

PRECISION AND RECALL FOR TIME SERIES



Nesime
Tatbul



Mejbah
Alam



Justin
Gottschlich



Tae Jun
Lee



Stan
Zdonik



BROWN

Motivation: Time Series Anomaly Detection

- **Anomaly:** Patterns that do not conform to expected behavior.
- Anomalies can have critical impact: loss of life, property damage, monetary loss, ...
- Applications of anomaly detection (AD) are numerous and diverse.

Autonomous Driving



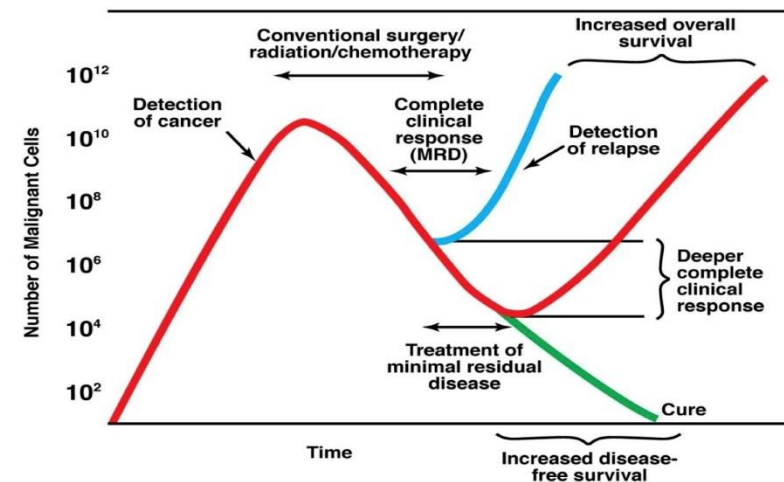
Six levels of autonomy:

- L0: No automation
- L1: Driver assistance
- L2: Partial automation
- L3: Conditional automation
- L4: High automation
- L5: Full automation

L3+ autonomy
requires robust
AD systems.

Source: Society of Automotive Engineers (SAE),
National Highway and Traffic Safety Administration (NHTSA)

Cancer Detection



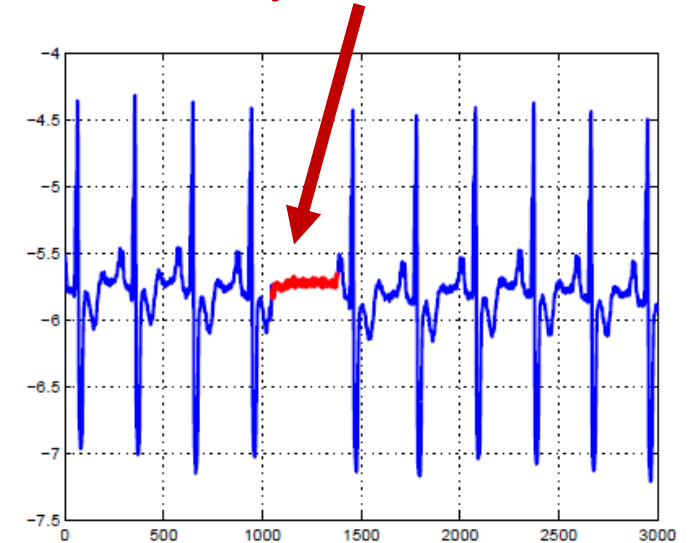
Anomalies
often occur
over a
period of time.

Source:
<http://www.vaccinogeninc.com/oncovax/science-due-diligence/overview-part-1>

Motivation: Range-based Anomalies

- Time series anomalies are **range based**, i.e., they occur over a period of time.
- There are **domain-specific application preferences**.
 - Cancer detection, Real-time systems:
 - Early response; Avoid false negatives!
 - Robotic defense systems:
 - Delayed response; Avoid false positives!
 - Emergency braking in self-driving cars:
 - Neither too early nor too late; Avoid false negatives!

