A Comparison Between Polynomial and Locally Weighted Regression for Fault Detection and Diagnosis of HVAC Equipment

Regunathan Radhakrishnan, Daniel Nikovski, Kadir Peker, Ajay Divakaran

Abstract

We investigate the accuracy of two predictive modeling methods for the purpose of Fault Detection and Diagnosis (FDD) for HVAC equipment. The comparison is performed within an FDD framework consisting of two steps. In the first step, a predictive regression model is build to represent the dependence of the internal state variables of the HVAC device on the external driving influences, under normal operating conditions. This regression model obtained from training data is used to predict expected readings for state variables, and compute deviations from these readings under various abnormal conditions. The object of the second step in the FDD framework is to learn to detect abnormalities based on regularities in computed deviations (residuals) from normal conditions. The accuracy of the first step (regression) is essential to the success of this method, since it disambiguates whether variations in observed state variables are due to faults or external driving conditions. In this paper, we present a comparison between locally weighted regression (a local non-linear model) and polynomial regression (a global non-linear model) in the context of fault detection and diagnosis of övercharged̈änd ünderchargedr̈efrigerant conditions in an HVAC device show that locally weighted regression clearly outperforms polynomial regression for this task.

IEEE International Conference on Industrial Electronics (IECON)

A Comparison between Polynomial and Locally Weighted Regression for Fault Detection and Diagnosis of HVAC Equipment

Regunathan Radhakrishnan
Daniel Nikovski
Kadir A. Peker
Ajay Divakaran

TR2006-94   November 2006

Abstract

We investigate the accuracy of two predictive modeling methods for the purpose of Fault Detection and Diagnosis (FDD) for HVAC equipment. The comparison is performed within an FDD framework consisting of two steps. In the first step, a predictive regression model is built to represent the dependence of the internal state vari- ables of the HVAC device on the external driving influences, under normal operating conditions. This regression model obtained from training data is used to predict expected read- ings for state variables, and compute deviations from these readings under various abnormal conditions. The objective of the second step in the FDD framework is to learn to detect abnormalities based on regularities in computed de- viations (residuals) from normal conditions. The accuracy of the first step (regression) is essential to the success of this method, since it disambiguates whether variations in observed state variables are due to faults or external driv- ing conditions. In this paper, we present a compar- ison be- tween locally weighted regression (a local non-linear model) and polynomial regression (a global non-linear model) in the context of fault detection and diagnosis of HVAC equip- ment. Our experimental results for detection and diagnosis of over- charged and undercharged refrigerant conditions in an HVAC device show that locally weighted regression clearly outperforms polynomial regression for this task.

Presented at the 32nd IEEE Conference on Industrial Electronics, Paris, France, November 7-10, 2006.

Publication History:–

1. First printing, TR2006-94, November 2006

A Comparison between Polynomial and Locally Weighted Regression for Fault Detection and Diagnosis of HVAC Equipment

Regunathan Radhakrishnan, Daniel Nikovski, Kadir Peker, Ajay Divakaran

Mitsubishi Electric Research Labs
201 Broadway, Cambridge, MA, 02139, USA
regu,nikovski,peker,ajayd@merl.com
www.merl.com

Abstract— **We investigate the accuracy of two predictive modeling methods for the purpose of Fault Detection and Diagnosis (FDD) for HVAC equipment. The comparison is performed within an FDD framework consisting of two steps. In the first step, a predictive regression model is built to represent the dependence of the internal state variables of the HVAC device on the external driving influences, under normal operating conditions. This regression model obtained from training data is used to predict expected readings for state variables, and compute deviations from these readings under various abnormal conditions. The objective of the second step in the FDD framework is to learn to detect abnormalities based on regularities in computed deviations (residuals) from normal conditions. The accuracy of the first step (regression) is essential to the success of this method, since it disambiguates whether variations in observed state variables are due to faults or external driving conditions. In this paper, we present a comparison between locally weighted regression (a local non-linear model) and polynomial regression (a global non-linear model) in the context of fault detection and diagnosis of HVAC equipment. Our experimental results for detection and diagnosis of "overcharged" and "undercharged" refrigerant conditions in an HVAC device show that locally weighted regression clearly outperforms polynomial regression for this task.**

