

使用 MolAICal 对虚拟筛选结果进行聚类

作者: Qifeng Bai (update 2021-10-16)

更多教程 (含英文教程) 请见如下:

MolAICal 官方主页: <https://molaical.github.io>

MolAICal 官方主页中国镜像: <https://molaical.gitee.io>

MolAICal 中文博客: <https://molaical.gitee.io/cntutorial.html>

1. 简介

有时虚拟筛选的结果具有相似的打分。排名靠前的分子具有非常相似的结构。选择代表性分子可以节省金钱和时间。本教程基于药物打分和结构相似性, 使用 MolAICal 对虚拟筛选结果进行聚类。欲了解更详细的 MolAICal, 请参考网址: <https://molaical.github.io>。

2. 材料

2.1. 软件要求

1) MolAICal: <https://molaical.github.io>

国内镜像 MolAICal: <https://molaical.gitee.io>

2.2. 示例文件

1) 所有必要的教程文件都从以下位置下载:

<https://gitee.com/molaical/tutorials/tree/master/017-clusterVSResults>

3. 程序

3.1 计算结构相似度

1) MolAICal 提供了两种结构相似度计算方式, 指纹相似度和 3D 结构相似度 (更多细节请查看 MolAICal 手册)。在这里, 本教程选择了 3D 结构相似性。

进入 017-clusterVSResults, 输入如下命令:

```
#> molaical.exe -tool 3Dcompare -i name.dat -s similarity.dat -f mol2list
```

它将生成一个名为 similarity.dat 的文件, 其中 name.dat 是分子名称列表。

2) 现在, 合并“similarity.dat”, “bindingScore.dat” (里面包含配体的打分值), 和“name.dat”。

输入命令如下:

```
#> molaical.exe -tool col -f similarity.dat -l bindingScore.dat -s " " -o tmp.dat
```

然后，将 tmp.dat 和 name.dat 合并到一个名为“pc.dat”的文件中。

```
#> molaical.exe -tool col -f tmp.dat -l name.dat -s " " -o pc.dat
```

注意：请按上述顺序合并文件。第一列是“similarity.dat”，第二列是“bindingScore.dat”，最后一列是“name.dat”。

3.2 聚类结果

MolAICal 使用 k-means 进行聚类，输入以下命令：

```
#> molaical.exe -tool kmeans -n 3 -i pc.dat -o results.dat
```

它将在文件“results.dat”中聚成 3 类。打开“results.dat”，如下图所示：

```
# 1th: the ligands name; 2th: Same number means same cluster; 3th: affir
lig_1.mol2      1      -3.27
lig_2.mol2      2      -7.67
lig_7.mol2      2      -7.56
lig_8.mol2      2      -7.46
lig_16.mol2     2      -7.63
lig_3.mol2      3      -6.95
lig_4.mol2      3      -6.42
lig_5.mol2      3      -6.43
lig_6.mol2      3      -6.5
lig_9.mol2      3      -6.66
lig_10.mol2     3      -6.45
lig_11.mol2     3      -6.42
lig_12.mol2     3      -6.54
lig_13.mol2     3      -6.65
lig_14.mol2     3      -6.9
lig_15.mol2     3      -6.96
lig_17.mol2     3      -6.95
lig_18.mol2     3      -6.47
lig_19.mol2     3      -7.05
lig_20.mol2     3      -6.43
lig_21.mol2     3      -6.72
lig_22.mol2     3      -6.75
lig_23.mol2     3      -6.4
lig_24.mol2     3      -7.0
```

Ligand Name	Cluster	Affinity Score
lig_1.mol2	1	-3.27
lig_2.mol2	2	-7.67
lig_7.mol2	2	-7.56
lig_8.mol2	2	-7.46
lig_16.mol2	2	-7.63
lig_3.mol2	3	-6.95
lig_4.mol2	3	-6.42
lig_5.mol2	3	-6.43
lig_6.mol2	3	-6.5
lig_9.mol2	3	-6.66
lig_10.mol2	3	-6.45
lig_11.mol2	3	-6.42
lig_12.mol2	3	-6.54
lig_13.mol2	3	-6.65
lig_14.mol2	3	-6.9
lig_15.mol2	3	-6.96
lig_17.mol2	3	-6.95
lig_18.mol2	3	-6.47
lig_19.mol2	3	-7.05
lig_20.mol2	3	-6.43
lig_21.mol2	3	-6.72
lig_22.mol2	3	-6.75
lig_23.mol2	3	-6.4
lig_24.mol2	3	-7.0