**部件/产品模型的相似度分析模型使用说明**

1. 模型用途

根据部件/产品模型属性寻找与该模型相似的同类模型。输入一个部件/产品模型属性对应的特征向量，输出与该部件/产品相似的其它部件/产品。

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，第一列为模型名称，其余n+1列为模型特征向量；另一个是需要进行相似度分析的模型，共有一行，其中第一个项为该模型名称，其余项为该模型的特征向量

输出数据为两列，一列是模型名称，另一列是输入的待分析模型与该模型相似度，从高到低排列。

1. 算法原理

部件/产品模型的相似度分析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。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。

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

部件/产品模型的相似度分析模型使用pandas工具将文件导入到DataFrame中，再转为numpy数组，遍历每一行求解相似度

1. 使用方法

具体输入命令：

**python model.py --db\_dir模型数据特征向量存放地址 –test\_dir 待分析模型特征向量存放地址 --result\_dir 结果保存地址 --topK 返回的结果个数**

例如：

**python model.py --db\_dir ./sample\_db.txt --test\_dir ./sample\_test.txt --result\_dir sample\_result.txt --topK 5**

分析结果会存放在**结果保存地址**中