lstm

April 9, 2019

0.0.1 Long Short-term Memory (LSTM)

```
In [1]: import sys
        sys.path.insert(0, '..')

        import d2l
        from mxnet import nd
        from mxnet.gluon import rnn

        (corpus_indices, char_to_idx, idx_to_char,
         vocab_size) = d2l.load_data_time_machine()
```

0.0.2 Initialize Model Parameters

```
In [2]: num_inputs, num_hiddens, num_outputs = vocab_size, 256, vocab_size
        ctx = d2l.try_gpu()

        def get_params():
            def _one(shape):
                return nd.random.normal(scale=0.01, shape=shape, ctx=ctx)

            def _three():
                return (_one((num_inputs, num_hiddens)),
                        _one((num_hiddens, num_hiddens)),
                        nd.zeros(num_hiddens, ctx=ctx))

            W_xi, W_hi, b_i = _three()  # Input gate parameters
            W_xf, W_hf, b_f = _three()  # Forget gate parameters
            W_xo, W_ho, b_o = _three()  # Output gate parameters
            W_xc, W_hc, b_c = _three()  # Candidate cell parameters
            # Output layer parameters
            W_hq = _one((num_hiddens, num_outputs))
            b_q = nd.zeros(num_outputs, ctx=ctx)
            # Create gradient
            params = [W_xi, W_hi, b_i, W_xf, W_hf, b_f, W_xo, W_ho, b_o, W_xc, W_hc,
                      b_c, W_hq, b_q]
            for param in params:
```

```
            param.attach_grad()
        return params
```

0.0.3 State initializer

```
In [3]: def init_lstm_state(batch_size, num_hiddens, ctx):
            return (nd.zeros(shape=(batch_size, num_hiddens), ctx=ctx),
                    nd.zeros(shape=(batch_size, num_hiddens), ctx=ctx))
```

0.0.4 LSTM Cell

```
In [4]: def lstm(inputs, state, params):
            [W_xi, W_hi, b_i, W_xf, W_hf, b_f, W_xo, W_ho, b_o, W_xc, W_hc, b_c,
             W_hq, b_q] = params
            (H, C) = state
            outputs = []
            for X in inputs:
                I = nd.sigmoid(nd.dot(X, W_xi) + nd.dot(H, W_hi) + b_i)
                F = nd.sigmoid(nd.dot(X, W_xf) + nd.dot(H, W_hf) + b_f)
                O = nd.sigmoid(nd.dot(X, W_xo) + nd.dot(H, W_ho) + b_o)
                C_tilda = nd.tanh(nd.dot(X, W_xc) + nd.dot(H, W_hc) + b_c)
                C = F * C + I * C_tilda
                H = O * C.tanh()
                Y = nd.dot(H, W_hq) + b_q
                outputs.append(Y)
            return outputs, (H, C)
```

0.0.5 Train the Model

```
In [5]: num_epochs, num_steps, batch_size, lr, clipping_theta = 160, 35, 32, 1e2, 1e-2
        pred_period, pred_len, prefixes = 40, 50, ['traveller', 'time traveller']
```

```
In [6]: d2l.train_and_predict_rnn(lstm, get_params, init_lstm_state, num_hiddens,
                                  vocab_size, ctx, corpus_indices, idx_to_char,
                                  char_to_idx, False, num_epochs, num_steps, lr,
                                  clipping_theta, batch_size, pred_period, pred_len,
                                  prefixes)
```

```
epoch 40, perplexity 7.948798, time 0.91 sec
 - traveller the the the the the the the the the the the the the t
 - time traveller the the the the the the the the the the the the t
epoch 80, perplexity 3.831706, time 0.92 sec
 - traveller, and why hand the time traveller, and the time tr
 - time traveller cand for in and filby, and why hand the time trav
epoch 120, perplexity 1.922319, time 0.90 sec
 - traveller shilby beed hionel re grimintid masion.'  'sol is
 - time traveller hefres, wh hall ngand to overlook this flectoond
epoch 160, perplexity 1.322308, time 0.90 sec
 - traveller the psychologist. 'you _can_ move about in all di
```

```
 - time traveller  'but bour filby, 'frar ather show back for any t
```

0.1 Gluon Implementation

```
In [7]: lstm_layer = rnn.LSTM(num_hiddens)
        model = d2l.RNNModel(lstm_layer, vocab_size)
        d2l.train_and_predict_rnn_gluon(model, num_hiddens, vocab_size, ctx,
                                        corpus_indices, idx_to_char, char_to_idx,
                                        num_epochs, num_steps, lr, clipping_theta,
                                        batch_size, pred_period, pred_len, prefixes)
```

```
epoch 40, perplexity 8.291799, time 0.48 sec
 - traveller and the tre the tre the tre the tre the tre the t
 - time traveller and the tre the tre the tre the tre the tre the t
epoch 80, perplexity 4.629873, time 0.55 sec
 - traveller another and whe the perention a four anotter and
 - time traveller a fourth dimension of space, and the peetter. the
epoch 120, perplexity 2.379147, time 0.48 sec
 - traveller the fourth dimension of space excent only mone so
 - time traveller cometry on a menttoncest on an shisses, an in,'is
epoch 160, perplexity 1.475115, time 0.49 sec
 - traveller. ''s all right ang mently wetter uloun--if the gr
 - time traveller came back, and so i neder that lime traveller cam
```