ISBN: 978-3-95450-209-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2019-THCPL07

EXPERIENCE USING NuPIC TO DETECT ANOMALIES IN CONTROLS DATA*

T. D'Ottavio†, P. S. Dyer, J. Piacentino, Jr., M. R. Tomko
Brookhaven National Laboratory, Upton, USA

Abstract

NuPIC (Numenta Platform for Intelligent Computing) is an open-source computing platform that attempts to mimic neurological pathways in the human brain. We have used the Python implementation to explore the utility of using this system to detect anomalies in both stored and real-time data coming from the controls system for the RHIC Collider at Brookhaven National Laboratory. This paper explores various aspects of that work including the types of data most suited to anomaly detection, the likelihood of developing false positive and negative anomaly results, and experiences with training the system. We also report on the use of this software for monitoring various parts of the controls system in real-time.

INTRODUCTION

An *anomaly* can be considered a point in time when the behavior of a system is significantly different from previous, normal behavior. For the purpose of this paper, we consider only anomalies in numeric time-series data. Most controls data fall into this category.

Why work on anomaly detection? For most subsystems within controls, we can usually identify data that experts would consider normal - at least for some period of time. Normal behavior might be defined as data within a certain range, or with a particular pattern, or changing within a prescribed rate, or some combination of all of these. Experts can usually pick out deviations from normal. But the volume of data accumulated in modern control systems does not allow for such review. We need computers and algorithms to detect these anomalies.

NUPIC DESCRIPTION AND SETUP

NuPIC stands for Numenta Platform for Intelligent Computing. It is a software system designed to mimic the neural algorithms used within the neocortex of the human brain [1]. Numenta's goal is to reverse-engineer the neocortex and apply that knowledge to the creation of machine intelligence [2]. The NuPIC software, which originated in 2013, is open source on GitHub [3]. Originally written in C++, the most popular implementation is in Python. There is also a third-party port to Java [4]. The Python implementation provides three APIs. At the highest level (easiest to implement, least customizable) is the Online Prediction Framework API, or OPF. At the lowest level (easiest to customize, most difficult to implement) is the Algorithm API. A middle ground is provided by the Network API. The work in this paper was done with the Python version of the NuPIC OPF 1.05. [5].

The NuPIC software uses a software algorithm called Hierarchical Temporal Memory (HTM), which uses stored data sequences to make predictions about future data. These predictions are then compared to the actual data delivered to calculate prediction errors which, along with prediction errors from surrounding data, are transformed into anomaly likelihood values. It is these likelihood values that ultimately determine if the software found an anomaly [6]. (We actually enhanced these values using an extension that will be described later.)

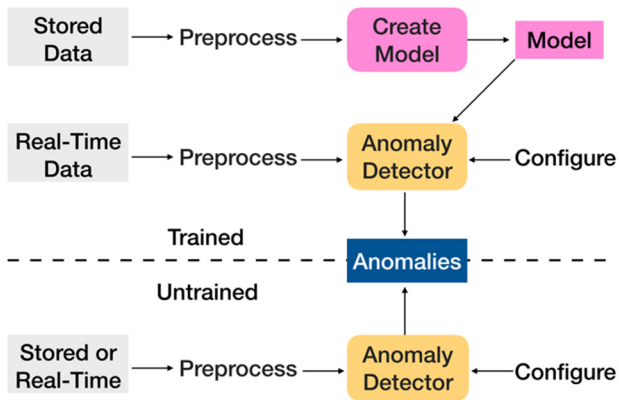

Figure 1: System diagram.

As seen in Fig. 1, the software system can be run in both a trained or untrained mode. When training, you create a NuPIC model using "normal" data. This model is then loaded into the NuPIC software prior to looking for anomalies. Once the model is loaded, you can instruct the software to continue learning from new data, or to turn learning off. In the untrained scenario, the NuPIC software is learning on new data as it arrives, though you can control which data is used for learning. Stored data, when it was used, was retrieved from our logging/archiving systems. Real-time data came from devices connected to the control system reporting at a rate of 1 Hz or slower. Preprocessing, when required, involved averaging, sampling or filtering of the data.

