



CSDN学院 IT实战派

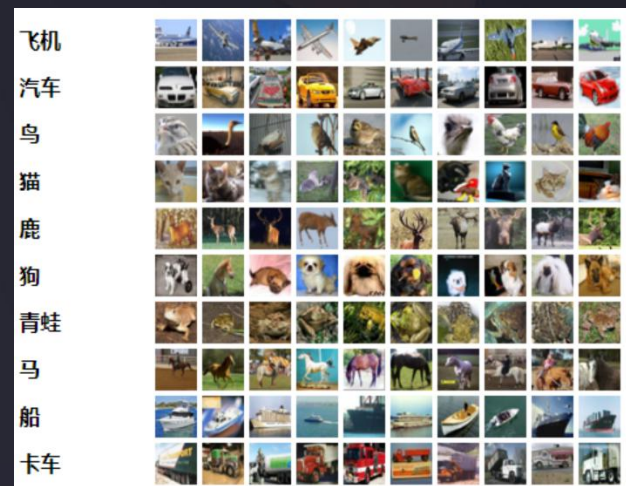
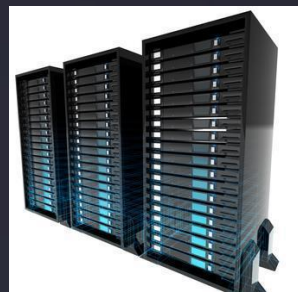
第十二章：自如应对之法-项目实战的经验4

上一章回顾

- 一个基于CIFAR数据集的计算视觉分类任务
- 数据增强和迁移学习
- 小数据量短周期训练



+



面试模拟：请介绍你经历过的一个实战项目？

如果是面试自然语言处理的岗位该如何介绍？

- 机器翻译
 - 智能客服
 - 语音合成
 - 知识库与知识图谱
 - 情感分析
 - 文本分类
- 等等

深度学习项目实战-网络评价的文本智能分析系统

电视比想象中的大好多，画面也很清晰，
✓ 系统很智能，更多功能还在摸索中

正面

送货时间非常长，外壳包装也损坏了，咨
✓ 询了客服完全没有反应

负面

LS TM + MLP

深度学习项目实战-案例实现重要代码讲解

标签及词汇表

```
labels, vocabulary = list(data['label'].unique()), list(data['evaluation'].unique())
```

构造字符级别的特征

```
string = ""
```

```
for word in vocabulary:
```

```
    string += word
```

