1 Compact Description For Anomalies

Providing explanations for anomalies is important and some recent work (among others) can be found in [1]. We illustrate a different and yet very simple approach when AAD is used with *Isolation Forest*.

The Isolation Forest version of AAD works by partitioning the space into smaller subspaces and weights each subspace to fit the expert feedback. The end result is: (1) a set of default (unsupervised) anomaly scores for each subspace and, (2) weights for each subspace inferred from feedback. We can generate compact descriptions for the anomalies using this information. The idea is to select some combination of (preferably) the smallest subspaces that **cover** all **discovered** anomalies. Therefore, we treat this as an instance of the *set covering* problem. We illustrate this on a toy dataset below.

Assume that there are total m subspaces across all Isolation Forest trees. We represent weights by $\mathbf{w} \in \mathbb{R}^m$, and the corresponding unsupervised anomaly scores by $\mathbf{d} \in \mathbb{R}^m$.

Now, the final anomaly scores for the **subspaces** (not instances), after incorporating feedback, is $\mathbf{a} = \mathbf{w} \circ \mathbf{d} \in \mathbb{R}^m$ where $\circ$ denotes the element wise (*Hadamard*) product.

We sort the scores in $\mathbf{a}$ in descending order and select (say) 30 top ranked subspaces. Next, we retain only those subspaces from this set which contain at least one anomaly **discovered** by the analyst. Let us denote the resulting set of subspaces by $\mathcal{S}$. Further, let $\mathcal{Z}$ be the set of discovered anomalies that belong to one or more subspaces in $\mathcal{S}$, and $|\mathcal{Z}| = n$. (Note that some of the discovered anomalies might not belong to any of the top ranked subspaces we selected.)

Let $|\mathcal{S}| = k$. Denote the *volumes* of the subspaces in $\mathcal{S}$ by the vector $\mathbf{v} \in \mathbb{R}^k$. Now, assume that a binary vector $\mathbf{x} \in \{0, 1\}^k$ contains 1 in locations corresponding to the subspaces in $\mathcal{S}$ which are included in the covering set, and 0 otherwise. Let $\mathbf{u}_z \in \{0, 1\}^k$ denote a vector for each anomaly $z \in \mathcal{Z}$ which contains 1 in all locations corresponding to the subspaces in $\mathcal{S}$ that z belongs to. Let $\mathbf{U} \in \{0, 1\}^{[n \times k]}$ represent the matrix of all $\mathbf{u}$s.

The selection of the compact set of subspaces to describe all discovered anomalies can be formulated as:

$$\underset{\mathbf{x} \in \{0,1\}^k}{\arg\min} \, \mathbf{x} \cdot \mathbf{v}^p \tag{1}$$

$$\text{s.t.} \, \mathbf{U} \cdot \mathbf{x} \geq \mathbf{1}$$

$$\text{where, } \mathbf{1} \text{ is the column vector of } n \text{ 1s, and}$$

$$p \text{ is an integer} \geq 1 \text{ (more below)}$$

The parameter p determines how severely to penalize larger volumes. This is usually 1. However, a larger value will strongly discourage bigger volume subspaces from being selected.

We apply this idea on a toy data shown in Figure 1a. The compact descriptions are shown in Figure 2c.

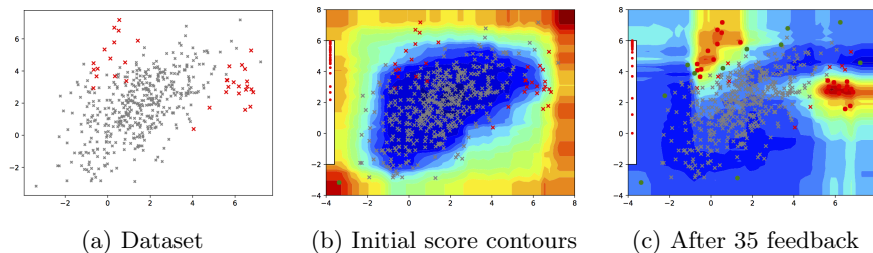| (a) Dataset | (b) Initial score contours | (c) After 35 feedback |

Figure 1: Dataset and score contours. Figure 1a shows a synthetic toy dataset. Figure 1b shows the initial Isolation Forest score contours. Figure 1c shows the score contours after 35 feedback iterations from the Oracle. The red dots are discovered anomalies (true positives). The green dots are discovered nominals (false positives). The red checks are undiscovered anomalies (false negatives).

Note that we have not considered labeled nominals in the optimization objective, but a more sophisticated objective might include them too. For instance, we could add another term in the objective which penalizes whenever a nominal is present in the selected subspace.

A different approach to generate descriptions might train a cost-based decision tree with the labeled anomalies and nominals. Once trained, the extracted rules would correspond to 'descriptions'. This approach is possibly more principled because the decision tree splits are [usually] based on information gain or gini-index and are therefore more informative. The downside is that this would be separate from the anomaly detector and therefore it would not offer much insight into the anomaly detector's behavior – ideally we would like the explanation or description of anomalies to reflect the detector's internals.

