

PROJECTION NEURAL NETWORKS

BACKGROUND

[0001] This specification relates to projection neural networks.

[0002] Neural networks are machine learning models that employ one or more layers of nonlinear units to predict an output for a received input. Some neural networks include one or more hidden layers in addition to an output layer. The output of each hidden layer is used as input to the next layer in the network, i.e., the next hidden layer or the output layer. Each layer of the network generates an output from a received input in accordance with current values of a respective set of parameters.

[0003] 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. An example of a recurrent neural network is a long short term (LSTM) neural network that includes one or more LSTM memory blocks. Each LSTM memory block can include one or more cells that each include an input gate, a forget gate, and an output gate that allow the cell to store previous states for the cell, e.g., for use in generating a current activation or to be provided to other components of the LSTM neural network.

SUMMARY

[0004] This specification describes a projection neural network implemented as computer programs on one or more computers in one or more locations.

[0005] According to a first aspect there is provided a system including one or more computers and one or more storage devices storing instructions that when executed by the one or more computers cause the one or more computers to implement a projection neural network. The projection neural network is configured to receive a projection network input and to generate a projection network output from the projection network input. The projection neural network includes a sequence of one or more projection layers, wherein each projection layer has a plurality of projection layer parameters and a plurality of projection layer functions having respective projection function parameters. Values of the plurality of projection layer parameters were adjusted during the training of the projection neural network, wherein the projection function parameters are random values determined before the projection neural network is trained and were not adjusted during training of

the projection neural network. Each projection layer is configured to receive a non-binary layer input and apply the plurality of projection layer functions to the non-binary layer input. Each projection layer function processes the non-binary layer input in accordance with the respective projection function parameters to generate a respective binary projection function output that projects the non-binary layer input to a lower dimensionality binary projection function output having a plurality of components that each assume a respective value from a set of two possible values.

Each projection layer generates a layer output by applying the projection layer parameters for the projection layer to the binary projection function outputs.

[0006] In some implementations, the projection neural network further includes an output layer configured to receive a layer output generated by a highest projection layer in the sequence and to generate the projection network output.

[0007] In some implementations, the output layer is a softmax output layer.

[0008] In some implementations, the projection neural network includes only the sequence of projection layers and the output layer.

[0009] In some implementations, a layer output of the highest projection layer in the sequence is the projection network output.

[0010] In some implementations, the projection neural network includes only the sequence of projection layers.

[0011] In some implementations, the layer input of a lowest projection layer in the sequence is a network input to the projection neural network.

[0012] In some implementations, the layer input of any projection layer other than the lowest projection layer is a layer output generated by the projection layer immediately below the projection layer in the sequence.

[0013] In some implementations, for each projection layer, each projection function is associated with a respective set of projection vectors. Applying each projection function to the layer input includes, for each projection vector: (i) determining a dot product between the layer input and the projection vector, (ii) when the dot product is negative, assigning a first value to a corresponding position in the projection function output, and (iii) when the dot product is positive, assigning a second value to the corresponding position in the projection function output.

[0014] In some implementations, for each projection layer, the projection functions are each encoded as sparse matrices and are used to generate a binary representation from the layer input.

[0015] In some implementations, the projection layer parameters include a parameter matrix and a bias vector. Generating the layer output by applying the projection layer parameters for the projection layer to the projection function outputs includes applying the parameter matrix to the projection function outputs and then adding the bias vector to the result.

[0016] According to a second aspect there is provided a method of training the projection neural network. The method includes receiving a training input and a target output for the training input. The training input is processed using the projection neural network in accordance with current values of the projection layer parameters, and the random values of the projection function parameters which are determined before the projection neural network is trained and are not adjusted during training of the projection neural network, to generate a projection network output for the training input. The training input is processed using a trainer neural network having multiple trainer neural network parameters. The trainer neural network is configured to process the training input in accordance with current values of the trainer neural network parameters to generate a trainer network output that is specific to the particular machine learning task. A gradient is determined with respect to the trainer neural network parameters of a loss function that depends on an error between the target output and the trainer network output. A gradient is determined with respect to the projection layer parameters of a loss function that depends on an error between the trainer network output and the projection network output. Updates to the current values of the trainer network parameters and the projection layer parameters are determined using the gradients.

[0017] In some implementations, the output generated by the trainer neural network is a soft target output.

[0018] In some implementations, the method further includes determining a gradient with respect to the trainer network parameters of the loss function that depends on the error between the trainer network output and the projection network output. The update to the current values of the trainer network parameters is also based on the gradient with respect to the trainer network parameters of the loss function that depends on the error between the trainer network output and the projection network output.

[0019] In some implementations, the method further includes determining a gradient with respect to the projection layer parameters of a loss function that depends on an error between the target output and the projection network output. The update to the current values of the projection layer parameters is also based on the gradient with respect to the projection layer parameters of the loss function that depends on the error between the target output and the projection network output.

