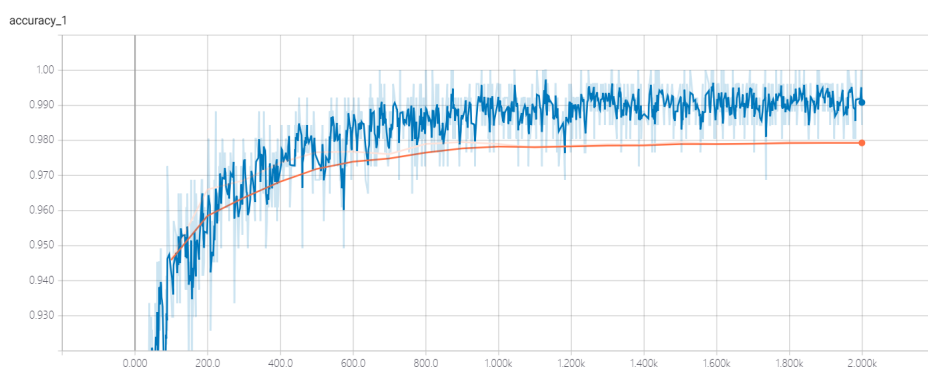


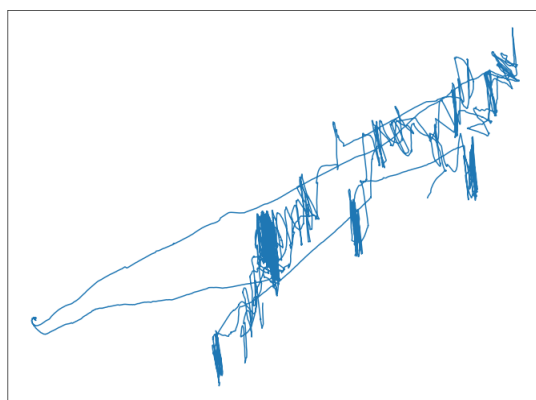
特征分析

可用于识别的特征类型

- 航速：拖网最高（ $2 < v < 6$ ），其次为刺网（ $1 < v < 2$ ），最低为张网（ $0 < v < 1$ ）。36所建议稿给出的各类型渔船航速的数据显示，总体速度波动较大，但基本都存在一个稳定的作业速度，所以可以先对速度检测曲线进行类似下图的平滑处理，减少速度曲线的波动。



- 航向（轨迹图）：经过观察分析发现各类型渔船的航向图有着各自类型独有的特点，比如拖网航向变化很频繁，轨迹图形成多个集中的区域；张网的航向变化很少，轨迹图多为直线；刺网兼有拖网及张网的特点，需要预处理后才能区分。可以尝试通过卷积神经网络对航向图像进行分类处理（已画出所有拖网船型的轨迹图，无明显可区分特征，未经预处理不适合卷积神经网络直接分类）。



- 航速比：相同类型船只的航速比去除掉一些奇异点之后的直方图拟合曲线具有一定的相似性，具有较强的可区分特征。可以尝试用卷积神经网络做分类（航速比直方图拟合曲线图多为背景信息，图像信息较少，可能结果不是很好，作为第二方案）。

