

Satellite-As-a-Sensor Neural Network Abnormality Classification Optimization

Michelle Hammond, Research & Development Lead
Air Force Center for Research Support (CERES)
730 Irwin Avenue, MLS CS10/STEC, Schriever AFB, CO 80912-7200; (719) 721-0473
michelle.hammond.ctr@afspc.af.mil

2d Lt Ryan Jobman, Satellite Operations Engineer
Air Force Center for Research Support (CERES)
730 Irwin Avenue, MLS CS10/STEC, Schriever AFB, CO 80912-7200; (719) 567-6233
ryan.jobman@afspc.af.mil

ABSTRACT – Neural networks and classification networks are used in commercial and government industries for data mining and pattern trend analysis. The commercial banking industry use neural networks to detect out of pattern spending habits of customers for identity theft purposes. An example of government use is the monitoring of satellite state-of-health measurements for pattern changes indicating possible sensor abnormality or onboard hardware failure in a real time environment.

Key words: neural network, abnormalities, clustering, satellite monitoring, data fusion

INTRODUCTION

Satellite-As-a-Sensor (SAS) neural network technology is currently used by the Center for Research Support (CERES) in Colorado Springs under the United States Air Force. CERES uses neural networks to monitor state-of-health telemetry to detect pattern changes. Using the neural network technology automates the process of finding pattern changes to insure maintenance actions can be performed to save the satellite from failures. This technology can be applied for future small satellite sensor telemetry monitoring and abnormality detection. The SAS research project explains the use of neural network processing and performs neural network classification with optimization.

A neural network is an artificial intelligence technology that learns from past data to recognize the pattern of a data source. For satellite telemetry purposes, historical telemetry from a satellite is used to train the neural network on ‘normal’ values. Normal values are chosen to be statistically similar to the average mean, standard deviation, and slope of a measurand over a selected period of time. Telemetry is then monitored real-time for pattern changes. The pattern changes are referred to as alerts or abnormalities and clustered into similar groups based on the alerting measurand values. The abnormality clustering, referred to as classification, is a data mining and data fusion technique for fusing neural

network output into status for the entire satellite. Classifications are built using an angular distance algorithm from the neural network error scores. The angular distance algorithm calculates the arccosine of the neural network output error score angular distance^[1]. Table 1 describes the angular distance calculation used when classifying neural network output error scores. This angular distance calculation of the classification creates a “classification cone” in n-space, n being the number of measurands from a particular satellite vehicle.

Table 1 Angular Distance Metric

Angular Distance Metric	Description
$d = \left \cos^{-1} \left[\frac{z \cdot x}{\ z\ \ x\ } \right] \right \leq \pi$	Z represents the current cluster center while X represents the class estimate.

The angular distance metric algorithm (d) is comparing the angular classification cone center to neighboring cone centers using a width parameter sigma. The sigma of the classification cone is used to determine at what value the outlier abnormalities should exceed to initiate a new classification. The trade-off of creating new classification cones is being precise enough to represent like abnormalities while limiting the amount of new classification cones for similar abnormalities. A graphical description of classification cones is in Figure 1 Classification cones

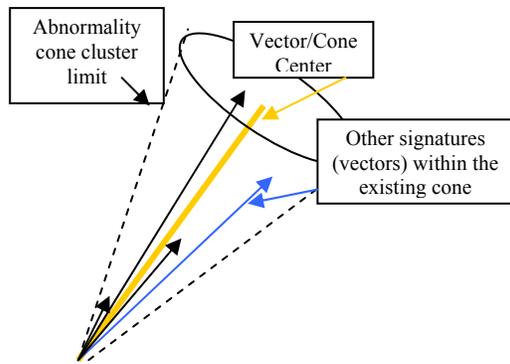


Figure 1 Classification Cones

As the neural networks and classification networks have been evaluated, some instances occurred where the classification clusters were so close in distance that the current sigma setting of .75 radians produces misclassifications. This misclassification occurs because the abnormality falls within an existing classification cones when it should create a new cone because the abnormality represents a different situation. Creating two classification cones for the previously mentioned situation will provide a more optimal neural network classification by reducing the amount of misclassifications. This research is focusing on the classification network parameter sigma for optimizing the classification network. The selected events to evaluate in this study have similar alerting telemetry values within the same subsystem, although users would like for these events to be represented in two separate classification cones. Because of these alerting value similarities, the classifications are falling within the same existing cone. The hypothesis of this work is by reducing the sigma setting, the probability of classifying abnormalities accurately will be above 50% while creating the least amount of new classification cones.

