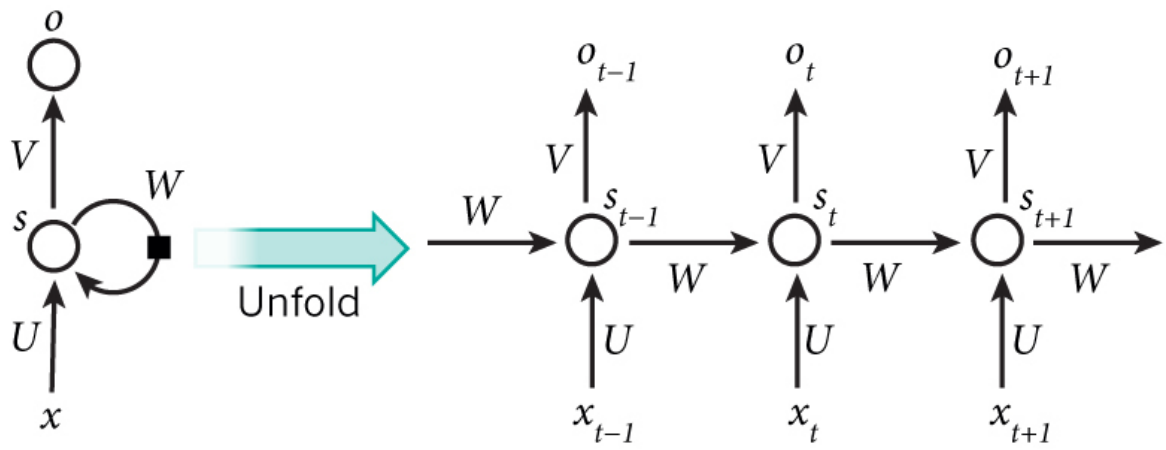
RNN 실습 - 텐서플로우에서 LSTM 및 GRU 사용

1. RNN 모형 및 구성방법
2. 텐서플로우에서 지원하는 RNN 'Cell' 유형
3. 유형별 특성 테스트 위한 코드 구성
4. Test #1 - Vanila RNN
5. Test #2 - Basic LSTM
6. Test #3 - GRU
7. Test #4 - LSTMCell + forget_bias
8. Test #5 - LayerNormBasicLSTMCell
9. Test #6 - LayerNormBasicLSTMCell - what's wrong?
10. 정리

In [1]:
```
!rm -fr logdir
!mkdir -p logdir
```

In [2]:
```
%load_ext do_not_print_href
%matplotlib inline
from __future__ import print_function, division
import sys
import time
import numpy as np
import tensorflow as tf
import matplotlib.pyplot as plt
```

RNN 모형 (복습)



이미지 출처: http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

$$s_t = tanh(Ux_t + Ws_{t-1})$$
$$o_t = softmax(Vs_t)$$

하지만 텐서플로우 RNN 은...

- ***V*** 에 해당하는 구조가 없음
- ***softmax()*** 도 없음
- ***V*** 와 ***softmax()*** 는 필요할 때만 만들어서 붙이면 됨 (`tf.layers.dense`, `tf.nn.softmax`)

RNN 모델 구성 (복습)

- tf.contrib.rnn.BasicRNNCell

```
`__init__`(
  num_units,
  activation=None,
  reuse=None
)
```

- tf.nn.dynamic_rnn

```
dynamic_rnn(
  cell,
  inputs,
  sequence_length=None,
  initial_state=None,
  dtype=None,
  parallel_iterations=None,
  swap_memory=False,
  time_major=False,
  scope=None
)
```

- RNN 구성

```
  cell            = tf.contrib.rnn.BasicRNNCell(
                      num_hidden_units)
  last, states    = tf.nn.dynamic_rnn(
                      cell,
                      inputs,
                      sequence_length=sequence_length,
                      dtype=tf.float32)
```

RNN Cell 종류

- tf.contrib.rnn.BasicRNNCell

```
  ...
```

- tf.contrib.rnn.BasicLSTMCell

```
`__init__`(
    num_units,
    forget_bias=1.0,
    state_is_tuple=True,
    activation=None,
    reuse=None
)
```

- tf.contrib.rnn.LSTMCell

```
`__init__`(
    num_units,
    use_peepholes=False,
    cell_clip=None,
    initializer=None,
    num_proj=None,
    proj_clip=None,
    num_unit_shards=None,
    num_proj_shards=None,
    forget_bias=1.0,
    state_is_tuple=True,
    activation=None,
    reuse=None
)
```