The heart of the system is the Anomaly Detector, which holds the NuPIC software. This software has about 30 parameters that can be tweaked to adjust the sensitivity and performance for various data sets [7]. However, Numenta delivers a set of parameters tuned specifically for anomaly detection as part of their Online Prediction Framework. With one exception described later, we used this set for all of the results that follow.

* Work supported by Brookhaven Science Associates, LLC under Contract No. DE-SC0012704 with the U.S. Department of Energy.
† dottavio@bnl.gov

ISBN: 978-3-95450-209-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2019-THCPL07

RESULTS

Anomalies can be classified into different types. For the purpose of evaluating the NuPIC software, we found it useful to separate anomalies into the four types shown in Fig. 2. In this section, we describe how well the NuPIC software performed with each of these different anomaly types.

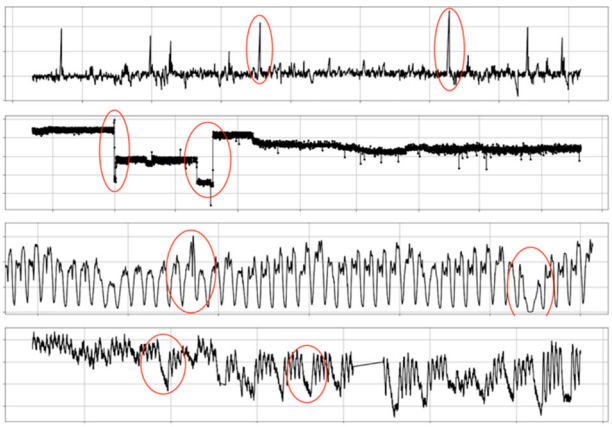

Figure 2: Types of anomalies. From top to bottom: Global Outlier, Level Change, Pattern Deviation, Pattern Change.

For each anomaly type, we present here only a representative example of the many data sets tested. In all of the cases shown here, the results were obtained by running the NuPIC software with stored data in an untrained mode, though most of the results would not have changed if using real-time data or running trained.

All of the results show a set of four stacked plots, with the following data (from bottom plot to top):

- The Raw Data (sometimes pre-processed) fed into the NuPIC software. In most cases, this is a single data set, but it is possible and sometimes desirable to also use the associated timestamp or another data set that could aid in the anomaly detection.
- The Anomaly Score, output from the NuPIC software. It is a measure of the difference between the actual value fed into the software compared with prediction value. A score of 0 indicates normal, predicted behaviour, while a score of 1 suggests anomalous behaviour.
- The Log of the Anomaly Likelihood. The Anomaly Likelihood, also an output from the NuPIC software, measures the probability that the Anomaly Score represents an anomaly given the historical distribution of Anomaly Scores. This helps to distinguish true anomalies from high anomaly scores caused by noise. This value varies from 0 to 1. Here we use the log of the anomaly likelihood as the meaningful values are very close to 1. In practice, the log value needs to be above 0.3 to be significant, corresponding to an anomaly likelihood above 0.999.
- The Consecutive Anomalies, a statistic tracking anomaly persistence. It is calculated as the number of consecutive Log Anomaly Likelihood values that have exceeded the 0.3 threshold at any point in time. In practice we have found that we can eliminate a number of false positives by counting consecutive values that pass this threshold. A value of 10 or greater is a very good sign that a real anomaly has been detected.

Global Outlier Anomalies

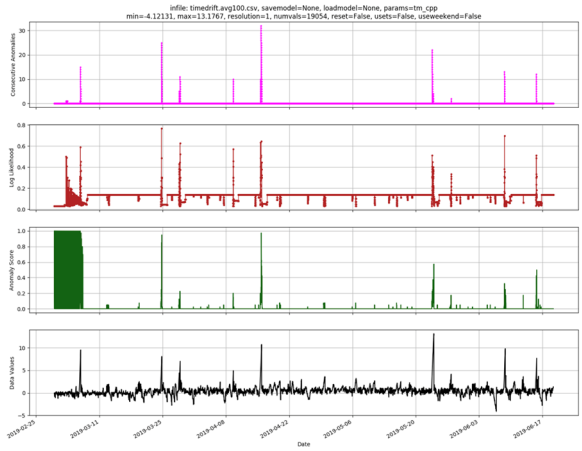

Figure 3: Global Outlier anomaly results.

