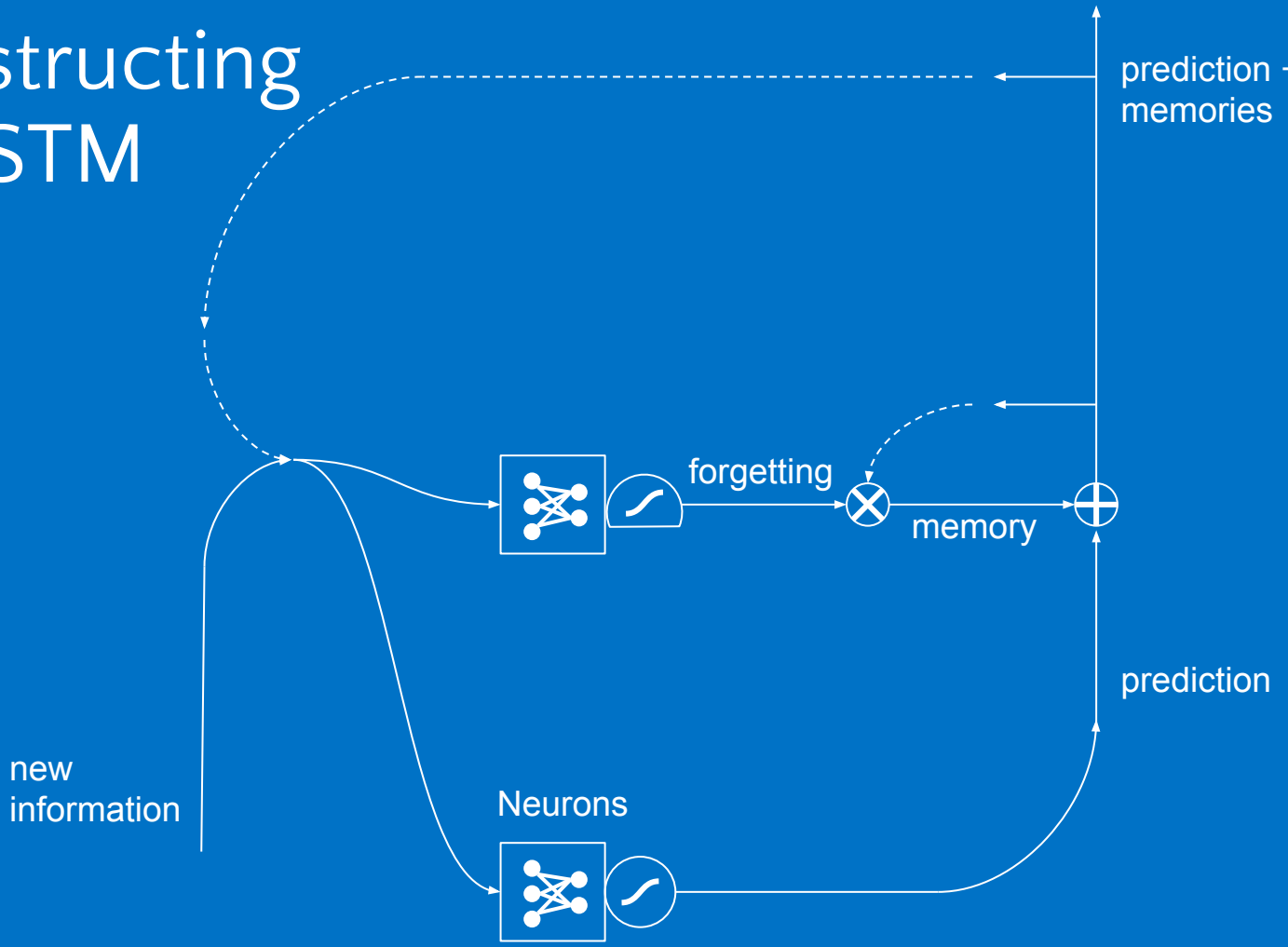
Jane saw Spot saw Doug saw ...

Spot. Doug. Jane.

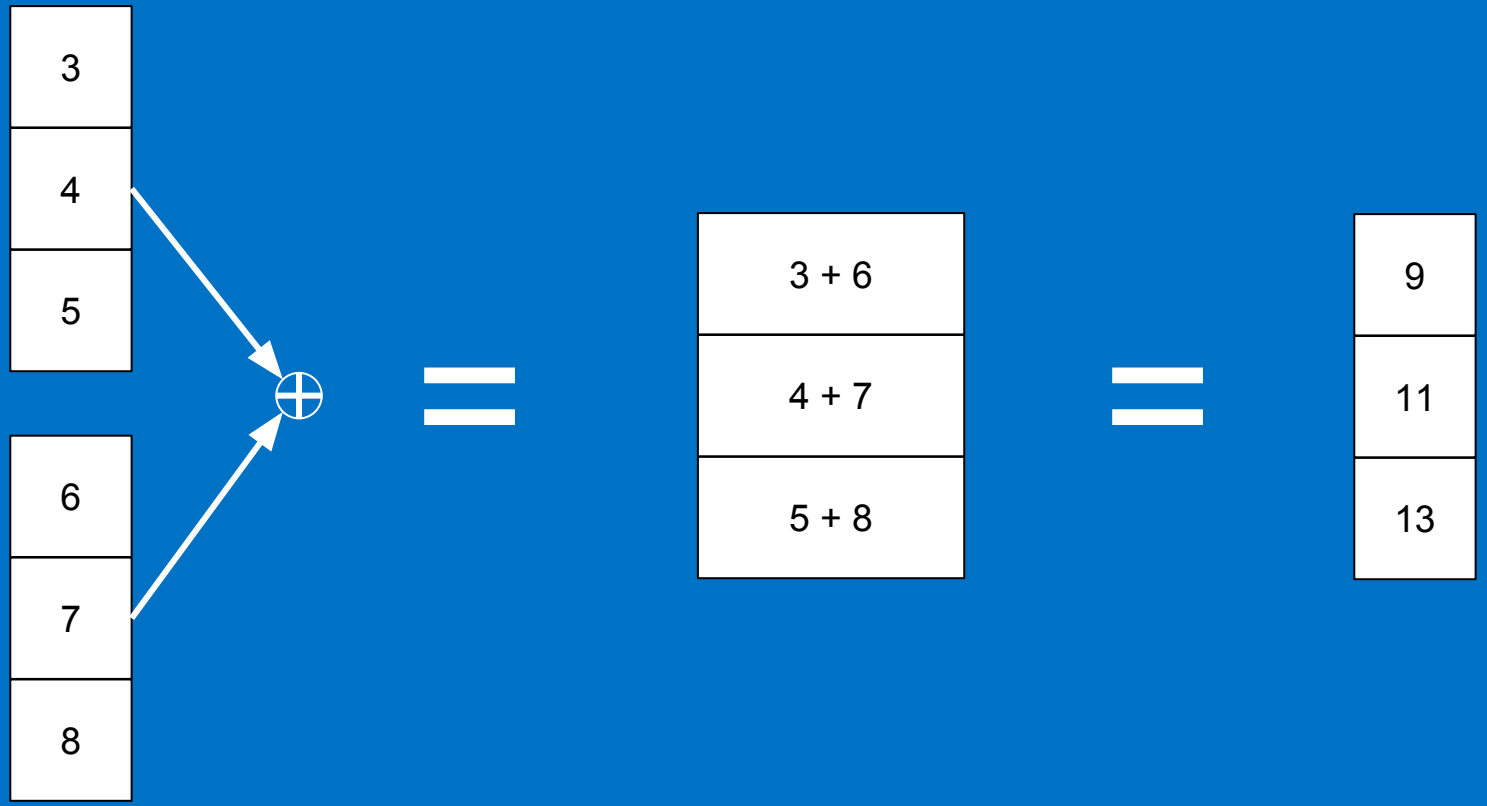
Moving into the world of LSTMs

- "Long short-term"
  - *Long* term memory refers to the learned weights
  - *Short term* memory refers to the values related to gates that change with each step through the LSTM
- LSTMs have a more complex structure than RNNs
- There are different types of LSTMs, that have different components or connections
- They extend the ability of RNNs to remember (and forget) deeper into the past
- LSTMs seek to address RNNs' exploding/vanishing gradients problem
  - Continuing to multiply a quantity by a number >1 or <1 can cause that quantity to become very large (exploding) or very small (vanishing)
  - Because the layers of RNNs relate to each other through multiplication, their derivatives can have this problem
  - If we don't know the gradients, we can't adjust the weights to continue learning

