**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。

**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：理论上说，pca适用于线性模型，tsne适用于非线性。实际情况

无论是从分析还是可视化展示，不方便。因而采用tsne的方式进行降维数据展示，但是处理成为2-3维并不能够作为特征进行后续处理（如果这里已经能够聚类了，那么和kmeans的聚类算法有什么区别）

目前降维成为2，是不是这边的components?pca与他的区别是什么晚点再看。其他的一些降维方法是不是可以适用。tsne的损失函数非凸，不同的初始值能得到不同的结果。

tsne有一系列参数，实验中经过tsne迭代次数为725次便终止。实验log中记录了迭代过程。并不是所有的参数都是725次，根据参数不同。

采用pca分析。pca无参数时候与加入components=2效果一致，这个说明需要转换到的维度是2维还是3维。pca的执行速度较快，tsne经过一次次迭代完成。

ipca的batch=10和78\*5=390效果好像差不多。

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。至此将维度降低了。但是这样子降维是否已经实现了聚类的功能？

Tsne在本例中的应用

这边的 tsne通过1355个维度为4170的数据，

sklearn中的tsne

**5、使用knn算法对中性问题和准绳问题进行聚类分析。如何聚类**

答：现在的特征数量比较多，有4170个特征， knn的话没有训练模型，直接拿测试数据去和当前维度的数据进行比较得出与之距离最近的k个邻居的分类情况，得出新有数据的分类情况。

**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：降维后的2维特征，可以看做是用最少的数据去模拟原先高维度的

**7、使用knn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。准绳问题c cluster center坐标(xc,yc),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。**

答：首先问题是，knn是一种分类算法，kmeans才是聚类算法，这边说的是不是kmeans？根据kmeans

knn是一种有监督学习，通过预先标记好的数据。然后计算