MULTI-TASK RECURRENT NEURAL NETWORKS

BACKGROUND

This specification relates to recurrent neural networks.

Neural networks are machine learning models that employ multiple layers of operations to predict one or more outputs from one or more inputs. Neural networks typically include one or more hidden layers situated between an input layer and an output layer. The output of each hidden layer is used as input to a next layer, e.g., a next hidden layer or the output layer.

Each layer of a neural network specifies one or more transformation operations to be performed on input to the layer. Some neural network layers have operations that are referred to as neurons. Each neuron receives one or more inputs and generates an output that is received by another neural network layer. Often, each neuron receives inputs from other neurons, and each neuron provides an output to one or more other neurons.

Each layer generates one or more outputs using the current values of a set of parameters for the layer. Training a neural network involves continually performing a forward pass on the input, computing gradient values, and updating the current values of the set of parameters for each layer. Once a neural network is trained, the final set of parameters can be used to make predictions in a production system.

Some neural networks are recurrent neural networks. A recurrent neural network is a neural network that receives an input sequence and generates an output sequence from the input sequence. In particular, a recurrent neural network can use some or all of the internal state of the network from a previous time step in computing an output at a current time step. This allows the recurrent neural network to exhibit dynamic temporal behavior. In other words, the recurrent neural network summarizes all information it has received up to a current time step and is capable of capturing long-term dependencies in sequences of data.

SUMMARY

This specification describes multi-task recurrent neural networks, and more specifically, multi-task recurrent neural networks that improve time-series predictions in microprocessors.

According to a first aspect of the present invention, there is provided a method comprising:

maintaining internal state data for a recurrent neural network that is trained to predict memory access addresses of a computer program in a plurality of memory regions, the internal state data specifying a separate internal state for each of the plurality of memory regions;

receiving a current input in a sequence of inputs, the sequence of inputs defining a sequence of prior memory access addresses of the computer program;

identifying a particular memory region of the memory access address defined by the current input in the sequence of inputs;

selecting, from the internal states specified in the maintained data, the internal state for the particular memory region;

processing, in accordance with the selected internal state for the particular memory region, the current input in the sequence of inputs using the recurrent neural network to:

generate an output, the output defining a probability distribution of a predicted memory access address, the predicted memory access address being a future memory access address of the computer program in the particular memory region, and

update the selected internal state of the particular memory region; and

associating the updated selected internal state with the particular memory region in the maintained data.

According to a second aspect of the present invention, there is provided a circuit comprising:

a recurrent neural network cell (RNN cell) configured to perform operations of a recurrent neural network;

a plurality of state registers, wherein each state register is configured to (i) store internal state data for the recurrent neural network and (ii) corresponds to a respective task of a plurality of tasks that the neural network has been trained to perform; and

circuitry configured to perform operations comprising:

receiving a current input in a sequence of inputs and sequence identifying data,

identifying a task of the plurality of tasks based on the sequence identifying data,

obtaining internal state data from the state register corresponding to the identified task,

providing, to the RNN cell, the obtained internal state data and the current input to cause the RNN cell to process the obtained internal state data and the current input to generate a neural network output and updated internal state data for the identified task,

receiving, from the RNN cell, the neural network output and the updated internal state data for the identified task, and

writing the updated internal state data to the state register corresponding to the identified task.

A circuit that implements a multi-task recurrent neural network includes a recurrent neural network cell ("RNN cell") configured to perform operations of a