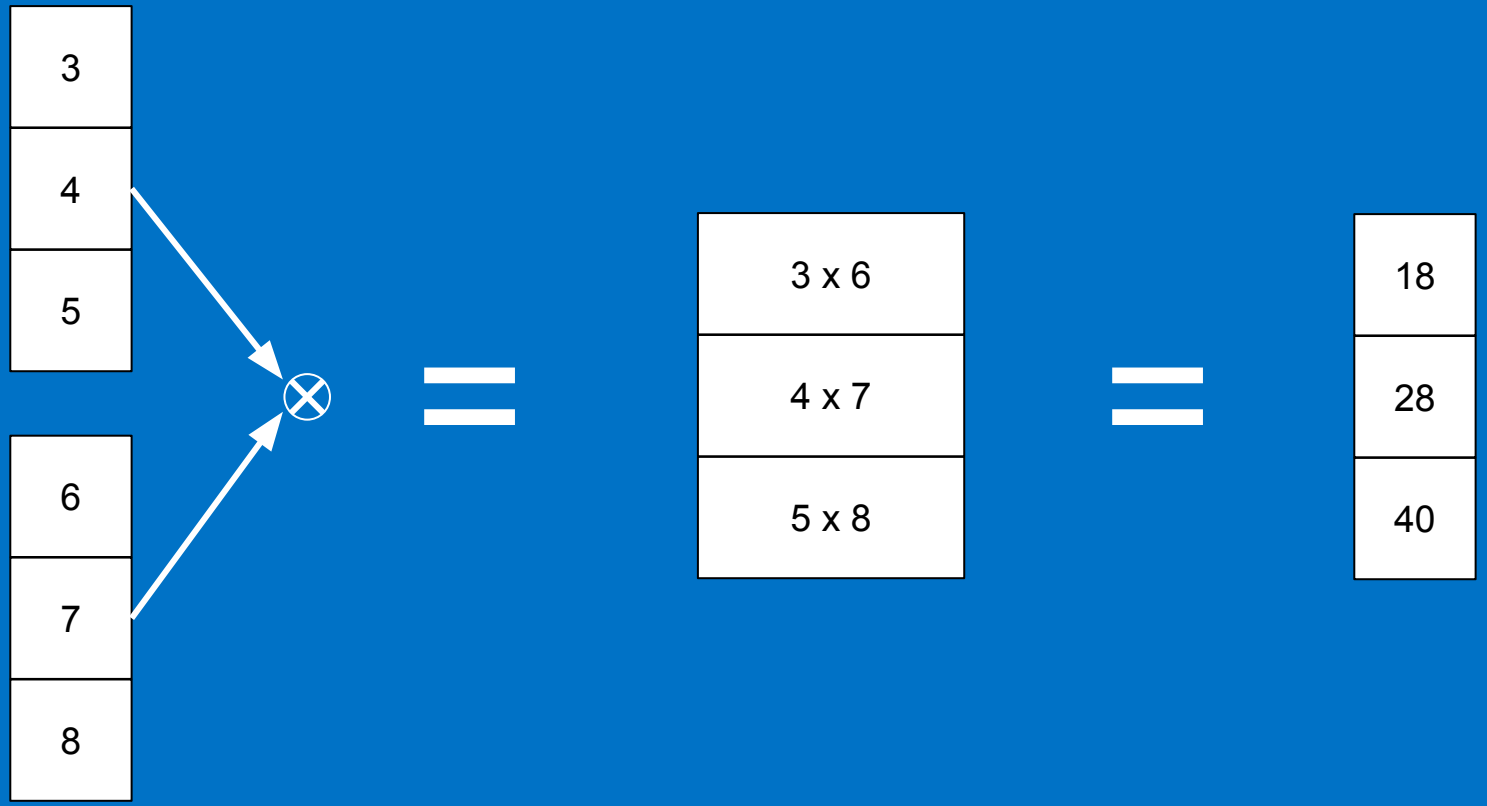
Constructing an LSTM



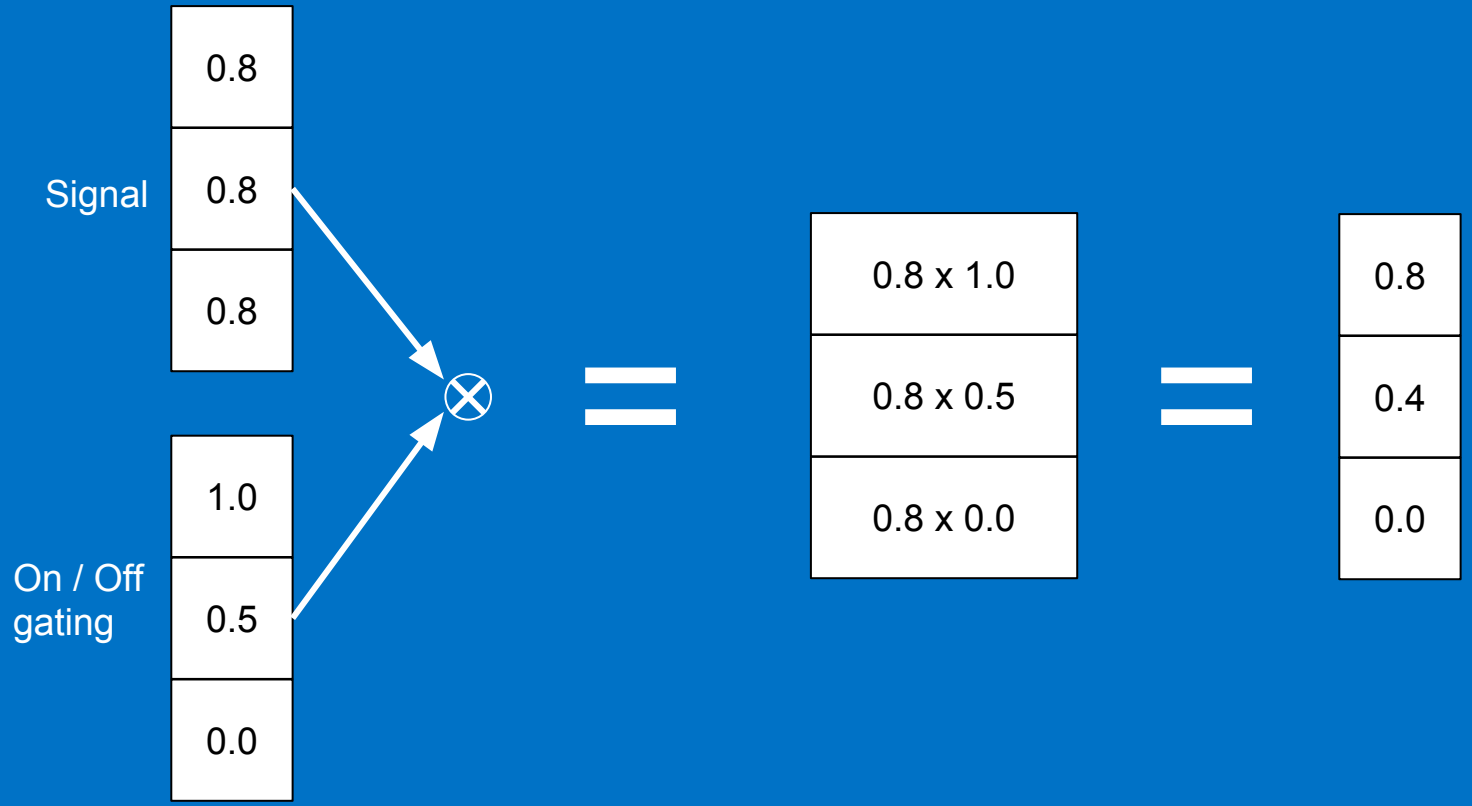
Plus junction: element-by-element addition