The least amount of new classification cones refers to the instantiation of classification cones when the width setting, sigma, parameter is decreased. When reducing the sigma, more abnormality classification cones are created while increasing the sigma setting reduces the number of abnormality classification cones^[1]. If the abnormality does not fall within an existing classification cone, a new cone is created. This type of an effect when changing sigma is important to the user of the system, the satellite operator. The more classification cones created, the more realistic the clusters represent for a particular situation, however if too many classification cones are created, the amount of time to research and determine what all the clusters

represent can be counterproductive. Figure 2 Classification width versus new cone creation describes the dependency of classification cone width to new cone instantiations.

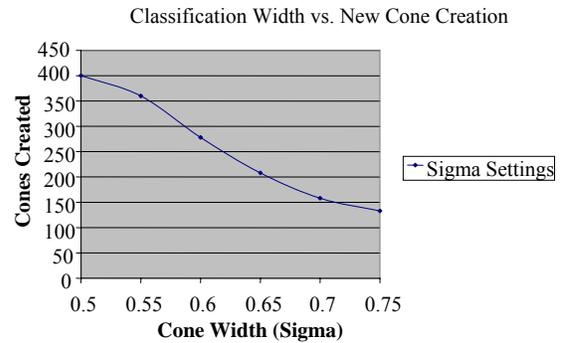


Figure 2 Classification Width vs. New Cone Creation

Classification cones are built upon a vector representing the absolute error score of all the neural network measurands, at one point in time, using an angular distance metric and angular threshold for clustering. The errors for each measurand are clustered into n-space classification cones (n being the number of satellite telemetry measurements evaluated). The sigma setting is the threshold in radians evaluating if the angular distance metric falls within an existing classification cluster. The hypothesis will be accepted if for a selected set of misclassified abnormalities, the probability of classifying these abnormalities accurately in separate abnormality classification cones is above 50%. The second criteria for accepting the hypothesis is that the lowest number of newly created abnormality classification cones exists while the probability of accurate abnormality classification cones is over 50%.

NEURAL NETWORK & CLASSIFICATION PROCESS

The neural networks being used as a base for classification networks has one hidden layer^[1]. The training technique includes using a genetic algorithm to select historical satellite state-of-health telemetry files. Based on the selected files, the neural networks learn the normal behavior of correlations and measurands. The correlation includes estimating the mean, standard deviation, linear regression slope, and slope change frequency at a particular time for the input training set. Those estimates along with the state-of-health telemetry measurands are used to train the neural networks on normal behavior. The training technique then creates a hidden layer and propagates the data through the neural networks again. The trained neural networks are put

into a real-time environment and used for pattern detection. In the real-time environment, the neural network produces error values representing the difference between the current value and the learned value from the training set. Those errors values are then input to the angular distance algorithm to create the classification. Classification processing flow is represented below in Figure 3 Classification Optimization Process. The abnormality detection is structured into the angular distance classification cone by using the weights for the output error scores of each measurand, static parameters, and variables to create the location in n-space of the classification cone.

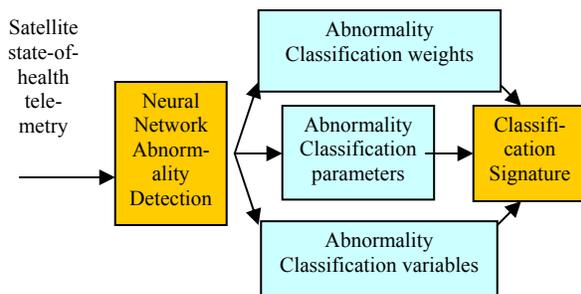


Figure 3 Classification Optimization Process

Neural network abnormality detection includes producing output error scores of all measurands for the satellite. The output error scores are then applied to creating an angular distance cluster, or classification cone using weights, parameters, and variables. These three classification components determine where in n-space the abnormality should be located. The resulting abnormality classification cone is compared in radian distance to existing classification cones to determine if the abnormality should be included in existing classification cones or create new classification cones.

APPROACH

The optimization experiment using sigma gathered information on pattern change detections, angular distance between abnormality classification cones, and historical events to use as benchmarks. Setup for data gathering included using two software applications, Abnet and TP3. These software applications are Government Off The Shelf (GOTS) tools built by Logical Designs Consulting, Inc with additional consulting assistance from Data Fusion & Neural Networks (DF&NN). Abnet is the software tool used for abnormality detection and classification for online and off-line purposes^[1]. Abnet applies the learned signatures estimates into clusters or new abnormality estimates while TP3 is an off-line tool for clustering of

signatures and vectors ^[1]. Historical telemetry will be used for evaluation purposes.