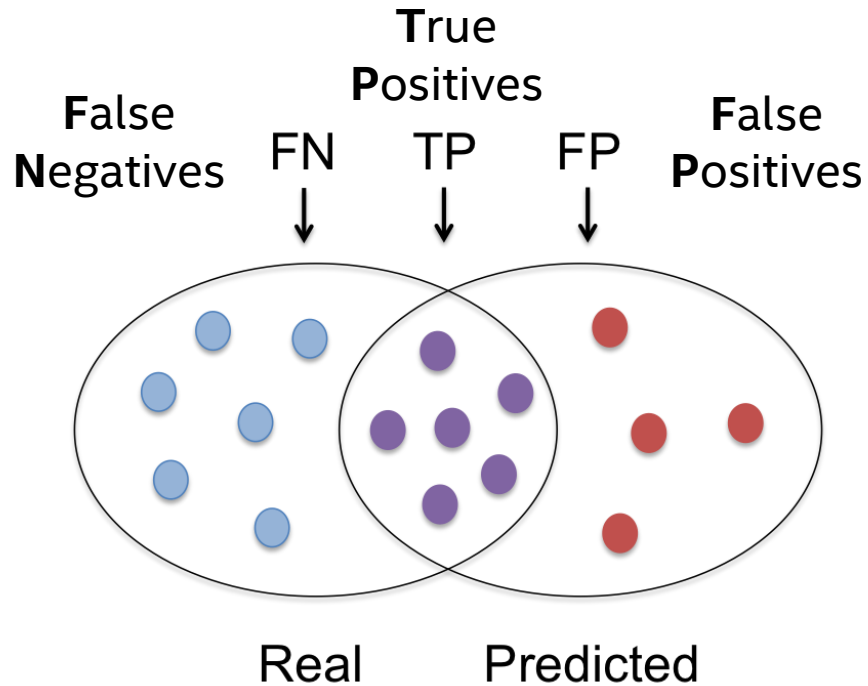
Atrial Premature Contraction
anomaly in human ECG



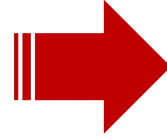
Source: Chandola et al., "Anomaly Detection: A Survey", ACM Computing Surveys, 41(3), 2009.

Problem: How to Measure Accuracy?

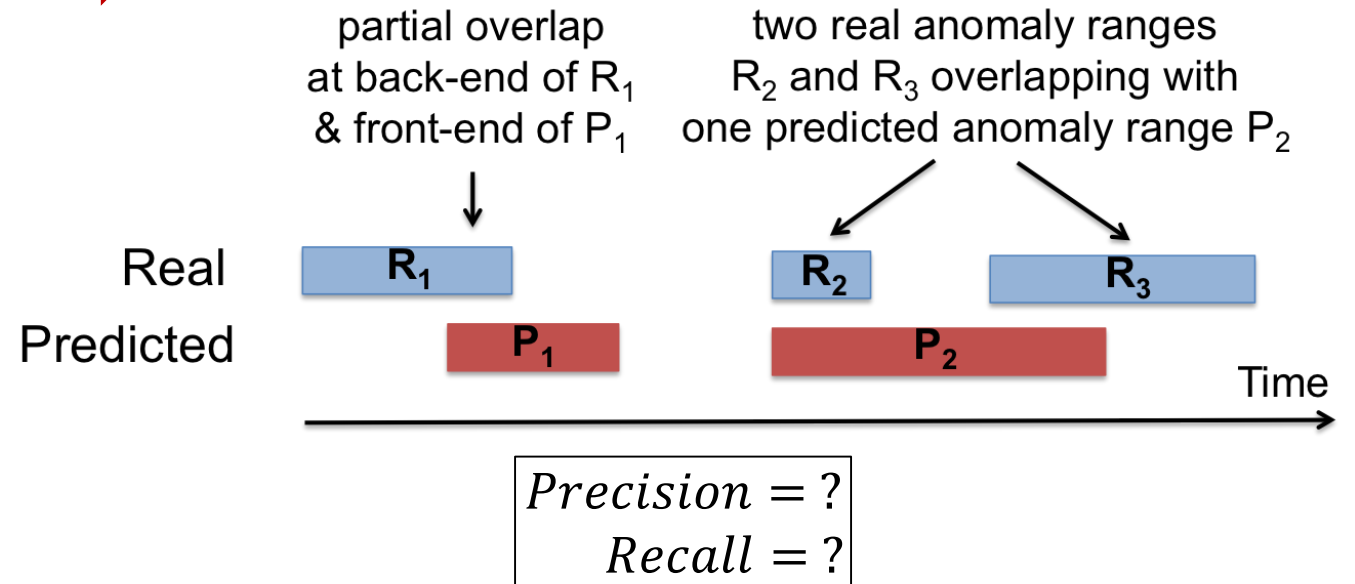
Point-based Anomalies



$$\begin{aligned} \text{Precision} &= TP \div (TP + FP) \\ \text{Recall} &= TP \div (TP + FN) \end{aligned}$$



Range-based Anomalies



- Must express **partial detection**
- Must support **flexible time bias**

State of the Art

- Classical Precision and Recall
 - Point-based anomalies
 - Precision penalizes FP, Recall penalizes FN
 - F_β -Score to combine and weight them
- Numenta Anomaly Benchmark (NAB) [2]
 - Point-based anomalies
 - Focuses specifically on early detection use cases
 - Difficult to use in practice (irregularities, ambiguities, magic numbers) [3]
- Activity recognition metrics
 - No support for flexible time bias

$$F_\beta = (1 + \beta^2) \times \frac{\text{Precision} \times \text{Recall}}{(\beta^2 \times \text{Precision}) + \text{Recall}}$$

β : relative importance of Recall to Precision

$\beta = 1$: evenly weighted (harmonic mean)

$\beta = 2$: weights Recall higher (i.e., no FN!)

