The Problem

- Anomaly detection (AD) is the process of finding patterns that do not conform to expected behavior.
- Time series anomalies occur over a range of time. Time bias is domain-specific.
- Goal: To design a model for evaluating, ranking, comparing the classification accuracy of time series AD algorithms.

Point-based Anomalies

- False Negatives
- True Positives
- FN
- TP
- FP
- FN
- TP
- FP

Real

Predicted

Range-based Anomalies

- partial overlap at back-end of R₁ & front-end of P₁
- two real anomaly ranges R₂ and R₃ overlapping with one predicted anomaly range P₂

Domain-specific Time Bias

- Recombinant range detections
- Domain-specific bias

The Solution

Range-based Recall

\[
\text{Recall}(R, P) = \frac{\sum_{R \in R} \text{Recall}(R, P)}{\sum_{P \in P} \text{Recall}(R, P)}
\]

\[
\text{Recall}(R, P) = \alpha \cdot \text{ExistenceReward}(R, P) + \beta \cdot \text{OverlapReward}(R, P)
\]

\[
\text{ExistenceReward}(R, P) = \begin{cases} 
1 \text{ if } \sum_{P \in P} |R \cap P| \geq 1, \\
0 \text{ otherwise.}
\end{cases}
\]

\[
\text{OverlapReward}(R, P) = \text{CardinalityFactor}(R, P) \cdot \sum_{P \in P} |R \cap P|, \\
\text{CardinalityFactor}(R, P) = \begin{cases} 
1 \text{ if } R \text{ overlaps with at most one } P \in P, \\
\frac{|R|}{|R|}, \text{ otherwise.}
\end{cases}
\]

Range-based Precision

\[
\text{Precision}(R, P) = \frac{\sum_{R \in R} \text{Precision}(R, P)}{\sum_{P \in P} \text{Precision}(R, P)}
\]

\[
\text{Precision}(R, P) = \text{CardinalityFactor}(P, R) \cdot \sum_{P \in P} |R \cap P|, \\
\text{CardinalityFactor}(P, R) = \begin{cases} 
1 \text{ if } P \text{ overlaps with at most one } R \in R, \\
\frac{|P|}{|P|}, \text{ otherwise.}
\end{cases}
\]

Experimental Results

- Datasets with labels:
  - Real: NAB Data Corpus
  - Synthetic: Paranom Tool
- Time series AD system:
  - LSTM on TensorFlow
- Our range-based metrics can be computed efficiently (see paper for cost analysis).

Our Model vs. the Classical Point-based Model

- Our model subsumes the classical point-based model, when:
  - all Rᵢ and Pⱼ are represented as unit-size ranges, and
  - α = 0, β = 1, γ(γ) = 1, ω(ω) is as defined above, and δ(δ) is flat.

Conclusion

- Our new accuracy model for time series AD is expressive, flexible, and extensible.
- Ongoing work includes:
  - developing new ML training strategies optimized for our model (see Greenhouse [SysML’18])
  - applying our model on real-world use cases (e.g., autonomous driving)
  - creating an open-source benchmarking suite for time series AD