- tf.contrib.rnn.GRUCell

```
`__init__`(
    num_units,
    activation=None,
    reuse=None,
    kernel_initializer=None,
    bias_initializer=None
)
```

- tf.contrib.rnn.LayerNormBasicLSTMCell

> LSTM unit with **layer normalization** and **recurrent dropout.**

- "Layer Normalization" Jimmy Lei Ba, Jamie Ryan Kiros, Geoffrey E. Hinton
  - https://arxiv.org/abs/1607.06450
- "Recurrent Dropout without Memory Loss" Stanislau Semeniuta, Aliaksei Severyn, Erhardt Barth.
  - https://arxiv.org/abs/1603.05118
- "Layer Normalization, Nishida Geio, 번역 김홍배 (slideshare.net)"
  - https://www.slideshare.net/ssuser06e0c5/normalization-72539464

```
`__init__`(
    num_units,
    forget_bias=1.0,
    input_size=None,
    activation=tf.tanh,
    layer_norm=True,
    norm_gain=1.0,
    norm_shift=0.0,
    dropout_keep_prob=1.0,
    dropout_prob_seed=None,
    reuse=None
)
```

- 기타 여러가지 Wrapper 지원 - https://www.tensorflow.org/api_guides/python/contrib.rnn

In [3]:
```
from tensorflow.examples.tutorials.mnist.input_data \
    import read_data_sets

mnist = read_data_sets('./mnist', one_hot=False)
```

```
Extracting ./mnist/train-images-idx3-ubyte.gz
Extracting ./mnist/train-labels-idx1-ubyte.gz
Extracting ./mnist/t10k-images-idx3-ubyte.gz
Extracting ./mnist/t10k-labels-idx1-ubyte.gz
```

In [4]:
```
INPUT_UNITS = 28
NUM_HIDDEN_UNITS = 31

BATCH_SIZE = 128
MAX_SEQ_LEN = 28
```

In [5]:
```
class MnistRnn:
    def __init__(self,
                 inputs,
                 labels,
                 input_units,
                 num_hidden_units,
                 batch_size,
                 max_seq_len,
                 rnn_cell_class = tf.contrib.rnn.BasicRNNCell,
                 # 여기 이후의 인수는 잠깐 무시 해 주세요
                 add_check = False,
                 lr = 0.001,
                 use_grad_clip = False):
        '''
        inputs: in shape [batch_size, max_seq_len, input_size]
        labels: in shape [batch_size]
        '''

        cell            = rnn_cell_class(num_hidden_units)
        sequence_length = [max_seq_len] * batch_size
        last, states    = tf.nn.dynamic_rnn(
                              cell,
```

```
                        inputs,
                        sequence_length=sequence_length,
                        dtype=tf.float32)

        # max_seq_len 축으로 0~27 까지 값 중에
        # 0~26 때의 출력 값은 사용하지 않음
        rnn_output = last[:,max_seq_len-1,:]
        # outputs shape will be: [batch_size, 10]
        outputs    = tf.layers.dense(rnn_output, 10)
        loss       = tf.losses.sparse_softmax_cross_entropy(
                        labels, outputs)

        if use_grad_clip:
            tvars_     = tf.trainable_variables()
            grads_, _  = tf.clip_by_global_norm(
                            tf.gradients(
                                loss,
                                tvars_),
                            5.0)
            optimize   = \
                tf.train.AdamOptimizer(learning_rate=lr). \
                            apply_gradients(zip(grads_, tvars_))
        else:
            optimize   = \
                tf.train.AdamOptimizer(learning_rate=lr). \
                            minimize(loss)

        # accuracy
        preds    = tf.argmax(outputs, axis=1)
        errors   = tf.count_nonzero(labels - preds)
        accuracy = 1.0 - tf.cast(errors,tf.float32) / \
                         tf.cast(tf.size(preds),tf.float32)

        # 클래스 객체 외부에서 참고할 수 있도록 속성으로 저장
        self.outputs        = outputs
        self.loss           = loss
        self.optimize       = optimize
        self.accuracy       = accuracy

        # check_numerics
        self.check = [tf.check_numerics(
                        t,
                        'check_numerics: {}'.format(t.name)) \
                      for t in tf.gradients(
                                  loss,
                                  tf.trainable_variables()) \
                      if t is not None] \
                     if add_check \
                     else tf.constant(1.0)
```

