

CSDN

博客 (http://feed.csdn.net/?ref=toolbar)学院 (http://edu.csdn.net/?ref=toolbar)

下载 (http://download.csdn.net/?ref=toolbar)更多 ▾

搜索

登录 (http://passport.csdn.net/account/login?ref=toolbar)注册 (http://passport.csdn.net/account/mobileregister?ref=toolbar&action=mobileRegister)

发布 (http://write.blog.csdn.net/postedited?ref=toolbar)活动 (http://activity?utm_source=csdnblog1)

一个基于TensorFlow的简单故事生成案例：带你了解LSTM

1789

数据娃握 (http://blog.cs...)

+关注

(http://blog.csdn.net/jdbc)

码云

原创93

粉丝353

喜欢170

51 (https://github.com/jifeng1989)

在深度学习中，循环神经网络（RNN）是一系列善于从序列数据中学习的神经网络。由于对长期依赖问题的鲁棒性，长短期记忆（LSTM）是一类已经有实际应用的循环神经网络。现在已有大量关于 LSTM 的文章和文献，其中推荐如下两篇：

- Goodfellow et.al.《深度学习》一书第十章：<http://www.deeplearningbook.org/>
- Chris Olah: 理解 LSTM: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

已存在大量优秀的库可以帮助你基于 LSTM 构建机器学习应用。在 GitHub 中，谷歌的 TensorFlow 在此文成文时已有超过 50000 次星，表明了其在机器学习从业者中的流行度。

与此形成对比，相对缺乏的似乎是关于如何基于 LSTM 建立易于理解的 TensorFlow 应用的优秀文档和示例，这也是本文尝试解决的问题。

假设我们想用 一个样本短故事来训练 LSTM 预测下一个单词，伊索寓言：

long ago , the mice had a general council to consider what measures they could take to outwit their common enemy , the cat . some said this , and some said that but at last a young mouse got up and said he had a proposal to make , which he thought would meet the case . you will all agree , said he , that our chief danger consists in the sly and treacherous manner in which the enemy approaches us . now , if we could receive some signal of her approach , we could easily escape from her . i venture , therefore , to propose that a small bell be procured , and attached by a ribbon round the neck of the cat . by this means we should always know when she was about , and could easily retire while she was in the neighbourhood . this proposal met with general applause , until an old mouse got up and said that is all very well , but who is to bell the cat ? the mice looked at one another and nobody spoke . then the old mouse said it is easy to propose impossible remedies .

Listing 1. 取自伊索寓言的短故事，其中有 112 个不同的符号。单词和标点符号都视作符号。

如果我们将文本中的 3 个符号以正确的序列输入 LSTM，以 1 个标记了的符号作为输出，最终神经网络将学会正确地预测下一个符号（Figure1）。

他的最新文章


更多文章 (http://blog.csdn.net/jdbc)

让PIP源使用国内镜像 (http://blog.csdn.net/jdbc/article/details/78510368)

TensorFlow下构建高性能神经网络模型 (http://blog.csdn.net/jdbc/article/details/78510019)

Python镜像&ubuntuTensorflow安装 (http://blog.csdn.net/jdbc/article/details/78509992)

90平米房屋装修



在线课程

SDCC 2017 前瞻技术实践线上峰会 (http://edu.csdn.net)

MYOM在美国点选酒类移动端的最佳实践 (http://edu.csdn.net/huiyiCourse/detail/608?utm_source=blog9)

C语言大型软件设计的面向对象 (http://edu.csdn.net/huiyiCourse/detail/594?utm_source=blog9)

