

## RF approach to identify bad meters

**1-Training time:** about one hour for more than 100 meters.

**2-Training approach:** Forecast regression (online forecasting), the reason can be seen the paper attached.

**3-Classifier approach:**

- 1)  $T_m < 100$ , considered as sc1, forecasting results are scr1.
- 2)  $100 < T_m < 120$ , considered as sc2, forecasting results are scr2.
- 3)  $120 < T_m$  considered as sc3, forecasting results are scr3.

**4-Feature choosing approach:**

$T_m$ ,  $T_w$ , Load, the past three hours data, and forecasting the following one hour.

**5-Results: The last line in every single text**

```
2014072014      138.2   3       3
2014072016      152.6   3       3
2014072019      129.2   3       3
p:1326  pr:1212 c1:1084 cr1:1079      c2:100  cr2:64  c3:94  cr3:69
```

**6-The whole forecasting results:**

Sp: 173921 spr: 165100 sc1: 150060 scr1:149704

sc2:16644 scr2:11350 sc3:7217 scr3:4046

The whole data dot is 173921, and the program can forecast 165100 data dot correctly.

(a)  $T_m < 100$ , the result: sc1: 150060 scr1: 149704

(b)  $100 < T_m < 120$ : sc2:16644 scr2:11350

(c)  $120 < T_m$ : sc3:7217 scr3:4046

**7-Conclusions and suggestions:**

This program can be suitable for millions of meter data disposal on line. I highly recommend this approach, especially in the engineering application.

The **advantages** of this approach can be as followings:

1-The data disposal and forecasting **speed are very fast, not easily over fitting**, especially for millions of meter data in applications.

2-Compared with ANN, the forecasting accuracy may be not so high, but some disadvantages of ANN cannot be easy to handle with in the practical applications, such as over fitting, lower disposal speed and so on.

**Suggestions:** If we want to get more convincing results, we need more meter data and choose more features to conduct more experiments.