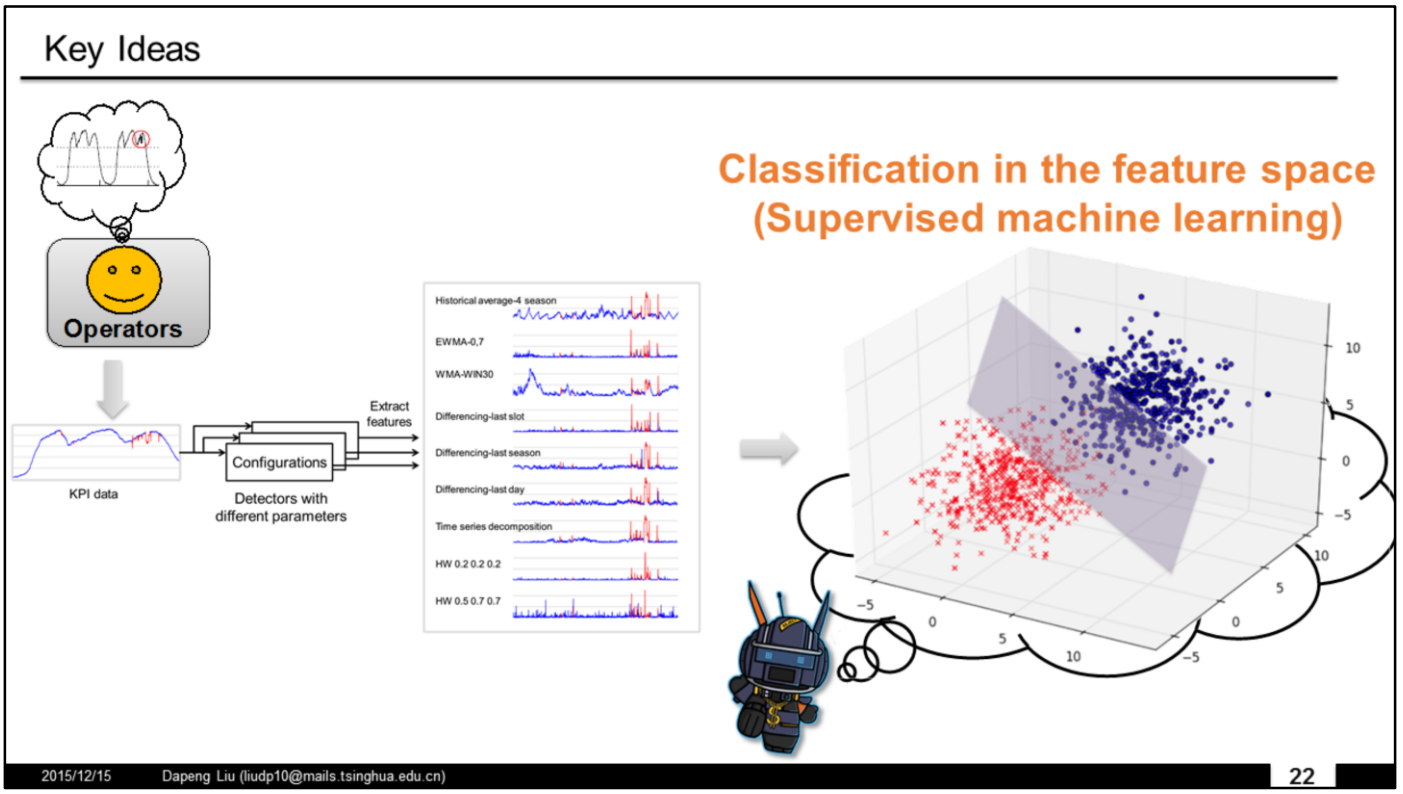
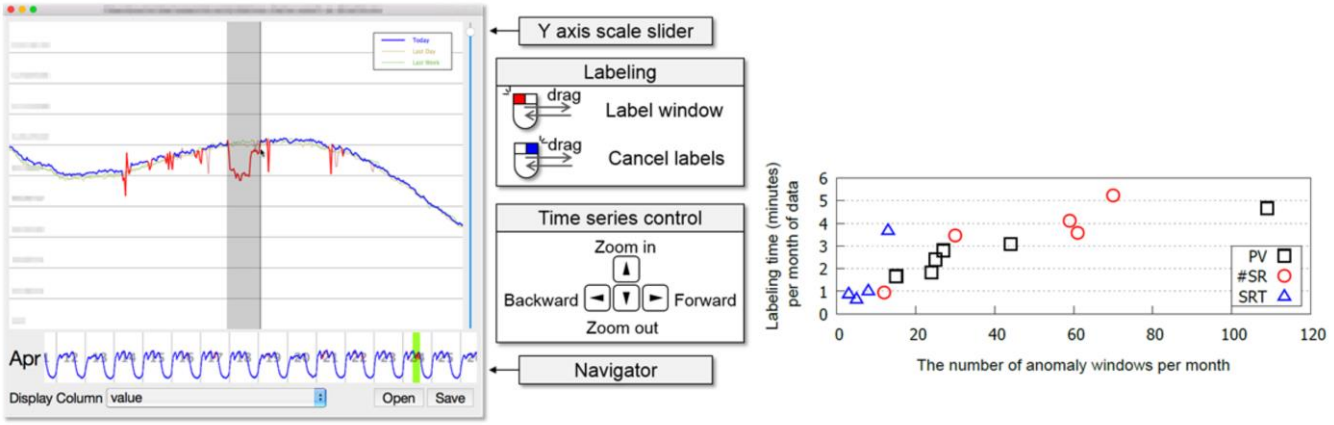


