

# 安装说明

## 1. 环境

1. CentOS Linux release 7.6.1810
2. Python==3.7.4
3. virtualenv==16.7.2
4. mongodb==4.0.1

## 2. 激活虚拟环境

1. source py37env/bin/activate(虚拟环境路径)

## 3. 安装第三方包

1. pip install -r requirements.txt
2. pip3 install https://download.pytorch.org/whl/cpu/torch-1.0.1.post2-cp37-cp37m-linuxx8664.whl
3. pip3 install torchvision

## 4. mongodb数据库示例

Document	
ObjectId("5d22c12be3a42916c95af318")	ObjectId
customer_id	Int32
site_id	Int32
client_type	String
ip	String
op_type	String
sum_method	String
content_type	String
create_time	Int32
device_token	String
mac	String
user_id	Int32
data_title	String
data_content_id	String
data_content_fromid	String
data_bundle_id	String
data_column_id	String
data_column_name	String
data_publish_time	Int32
origin_create_at	Int32

# 配置说明

## Config.ini设置

### 1. 数据流入库日志设置(LOG)

path:设置日志目录

logname:设置日志名称

max\_bytes:设置日志最大字节数

backup\_count:设置日志文件份数

### 2. 数据库设置(MONGODB)

ip:数据库地址

port:数据库端口号

database:待分区数据名称

collection:待分区数据集合名称

r\_database:分区后数据名

cache\_database:缓存数据流数据名称

cache\_collection = 缓存数据流集合名称

### 3. HTTP接口设置(HTTP)

ip:HTTP服务器地址

port:HTTP服务器端口号

url:获取缓存数据库总记录数地址(如http://ip:port/get\_cache\_counts)

merge\_time\_interval:设置合并数据流的时间间隔

merge\_thresholds:设置合并数据流的记录数阈值

### 4. 训练超参数(hyperparameter)

loss:设置损失函数类型

n\_iter:设置迭代次数

batch\_size:设置训练批次大小

learning\_rate:设置学习率

l2:设置归一化方式

use\_cuda:是否使用GPU训练

top\_k:设置推荐数目

### 5. 其它项[other]

calculate\_time\_interval:检测是否训练的时间间隔

records\_thresholds:设置总记录数的下限阈值

valid\_records\_thresholds:设置满足训练的记录数下限阈值

directory:设置各区数据保存目录

tmp:设置缓存数据保存目录

save\_k:设置区域内热点新闻保存数目

## 程序启动

```
cd ProjectPath  
python manage.py > log.txt 2>&1 &  
python http_server.py > http_log.txt 2>&1 &
```

