APPENDIX

**1. Additional Experimental Results**

This section contains the full results and visuals for experiments 1 thru 4. Some of the results are reproduced from the previous discussion, but additional results are also included for each experiment. For each experiment, the following are included:

* Curves for accuracy, recall, precision, and f1-measure
* Curves for raw true positive, true negative, false positive, and false negative counts
* Variance for accuracy, recall, precision, and f1-measures.

**a. Experiment One Full Results**

|  |  |
| --- | --- |
| XX.1 | XX.2 |
| XX.3 | XX.4 |
| XX.5 | Figure-set XX. Dataset D1, experiment 1 result plots for accuracy (1), f1-measure (2), recall (3), and precision (4). The x/y axes of all curves were oriented for best visualization, so note their orientation carefully. ROC curve (5) is shown at left. |

Performance metrics for accuracy, recall, precision, and f1-measure are shown above. Additionally, true positive, true negative, false positive, and false negative counts are shown below.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure XX: Clockwise from top left, true positive, true negative, false positive, and false negative count

The raw true positive, true negative, false positive, and false negative counts are shown above in figure XX. These curves elucidate the previous performance metrics, since the performance metrics are directly based on these values. As shown, the small values of had the strongest influence in terms of either finding or failing to find anomalous traces. A slight bump along the theta-trace axis of the false-negative count curve, in the span of and demonstrates the algorithm suffering slightly due to greater trace diversity at these values.

The variances of the performance metrics are also shown below for experimentalendpoint values of 0.5 and 0.9.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure XX: From top-left, accuracy variance, f1-measure variance, precision variance, and recall variance, plotted over the range of under test, for .

The variances for different performance metrics are shown above in figure XX. As shown, accuracy variance is quite low for reasonable values in the approximate range [0.05,0.15], when averaged over the 60 models tested for each value. The importance of including these results is given by the recall and precision charts (and to the f1-measure to a lesser extent, since the f1-measure is calculated using recall and precision as input). The variance of recall and precision is quite significant, above 0.20, in the range The cause of this variance is the nature of the data: if the algorithm fails to find a certain *type* of anomaly, that anomaly is usually shared by several or more traces. Thus, by failing to find any specific instance of an anomaly, a whole set of graphical traces is missed, resulting in a large punishment and significant variance. This effect is expected given the way that the data was generated, but it is worth noting given how it helps to characterize the performance of the algorithm in the context of the structural data with which it was evaluated.

**b. Experiment Two Full Results**

|  |  |
| --- | --- |
| XX.1 | XX.2 |
| XX.3 | XX.4 |
| XX.5 | Figure XX. Dataset D1, experiment 2 results for (from top-left) accuracy (1), f1-measure (2), recall (3), and precision (4). The x/y axes of all curves were oriented for best visualization, so note their orientation carefully. ROC curve is shown at left (5). |

Performance metrics for accuracy, recall, precision, and f1-measure are shown above. Additionally, true positive, true negative, false positive, and false negative counts are shown below.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure XX: Clockwise from top left, true positive, true negative, false positive, and false negative counts plotted over a range of values.

The raw true positive, true negative, false positive, and false negative counts are shown above in figure XX. These curves explain the previous performance metrics, since the performance metrics are directly based on these counts. As shown, the small values of had the strongest influence in terms of either finding or failing to find anomalous traces. The true positive/negative counts show steady performance even for large values of , indicating the method continued to find anomalies even when they became very dense in the data. Perhaps the most important curve here is the false-negative count (bottom right, above), showing a linear increase along the axis. The linearity along this axis implies the method’s performance decayed in terms of a constant, however it also shows the potential for the method to provide unsatisfactory performance on certain datasets. Values of are a bit unreasonable since they imply data for which the distribution of traces and anomalies is very nearly the same, but they provide context for expected algorithmic performance in such extreme cases. In cases where a false-negative (type II error) is very high cost, and where the data exhibited an unusually high anomaly probability, one would need to use a high value to defend against false-negatives.