2 Querying Diversity

We can use the anomaly description to diversify our queries when we have the option to query labels for more than one instance per round of feedback. Following is how we go about it:

1. Select the top ranked m (say, $m = 15$) instances. Denote this set by $\mathcal{C}$.

2. Next, select the top (say) 5 most anomalous regions for each instance in $\mathcal{C}$. Let $\mathcal{F}$ be the union of all these regions.

3. Now, *compactly* describe all instances in $\mathcal{C}$ using the regions in $\mathcal{F}$.

4. Now, start with the most anomalous instance, and select one-by-one k (say, $k = 5$) instances which have the fewest overlapping regions for querying.
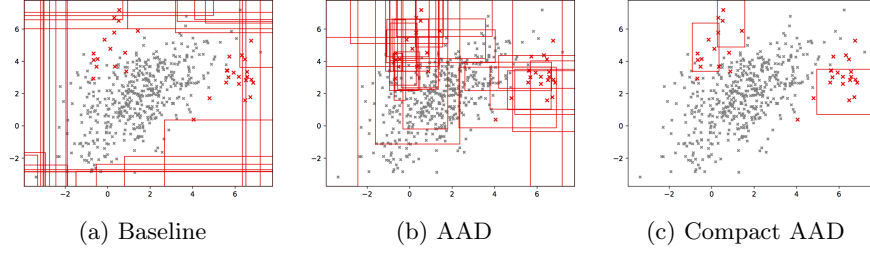
(a) Baseline   (b) AAD   (c) Compact AAD

Figure 2: Top 30 subspaces ranked by $\mathbf{w} \circ \mathbf{d}$. Figure 2a shows the top 30 subspaces with no feedback. We can see that most of these are just the exterior regions of the dataset. Figure 2b shows that after feedback, the most anomalous subspaces are those around the labeled anomalies. Figure 2c shows the most compact set of subspaces (for AAD) which cover all labeled anomalies. These are computed with Equation 1. Note that the compact subspaces only cover anomalies that were discovered in the 35 feedback iterations. Anomalies which were not detected are likely to fall outside these compact subspaces.

References

[1] Meghanath Macha and Leman Akoglu *X-PACS: eXPlaining Anomalies by Characterizing Subspaces*, 2017, http://arxiv.org/abs/1708.05929.

(a) Candidate regions and instances

(b) Simple select-top query
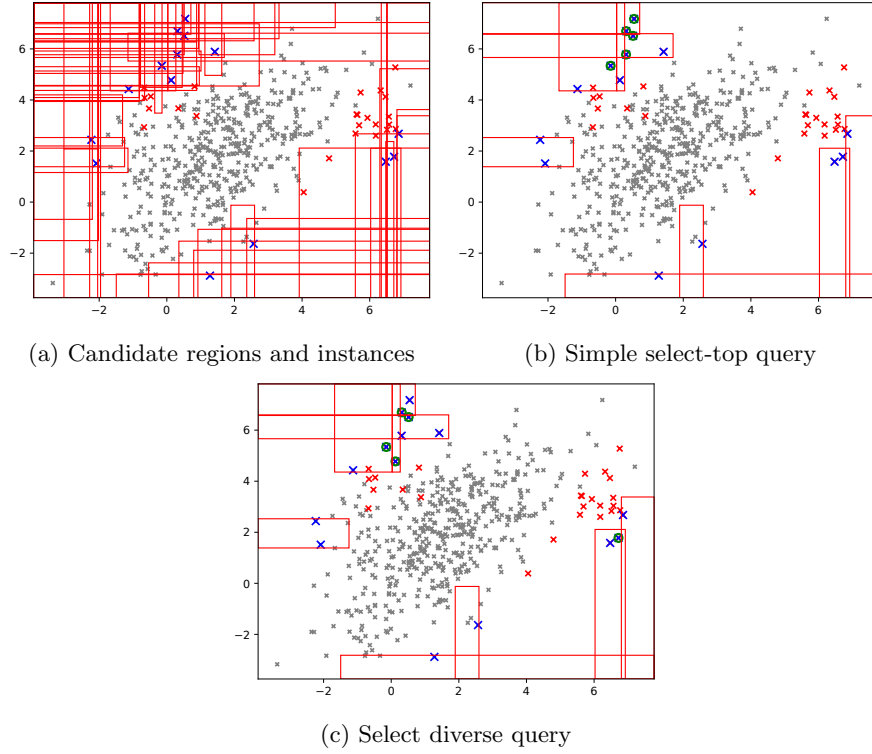
(c) Select diverse query

Figure 3: Diverse querying strategy using sub-space descriptions. These figures show the 15 most anomalous instances (ranked by score) in blue. We assume that the true labels are not known for these instances. We find the top 5 most anomalous regions for each of the 15 instances and the union of all these rectangles is shown in red in Figure 3a. Figure 3b and Figure 3c show the regions which compactly contain (describe) the top-ranked instances. The instances circled in green are the ones selected for querying. The instances (in green) in Figure 3b are merely the top-ranked 5 instances without taking into account any diversity. Figure 3c shows that when we select instances which have different descriptions (minimum region overlap), they are more diverse.

4