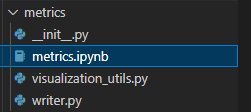
## 老师建议

尝试the number of the client: 2

different data



修改可视化的嘛

tensorflow 里面的固定东西，最后也能修改



也能修改

主要是main.py调用了server.py等，然后就慢慢来做吧

修改其中的一些原理即可了

## overleaf 周报

maybe I didn't use the GPU! actually Even when I use the GPU, the same error appears. Anyway, downloading data doesn't need the GPU.

First, I try the femnist

\newline using the python=3.7, I've been stuck in the Line command in the shell: %leaf/paper\_experiments $> ./sent140.sh <result-output-dir>,

\\the platform I choose is a cloud computing platform. after I input that line command, the server processed the information but the terminal went out after a while. Maybe it is the reason of the platform.

The following is the record of debugging the error.

it keeps producing the same error and I'm working on it. Maybe other students have met the same problem before and I can resort to them, while I cannot find solutions on github issues, yet.

OK, then the setup problem can be solved.

\\Now, I'm researching about the femnist data. follow some README.MD files, actually, I'm handling some other errors. Anyway, it is the nature of experiment. when I try to run the " ./preprocess.sh -s niid --sf 0.05 -k 0 -t sample", the shell gives me 2 errors

\newline"KeyError: 'd41d8cd98f00b204e9800998ecf8427e' finished assigning class labels to write images\newline FileNotFoundError: [Errno 2] No such file or directory: '/hy-tmp/leaf/data/femnist/data/intermediate/write\_with\_class.pkl'"

\\[Errno 2]the file "write\_with\_class.pkl" doesn't exist, but [Errno 1]

\\then I find that I need to use the python=3.8 because one error reminds of that. Actually, I really want to use the same setup as Prof. Yang. Oh! Let me just ask Amy what environment she uses, then it will be quite easy! But still comes the same error. \newline KeyError: 'd41d8cd98f00b204e9800998ecf8427e' should be the problem.

环境-3.7必定成功

（没问题：）

conda create -n leaf7 python=3.7

conda create -n leaf python=3.7

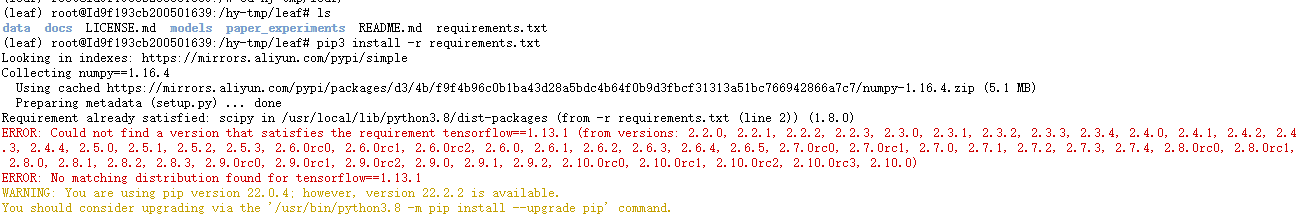
conda remove -n leaf --all

conda install python=3.7

有点奇怪

conda create -n leaf python=3.5

readme.md其实说明了一切，只是咱



pip3 install -r requirements.txt

不行



pip install 可以

pip3 install -r requirements.txt

会不会是当时安装用到的文件有了错误罢了？

## 1 操作开始

cd /hy-tmp/leaf/

conda activate leaf7

cd /hy-tmp/leaf/

conda activate ISSBA

conda activate TF2.1

cd models/metrics

./femnist.sh result\_femnist/

python main.py -dataset femnist -model cnn -lr 0.06 --minibatch 0.1 --clients-per-round 3 --num-rounds 200

python metrics.py

#runipy metrics.ipynb

如何在linux命令行（终端）执行ipynb 文件。

:

安装 runipy

pip install runipy

终端执行ipynb

runipy <YourNotebookName>.ipynb

runipy metrics.ipynb

来自 <<https://blog.csdn.net/poson/article/details/103933786>>

Twitter Sentiment Analysis

:useful!

