

自然语言处理

Project 1 Task 2

院系：人工智能学院

姓名：王蔚昕

学号：211300042

2024 年 4 月 23 日

目录

1 子问题 1	3
1.1 难点	3
2 子问题 2	3
2.1 实验一	3
2.2 实验二	5

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 分类方法不合理

想到这一点的原因是，在 KNN 的实验中，因为我采用了将性格分为四类分别预测的方式，最后发现哪怕 KNN 这种相对原始的方法对于每一种人格预测的分辨率也能达到 60% 以上，如果使用神经网络进行这种程度的二分类问题，想必效果会出乎意料的好。所以我认为，这种过于细致的分类可能本身之间也不存在明显地界限，结合我经历过的 MBTI 性格测试题目以及最终得出结论的方法来看，这种方法只是给每个题目的选择在不同维度打分，综合计算各个维度的分数而已，所以四个维度有明显清晰地界限的也仅仅是他们对应的两个性格而已，直接进行 16 种性格的分类似乎是不合理的，于是我决定对于这一点展开实验。

2 子问题 2

2.1 实验一

2.1.1 实验方法

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。

2.1.2 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%。

2.1.3 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。

2.1.4 一些思考

出现这个结果也不难预料，因为总有一些高频词汇，不论什么人说话总要用到，这样的词汇往往参考价值不大，结合上一个报告中的思考，面临一个境地，数量稀少的词汇到底应该怎么认定？这些词汇是只有这类人才会说的还是说这些词汇只是这几个人会说的？如果是前者无疑可以对模型起到很好地作用，但是后者反而可能对预测造成干扰。当然，这是基于词频统计的传统学习方法来讲，深度学习的问题可能要涉及语义理解了。但是语义就一定可以表达一个人的性格吗？一个极端的例子，如果所有人围绕 $1+1$ 等于几这个问题来讨论，那么神经网络找到的语义一定是大差不差的，截然不同的语义可能是某些调皮的性格的人，也可能的的确是所谓的“九漏鱼”。所以，我还是认为，仅仅用文本来判断区分如此细致的 MBTI 性格划分是不太合理的。

2.2 实验二

2.2.1 实验方法

为了明确一种相对独立的分类，我首先统计了数据集的各种单独人格的比例，找到两个对立且数量相似的人格分别是 J 和 P, 随后将这两个类别视作二分类任务，使用 Bert 模型进行训练，然后观测实验结果。

2.2.2 实验结果

实验结果表明，单独统计这两个类别效果并非达到预期，在 100 样例，100 验证集的情况下，仅仅达到 54% 的准确率，事实上这个准确率并不如 KNN。

2.2.3 实验结果分析

这个实验表明，猜想可能是错误的，或者说 JP 的分类仍然不够确切。但是也不能排除实验因素，因为仅仅使用了 100 样例，经测试 1000 样例的神经网络 8 个小时只能完成一个 EPOCH 的百分之十四，故选取了 100 样例作为代表性实验。预测效果不够好的可能也包括了实验无法完全进行。

2.2.4 一些思考

首先在实验中我犯了一个很简单的缓存错误，预实验内容无关，这个错误曾经严重影响了实验的效率。

其次，发现了一个在 tsak-1 中出现的一个问题，已经在 github 中完成代码的修正。

在进行完实验后，我确信神经网络分类效果一般的原因是样本问题，因为一个简单的二分类问题神经网络没有理由不如 KNN，但是需要大量样本也说明了这两个性格标签的差距似乎也不是很大，于是我回顾了 task-1 中 KNN 的分类结果，我发现 KNN 中 I, E 性格的分类准确度是遥遥领先与其他相对人格的分类器的，所以我认为，并不是所有人格都在言语上有重大区别，或者说不同方面的性格在言语上的差别不同，比如 IE 的区别可能会导致在语言上有更大的差距。所以证明了一点，用文本来分类 MBTI 性格的确不是很科学，因为某些方面的分类对于文本的影响可能很小。