EE7207 Lecture 8

Modern Recurrent Neural Networks

About me

Just call me Nick!

Nick LUO Wuqiong

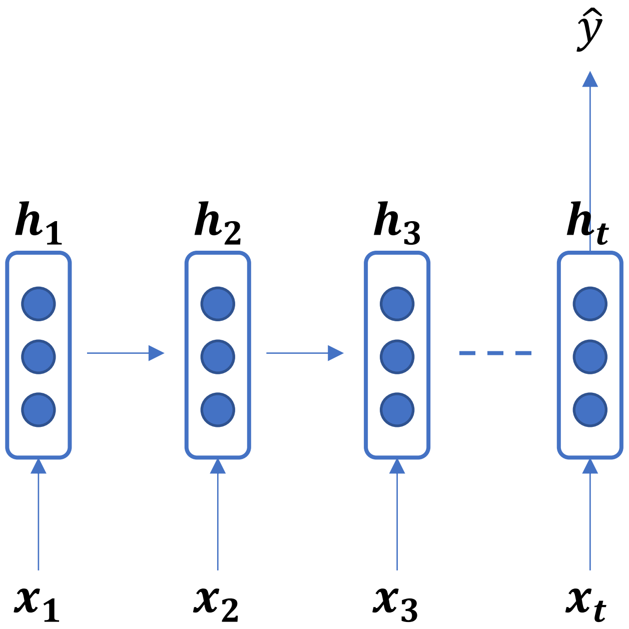
Vice President
Data Science Lead
OCBC AI Lab

Examples of sequence data in applications

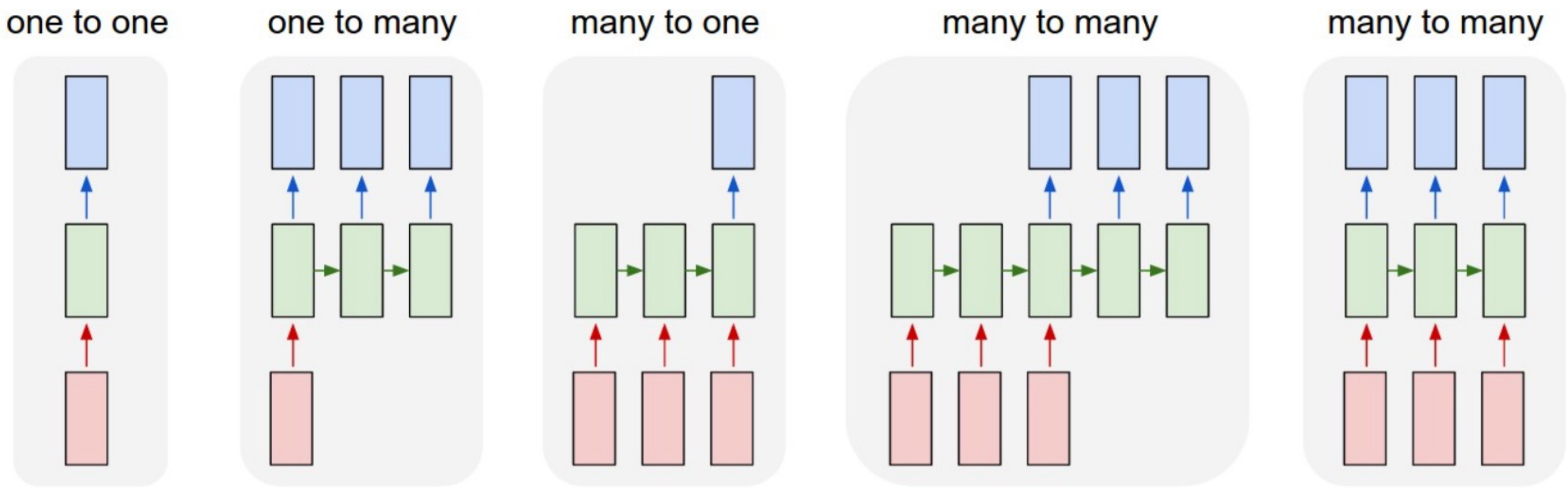
Language Model	Speech Recognition	Machine Translation	Stock Prediction
Sequence to one	Sequence to sequence	Sequence to sequence	Sequence to one
X: text sequence Y: next word	X: wave sequence Y: text sequence	X: text sequence (in one language) Y: text sequence (in another language)	X: sequence of market data Y: next day/year price/direction

