

KPID算法：KPI异常检测中的 “PID”算法

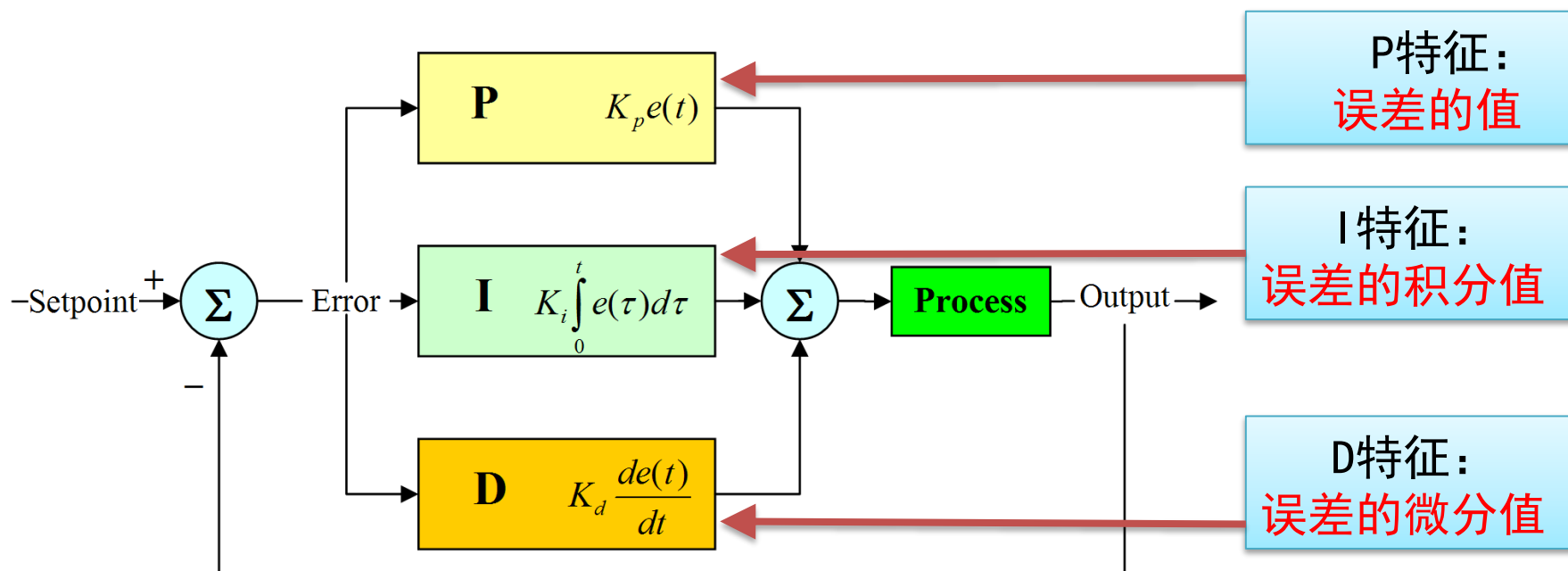
队伍：火眼金睛

5月19日·北京·清华大学

答辩大纲

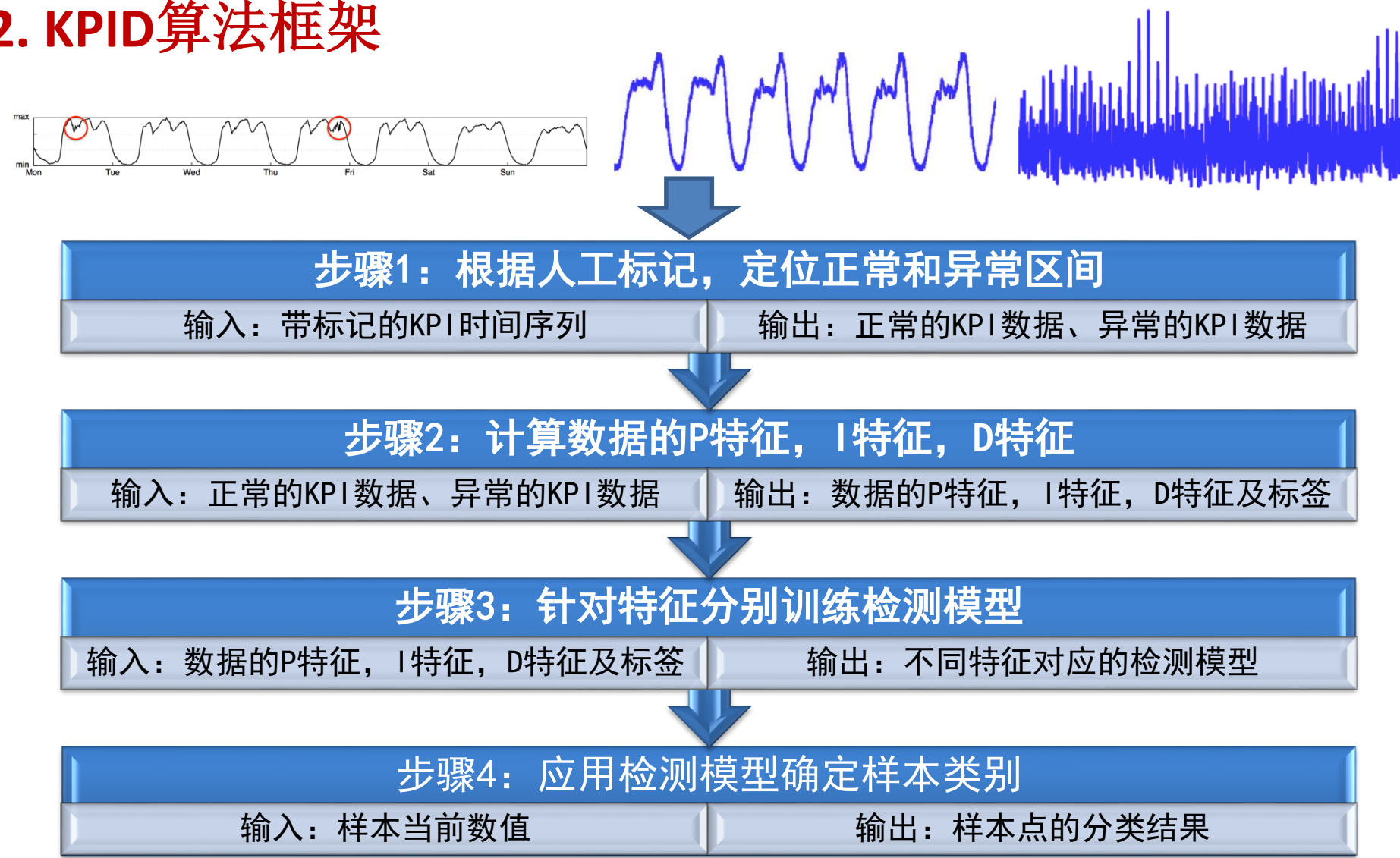
1. PID算法简介
2. KPID算法框架
3. KPID算法示例
4. KPID算法特点

1. PID算法简介



- PID控制算法：误差数据 \rightarrow 控制动作
- KPI检测：KPI数据 \rightarrow KPI标签
- 三种数据特征
- 基于PID特征的KPI异常检测算法：KPID算法