来自 <<https://leaf.cmu.edu/build/html/tutorials/sent140-md.html>>

leaf/result\_Twitter

./sent140.sh ../result\_Twitter

（运行2次，都没用）

./sent140.sh result\_Twitter

## 2 继续

python main.py -dataset femnist -model cnn -lr 0.06 --minibatch 0.1 --clients-per-round 3 --num-rounds 200

老师说的metrics文件包，就好像有plots函数！

文件跑完，全都是名字命名反了的！真离谱

去main.py里修改一下

~~cp metrics\_stat.csv stat\_metrics.csv~~

~~cp metrics\_sys.csv sys\_metrics.csv~~

finished converting data to .json format

------------------------------

sampling data

Using seed 1666495910

/hy-tmp/leaf/data/femnist/meta

- random seed written out to /hy-tmp/leaf/data/femnist/meta/sampling\_seed.txt

writing all\_data\_0\_niid\_01.json

./preprocess.sh: line 186: 1064 Killed LEAF\_DATA\_META\_DIR=${META\_DIR} python3 sample.py $NAMETAG $SFRACTAG --seed ${SEED\_ARGUMENT}

------------------------------

removing users with less than 0 samples

writing all\_data\_0\_niid\_01\_keep\_0.json

------------------------------

generating training and test sets

- random seed written out to /hy-tmp/leaf/data/femnist/meta/split\_seed.txt

splitting data by sample

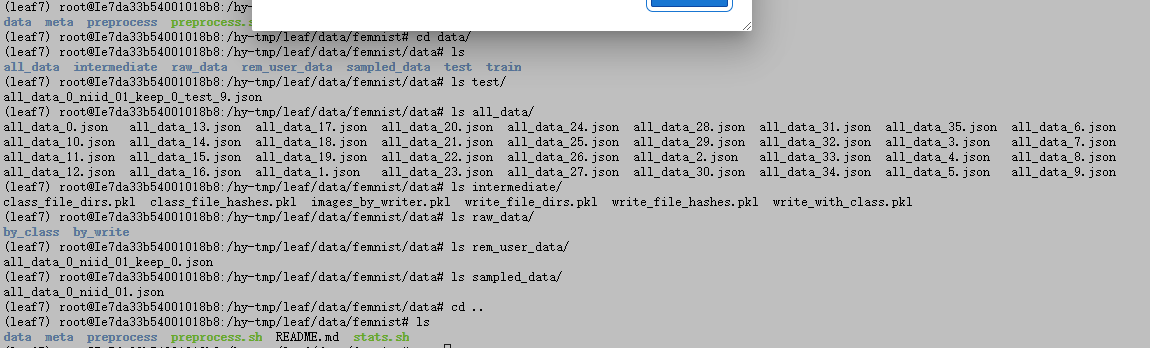
writing all\_data\_0\_niid\_01\_keep\_0\_train\_9.json

writing all\_data\_0\_niid\_01\_keep\_0\_test\_9.json

------------------------------

calculating JSON file checksums

checksums written to meta/dir-checksum.md5



基本都是.json文件了，

model exe

=========================Options=============================

-max\_depth 10000

-min\_bytes 0

-min\_peak\_bytes 0

-min\_residual\_bytes 0

-min\_output\_bytes 0

-min\_micros 0

-min\_accelerator\_micros 0

-min\_cpu\_micros 0

-min\_params 0

-min\_float\_ops 1

-min\_occurrence 0

-step -1

-order\_by float\_ops

-account\_type\_regexes .\*

-start\_name\_regexes .\*

-trim\_name\_regexes

-show\_name\_regexes .\*

-hide\_name\_regexes

-account\_displayed\_op\_only true

-select float\_ops

-output stdout:

## 3 图片修改

(ISSBA) root@Ie7da33b54001018b8:/hy-tmp/leaf/data/femnist/data/raw\_data# ls by\_class/

30 32 34 36 38 41 43 45 47 49 4b 4d 4f 51 53 55 57 59 61 63 65 67 69 6b 6d 6f 71 73 75 77 79

31 33 35 37 39 42 44 46 48 4a 4c 4e 50 52 54 56 58 5a 62 64 66 68 6a 6c 6e 70 72 74 76 78 7a

