Introduction to Deep Learning (CS474)

Lecture 21

Rupam

Outline

- **Module 3**
  - **Introduction to Recurrent Neural Network (RNN)**

Rupam

Introduction to Recurrent Neural Network (RNN)

- Recurrent neural networks or RNNs are a family of neural networks for processing sequential data.

- Much as a convolutional network is a neural network that is specialized for processing a grid of values X such as an image, a recurrent neural network is a neural network that is specialized for processing a sequence of values $x^{(1)}, \ldots, x^{(\tau)}$.

- Just as convolutional networks can readily scale to images with large width and height, and recurrent networks can scale to much longer sequences.

Slide credit: I. Goodfellow, Y. Bengio, A. Courville

Rupam

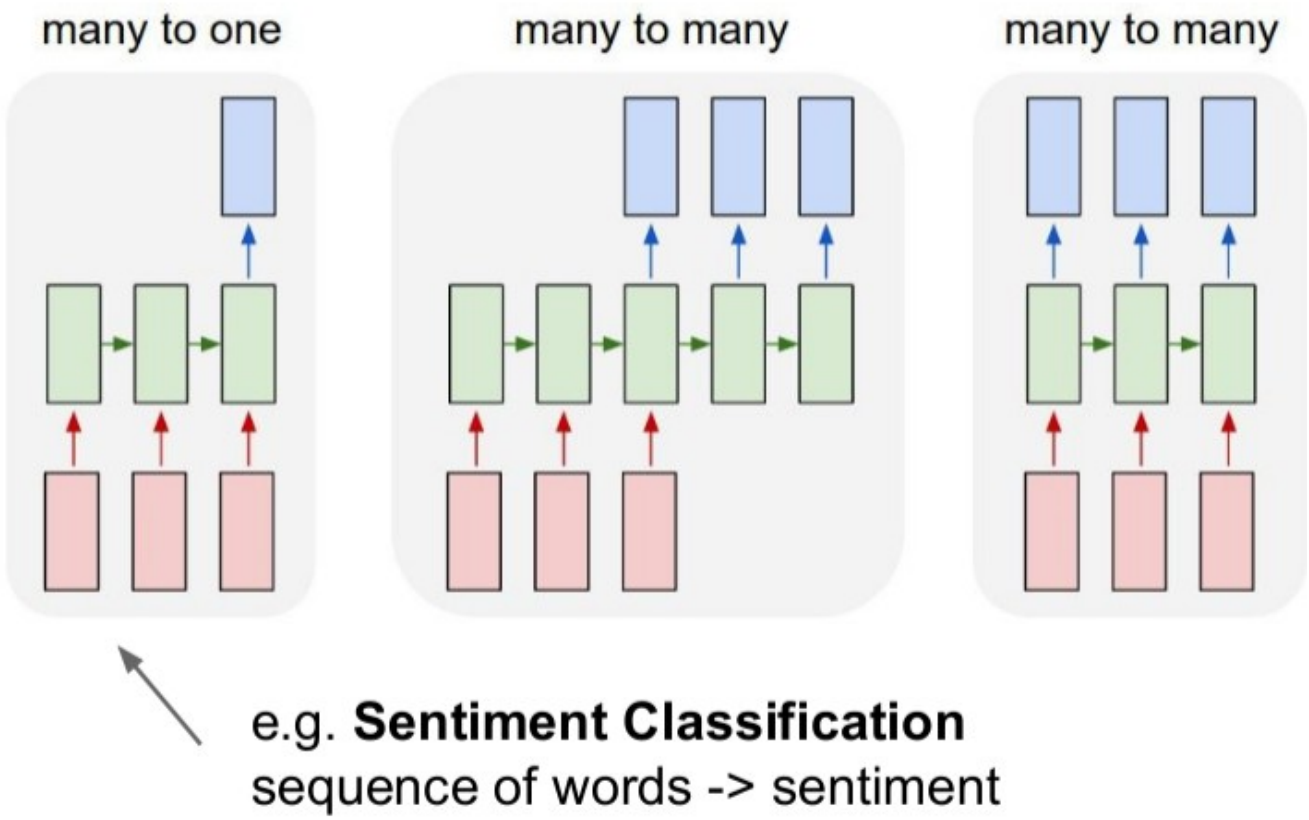
Examples of Sequence Data

Sentiment classification	"There is nothing to like in this movie."	→	★☆☆☆☆
DNA sequence analysis	AGCCCCTGTGAGGAACTAG	→	AGCCCCTGTGAGGAACTAG
Machine translation	Voulez-vous chanter avec moi?	→	Do you want to sing with me?
Video activity recognition		→	Running
Name entity recognition	Yesterday, Harry Potter met Hermione Granger.	→	Yesterday, Harry Potter met Hermione Granger.

