Title: Gated recurrent unit

URL: https://en.wikipedia.org/wiki/Gated_recurrent_unit

PageID: 50569499

Categories: Category:2014 in artificial intelligence, Category:2014 software, Category:Artificial neural networks, Category:Neural network architectures

---

Supervised learning

Unsupervised learning

Semi-supervised learning

Self-supervised learning

Reinforcement learning

Meta-learning

Online learning

Batch learning

Curriculum learning

Rule-based learning

Neuro-symbolic AI

Neuromorphic engineering

Quantum machine learning

Classification

Generative modeling

Regression

Clustering

Dimensionality reduction

Density estimation

Anomaly detection

Data cleaning

AutoML

Association rules

Semantic analysis

Structured prediction

Feature engineering

Feature learning

Learning to rank

Grammar induction

Ontology learning

Multimodal learning

Apprenticeship learning

Decision trees

Ensembles Bagging Boosting Random forest

Bagging

Boosting

Random forest

k -NN

Linear regression

Naive Bayes

Artificial neural networks

Logistic regression

Perceptron

Relevance vector machine (RVM)

Support vector machine (SVM)

BIRCH

CURE

Hierarchical

k -means

Fuzzy

Expectation–maximization (EM)

DBSCAN

OPTICS

Mean shift

Factor analysis

CCA

ICA

LDA

NMF

PCA

PGD

t-SNE

SDL

Graphical models Bayes net Conditional random field Hidden Markov

Bayes net

Conditional random field

Hidden Markov

RANSAC

k -NN

Local outlier factor

Isolation forest

Autoencoder

Deep learning

Feedforward neural network

Recurrent neural network LSTM GRU ESN reservoir computing

LSTM

GRU

ESN

reservoir computing

Boltzmann machine Restricted

Restricted

GAN

Diffusion model

SOM

Convolutional neural network U-Net LeNet AlexNet DeepDream

U-Net

LeNet

AlexNet

DeepDream

Neural field Neural radiance field Physics-informed neural networks

Neural radiance field

Physics-informed neural networks

Transformer Vision

Vision

Mamba

Spiking neural network

Memtransistor

Electrochemical RAM (ECRAM)

Q-learning

Policy gradient

SARSA

Temporal difference (TD)

Multi-agent Self-play

Self-play

Active learning

Crowdsourcing

Human-in-the-loop

Mechanistic interpretability

RLHF

Coefficient of determination

Confusion matrix

Learning curve

ROC curve

Kernel machines

Bias–variance tradeoff

Computational learning theory

Empirical risk minimization

Occam learning

PAC learning

Statistical learning

VC theory

Topological deep learning

AAAI

ECML PKDD

NeurIPS

ICML

ICLR

IJCAI

ML

JMLR

Glossary of artificial intelligence

List of datasets for machine-learning research List of datasets in computer vision and image processing

List of datasets in computer vision and image processing

Outline of machine learning

v

t

e

In artificial neural networks , the gated recurrent unit (GRU) is a gating mechanism used 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 gating mechanism to input or forget certain features, [2] but lacks a context vector or output gate, resulting in fewer parameters than LSTM. [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 showed that gating is indeed helpful in general, and Bengio 's team came to no concrete conclusion on which of the two gating units was better. [6] [7]

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]

In the following, the operator ■ {\displaystyle \odot } denotes the Hadamard product .

Fully gated unit

Initially, for t = 0 {\displaystyle t=0} , the output vector is h 0 = 0 {\displaystyle h_{0}=0} .

Variables (d {\displaystyle d} denotes the number of input features and e {\displaystyle e} the number of output features):

x t ∈ R d {\displaystyle x_{t}\in \mathbb {R} ^{d}} : input vector

h t ∈ R e {\displaystyle h_{t}\in \mathbb {R} ^{e}} : output vector

h ^ t ∈ R e {\displaystyle {\hat {h}}_{t}\in \mathbb {R} ^{e}} : candidate activation vector

z t ∈ (0 , 1) e {\displaystyle z_{t}\in (0,1)^{e}} : update gate vector

r t ∈ (0 , 1) e {\displaystyle r_{t}\in (0,1)^{e}} : reset gate vector

W ∈ R e × d {\displaystyle W\in \mathbb {R} ^{e\times d}} , U ∈ R e × e {\displaystyle U\in \mathbb {R} ^{e\times e}} and b ∈ R e {\displaystyle b\in \mathbb {R} ^{e}} : parameter matrices and vector which need to be learned during training

Activation functions

σ {\displaystyle \sigma } : The original is a logistic function .

φ {\displaystyle \phi } : The original is a hyperbolic tangent .

Alternative activation functions are possible, provided that σ (x) ∈ [0 , 1] {\displaystyle \sigma (x)\in [0,1]} .