This type of anomaly is very common in controls data. It occurs when a data stream is normally contained within a certain range and one or more values have fallen outside of that range. The NuPIC software works very well for this type of anomaly. A typical run of the NuPIC software for this type of data is shown in Fig. 3. This data set has about 19k points and was input to untrained NuPIC software. As can be seen in the Anomaly Score plot, untrained NuPIC software needs about 500 points before it outputs any meaningful results. Note that the "resolution" parameter of the NuPIC software was altered from its default value in order to obtain these results. See the Discussion section for more information about how to best set the resolution parameter.

Level Change Anomalies

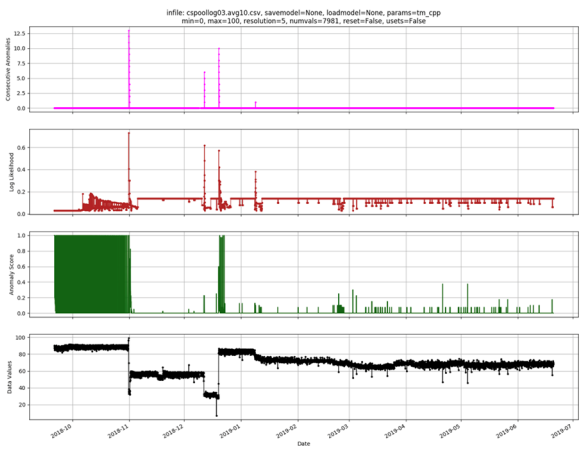

Figure 4: Level Change anomaly results.

Figure 4 shows data that goes through several level changes. That is, the data has a somewhat constant value for a significant period of time, then this level changes to some new value. The NuPIC software works well with this

ISBN: 978-3-95450-209-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2019-THCPL07

type of anomaly. These anomalies are not always a problem, but it is sometimes useful to be notified about level changes. The example shown here is idle time data from a computer. Note that this is a data set of about 8k points, but 10:1 averaging had been applied in a pre-processing stage. See the Discussion section for information about averaging and pre-processing data.

Pattern Deviation Anomalies

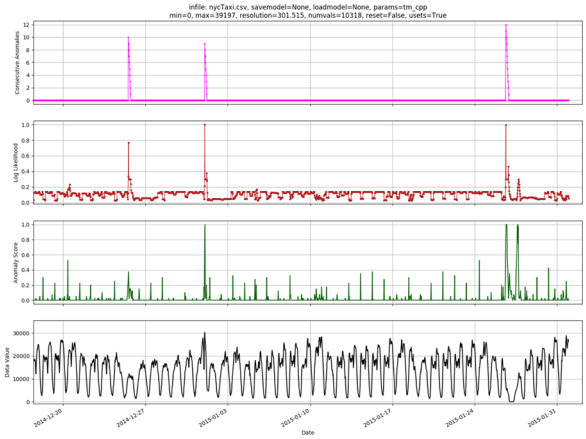

Figure 5: Pattern Deviation anomaly results 1.

A Pattern Deviation anomaly occurs when data that has some regular, repeating pattern shows a change, not in the pattern, but in the values of the data over that pattern. The NuPIC software yielded mixed results with this type of anomaly. Figure 5 shows the number of New York City taxi rides every 30 minutes for a 40-day window in the winter of 2014. Note that the NuPIC software had been trained with data for all of 2014 before getting to this point. It does an excellent job of picking out several anomalies, which occurred on Christmas, New Year's Eve, and during a snowstorm in January. In this case, the date and time were encoded into the software as well as the taxi data. This allows the software to pick out diurnal as well as weekday/weekend patterns.