第一方案（特征提取及分类）

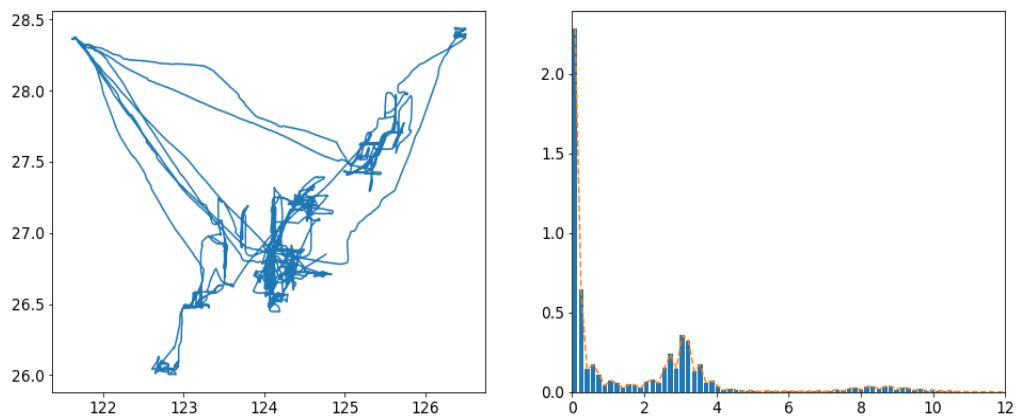
拖网数据分析及特征提取

- 拖网类型船只的总样本数（船号数）为 **445**，总的的数据组数为 **17,730,700**，其中船号为 **47292** 的拖网渔船拥有最少记录数据 **20,922** 组，拥有最多记录数据的船号为 **50029**，包含 **63,356** 组：

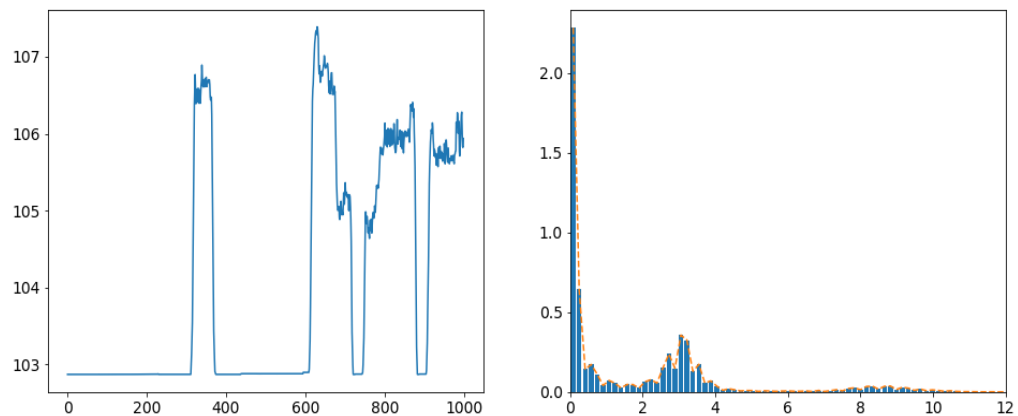
船号	数据总数（组）
47292（最少数据）	20,922
50029（最多数据）	63,356
所有拖网船只	17,730,700

其中，船号31436，第3411、3412航速数据错误。

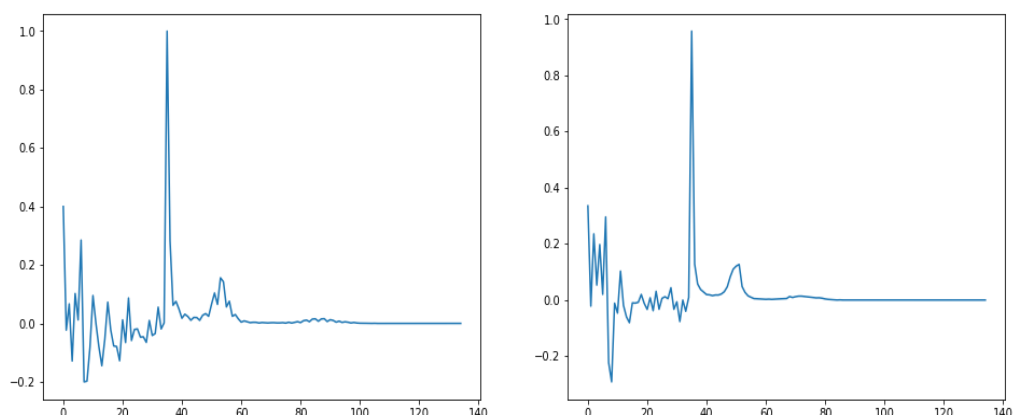
- 拖网数据特点：航速在 $v = 0$ 及 $v = 2$ 附近的权重较大，这点与其它两种类型的船只差别较大，因而该航速比直方图特征可直接作为部分的拖网船型的特征（最好做一下归一化处理）。



- 通过先对经、纬度坐标信息做降维去相关处理，然后可以得到每条渔船的 $1 \times N$ 维的特征，随后通过对特征固定数量的采样确保每条渔船的特征维数（暂时取 **1000** 维）都相同。