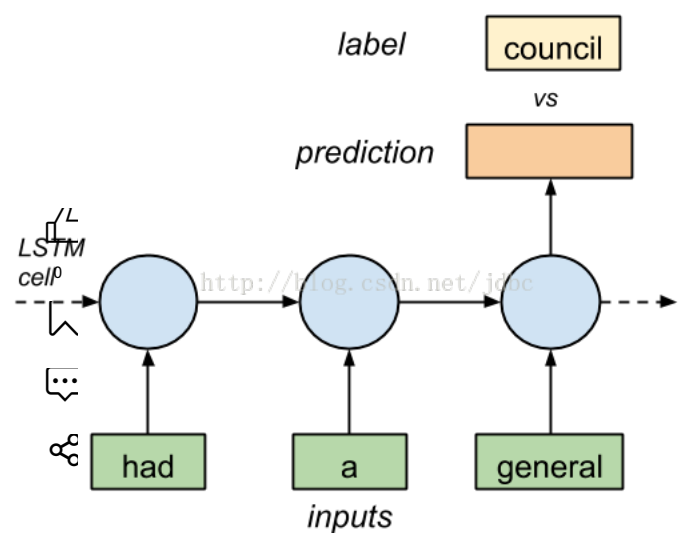


图 1. 有 3 个输入和 1 个输出的 LSTM 单元

严格说来，LSTM 只能理解输入的实数。一种将符号转化为数字的方法是基于每个符号出现的频率为其分配一个对应的整数。例如，上面的短文中有 112 个不同的符号。如列表 2 所示的函数建立了一个有如下条目 [「,」: 0][「the」: 1], ..., [「council」: 37], ..., [「spoke」= 111] 的词典。而为了解码 LSTM 的输出，同时也生成了逆序字典。

```
[python]
1. def build_dataset(words):
2.     count = collections.Counter(words).most_common()
3.     dictionary = dict()
4.     for word, _ in count:
5.         dictionary[word] = len(dictionary)
6.     reverse_dictionary = dict(zip(dictionary.values(), dictionary.keys()))
7.     return dictionary, reverse_dictionary
```

Listing 2. 建立字典和逆序字典的函数

类似地，预测值也是一个唯一的整数值与逆序字典中预测符号的索引相对应。例如：如果预测值是 37，预测符号便是「council」。

输出的生成看起来似乎简单，但实际上 LSTM 为下一个符号生成了一个含有 112 个元素的预测概率向量，并用 softmax() 函数归一化。有着最高概率值的元素的索引便是逆序字典中预测符号的索引值（例如：一个 one-hot 向量）。图 2 给出了这个过程。

热门文章

- TensorFlow入门教程集合 (http://blog.csdn.net/jdbc/article/details/8299) 29626
- VMware Workstation 12 安装 Ubuntu 16.04 LTS (http://blog.csdn.net/jdbc/article/details/8299) 19293
- 解决MyEclipse开启后总是不停的在Update index (http://blog.csdn.net/jdbc/article/details/8299) 12209
- 文本数据的机器学习自动分类方法(上) (http://blog.csdn.net/jdbc/article/details/8299) 9599
- tensorflow自学之前的bigpicture (http://blog.csdn.net/jdbc/article/details/8299) 9418

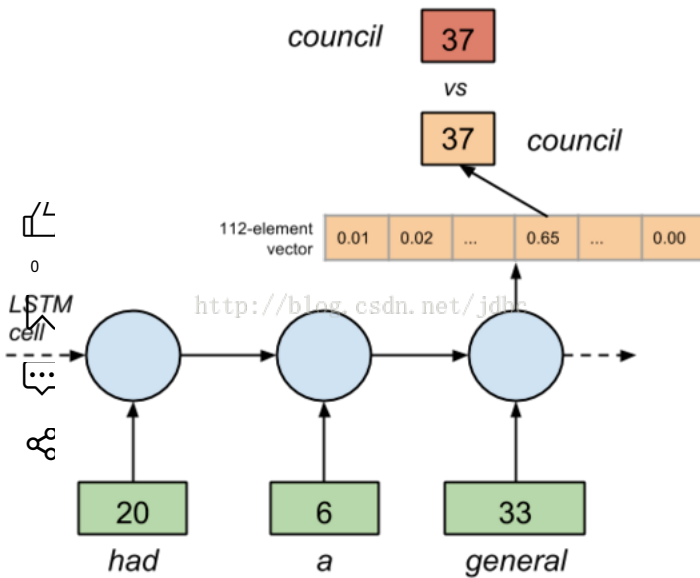


图2. 每一个输入符号被分配给它的独一无二的整数值所替代。输出是一个表明了预测符号在反向词典中索引的 **one-hot** 向量。

LSTM 模型是这个应用的核心部分。令人惊讶的是，它很易于用 TensorFlow 实现：

```
[python]
1. def RNN(x, weights, biases):
2.
3.     # reshape to [1, n_input]
4.     x = tf.reshape(x, [-1, n_input])
5.
6.     # Generate a n_input-element sequence of inputs
7.     # (eg. [had] [a] [general] -> [20] [6] [33])
8.     x = tf.split(x, n_input, 1)
9.
10.    # 1-layer LSTM with n_hidden units.
11.    rnn_cell = rnn.BasicLSTMCell(n_hidden)
12.
13.    # generate prediction
14.    outputs, states = rnn.static_rnn(rnn_cell, x, dtype=tf.float32)
15.
16.    # there are n_input outputs but
17.    # we only want the last output
18.    return tf.matmul(outputs[-1], weights['out']) + biases['out']
```