2. KPID算法框架

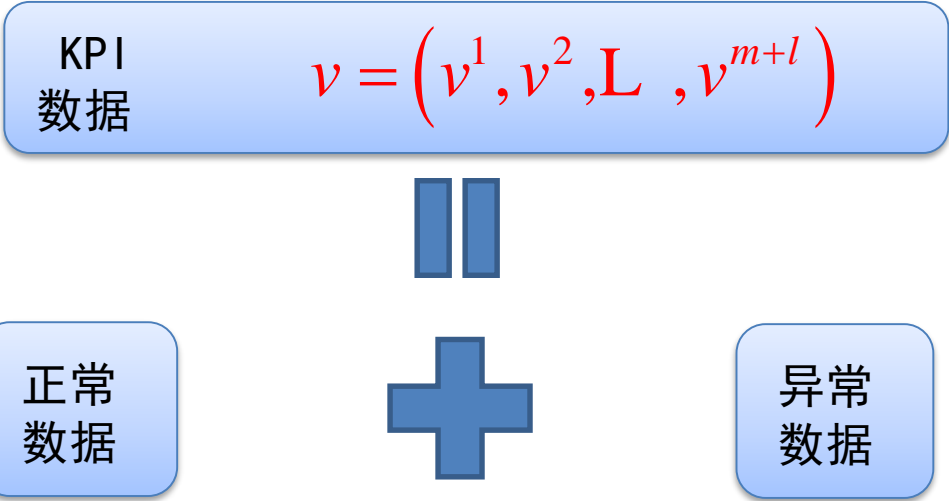


2. KPID算法框架

步骤1：根据人工标记，定位正常和异常区间

输入：带标记的KPI时间序列

输出：正常的KPI数据、异常的KPI数据



$$v_{normal} = (v^1_{normal}, v^2_{normal}, L, v^m_{normal}) \quad v_{abnormal} = (v^1_{abnormal}, v^2_{abnormal}, L, v^l_{abnormal})$$

2. KPID算法框架

步骤二：计算数据的P特征，I特征，D特征

输入：正常的KPI数据、异常的KPI数据

输出：数据的P特征，I特征，D特征及标签

正常
数据

$$v_{normal} = (v_{normal}^1, v_{normal}^2, L, v_{normal}^m)$$



P特征: $P_{normal} = (v_{normal}^1, v_{normal}^2, L, v_{normal}^m)$

I特征: $I_{normal} = \left(\sum_1^{1+T} v_{normal}^i, L, \sum_{m-T}^m v_{normal}^i \right)$

D特征: $D_{normal} = (v_{normal}^2 - v_{normal}^1, L, v_{normal}^m - v_{normal}^{m-1})$

异常特征来源分两类：

1. 计算特征：从异常数据中计算并筛选出来的
2. 人造特征：超过正常数据特征分布的人造特征

输入：正常

D特征及标签

异常特征数据特征分两类

1. 超过正常数据特征上界的特征
2. 低于正常数据特征下界的特征

P特征：

$$P_{abnormal} = \{v_{abnormal}^i \mid v_{abnormal}^i > c1, v_{abnormal}^i < c2\} \cup \{c3, c4\}$$

I特征：

$$I_{abnormal} = \left\{ \sum_{j=i}^{j=i+T} v_{abnormal}^j \mid \sum_{j=i}^{j=i+T} v_{abnormal}^j > c5, \sum_{j=i}^{j=i+T} v_{abnormal}^j < c6 \right\} \cup \{c7, c8\}$$

D特征：

$$D_{abnormal} = \{v_{abnormal}^i - v_{abnormal}^{i-1} \mid v_{abnormal}^i - v_{abnormal}^{i-1} > c9, v_{abnormal}^i - v_{abnormal}^{i-1} < c10\} \cup \{c11, c12\}$$