Times junction: element-by-element multiplication

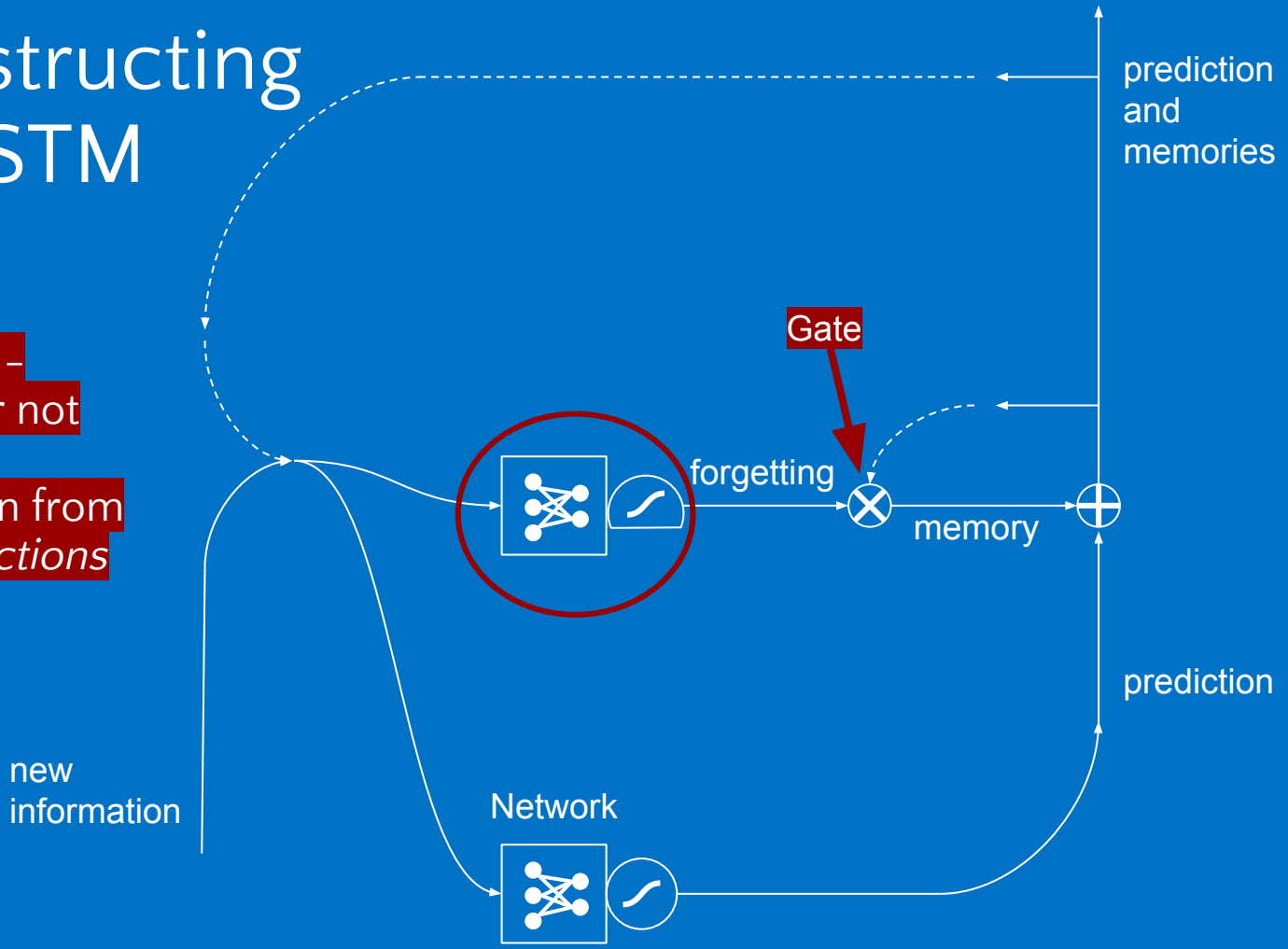


Gating


Signal
0.8
0.8
0.8
On / Off gating
1.0
0.5
0.0
0.8 x 1.0
0.8 x 0.5
0.8 x 0.0
0.8
0.4
0.0