I. Introduction

Our approach to fault detection and diagnosis (FDD) of HVAC equipment is based on statistical machine learning (SML) techniques, and follows the methodology suggested by several extensive studies in this field, such as that by Braun and Li [1] and Zogg [6]. In this approach, fault detection and fault diagnosis are addressed as two separate problems that are solved in sequence by means of two different classes of SML methods. For the purpose of fault detection, a statistical model of sensor readings during *normal* operation of HVAC equipment is learned from data collected from *healthy* equipment. Once such a model is available, it is used to detect signal readings that deviate from the expected values for normal operation — that is, fault detection is reduced to a problem of *anomaly detection*. Once a faulty state has been detected, the objective of fault diagnosis is to establish exactly what kind of fault this is. This problem is also solved by means of SML methods, although in this case the problem is reduced to learning a classification model from data collected from *faulty* equipment in *abnormal* conditions. The two types of SML

methods — those used for fault detection (anomaly detection) and those used for fault classification — are quite different, and will be described in more detail below.

A. Fault Detection

Let the vector $\mathbf{Y} = [\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_n}]$ describe all n sensor readings y_i, $i = 1, n$ and the vector $\mathbf{x} = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_m}]$ describe all the external driving condition variables. Examples of internal state variables $\mathbf{Y}$ for HVAC devices are superheat, evaporator temperature, etc, while typical external variables $\mathbf{x}$ are outside air temperature, humidity, etc. For such devices, fault detection can be solved by means of a model f that relates the external driving conditions $\mathbf{x}$ to normal (expected) operating states $\underline{\mathbf{Y}}$: $\underline{\mathbf{Y}} = f(\mathbf{x})$. If such a model is available, fault detection reduces to comparing actual readings $\mathbf{Y}$ with their expected values $\underline{\mathbf{Y}} = f(\mathbf{x})$ for the current external conditions $\mathbf{x}$, and signaling a fault when their difference exceeds a threshold.

However, obtaining such a model for practical HVAC equipment either analytically or via simulation is usually extremely difficult, due to the complex thermodynamical relationship between external driving conditions and the construction details of HVAC devices. Existing physical models, such as ACMODEL developed by Rossi and Braun [4] have had limited success: their accuracy is low, they require detailed mechanical descriptions of the equipment, and they are not probabilistic. So, even if they could produce an accurate estimate of the expected operating state $\underline{\mathbf{Y}}$ as a function of the driving conditions $\mathbf{x}$, such estimates are deterministic, and do not show what kind of deviations from expected state are acceptable and what are not.

For this reason, research in FDD has turned towards SML models, also known as "black-box models". Such models ignore completely the physical nature of the relationship between driving conditions, device construction, and normal operating states, and instead learn that relationship from data. SML is the area that is concerned with the development of computational learning algorithms that can do that. So far, a fairly large number of learning algorithms have been applied to the problem of normal state determination — successful examples include linear regression, polynomial regression, neural networks, radial

basis function networks, memory-based learning, locally weighted regression, etc. [1]. Thus, the problem of determining the expected normal operating conditions given measurements of external driving conditions is a **regression** task. In addition, since the problem of fault detection reduces to learning a **conditional probability density** model relating driving conditions to the probability that a sensor reading is normal, we are interested in regression methods that also produce estimates of the error on their predictions. Furthermore, both simulation models and exploratory data analysis of HVAC data suggest that the relationship f between external conditions $\mathbf{x}$ and expected readings $\underline{\mathbf{Y}}$ is non-linear, so research has focused on non-linear SML methods.

B. Fault Classification

Once a fault has been detected, its type must be determined as precisely as possible. The common approach is to base this classification on the *deviations* (residuals) $\Delta\mathbf{Y} = \mathbf{Y} - \underline{\mathbf{Y}}$ from expected normal operating conditions $\underline{\mathbf{Y}}$, rather than directly on sensor readings $\mathbf{Y}$. This corresponds to the assumption that the manifestation of a fault (residuals $\Delta\mathbf{Y}$) is largely independent of actual driving conditions $\mathbf{x}$. The practical consequences from this assumption is that training data can be collected from faulty equipment under a large number of driving conditions $\mathbf{x}$, as long as the predictive model identifies correctly $\underline{\mathbf{Y}}$ from $\mathbf{x}$.