In [6]:
```
train_loop_count = mnist.train.num_examples // BATCH_SIZE
test_loop_count  = mnist.test.num_examples // BATCH_SIZE

train_loop_count, test_loop_count
```

Out[6]: (429, 78)

In [7]:

```
def train(inputs, labels, max_epochs, train_writer, test_writer):
    step = 0
    for ep in range(max_epochs):

        train_elapsed = []
        train_losses = []
        train_accuracy = []
        for i in range(train_loop_count):
            t_start     = time.time()
            offs        = i * BATCH_SIZE
            batch_input = \
                    mnist.train.images[offs:offs+BATCH_SIZE,:]
            batch_input = \
                    batch_input.reshape(
                            [BATCH_SIZE,
                               MAX_SEQ_LEN,
                               INPUT_UNITS])
            batch_label = \
                    mnist.train.labels[offs:offs+BATCH_SIZE]
            optimize, loss, accuracy, _ = \
                sess.run([model.optimize,
                          model.loss,
                          model.accuracy,
                          model.check],
                         feed_dict = {
                          inputs: batch_input,
                          labels: batch_label })
            train_losses.append(loss)
            train_accuracy.append(accuracy)
            t_elapsed   = time.time() - t_start
            train_elapsed.append(t_elapsed)

            step += 1
            summary = tf.Summary(
                value=[
                    tf.Summary.Value(
                        tag='train_accuracy',
                        simple_value=accuracy
                    ),
                    tf.Summary.Value(
                        tag='loss',
                        simple_value=loss
                    ),
                ]
            )
            train_writer.add_summary(summary,global_step=step)

            if step % 250 == 0:
                print(('[trn] ep {:d}, step {:d}, ' +
                       'loss {:f}, accu {:f}, ' +
                       'sec/iter {:f}').format(
                    ep + 1,
                    step,
                    np.mean(train_losses),
                    np.amin(train_accuracy),
                    np.mean(train_elapsed)))
                train_losses = []
```

```
            train_accuracy = []
            train_elapsed = []

        train_writer.flush()

        test_elapsed  = []
        test_accuracy = []

        for i in range(test_loop_count):
            t_start     = time.time()
            offs        = i * BATCH_SIZE
            batch_input = mnist.test.images[offs:offs+BATCH_SIZE,:]
            batch_input = batch_input.reshape(
                            [BATCH_SIZE,
                               MAX_SEQ_LEN,
                               INPUT_UNITS])
            batch_label = mnist.test.labels[offs:offs+BATCH_SIZE]
            accuracy, = \
                sess.run([model.accuracy],
                         feed_dict = {
                          inputs: batch_input,
                          labels: batch_label })
            test_accuracy.append(accuracy)
            t_elapsed   = time.time() - t_start
            test_elapsed.append(t_elapsed)

            step += 1

        if len(test_accuracy) > 0:
            print(('[tst] ep {:d}, step {:d}, ' +
                   'accu {:f}, sec/iter {:f}').format(
                ep + 1,
                step,
                np.amin(test_accuracy),
                np.mean(test_elapsed)))

            summary = tf.Summary(
                value=[
                    tf.Summary.Value(
                        tag='test_accuracy',
                        simple_value=np.amin(test_accuracy)
                    ),
                ]
            )
            test_writer.add_summary(summary,global_step=step)
            test_writer.flush()
```

Test #1 - Vanila RNN

```
cell = tf.contrib.rnn.BasicRNNCell(num_hidden_units)
```

In [8]:
```
tf.reset_default_graph()

inputs_ = tf.placeholder(tf.float32,
                         [BATCH_SIZE, MAX_SEQ_LEN, INPUT_UNITS],
                         name='inputs')
labels_ = tf.placeholder(tf.int64,
                         [BATCH_SIZE],
                         name='labels')


model = MnistRnn(inputs_,
                 labels_,
                 INPUT_UNITS,
                 NUM_HIDDEN_UNITS,
                 BATCH_SIZE,
                 MAX_SEQ_LEN,
                 tf.contrib.rnn.BasicRNNCell)
```