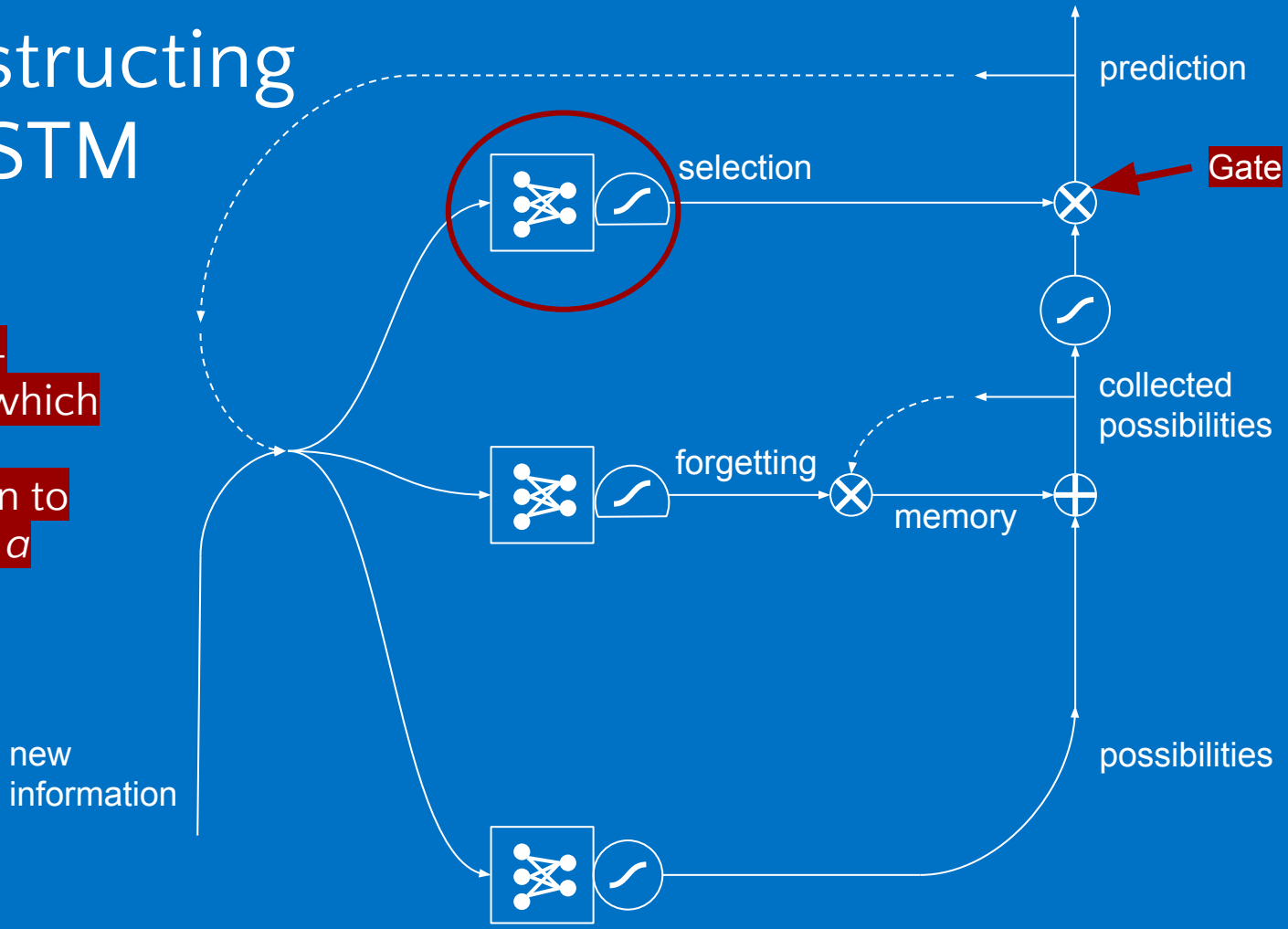
Constructing an LSTM

Forgetting – Keeping or not keeping information from *past predictions*



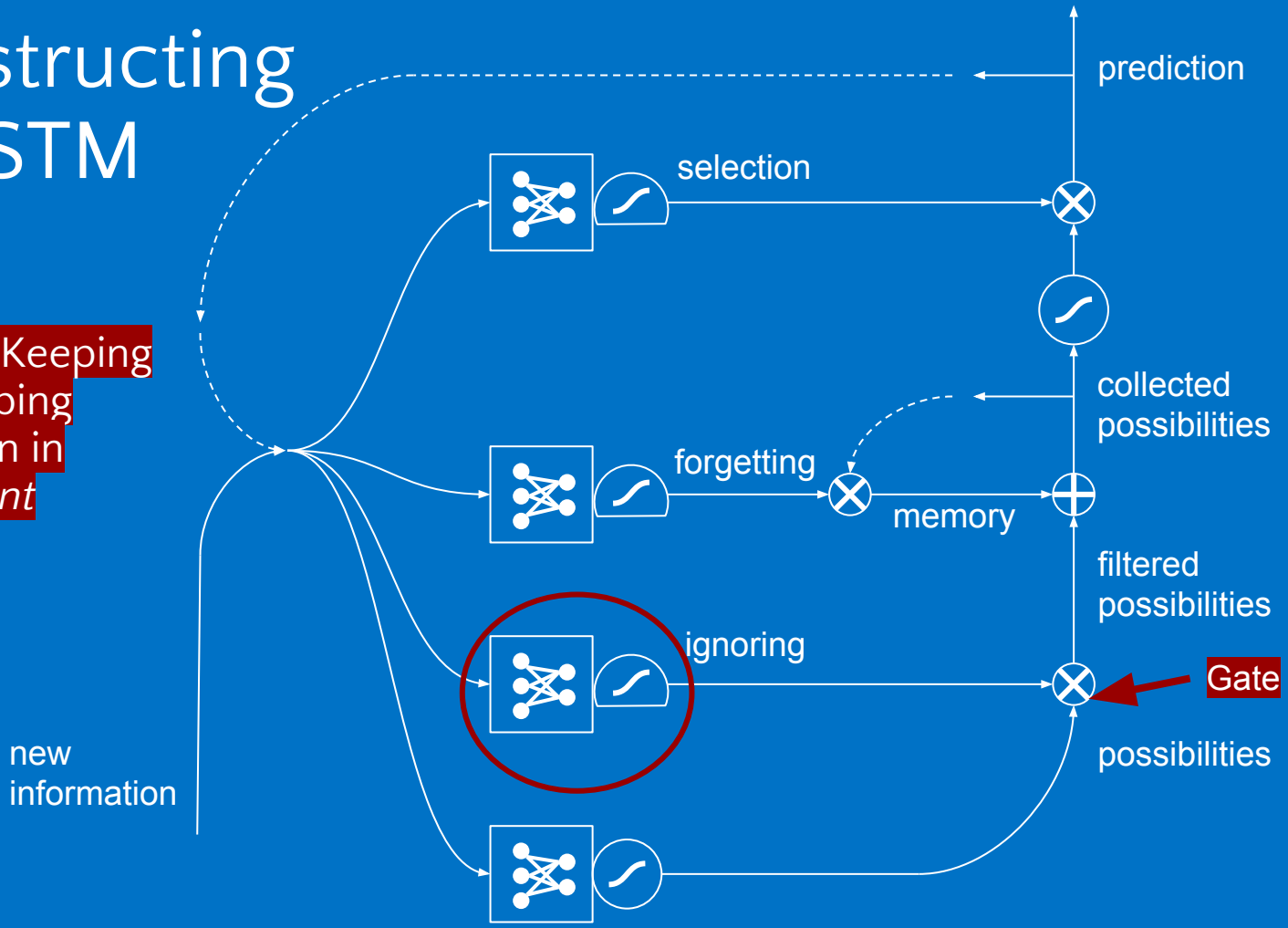
Constructing an LSTM

Selection – Selecting which internal information to *release as a prediction*

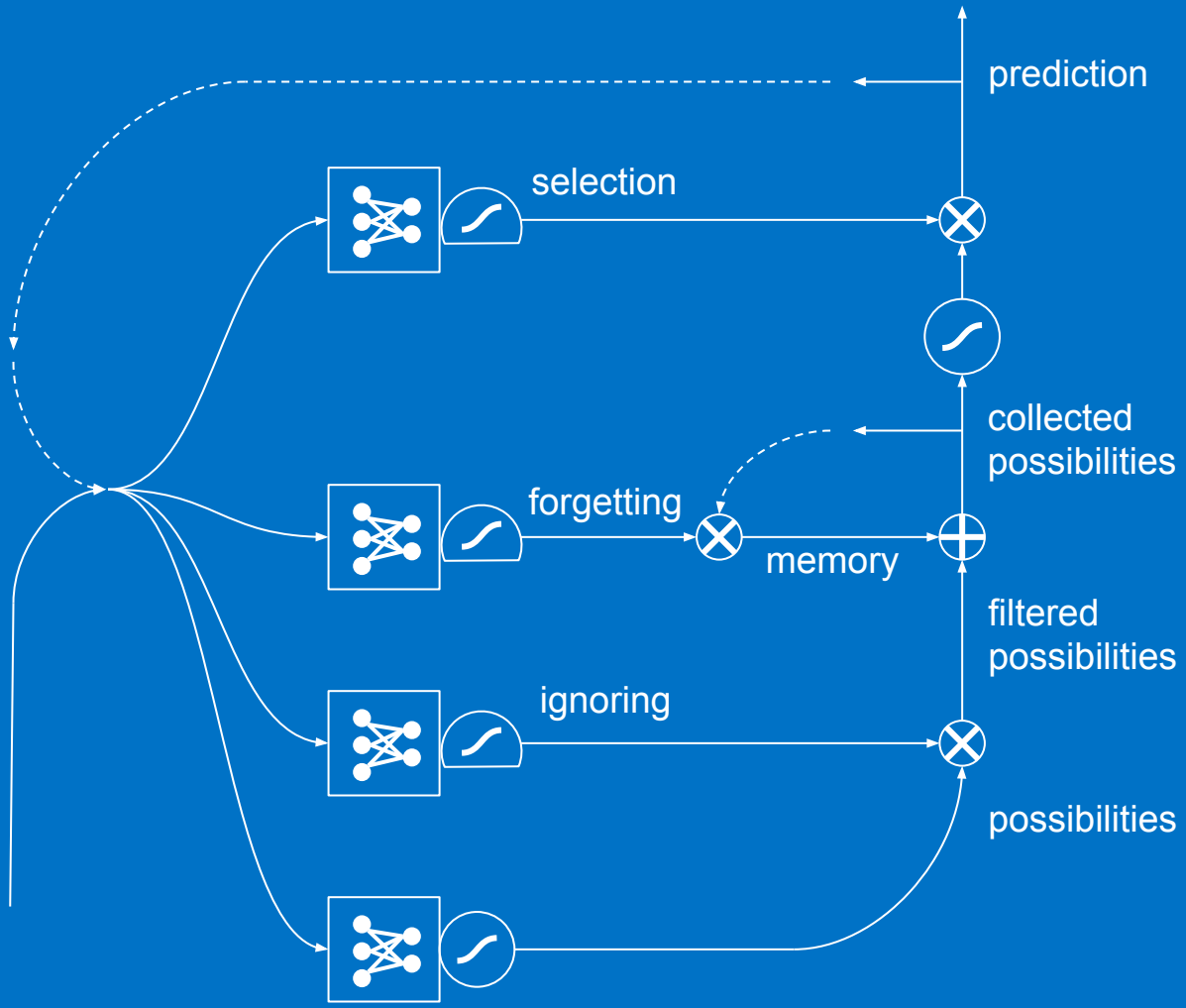


Constructing an LSTM

Ignoring – Keeping or not keeping information in *your current prediction*

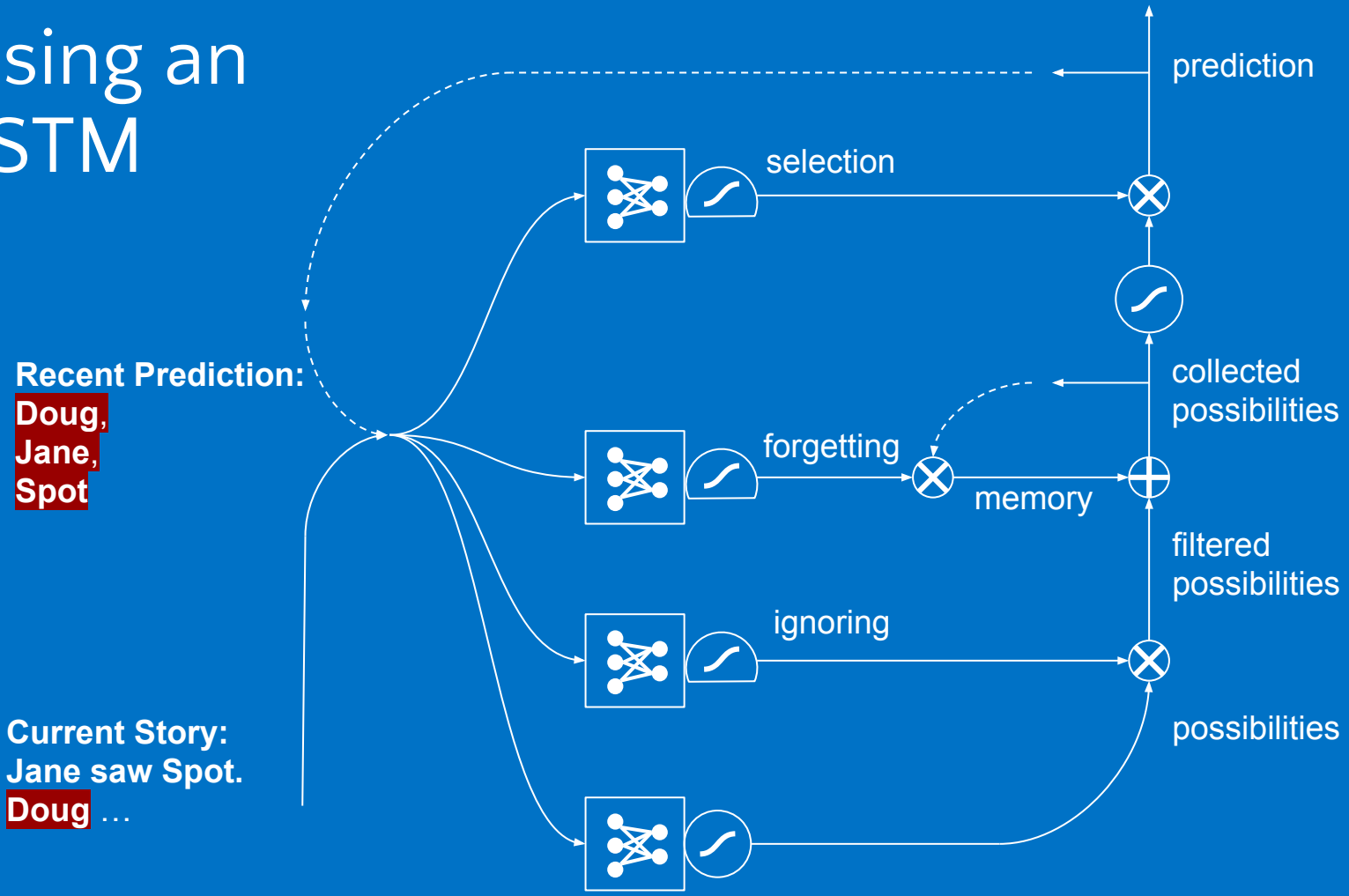


Using an LSTM