(ISSBA) root@Ie7da33b54001018b8:/hy-tmp/leaf/data/femnist/data/raw\_data# ls by\_write/

hsf\_0 hsf\_1 hsf\_2 hsf\_3 hsf\_4 hsf\_6 hsf\_7

(ISSBA) root@Ie7da33b54001018b8:/hy-tmp/leaf/data/femnist/data/raw\_data# ls by\_class/30

hsf\_0 hsf\_0.mit hsf\_1 hsf\_1.mit hsf\_2 hsf\_2.mit hsf\_3 hsf\_3.mit hsf\_4 hsf\_4.mit hsf\_6 hsf\_6.mit hsf\_7 hsf\_7.mit train\_30

(ISSBA) root@Ie7da33b54001018b8:/hy-tmp/leaf/data/femnist/data/raw\_data#



然后再preprocess

操作经验：

(leaf7) root@Ie7da33b54001018b8:/hy-tmp/leaf/data/femnist# cp -r /zt/leaf/data/femnist/preprocess .

(leaf7) root@Ie7da33b54001018b8:/hy-tmp/leaf/data/femnist# cp /zt/leaf/data/femnist/preprocess.sh .

## 修改虚拟环境

conda install python=3.8

pip install/uninstall

pip3 install ...:支持py3.0

conda install tensorflow

conda deactivate

### #虚拟环境命令+包的安装

**查看所有的虚拟环境**

conda info -e

结论：

如果用conda 安装不上的话试试pip。 pip安装不上的试试conda

1.第一步：去查看pip3 list 中发现确实没有gensim这个包，ok

pip3 install gensim --user 之后成功安装，但是jupyter 还是报同样的错误，

2.第二步：去Anaconda页面搜索gensim 发现确实没安装，点击apply 也一直没反应

————————————————

版权声明：本文为CSDN博主「人鱼线」的原创文章，遵循CC 4.0 BY-SA版权协议，转载请附上原文出处链接及本声明。

原文链接：<https://blog.csdn.net/qfikh/article/details/103481343>

## 教程理解

Quickstart script原来运行的.sh文件，真的就是先运行一下罢了，现在咱要修改，从原理上去修改，

就像人家造飞行汽车一样，要从原理上修改呀！！！

最近在看

<https://zhuanlan.zhihu.com/p/164659496#:~:text=%E5%85%B8%E5%9E%8B%E8%BF%87%E7%A8%8B>

## 理论理解

Specifically, we investigate the effect of varying the minimum number of samples per user (for training) on model accuracy when training using FedAvg algorithm, using the LEAF framework.

：just to vary the number of samples on model accuracy

改变每个用户的最小样本数量(用于训练)的模型准确性

the experiment requires pre-trained word embeddings. Thus, it still requires a script to process the data.

### error 1

{

./preprocess.sh -s niid --sf 0.05 -k 0 -t sample

KeyError: 'd41d8cd98f00b204e9800998ecf8427e'

finished assigning class labels to write images

FileNotFoundError: [Errno 2] No such file or directory: '/hy-tmp/leaf/data/femnist/data/intermediate/write\_with\_class.pkl'

}

:错误提示的意思是找不到对应的键值。

data有文件存在，所以有这个报错

preprocess文件可以仔细看，

问题描述

使用pickle.load(f)读取pkl文件的时候，报错ValueError: unsupported pickle protocol: 5

解决方案

删除data下的其他文件

原因是由于保存pkl文件的版本是python3.8，而读取的时候用了python3.7来读取，所以才会报错。

同样的error{

assigning class labels to write images

Traceback (most recent call last):

File "match\_hashes.py", line 25, in <module>

write\_classes.append((w, f, class\_hash\_dict[h][0]))

KeyError: 'd41d8cd98f00b204e9800998ecf8427e'

finished assigning class labels to write images

------------------------------

grouping images by writer

Traceback (most recent call last):

File "group\_by\_writer.py", line 13, in <module>

write\_class = util.load\_obj(wwcd)

File "/hy-tmp/leaf/data/utils/util.py", line 10, in load\_obj

with open(name + '.pkl', 'rb') as f:

FileNotFoundError: [Errno 2] No such file or directory: '/hy-tmp/leaf/data/femnist/data/intermediate/write\_with\_class.pkl'

finished grouping images by writer

------------------------------

converting data to .json format

Traceback (most recent call last):

File "data\_to\_json.py", line 43, in <module>

writers = util.load\_obj(by\_writer\_dir)