Alternate forms can be created by changing z t {\displaystyle z_{t}} and r t {\displaystyle r_{t}} [9]

Type 1: each gate depends only on the previous hidden state and the bias. z t = σ (U z h t – 1 + b z) r t = σ (U r h t – 1 + b r) {\displaystyle {\begin{aligned}z_{t}&=\sigma (U_{z}h_{t-1}+b_{z})\\r_{t}&=\sigma (U_{r}h_{t-1}+b_{r})\\\end{aligned}}}

Type 2: each gate depends only on the previous hidden state. z t = σ (U z h t – 1) r t = σ (U r h t – 1) {\displaystyle {\begin{aligned}z_{t}&=\sigma (U_{z}h_{t-1})\\r_{t}&=\sigma (U_{r}h_{t-1})\\\end{aligned}}}

Type 3: each gate is computed using only the bias. z t = σ (b z) r t = σ (b r) {\displaystyle {\begin{aligned}z_{t}&=\sigma (b_{z})\\r_{t}&=\sigma (b_{r})\\\end{aligned}}}

Minimal gated unit

The minimal gated unit (MGU) 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]

Variables

x t {\displaystyle x_{t}} : input vector

h t {\displaystyle h_{t}} : output vector

h ^ t {\displaystyle {\hat {h}}_{t}} : candidate activation vector

f t {\displaystyle f_{t}} : forget vector

W {\displaystyle W} , U {\displaystyle U} and b {\displaystyle b} : parameter matrices and vector

Light gated recurrent unit

The light gated recurrent unit (LiGRU) [4] removes the reset gate altogether, replaces tanh with the ReLU activation, and applies batch normalization (BN):

LiGRU has been studied from a Bayesian perspective. [11] This analysis yielded a variant called light Bayesian recurrent unit (LiBRU), which showed slight improvements over the LiGRU on speech recognition tasks.

References

v

t

e

History timeline

timeline

Companies

Projects

Parameter Hyperparameter

Hyperparameter

Loss functions

Regression Bias–variance tradeoff Double descent Overfitting

Bias–variance tradeoff

Double descent

Overfitting

Clustering

Gradient descent SGD Quasi-Newton method Conjugate gradient method

SGD

Quasi-Newton method

Conjugate gradient method

Backpropagation

Attention

Convolution

Normalization Batchnorm

Batchnorm

Activation Softmax Sigmoid Rectifier

Softmax

Sigmoid

Rectifier

Gating

Weight initialization

Regularization

Datasets Augmentation

Augmentation

Prompt engineering

Reinforcement learning Q-learning SARSA Imitation Policy gradient

Q-learning

SARSA

Imitation

Policy gradient

Diffusion

Latent diffusion model

Autoregression

Adversary

RAG

Uncanny valley

RLHF

Self-supervised learning

Reflection

Recursive self-improvement

Hallucination

Word embedding

Vibe coding

Machine learning In-context learning

In-context learning

Artificial neural network Deep learning

Deep learning

Language model Large language model NMT

Large language model

NMT

Reasoning language model

Model Context Protocol

Intelligent agent

Artificial human companion

Humanity's Last Exam

Artificial general intelligence (AGI)

AlexNet

WaveNet

Human image synthesis

HWR

OCR

Computer vision

Speech synthesis 15.ai ElevenLabs

15.ai

ElevenLabs

Speech recognition Whisper

Whisper

Facial recognition

AlphaFold

Text-to-image models Aurora DALL-E Firefly Flux Ideogram Imagen Midjourney Recraft Stable Diffusion

Aurora

DALL-E

Firefly

Flux

Ideogram

Imagen

Midjourney

Recraft

Stable Diffusion

Text-to-video models Dream Machine Runway Gen Hailuo AI Kling Sora Veo

Dream Machine

Runway Gen

Hailuo AI

Kling

Sora

Veo

Music generation Riffusion Suno AI Udio

Riffusion

Suno AI

Udio

Word2vec

Seq2seq

GloVe

BERT

T5

Llama

Chinchilla AI

PaLM

GPT 1 2 3 J ChatGPT 4 4o o1 o3 4.5 4.1 o4-mini 5

1

2

3

J

ChatGPT

4

4o

o1

o3

4.5

4.1

o4-mini

5

Claude

Gemini Gemini (language model) Gemma

Gemini (language model)

Gemma

Grok

LaMDA

BLOOM

DBRX

Project Debater

IBM Watson

IBM Watsonx

Granite

PanGu-Σ

DeepSeek

Qwen

AlphaGo

AlphaZero

OpenAI Five

Self-driving car

MuZero

Action selection AutoGPT

AutoGPT

Robot control

Alan Turing

Warren Sturgis McCulloch

Walter Pitts

John von Neumann

Claude Shannon
Shun'ichi Amari
Kunihiko Fukushima
Takeo Kanade
Marvin Minsky
John McCarthy
Nathaniel Rochester
Allen Newell
Cliff Shaw
Herbert A. Simon
Oliver Selfridge
Frank Rosenblatt
Bernard Widrow
Joseph Weizenbaum
Seymour Papert
Seppo Linnainmaa
Paul Werbos
Geoffrey Hinton
John Hopfield
Jürgen Schmidhuber
Yann LeCun
Yoshua Bengio
Lotfi A. Zadeh
Stephen Grossberg
Alex Graves
James Goodnight
Andrew Ng
Fei-Fei Li
Alex Krizhevsky
Ilya Sutskever
Oriol Vinyals
Quoc V. Le
Ian Goodfellow
Demis Hassabis
David Silver
Andrej Karpathy
Ashish Vaswani
Noam Shazeer

Aidan Gomez

John Schulman

Mustafa Suleyman

Jan Leike

Daniel Kokotajlo

François Chollet

Neural Turing machine

Differentiable neural computer

Transformer Vision transformer (ViT)

Vision transformer (ViT)

Recurrent neural network (RNN)

Long short-term memory (LSTM)

Gated recurrent unit (GRU)

Echo state network

Multilayer perceptron (MLP)

Convolutional neural network (CNN)

Residual neural network (RNN)

Highway network

Mamba

Autoencoder

Variational autoencoder (VAE)

Generative adversarial network (GAN)

Graph neural network (GNN)

Category