- Most machine learning models can only handle structured data in a tabular form
- It’s difficult to deal with unstructured sequence data
- Earlier attempt of converting unstructured sequence data into structured form:
    - Bag-of-words: the text sequence is represented as the bag of its words, discarding the word order

Recurrent Neural Network

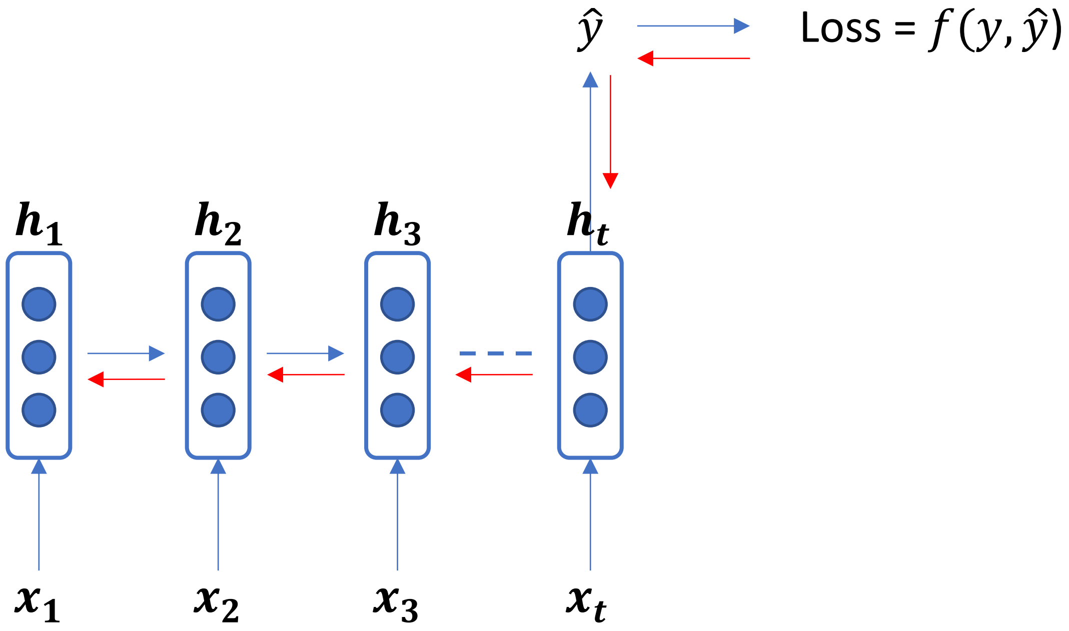


many to one example



https://karpathy.github.io/2015/05/21/rnn-effectiveness/

Backpropagation through time



Vanishing gradients and exploding gradient problem

The chain rule: $\sigma'(\boldsymbol{h_t}) \times \sigma'(\boldsymbol{h_{t-1}}) \times \cdots \times \sigma'(\boldsymbol{h_1})$

The value becomes very large if each of them is greater than 1: exploding gradients problem

- Gradient clipping: cap the gradient at a predefined value

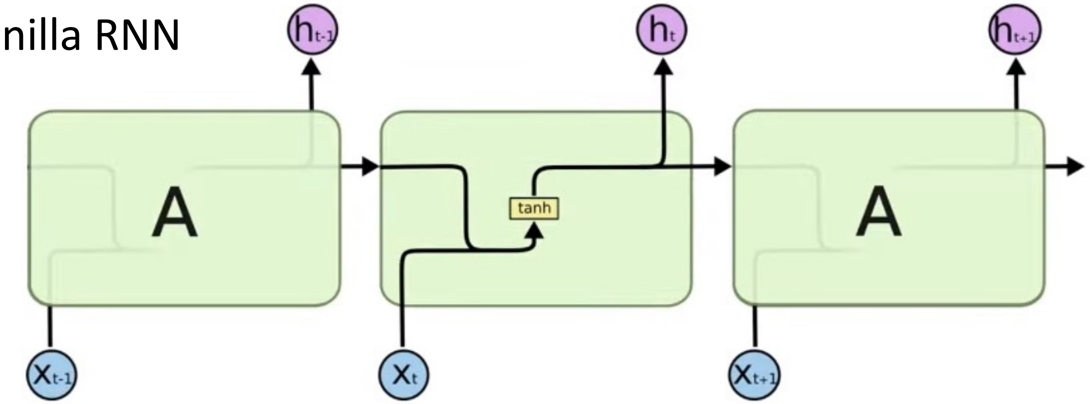
The value becomes 0 fast if each of them is less than 1: vanishing gradients problem

- No easy way to handle this for vanilla RNN, we'll be introducing LSTM and GRU that can (partially) address this issue

Vanilla RNN is not good at capturing long-term dependencies.