From the point of view of SML, this step reduces to building a classifier from training data. One fairly simple approach is to build such a classifier manually, in the form of explicit rules involving only the sign of the residuals. An example of such a rule is "If the superheat is higher than normal, and the temperature of the evaporator is lower than normal, and the sub-cooling temperature is lower than normal, then the fault is due to refrigerant leakage". Clearly, such a rule is pretty crude, because it ignores the exact magnitude of the deviations, as well as their mutual dependency.

A much better approach is to use sensor readings from faulty equipment with correctly diagnosed faults (by a human expert), in order to learn a classifier from data. For this approach, the classifier learns a function that maps the vector of residuals $\Delta\mathbf{Y}$ to a set of discrete class labels, one for each possible fault. In essence, such a classifier would learn how to discriminate between the residuals typical of different faults. Just like the case for prediction under normal operating conditions, many available SML methods exist. In practice, most of the SML models described in the previous subsection can be used for classification just as well, with certain modifications. For example, linear regression is replaced by logistic regression; the output units of neural nets are sigmoidal rather than linear, etc.

Clearly, the accuracy of the regression step influence directly the success of the classification step. Any modeling errors in the regression step would lead to significant residuals in the classification step, and would result in incorrect classifiers. In order to answer the question what kind of regression techniques are good for modeling HVAC equip-

ment, in this paper we present a comparison between one local and one global model for the regression step in the context of fault detection and diagnosis of HVACs. We use locally weighted regression for the local model and polynomial regression for the global model. The rest of the paper is organized as follows. In section 2, we give a brief introduction to the FDD framework followed by us. In section 3, we present the experimental procedure and results with locally weighted regression and polynomial regression. In section 4, we present our conclusions.

II. Fault Detection and Diagnosis Framework

Figure 1 illustrates the regression procedure. As mentioned in the previous section, we learn regression models from normal data relating the external driving conditions (such as outdoor temperature, humidity, and indoor temperature) to the state variables of the HVAC device. Here $\mathbf{x}$ represents the input driving conditions and $\mathbf{y}$ represents the observed state variables under normal operating conditions of the HVAC equipment. Given these data, the task of the regression step is to find a function $\mathbf{f}(\mathbf{x})$ such that an objective function (e.g mean squared error) to measure the goodness of fit is optimized. This allows us to predict the value of state variables ($\mathbf{y}$) under normal operation of the HVAC equipment, given the input conditions $\mathbf{x}$. Such a model can either be a global or a local model. In the global model such as polynomial regression, each training point has the same influence on the model. In the local model such as locally weighted regression, the training points closer to the query have more influence on the model than the farther ones.

Figure 2 illustrates the classification procedure. In this step, we generate prediction residuals based on the regression models and observed values of the state variables. Here $\mathbf{x}$ and $\mathbf{y}$ represent the input conditions and state variables while the equipment is under normal operation. $\mathbf{x_1}$ and $\mathbf{y_1}$ represent the input conditions and state variables while the equipment is operating under a particular fault type (e.g., for the task of detecting incorrect refrigerant level, it can be type 1: Overcharged). The variables $\mathbf{x_2}$ and $\mathbf{y_2}$ represent the input conditions and state variables while the equipment is operating under another fault type (e.g. type 2: Undercharged). If the HVAC is operating normally, the difference between the predicted and observed state variables (residuals) would be small and centered around 0. This residual is shown as ($\underline{\mathbf{Y}} - \mathbf{y}$). Under faulty conditions, the residuals would be large and in a certain direction (positive or negative) depending on the type of fault. The residuals under faulty conditions are indicated as ($\underline{\mathbf{Y_1}} - \mathbf{y_1}$) and ($\underline{\mathbf{Y_2}} - \mathbf{y_2}$). Then, we use these prediction residuals as features and train classifiers to distinguish between three conditions of operation of the equipment (normal, overcharged and undercharged). These three classes (conditions) are indicated as C_N, C_{t1}, C_{t2}.