**c. Experiment 3 Full Results**

The full results for experiment 3 are nearly identical to those of experiment 1, since they tested the same model parameter, , but for models with the modification of reusing existing activities for anomalies. The results are presented below with minimal discussion, since the comments in (a) above apply in the same manner.

|  |  |
| --- | --- |
| XX.1 | XX.2 |
| XX.3 | XX.4 |
| XX.5 | Figure-set XX.x. Dataset D2, experiment 1 results for (from top-left) accuracy (1), recall (2), precision (3), and f1-measure. The x/y axes of all curves were oriented for best visualization, so note their orientation carefully. ROC curve is shown at left (5). |

Experiment 3 performance curves are shown above and reflect the same results as in Appendix 1.a. above.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure XX: Clockwise from true positive, true negative, false positive, and false negative counts for experiment 3.

True positive, true negative, false positive, and false negative curves are shown above, averaged over the 60 logs tested each value, and show only slightly altered performance compared with experiment 1 results listed in Appendix 1.a.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure XX: Accuracy, f1-measure, recall, and precision variance for experiment 3.

Test variance for accuracy, f1-measure, recall, and precision are shown above, and again differ little from experiment 1 results listed in Appendix 1.a.

**d. Experiment Four Full Results**

The results for experiment 4 differ very little from those of experiment 1 listed in section 1.b., and don’t require additional comment. See section 1.c. for lengthier comment.

|  |  |
| --- | --- |
| XX.1 | XX.2 |
| XX.3 | XX.4 |
| XX.5 | Figure-set XX.x. Performance results for D4, experiment 2 over a range of theta-anomaly values for (from top-left) accuracy (1), f1-measure (2), recall (3), and precision (4). Again, note that the horizontal x/y axes have been oriented to improve visualization. ROC curve is shown at left (5). |

The raw true positive, true negative, false positive, and false negative curves differ very little from those found in section 1.c., as shown below.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure XX: Clockwise from top left, true positive, true negative, false positive, and false negative curves for experiment 4.

**e. Experiment Five Full Results**

Performance curves for accuracy, recall, precision, and f1-measure are displayed below for experiment five over a range of model anomaly quantities and . The results were averaged over the 30 models generated and tested for each number of anomaly structures in .

|  |  |
| --- | --- |
|  |  |
|  |  |
|  | Figure-set XX.x. Performance results for D3, experiment 3 over a range of k-anomaly values for (from top-left) accuracy (1), f1-measure (2), recall (3), and precision (4). These visuals were oriented to convey the curvature of each metric, so attention must be paid to the orientation of the x/y axes. The *k* parameter is given by the axes labeled “Anomalous Structures.” ROC curve is shown at left (5). |

Below are curves derived by averaging the true positive, true negative, false positive, and false negative counts over the 30 models tested for each number of anomalous structures.

|  |  |
| --- | --- |
|  |  |
|  |  |

Table 1 Clockwise from top left, true positive, true negative, false positive, and false negative counts per number of anomalous structures and .

The true positive, true negative, false positive, and false negative rates shown above are particularly important for this experiment, since the previous performance metrics are derived directly from these values. The dynamics shown in these curves give direct insight about the characteristic advantages and weaknesses of the method in experiment five, particularly for the reasonable parameter span defined by and between 0 and perhaps 8 anomalous structures. Values outside this range were exhaustive for the sake of evaluation but are clearly extreme. The false-negative curve again shows a characteristic increase along the anomalous-structures axis in the direction of more numerous anomalous structures, but the curve above is shown in powers of two. In linear space, the curves are linear, as shown below for a slice across the anomalous structure axis at a reasonable value of 0.06.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure XX: Performance curves in linear space for a selected value of 0.06. Clockwise from top-left, accuracy, f1-measure, precision, and recall.The plots are merely a slice of the performance curves from prior for a fixed value.

As shown above, the performance of the method maximizes for lower numbers of anomalous structures, less than 8 or so. The results demonstrate the level of effectiveness of this method for a modest density of anomalous structures; 16 or more anomalous structures are extreme values for the sake of exhaustive evaluation of the method.