WIKIPEDIA

Gated recurrent unit

Gated recurrent units (**GRU**s) are a gating mechanism in recurrent neural networks, introduced in 2014 by Kyunghyun Cho et al.[1] The GRU is like a long short-term memory (LSTM) with a forget gate,[2] but has fewer parameters than LSTM, as it lacks an output gate.[3] GRU's performance on certain tasks of polyphonic music modeling, speech signal modeling and natural language processing was found to be similar to that of LSTM.[4][5] GRUs have been shown to exhibit better performance on certain smaller and less frequent datasets.[6][7]

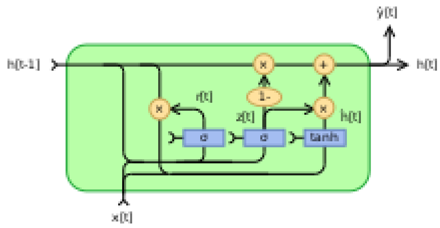
Contents

Architecture

There are several variations on the full gated unit, with gating done using the previous hidden state and the bias in various combinations, and a simplified form called minimal gated unit.[8]

The operator $\odot$ denotes the Hadamard product in the following.

Fully gated unit



Gated Recurrent Unit, fully gated version

Initially, for $t = 0$, the output vector is $h_0 = 0$.

$$\begin{aligned} z_t &= \sigma_g(W_z x_t + U_z h_{t-1} + b_z) \\ r_t &= \sigma_g(W_r x_t + U_r h_{t-1} + b_r) \\ \hat{h}_t &= \phi_h(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t \end{aligned}$$

Variables

- x_t: input vector
- h_t: output vector
- $\hat{h}_t$: candidate activation vector
- z_t: update gate vector
- r_t: reset gate vector
- W, U and b: parameter matrices and vector

Activation functions

- σ_g: The original is a sigmoid function.
- ϕ_h: The original is a hyperbolic tangent.

Alternative activation functions are possible, provided that $\sigma_g(x) \in [0, 1]$.

Alternate forms can be created by changing z_t and r_t[9]

- Type 1, each gate depends only on the previous hidden state and the bias.

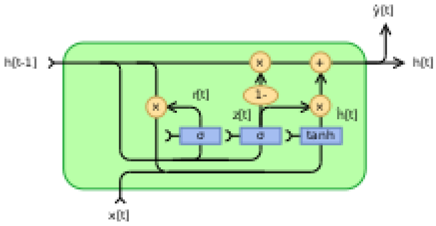
$$
\begin{aligned}
z_t &= \sigma_g(U_z h_{t-1} + b_z) \\
r_t &= \sigma_g(U_r h_{t-1} + b_r)
\end{aligned}
$$

- Type 2, each gate depends only on the previous hidden state.

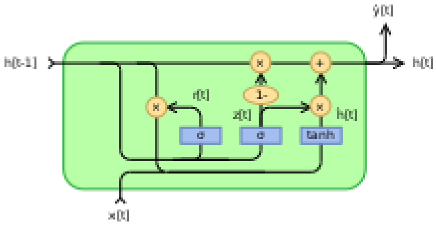
$$
\begin{aligned}
z_t &= \sigma_g(U_z h_{t-1}) \\
r_t &= \sigma_g(U_r h_{t-1})
\end{aligned}
$$

- Type 3, each gate is computed using only the bias.

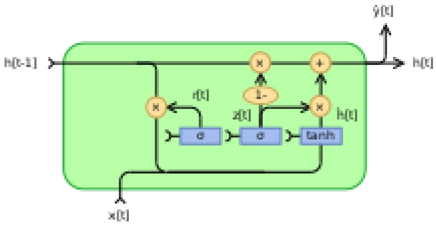
$$
\begin{aligned}
z_t &= \sigma_g(b_z) \\
r_t &= \sigma_g(b_r)
\end{aligned}
$$