Proving the hypothesis required analysis of the effects on abnormality classification clustering when the sigma value is changed. Based on user knowledge and referencing existing research in this area, the width parameter of classification cones can act to optimize results from an automated decision making system ^[2-6]. The current neural network was used to evaluate thirty-nine groups of historical telemetry representing abnormalities at the sigma values of .75, .7, .65, .6, .55, and .5 radians. Of the 39 data sets, 1,120 seconds of abnormalities were detected. The abnormalities were evaluated to select the set of abnormalities misclassified at the highest rate. The selected misclassifications used as a benchmark were the abnormalities of the ground antenna starting communication with the satellite referred to as event AA and ending communication with the satellite referred to as event AB. These events are detected as pattern changes because the user wants to be aware of all communication with the vehicle. The alerting values are similar; however, the users want to see event AA and event AB fall into two separate classification cones. The selected abnormality classification cones were evaluated at each sigma setting to quantify which sigma reduced misclassifications of AA and AB while creating the fewest number of newly created abnormality classification cones across the remainder of the classification network.

The first method of evaluation included the distance between event AA and event AB classification cones. The closer in distance the existing classification cones are to each other, the higher chance of overlapping cones and misclassifications. If the distance between the classification cones can be increased, a more accurate classification cone is created increasing the confidence of the situation. The research project used the probability of accurate classifications versus misclassifications for the selected set of abnormalities to be the proof criteria. A comparison of the newly created cones versus the accurate classifications provided the needed information to select an optimal sigma setting for the classification network.

RESULTS

The 1,120 seconds of abnormalities in the 39 groups of telemetry sets fell into forty-five existing classification cones when sigma is at the original .75 radians. When decreasing the sigma of the clustering network to .65 radians, nine seconds of abnormalities fell within newly created classification cones. As the sigma is reduced, the new abnormality classification cones increase.

Twenty-one of the original 39 input blocks included AA and AB abnormalities. Sigma setting of .75 and .7 produced the highest number of misclassifications between abnormality AA and AB for the 21 clusters with the abnormalities.

Abnormality events AA and AB are similar abnormal conditions; however, it is integral to the data analysis that these two event types be represented separately. Evaluation of the similarity between these two abnormality classification cones was used to validate why a higher number of misclassifications are occurring. Evaluation of classification cone distance away from neighboring classification cones within .2 radians provided some insight into the reason of misclassifications. When the sigma value is at .75 and .7 radians, the distance of neighboring AA and AB existing cones was within .2 radians. Since not only the highest output error scores are considered, all the output error scores can shift the cluster into a neighboring classification cone that represents a misclassified event. The distance between event AA and event AB increased to between .3 and .2 radians starting with a .65 radian sigma. Figure 4 shows the existing cone similarities within .2 radians of neighboring classification cones. The results include evaluating the distance in radians between the abnormality classification cone using sigma at .75, .7, .65, .6, .55, and .5 radians.

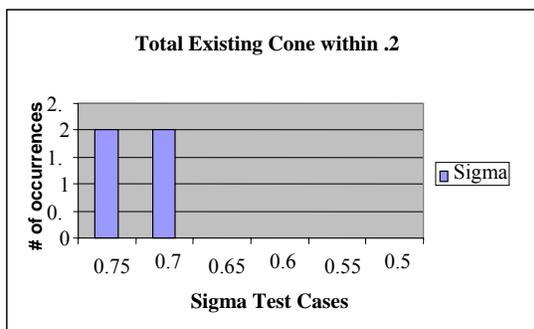


Figure 4 Distance between Existing Cones

At sigma .65 radians the abnormality cluster centers of event AA and AB are further than .2 angular distances away from each other. This separation of distance between similar cone clusters increases the probability that event AA and AB would fall into separate classification clusters. This would increase the probability of accurate classifications for the selected set of abnormalities and optimize the classification networks.

The sigma settings were then evaluated on historical data to determine what the percentage of correct classification was and the number of new cones created.