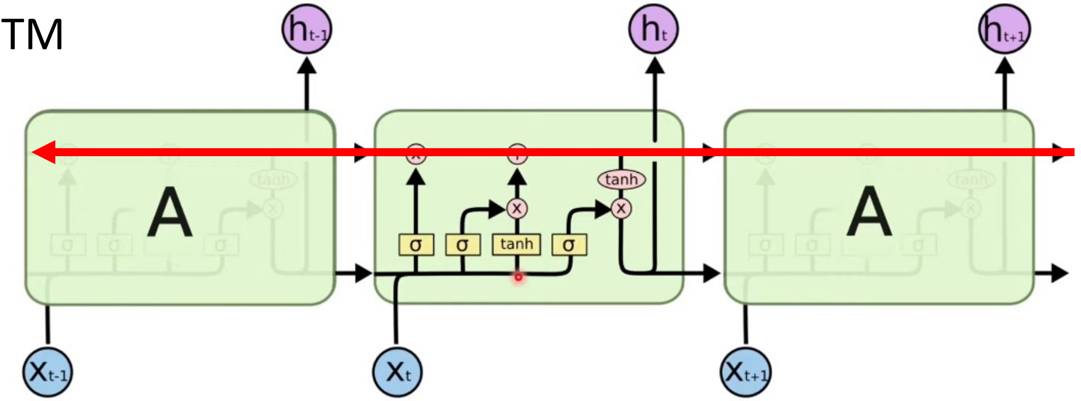
Long Short-Term Memory (LSTM) Networks

Vanilla RNN



- LSTM has gates to optionally let information through
- LSTM can decide how much old information to forget and how much new information to remember