Type 1



Type 2



Type 3

Minimal gated unit

The minimal gated unit is similar to the fully gated unit, except the update and reset gate vector is merged into a forget gate. This also implies that the equation for the output vector must be changed:[10]

$$
\begin{aligned}
f_t &= \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \\
\hat{h}_t &= \phi_h(W_h x_t + U_h(f_t \odot h_{t-1}) + b_h) \\
h_t &= (1 - f_t) \odot h_{t-1} + f_t \odot \hat{h}_t
\end{aligned}
$$

Variables

- x_t: input vector
- h_t: output vector
- $\hat{h}_t$: candidate activation vector
- f_t: forget vector
- W, U and b: parameter matrices and vector

References

1. Cho, Kyunghyun; van Merrienboer, Bart; Gulcehre, Caglar; Bahdanau, Dzmitry; Bougares, Fethi; Schwenk, Holger; Bengio, Yoshua (2014). "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation" (https://arxiv.org/abs/1406.1078). arXiv:1406.1078 (https://arxiv.org/abs/1406.1078).

2. Felix Gers; Jürgen Schmidhuber; Fred Cummins (1999). "Learning to Forget: Continual Prediction with LSTM" (https://ieeexplore.ieee.org/document/818041). *Proc. ICANN'99, IEE, London*. **1999**: 850–855. doi:10.1049/cp:19991218 (https://doi.org/10.1049%2Fcp%3A19991218). ISBN 0-85296-721-7.
3. "Recurrent Neural Network Tutorial, Part 4 – Implementing a GRU/LSTM RNN with Python and Theano – WildML" (http://www.wildml.com/2015/10/recurrent-neural-network-tutorial-part-4-implementing-a-grulstm-rnn-with-python-and-theano/). *Wildml.com*. 2015-10-27. Retrieved May 18, 2016.
4. Ravanelli, Mirco; Brakel, Philemon; Omologo, Maurizio; Bengio, Yoshua (2018). "Light Gated Recurrent Units for Speech Recognition". *IEEE Transactions on Emerging Topics in Computational Intelligence*. **2** (2): 92–102. arXiv:1803.10225 (https://arxiv.org/abs/1803.10225). doi:10.1109/TETCI.2017.2762739 (https://doi.org/10.1109%2FTETCI.2017.2762739). S2CID 4402991 (https://api.semanticscholar.org/CorpusID:4402991).
5. Su, Yuahang; Kuo, Jay (2019). "On Extended Long Short-term Memory and Dependent Bidirectional Recurrent Neural Network". arXiv:1803.01686 (https://arxiv.org/abs/1803.01686).
6. Su, Yuanhang; Kuo, Jay (2014). "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling". arXiv:1412.3555 (https://arxiv.org/abs/1412.3555) [cs.NE (https://arxiv.org/archive/cs.NE)].
7. Gruber, N.; Jockisch, A. (2020), "Are GRU cells more specific and LSTM cells more sensitive in motive classification of text?", *Frontiers in Artificial Intelligence*, **3**, doi:10.3389/frai.2020.00040 (https://doi.org/10.3389%2Ffrai.2020.00040), S2CID 220252321 (https://api.semanticscholar.org/CorpusID:220252321)
8. Chung, Junyoung; Gulcehre, Caglar; Cho, KyungHyun; Bengio, Yoshua (2014). "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling". arXiv:1412.3555 (https://arxiv.org/abs/1412.3555) [cs.NE (https://arxiv.org/archive/cs.NE)].
9. Dey, Rahul; Salem, Fathi M. (2017-01-20). "Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks". arXiv:1701.05923 (https://arxiv.org/abs/1701.05923) [cs.NE (https://arxiv.org/archive/cs.NE)].
10. Heck, Joel; Salem, Fathi M. (2017-01-12). "Simplified Minimal Gated Unit Variations for Recurrent Neural Networks". arXiv:1701.03452 (https://arxiv.org/abs/1701.03452) [cs.NE (https://arxiv.org/archive/cs.NE)].

Retrieved from "https://en.wikipedia.org/w/index.php?title=Gated_recurrent_unit&oldid=997015931"

This page was last edited on 29 December 2020, at 16:52 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.