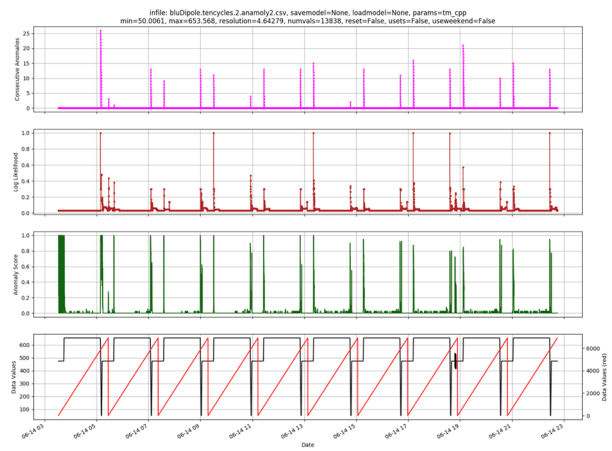

Figure 6: Pattern Deviation anomaly results 2.

An example where the NuPIC software did not do well with a Pattern Deviation anomaly is shown in Fig. 6. Here both a repeating magnet cycle (black) and the time from the start of the cycle (red) are encoded into the NuPIC software. Even after repeating the same cycle hundreds of times, the software indicates an anomaly anytime there is a large swing in the magnet current. Even when a real anomaly was artificially introduced (see Fig. 6), it did not consider it significant.

Pattern Change Anomalies

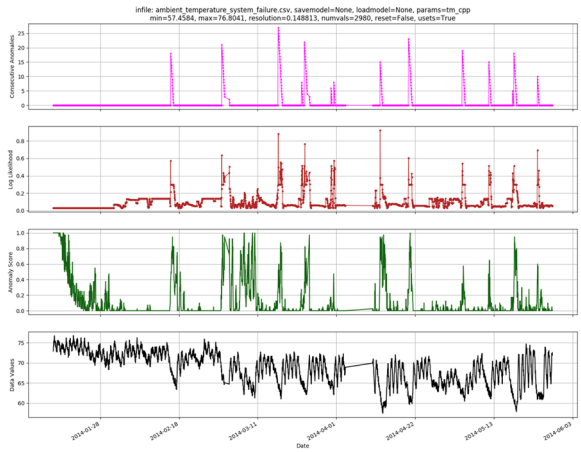

Figure 7: Pattern Change anomaly results 1.

A Pattern Change anomaly occurs when a change is detected in a pattern that exists when the data is considered normal. Figure 7 shows hourly temperature data taken over a time period of five months. A normal diurnal temperature pattern is interrupted occasionally whenever an air conditioner is turned on. The NuPIC software picks up this change each time, again showing that the software is particularly strong in picking up changes in diurnal patterns.

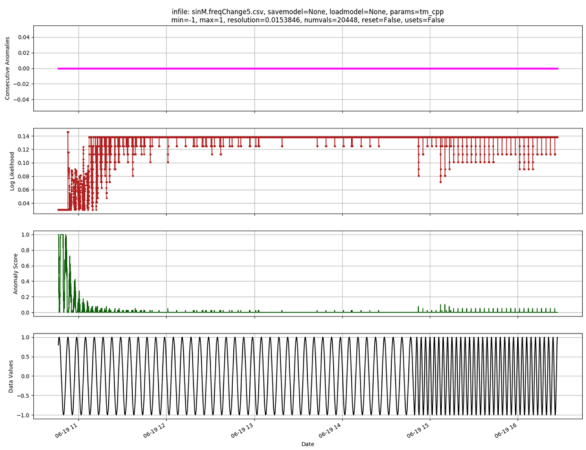

Figure 8: Pattern Change anomaly results 2.

On the other hand, the NuPIC software can fail dramatically on pattern changes that are very obvious as shown in Fig. 8. In this case a sin wave data set was generated with 360 points per cycle. After many iterations of the same cycle, the frequency was increased by a factor of two. In this case the NuPIC software barely noticed the change. For some reason, it did better or worse depending on how many points were encoded into each cycle, showing good

ISBN: 978-3-95450-209-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2019-THCPL07

detection of the change when the number of points per cycle was reduced to 60. Perhaps knowledge of the internals and adjustments to one or more of the configuration parameters would have solved this problem, but that was beyond the scope of this investigation.