Listing 3. 有 512 个 LSTM 单元的网络模型

最难部分是以正确的格式和顺序完成输入。在这个例子中，LSTM 的输入是一个有 3 个整数的序列（例如：1x3 的整数向量）

网络的常量、权值和偏差设置如下：

```
[python]
1. vocab_size = len(dictionary)
2. n_input = 3
3.
4. # number of units in RNN cell
5. n_hidden = 512
6.
7. # RNN output node weights and biases
8. weights = {
9.     'out': tf.Variable(tf.random_normal([n_hidden, vocab_size]))
10. }
11. biases = {
```

```

12.         'out': tf.Variable(tf.random_normal([vocab_size]))
13.     }

```

Listing 4. 常量和训练参数

训练过程中的每一步，3 个符号都在训练数据中被检索。然后 3 个符号转化为整数以形成输入向量。

```

[python]
0.
1. symbols_in_keys = [ [dictionary[ str(training_data[i])]] for i in range(offset, offset+n_input) ]

```

Listing 5. 将符号转化为整数向量作为输入

训练标签是一个位于 3 个输入符号之后的 one-hot 向量。

```

[python]
1. symbols_out_onehot = np.zeros([vocab_size], dtype=float)
2. symbols_out_onehot[dictionary[str(training_data[offset+n_input])]] = 1.0

```

Listing 6. 单向量作为标签

在转化为输入词典的格式后，进行如下的优化过程：

```

[python]
1. _, acc, loss, onehot_pred = session.run([optimizer, accuracy, cost, pred], feed_dict={x: symbols_in_keys, y: symbols_out_onehot})

```

Listing 7. 训练过程中的优化

精度和损失被累积以监测训练过程。通常 50,000 次迭代足以达到可接受的精度要求。

```

[python]
1. ...
2. Iter= 49000, Average Loss= 0.528684, Average Accuracy= 88.50%
3. ['could', 'easily', 'retire'] - [while] vs [while]
4. Iter= 50000, Average Loss= 0.415811, Average Accuracy= 91.20%
5. ['this', 'means', 'we'] - [should] vs [should]

```

Listing 8. 一个训练间隔的预测和精度数据示例（间隔 1000 步）

代价是标签和 softmax() 预测之间的交叉熵，它被 RMSProp 以 0.001 的学习率进行优化。在本文示例的情况下，RMSProp 通常比 Adam 和 SGD 表现得更好。

```

[python]
1. pred = RNN(x, weights, biases)
2.
3. # Loss and optimizer
4. cost = tf.reduce_mean(tf.nn.softmax_cross_entropy_with_logits(logits=pred, labels=y))
5. optimizer = tf.train.RMSPropOptimizer(learning_rate=learning_rate).minimize(cost)

```

Listing 9. 损失和优化器

LSTM 的精度可以通过增加层来改善。

```
[python]
1. rnn_cell = rnn.MultiRNNCell([rnn.BasicLSTMCell(n_hidden),rnn.BasicLSTMCell(n_hidden)])
```

Listing 10. 改善的 LSTM

现在，到了有意思的部分。让我们通过将预测得到的输出作为输入中的下一个符号输入 LSTM 来生成一个故事吧。示例输入是「had a general」，LSTM 给出了正确的输出预测「council」。然后「council」作为新的输入「a general council」的一部分输入神经网络得到下一个输出「to」，如此循环下去。令人惊讶的是，LSTM 创作出了一个有一定含义的故事。

had a general council to consider what measures they could take to outwit their common enemy, the cat. some said this, and some said that but at last a young mouse got



