Pipeline Alarm Forecasting

Jaired Collins[1], Colin Quinn[2],
George Corliss[2], Richard Povinelli[2]
[1]Missouri Southern State University; [2]Marquette University

Research Question

Is it possible to forecast when an alarm goes off in a natural gas pipeline?

Background

Natural Gas Pipeline Alarms

- Sensors
- Pressure, Temperature, H2S, H2O, Flow
- Alarms are triggered when thresholds are exceeded
- High-High, High, Low, Low-Low alarms

Support Vector Machines (SVM)

- Used for classification

$$\min_{w,e,b} \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{M} \xi_i$$

$$\text{s.t.} \begin{cases} y_i[w^T \varphi(x_i) + b] = 1 - e_i, & i = 1, \dots, M \\ \xi_i \geq 0, & i = 1, \dots, M \end{cases}$$

Support Vector Machines for Regression

- Changes an SVM slightly to include points instead of avoiding them [1]

$$\min_{w,\xi} \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{M} (\xi_i + \xi_i^*)$$

$$\text{s.t.} \begin{cases} y_i - \langle w, x_i \rangle - b & \leq & \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i & \leq & \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq & 0 \end{cases}$$

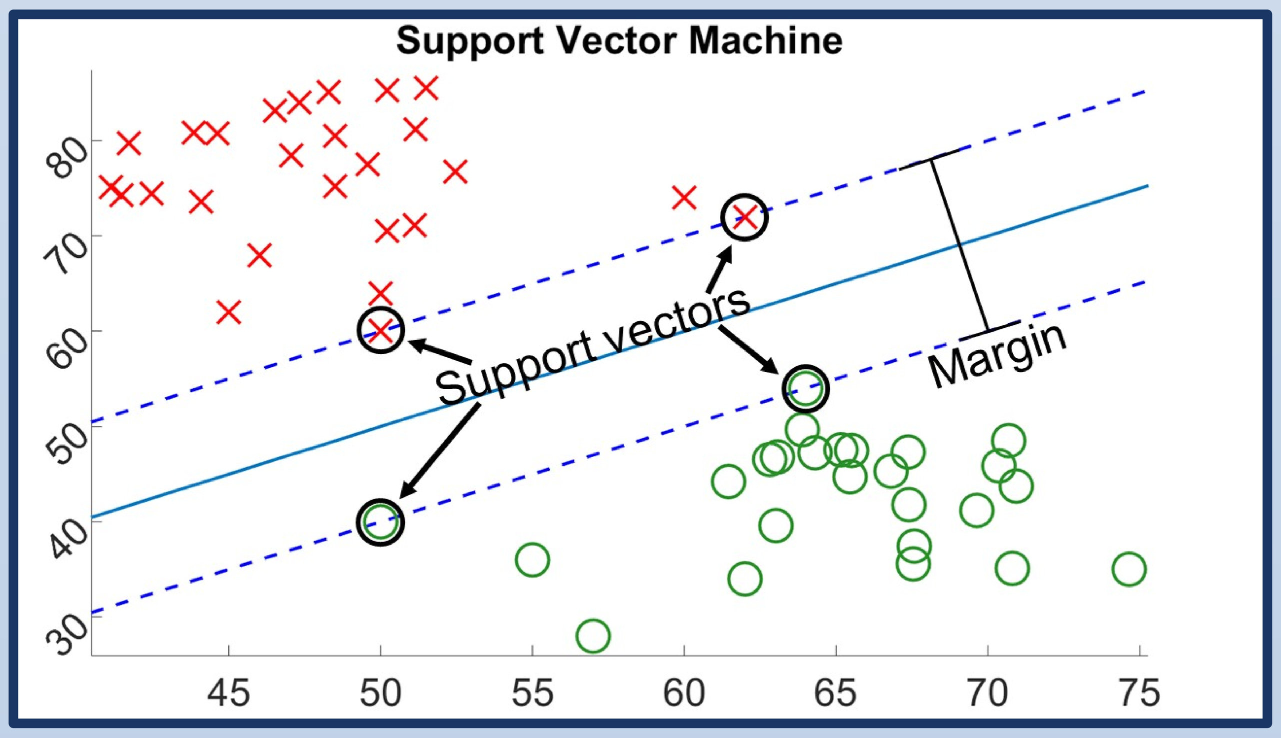

Figure 1: The line with the highest margin separates the data.