File "/hy-tmp/leaf/data/utils/util.py", line 10, in load\_obj

with open(name + '.pkl', 'rb') as f:

FileNotFoundError: [Errno 2] No such file or directory: '/hy-tmp/leaf/data/femnist/data/intermediate/images\_by\_writer.pkl'

finished converting data to .json format

}

Python=3.8又给我给炸了，我就用了pip install tensorflow

readme用了pip3

卸载2.0(因为python版本是3.0以上所以是pip3)：

pip3 uninstall tensorflow

我是用pip安装的

Collecting tensorflow

Killed

### error 2 solved

运行main.py:

Clients in Total: 5

--- Random Initialization ---

train\_accuracy: 0.00500833, 10th percentile: 0, 50th percentile: 0.00304878, 90th percentile 0.0118636

train\_loss: 4.13098, 10th percentile: 4.12663, 50th percentile: 4.13026, 90th percentile 4.13544

test\_accuracy: 0.0148149, 10th percentile: 0, 50th percentile: 0, 90th percentile 0.0368978

test\_loss: 4.12193, 10th percentile: 4.10857, 50th percentile: 4.12467, 90th percentile 4.13296

--- Round 1 of 2000: Training 3 Clients ---

Traceback (most recent call last):

File "main.py", line 186, in <module>

main()

File "main.py", line 83, in main

sys\_metrics = server.train\_model(num\_epochs=args.num\_epochs, batch\_size=args.batch\_size, minibatch=args.minibatch)

File "/hy-tmp/leaf/models/server.py", line 60, in train\_model

comp, num\_samples, update = c.train(num\_epochs, batch\_size, minibatch)

File "/hy-tmp/leaf/models/client.py", line 39, in train

comp, update = self.model.train(data, num\_epochs, num\_data)

File "/hy-tmp/leaf/models/model.py", line 88, in train

self.run\_epoch(data, batch\_size)

File "/hy-tmp/leaf/models/model.py", line 96, in run\_epoch

for batched\_x, batched\_y in batch\_data(data, batch\_size, seed=self.seed):

File "/hy-tmp/leaf/models/utils/model\_utils.py", line 18, in batch\_data

np.random.shuffle(data\_x)

File "mtrand.pyx", line 4865, in mtrand.RandomState.shuffle

File "mtrand.pyx", line 4868, in mtrand.RandomState.shuffle

TypeError: 'tuple' object does not support item assignment

解决方案：

np.random.shuffle(list(data\_x)) #we must change this line of the code

    np.random.set\_state(rng\_state)

    np.random.shuffle(list(data\_y))

I just search “File "mtrand.pyx", line 4865, in mtrand.RandomState.shuffle File "mtrand.pyx", line 4868, in mtrand.RandomState.shuffle TypeError: 'tuple' object does not support item assignment”

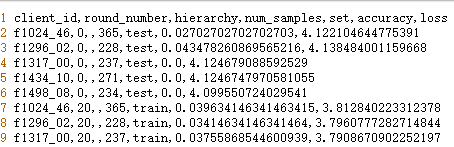
and take a try about”https://www.5axxw.com/questions/content/phf1qa”