Slide credit: Prof. A. Ng

Rupam

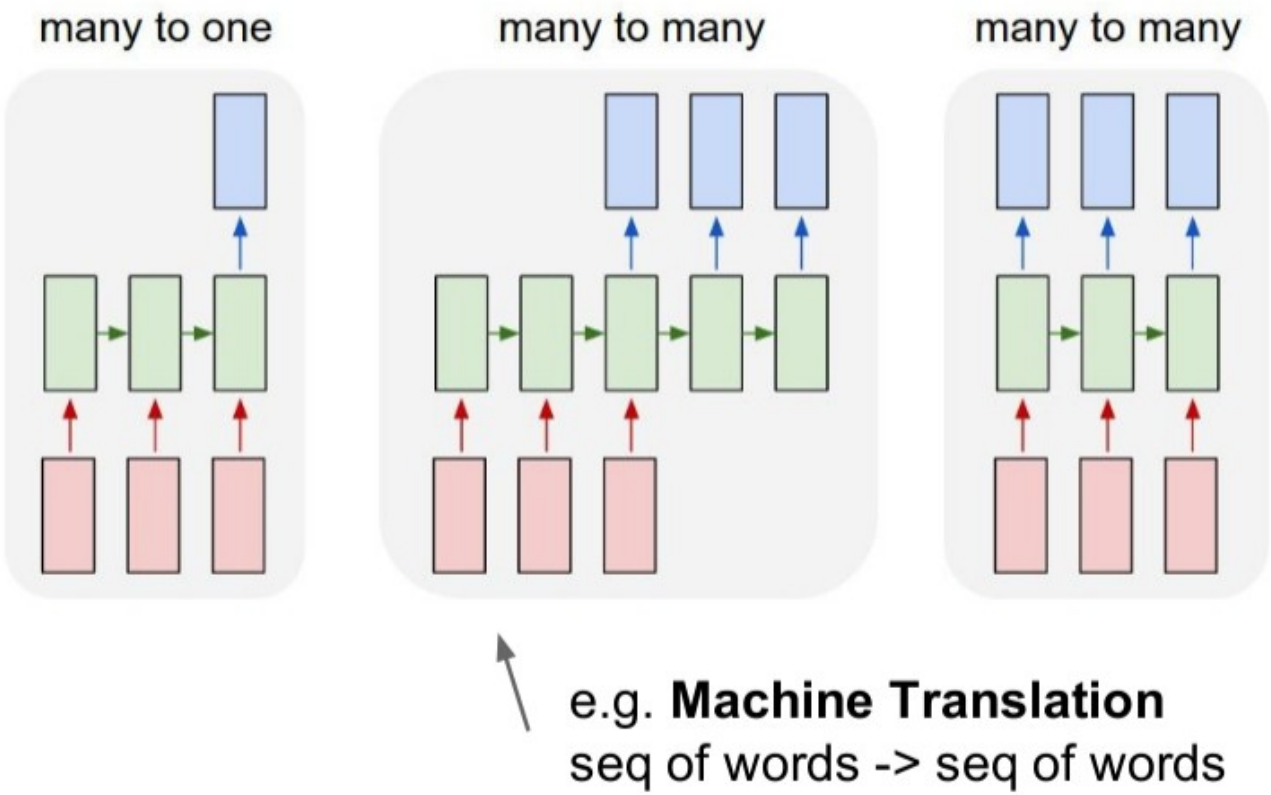
Process Sequences



Slide credit: F. Li & J. Johnson & S. Yeung

Rupam

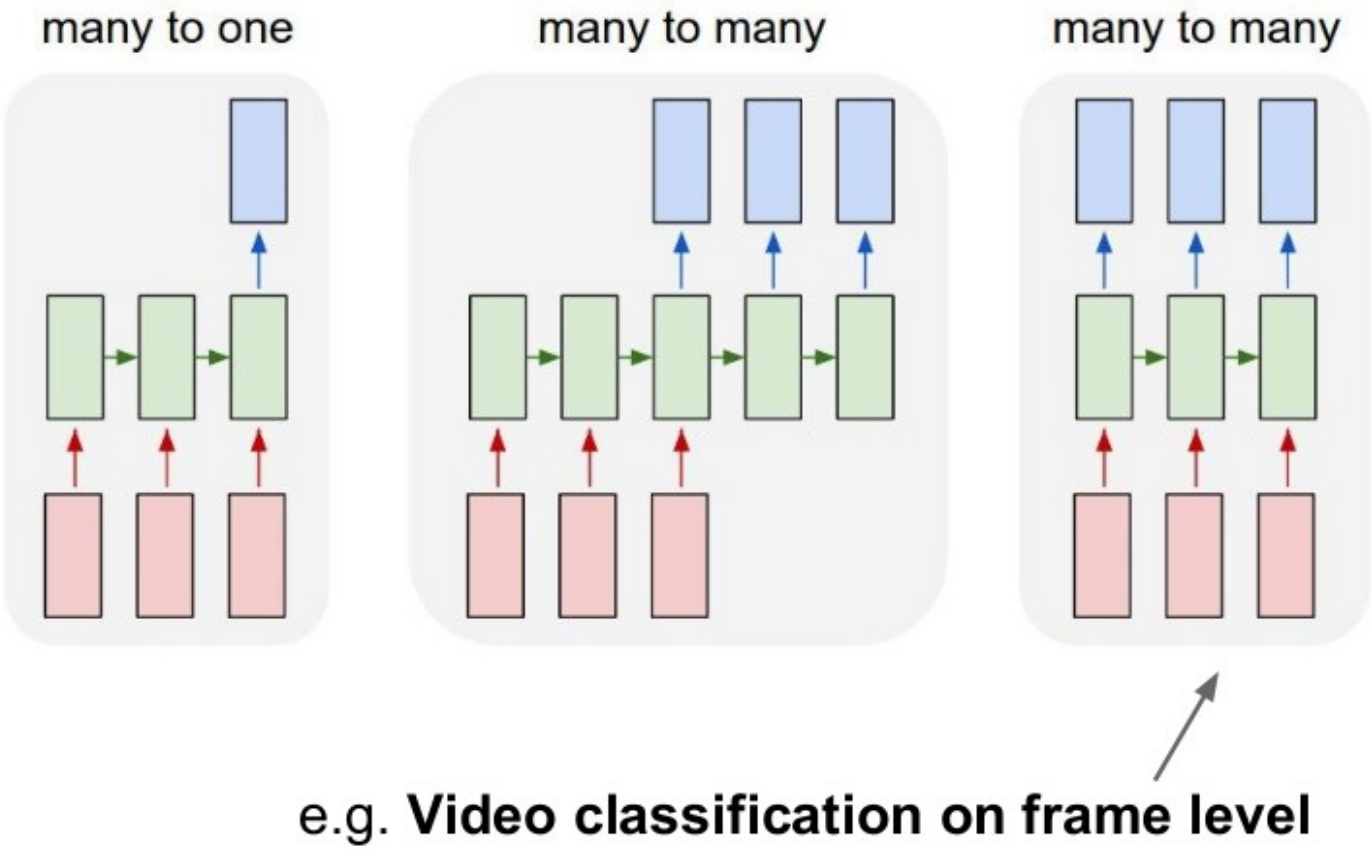
Process Sequences



Slide credit: F. Li & J. Johnson & S. Yeung

Rupam

Process Sequences



Slide credit: F. Li & J. Johnson & S. Yeung

Rupam

Example

x: (Harry Potter) and (Hermione Granger) invented a new spell.

