**部件/产品资源模型的检索模型使用说明**

1. 模型用途

给定一个部件/产品模型，获知制造该部件/产品所需的时间、设备、材料信息等。输入一个部件/产品模型对应的特征向量，输出该部件/产品所需的时间、设备、材料信息等（资源模型）。

1. 数据说明

输入数据需要两个文件，一个是模型相关数据库，共有m行n列，每一行代表一个模型，模型的前n列为模型特征向量在每个维度的值，接下来分为三个部分，第一个部分是时间，占一列，第二个部分和第三个部分分别是所需设备和原材料，每一列代表一种设备或材料，0代表需要该设备或材料，1代表不需要该设备或材料；另一个是需要进行检索部件的特征向量，共有一行，每一项之间用逗号分隔。

输出数据为k个部分，每一部分代表检索到的一个结果，及其所需的时间设备和原材料。

1. 算法原理

部件/产品资源模型的检索模型的核心思想是利用余弦相似度对模型之间的相似度进行度量，按相似度从大到小排列，返回靠前的结果，进行检索。余弦相似度，又称为余弦相似性，是通过计算两个向量的夹角余弦值来评估他们的相似度。对于两个向量，可以想象成空间中的两条线段，都是从原点（[0, 0, ...]）出发，指向不同的方向。两条线段之间形成一个夹角：如果夹角为0度，则意味着方向相同、线段重合；如果夹角为90度，意味着形成直角，方向完全不相似；如果夹角为180度，意味着方向正好相反。因此，可以通过夹角的大小，来判断向量的相似程度。夹角越小，就代表越相似。

1. 数据预处理

部件/产品资源模型的检索模型使用pandas工具将文件导入到DataFrame中,将特征向量和检索结果信息分隔开来，用特征向量求相似度后和结果信息重新连接，排序，得到最终结果。

1. 数学模型

对n维向量A，B，假设A= [A1, A2, ..., An] ，B= [B1, B2, ..., Bn] ，则A与B的夹角θ的余弦等于：



余弦值的范围在[-1,1]之间，值越趋近于1，代表两个向量的方向越接近；越趋近于-1，他们的方向越相反；接近于0，表示两个向量近乎于正交。一般情况下，相似度都是归一化到[0,1]区间内，因此余弦相似度表示为 cosine\_similarity = 0.5cosθ + 0.5

1. 使用方法

具体输入命令：

**python model.py --db\_dir部件/产品特征向量存放地址 --search\_item\_dir 待检索部件/产品特征向量存放地址 --result\_dir 结果保存地址 --topK 返回的结果个数 --feature\_length 特征向量维度 --device\_num 设备种类数 --material\_num 材料种类数**

例如：

**python model.py --db\_dir ./sample\_db.txt --search\_item\_dir ./sample\_item.txt --result\_dir sample\_result.txt --topK 5 --feature\_length 20 --device\_num 10 --material\_num 10**

分析结果会存放在**结果保存地址**中