modify the code similarly

### error 3 solved

使用自制添加目录法，实现结果了。

vim stat\_metrics.csv



error 3画图

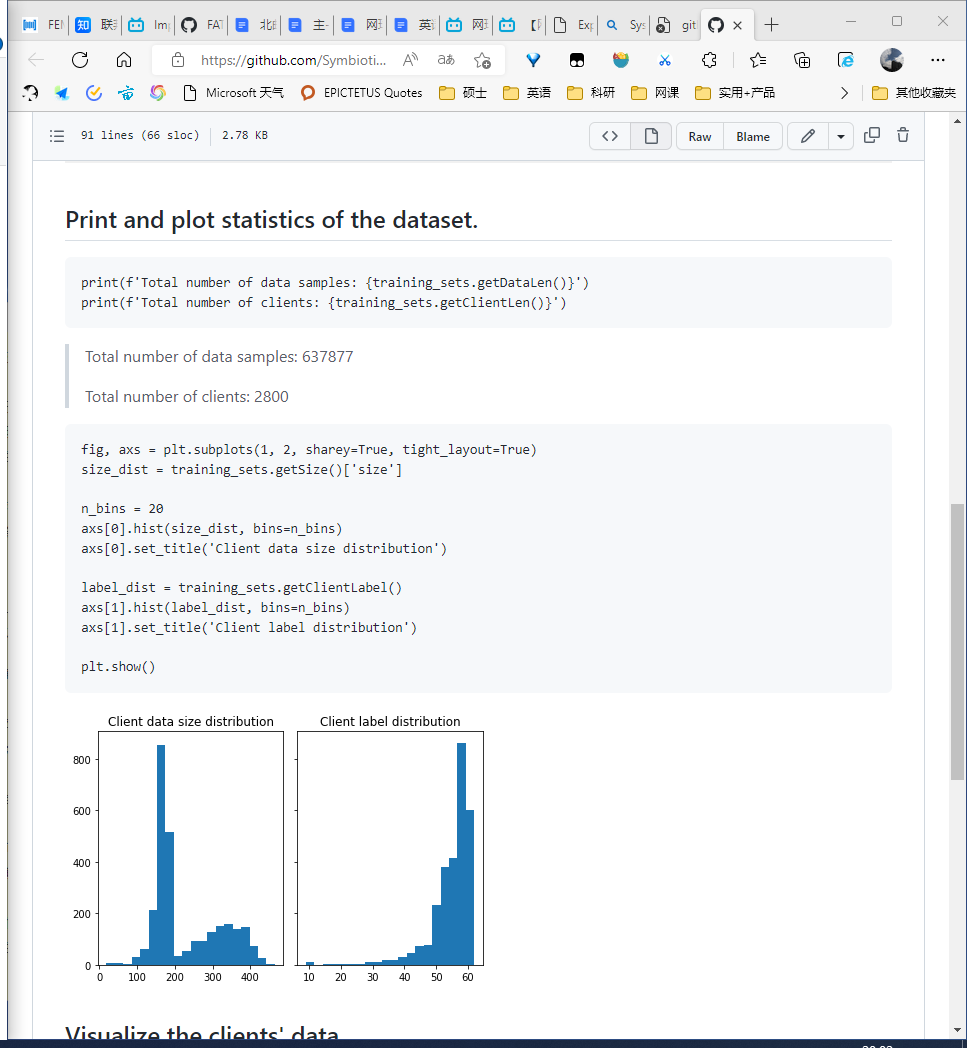
拥有的数据：

vim stat\_metrics.csv

stat\_metrics: 目测是client\_id, round\_number,hierachy, num\_samples,set，1,2,3（后面3个大数字：bytes\_written,client computations,bytes\_read）

bytes\_written,client computations,bytes\_read

<https://github.com/SymbioticLab/FedScale/blob/master/docs/Femnist_stats.md>



(leaf7) root@Ie5413bde100401a26:/hy-tmp/leaf/models/metrics# vim sys\_metrics.csv

目测是client\_id, round\_number,hierachy, num\_samples,set，1,2,3（后面3个大数字：bytes\_written,client computations,bytes\_read）：一共8个sys\_metrics, c\_groups, c~~\_num\_samples;~~ 我认为是否则数字不可能这么大的呀！

Return:

            bytes\_written: number of bytes written by each client to server

                dictionary with client ids as keys and integer values.

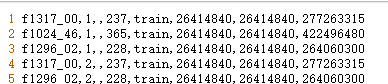
            client computations: number of FLOPs computed by each client

                dictionary with client ids as keys and integer values.

            bytes\_read: number of bytes read by each client from server

                dictionary with client ids as keys and integer values.

每次写入是8个，那么从写入的Parameter去看



train这里的column, ，似乎多了一行！

明天，打开看下几行几列来着，估计是输入没有进入

python 提示 keyError 的4种解决方法

来自 <<https://blog.csdn.net/Allure_LoveU/article/details/118927734>>

Traceback (most recent call last):

File "metrics.py", line 11, in <module>

stat\_metrics, sys\_metrics = visualization\_utils.load\_data(stat\_file, sys\_file)

File "/hy-tmp/leaf/models/metrics/visualization\_utils.py", line 37, in load\_data

sys\_metrics.sort\_values(by=NUM\_ROUND\_KEY, inplace=True)

File "/usr/local/miniconda3/envs/leaf7/lib/python3.7/site-packages/pandas/core/frame.py", line 5298, in sort\_values

k = self.\_get\_label\_or\_level\_values(by, axis=axis)