- 接着使用主成分分析法从这些 **1000** 维特征中提取出 M 维的特征向量（暂时先取 **35** 维）并做归一化处理，最后将该特征与速占比特征（**100** 维）结合得到该条渔船最终的特征（**135** 维）。
- 下图是前两条拖网渔船的最终特征。

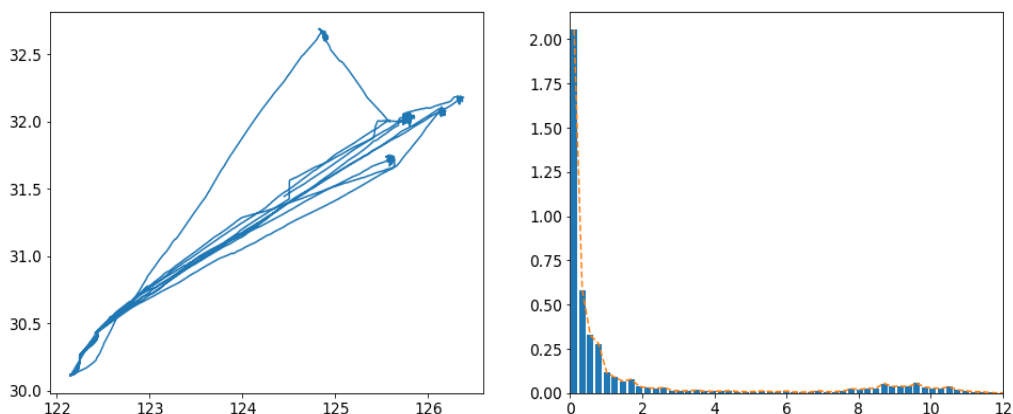


张网数据分析及特征提取

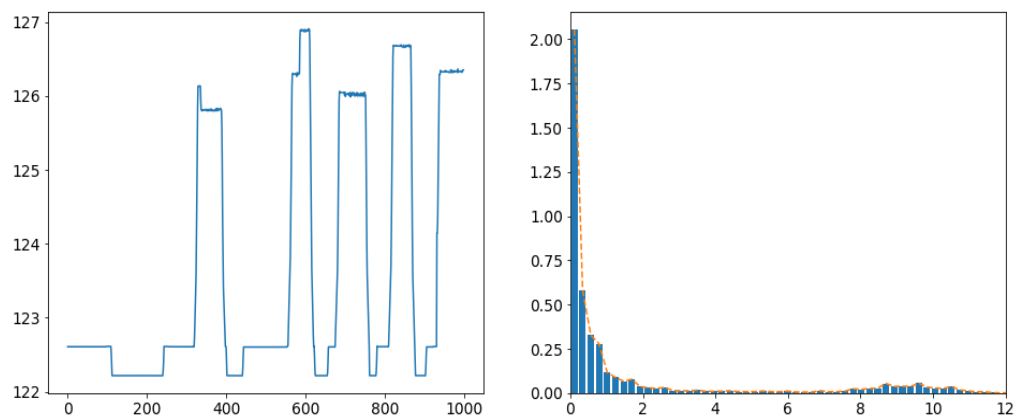
- 张网类型船只的总样本数（船号数）为 **77**，总的的数据组数为 **2,875,900**，其中船号为 **35234** 的张网渔船拥有最少记录数据 **20,557** 组，拥有最多记录数据的船号为 **55430**，包含 **57,307** 组：

