Using command-line to label data with aizynthfinder

Installation

```
conda env create -f
https://raw.githubusercontent.com/MolecularAI/aizynthfinder/master/env-users.yml

conda env update -n aizynth-env -f
https://raw.githubusercontent.com/MolecularAI/aizynthfinder/master/env-users.yml

conda activate aizynth-env
```

Generate yml file

```
download_public_data my_folder
```

- `my_folder` is the folder that you want download to. This will creat a `config.yml` file.
- However, to keep the setting the same as the paper, we have to add more information.
  - open your `config.yml` file
  - ```
    properties:
      iteration_limit: 100
      return_first: true
      time_limit: 60
      C: 1.4
      cutoff_cumulative: 0.995
      cutoff_number: 50
      max_transforms: 6
    ```

    Add the code above on the top of the yml file, keep other parts unchanged.
  - the final `config.yml` file should be like:

    ```
    properties:
      iteration_limit: 100
      return_first: true
      time_limit: 60
      C: 1.4
      cutoff_cumulative: 0.995
      cutoff_number: 50
      max_transforms: 6
    policy:
      files:
        uspto:
          - D:\aizynthfinder\uspto_model.hdf5
          - D:\aizynthfinder\uspto_templates.hdf5
    filter:
      files:
    ```
    ```

```
    uspto: D:\aizynthfinder\uspto_filter_model.hdf5
stock:
  files:
    zinc: D:\aizynthfinder\zinc_stock.hdf5
```

- **Warning: yml files are sensitive to the space bar and cannot replace spaces with indentation**

Run script to generate label

Prerequisite

- the tool.py code is shown below:

```
import os
import pandas as pd


data = pd.read_csv("smiles.csv", encoding='utf-8').iloc[:, 0]
data = data.dropna(axis=0, how='any') # read_csv
count = len(data) # lines of txt file
batch = 100 # set numbers of smiles in a batch for saving

with open("smiles.txt", "w+") as f:
    for i in range(0, count):
        f.write(data[i] + "\n")
        if ((i+1) % batch == 0) | ((i+1) >= count):
            f.close()
            os.system("aizynthcli --config config.yml --smiles smiles.txt")
            pd.read_hdf("output.hdf5", "table")
[["target","is_solved"]].to_csv("label.csv", mode="a", header=False,
index=None, sep = ",")
            f = open("smiles.txt", "w+")
```

- **"smiles.csv" is the imported csv. You can also change it into the absolute path in the tool.py script. Make sure the first column of the csv file is the smiles.**
- **make sure the operating environment is aizynth-env**
- **the default number of smiles for saving each time is 100, which can be changed in the script**

Run script

- Run in terminal:

```
python tool.py
```

- Then, the script will generate a csv file.
- **Do not keep the csv open when the script is running.**