In [9]:
```
config = tf.ConfigProto(gpu_options={'allow_growth':True})
sess = tf.InteractiveSession(config=config)

tf.global_variables_initializer().run()

train_writer = tf.summary.FileWriter('logdir/train_basic_rnn',
                                     graph=tf.get_default_graph())
test_writer  = tf.summary.FileWriter('logdir/test_basic_rnn',
                                     graph=tf.get_default_graph())

train(inputs_, labels_, 10, train_writer, test_writer)
```

```
[trn] ep 1, step 250, loss 1.550250, accu 0.085938, sec/iter 0.007
951
[tst] ep 1, step 507, accu 0.468750, sec/iter 0.002797
[trn] ep 2, step 750, loss 0.987132, accu 0.562500, sec/iter 0.007
306
[tst] ep 2, step 1014, accu 0.539062, sec/iter 0.002285
[trn] ep 3, step 1250, loss 0.765099, accu 0.625000, sec/iter 0.00
7807
[tst] ep 3, step 1521, accu 0.648438, sec/iter 0.004670
[trn] ep 4, step 1750, loss 0.632159, accu 0.648438, sec/iter 0.00
7768
[tst] ep 4, step 2028, accu 0.656250, sec/iter 0.003359
[trn] ep 5, step 2250, loss 0.562881, accu 0.664062, sec/iter 0.00
9445
[tst] ep 5, step 2535, accu 0.671875, sec/iter 0.005474
[trn] ep 6, step 2750, loss 0.504307, accu 0.695312, sec/iter 0.00
8914
[tst] ep 6, step 3042, accu 0.679688, sec/iter 0.003352
[trn] ep 7, step 3250, loss 0.462129, accu 0.718750, sec/iter 0.00
9098
[tst] ep 7, step 3549, accu 0.664062, sec/iter 0.005065
[trn] ep 8, step 3750, loss 0.429448, accu 0.757812, sec/iter 0.00
9314
[tst] ep 8, step 4056, accu 0.710938, sec/iter 0.003061
[trn] ep 9, step 4250, loss 0.401605, accu 0.750000, sec/iter 0.00
8644
[tst] ep 9, step 4563, accu 0.710938, sec/iter 0.004161
[trn] ep 10, step 4750, loss 0.384211, accu 0.781250, sec/iter 0.0
08817
[tst] ep 10, step 5070, accu 0.726562, sec/iter 0.003440
```

In [10]:
```
# !tensorboard --ip 0.0.0.0 --logdir logdir
```

Test #2 - Basic LSTM

```
cell = tf.contrib.rnn.BasicLSTMCell(num_hidden_units)
```

In [11]:
```
tf.reset_default_graph()

inputs_ = tf.placeholder(tf.float32,
                         [BATCH_SIZE, MAX_SEQ_LEN, INPUT_UNITS],
                         name='inputs')
labels_ = tf.placeholder(tf.int64,
                         [BATCH_SIZE],
                         name='labels')

model = MnistRnn(inputs_,
                 labels_,
                 INPUT_UNITS,
                 NUM_HIDDEN_UNITS,
                 BATCH_SIZE,
                 MAX_SEQ_LEN,
                 tf.contrib.rnn.BasicLSTMCell)
```

In [12]:
```
config = tf.ConfigProto(gpu_options={'allow_growth':True})
sess = tf.InteractiveSession(config=config)

tf.global_variables_initializer().run()

train_writer = tf.summary.FileWriter('logdir/train_basic_lstm',
                                     graph=tf.get_default_graph())
test_writer  = tf.summary.FileWriter('logdir/test_basic_lstm',
                                     graph=tf.get_default_graph())

train(inputs_, labels_, 10, train_writer, test_writer)
```

```
[trn] ep 1, step 250, loss 1.499541, accu 0.039062, sec/iter 0.013
030
[tst] ep 1, step 507, accu 0.695312, sec/iter 0.005082
[trn] ep 2, step 750, loss 0.469044, accu 0.671875, sec/iter 0.013
055
[tst] ep 2, step 1014, accu 0.796875, sec/iter 0.004687
[trn] ep 3, step 1250, loss 0.304381, accu 0.773438, sec/iter 0.01
3056
[tst] ep 3, step 1521, accu 0.851562, sec/iter 0.006747
[trn] ep 4, step 1750, loss 0.228537, accu 0.835938, sec/iter 0.01
3932
[tst] ep 4, step 2028, accu 0.867188, sec/iter 0.005018
[trn] ep 5, step 2250, loss 0.196097, accu 0.843750, sec/iter 0.01
2640
[tst] ep 5, step 2535, accu 0.867188, sec/iter 0.004851
[trn] ep 6, step 2750, loss 0.166952, accu 0.882812, sec/iter 0.01
2917
[tst] ep 6, step 3042, accu 0.875000, sec/iter 0.004824
[trn] ep 7, step 3250, loss 0.147182, accu 0.898438, sec/iter 0.01
4815
[tst] ep 7, step 3549, accu 0.867188, sec/iter 0.005175
[trn] ep 8, step 3750, loss 0.133158, accu 0.906250, sec/iter 0.01
3107
[tst] ep 8, step 4056, accu 0.867188, sec/iter 0.005694
[trn] ep 9, step 4250, loss 0.119845, accu 0.914062, sec/iter 0.01
3864
[tst] ep 9, step 4563, accu 0.906250, sec/iter 0.005666
[trn] ep 10, step 4750, loss 0.107291, accu 0.914062, sec/iter 0.0
15189
[tst] ep 10, step 5070, accu 0.906250, sec/iter 0.004921
```

