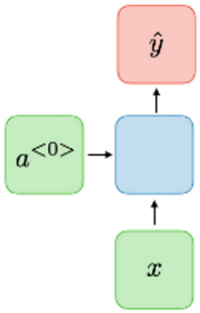
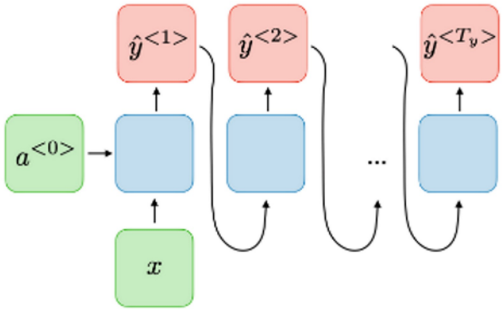
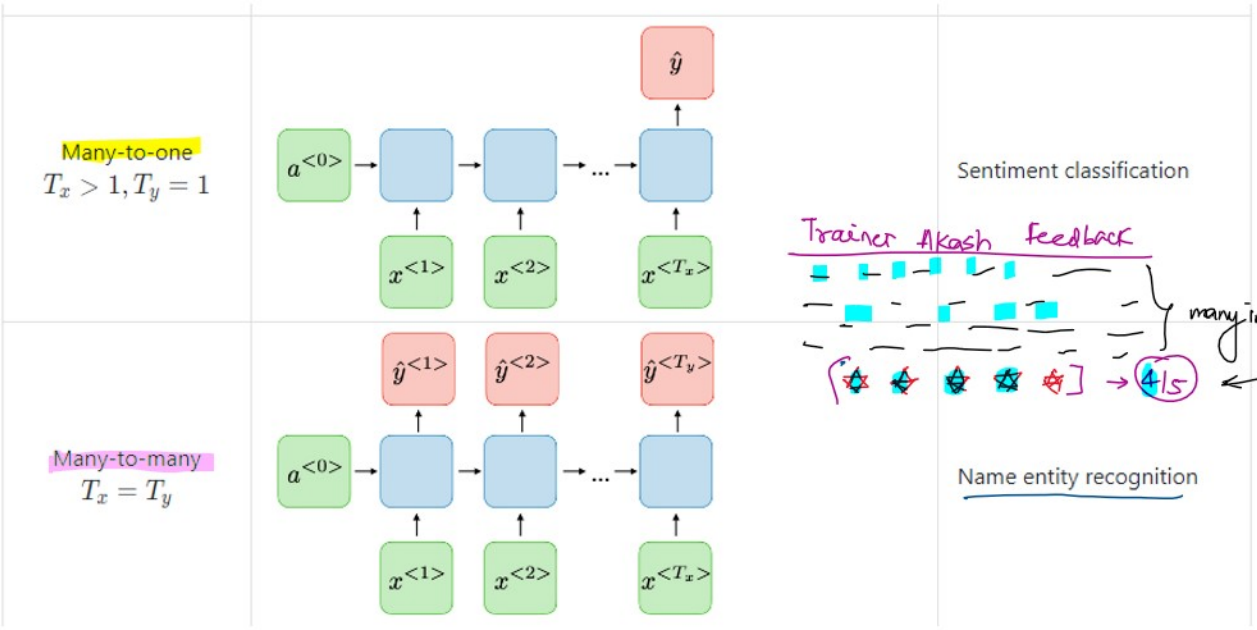
Recurrent Connection

27 October 2024 21:48

Type of RNN	Illustration	Example
One-to-one $T_x = T_y = 1$		Traditional (recurrent) neural network
One-to-many $T_x = 1, T_y > 1$		Music generation Once upon → 5 words input not "C" → a sequence

A RNN is given an initial note or theme as input and then it generates a full sequence of musical notes based on that input.

Example: MuseNet and Magenta use neural networks to expand on a small piece of input note and create longer compositions



Named Entity Recognition

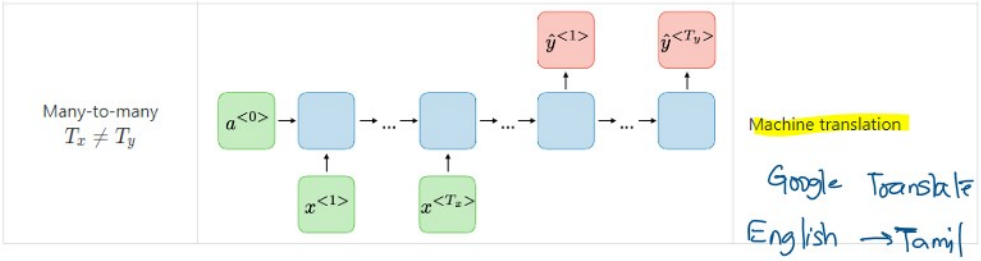
- a common task in NLP (Natural Language Processing) that involves identifying and classifying entities in texts into predefined such as names of persons, organizations, locations, dates and many more.

Input: Apple Inc. is looking at buying UK startup for $1 billion.

RNN

Apple Inc. (organization) is looking (actions) at buying UK (Location (country)) startup (organization) for $1 billion. (Money → currency)

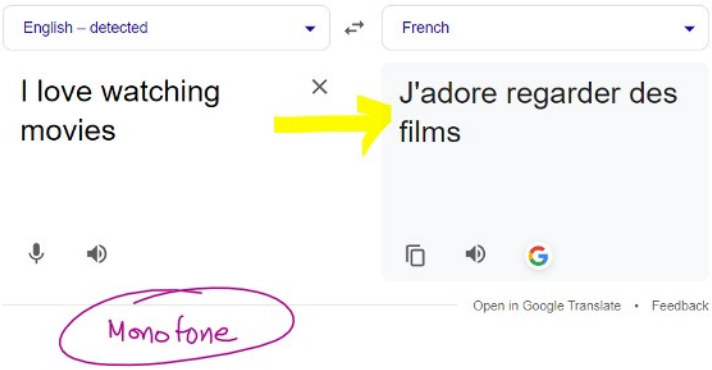
Person (e.g., Steve Jobs)
Organization (e.g., Apple Inc.)
Location (e.g., Cupertino)
Date (e.g., August 3, 2024)
Time (e.g., 10:00 AM)
Money (e.g., $500)
Percentage (e.g., 15%)
Product (e.g., iPhone)
Event (e.g., World Cup)
Facility (e.g., Golden Gate Bridge)



Google Translate
English → Tamil

Scenario: English to French Translation

sequence-to-sequence (seq2seq) model with attention processes an input sentence in English and generates an equivalent sentence in French. The encoder reads and understands the input sequence, the decoder generates the corresponding translation word-by-word.



Monofone

Limitations of RNNs

① Vanishing and Exploding Gradient Problems

Issue: During backpropagation, the gradients of loss w.r.t. to earlier layers become extremely small (vanishing) or large (exploding)

This issue is especially severe in long sequences.

Impact: When the gradients become very small, earlier layers receive very little information during training, making it difficult for network to learn from earlier time steps

2. Difficulty with long-term dependencies

Issue: RNNs struggle to capture long-term dependencies in data. As the length of the input sequence increases, RNNs tend to forget earlier information due to their short term memory capacity → Remedy is LSTM.

Impact: Tasks like language modelling or understanding meaning across longer texts become less effective with standard RNNs.

3. Lack of Parallelization:

Issue: Unlike CNNs, where operations on different pixels can be processed in parallel, RNNs process one time step at a time which prevents parallel processing

Impact: Longer training time and inefficient utilization of machines (hardware)