Validating An Online Adaptive System Using Support Vector Machines

Yan Liu yanliu@csee.wvu.edu
Srikanth Gururajan srikanth@cemrweb.cemr.wvu.edu
Bojan Cukic cukic@csee.wvu.edu
West Virginia University,
Morgantown, WV 26505, U.S.A.

Abstract **- In this paper, we propose a novelty detection method based on Support Vector Machines as a candidate approach for validating online adaptive systems. As a one-class classifier, the support vector data description is able to form a decision boundary around the learned data domain with very little or no knowledge of data points outside the boundary (outliers). Preliminary studies on an actual online adaptive system show that the support vector data description can be adopted as an effective tool for finding indications of the safe region for the learned domain, whereby we are able to separate faulty behavior from normal events.**

I. Introduction

The unusual plasticity of adapting promptly to unforseen situations of online adaptive systems has attracted increasing attention in many industrial applications. However, it also poses a very challenging problem in terms of software verification and validation since the system will react distinctively with respect to specific data, among which novel data might cause unstable states and potential failures. From safety and reliability assurance points of view, we propose an online validation framework consisting of: 1) Pre-filter, which should be able to block the penetration of unreliable or unreasonable training data from entering the online adaptive component and provide backward and forward recovery capabilities, 2) Run-time monitor, dealing with the real-time stability and convergence analysis of the learning device indicating the current confidence measure, and 3) Post-filter, which should disallow the propagation of unsafe adjustments into the controller, based on its knowledge of system safety requirements. Figure 1 illustrates such an architecture for an online adaptive system.

Our work on Step 2 in this validation framework has been focusing on the run-time performance checks using the Lyapunov theory of stability. Although not sufficient, Step 2 is a necessary part of this research ensuring learning (adaptation) stability and peak performance. In recent work [1], we proved that when an online neural network trains on a fixed feature manifold, the evolving state of the system due to the network's position adjustment is self-stabilizing in a globally asymptotically stable manner. This stability measure, together with the properly applied statistical analysis performed by the pre-adaptation filter, should give us a reliable indication of the validity and safety of the resulting system outputs (control adjustments). Steps 1 and 3 in our validation framework seek to define a novelty detection technique suitable for alerting to anomalous system adaptation and false prediction (control adjustment).

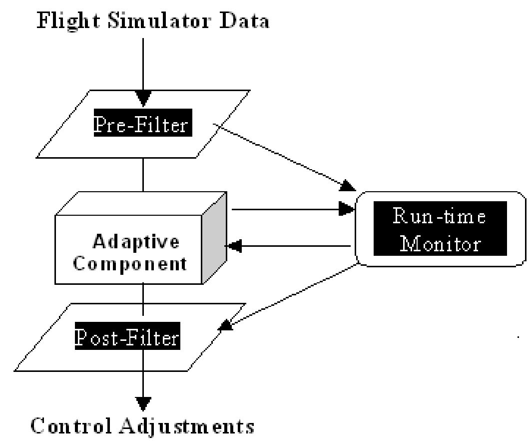

Fig. 1. A Validation Framework.

II. Novelty Detection Using SVMs

As a key step to realize our validation framework, we seek a novelty detection technique suitable for implementing both pre-filter and post-filter. Rather than trying to detect outliers by estimating the density function, which is impractical to be implemented for real time monitoring, we attempt to find a technique that is able to model the support of the data distribution. Having investigated the data mining literature, we propose a novel approach using support vector machines as our novelty detection tool.

Most real-world classification problems face the difficulty caused by non-separable nonlinear data space. Other than designing a complex computational model to approximate the separation, Vapnik [2] derived the support vector machine based on the idea of mapping the data into a higher dimensional space, called the feature space, where we can define a separating hyperplane. For the purpose of detecting novel data, a classification function needs to be defined. This function should have positive values in those regions of input space where the data predominately lies and negative values elsewhere. In another word, a "safe region" boundary needs to be formed. The method of support vector domain description proposed by Tax and Duin [3] is such an approach that finds a sphere with the minimal volume to contain all data. Given a data set S consisting of N examples $x_i, i = 1, ..N$, the objective is to minimize an error function containing the volume of this sphere. With the constraint that all data points must be within the sphere, which is defined by its radius R and its center a, the objective function can be translated into the following form:

$$L(R, a, \alpha_i) = R^2 - \sum_i \alpha_i \{R^2 - (x^2 - 2ax_i + a^2)\}$$

where $\alpha_i > 0$ is the Lagrange multiplier. L is to be minimized with respect to R and a and maximized with respect to α_i. By solving the partial derivatives of L, we also have:

$$\sum_i \alpha_i = 1; a = \sum_i \alpha_i x_i,$$

which gives the Lagrangian with respect to α_i:

$$L = \sum_i \alpha_i (x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j)$$

where $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$. According to the solution to maximize L, there will be a large portion of α_i's become zero, which leaves the rest of α_i's that are greater than zero. Their corresponding objects are those called support objects, which lie on the boundary that forms a sphere to contain the data. Hence, object z is accepted by the description when:

$$\|z - a\|^2 = (z - \sum_i \alpha_i x_i)(z - \sum_i \alpha_i x_i) \leq R^2.$$

By replacing some kernel functions $K(x, y)$ with the product of (x, y) in the above equations, the object z is declared as novel when:

$$K(z, z) - 2 \sum_i \alpha_i K(z, x_i) + \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \leq R^2.$$

The use of kernel functions offers a more flexible description of the boundary other than the sphere shaped description. Moreover, it avoids the exponentially increasing computational cost for high-dimensional datasets.

III. Experiments

We apply the SVM method to novelty detection in the context of an actual flight control system and conduct a series of experiments based on eight data segments obtained from its regular mode simulation. By applying the support vector data description method, the nominal domain is well learned and thus a certain data description is inferred. This description is then used for examining outliers collected from five different failure mode simulations. In these six datasets, each data point has multiple dimensions, which represent different maneuvering and flight condition parameters such as pitch rate, altitude, velocity, etc. Among those parameters, some of them are strongly correlated with respect to the flight control system design specification. In order to provide better visual demonstration, we monitor the boundary of pairwise parameters with high correlations. For each sequence of data we monitor, the Gaussian kernel function is applied.

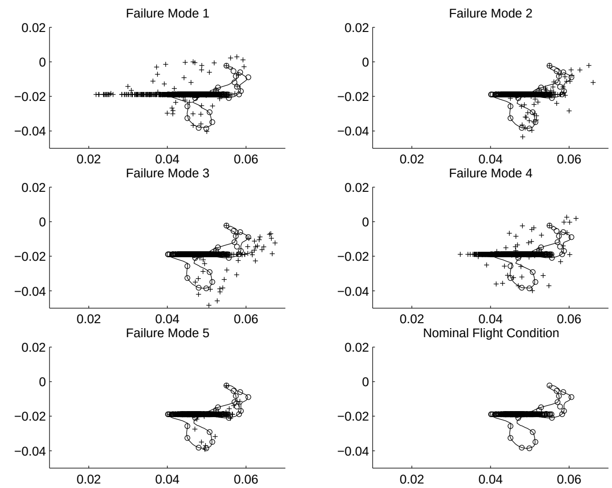

Fig. 2. 2D plots of novelty detection of IFCS simulation data testing on five failure modes.

Figure 2 shows one set of novelty detection results on these five failure mode. The crosses represent the testing data points while the circles represent the learned data points. The closed line surrounding some of the data points is the decision boundary obtained through the SVMs learning on the nominal flight data. The data points falling outside the boundary are those considered novelties. Accordingly, these outliers are effectively detected and reported, consistent with the data inspection by aerospace engineers. Although this is a preliminary demonstration of implementing of the pre-filter (See Figure 1) using the SVM method, its success inspires further investigation.

IV. Conclusion

This paper proposes a novel approach to validating online adaptive systems using support vector machines. We offer a validation framework comprising of two filters that check the validity of inputs and classifications. Empirical results are obtained from running the SVMs for detecting novelties on a sequence of simulation data for an actual flight control system. We conclude that the method provides us as a reasonably good technique for separating nominal flight conditions from the unknown possible failures. Its efficiency also gives us confidence in applying the method for online monitoring of the adaptive learner. In the future, we expect to be able to test our method in an online fashion.

References

[1] S. Yerramalla, B. Cukic and E. Fuller. Lyapunov Stability Analysis of Quantization Error for DCS Neural Networks. *Int'l Joint Conference on Neural Networks*, July 2003.
[2] V. N. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
[3] D.M.J. Tax and R.P.W. Duin, Support vector domain description, *Pattern Recognition Letters*, vol. 20, no. 11-13, 1999, 1191-1199.