船号	数据数（组）
35234（最少数据）	20,557
55430（最多数据）	57,307
所有张网船只	2,875,900

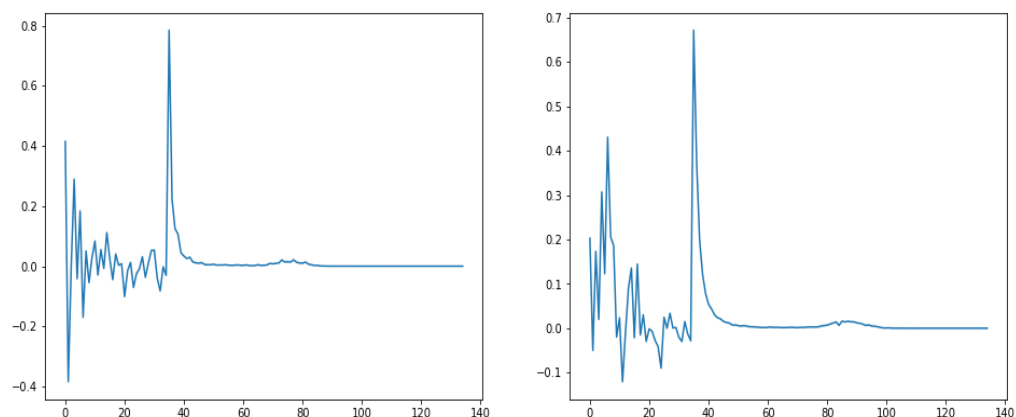
- 张网数据特点：航速在 $0 < v < 2$ 之间分布较多，这点与刺网相似，但是张网的航线轨迹图的目标地点更加的明确，侧面反映该类渔船的航向变化不是很频繁，这点与其它两种类型的船只差别较大。



- 通过与拖网渔船相同的操作，可以得到该条张网渔船的采样特征。该特征与其它两类船型有着明显的差异性，在一定区域内较为平稳，且有着近似周期的性质。



- 前两条张网渔船的最终 135 维特征。

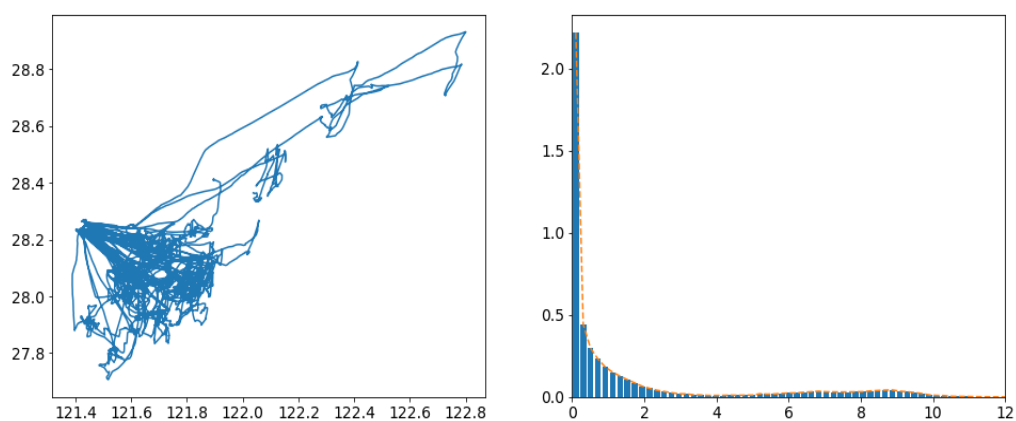


刺网数据分析及特征提取

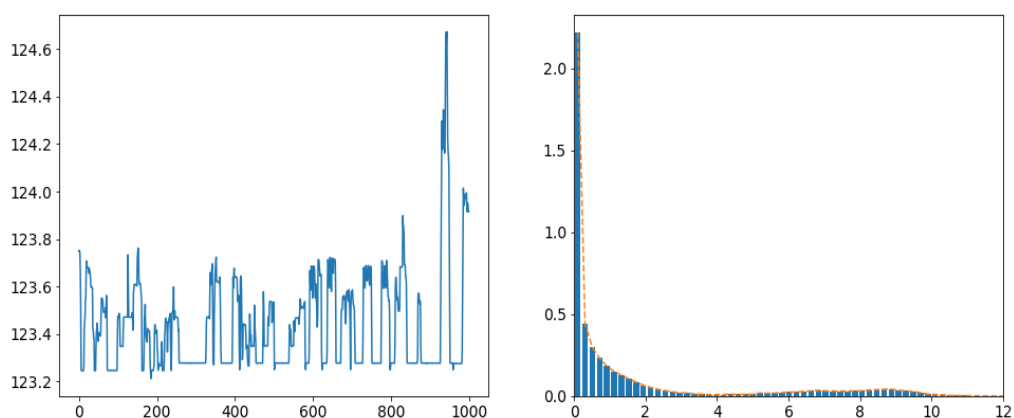
- 刺网类型船只的总样本数（船号数）为 173，总的的数据组数为 8,481,908，其中船号为 52251 的刺网渔船拥有最少记录数据 25,871 组，拥有最多记录数据的船号为 28824，包含 63,410 组：