Listing 11. 截取了样本故事生成的故事中的前 32 个预测值

如果我们输入另一个序列（例如：「mouse」，「mouse」，「mouse」）但并不一定是这个故事中的序列，那么会自动生成另一个故事。

mouse mouse mouse, neighbourhood and could receive a outwit always the neck of the cat. some said this, and some said that but at last a young mouse got up and said

Listing 12. 并非来源于示例故事中的输入序列

示例代码可以在这里找到：https://github.com/tifoft/Deep-Learning-Experiments/tree/master/Experiments/Tensorflow/RNN/rnn_words.py (https://github.com/tifoft/Deep-Learning-Experiments/tree/master/Experiments/Tensorflow/RNN/rnn_words.py)

示例文本的链接在这里：https://github.com/tifoft/Deep-Learning-Experiments/tree/master/Experiments/Tensorflow/RNN/belling_the_cat.txt (https://github.com/tifoft/Deep-Learning-Experiments/tree/master/Experiments/Tensorflow/RNN/belling_the_cat.txt)

小贴士：

1. 用整数值编码符号容易操作但会丢失单词的意思。本文中将符号转化为整数值是用来简化关于用 TensorFlow 建立 LSTM 应用的讨论的。更推荐采用 Word2Vec 将符号编码为向量。
2. 将输出表达成单向量是效率较低的方式，尤其当我们有一个现实的单词量大小。牛津词典有超过 170,000 个单词，而上面的例子中只有 112 个单词。再次声明，本文中的示例只为了简化讨论。
3. 这里采用的代码受到了 Tensorflow-Examples 的启发：https://github.com/aymericdamien/TensorFlow-Examples/blob/master/examples/3_NeuralNetworks/recurrent_network.py
4. 本文例子中的输入大小为 3，看一看当采用其它大小的输入时会发生什么吧（例如：4, 5 或更多）。
5. 每次运行代码都可能生成不同的结果，LSTM 的预测能力也会不同。这是由于精度依赖于初始参数的随机设定。训练次数越多（超过 150,000 次）精度也会相应提高。每次运行代码，建立的词典也会不同
6. Tensorboard 在调试中，尤其当检查代码是否正确地建立了图时很有用。
7. 试着用另一个故事测试 LSTM，尤其是用另一种语言写的故事。

原文链接：<https://medium.com/towards-data-science/lstm-by-example-using-tensorflow-feb0c1968537>



相关文章推荐



tensorflow笔记：多层LSTM代码分析 (<http://blog.csdn.net/u014595019/article/details/52...>)

tensorflow笔记系列：（一） tensorflow笔记：流程，概念和简单代码注释（二） tensorflow笔记：多层CNN代码分析之前讲过tensorflow中CNN的示例代码：...



u014595019 (<http://blog.csdn.net/u014595019>) 2016年10月08日 17:33 150809



tensorflow学习笔记（六）： LSTM 与 GRU (<http://blog.csdn.net/u012436149/article/det...>)

LSTM & GRU基本LSTMtensorflow提供了LSTM实现的一个basic版本，不包含lstm的一些高级扩展，同时也提供了一个标准接口，其中包含了lstm的扩展。分别为：tf.nn.rnn...



u012436149 (<http://blog.csdn.net/u012436149>) 2016年10月21日 21:20 16916



11.11 程序员该买些什么犒劳自己？

这是一份11.11程序员版爆款剁手清单！万万没想到，他们买了这些...

(http://www.baidu.com/cb.php?c=lgF_pyfqhHmknjDLnjT0lZ0qnfK9ujYzP1nsrjDz0Aw-5Hc3rHnYnHb0TAq15HfLPWRznjb0T1YzuH6kPW63Pvn3PH64nWbk0AwY5HDdnH0vnWRzP1D0lgF_5y9YIZ0lQzq-uZR8mLPbUB48ugfEXyN9T-KzUvdEIA-EUBqbugw9pysEn1qdlAdxTvqdThP-5yF_UvTkn0KzujY4rHb0mhYqn0KsTWYs0ZNGujYkPHTYn1mk0AqGujYknWb3rjDY0APGujYLnWm4n1c0ULI85H00TZbqnW0v0APzm1YvPj0sPs)

详细解读简单的lstm的实例 (<http://blog.csdn.net/zjm750617105/article/details/51321889>)

