

使用 PyInstaller 打包 MolAICal 软件深度学习模型“AI GenMols”的例子

作者：MolAICal (update 2020-08-02)

更多教程（含英文教程）请见如下：

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

MolAICal 文章介绍：<https://doi.org/10.1093/bib/bbaa161>

MolAICal 中文博客：<https://molaical.github.io/centutorial.html>

MolAICal blogspot：<https://qblab.blogspot.com>

1. 介绍

“AI GenMols”有时在 Linux 操作系统中没有很好的兼容性。在这种情况下，它需要再次生成二进制“AI GenMols”。在本教程中，选择了 ORGAN (<https://github.com/gablg1/ORGAN>) 来安装“AI GenMols”。此外，您可以根据自己训练的深度学习模型来打包“AI GenMols”

2. 材料

2.1. 所需软件

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

2) Anaconda: <https://www.anaconda.com/open-source>

Anaconda 的版本选择应基于 Python 3.x 而不是 Python 2.x

2.2. 示例文件

1) 所有必需的教程文件均可从以下网址下载:

<https://github.com/MolAICal/specialtopic/tree/master/012-AIGenMols>

3. 步骤

3.1. 安装 Anconda (<https://www.anaconda.com/distribution>).

确保完全安装并且正确设置了环境变量。

3.2. 下载 “spec-list.txt”并创建一个独立的环境

```
#> conda create --name AIGen --file spec-list.txt
```

```
#> conda activate AIGen
```

注意： 如果发生某些错误，例如：*CondaHTTPError: HTTP 404 NOT FOUND for url <https://conda.anaconda.org/conda-forge/linux-64/xxxx-xxxx>*，您可以在“spec-list.txt”中删除相应的软件包。然后重复上述步骤。可以通过以下类似步骤安装缺少的软件包，而无需安装版本号：

```
#> conda install -c conda-forge xxxx
```

3.3. 安装必需的库

```
#> conda install -c conda-forge openblas
```

```
#> conda install -c conda-forge openssl
```

```
#> pip install editdistance==0.3.1
```

3.4. 最后，生成二进制“深度学习模型”

解压缩“organ.zip”

```
#> unzip organ.zip
```

```
#> cd organ
```

```
#> pyinstaller --add-data="organ/NP_score.pkl.gz:organ"
--add-data="organ/SA_score.pkl.gz:organ" --add-data="organ/data/FDA-H.csv:organ/data"
--add-data="organ/checkpoints/FDA-H/FDA-H_99.ckpt.data-00000-of-00001:organ/checkpoints/
FDA-H"
--add-data="organ/checkpoints/FDA-H/FDA-H_99.ckpt.index:organ/checkpoints/FDA-H"
--add-data="organ/checkpoints/FDA-H/FDA-H_99.ckpt.meta:organ/checkpoints/FDA-H"
--add-data="organ/data/FDA1884.csv:organ/data"
--add-data="organ/checkpoints/FDA1884/FDA1884_119.ckpt.data-00000-of-00001:organ/checkp
oints/FDA1884"
--add-data="organ/checkpoints/FDA1884/FDA1884_119.ckpt.index:organ/checkpoints/FDA1884
"
--add-data="organ/checkpoints/FDA1884/FDA1884_119.ckpt.meta:organ/checkpoints/FDA1884"
--add-data="organ/data/zinc.csv:organ/data"
--add-data="organ/checkpoints/ZINC/ZINC_99.ckpt.data-00000-of-00001:organ/checkpoints/ZIN
C" --add-data="organ/checkpoints/ZINC/ZINC_99.ckpt.index:organ/checkpoints/ZINC"
--add-data="organ/checkpoints/ZINC/ZINC_99.ckpt.meta:organ/checkpoints/ZINC" -F main.py
```

打开文件夹“organ/dist”

```
#> cd dist
```

您将看到名为“main”的文件，将其重命名为“AIGenMols”

```
#> mv main AIGenMols
```

```
#> chmod +x AIGenMols
```

此时完全生成了名为“AIGenMols”的二进制深度学习模型。将其直接复制或移动到 MolAIcal 软件文件夹中即可。

安装过程分析:

```
-----  
--add-data="organ/data/zinc.csv:organ/data"  
--add-data="organ/checkpoints/ZINC/ZINC_99.ckpt.data-00000-of-00001:organ/checkpoints/ZINC  
C" --add-data="organ/checkpoints/ZINC/ZINC_99.ckpt.index:organ/checkpoints/ZINC"  
--add-data="organ/checkpoints/ZINC/ZINC_99.ckpt.meta:organ/checkpoints/ZINC" -F main.py  
-----
```

以上虚线部分是打包关键的部分，您可以使用自己训练好的深度学习模型替换以上数据

注意:

- 1). 安装 Anaconda 时，Anaconda 的安装路径应确保没有空格。例如：`/home/test install/anaconda`。字符“test”和“install”之间的空格可能会导致“AI GenMols”打包过程失败。
- 2). “AI GenMols”的命令用法应如下所示，才能保证打包的模型可以直接被 MolAICal 调用:

```
AI GenMols -f test.txt -s FDAFrag -n 100
```

```
AI GenMols -f test.txt -s ZINCMol -n 100
```

-f: 表示生成包含 SMILES 格式分子的文件名。

-s: 代表选择的深度学习模型名称，如：ZINCMol 代表使用 ZINC 数据库的分子训练出的深度药物生长模型

-n: 代表产生的分子数

这只是打包“AI GenMols”的一个示例。您可以对自己训练的模型使用类似的打包方法。我们将及时更新深度学习模型。