DISCUSSION

The examples in the Results section are representative of the results using many other data sets that we have fed into the NuPIC software. It does very well with Global Outlier and Level Change anomalies but has a mixed record with the Pattern Deviation and Pattern Change anomalies. In this section, we'll explore some of the details on how to best setup and use this software.

Data Pre-processing

The NuPIC software works best with data sets that have about 10k points, with a working minimum of about 1k points. Using more points is fine but leads to longer processing times and will not necessarily lead to better results. In some cases, the results will be worse, especially if the data is very noisy. So, if large data sets are to be analyzed, in most cases its best to average the data before running it into NuPIC.

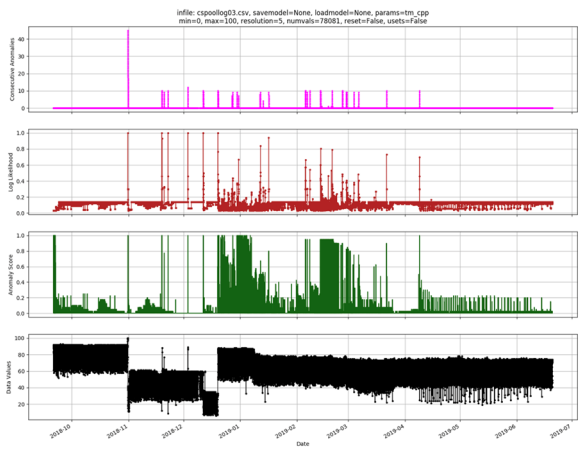

Figure 9: Level Change results without averaging.

To see this effect, compare the results shown in Fig. 9 with the results shown in Fig. 4. The only difference is that 80k points shown in Fig. 9 were reduced to the 8k points in Fig. 4 using 10:1 averaging.

In other cases, filtering of the data is necessary to get meaningful results. This is true when data streams contain real data mixed with noise. In our case, this often happens when ion beams are turned on and off. Filtering out the "off" data has a huge benefit in detecting anomalies in the "on" or normal data.

Software Configuration

As mentioned in the Introduction, the NuPIC software has about 30 configuration parameters that can be used to fine tune the system for a particular data set. However, the experts at Numenta have delivered the software with a set that they say is tuned for anomaly detection for most types of data. The results shown here use this pre-configured setup. It should be noted that this configuration is biased toward data wherein the time of day contains information about the anomalous nature of the data, which is evidenced in part by our results showing good performance on data with diurnal patterns.

The one adjustable parameter that we did adjust is the "resolution". This parameter determines the number of data buckets that the software sets up for the data. By default, the number of buckets is 130 and the resolution is determined by the formula (max – min) / 130 where min and max are the expected (or measured) min and max for the data set. Often, it is useful to increase the resolution to reduce the sensitivity of the software to small changes. This is an easily adjustable parameter in the NuPIC software and can greatly affect the number of false positives or negatives. Compare the results in Fig. 10 (using the default resolution = 0.13) with the results in Fig. 3, where the resolution was increased to 1.0.

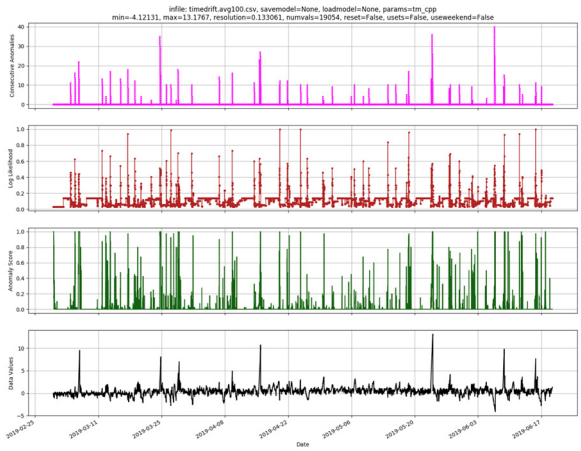

Figure 10: Global Outlier results without a change to the resolution parameter.

Learning, Training and Models

In its normal running mode, the NuPIC software is learning with each new point added. If started in its untrained mode (see Fig. 1), the software needs about 500 data points before it can make useful predictions. During this time the software is storing the temporal data sequences that it will ultimately use to make predictions.