本文是初学keras这两天来，自己仿照addition_rnn.py，写的一个实例，数据处理稍微有些不同，但是准确性相比addition_rnn.py 差一点，下面直接贴代码，解释和注释都在代码里...



zjm750617105 (<http://blog.csdn.net/zjm750617105>) 2016年05月05日 15:31 11452

TensorFlow的55个经典案例 (http://blog.csdn.net/xzy_thu/article/details/76220654)

导语: 本文是TensorFlow实现流行机器学习算法的教程汇集，目标是让读者可以轻松通过清晰简明的案例深入了解 TensorFlow。这些案例适合那些想要实现一些 TensorFlow 案例的初学者...



xzy_thu (http://blog.csdn.net/xzy_thu) 2017年07月27日 20:51 1774

TensorFlow使用的一个小案例 (<http://blog.csdn.net/DilemmaVF/article/details/66476862>)

TensorFlow使用的一个小案例上一篇博客我们已经安装了TensorFlow，有很多人可能会对机器学习感到害怕，但其实有好多原理我们现在不需要懂，我们先搞清楚TensorFlow有什么作用就行，那...



DilemmaVF (<http://blog.csdn.net/DilemmaVF>) 2017年03月26日 16:56 3764

人人都能看懂的 AI 入门课

本课程将讲述人工智能的现状、应用场景和入门方法，并通过运用 TensorFlow，使得受众能清晰了解人工智能的运作方



式。

(http://www.baidu.com/cb.php?c=lgF_pyfqHmknjcvPjm0lZ0qnfK9ujYzP1ndnHfz0Aw-5Hc4nj6vPjm0TAq15Hf4rjn1n1b0T1YsnAfzPHu-nWNBry6dmvfk0AwY5HDdnH0vnWRzP1c0lgF_5y9YIZ0lQzqMpgwBUvqoQhP8QvIGIAPCmgfEmvq_lyd8Q1R4uWI-n16kPWK6WrHnvnHRvvnNBuD4PHqdlAdxTvqdThP-5HDknHK4mhkEusKzujY4rHb0mhYqn0KsTWYs0ZNGujYkPHTYn1mk0AqGujYkn10snjf10APGujYLnWm4n1c0ULI85H00TZbqnW0v0APzm1Y1rjmYnf)

数十种TensorFlow实现案例汇集：代码+笔记 (<http://blog.csdn.net/baolibin528/article/d...>)

这是使用 TensorFlow 实现流行的机器学习算法的教程汇集。本汇集的目标是让读者可以轻松通过案例深入 TensorFlow。这些案例适合那些想要清晰简明的 TensorFlow 实现案例...

baolibin528 (<http://blog.csdn.net/baolibin528>) 2017年01月12日 11:05 1604

数十种TensorFlow实现案例汇集：代码+笔记 (<http://blog.csdn.net/zouyu1746430162/ar...>)

//QQ: 174643016 QQ群: 532629018 CAFFE 深度学习交流 // <http://bbs.21ic.com/icview-759778-1-1.html> (项目展示链接)

zouyu1746430162 (<http://blog.csdn.net/zouyu1746430162>) 2016年11月23日 18:21 2806

数十种TensorFlow实现案例汇集：代码+笔记 (<http://blog.csdn.net/dj0379/article/details...>)

2016-10-17 13:08:31 开源 0 0 0 这是使用 TensorFlow 实现流行的机器学习算法的教程汇集。本汇集的目标是让读者可以轻松通过案例...

dj0379 (<http://blog.csdn.net/dj0379>) 2016年10月18日 17:48 3109

TensorFlow实现案例汇集：代码+笔记 (<http://blog.csdn.net/Andrewseu/article/details/6...>)

选自 Github 机器之心编译参与：吴攀、李亚洲这是使用 TensorFlow 实现流行的机器学习算法的教程汇集。本汇集的目标是让读者可以轻松通过案例深入 TensorFlow...

Andrewseu (<http://blog.csdn.net/Andrewseu>) 2017年03月09日 15:18 806

数十种TensorFlow实现案例汇集：代码+笔记 (http://blog.csdn.net/mydear_11000/articl...)