2. KPID算法框架

参数选择

$$c1 = \max(P_{normal})$$

$$c2 = \min(P_{normal})$$

$$c3 = \max(P_{normal}) + 0.25 * (\max(P_{normal}) - \min(P_{normal}))$$

$$c4 = \min(P_{normal}) - 0.25 * (\max(P_{normal}) - \min(P_{normal}))$$

T:10

$$c5 = \max(I_{normal})$$

$$c6 = \min(I_{normal})$$

$$c7 = \max(I_{normal}) + 0.25 * (\max(I_{normal}) - \min(I_{normal}))$$

$$c8 = \min(I_{normal}) - 0.25 * (\max(I_{normal}) - \min(I_{normal}))$$

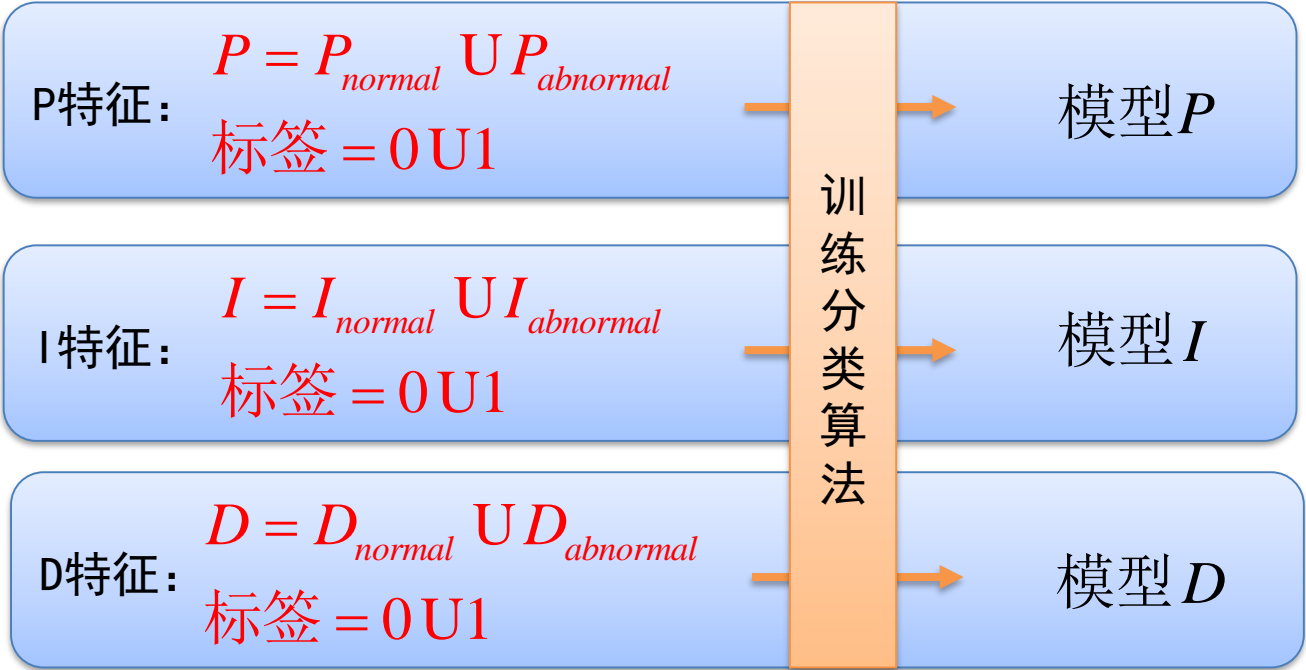
$$c9 = -c10$$

$c9$ 通过网格搜索方法确定，搜索范围 $[0, 10 * \text{mean}(\text{abs}(D_{normal}))]$

$$c11 = \max(D_{normal})$$

$$c12 = \min(D_{normal})$$

2. KPID算法框架



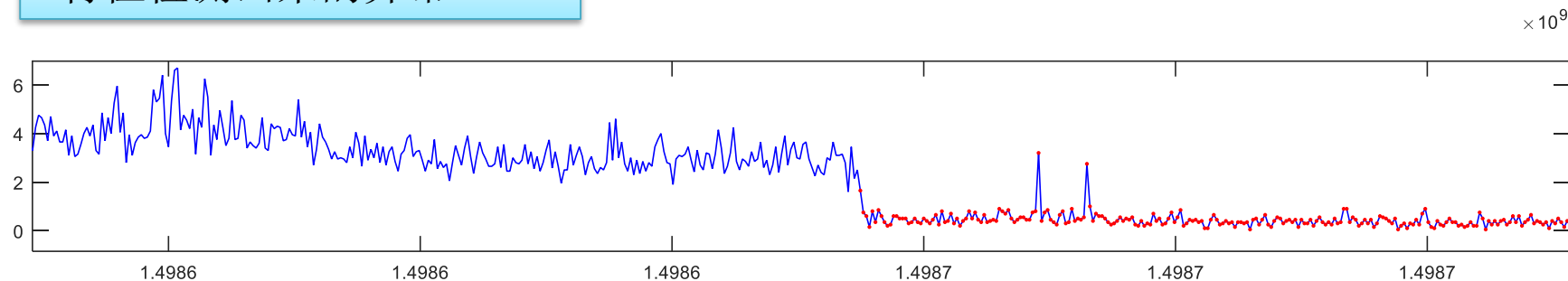
1. sklearn.ensemble : AdaBoostClassifier
2. sklearn.ensemble: RandomForestClassifier