File "/usr/local/miniconda3/envs/leaf7/lib/python3.7/site-packages/pandas/core/generic.py", line 1563, in \_get\_label\_or\_level\_values

raise KeyError(key)

KeyError: 'round\_number'

---------------------------------------------------------------------------

KeyError Traceback (most recent call last)

/tmp/ipykernel\_2768/3587986460.py in <module>

6 sys\_file = 'metrics\_sys.csv' # change to None if desired

7

----> 8 stat\_metrics, sys\_metrics = visualization\_utils.load\_data(stat\_file, sys\_file)

/hy-tmp/leaf/models/metrics/visualization\_utils.py in load\_data(stat\_metrics\_file, sys\_metrics\_file)

33 stat\_metrics.sort\_values(by=NUM\_ROUND\_KEY, inplace=True)

34 if sys\_metrics is not None:

---> 35 sys\_metrics.sort\_values(by=NUM\_ROUND\_KEY, inplace=True)

36

37 return stat\_metrics, sys\_metrics

/usr/local/miniconda3/envs/leaf7/lib/python3.7/site-packages/pandas/core/frame.py in sort\_values(self, by, axis, ascending, inplace, kind, na\_position, ignore\_index, key)

5296

5297 by = by[0]

-> 5298 k = self.\_get\_label\_or\_level\_values(by, axis=axis)

5299

5300 # need to rewrap column in Series to apply key function

/usr/local/miniconda3/envs/leaf7/lib/python3.7/site-packages/pandas/core/generic.py in \_get\_label\_or\_level\_values(self, key, axis)

1561 values = self.axes[axis].get\_level\_values(key).\_values

1562 else:

-> 1563 raise KeyError(key)

1564

1565 # Check for duplicates

KeyError: 'round\_number'

and try to see something. I've met a new keyerror:

**How to Fix: KeyError in Pandas**

**来自 <<https://www.geeksforgeeks.org/how-to-fix-keyerror-in-pandas/>>**

建立一个改名字文件+.csv修改脚本：读取第一行，然后添加到其他文件第一行

主要是修改了存储文件名

### error 4 solved

bug

femnist 没数据的，报错

Clients in Total: 0

--- Random Initialization ---

--- Round 1 of 400: Training 2 Clients ---

Traceback (most recent call last):

File "main.py", line 186, in <module>

main()

File "main.py", line 87, in main

server.update\_model()

File "/hy-tmp/leaf/models/server.py", line 72, in update\_model

base = [0] \* len(self.updates[0][1])

IndexError: list index out of range

### error 5 solved

杨老师操作：

python3 main.py -dataset femnist -model cnn

报错，no module，

但是 python main.py -dataset femnist -model cnn 可行

### error6 doing

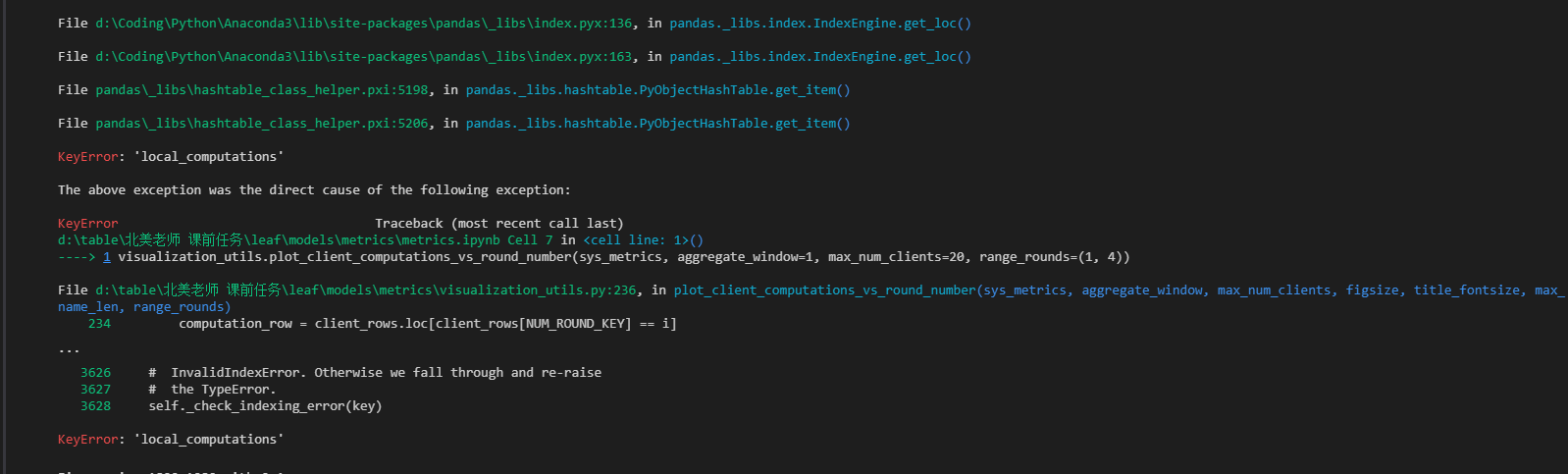
Output exceeds the [size limit](command:workbench.action.openSettings?[). Open the full output data [in a text editor](command:workbench.action.openLargeOutput?1a3cd4d1-798a-4679-982e-fb3786d3c4e6)

