**CQT专利文档说明**

**1、三类问题.：中性问题I ，准绳问题C，相关问题R是通过人工还是通过软件构建问题的组合方式**

答：通过软件构建的方式，但是需要满足一定规则。固定题位地方需要固定。其他的分为不同种类的题目数量，比如每道题目至少三个以上。其他位置的题目通过软件方式构建，如同类问题不得相邻。

**2、CQT测试法中题目的排序有着特殊要求，每一遍测试序列中，第一题，第二题最后一题为事件不相关的固定题位，同类问题不得相邻。**

答：“不得相邻”的要求，是为了能明确被测试者在面对不同问题时候的表情变化。比如位置三处为C准绳问题，用户做出了相关反应。那么题位四处应当选择R相关问题，或者I无关问题，以查看和问题三时候用户反应的差别。如果这三类问题让被测试者的心理活动发生变化，那么相邻题目属于不相同类别的要求，会使得用户有较明显的情绪变化。

**3、情绪和语音抽取成特征，例如说话声音是否颤抖，皱眉次数，眼睛向下看事件。特征维度有多高。**

答：数据集情况为三张csv表格，每张表格特征均为4170维度。R表中包含263个样本，C中561个样本，I中528个样本。三个数据集包含的4170个特征，为如下三类：眼睛转动的8个方向、微表情（来自于actionUnit URL: http://www.cs.cmu.edu/~face/facs.htm中抽取的65个表情特征）、头转动的方向。特征中的单位分为两类：特征发生的次数/回答问题的时长（次/s），特征发生的时间/回答问题的时长（归一化到0-1区间，无量纲）。

**4、降维降成多少维度，为什么选取tsne? 如采用其他降维方式，例如pca可以嘛。**

#### 答：为了方便可视化展示，目前降维成为2维。tsne主要的功能是将原先高纬度空间中的欧式距离转换为条件概率，然后通过Kl方式，转换为低纬度空间比如二维度，计算条件概率相类似的情况。选用tsne，因为python社区对于采用该方法降维的效果评价较高。且在其他数据集上通过tsne方法测试，取得的效果相比较其他降维方法的效果要好。Pca与其他的降维方式的效果相比较tsne较差一些。但是pca的执行速度较快，tsne经过多次迭代完成，计算速度较慢。

**5、使用kmeans聚类算法对中性问题和准绳问题进行分析。如何聚类。**

答：对于三类问题，中性问题I，准绳问题C，相关问题R，有不同的label 0,1,2。被测试者应该有不同的心理活动。已经得到的数据是被测试者的4170个特征。首先通过采用特征筛选方式，挑选出较具代表性的一组特征。再用tsne方法降维。针对低纬特征，同一类别的数据便是一个簇。

**6、假设降维后，特征空间为2维。那么这里特征1和特征2分别代表什么特征？**

答：通过tsne方法降维成2维，主要为了可是化展示。这里的2个特征是用原先4170维的特征向低维空间投影，所以维数减少了。在此过程中，特征发生了根本性的变化，原始的特征消失了。（虽然新的特征也保持了原特征的一些性质）。因而特征1与特征2并不像原先的4170个特征一样，有具体的物理意义。它们是从原始高维特征中提取出来，能较好表达原始特征的两个特征。

**7、使用kmeans算法，得到中性问题类别的cluster center坐标是（xi,yi）,如何理解一个问题的坐标。**

答：这里提及的聚类算法可能是kmeans。通过标签分类，采用均值方法计算每类样本点的中心，（xi,yi），（xc,yc）。测试新用户时，根据被测试用户针对相关问题的R的反应特征，分别计算相关问题(xr,yr)与两个cluster center的欧式距离。如果d\_ri>d\_rc。则贷款者有隐瞒倾向，相反则为真实回答。