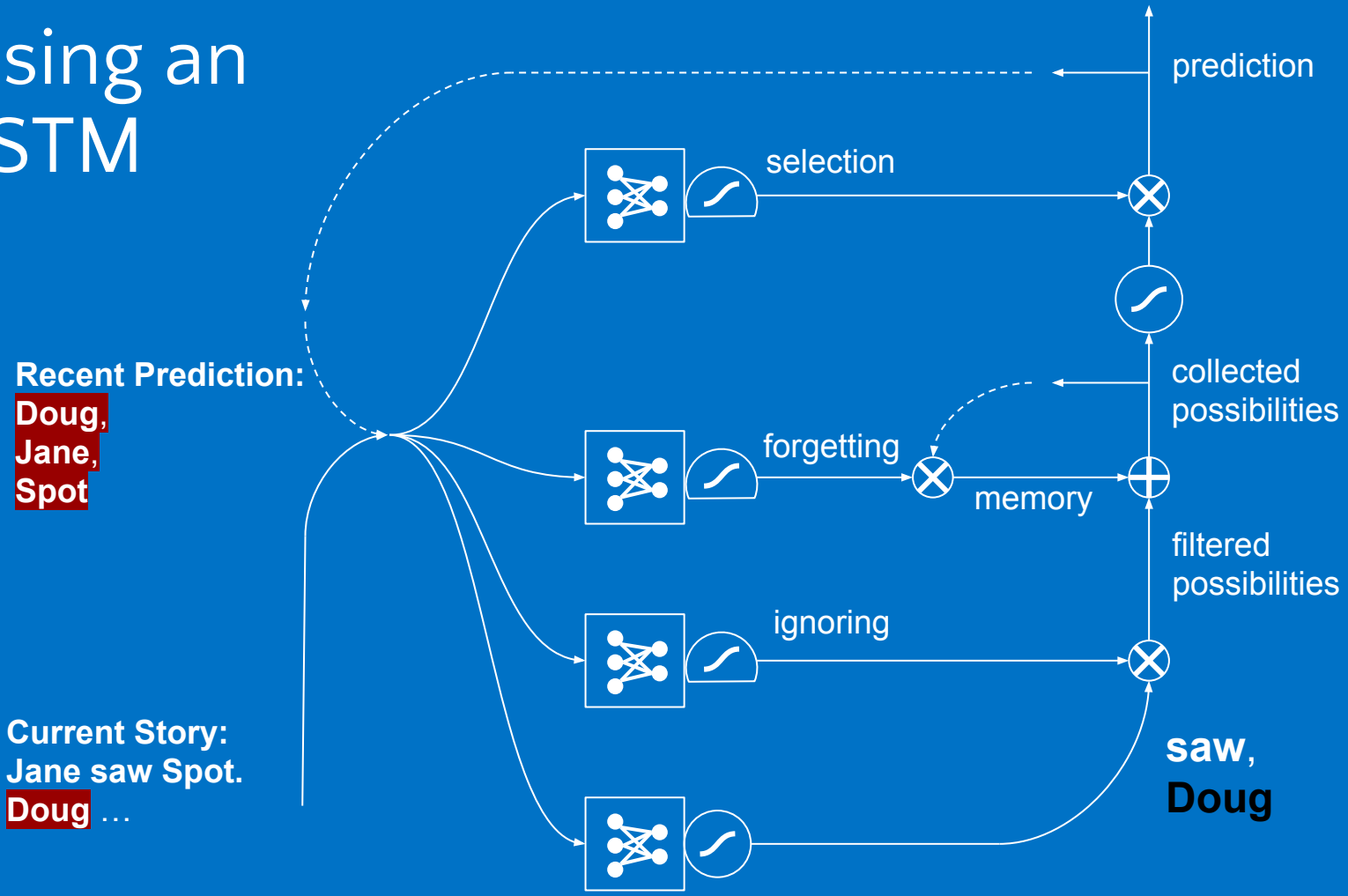


Current Story:
Jane saw Spot.
Doug …

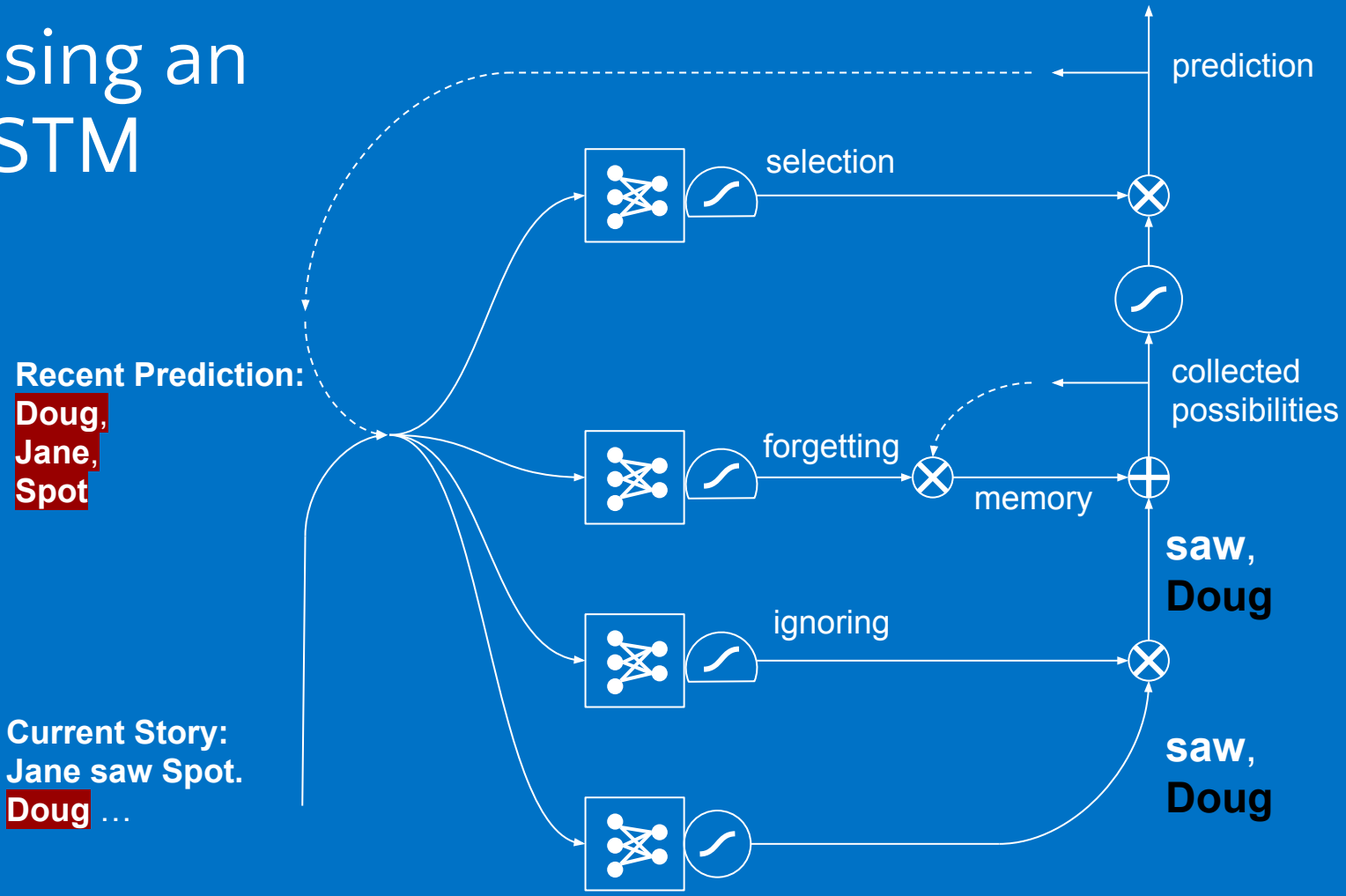
Using an LSTM



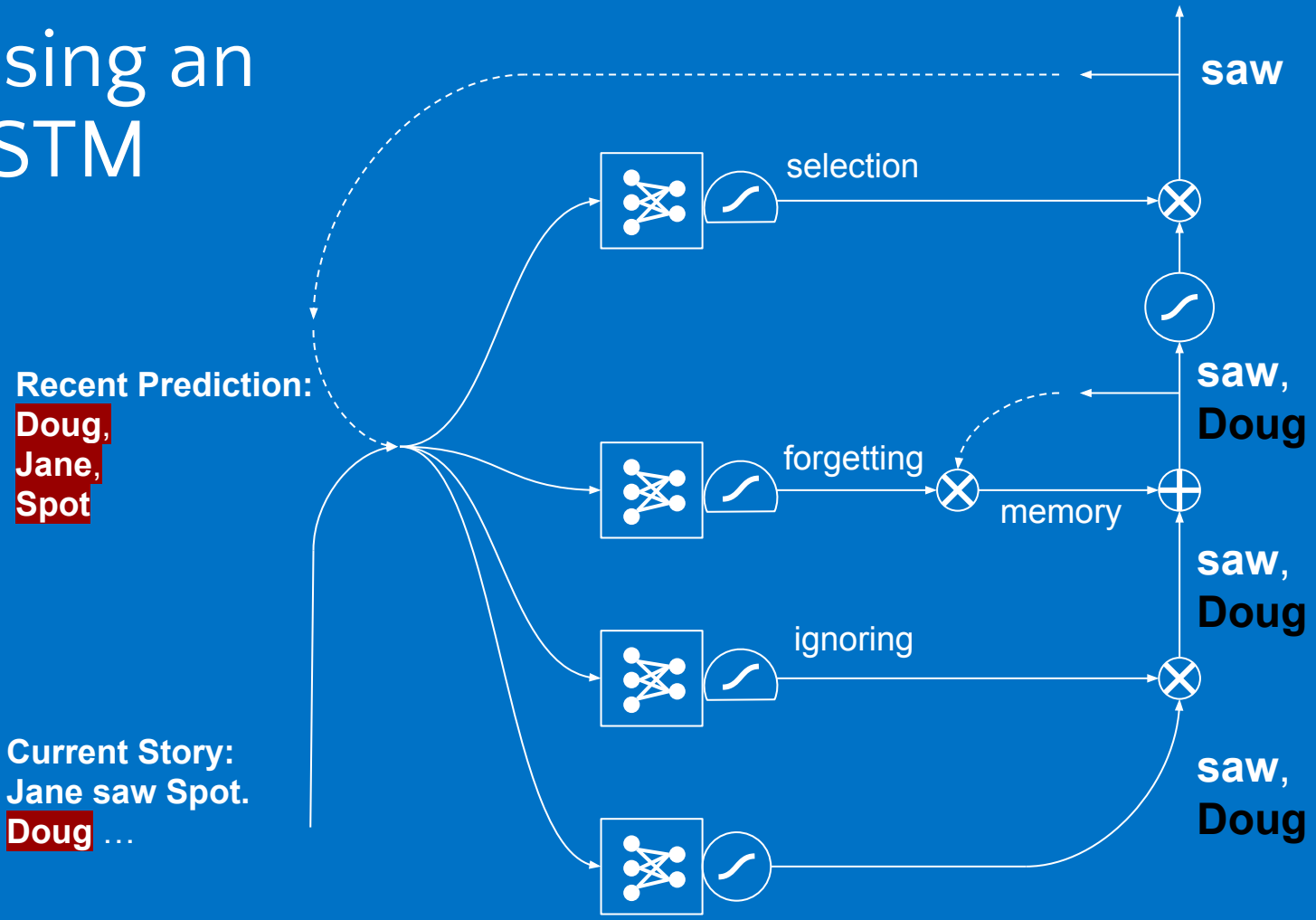
Using an LSTM



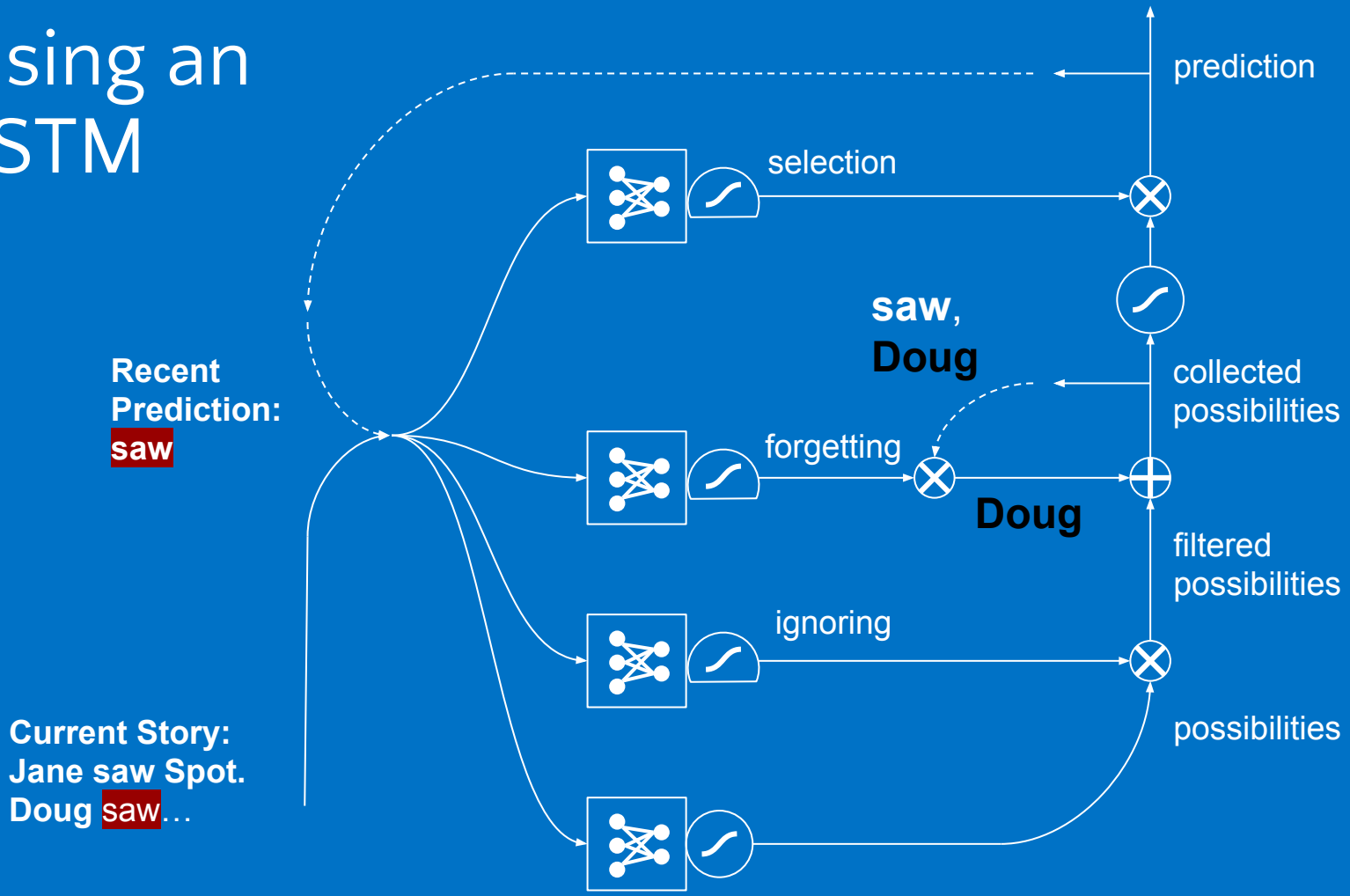
Using an LSTM



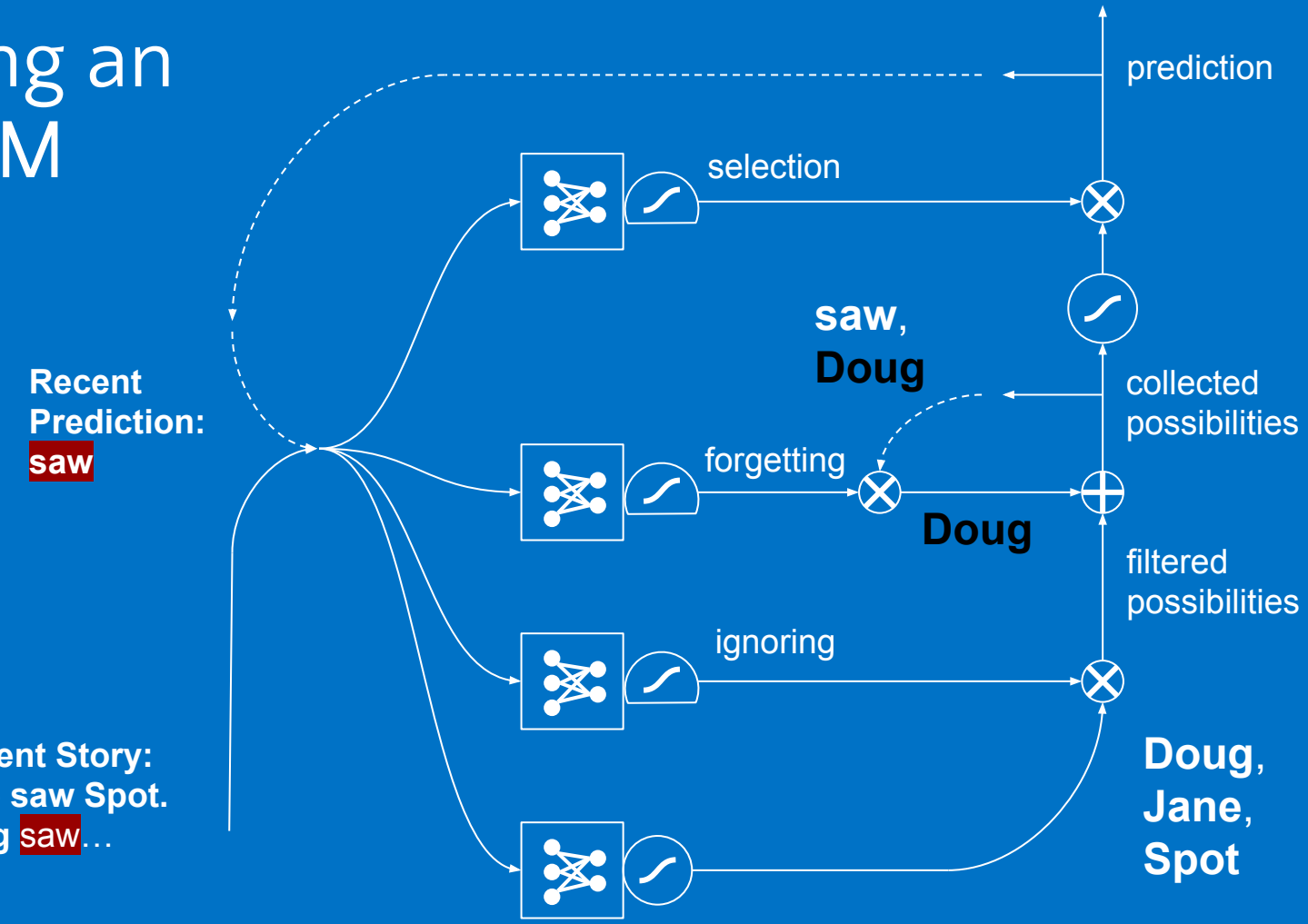
Using an LSTM



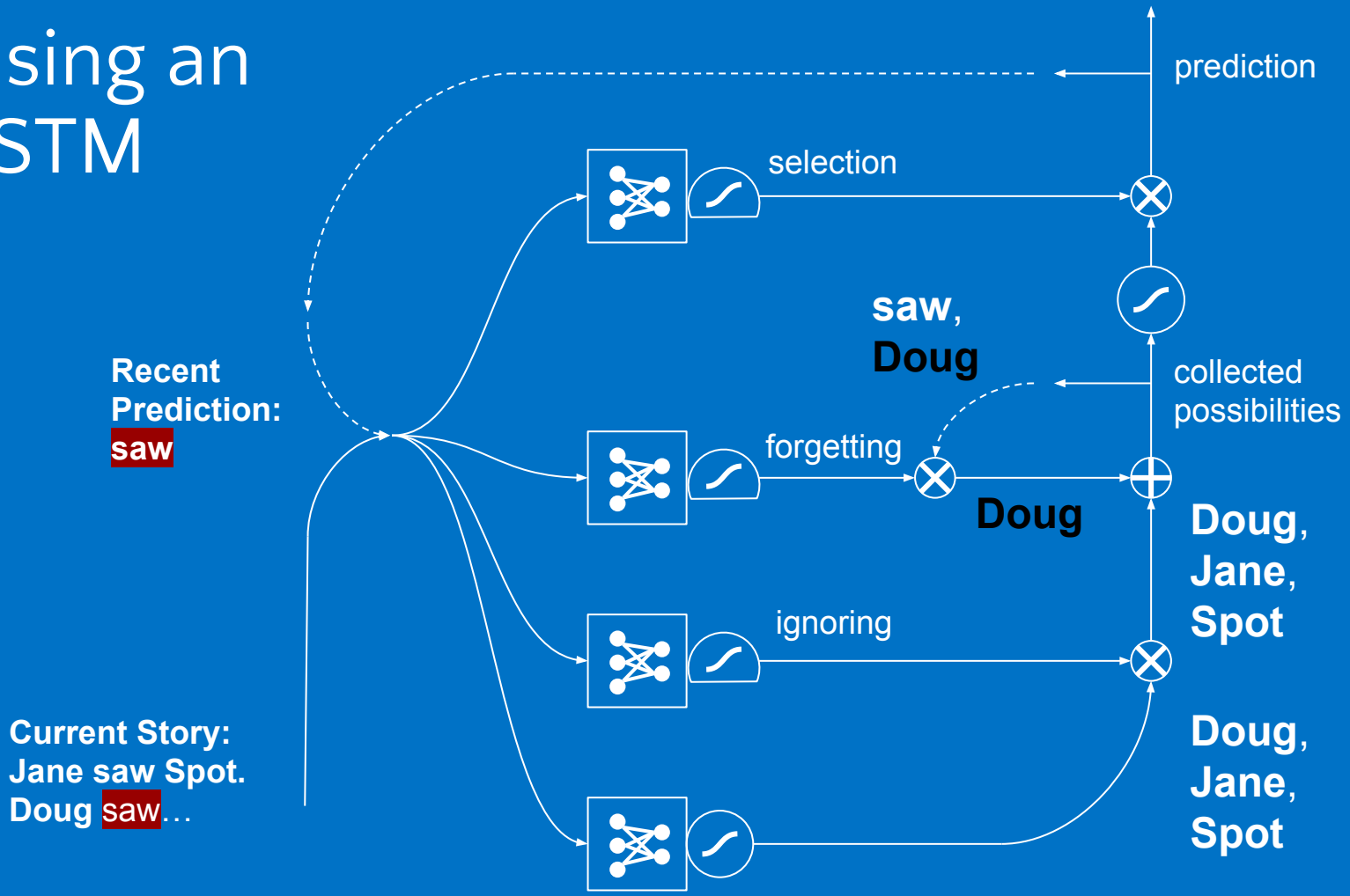
Using an LSTM



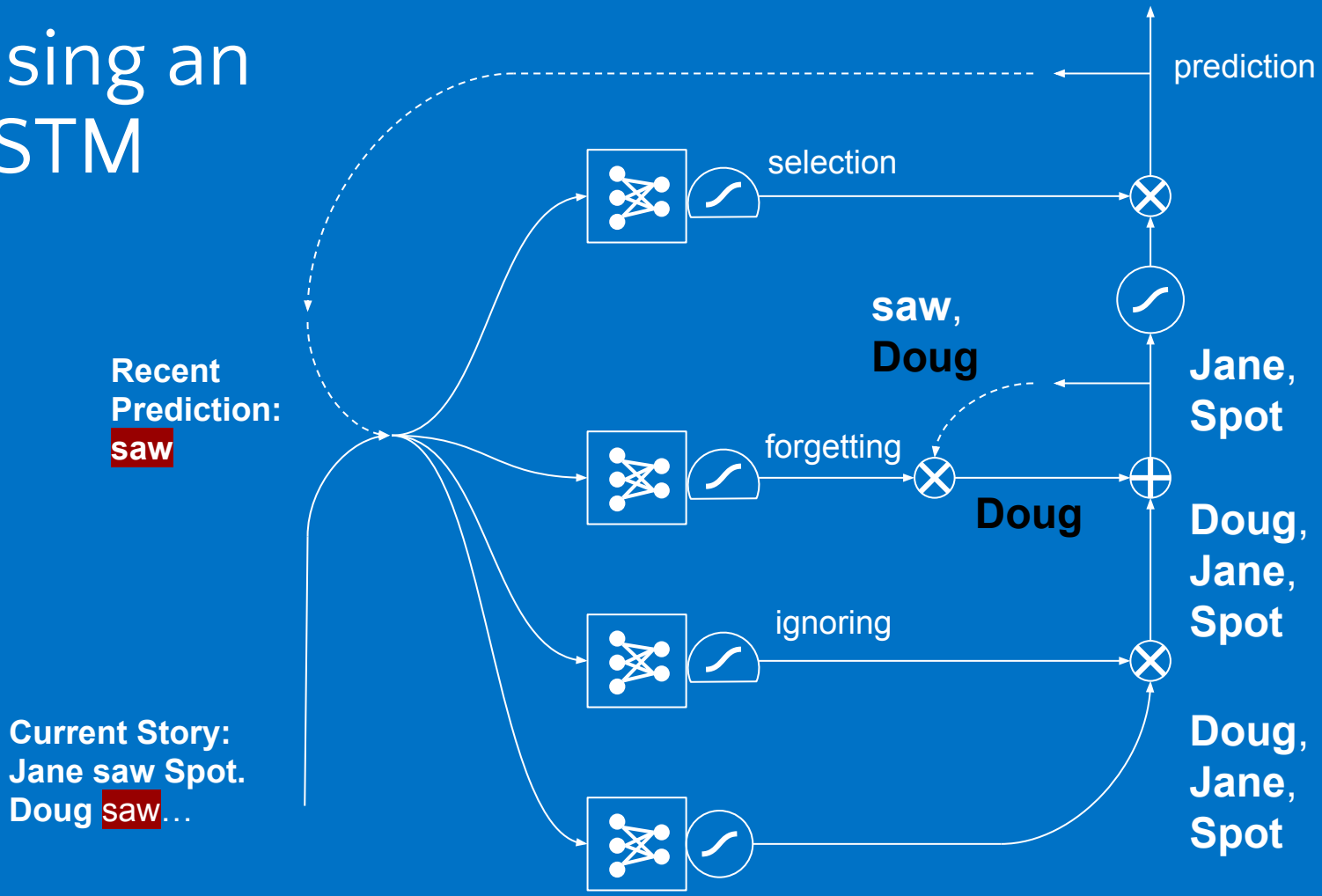
Using an LSTM



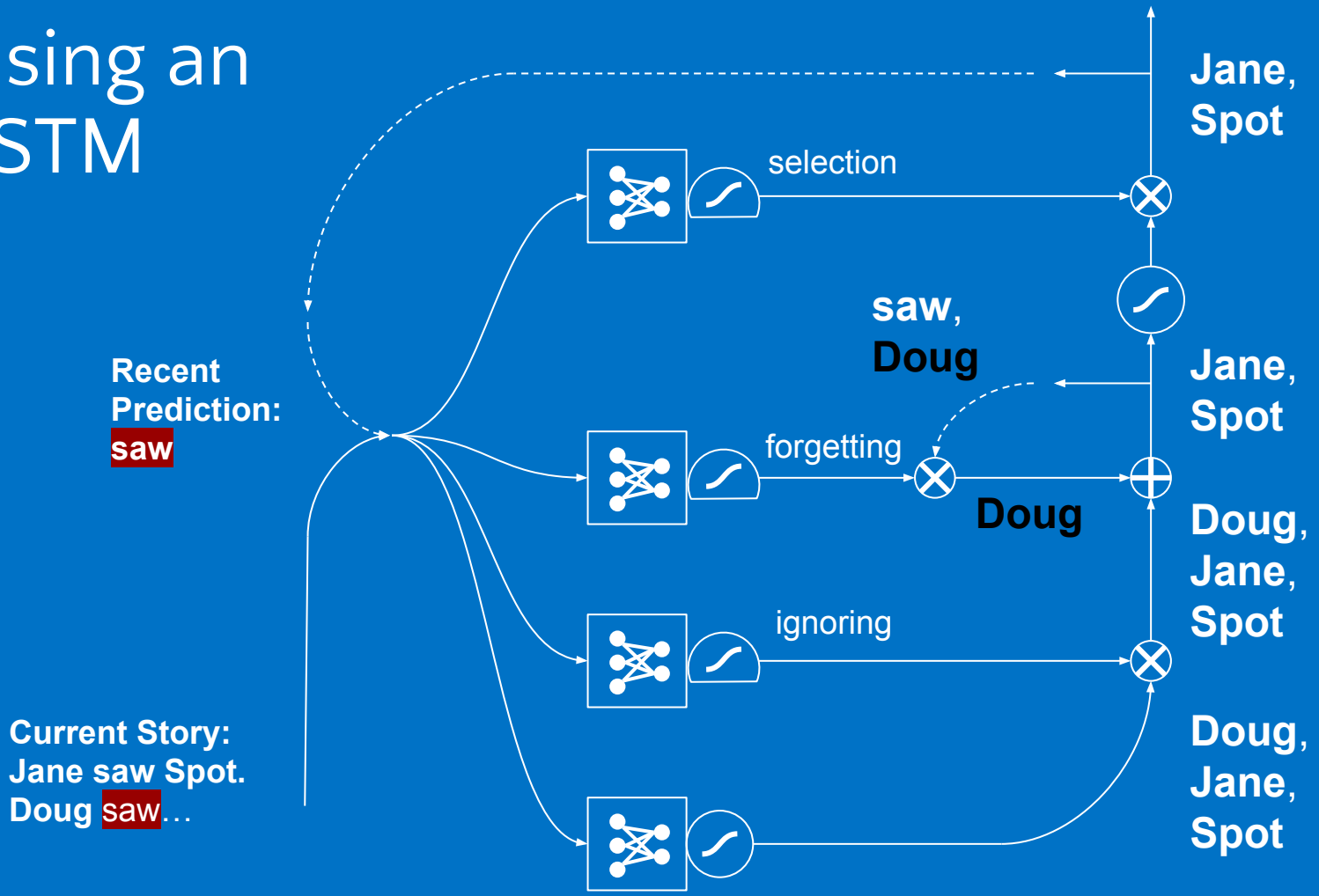