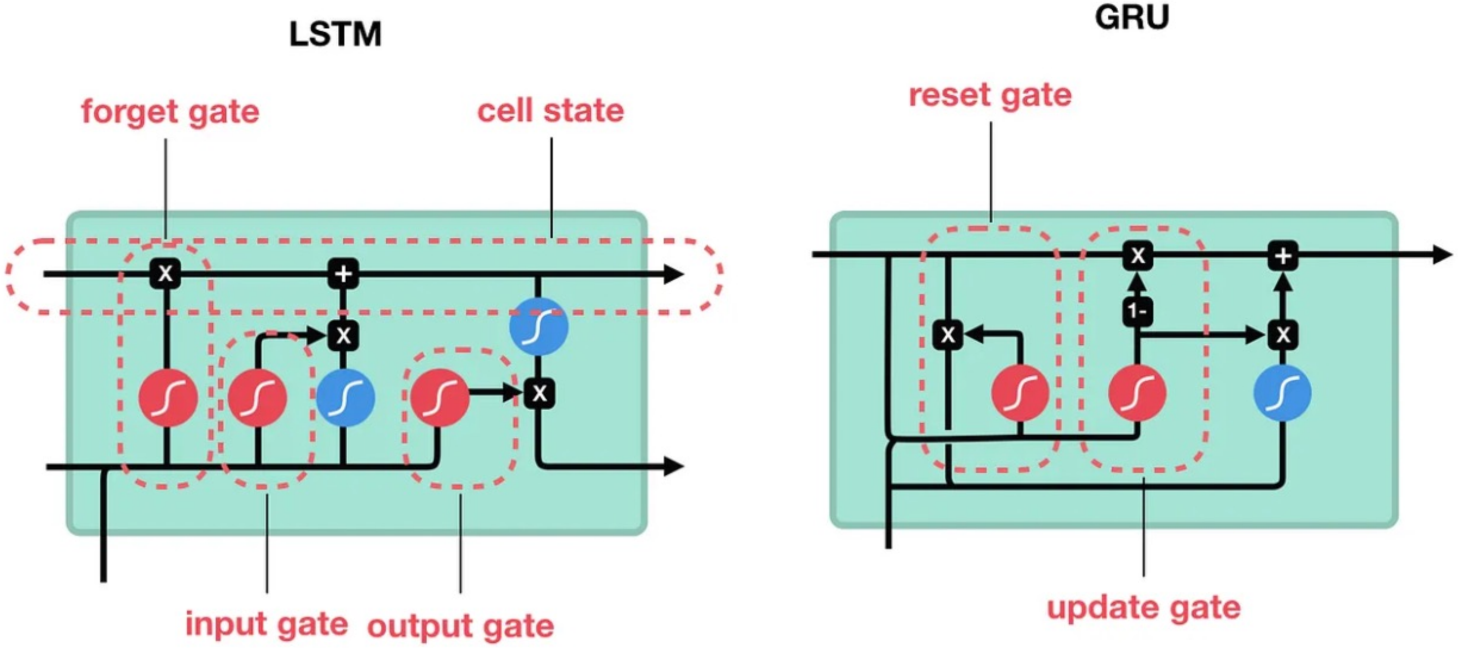
LSTM



- A highway for gradients to pass through
- Similar to ResNet for computer vision

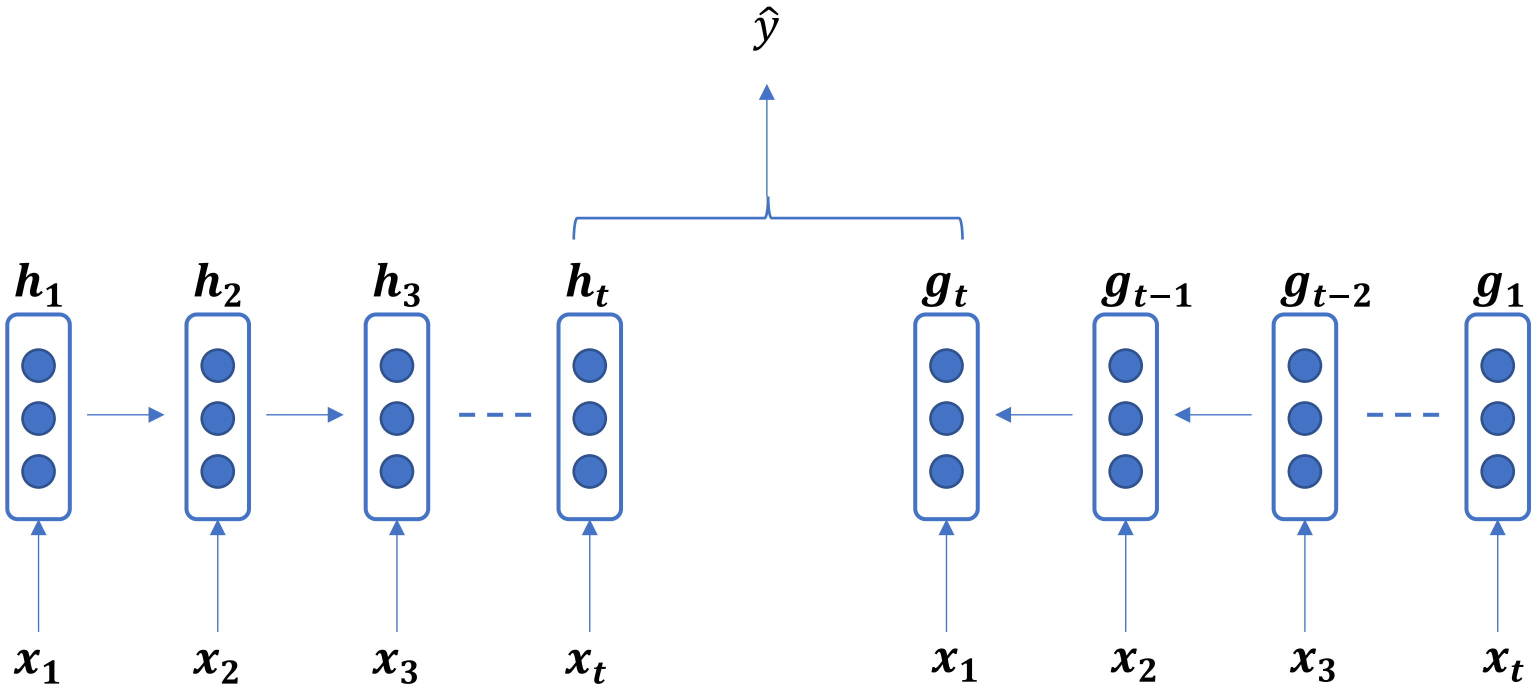
Gated Recurrent Units (GRUs)

