Precision and Recall for Range-Based Anomaly Detection

Tae Jun Lee
Microsoft

Justin Gottschlich, Nesime Tatbul
Intel Labs

Eric Metcalf, Stan Zdonik
Brown University

ABSTRACT

Classical anomaly detection is principally concerned with point-based anomalies, anomalies that occur at a single data point. In this paper, we present a new mathematical model to express range-based anomalies, anomalies that occur over a range (or period) of time.

1 INTRODUCTION

Anomaly detection (AD) seeks to identify atypical events. Anomalies tend to be domain or problem specific, and many occur over a period of time. We refer to such events as range-based anomalies, as they occur over a range (or period) of time\(^1\). Therefore, it is critical that the accuracy measures for anomalies, and the systems detecting them, capture events that occur over a range of time. Unfortunately, classical metrics for anomaly detection were designed to handle only fixed-point anomalies [1]. An AD algorithm behaves much like a pattern recognition and binary classification algorithm: it recognizes certain patterns in its input and classifies them as either normal or anomalous. For this class of algorithms, Recall and Precision are widely used for evaluating the accuracy of the result. They are formally defined as in Equations 1 and 2, where \( TP \) denotes true positives, \( FP \) denotes false positives, and \( FN \) denotes false negatives.

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (1) \\
\text{Precision} = \frac{TP}{TP + FP} \quad (2)
\]

While useful for point-based anomalies, classical recall and precision suffer from their inability to capture, and bias, classification correctness for domain-specific time-series anomalies. Because of this, many time-series AD systems’ accuracy are being misrepresented, as point-based recall and precision are used to measure their effectiveness [9]. Furthermore, the need to accurately identify time-series anomalies is growing due to the explosion of streaming and real-time systems [2, 4, 7, 8, 10]. To address this, we redefine recall and precision to encompass range-based anomalies. Unlike prior work [2, 6], our mathematical definitions are a superset of the classical definitions, enabling our system to subsume point-based anomalies. Moreover, our system is broadly generalizable, providing specialization functions to control a domain’s bias along a multi-dimensional axis that is necessary to accommodate the needs of specific domains.

In this short paper, we present novel formal definitions of recall and precision for range-based anomaly detection that both subsume those formerly defined for point-based anomaly detection as well as being customizable to a rich set of application domains. Empirical data has been obtained to meet the venue’s compressed format.

2 RANGE-BASED RECALL

Table 1: Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R )</td>
<td>set of real anomaly ranges</td>
</tr>
<tr>
<td>( R_i )</td>
<td>the ( i )-th real anomaly range</td>
</tr>
<tr>
<td>( P )</td>
<td>set of predicted anomaly ranges</td>
</tr>
<tr>
<td>( P_j )</td>
<td>the ( j )-th predicted anomaly range</td>
</tr>
<tr>
<td>( N_r )</td>
<td>number of real anomaly ranges</td>
</tr>
<tr>
<td>( N_p )</td>
<td>number of predicted anomaly ranges</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>relative weight of existence reward</td>
</tr>
<tr>
<td>( \beta )</td>
<td>relative weight of overlap reward</td>
</tr>
<tr>
<td>( \gamma() )</td>
<td>overlap cardinality function</td>
</tr>
<tr>
<td>( \omega() )</td>
<td>overlap size function</td>
</tr>
<tr>
<td>( \delta() )</td>
<td>positional bias function</td>
</tr>
</tbody>
</table>

Classical Recall rewards an AD system when anomalies are successfully identified (i.e., TP) and penalizes it when they are not (i.e., FN). It is computed by counting the number of anomalous points successfully predicted and then dividing that number by the total number of anomalous points. However, it is not sensitive to domains where a single anomaly can be represented as a range of contiguous points. In this section, we propose a new way to compute recall for such range-based anomalies. Table 1 summarizes our notation.

Given a set of real anomaly ranges \( R = \{R_1, \ldots, R_{N_r}\} \) and a set of predicted anomaly ranges \( P = \{P_1, \ldots, P_{N_p}\} \), our \( \text{Recall} \) formulation iterates over the set of all real anomaly ranges \( R \), computing a recall score for each real anomaly range \( (R_i \in R) \) and adding them up into a total recall score. This total score is then divided by the total number of real anomalies \( (N_r) \) to obtain an average recall score for the whole time-series.

\[
\text{Recall}(R, P) = \frac{\sum_{i=1}^{N_r} \text{Recall}(R_i, P)}{N_r} \quad (3)
\]

When computing the recall score \( \text{Recall}(R_i, P) \) for a single real anomaly range \( R_i \), we take the following aspects into account:

- **Existence**: Identifying an anomaly (even by a single point in \( R_i \)) may be valuable in some application domains.
- **Size**: The larger the size of the correctly predicted portion of \( R_i \), the higher the recall score will likely be.
- **Position**: In some cases, not only size, but also the relative position of the correctly predicted portion of \( R_i \) may be important to the application (e.g., early and late biases).
- **Cardinality**: Detecting \( R_i \) with a single predicted anomaly range \( P_j \in P \) may be more valuable to an application than doing so with multiple different ranges in \( P \).