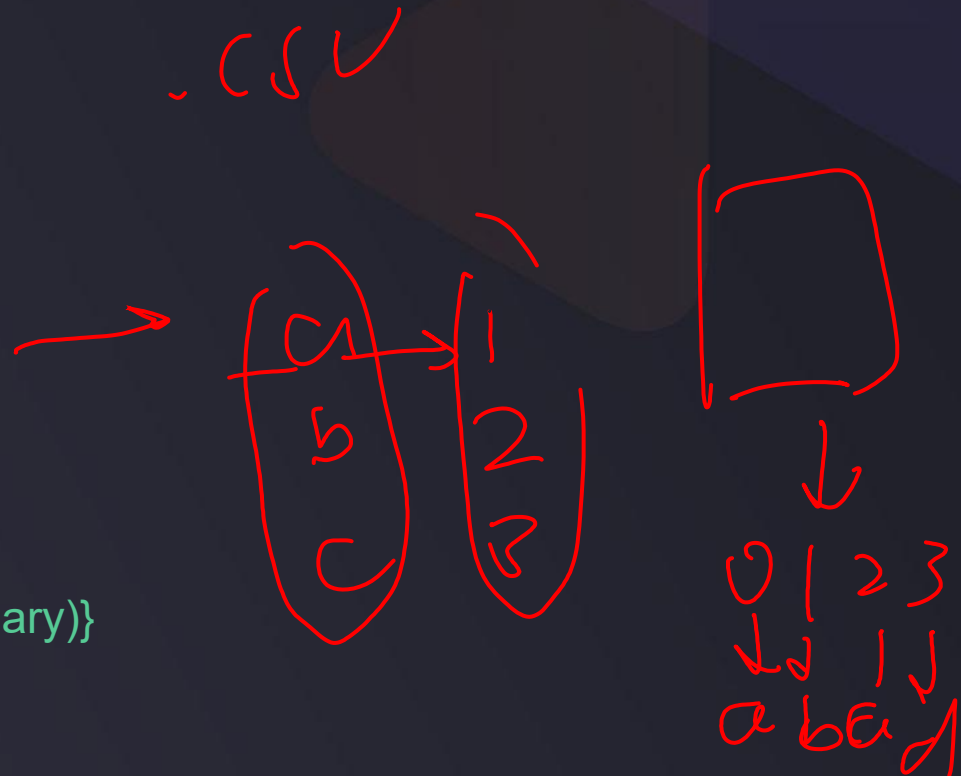
```
vocabulary = set(string)
```

制作字典列表，用在人机交互

```
word_dictionary = {word: i+1 for i, word in enumerate(vocabulary)}
```

```
inverse_word_dictionary = {i+1: word for i, word in enumerate(vocabulary)}
```

```
print(word_dictionary)
```



深度学习项目实战-案例实现重要代码讲解

#制作训练输入和训练标签

#对于训练输入要进行序列填充, 按input_shape填充, 长度不足的按0补充

```
x = [[word_dictionary[word] for word in sent] for sent in data['evaluation']]
```

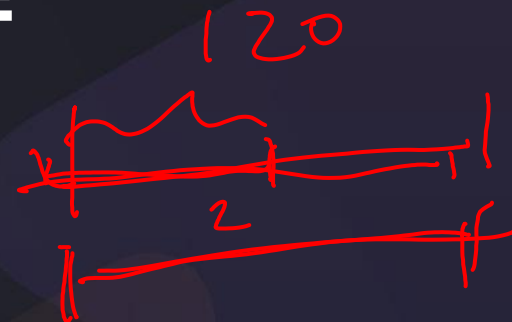
```
x = pad_sequences(maxlen=180, truncating="pre", sequences=x, padding='post', value=0)
```

#对于训练目标进行ONE-HOT-CODING

```
y = [[label_dictionary[sent]] for sent in data['label']]
```

```
y = [np_utils.to_categorical(label, num_classes=label_size) for label in y]
```

```
y = np.array([list(_[0]) for _ in y])
```



0 1 1

(2)

Handwritten diagram showing a sequence of 100 elements, with a bracket indicating a segment of 80 elements.

Handwritten diagram showing a sequence of 180 elements, with a bracket indicating a segment of 180 elements.

深度学习项目实战-案例实现重要代码讲解

#搭建模型

```
model = Sequential()
```

```
model.add(Embedding(input_dim=vocab_size + 1, output_dim=20, input_length=180, mask_zero=True))
```

```
model.add(LSTM(100))
```

```
model.add(Dropout(0.2))
```

```
model.add(Dense(label_size, activation='softmax'))
```

```
model.summary()
```

#编译模型和训练模型

```
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
```

```
model.fit(train_x, train_y, epochs=5, batch_size=32, validation_data=(test_x, test_y))
```

$m \times n \times k$

$n \times 20$

many to 1

A	1	1 0 0 0 0 0 2
B	2	0 (- - - -)
C	3	L
D	4	C
...
W	2000	

0.1 0.2 ... 0.2
0.2

过了若干时间后。

面试模拟：请介绍你经历过的一个实战项目？

1. **任务和模型定义：** 对于淘宝上面的网络评语进行智能分析
2. **数据处理与特征工程：** 对数据进行了序列填充，独热编码和word embedding
3. **模型搭建与训练：** 搭建了LSTM+MLP的模型，循环神经网络的类型是MANY TO ONE
4. **模型测试及优化：** 使用Adam优化器，通过了MINI-BATCH的方式进行训练，仅训练5个epoch就能达到90%+的准确度，而且用自己手打的评语进行测试表现也蛮不错的

面试模拟：自然语言处理岗位可能需要的其他知识点

- SGD, naive-bayes, GRU, 和LSTM
- LAMB, transformer, 和BERT
- ELMO 和GPT
- Attention注意力机制
- 序列模型
- Sequence2Sequence
- 语言模型（英文，中文）与句法分析

| 本章回顾

- 一个网络评价的文本智能分析系统
- Word Embedding
- LSTM+MLP

课程相关资料





欢迎大家扫码或者添加微信好友ai_flare（学习小助手），加入学习群，老师会在群里帮大家解答学习、职业发展与求职问题（名额有限、人满即止）