- GRU is simpler than LSTM, and can be used to build much bigger networks
- LSTM is more general and powerful
- Both LSTM and GRU employs **Gating Mechanism** to address the issue of long term dependencies

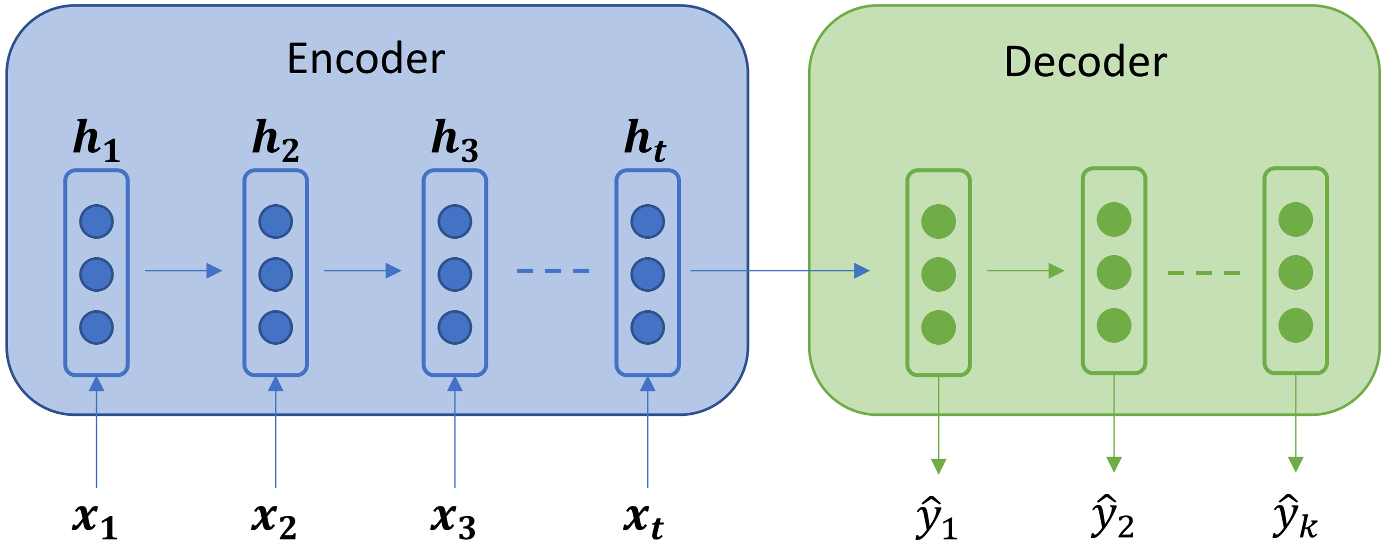


https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

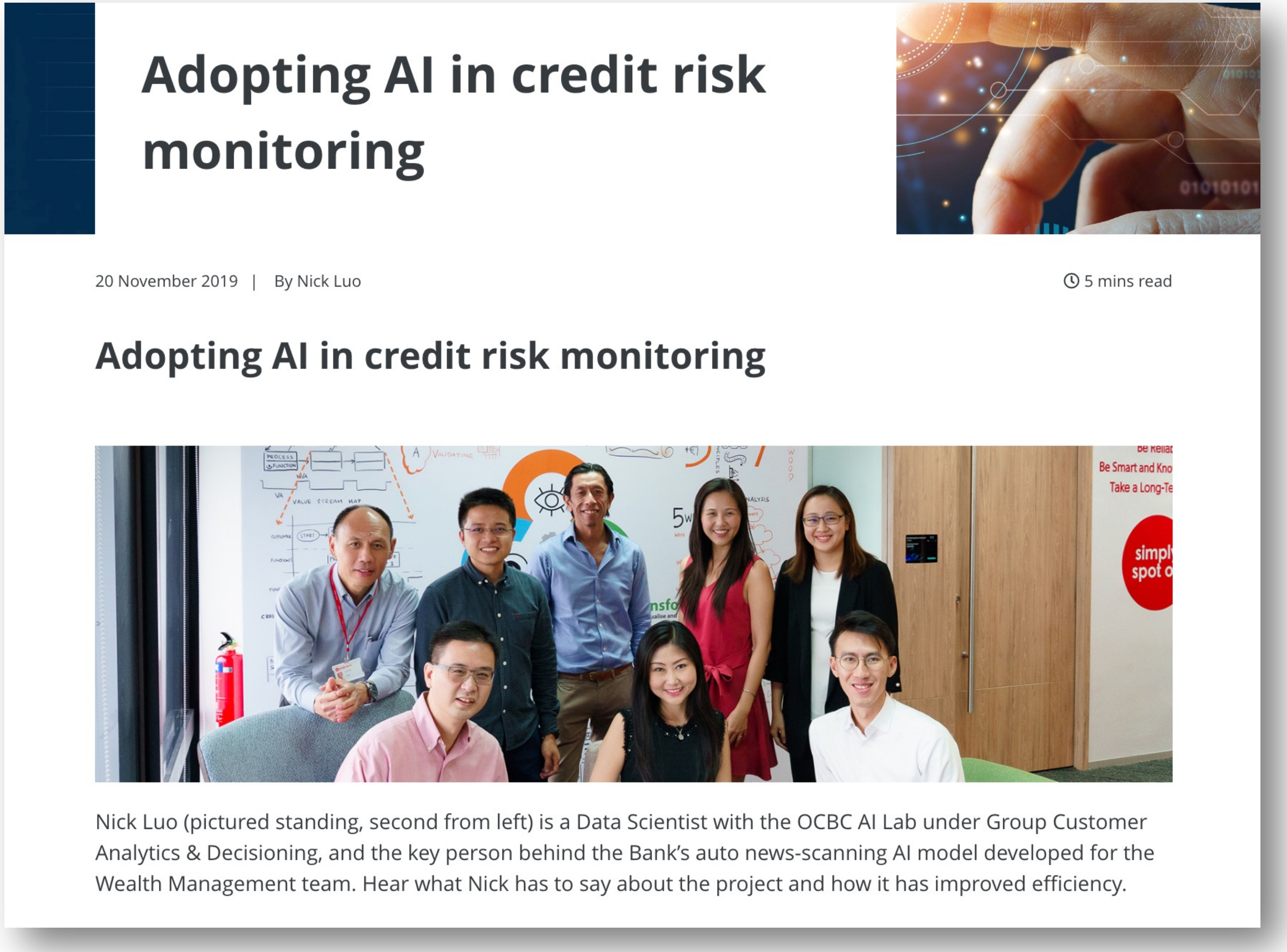
Bidirectional Recurrent Neural Networks (Bi-RNNs)