Using an LSTM



Using an LSTM

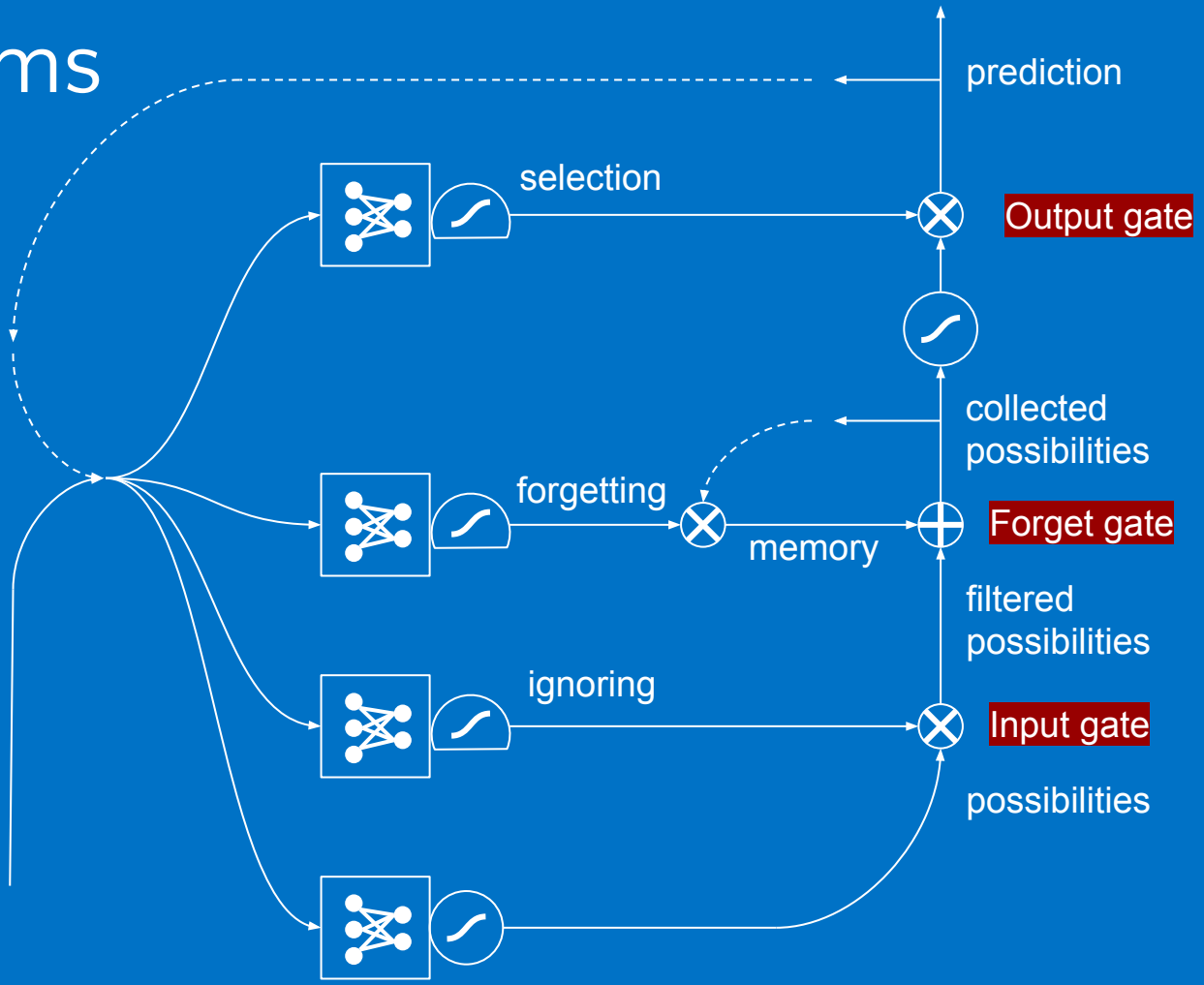


Using an LSTM



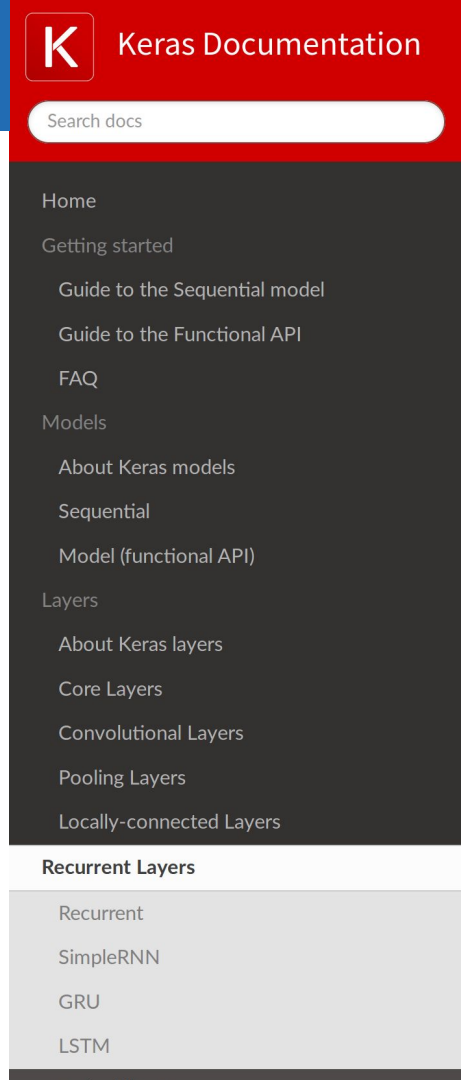
LSTM Terms

- Input gate - how much of the newly computed ideas you want to let through
- Forget gate - how much of the previous information you want to let through
- Output gate - how much of the internal information you want to move forward to the next steps



Implementing RNNs & LSTMs using Keras

- Keras is a high-level neural networks API, written in Python and capable of running on top of either TensorFlow, CNTK or Theano.
- We use it in the DSI for capstone projects. MLPs, CNNs, RNNs, some Reinforcement Learning too.
- Will become TensorFlow's default API.
- Available recurrent layers:
  - Recurrent
  - SimpleRNN
  - Long Short-Term Memory (LSTM)
  - Gated Recurrent Unit (GRU)



Check-In Questions

- What are some applications of RNNs & LSTMs?
- Describe the structure of an RNN.
- What is the problem with RNNs that LSTMs improve upon?
- What is gating & why is it used?
- What are three types of gates in LSTMs?

Additional Resources

- Chris Olah's tutorial
- Andrej Karpathy's
  - Blog post
  - RNN code
  - Stanford CS231n lecture
  - NN course
  - Github
- DeepLearning 4J tutorial - helpful discussion & list of good resources
- How Neural Networks Work [video]
- Ian Goodfellow's Deep Learning book
- Iamtrask's blog
- Jakob Aungiers's blog & Github