

自然语言处理

Project 1 Task 1

院系：人工智能学院

姓名：王蔚昕

学号：211300042

2024 年 4 月 21 日



目录

1 子问题 1	3
1.1 难点	3
1.2 实验	3

1 子问题 1

1.1 难点

1.1.1 言语与性格的相关度

首先必须承认的是，言语一定与性格相关，但是在这里要说的事，言语与性格能有多相关？因为我认为语言对于性格来讲并非一个强有力的证明因素，他可能只是一个更容易收集到的因素，而且与性格有关联。所以我提出的猜想是，仅仅凭借言语可能不足以完全确定一个人的性格。一个人格可能有代表性行为，代表性处理事情的偏向，但是语言收到影响的东西太多了，同样的性格，在不同的环境中成长表现出来的言语可能是不同的，甚至截然相反。因为人们对于语言的学习大多数来自于身边的人，说话习惯也可能和身边的某个人或某些人有相似之处。所以我决定验证不同性格人之间言语的相似度。

1.1.2

1.2 实验

1.2.1 实验方法方法


为了调查不同性格之间的话语相似度，我们预计统计两种完全对立的性格中使用的单词数目，计算比例，然后用内积来判断这两种性格语言的相似度。

1.2.2 实验结果

实验结果表明，哪怕是完全相反的两种人格他们的语言之间的相似度也高达 98%。

1.2.3 实验结果分析

通过以上实验结果，可以发现的是，不同人格之间说的话或许并没有显著差距，或者说这个差距很小，这一点可能会降低神经网络分类器的性能。这一点更加证明了 KNN 在实验中取得效果并不好的原因，因为 KNN 的度量函数采用普通的二范数来度量，而已经证明了话语之间的不同并不完全来自性格，可能更多来源于个人，所以这种表面的学习方法不能很好理解话

A cluster of overlapping translucent squares in light blue and light green colors, located in the top right corner of the page.

语深处可能的表达不同，导致 KNN 效果远远低于神经网络。神经网络可能在语言理解上超过了 KNN，但是，但是由于语言和性格之间关联并非十分强烈，所以哪怕是有能力理解语义的神经网络也不能完美区分性格。