In [13]:
```
# !tensorboard --ip 0.0.0.0 --logdir logdir
```

Test #3 - GRU

```
cell = tf.contrib.rnn.GRUCell(num_hidden_units)
```

In [14]:
```
tf.reset_default_graph()

inputs_ = tf.placeholder(tf.float32,
                         [BATCH_SIZE, MAX_SEQ_LEN, INPUT_UNITS],
                         name='inputs')
labels_ = tf.placeholder(tf.int64,
                         [BATCH_SIZE],
                         name='labels')

model = MnistRnn(inputs_,
                 labels_,
                 INPUT_UNITS,
                 NUM_HIDDEN_UNITS,
                 BATCH_SIZE,
                 MAX_SEQ_LEN,
                 tf.contrib.rnn.GRUCell)
```

In [15]:
```
config = tf.ConfigProto(gpu_options={'allow_growth':True})
sess = tf.InteractiveSession(config=config)

tf.global_variables_initializer().run()

train_writer = tf.summary.FileWriter(
                'logdir/train_gru',
                graph=tf.get_default_graph())
test_writer  = tf.summary.FileWriter(
                'logdir/test_gru',
                graph=tf.get_default_graph())

train(inputs_, labels_, 10, train_writer, test_writer)
```

```
[trn] ep 1, step 250, loss 1.550183, accu 0.117188, sec/iter 0.014
488
[tst] ep 1, step 507, accu 0.679688, sec/iter 0.005059
[trn] ep 2, step 750, loss 0.464381, accu 0.703125, sec/iter 0.014
379
[tst] ep 2, step 1014, accu 0.820312, sec/iter 0.004900
[trn] ep 3, step 1250, loss 0.284551, accu 0.820312, sec/iter 0.01
4348
[tst] ep 3, step 1521, accu 0.859375, sec/iter 0.004766
[trn] ep 4, step 1750, loss 0.216922, accu 0.828125, sec/iter 0.01
5113
[tst] ep 4, step 2028, accu 0.882812, sec/iter 0.004711
[trn] ep 5, step 2250, loss 0.180984, accu 0.851562, sec/iter 0.01
4021
[tst] ep 5, step 2535, accu 0.890625, sec/iter 0.005294
[trn] ep 6, step 2750, loss 0.158896, accu 0.835938, sec/iter 0.01
5809
[tst] ep 6, step 3042, accu 0.882812, sec/iter 0.004615
[trn] ep 7, step 3250, loss 0.140687, accu 0.875000, sec/iter 0.01
5920
[tst] ep 7, step 3549, accu 0.890625, sec/iter 0.007297
[trn] ep 8, step 3750, loss 0.126121, accu 0.898438, sec/iter 0.01
8148
[tst] ep 8, step 4056, accu 0.906250, sec/iter 0.005170
[trn] ep 9, step 4250, loss 0.115712, accu 0.914062, sec/iter 0.01
3839
[tst] ep 9, step 4563, accu 0.906250, sec/iter 0.004460
[trn] ep 10, step 4750, loss 0.106273, accu 0.914062, sec/iter 0.0
13993
[tst] ep 10, step 5070, accu 0.906250, sec/iter 0.004573
```

In [16]:
```
# !tensorboard --ip 0.0.0.0 --logdir logdir
```

Test #4 - LSTMCell + forget_bias

```
cell = tf.contrib.rnn.BasicLSTMCell(
        num_hidden_units,
        forget_bias=lstm_forget_bias)
```

