

词嵌入

下面我们看看 pytorch 中如何调用词向量

PyTorch 实现

词嵌入在 pytorch 中非常简单，只需要调用 `torch.nn.Embedding(m, n)` 就可以了，m 表示单词的总数目，n 表示词嵌入的维度，其实词嵌入就相当于是一个大矩阵，矩阵的每一行表示一个单词

```
import torch
from torch import nn
from torch.autograd import Variable
```

```
# 定义词嵌入
embeds = nn.Embedding(2, 5) # 2 个单词，维度 5
```

```
# 得到词嵌入矩阵
embeds.weight
```

```
Parameter containing:
-1.3426  0.7316 -0.2437  0.4925 -0.0191
-0.8326  0.3367  0.2135  0.5059  0.8326
[torch.FloatTensor of size 2x5]
```

我们通过 `weight` 得到了整个词嵌入的矩阵，注意，这个矩阵是一个可以改变的 parameter，在网络的训练中会不断更新，同时词嵌入的数值可以直接进行修改，比如我们可以读入一个预训练好的词嵌入等等

```
# 直接手动修改词嵌入的值
embeds.weight.data = torch.ones(2, 5)
embeds.weight
```

```
Parameter containing:
  1  1  1  1  1
  1  1  1  1  1
[torch.FloatTensor of size 2x5]
```

```
# 访问第 50 个词的词向量
embeds = nn.Embedding(100, 10)
single_word_embed = embeds(Variable(torch.LongTensor([50])))
```

```
single_word_embed
```

```
Variable containing:
-1.4954 -1.8475  0.2913 -0.9674 -2.1250 -0.5783 -0.6717  0.5638  0.7038  0.4437
[torch.FloatTensor of size 1x10]
```

可以看到如果我们要访问其中一个单词的词向量，我们可以直接调用定义好的词嵌入，但是输入必须传入一个 Variable，且类型是 LongTensor

下次课，我们会讲一讲词嵌入到底有什么用。