$\beta = 0.5$: weights Precision higher (i.e., no FP!)



[2] Lavin and Ahmad, "Evaluating Real-Time Anomaly Detection Algorithms – The Numenta Anomaly Benchmark", IEEE ICMLA, 2015.

[3] Singh and Olinsky, "Demistifying Numenta Anomaly Benchmark", IEEE IJCNN, 2017.

Precision and Recall for Time Series

Customizable
parameters



Notation	Description
R, R_i	set of real anomaly ranges, the i^{th} real anomaly range
P, P_j	set of predicted anomaly ranges, the j^{th} predicted anomaly range
N, N_r, N_p	number of all points, number of real anomaly ranges, number of predicted anomaly ranges
α	relative weight of existence reward
$\gamma(), \omega(), \delta()$	overlap cardinality function, overlap size function, positional bias function

Range-based
Recall

$$Recall_T(R, P) = \frac{\sum_{i=1}^{N_r} Recall_T(R_i, P)}{N_r}$$

$$Recall_T(R_i, P) = \alpha \times ExistenceReward(R_i, P) + (1 - \alpha) \times OverlapReward(R_i, P)$$

$$ExistenceReward(R_i, P) = \begin{cases} 1, & \text{if } \sum_{j=1}^{N_p} |R_i \cap P_j| \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$OverlapReward(R_i, P) = CardinalityFactor(R_i, P) \times \sum_{j=1}^{N_p} \omega(R_i, R_i \cap P_j, \delta)$$

$$CardinalityFactor(R_i, P) = \begin{cases} 1, & \text{if } R_i \text{ overlaps with at most one } P_j \in P \\ \gamma(R_i, P), & \text{otherwise} \end{cases}$$

Range-based
Precision

$$Precision_T(R, P) = \frac{\sum_{i=1}^{N_p} Precision_T(R, P_i)}{N_p}$$

$$Precision_T(R, P_i) = CardinalityFactor(P_i, R) * \sum_{j=1}^{N_r} \omega(P_i, P_i \cap R_j, \delta)$$

- We extend classical Precision and Recall to measure ranges.
- Our model is:
 - expressive
 - flexible
 - extensible

Customization Examples

Overlap Size $\omega()$

```
function  $\omega$ (AnomalyRange, OverlapSet,  $\delta$ )
  MyValue  $\leftarrow$  0
  MaxValue  $\leftarrow$  0
  AnomalyLength  $\leftarrow$  length(AnomalyRange)
  for i  $\leftarrow$  1, AnomalyLength do
    Bias  $\leftarrow$   $\delta(i, \text{AnomalyLength})$ 
    MaxValue  $\leftarrow$  MaxValue + Bias
    if AnomalyRange[i] in OverlapSet then
      MyValue  $\leftarrow$  MyValue + Bias
  return MyValue/MaxValue
```

Positional Bias $\delta()$

```
function  $\delta(i, \text{AnomalyLength})$  ▷ Flat bias
  return 1
function  $\delta(i, \text{AnomalyLength})$  ▷ Front-end bias
  return AnomalyLength - i + 1
function  $\delta(i, \text{AnomalyLength})$  ▷ Back-end bias
  return i
function  $\delta(i, \text{AnomalyLength})$  ▷ Middle bias
  if i  $\leq$  AnomalyLength/2 then
    return i
  else
    return AnomalyLength - i + 1
```

Cancer Detection:

- Set $\delta() = \text{Front-end}$, $\beta = 2$

Robotic Defense:

- Set $\delta() = \text{Back-end}$, $\beta = 0.5$

Emergency Braking:

