

品牌识别项目

问题

例如识别“后”这个品牌，规则匹
配比较困难

例句

伊丽莎白雅顿，绿茶蜂蜜身体
乳。真的！味道清新，上身不油
腻。后的洗面奶。Aesop二重奏
爽肤水。契尔氏高保湿霜。倩碧
黄油无油版。娇韵诗双管精华。
菲洛嘉眼霜。 我的肤质，混油偏
干。

模型需要这个句子中出现的品牌
词和位置：伊丽莎白雅顿，后，
Aesop，倩碧，娇韵诗，菲洛嘉

任务类型： 序列分类

Token级别分类，简单理解为对
一句话中的每个字进行分类

[https://zhuanlan.zhihu.com/p/
197488236](https://zhuanlan.zhihu.com/p/197488236)

任务类似论文

[https://zhuanlan.zhihu.com/p/
326135985](https://zhuanlan.zhihu.com/p/326135985)

[https://zhuanlan.zhihu.com/p/
307454757](https://zhuanlan.zhihu.com/p/307454757)

标注工具: label-studio

使用的模型

多任务模型

[https://zhuanlan.zhihu.com/p/
408851910](https://zhuanlan.zhihu.com/p/408851910)

[https://zhuanlan.zhihu.com/p/
383137481](https://zhuanlan.zhihu.com/p/383137481)

训练时模型输入

例如：['B-PIN', 'I-PIN', 'O', 'O',
'O', 'O', 'O', 'O', 'O'] ['欧', '树',
'小', '金', '油', '看', '着', '不', '错']

使用的标注方式BIO

B代表实体的开始，I表示实体的
中间或结尾，O表示其它实体

PIN是代表品牌，当然也可以添
加其它

模型的性能

f1 score达到0.9左右

f1 score定义 $f1 = \frac{2 \times P \times R}{P + R}$

P代表精确率（预测为正的样本
中有多少是真正的正样本），R
代表召回率（样本中的正例有多
少被预测正确了）

ACC: 86.254, F1: 89.967

模型需要改成预测是不是品牌词
的模式

给定句子+品牌词

预测在句子中是否是品牌词

如果句子中有多个相同的品牌
词，那么如果有一个是品牌词，
那么就给判断是

准确率：98.076, F1: 90.966

标签：是，否