We capture these aspects as a sum of two reward terms weighted by \( \alpha \) and \( \beta \), respectively, where \( 0 \leq \alpha, \beta \leq 1 \) and \( \alpha +\beta = 1 \). \( \alpha \) represents the relative importance of rewarding existence, whereas \( \beta \) represents the relative importance of rewarding size, position, and

\(^1\) The work was done while a Brown student.

\(^2\) Range-based anomalies are a specific type of collective anomalies [3]. Moreover, range-based anomalies are similar, but not identical, to sequence anomalies [11].
Additionally, an overlap reward, dependent upon three application-defined functions \(0 \leq \gamma() \leq 1\), \(0 \leq \omega() \leq 1\), and \(\delta() \geq 1\), can be earned. These functions capture the cardinality \((\gamma)\), size \((\omega)\), and position \((\delta)\) of the overlap. The cardinality term serves as a scaling factor for the rewards earned from size and position of the overlap.

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

The cardinality factor is largest \((i.e., 1)\), when \(R_i\) overlaps with at most one predicted anomaly range \(i.e., it is identified by a single prediction range\). Otherwise, it receives a value \(0 \leq \gamma() \leq 1\) defined by the application.

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

The Recall\(_T\) constants \((\alpha\) and \(\beta\)) and functions \((\gamma(), \omega(), \text{ and } \delta())\) are tunable according to the needs of the application. Next, we illustrate how they can be customized with examples.

The cardinality factor should generally be inversely proportional to \(\text{Card}(R_i)\), i.e., the number of distinct prediction ranges that a real anomaly range \(R_i\) overlaps. For example, \(\gamma(R_i, P)\) can simply be set to \(1/\text{Card}(R_i)\).

Figure 1a shows an example for the \(\omega()\) function for size, which can be used with many different \(\delta()\) functions for positional bias as shown in Figure 1b. If all index positions are equally important, then the flat bias function should be used. If earlier ones are more important than later ones \(e.g.,\) early cancer detection \([5]\), real-time apps \([2]\), then the front-end bias function should be used. Finally, if later index positions are more important \(e.g.,\) delayed response in robotic defense), then the tail-end bias function should be used.

Our recall formula for range-based anomalies subsumes the classical one for point-based anomalies \(\text{Recall}_T \equiv \text{Recall}\) when:

(i) all \(R_i \in R\) and \(P_j \in P\) are represented as single-point ranges \(\text{e.g., range } [1, 3] \text{ represented as } [1, 1], [2, 2], [3, 3]\), and

(ii) \(\alpha = 0, \beta = 1, \gamma() = 1, \omega()\) is as in Figure 1a, and \(\delta()\) returns flat positional bias as in Figure 1b.

### 3 RANGE-BASED PRECISION

Classical \(\text{Precision}\) is computed by counting the number of successful prediction points \(\text{i.e., TP}\) in proportion to the total number of prediction points \(\text{i.e., TP + FP}\). The key difference between \(\text{Precision}\) and \(\text{Recall}\) is that \(\text{Precision}\) penalizes FPs. In this section, we extend classical precision to handle range-based anomalies. Our formulation follows a similar structure as \(\text{Recall}\).

Given a set of real anomaly ranges \(R = \{R_1, \ldots, R_{N_p}\}\) and a set of predicted anomaly ranges \(P = \{P_1, \ldots, P_{N_p}\}\), \(\text{Precision}_T(R, P)\) iterates over the set of predicted anomaly ranges \(P\), computing a precision score for each range \(P_i \in P\) and then sums them. This sum is then divided by the total number of predicted anomalies \(N_p\), averaging the score for the whole time-series.

\[
\text{Precision}_T(R, P) = \frac{\sum_{i=1}^{N_p} \text{Precision}(R, P_i)}{N_p}
\]

When computing \(\text{Precision}_T(R, P_i)\) for a single predicted anomaly range \(P_i\), there is no need for an \(\text{existence}\) reward, because precision by definition emphasizes precision quality, and existence by itself is too low a bar for judging the quality of a prediction. This removes the need for \(\alpha\) and \(\beta\) constants. Therefore:

\[
\text{Precision}(R, P_i) = \text{CardinalityFactor}(P_i, R) \\
\times \sum_{j=1}^{N_c} \omega(P_i, P_j \cap R_j, \delta)
\]

\(\gamma(), \omega(), \text{ and } \delta()\) are customizable as before. Furthermore, \(\text{Precision}_T \equiv \text{Precision}\) under the same settings as in Section 2 (except \(\alpha\) and \(\beta\) are not needed). Note that, while \(\delta()\) provides a potential knob for positional bias, we believe that in many domains a flat bias function will suffice for \(\text{Precision}_T\), as an FP is typically considered uniformly bad wherever it appears in a prediction range.

### 4 CONCLUSION

In this paper, we note that traditional recall and precision were invented for point-based analysis. In range-based anomaly detection, anomalies are not necessarily single points, but are, in many cases, ranges. In response, we offered new recall and precision definitions that take ranges into account.

**Acknowledgments.** This research has been funded in part by Intel.
REFERENCES