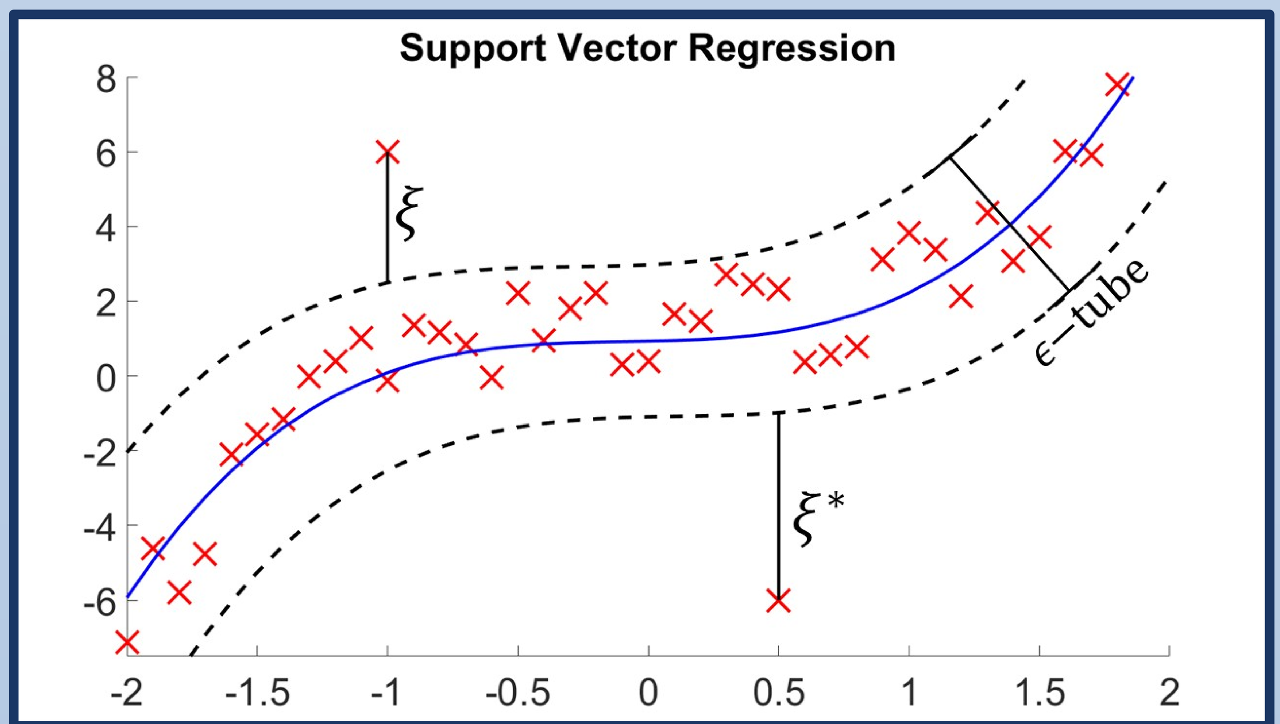

Figure 2: An ε-tube is analogous to an SVM's margin, but instead of precluding data, an ε-tube seeks to contain them.

Data

Scope Restriction

- Only pressure was used
- Data was originally nonuniformly sampled
  - Resampled at 1 minute with zero-order hold
- Anomalous data was removed
- Data was obtained from a gas company in southwest America. To protect their interests, dates were removed and all figures in this poster are rescaled to 100.

Model Selection

Least-Squares Support Vector Machine

- An implementation of an SVM
- Includes a squared error term
- Can also be turned into regression
- Can be turned into a system of linear equations [2]
- Lagrange multipliers α replace weights w and removes error e

$$\begin{bmatrix} 0 & 1_M^T \\ 1_M & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$$

Nonlinear Autoregressive Model (NAR)

- The forecasted value depends nonlinearly on its previous values
- Can be produced with an LS-SVM by

$$\hat{y}_{t+n} = \alpha^T \varphi(y_t, y_{t-}, \dots, y_{t-p}) + b$$

Methods

Rule-Based Ensembling

- Used a least-squares support vector machine for regression (LS-SVR)
- About 10% of data was used for training
- Three different models, one each for trough, normal, and peak data
- When peak values rise above a threshold, the peak values replace the normal model's data
- Trough values replace normal values in a similar fashion but with a low threshold

Alarm Forecasting

- Alarms are binarized according to the defined thresholds
- When a forecasted value exceeds a threshold, an alarm is predicted

Results

Regression and Classification

- Figure 3 shows predicted versus actual values.
- The tables below show metrics of the regression and alarm classification

	Ensemble	Naïve
MAE	1.1258	1.3306
MAPE	0.0995	0.1171
RMSE	1.6446	1.9877

Table 1: Regression performance metrics.