- So we see, the key idea behind this design is that, operators actually have the domain knowledge of anomalies in their mind, but it is difficult for them to formally define it. So, we use machine learning to model those anomaly concepts from historical cases.
- This space is just like the brain of Opprentice.
- In this process, machine learning will automatically find the classification boundaries in the feature space. In other words, it can select which detectors and parameters are suitable for detecting those anomalies cared by operators.

Address Challenges of Designing Opprentice

- Labeling overhead
  - Solution: an effective labeling tool



There are several challenges when designing Opprentice.

- The first one is how to solve the labeling overhead? Our solution is to develop an effective labeling tool. This is the user interface of our labeling tool.

- Labeling with this tool won't cost much time, actually less than 6 minutes for each month of data according to our use experience.

Address Challenges of Designing Opprentice

- Labeling overhead
  - Solution: an effective labeling tool
- Incomplete anomaly types in the historical data
  - Solution: incremental re-training with new data

- The second one is that the historical data may not contain all kinds of anomaly types. Our solution is to incrementally re-train Opprentice with the latest data

Address Challenges of Designing Opprentice

- Labeling overhead
  - Solution: an effective labeling tool
- Incomplete anomaly types in the historical data
  - Solution: incremental re-training with new data
- Class imbalance problem
  - Solution: adjusting classification threshold (cThld) based on the preference

- The third challenge is class imbalance problem. That is, anomalies are infrequent in the data, so they are often ignored by classification algorithms.
- To solve this challenge, we adjust the classification threshold based on the accuracy preference to give higher priority of the anomaly class

PC-Score: Adjust the Classification Threshold (cThld)

$$Precision = \frac{\#\ of\ true\ anomlies\ detected}{\#\ of\ anomlies\ detected}$$

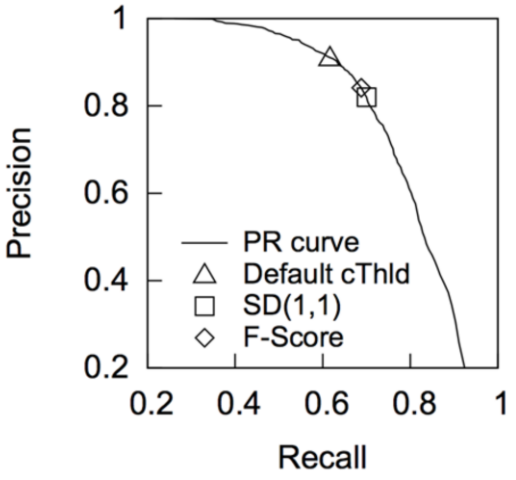
$$Recall = \frac{\#\ of\ true\ anomlies\ detected}{\#\ of\ true\ anomlies}$$

Precision and recall are two important metrics to evaluate the detection accuracy. Precision is how many anomalies detected are the true anomalies. Recall is how many true anomalies are detected. Empirically, precision and recall are often a trade-off.

PC-Score: Adjust the Classification Threshold (cThld)

$$Precision = \frac{\#\ of\ true\ anomlies\ detected}{\#\ of\ anomlies\ detected}$$

$$Recall = \frac{\#\ of\ true\ anomlies\ detected}{\#\ of\ true\ anomlies}$$



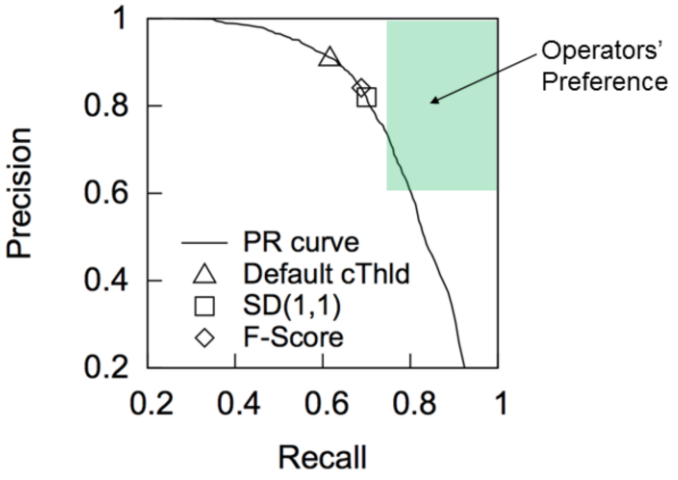
This curve is the precision-recall curve, or PR-curve of a random forest, each point on the curve is derived from a different classification threshold. To select a proper point on the curve, there are several methods, such as using the default cThld 0.5, selecting the point that has the shortest distance to the top-right corner, or the point that can maximize the F-Score.