$x^{\langle 1\rangle}$ $x^{\langle 2\rangle}$ $x^{\langle 3\rangle}$ .... $x^{\langle t\rangle}$ .... $x^{\langle 9\rangle}$

$T_x = 9$

→ y: 1 1 0 1 1 0 0 0 0

$y^{\langle 1\rangle}$ $y^{\langle 2\rangle}$ $y^{\langle 3\rangle}$ .... $y^{\langle 9\rangle}$

$T_y = 9$

$X^{(i)\langle t\rangle}$ $T_x^{(i)}$

$y^{(i)\langle t\rangle}$ $T_y^{(i)}$

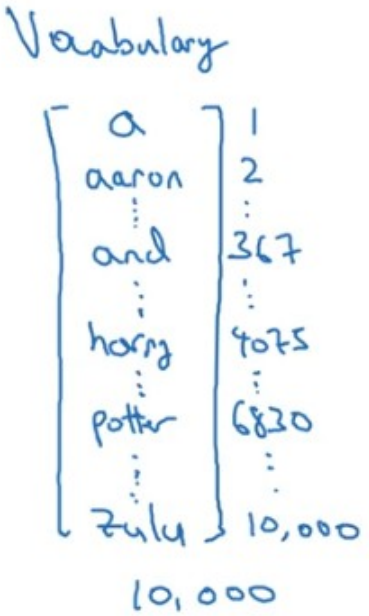
Slide credit: Prof. A. Ng

Rupam

Example

x: Harry Potter and Hermione Granger invented a new spell.