--------------------------------------------------------------------------- KeyError Traceback (most recent call last) File d:\Coding\Python\Anaconda3\lib\site-packages\pandas\core\indexes\base.py:3621, in Index.get\_loc(self, key, method, tolerance) 3620 try: -> 3621 return self.\_engine.get\_loc(casted\_key) 3622 except KeyError as err: File d:\Coding\Python\Anaconda3\lib\site-packages\pandas\\_libs\index.pyx:136, in pandas.\_libs.index.IndexEngine.get\_loc() File d:\Coding\Python\Anaconda3\lib\site-packages\pandas\\_libs\index.pyx:163, in pandas.\_libs.index.IndexEngine.get\_loc() File pandas\\_libs\hashtable\_class\_helper.pxi:5198, in pandas.\_libs.hashtable.PyObjectHashTable.get\_item() File pandas\\_libs\hashtable\_class\_helper.pxi:5206, in pandas.\_libs.hashtable.PyObjectHashTable.get\_item() KeyError: 'local\_computations' The above exception was the direct cause of the following exception: KeyError Traceback (most recent call last) d:\table\北美老师 课前任务\leaf\models\metrics\metrics.ipynb Cell 7 in <cell line: 1>() ----> [1](vscode-notebook-cell:/d:/table/%E5%8C%97%E7%BE%8E%E8%80%81%E5%B8%88 %E8%AF%BE%E5%89%8D%E4%BB%BB%E5%8A%A1/leaf/models/metrics/metrics.ipynb" \l "W6sZmlsZQ==?line=0) visualization\_utils.plot\_client\_computations\_vs\_round\_number(sys\_metrics, aggregate\_window=1, max\_num\_clients=20, range\_rounds=(1, 4)) File d:\table\北美老师 课前任务\leaf\models\metrics\visualization\_utils.py:236, in plot\_client\_computations\_vs\_round\_number(sys\_metrics, aggregate\_window, max\_num\_clients, figsize, title\_fontsize, max\_name\_len, range\_rounds) 234 computation\_row = client\_rows.loc[client\_rows[NUM\_ROUND\_KEY] == i]

...

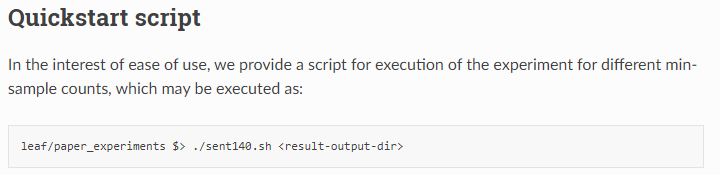
3626 # InvalidIndexError. Otherwise we fall through and re-raise 3627 # the TypeError. 3628 self.\_check\_indexing\_error(key) KeyError: 'local\_computations'

<Figure size 1800x1080 with 0 Axes>



## 实验继续

运行



TypeError: Descriptors cannot not be created directly.

If this call came from a \_pb2.py file, your generated code is out of date and must be regenerated with protoc >= 3.19.0.

