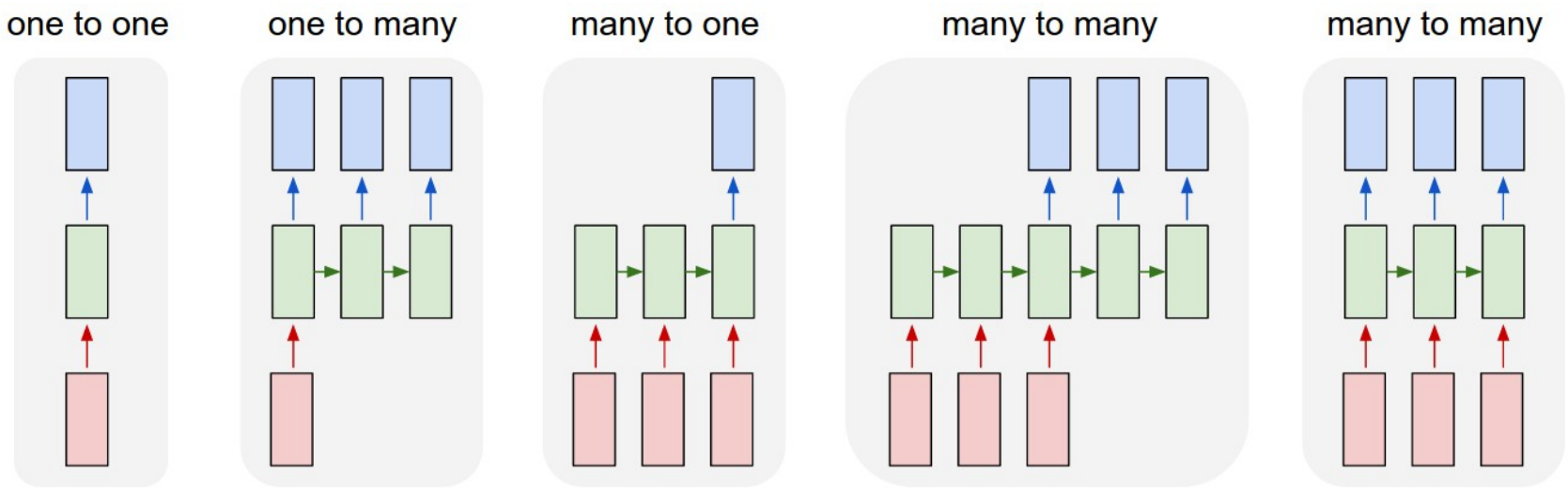
Recurrent networks

Tim Lawson

November 5, 2023

Working with sequences[1]



1. One-to-one: image classification
2. One-to-many: image captioning
3. Many-to-one: sentiment analysis
4. Many-to-many: machine translation
5. Many-to-many: video classification

[1] Karpathy 2015.

Recurrent vs. feedforward networks[2]

- Feedforward networks represent history by **context**, recurrent networks represent history by recurrent network **connections**.
- Feedforward networks have a **fixed** history length, recurrent networks have an **unlimited** history length.
- Feedforward networks compress single words, recurrent networks can compress history (**sequences** of words).
- Recurrent networks can form **short-term memory**.

[2] Mikolov 2010.

Vanishing gradients

"In theory, the time dependency allows [a recurrent network] in each iteration to know about every part of the sequence that came before. However, this time dependency typically causes a **vanishing gradient** problem that results in **long-term** dependencies being **ignored** during training."[3]

Long short-term memory (**LSTM**) networks[4] and gated recurrent units[5] (**GRUs**) are popular solutions to this problem.

[3] Madsen 2019; Pascanu, Mikolov, and Bengio 2013.
[4] Hochreiter and Schmidhuber 1997; Olah 2015.
[5] Cho et al. 2014.

Bibliography

Cho, Kyunghyun et al. (2014). *Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation*. URL: http://arxiv.org/abs/1406.1078.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long Short-Term Memory". In: *Neural Computation* 9.8, pp. 1735–1780. URL: https://direct.mit.edu/neco/article/9/8/1735-1780/6109.

Karpathy, Andrej (2015). *The Unreasonable Effectiveness of Recurrent Neural Networks*. URL: https://karpathy.github.io/2015/05/21/rnn-effectiveness/.

Madsen, Andreas (2019). "Visualizing Memorization in RNNs". In: *Distill* 4.3, e16. URL: https://distill.pub/2019/memorization-in-rnns.

Mikolov, Tomáš (2010). *Recurrent Neural Network Based Language Model*. URL: https://www.fit.vutbr.cz/research/groups/speech/servite/2010/rnnlm_mikolov.pdf.

Olah, Christopher (2015). *Understanding LSTM Networks – Colah's Blog*. URL: https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

Pascanu, Razvan, Tomáš Mikolov, and Yoshua Bengio (2013). "On the Difficulty of Training Recurrent Neural Networks". In: *Proceedings of the 30th International Conference on Machine Learning*. PMLR, pp. 1310–1318. URL: https://proceedings.mlr.press/v28/pascanu13.html.