Once we have completed these two steps, we can monitor the condition of any HVAC equipment in the field using the algorithm shown in figure 3. Using the driving conditions in the field ($\mathbf{x_f}$) as input to the regression model ($\mathbf{f}(\mathbf{x})$)
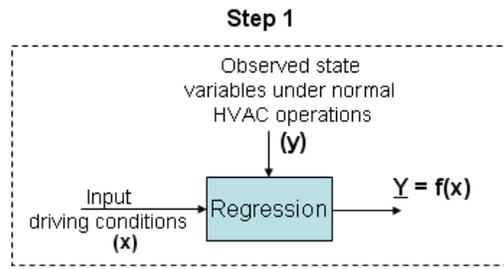
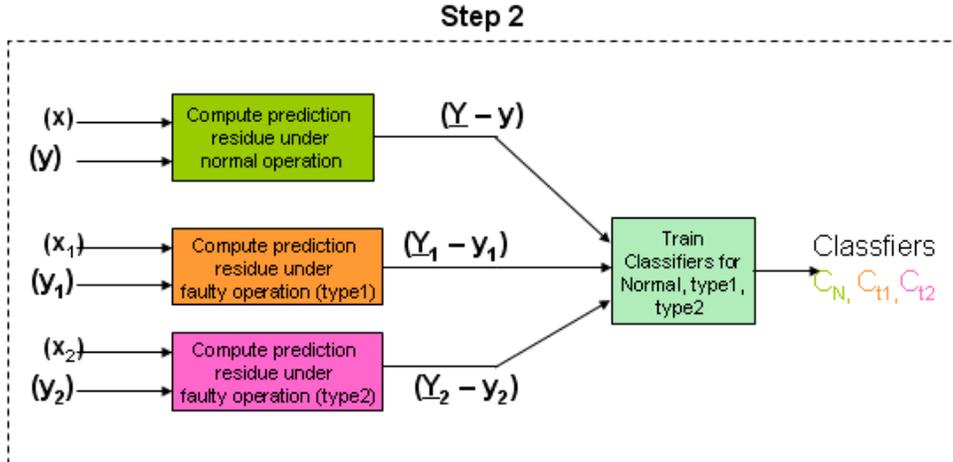Fig. 1. Regression step to model normal operation of HVAC equipment

Fig. 2. Training fault classifier based on residuals

we first predict the value of state variables. The predicted values are represented as $\mathbf{\underline{Y}_f}$. We also have sensor readings for these state variables $(\overline{\mathbf{y_f}})$ from the HVAC equipment in the field. Then, we can generate the residuals $(\mathbf{\underline{Y}_f} - \mathbf{y_f})$ and classify them using the trained classifiers to decide on the condition of the equipment.

In the following section, we present the experimental results for both polynomial regression and locally weighted regression in the context of this FDD framework.

III. EXPERIMENTAL PROCEDURE AND RESULTS

The experimental data was collected from several devices of the same class, installed in different locations. The dataset included sensor readings for the following two driving condition variables: the outside air temperature x_1, and the mixed-air temperature x_2. Accordingly, the driving condition vector $(\mathbf{x})$ consisted of these two variables.

We use the following variables as state variables as suggested by [1]: *Evaporating temperature (iTh_3), Discharge line temperature (Th_1), Condensing temperature (Th_5), Sub-cooling temperature (Th_7), Condenser air temperature difference ($Th_5 - Th_7$), Evaporator air temperature difference ($iTh_3 - iTh_2$)* . Then, the state variable vector $(\mathbf{Y})$ is simply a vector consisting of all these sensor readings. Now that we know the driving conditions $(\mathbf{x})$ and the state variable vector $(\mathbf{Y})$, the first task is to learn the relationship between the two under normal operating conditions. We use the sensor readings in two different installations for

data under normal operating conditions. Note that since these installations were in different locations and readings were collected on different days, the two devices operated under fairly different conditions. However, these readings were not collected under laboratory conditions to ensure thorough coverage of conditions space, as in [1]. One practical consequence from this is that when the regression model is used for prediction, it is much more likely to have to extrapolate external conditions, rather than interpolate among readings experienced during training.

In the following subsections, we present the results of modelling the relationship between $\mathbf{Y}$ and $\mathbf{x}$ using two promising methods. The first method uses polynomial regression, and the second method uses locally weighted regression.