In [17]:
```
tf.reset_default_graph()

inputs_ = tf.placeholder(tf.float32,
                         [BATCH_SIZE, MAX_SEQ_LEN, INPUT_UNITS],
                         name='inputs')
labels_ = tf.placeholder(tf.int64,
                         [BATCH_SIZE],
                         name='labels')

lstm_with_forget_bias = lambda num_hidden_units: \
    tf.contrib.rnn.BasicLSTMCell(
            num_hidden_units,
            forget_bias=5.0)

model = MnistRnn(inputs_,
                 labels_,
                 INPUT_UNITS,
                 NUM_HIDDEN_UNITS,
                 BATCH_SIZE,
                 MAX_SEQ_LEN,
                 lstm_with_forget_bias)
```

In [18]:
```
config = tf.ConfigProto(gpu_options={'allow_growth':True})
sess = tf.InteractiveSession(config=config)

tf.global_variables_initializer().run()

train_writer = tf.summary.FileWriter('logdir/train_lstm_forget_bias
',
                                     graph=tf.get_default_graph())
test_writer  = tf.summary.FileWriter('logdir/test_lstm_forget_bias'
,
                                     graph=tf.get_default_graph())
```

In [19]:
```
train(inputs_, labels_, 10, train_writer, test_writer)
```

```
[trn] ep 1, step 250, loss 1.470503, accu 0.054688, sec/iter 0.013
168
[tst] ep 1, step 507, accu 0.617188, sec/iter 0.005363
[trn] ep 2, step 750, loss 0.543548, accu 0.664062, sec/iter 0.013
188
[tst] ep 2, step 1014, accu 0.703125, sec/iter 0.007593
[trn] ep 3, step 1250, loss 0.369156, accu 0.757812, sec/iter 0.01
5381
[tst] ep 3, step 1521, accu 0.796875, sec/iter 0.005513
[trn] ep 4, step 1750, loss 0.286988, accu 0.828125, sec/iter 0.01
3617
[tst] ep 4, step 2028, accu 0.820312, sec/iter 0.005127
[trn] ep 5, step 2250, loss 0.240008, accu 0.851562, sec/iter 0.01
5156
[tst] ep 5, step 2535, accu 0.835938, sec/iter 0.005836
[trn] ep 6, step 2750, loss 0.208887, accu 0.867188, sec/iter 0.01
3542
[tst] ep 6, step 3042, accu 0.859375, sec/iter 0.005498
[trn] ep 7, step 3250, loss 0.183400, accu 0.882812, sec/iter 0.01
5910
[tst] ep 7, step 3549, accu 0.851562, sec/iter 0.007850
[trn] ep 8, step 3750, loss 0.164586, accu 0.882812, sec/iter 0.01
4953
[tst] ep 8, step 4056, accu 0.867188, sec/iter 0.006749
[trn] ep 9, step 4250, loss 0.149524, accu 0.890625, sec/iter 0.01
5110
[tst] ep 9, step 4563, accu 0.882812, sec/iter 0.006931
[trn] ep 10, step 4750, loss 0.136078, accu 0.898438, sec/iter 0.0
15293
[tst] ep 10, step 5070, accu 0.898438, sec/iter 0.008359
```

In [20]:
```
# !tensorboard --ip 0.0.0.0 --logdir logdir
```

Test #5 - LayerNormBasicLSTMCell

```
cell = tf.contrib.rnn.LayerNormBasicLSTMCell(
                    num_hidden_units)
```

In [21]:
```
tf.reset_default_graph()

inputs_ = tf.placeholder(tf.float32,
                         [BATCH_SIZE, MAX_SEQ_LEN, INPUT_UNITS],
                         name='inputs')
labels_ = tf.placeholder(tf.int64,
                         [BATCH_SIZE],
                         name='labels')

model = MnistRnn(inputs_,
                 labels_,
                 INPUT_UNITS,
                 NUM_HIDDEN_UNITS,
                 BATCH_SIZE,
                 MAX_SEQ_LEN,
                 tf.contrib.rnn.LayerNormBasicLSTMCell)
```

In [22]:
```
config = tf.ConfigProto(gpu_options={'allow_growth':True})
sess = tf.InteractiveSession(config=config)

tf.global_variables_initializer().run()

train_writer = tf.summary.FileWriter('logdir/train_ln_basic_lstm',
                                     graph=tf.get_default_graph())
test_writer  = tf.summary.FileWriter('logdir/test_ln_basic_lstm',
                                     graph=tf.get_default_graph())

train(inputs_, labels_, 1, train_writer, test_writer)
```

```
[trn] ep 1, step 250, loss nan, accu 0.039062, sec/iter 0.063241
[tst] ep 1, step 507, accu 0.054688, sec/iter 0.011702
```

In [23]:
```
# !tensorboard --ip 0.0.0.0 --logdir logdir
```

Test #6 - LayerNormBasicLSTMCell - what's wrong?