[0020] According to a third aspect there is provided a system including one or more computers and one or more storage devices storing instructions that when executed by the one or more

CLAIMS

1. A system comprising one or more computers and one or more storage devices storing instructions that when executed by the one or more computers cause the one or more computers to implement:

a projection neural network configured to receive a projection network input and to generate a projection network output from the projection network input, the projection neural network comprising a sequence of one or more projection layers, wherein each projection layer has a plurality of projection layer parameters, and a plurality of projection layer functions having respective projection function parameters, wherein values of the plurality of projection layer parameters were adjusted during the training of the projection neural network, wherein the projection function parameters are random values determined before the projection neural network is trained and were not adjusted during training of the projection neural network, and wherein each projection layer is configured to:

receive a non-binary layer input;

apply the plurality of projection layer functions to the non-binary layer input, wherein: each projection layer function processes the non-binary layer input in accordance with the respective projection function parameters to generate a respective binary projection function output that projects the non-binary layer input to a lower dimensionality binary projection function output having a plurality of components that each assume a respective value from a set of two possible values, and

generate a layer output by applying the projection layer parameters for the projection layer to the binary projection function outputs.

2. The system of claim 1, wherein the projection neural network further comprises:

an output layer configured to receive a layer output generated by a highest projection layer in the sequence and to generate the projection network output.

3. The system of claim 2, wherein the output layer is a softmax output layer.

4. The system of any one of claims 2 or 3, wherein the projection neural network includes only the sequence of projection layers and the output layer.

5. The system of claim 1, wherein a layer output of the highest projection layer in the sequence is the projection network output.
6. The system of claim 5, wherein the projection neural network includes only the sequence of projection layers.

7. The system of any one of claims 1-6, wherein the layer input of a lowest projection layer in the sequence is a network input to the projection neural network.
8. The system of claim 7, wherein the layer input of any projection layer other than the lowest projection layer is a layer output generated by the projection layer immediately below the projection layer in the sequence.
9. The system of any one of claims 1-8, wherein, for each projection layer, the projection function parameters associated with each projection layer function of the projection layer comprise a respective set of projection vectors, and wherein applying each projection layer function to the non-binary layer input comprises:
- for each projection vector:
 - determining a dot product between the non-binary layer input and the projection vector;
 - when the dot product is negative, assigning a first value from the set of two possible values to a corresponding component of the binary projection function output; and
 - when the dot product is positive, assigning a second value from the set of two possible values to the corresponding component of the binary projection function output.
10. The system of any one of claims 1-9, wherein, for each projection layer, the respective projection function parameters associated with each projection layer function of the projection layer are encoded as sparse matrices.
11. The system of any one of claims 1-10, wherein the projection layer parameters include a parameter matrix and a bias vector, and wherein generating the layer output by applying the projection layer parameters for the projection layer to the binary projection function outputs comprises:
- applying the parameter matrix to the binary projection function outputs and then adding the bias vector to the result.
12. A method of training the projection neural network of any one of claims 1-11, the method comprising:
- receiving a training input and a target output for the training input;
 - processing the training input using the projection neural network in accordance with

current values of the projection layer parameters, and the random values of the projection function parameters which are determined before the projection neural network is trained and are not adjusted during training of the projection neural network, to generate a projection network output for the training input;

processing the training input using a trainer neural network having a plurality of trainer neural network parameters, wherein the trainer neural network is configured to process the training input in accordance with current values of the trainer neural network parameters to generate a trainer network output that is specific to the particular machine learning task; and

determining a gradient with respect to the trainer neural network parameters of a loss function that depends on an error between the target output and the trainer network output;

determining a gradient with respect to the projection layer parameters of a loss function that depends on an error between the trainer network output and the projection network output; and

determining updates to the current values of the trainer network parameters and the projection layer parameters using the gradients.

13. The method of claim 12, wherein the output generated by the trainer neural network is a soft target output.

14. The method of any one of claims 12 or 13, further comprising:

determining a gradient with respect to the trainer network parameters of the loss function that depends on the error between the trainer network output and the projection network output, wherein the update to the current values of the trainer network parameters is also based on the gradient with respect to the trainer network parameters of the loss function that depends on the error between the trainer network output and the projection network output.

15. The method of any one of claims 12-14, further comprising:

determining a gradient with respect to the projection layer parameters of a loss function that depends on an error between the target output and the projection network output, wherein the update to the current values of the projection layer parameters is also based on the gradient with respect to the projection layer parameters of the loss function that depends on the error between the target output and the projection network output.

16. A system comprising one or more computers and one or more storage devices storing instructions that when executed by the one or more computers cause the one or more computers to perform the operations of the respective method of any one of claims 12-15.

17. One or more computer storage media storing instructions that when executed by one or more computers cause the one or more computers to perform the operations of the respective method of any one of claims 12-15.

18. The system of claim 1, the one or more storage devices corresponding to one or more computer storage media.