py 项目中学到的知识梳理

作者：MaidongAndYida
原文链接：https://ld246.com/article/1560505239180
来源网站：链滴

两个月前需求：使用python3做一个将观测数据编译产出成bufr数据的一个工具

刚刚完成初版，其中的数据文件路径和数据内容格式还需要仔细核对，但整体逻辑已实现，剩下的工时间可能会用来完善它

Anaconda3

The open-source Anaconda Distribution is the easiest way to perform Python/R data science nd machine learning on Linux, Windows, and Mac OS X. With over 11 million users worldwide, it is the industry standard for developing, testing, and training on a single machine, enabling i dividual data scientists to:

- Quickly download 1,500+ Python/R data science packages
- Manage libraries, dependencies, and environments with Conda
- Develop and train machine learning and deep learning models with scikit-learn, TensorFl w, and Theano
- Analyze data with scalability and performance with Dask, NumPy, pandas, and Numba
- Visualize results with Matplotlib, Bokeh, Datashader, and Holoviews

python读取xml

```
from xml.dom import minidom


def readXmlByTagName(path):
    with open(path, 'r', encoding='utf8') as fh:
        # 获取根节点
        root = minidom.parse(fh).documentElement
        # 节点类型：'ELEMENT_NODE'，元素节点； 'TEXT_NODE'，文本节点； 'ATTRIBUTE_NODE
```

```
，属性节点
        #print('节点类型:')
        return root
def getElementsByTagName(root,tagName):
    return root.getElementsByTagName(tagName)[0].childNodes[0].data
```

DataFrame --- pandas

pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

pandas的使用效果很腻害，在项目中主要用来读取如下图格式数据：

```
54421, 2222, 2019, 144, 1700, 0, 01, -999. 9, 140. 992, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, 26. 069, 27. 303, 34. 686, 969. 470
54421, 2222, 2019, 144, 1705, 0, 01, -999. 9, 121. 433, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, 26. 043, 27. 327, 34. 453, 969. 504
54421, 2222, 2019, 144, 1710, 0, 01, -999. 9, 115. 159, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, 26. 038, 27. 393, 34. 406, 969. 497
54421, 2222, 2019, 144, 1715, 0, 01, -999. 9, 130. 936, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, 26. 051, 27. 458, 34. 070, 969. 484
54421, 2222, 2019, 144, 1720, 0, 01, -999. 9, 142. 536, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, 26. 058, 27. 461, 34. 185, 969. 463
54421, 2222, 2019, 144, 1725, 0, 01, -999. 9, 163. 731, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, 26. 057, 27. 410, 34. 013, 969. 450
54421, 2222, 2019, 144, 1730, 0, 01, -999. 9, 115. 613, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, 26. 044, 27. 340, 34. 097, 969. 362
54421, 2222, 2019, 144, 1735, 0, 01, -999. 9, 129. 820, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, 26. 023, 27. 305, 34. 060, 969. 321
54421, 2222, 2019, 144, 1740, 0, 01, -999. 9, 134. 829, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, 26. 011, 27. 327, 34. 039, 969. 341
54421, 2222, 2019, 144, 1745, 0, 01, -999. 9, 160. 241, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, 26. 000, 27. 397, 34. 740, 969. 307
54421, 2222, 2019, 144, 1750, 0, 01, -999. 9, 151. 460, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, 26. 014, 27. 459, 34. 870, 969. 328
54421, 2222, 2019, 144, 1755, 0, 01, -999. 9, 142. 488, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, -999. 9, 26. 018, 27. 441, 34. 563, 969. 273
```

用到的pandas语法大概有:

- pandas.read_table(data_path, sep=',',dtype = 'str') 用来将数据读取出来
- .shape[0]用来获取数据的行数
- .iloc根据x和y轴来定位元素
- 文档地址

十进制转二进制

```
def Number2BinStr(num, size):

    '''
    整形转二进制字符的方法;
    :param num: 需要变换的整数;
    :param size:设定二进制宽度
    :return:
    '''
    fmt = '{0:0%db}' % size
    return fmt.format(num),size
```

字符串转二进制

```
def encode(s='', size=8):
   str_len = len(s)
   if str_len*8 <size:
      for i in range(0, int((size - str_len*8)/8)):
```

```
        s = s + ' '
    elif str_len*8 >size:
        pass
        # s = s
        # for i in range(0, int((str_len*8 - size)/8)):
    strs = ''
    for c in s:
        str_byte = bin(ord(c)).replace('0b', '')
        b = 8 - len(str_byte)
        for i in range(0, b):
            str_byte = '0'+str_byte
        strs = strs + str_byte
    return strs, size
```

求数据乘以比例因子加系数

```
def data_trasform_func(data, x, b):
    '''
    求数据乘以比例因子加系数的方法;
    :param data: 数据值;
    :param x:比例因子
    :param b:基准值
    :return:返回转换后的值;
    '''
    return int(data*math.pow(10, x) + b)
```

判断某文件夹下是否包含某个名称的文件，仅支持单个词模糊查询

```
#判断是否有数据文件
def search(path=".", name=""):
    result = []
    for item in os.listdir(path):
        item_path = os.path.join(path, item)
        if os.path.isdir(item_path):
            search(item_path, name)
        elif os.path.isfile(item_path):
            if name in item:
                result.append(item_path)
    return result
```

bufr编码格式附件

bufr编码格式规则.rar