2. KPID算法框架

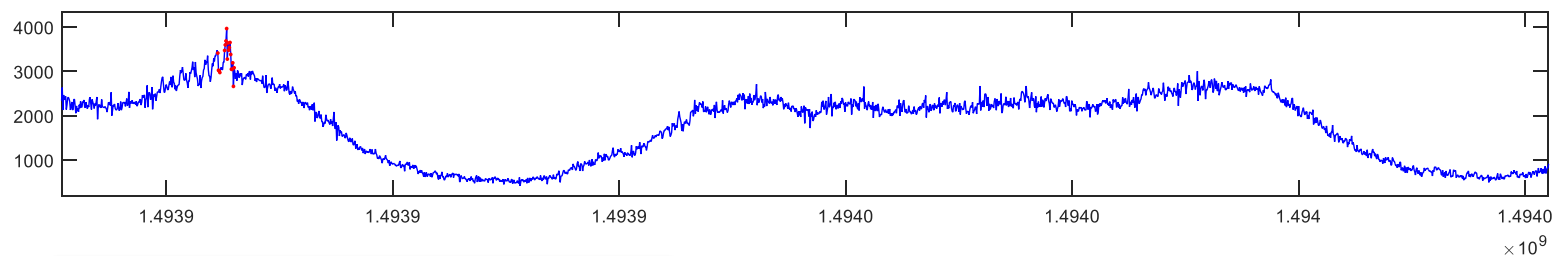


3. KPID算法应用示例

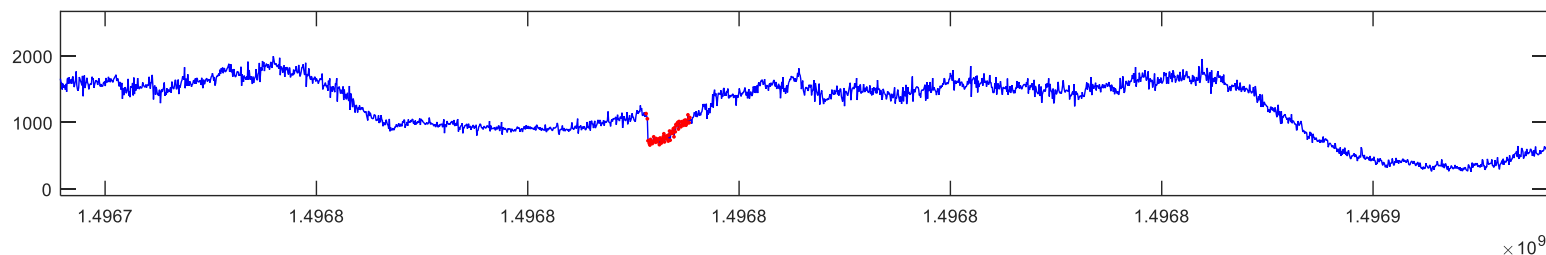
P特征检测出来的异常



I特征检测出来的异常



D特征检测出来的异常



4. KPID算法特点

特征简单

- 三种特征：P特征，I特征，D特征
- 物理含义明确，易于一线运维人员理解与接受

复杂度低

- 每个特征为一维特征
- 特征通过加减运算即可获取
- 特征之间，模型之间无耦合

调参容易

- 只需对一个参数调参
- 根据经验， $[0,1]$ 间，0.1的搜索步长

弱假设

- 只需假设异常数据特征在异常数据特征的数值区间之外
- 分类结果对分类算法鲁棒

Thank you !