Encoder-Decoder Architecture



Real-world case study: sentiment classification on external news



Adopting AI in credit risk monitoring

20 November 2019 | By Nick Luo

5 mins read

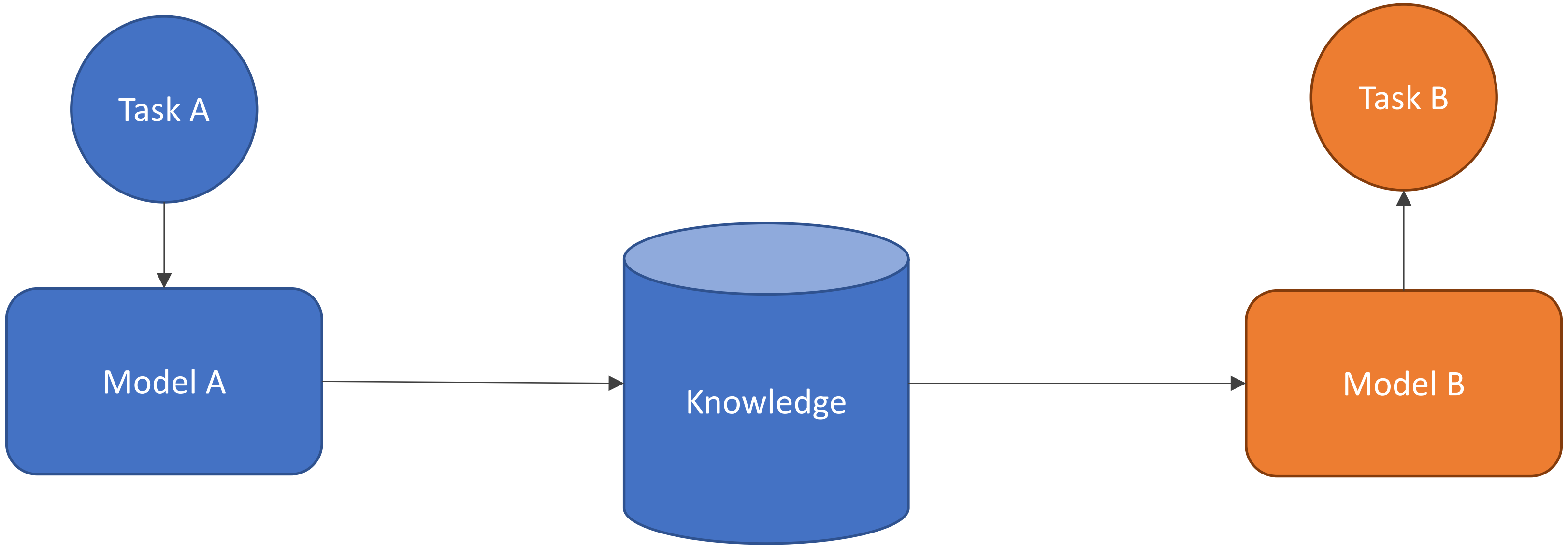
Adopting AI in credit risk monitoring

Nick Luo (pictured standing, second from left) is a Data Scientist with the OCBC AI Lab under Group Customer Analytics & Decisioning, and the key person behind the Bank's auto news-scanning AI model developed for the Wealth Management team. Hear what Nick has to say about the project and how it has improved efficiency.

https://www.ocbc.com/group/careers/blog/adopting-ai-in-credit-risk-monitoring.page

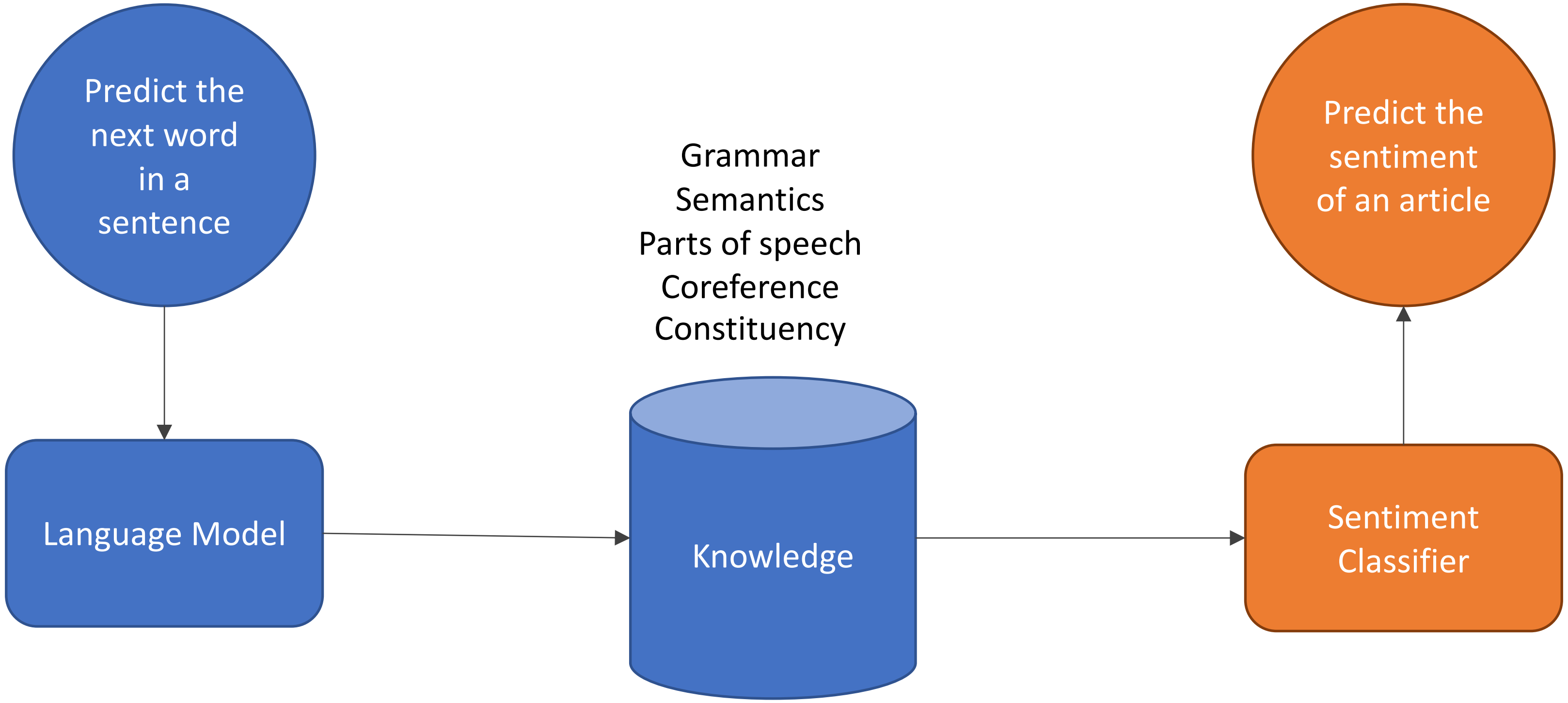
Finetune Language Models for Sentiment Analysis