船号	数据数（组）
52251（最少数据）	25,871
28824（最多数据）	63,410
所有刺网船只	8,481,908

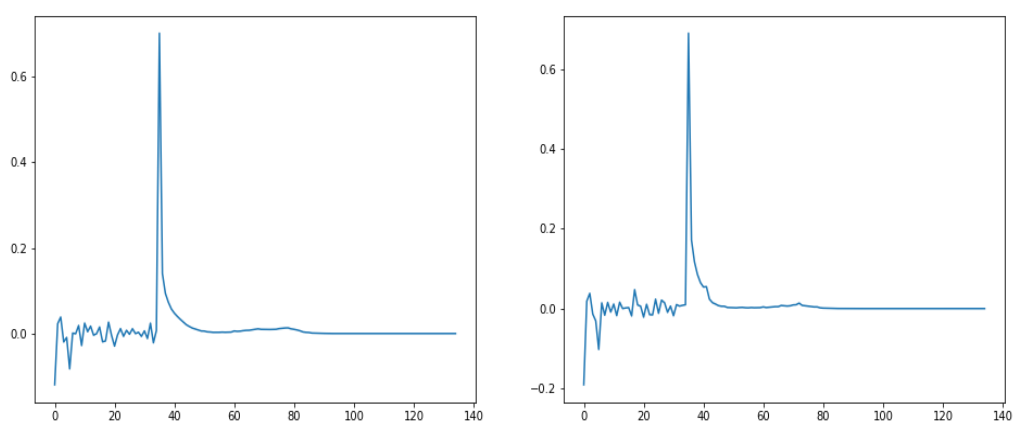
- 刺网数据特点：速度直方图分布与张网类型的船只很像，在航线轨迹图上与拖网十分的相似（都在频繁的改变航向）。