这些案例适合那些想要清晰简明的 TensorFlow 实现案例的初学者。本教程还包含了笔记和带有注解的代码。项目地址：<https://github.com/aymericdamien/Te...>

mydear_11000 (http://blog.csdn.net/mydear_11000) 2016年10月21日 09:08 3636



深度学习-LSTM网络-代码-示例 ([u012609509/article/details/51910405](http://blog.csdn.net/u012609509/article/details/51910405))

一、LSTM网络原理要点介绍 (1) LSTM网络用来处理带“序列”(sequence)性质的数据，比如时间序列的数据，像每天的股价走势情况，机械振动信号的时域波形，以及类似于自然语言这种本身带有...

u012609509 (<http://blog.csdn.net/u012609509>) 2016-07-14 16:20 6025

LSTM实现详解 (/real_myth/article/details/51275869)



LSTM实现详解发表于2015-09-14 16:58| 5021次阅读| 来源Apaszke Github| 3 条评论| 作者Adam Paszke LSTM神经网络R N...

 Real_Myth (http://blog.csdn.net/Real_Myth) 2016-04-28 16:58  2421



TensorFlow入门和示例分析 (/dabokele/article/details/54293560)

本文以TensorFlow源码中自带的手写数字识别Example为例，引出TensorFlow中的几个主要概念。并结合Example源码一步步分析该模型的实现过程。一、什么是TensorFlow 在...

 dabokele (<http://blog.csdn.net/dabokele>) 2017-01-09 22:01  15155





Tensorflow实例：利用LSTM预测股票每日最高价（一） (/mylove0414/article/details/55...

RNN与LSTM这一部分主要涉及循环神经网络的理论，讲的可能会比较简略。什么是RNNRNN全称循环神经网络（Recurrent Neural Networks），是用来处理序列数据的。在传统的神经网络...

 mylove0414 (<http://blog.csdn.net/mylove0414>) 2017-02-19 17:28  23800

TensorFlow学习笔记之三——适合入门的一些资源 (/snsn1984/article/details/51371970)

Github上除了有TensorFlow的源码库之外，还有一些很不错的适合入门的资源。现在将目前已经解除到的资源整理出来，和大家分享。1、TensorFlow源码库<https://github.com...>

 sns1984 (<http://blog.csdn.net/snsn1984>) 2016-05-11 10:16  16739



TensorFlow最佳实践样例程序 (/nnnnnnnnnnny/article/details/70199530)

此例程出自《TensorFlow实战Google深度学习框架》第5章 MNIST数字识别问题，很好的一个例子。工程目录：-mnist_best-dataset //存放数据集的文件夹，可以...

 NNNNNNNNNNNY (<http://blog.csdn.net/NNNNNNNNNNNY>) 2017-04-16 23:23  3428



LSTM实现详解 (/zdy0_2004/article/details/49977423)

<http://dataunion.org/20778.html> 相关阅读：深入浅出LSTM神经网络前言在很长一段时间里，我一直忙于寻找一个实现LSTM网络的好教程。它们似乎很复杂，而...

 zdy0_2004 (http://blog.csdn.net/zdy0_2004) 2015-11-22 14:13  21445



数十种TensorFlow实现案例汇集 (/zhaoyu106/article/details/53157818)

这是使用 TensorFlow 实现流行的机器学习算法的教程汇集。本汇集的目标是让读者可以轻松通过案例深入 TensorFlow。这些案例适合那些想要清晰简明的 TensorFlow 实现案...

 zhaoyu106 (<http://blog.csdn.net/zhaoyu106>) 2016-11-14 12:35  2365



TensorFlow个人学习（下载安装&简单代码例子） (/q17686617156/article/details/72934...

（由于项目最终决定使用TensorFlow，所以博客来总结学习TensorFlow）下载和安装：由于TensorFlow也官方支持windows了，省去了很多功夫，安装就更简单了。Te...

 q17686617156 (<http://blog.csdn.net/q17686617156>) 2017-06-08 19:18  1128

Tensorflow-简单的例子 (/m0_37306360/article/details/74908947)

TensorFlow提供了优化器，可以缓慢地更改每个变量，以便最小化损失函数。最简单的优化器是梯度下降。它根据相对于该变量的损失导数的大小修改每个变量。通常，手动计算符号导数是冗长乏味且容易出错的。因...

 m0_37306360 (http://blog.csdn.net/m0_37306360) 2017-07-10 09:54  325



0