- Huge amount of labelled data is needed to train a big neural network from scratch
- Transfer learning can significantly reduce the amount of labelled data
- Transfer learning refers to the use of a model that has been trained to solve one problem as the starting point to solve another related problem



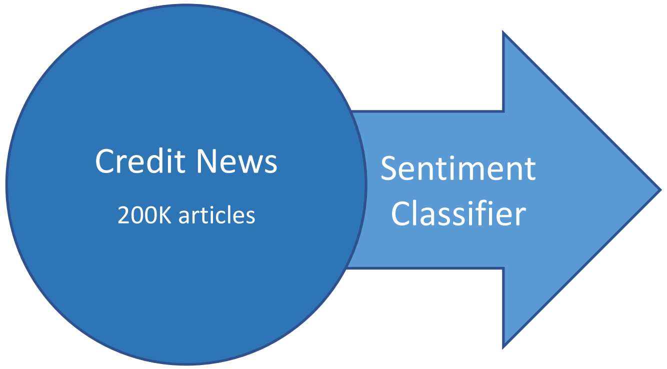
Finetune Language Models for Sentiment Analysis

- Use a trained language model as the starting point to build a sentiment classifier



Transfer learning helps reduce the amount of labelled data needed

- Without transfer learning



Credit News

200K articles

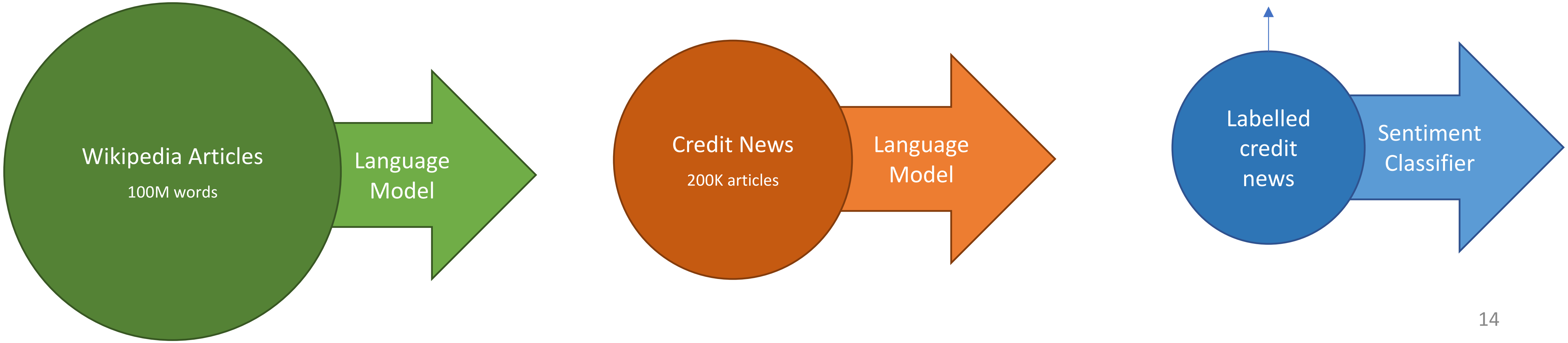
Sentiment Classifier

- Need to label 200K articles
- 400 articles per week
- 500 weeks ≈ 10 years!

- With transfer learning

5 weeks

To label 2K articles



Wikipedia Articles

100M words

Language Model

Credit News

200K articles

Language Model

Labelled credit news

Sentiment Classifier

Assignment: Sentiment Classification Model for Movie Reviews