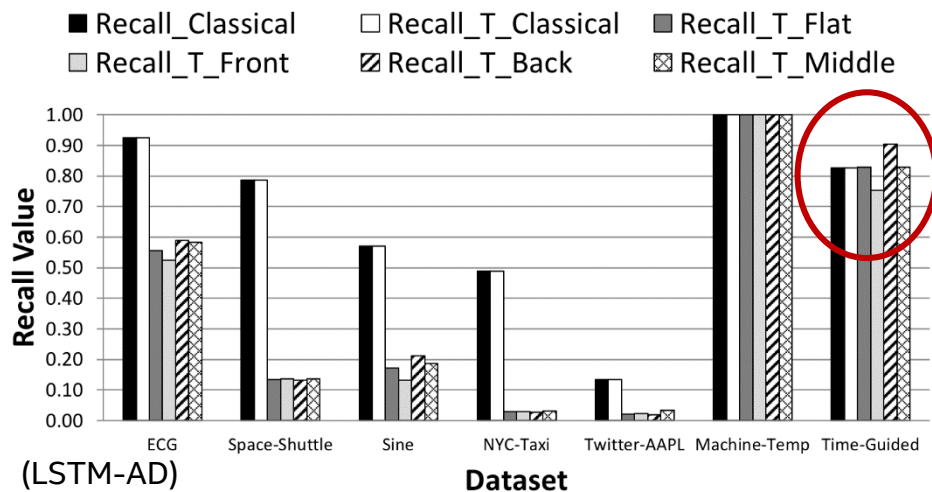
- Set $\delta() = \text{Middle}$, $\beta = 1.5$

Our model subsumes the classical point-based model, when:

- all ranges are represented as unit-size ranges, and
- $\alpha=0$, $\gamma=1$, $\omega()$ is as above, and $\delta() = \text{Flat}$

Selected Experimental Results

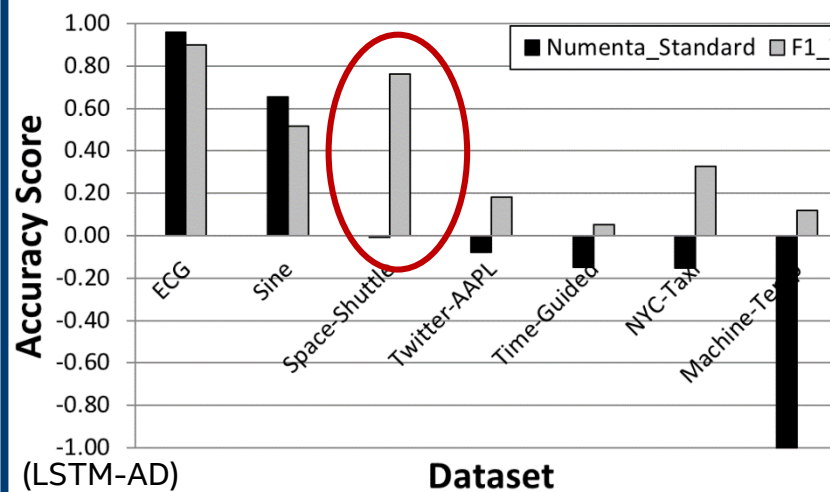
Comparison to Classical model



Our model

- subsumes the classical model
- is sensitive to positional bias

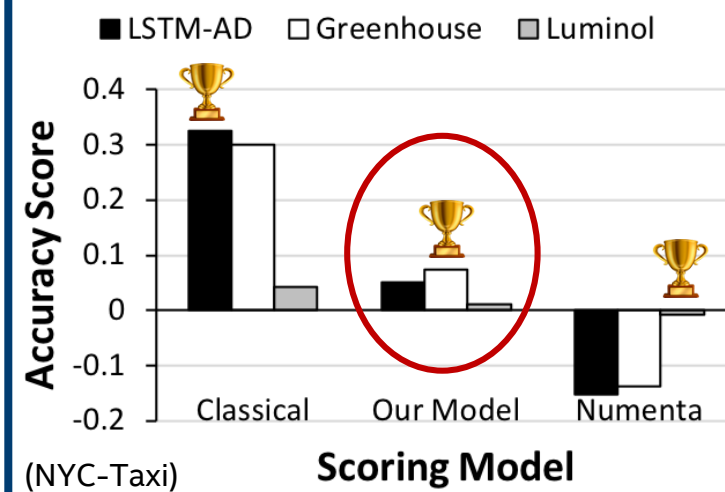
Comparison to Numenta model



Our model can

- mimic the Numenta model
- catch additional intricacies

Multiple Anomaly Detectors



Our model is more effective in

- evaluating multiple detectors
- capturing subtleties in data

Please see our paper for details of this experimental study and additional results.

Key Takeaways

- This work extends the classical Precision and Recall model to time series data.
- We provide tunable parameters to capture domain-specific application preferences.
- Experiments with diverse datasets and anomaly detectors prove the benefits of our approach.
- Future work includes:
 - designing new training strategies for range-based anomaly detection
 - exploring use in other time series classification tasks and applications

More Information

Watch our short video:

<https://www.youtube.com/watch?v=K5f-dUBiQP4>

Read our paper:

<https://arxiv.org/abs/1803.03639/>

Download our tool:

<https://github.com/IntelLabs/TSAD-Evaluator/>

Visit our poster session at NeurIPS'18:

Today at 5:00 - 7:00 PM in Room 210 & 230 AB #116

Thanks to Intel and NSF for funding this research.