
金融市场交易系统

闫涛

上海转晴科技有限公司

北京 2021.08.18

{yt7589}@qq.com

第一篇数据预处理

第一篇概率机器学习导论

第 1 章行情数据处理

Abstract

在本章中我们将通过 AKshare 库，获取 A 股分钟级行情数据，并将其进行预处理，变为深度学习可用的数据集。

1 行情数据处理概述

1.1 获取原始行情数据

我们首先通过 `apps.fmts.ds.akshare_data_source.AkshareDataSource` 获取原始的行情数据，将其保存到 csv 文件中。如果存在该 csv 文件，则直接从该文件中读出数据并返回。数据格式为：

```
1 .....
2 ['20210817 14:55:00', 30.339999999999996, 30.339999999999996,
   30.339999999999996, 30.339999999999996, 200.0]
3 ['', 30.339999999999996, 30.339999999999996, 30.339999999999996,
   30.339999999999996, 200.0]
4 .....
```

Listing 1: 行情数据格式

1.2 我

我们以收盘价为例，收盘价的折线图如下所示：

```
1 class OmniglotDs(Dataset):
2     def __init__(self, data_dir, k_way, q_query):
3         self.file_list = [f for f in glob.glob(data_dir +
4             '**/character*', recursive=True)]
5         self.transform = transforms.Compose([transforms.ToTensor()])
6         self.n = k_way + q_query
7
8     def __getitem__(self, idx):
9         sample = np.arange(20)
10        np.random.shuffle(sample)
11        img_path = self.file_list[idx]
12        img_list = [f for f in glob.glob(img_path + "**/*.png",
13            recursive=True)]
14        img_list.sort()
15        imgs = [self.transform(Image.open(img_file))
16            for img_file in img_list]
17        imgs = torch.stack(imgs)[sample[:self.n]]
18        return imgs
19
20    def __len__(self):
21        return len(self.file_list)
```

Listing 2: Omniglot 数据集类

2 附录 X