Validation for proving that a sigma lower than .75 would reduce misclassifications for the classification network included using the original 21 clusters of abnormal events and evaluating the number of event AA and event AB falling into two separate classification clusters. From the previous analysis we knew at .65 sigma and below the distance between the AA and AB clusters increased above .2 radians. Therefore, .65 and .6 were the candidates that would match the hypothesis of correct classification over 50% of the time between event AA and event AB while keeping to a minimum newly created cones. Table 2 Sigma Setting Validation shows the results of tests performed on the candidate sigma settings and their corresponding percentage of correct classifications.

Table 2 Sigma Setting Validation

Sigma Setting	% of Correct Classification
.75	9.5%
.7	9.5%
.65	67%
.6	67%

With the sigma value at .75 and .7 radian, the probability of event AA and event AB falling within two classification clusters was 9.5% while sigma at .65 radians produced a 67% probability. Sigma at .65 clearly provided a more optimal classification network above our 50% threshold. The second criteria of keeping the newly created cones to a minimum supported .65 as the optimal setting. Sigma values below .65 also had a probability above 50% for separating clusters, but the number of newly created clusters was much more than the .65 setting making them counterproductive to the satellite operator. The new cone creations per sigma settings of .75, .7, .65, .6, .55, and .5 are exhibited in Figure 5.

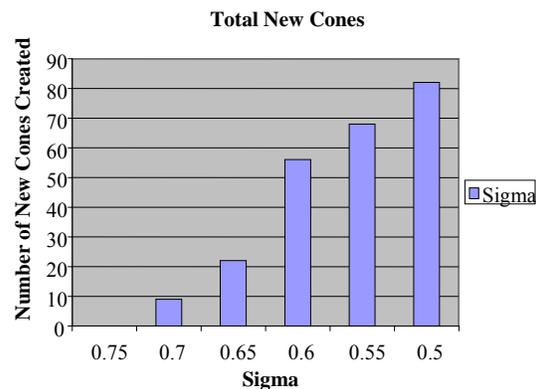


Figure 5 New Cones Created with Varying Sigma

CONCLUSION

The research project of SAS neural network classification optimization proved that by reducing the classification sigma parameter from .75 to .65 a more optimal classification network was created. An optimal classification network reduces overlapping classification cones resulting in a reduction of misclassifications. The criteria of properly classifying neural network detections of event AA and event AB over 50% of the time while creating the least amount of new cones for the remaining classification network was met by .65 sigma. This research was performed on a GEO orbiting satellite, although could be applied to small satellites within LEO or MEO orbits. For small satellites in the future, this neural network pattern recognition and classification technology can be applied to ensure the health of the vehicle during real time, possibly saving the vehicle and the mission.

FUTURE RESEARCH

Future research furthering the optimization of classification networks could include using dynamic clustering techniques. The added ability of dynamic clustering is to define the attributes and variables for each abnormality cluster based on the abnormality subsystem or most offending measurands. Another approach for future research could be to evaluate the differences of the angular distance clustering algorithm against the industry standard k-means algorithm. The k-means algorithm is a centroid clustering technique that is said to be faster, more scalable than comparable clustering algorithm, and easier for the users to understand^[6].

For small satellite purposes, future research of applying this technology could be performed at CERES taking into consideration the small satellite vehicle challenges. Challenges include determining the appropriate size and type of training data to select for training neural networks. The training data is important to the success of the neural network detection capability. Other research may include understanding how to provide the most real time data as possible for a small satellite based on the orbit and antenna locations. This challenge would include using the AFSCN or dedicated communication lines for the telemetry to be provided to CERES or another military agency for the neural networks to process real time.

REFERENCES

1. Bowman, D.C., *Abnormality Detection, Classification, and Evaluation Tools Description*. 2003.
2. Bernier, H.G.M., *The effect of some internal neural network parameters on SAR texture classification performance*. IEEE, 2004: p. 3845 - 3848.
3. Brookshear, G.J., *Computer Science an overview*. 8 ed. 2005: Addison-Wesley.
4. Manikopoulos, Z.Z.a.C., *Investigation of Neural Network Classification of Computer Network Attacks*. Information & Technology: Research and Education, 2003. Proceedings. ITRE2003, 2003: p. 590-594.
5. Tsui, C.-k.C.H.-t., *Agent swarm classification network ASCN*. IEEE International Conference on Systems, Man, and Cybernetics, 2004. **6**: p. 5604-5608.
6. Benjamin J. Anderson, D.R.M., Ann M. Ritz, Andrew Ault, Deborah Gross, Melanie Yuen, Markus Galli, *User-Friendly Clustering for Atmospheric Data Analysis*, Carleton College. p. 8.