However, these methods do not take operators' preference into considerations.

PC-Score: Adjust the Classification Threshold (cThld)

$$Precision = \frac{\#\ of\ true\ anomlies\ detected}{\#\ of\ anomlies\ detected}$$

$$Recall = \frac{\#\ of\ true\ anomlies\ detected}{\#\ of\ true\ anomlies}$$

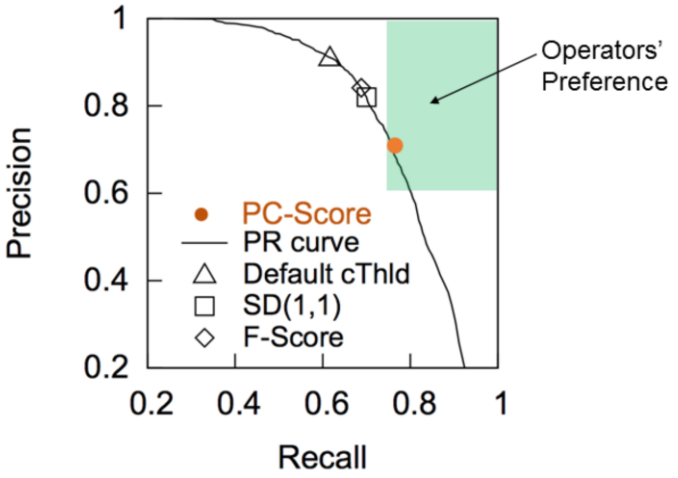


For example, suppose that this green region represents operators' accuracy preferences. We see the PR-curve in fact has points inside these regions, which means it can satisfy the preference. But, the traditional methods will ignore the preferences and won't select the correct points or the corresponding cThld.

PC-Score: Adjust the Classification Threshold (cThld)

$$Precision = \frac{\#\ of\ true\ anomlies\ detected}{\#\ of\ anomlies\ detected}$$

$$Recall = \frac{\#\ of\ true\ anomlies\ detected}{\#\ of\ true\ anomlies}$$



$$\boldsymbol{PC\text{-}Score} = \begin{cases} F\text{-}Score + 1 & ,if\ the\ point\ satisfies\ the\ preference \\ F\text{-}Score & ,otherwise \end{cases}$$

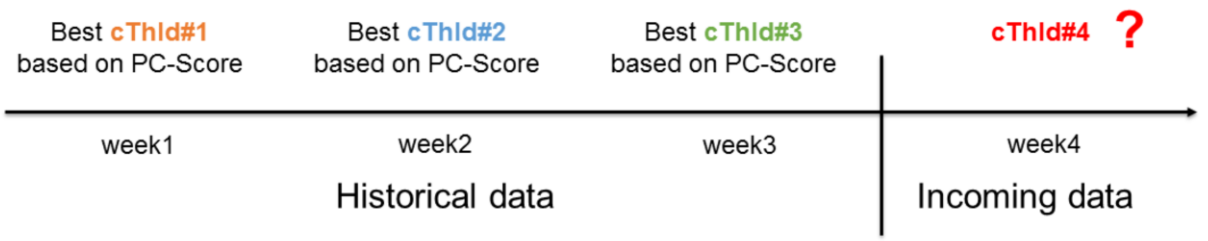
(Preference-centric Score)

2015/12/15 Dapeng Liu (liudp10@mails.tsinghua.edu.cn) 29

Therefore, we propose the Preference-centric score, or the PC-Score. PC-Score is based on F-Score, but it takes operators' preference into considerations. The main idea is intuitive. We give higher priority to those points that satisfy the preference by adding the F-Score by one.

So the cThld selected by the PC-Score can satisfy the preference if possible.

EWMA: Predict cThld based on PC-Score



Exponentially weighted moving average (EWMA)

$$\text{cThld}_i^p = \begin{cases} \alpha \cdot \text{cThld}_{i-1}^b + (1-\alpha) \cdot \text{cThld}_{i-1}^p & , i > 1 \\ \text{5-fold prediction} & , i = 1 \end{cases}$$

OK, now based on the PC-Score, we can determine the best cThld for each week if we have the data of that week.

**Click
Then, a question is that for online detection, how can we determine the best cThld for a future week, where we do not have the data yet.

**Click
To solve this problem, we here use **exponentially weighted MA** to predict the best cThld of next week based on those historical best cThld.

Address Challenges of Designing Opprentice

- Labeling overhead
  - Solution: an effective labeling tool
- Incomplete anomaly types in the historical data
  - Solution: incremental re-training with new data
- Class imbalance problem
  - Solution: adjusting classification threshold (cThld) based on the preference
- Irrelevant and redundant features
  - Solution: random forests

- And Last, because we want to save manual efforts, the detectors and their parameters are used without carefully evaluation, so many of the features provided by these detectors could be irrelevant or redundant.

- To solve this problem, we conducted pilot experiements on different machine learning algorithms, and find that the random forest, an ensemble based machine learning algorithm, turns out to be more accurate and robust to irrelevant and redundant features. The last speaker has also confirmed this in their paper.