If you cannot immediately regenerate your protos, some other possible workarounds are:

1. Downgrade the protobuf package to 3.20.x or lower.

2. Set PROTOCOL\_BUFFERS\_PYTHON\_IMPLEMENTATION=python (but this will use pure-Python parsing and will be much slower).

More information: <https://developers.google.com/protocol-buffers/docs/news/2022-05-06#python-updates>

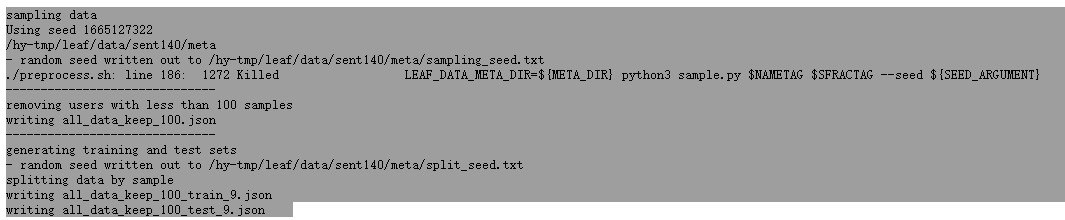
pip uninstall protobuf

pip install protobuf==3.20.1

pip 安装包罢了，一些是依赖问题

卸载了

protobuf 4.21.7



sent140/embs.json'

HTTP request sent, awaiting response... 200 OK

Length: 81363704 (78M) [application/zip]

都是这样大小的

同样的错误

FileNotFoundError: [Errno 2] No such file or directory: 'sent140/embs.json'

/hy-tmp/leaf /hy-tmp/leaf/paper\_experiments

/hy-tmp/leaf/models/metrics /hy-tmp/leaf /hy-tmp/leaf/paper\_experiments

mv: cannot stat 'sys\_metrics.csv': No such file or directory

mv: cannot stat 'stat\_metrics.csv': No such file or directory

### 2阶段尝试

root@Id9f193cb200501639:/hy-tmp/leaf/data/sent140# ls data/

all\_data intermediate raw\_data rem\_user\_data sampled\_data test train

root@Id9f193cb200501639:/hy-tmp/leaf/data/sent140/data# ls test/

all\_data\_keep\_3\_test\_9.json

root@Id9f193cb200501639:/hy-tmp/leaf/data/sent140/data# ls train/

all\_data\_keep\_3\_train\_9.json

无效的再处理：

root@Id9f193cb200501639:/hy-tmp/leaf/data/sent140# ./preprocess.sh -s niid --sf 0.05 -k 3 -t sample

------------------------------

sampling data

Using seed 1665369804

/hy-tmp/leaf/data/sent140/meta

- random seed written out to /hy-tmp/leaf/data/sent140/meta/sampling\_seed.txt

writing all\_data\_niid\_05.json

------------------------------

calculating JSON file checksums

checksums written to meta/dir-checksum.md5

Data for one of the specified preprocessing tasks has already been

generated. If you would like to re-generate data for this directory,

please delete the existing one. Otherwise, please remove the

respective tag(s) from the preprocessing command.

./preprocess.sh -s niid --sf 0.05 -k 3 -t sample

GPU开始

############################## sent140.stacked\_lstm ##############################

Traceback (most recent call last):

File "main.py", line 186, in <module>

main()

File "main.py", line 58, in main

client\_model = ClientModel(args.seed, \*model\_params)

File "/hy-tmp/leaf/models/sent140/stacked\_lstm.py", line 21, in \_\_init\_\_

\_, self.indd, vocab = get\_word\_emb\_arr(VOCAB\_DIR)

File "/hy-tmp/leaf/models/utils/language\_utils.py", line 119, in get\_word\_emb\_arr

with open(path, 'r') as inf:

FileNotFoundError: [Errno 2] No such file or directory: 'sent140/embs.json'

## 无效操作

只盯着代码、函数，还是不行的，结合了论文本身，那么更容易理解了！

operation

## 实验具有的模型

运行实验

当前LEAF数据集所进行的实验为FedAvg的cross machine下的单机多进程场景，目前已完成femnist和shakespeare两类数据集的测试。

通过运行fedlab\_benchmarks/fedavg/cross\_machine/LEAF\_test.sh可快速执行LEAF数据集下FedAvg的模拟实验。

来自 <<https://zhuanlan.zhihu.com/p/410131654>>