	Accuracy	Sensitivity	Specificity
HH	0.9967	0.9417	0.9994
H	0.9909	0.9583	0.9945
L	0.9781	0.9127	0.9920
LL	0.9970	0.8333	1.0000

Table 2: Ensemble alarm classification metrics.

	Accuracy	Sensitivity	Specificity
HH	0.9940	0.9353	0.9969
H	0.9870	0.9347	0.9928
L	0.9725	0.9218	0.9833
LL	0.9970	0.9180	0.9984

Table 3: Naïve alarm classification metrics.

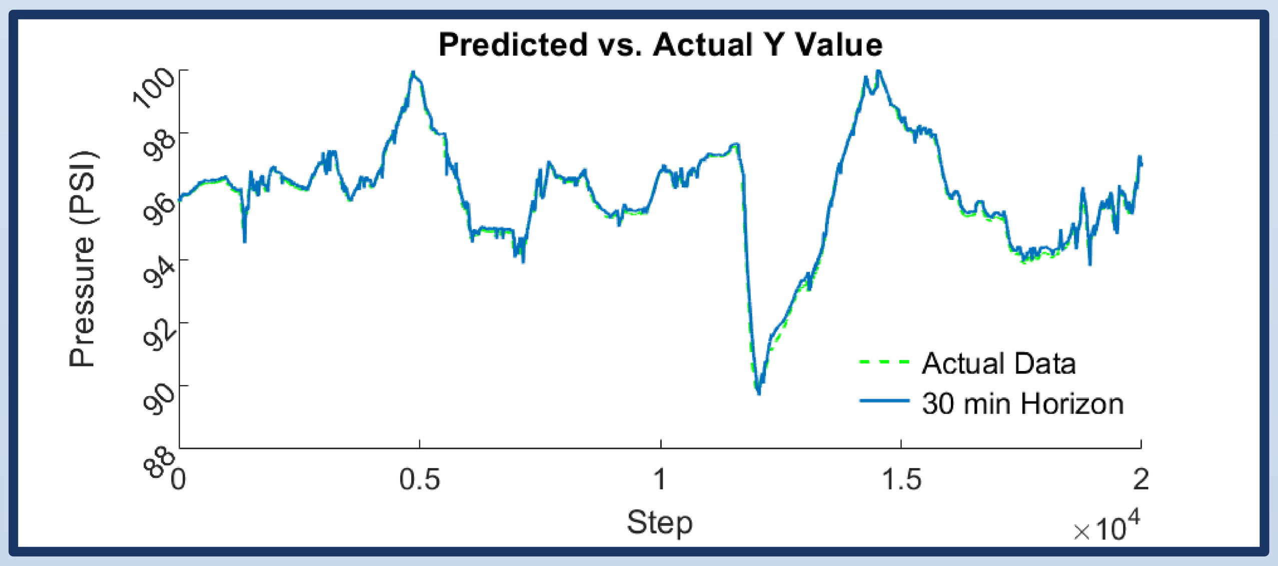

Figure 3: Predicted versus actual pressure values over a span of about 14 days.

Conclusion

Regression

- Trough values are difficult to forecast
- Ensemble model generally works better than naïve in non-troughed values

Alarm Forecasting

- Sensitivity in ensemble is higher in H/HH but not L/LL
- Need to improve trough regression to improve L/LL alarm prediction

References

Alex J. Smola and Bernhard Schölkopf. "A Tutorial on Support Vector Regression". Statistics and Computing, 14(3):199–222, 2004.

J. A. K. Suykens. "Least Squares Support Vector Machines". World Scientific, 2005.

This work was sponsored in part by National Science Foundation REU Site grant #ACI-1461264, 'Computation Across the Disciplines', at Marquette University