When the software is learning, it is also being trained. The training at any point in time can be captured and stored in what NuPIC calls a model file. When the software is started again, it can then read this model file and recapture the contained training. At that point, learning can continue, or it can be turned off. This is under software control. You might want to turn learning off if you think that you have captured a good set of "normal" data and if you think that the normal behavior will remain stable. In this way, you can guarantee an outcome based on the training. The disadvantage is that creating, storing, and loading a model is time consuming and resource intensive and does not necessarily produce better results. Also, in many cases, you may want the software to learn about new normal modes as they appear. This is the case with, for example, Level Change anomalies. In practice, we have found that only a

ISBN: 978-3-95450-209-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2019-THCPL07

small percentage of the data sets benefit from model training.

As mentioned above, it is possible to turn learning on and off dynamically. So, you could, for example, turn learning off when you know the data is not normal, then turn learning back on when it is. This is a useful strategy to use in lieu of pre-filtering the data.

Getting the Most Out of NuPIC

After running hundreds of data sets through the NuPIC software, we recommend the following procedure for getting the most out of this software.

- Preprocess the data as needed. Use averaging to reduce the number of points or to reduce noise. Filter out undesired data. Try to gather around 10k points. More is OK but consider 1k to be the minimum.
- Adjust the sensitivity using the resolution parameter. Larger values make the software less sensitive. Use 10% of the noise envelope as a good starting point.
- If necessary, train and save a model. Use a model if you have a good normal data set and you consider it to be very stable.
- Run the anomaly detector, loading the model if saved.
- An anomaly is found if you generate 10 consecutive Log Anomaly Likelihood values greater than 0.3.

Running NuPIC With Real-Time Data

Using the NuPIC software with real-time data is virtually identical to using it with stored data. In either case, for every value you feed into the system, you get out one predicted value upon which the anomaly score is based. For real-time data, consider running the software in its trained mode. Otherwise, you will have to wait for the software to collect about 500 points before giving useful results.

Real-time monitoring on a modern Linux computer requires about 2.5% CPU usage to monitor data returning at once/second. If using stored data, running through 10k points takes about 1 minute.

CONCLUSIONS

The NuPIC software can be a useful tool for detecting anomalies with some data sets. With minimal configuration, it does well with anomalies that are of the Global Outlier or Level Change type. Results were mixed for Pattern Deviation and Pattern Change anomalies. For the latter types, you should analyze some stored data first before expecting good results. This is probably necessary in general to at least determine the level of pre-processing and the resolution setting necessary to get good results.

How does the NuPIC software compare with other Anomaly Detection systems? We can't say and spent no time comparing this system with others. However, Numenta has published extensive results on this topic [6]. Here is how we would summarize the Pros and Cons of the NuPIC software as we have used it for this publication.

NuPIC Pros

- Can be set up to run with a fixed set of training data, or as an online learner, or in combination
- General purpose - can be used for a wide range of data
- Can be used with both stored and real-time data
- Works well for Global Outlier and Level Change anomalies
- Works well when there are diurnal patterns in the data

NuPIC Cons

- May require some preprocessing and/or configuration for different data sets
- Mixed results with Pattern Deviation and Pattern Change anomalies
- Can be slow for batch processing of large data sets
- Can detect anomalies for only one data set at a time
- Custom algorithms for a particular data set may work just as well or better

REFERENCES

[1] Numenta, https://numenta.org

[2] Numenta, https://numenta.com

[3] Numenta Platform for Intelligent Computing, https://github.com/numenta/nupic

[4] htm.java, https://github.com/numenta/htm.java

[5] OPF Guide, http://nupic.docs.numenta.org/stable/guides/opf.html

[6] S. Ahmad, A. Lavin, S. Purdy, Z. Agha, "Unsupervised Real-Time Anomaly Detection for Streaming Data", *Neurocomputing*, vol. 262, Nov. 2017, pp. 134-147. doi:10.1016/j.neucom.2017.04.070

[7] Example Model Params, http://nupic.docs.numenta.org/1.0.0/quick-start/example-model-params.html