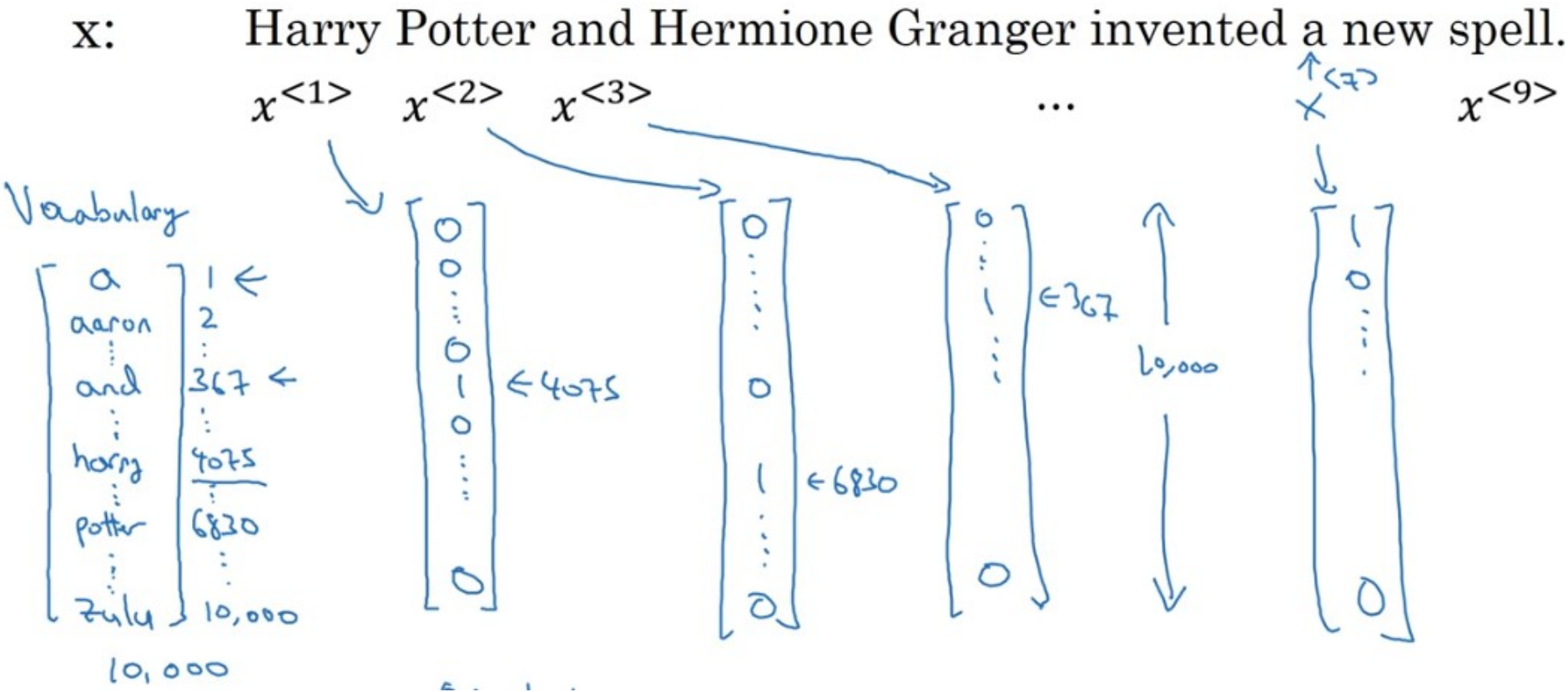
$x^{<1>}$ $x^{<2>}$ $x^{<3>}$ ... $x^{<9>}$



Slide credit: Prof. A. Ng

Rupam

Example



Slide credit: Prof. A. Ng

Rupam

Problem with the Standard Network

- For example, consider the two sentences “I went to Nepal in 2009” and “In 2009, I went to Nepal.”
- If we ask a machine learning model to read each sentence and extract the year in which the narrator went to Nepal, we would like it to recognize the year 2009 as the relevant piece of information, whether it appears in the sixth word or the second word of the sentence.
- Suppose that we trained a feedforward network that processes sentences of fixed length. A traditional fully connected feedforward network would have separate parameters for each input feature.
- Parameter sharing makes it possible to extend and apply the model to examples of different forms (different lengths, here) and generalize across them.

Slide credit:I. Goodfellow, Y. Bengio, A. Courville

References

- All the contents present in the slides are taken from various online resources. Due credit is given in the respective slides.These slides are used for *academic* purposes only.