- 通过同样的数据预处理操作，可以得到该条刺网渔船的采样特征。该采样特征与其它两类船型的采样特征相比，波动性很明显，是非常好的可区分特征。

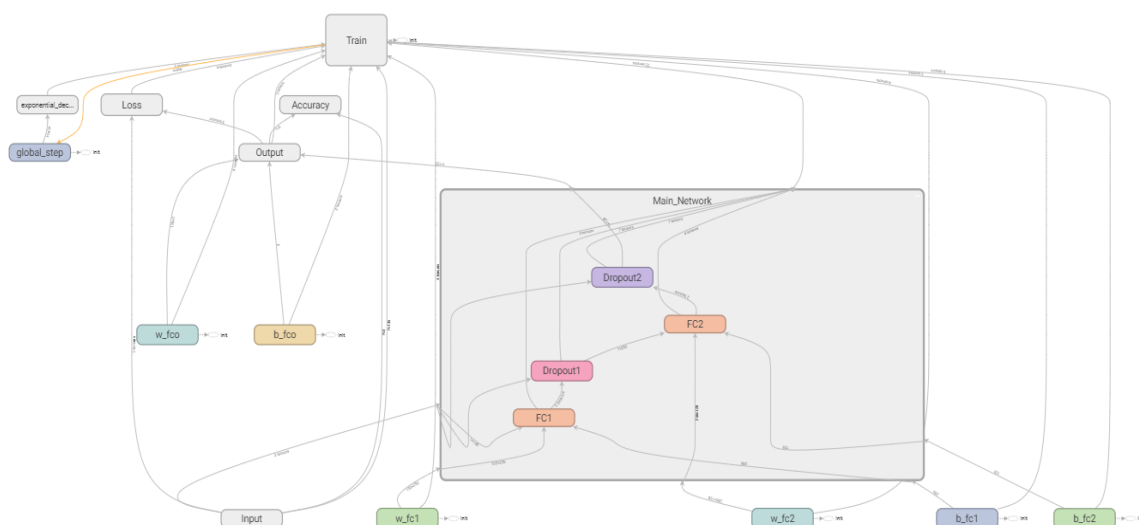


- 前两条刺网渔船的最终 135 维特征。



特征分类

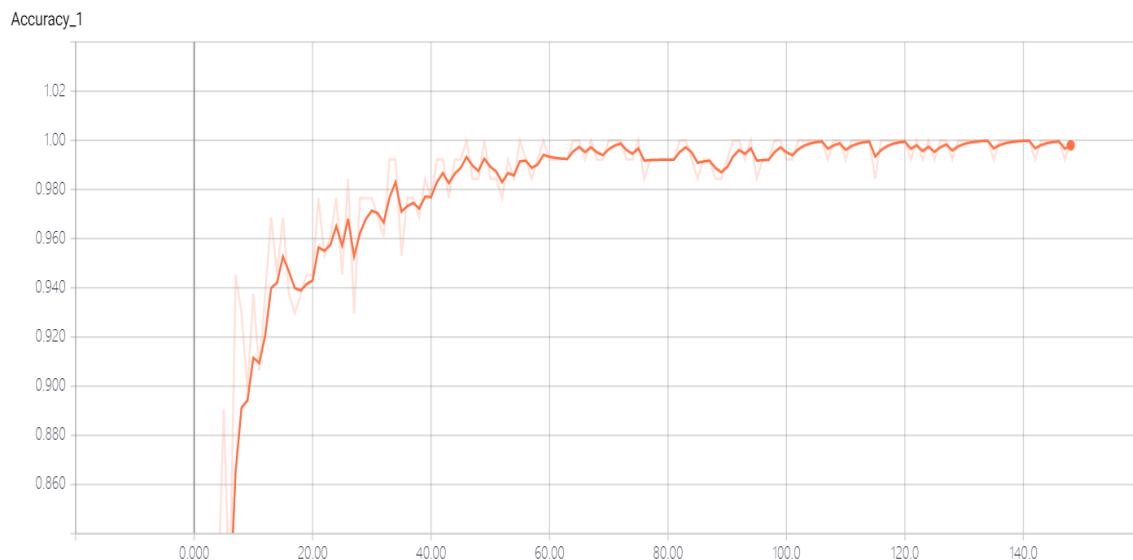
对上述提取的特征采用包含 2 个隐层的神经网络做分类，网络结构如下所示：



对 3 种类型共 695 条渔船信息做分类识别的结果如下所示（由于尚未对特征参数及网络结构优化，所以训练后期仍存在 0.5% 的较小精度波动）。

```
1 Epoch 1, cross_entropy: 0.8688, accuracy: 64.84%.
2 Epoch 10, cross_entropy: 0.2746, accuracy: 89.84%.
3 Epoch 20, cross_entropy: 0.1436, accuracy: 94.53%.
4 Epoch 30, cross_entropy: 0.0974, accuracy: 97.66%.
5 Epoch 40, cross_entropy: 0.0647, accuracy: 98.44%.
6 Epoch 50, cross_entropy: 0.0174, accuracy: 100.00%.
7 Epoch 60, cross_entropy: 0.0215, accuracy: 100.00%.
8 Epoch 70, cross_entropy: 0.0161, accuracy: 99.22%.
9 Epoch 80, cross_entropy: 0.0323, accuracy: 99.22%.
10 Epoch 90, cross_entropy: 0.0462, accuracy: 98.44%.
11 Epoch 100, cross_entropy: 0.0179, accuracy: 100.00%.
12 Epoch 110, cross_entropy: 0.0072, accuracy: 100.00%.
13 Epoch 120, cross_entropy: 0.0125, accuracy: 100.00%.
14 Epoch 130, cross_entropy: 0.0077, accuracy: 100.00%.
15 Epoch 140, cross_entropy: 0.0119, accuracy: 100.00%.
16 Epoch 150, cross_entropy: 0.0078, accuracy: 100.00%.
17 Training completed.
18
19 Cross_entropy on the whole training set: 0.0034,
    accuracy: 100.00%.
```

通过 **Tensorboard** 进行可视化处理后得到识别精度随迭代次数的变化：



以及交叉熵损失随迭代次数的变化：

Cross_entropy