- tf.check_numerics()

  > When run, reports an `InvalidArgument` error if tensor has any values that are not a number (NaN) or infinity (Inf). Otherwise, passes tensor as-is.

```
check_numerics(
    tensor,
    message,
    name=None
)
```

- tf.gradients()

  > Constructs symbolic partial derivatives of sum of `ys` w.r.t. x in `xs`

```
gradients(
    ys,
    xs,
    grad_ys=None,
    name='gradients',
    colocate_gradients_with_ops=False,
    gate_gradients=False,
    aggregation_method=None
)
```

- 사용 예:

```
# check_numerics
self.check = [tf.check_numerics(t,
                'check_numerics: {}'.format(t.name)) \
              for t in tf.gradients(
                          loss,
                          tf.trainable_variables()) \
              if t is not None]
...
summary, optimize, loss, accuracy, _ = \
    sess.run([model.train_summary,
              model.optimize,
              model.loss,
              model.accuracy,
              model.check],
              ...
```

In [24]:
```
tf.reset_default_graph()

inputs_ = tf.placeholder(tf.float32,
                         [BATCH_SIZE, MAX_SEQ_LEN, INPUT_UNITS],
                         name='inputs')
labels_ = tf.placeholder(tf.int64,
                         [BATCH_SIZE],
                         name='labels')

model = MnistRnn(inputs_,
                 labels_,
                 INPUT_UNITS,
                 NUM_HIDDEN_UNITS,
                 BATCH_SIZE,
                 MAX_SEQ_LEN,
                 tf.contrib.rnn.LayerNormBasicLSTMCell,
                 add_check = True)

# 여기서 NaN 문제가 생기는 걸 확인했으면, 다음 방법들을 시도 해 보세요
#  1. learning_rate 를 줄여
#       e.g.: lr = 0.0001
#  2. 적당한 값으로 gradient clipping
#       e.g.: use_grad_clip = True
#  3. 사용한 컴퍼넌트에 smoothing 할 수 있는 파라메터가 있는지 확인하고 적용
#     rnn_cell_class = lambda num_hidden_units: \
#        tf.contrib.rnn.LayerNormBasicLSTMCell(
#                                num_hidden_units,
#                                norm_gain=0.85,
#                                norm_shift=0.15)
```

In [ ]:
```
config = tf.ConfigProto(gpu_options={'allow_growth':True})
sess = tf.InteractiveSession(config=config)

tf.global_variables_initializer().run()

train_writer = tf.summary.FileWriter(
                'logdir/train_ln_basic_lstm_2',
                graph=tf.get_default_graph())
test_writer  = tf.summary.FileWriter(
                'logdir/test_ln_basic_lstm_2',
                graph=tf.get_default_graph())

train(inputs_, labels_, 10, train_writer, test_writer)
```

``` InvalidArgumentError: check_numerics: gradients_1/rnn/while/rnn/layer_norm_basic_lstm_cell/state_1/batchnorm/sub/Enter_grad/b_acc_3:0 : Tensor had NaN values [[Node: CheckNumerics_10 = CheckNumerics[T=DT_FLOAT, message="check_numerics: gradients_1/rnn/while/rnn/layer_norm_basic_lstm_cell/state_1/batchnorm/sub/Enter_grad/b_acc_3:0", _device="/job:localhost/replica:0/task:0/cpu:0"] (gradients_1/rnn/while/rnn/layer_norm_basic_lstm_cell/state_1/batchnorm/sub/Enter_grad/b_acc_3)]] ```
```

In [26]:
```
# !tensorboard --ip 0.0.0.0 --logdir logdir
```

In [27]:
```
tf.reset_default_graph()

inputs_ = tf.placeholder(tf.float32,
                         [BATCH_SIZE, MAX_SEQ_LEN, INPUT_UNITS],
                         name='inputs')
labels_ = tf.placeholder(tf.int64,
                         [BATCH_SIZE],
                         name='labels')

model = MnistRnn(inputs_,
                 labels_,
                 INPUT_UNITS,
                 NUM_HIDDEN_UNITS,
                 BATCH_SIZE,
                 MAX_SEQ_LEN,
                 lambda num_hidden_units: \
                       tf.contrib.rnn.LayerNormBasicLSTMCell(
                                        num_hidden_units,
                                        norm_gain=0.85,
                                        norm_shift=0.15),
                 add_check = True,
                 lr = 0.0001,
                 use_grad_clip = True)
```

