

梁博云实习第三个项目：写一个文本特征提取代码，可以用任何语言，比谁的实现最简洁，代码最易懂。代码行数最少，可读性最强获胜。

<http://116.196.101.207:8080/pullword/>

特征提取配置文件见附件，我们举几个例子，和输出，帮助大家理解这个批评过程。

段落 1：这个老板不怎么骗人，但是一次上门体验的经历还是让我非常后悔。

首先这个把这个段落分为两个句子

1) 这个老板不怎么骗人，

【这个句子匹配了 不.骗人 -1 这个特征,其中点 (.) 通配了“怎么”】，因此输出 不.骗人 -1

2) 但是一次上门体验的经历还是让我非常后悔。

【这个句子匹配了 非常.后悔 5 这个特征】，因此输出 非常.后悔 5

段落 2：这块蛋糕没有异味。

这个句子既可以匹配（没.异味），也似乎可以匹配（有.异味）但是按照匹配规则，没.异味在前面，一旦匹配了没.异味就不会再向后匹配，所以这个句子只匹配（没.异味）输出 没.异味

再给大家几个样例句子，和输出，供大家检查算法鲁棒性。

段落 3：这个电饭煲感觉不怎么有惊喜，买来以后感觉很后悔。

输出：不.有.惊喜 1；后悔 2

段落 4：这台车的发动机有很明显的且让人厌烦的异响，我本来高高兴兴的，结果让我喜爱不起来

输出：有.异响 1；本来.高兴 1;喜爱-1

段落 5：我真是后悔不早点买下来，不是我不爱你，是因为一直没赚到钱。

输出：悔.不.早.买 -5；不.不.爱你 -0.01

段落 6：

输出：~~非常.放心~~-1.5; ~~没.问题~~-1

输出：没.问题 -1; 非常.放心 -1.5

注：一个句子可能匹配不同配置行的特征，但是每一配置行只要匹配到一个特征就不往后匹配。

本作业为 A 班 B 班的共同作业，A 班和 B 班的提交时间均为 5 月 22 日

### 【附件】

特征提取配置文件：

特征 1[空格]权重 2:特征 2[空格]权重 2: 特征 3[空格]权重 3...特征 n[空格]权重 n

冒号作为区分，以句子为单位，从第一行开始匹配，按行匹配，直到匹配完毕，在每行的批评中如果符合了特征 x，则提取对应的权重 x，输出，并且不再匹配 x+1 以后的特征。

特中的点号(.)匹配中间出现的所有字符。

不骗人 -1:不.骗人 -1:骗人 1

悔.没.早.买 -5:悔.不.早.买 -5:别.后悔 -0.5:没.后悔 -1:不.后悔 -1:后悔.买.晚 -2:后悔.入.晚

-2:后悔.没 -0.5:非常.后悔 5:后悔 2

耗电.高 1:耗电.大 1:耗电.多 1:很.耗电 1:非常.耗电 2:不.耗电 -1:耗电量 0.00001:耗电 1

没想.有.惊喜 -1:没有.惊喜 1:不.有.惊喜 1:有.惊喜 -2:惊喜 -1

没.毛病 -2:不.毛病 -1:毛病 1

不漂亮 2:很漂亮 -2:漂亮 -1

没.问题 -1:所有.问题 0.1:有.问题.可 -0.1:有.问题 1

充电.不.烫 -1:充电.烫 1

不.延迟 -1:延迟 1

没.掉色 -1:不.掉色 -1:掉色 1

不.显瘦 0.1:显瘦 -1

不厚实 1:不.厚实 1:厚实 -1

没.异味 -1:有.异味 1:无.异味 -1:除.异味 -0.5:分解.异味 -0.5

挺.放心 -1:非常.放心 -1.5:不.放心 1:很.放心 -1:放心 -0.5

不.紧绷 -1

没.隔阂 -1:不.隔阂 -1:隔阂 1

非常.享受 -2:不.享受 1:没.享受 1:很.享受 -1.2:享受 -1

无.异响 -1:没.异响 -1:有.异响 1:异响 0.5

真.不行 1:真.行 -1

真.不.脆 1:真.脆 -1

有.耐心 -1:没.耐心 0.5

不喜爱 1:不.喜爱 1:喜爱 -1:喜爱.不起来 2

不幸福 1:不.幸福 1:很.幸福 -2

可不可爱 0.1:不.可爱 1:很.可爱 -2:可爱 -0.5

不.不.爱你 -0.01:不.爱你 0.5:爱你 -0.5

不自私 -0.5:不.自私 -0.5:自私 1

很不好 1:不.很好 0.6:没.很好 0.2:很好 -0.4:不好意思 0.1:不好 0.3

不幸 1.5

不美好 1:不.美好 1:美好 -1

不.准备.入手 1:准备.入手 -1

不是.正品 1:不.正品 0.5:无.正品 0.5:没.正品 0.5:是.正品 -0.5:正品 -0.5

不.担心 -1:担心 0.3

没.注意 0.5

不会.好评 0.5:还想.好评 0.5:好评.刷 1:好评 -1

没.色差 -1:有.色差 1:色差.小 -0.3:色差.不大 -0.2:色差 1