In [28]:
```
config = tf.ConfigProto(gpu_options={'allow_growth':True})
sess = tf.InteractiveSession(config=config)

tf.global_variables_initializer().run()

train_writer = tf.summary.FileWriter(
                'logdir/train_ln_basic_lstm_2',
                graph=tf.get_default_graph())
test_writer  = tf.summary.FileWriter(
                'logdir/test_ln_basic_lstm_2',
                graph=tf.get_default_graph())

train(inputs_, labels_, 10, train_writer, test_writer)
```

```
[trn] ep 1, step 250, loss 2.108183, accu 0.046875, sec/iter 0.104
071
[tst] ep 1, step 507, accu 0.421875, sec/iter 0.012975
[trn] ep 2, step 750, loss 1.525936, accu 0.429688, sec/iter 0.104
317
[tst] ep 2, step 1014, accu 0.515625, sec/iter 0.009750
[trn] ep 3, step 1250, loss 1.210164, accu 0.539062, sec/iter 0.10
3209
[tst] ep 3, step 1521, accu 0.601562, sec/iter 0.008972
[trn] ep 4, step 1750, loss 0.963128, accu 0.632812, sec/iter 0.08
8096
[tst] ep 4, step 2028, accu 0.671875, sec/iter 0.008804
[trn] ep 5, step 2250, loss 0.794114, accu 0.664062, sec/iter 0.08
7969
[tst] ep 5, step 2535, accu 0.687500, sec/iter 0.010850
[trn] ep 6, step 2750, loss 0.673972, accu 0.726562, sec/iter 0.09
1234
[tst] ep 6, step 3042, accu 0.726562, sec/iter 0.008794
[trn] ep 7, step 3250, loss 0.574158, accu 0.750000, sec/iter 0.09
2356
[tst] ep 7, step 3549, accu 0.742188, sec/iter 0.008645
[trn] ep 8, step 3750, loss 0.494440, accu 0.750000, sec/iter 0.08
8173
[tst] ep 8, step 4056, accu 0.757812, sec/iter 0.008989
[trn] ep 9, step 4250, loss 0.431050, accu 0.757812, sec/iter 0.08
3751
[tst] ep 9, step 4563, accu 0.789062, sec/iter 0.008605
[trn] ep 10, step 4750, loss 0.378663, accu 0.765625, sec/iter 0.0
85223
[tst] ep 10, step 5070, accu 0.804688, sec/iter 0.008551
```

정리해 봅시다

텐서플로우에서 지원하는 RNN Cell 유형

- tf.contrib.rnn.BasicRNNCell()
- tf.contrib.rnn.BasicLSTMCell()
- tf.contrib.rnn.LSTMCell()
- tf.contrib.rnn.GRUCell()
- tf.contrib.rnn.LayerNormBasicLSTMCell()

RNN 구성

- tf.nn.dynamic_rnn()

```
cell          = tf.contrib.rnn.BasicRNNCell(
                   num_hidden_units)
last, states = tf.nn.dynamic_rnn(
                        cell,
                        inputs,
                        sequence_length=sequence_length,
                        dtype=tf.float32)
```

실행중 NaN, Inf 등이 발생하면?

1. tf.check_numerics()로 문제가 되는 컴퍼넌트 파악
2. learning_rate 를 줄여본다
3. 적당한 값으로 gradient clipping 해 본다
4. 사용한 컴퍼넌트에 smoothing 할 수 있는 파라메터가 있는지 확인하고 적용해 본다

gradient clipping?

- tf.clip_by_global_norm

```
clip_by_global_norm(
    t_list,
    clip_norm,
    use_norm=None,
    name=None
)
```

- tf.gradients

```
gradients(
    ys,
    xs,
    grad_ys=None,
    name='gradients',
    colocate_gradients_with_ops=False,
    gate_gradients=False,
    aggregation_method=None
)
```

- optimizer.apply_gradients

```
apply_gradients(
    grads_and_vars,
    global_step=None,
    name=None
)
```

- 사용예:

```
tvars_     = tf.trainable_variables()
grads_, _  = tf.clip_by_global_norm(
                tf.gradients(loss,tvars_),
                5.0)
optimize   = tf.train.AdamOptimizer(
                learning_rate=learning_rate) \
                .apply_gradients(zip(grads_, tvars_))
```