A. Results with polynomial regression

Polynomial regression is a fairly direct extension of ordinary linear regression, with some modifications that allow it to represent non-linear relationships. Rather than computing a linear relationship between input variables $\mathbf{x}$ and output variables $\mathbf{Y}$, a polynomial function of the driving conditions $(g_i(\mathbf{x}))$ is chosen first. After that, the optimal linear mapping between $g_i(\mathbf{x})$) and each individual output variable y_i is computed, using the computational machinery of ordinary linear regression. We conducted our experiments in the statistical environment R, so coefficients were computed using the S-language function lm [5].
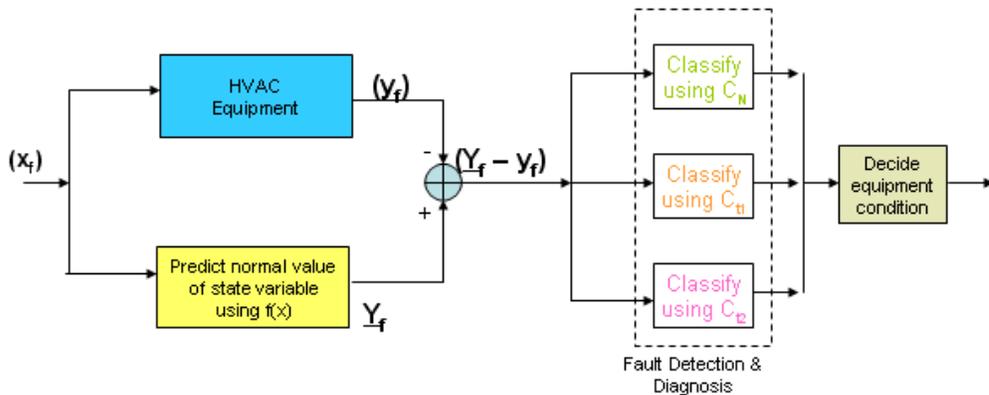
Fig. 3. Fault Detection and Diagnosis Algorithm for HVACs

state variable (y_i)	polynomial function $(g_i(\mathbf{x}))$
iTh_3	1st order
Th_1	3rd order with cross terms
Th_5	1st order
Th_7	2nd order with cross terms
$Th_5 - Th_7$	$1^{st} order$
$iTh_3 - iTh_2$	2nd order with cross terms

TABLE I
POLYNOMIAL FUNCTION ORDERS FOR EACH STATE VARIABLE

Note that the suitable polynomial function $g_i(\mathbf{x})$ might be different for each output variable y_i. We use the polynomial functions suggested in [1] for each state variable. The order of the polynomial is shown in Table I.

B. Results with Locally Weighted Regression

In the previous section, we described how polynomial regression can be used for modeling of steady-state behavior. What is specific to polynomial regression is that the form of the polynomial function $(g_i(\mathbf{x}))$ assumed is global, i.e. used throughout the input space. This is a fairly strong assumption, and most likely not true for most real devices.

One alternative method for non-linear regression that also has close ties with ordinary linear regression is locally-weighted regression (LWR) [3]. LWR operates by computing a custom model for each specific query point in input space, only after this query point is known. While in polynomial regression each training point has the same influence in determining the coefficients of the global model, in LWR the training points nearer to the query point have much more influence on the coefficients of the polynomial than the training points that are farther away. Note that the local model of LWR can also be of order higher than one, just like in PR. However, in our experiments, we used locally linear regression (order one), again using the the R environment and the package LOCFIT [3].

Figure 4 compares the prediction performance of polynomial regression and locally weighted regression on the training set (data from a normally operating HVAC equipment). Note that the prediction errors for the state variable (Condensing Temperature (Th_5)) in case of locally weighted regression are much smaller than the prediction errors in the case of polynomial regression.

C. Comparison of cross-validation performances using residuals from locally weighted regression and polynomial regression

Figure 5 shows the density of prediction residuals using locally weighted regression for one of the state variables (the condensing temperature (Th_5) under three operating conditions of the HVAC equipment. The densities of residuals are plotted for the following three scenarios: normally charged refrigerant, varying amounts of overcharged refrigerant and varying levels of undercharged-ness. Note the prediction residuals are centered around zero for normally charged HVAC equipment while they are shifted in opposite directions for overcharged and undercharged scenarios of the HVAC equipment. Since the model is fitted on training data from a "healthy" machine, we expect little discrepancy between the predicted and observed state variable if the machine is operating under normal conditions. If there is a fault (overcharged or undercharged refrigerant), we expect to see a discrepancy between the observed value and the predicted value. The direction and the value of the discrepancy may be indicative of the type of fault. Thus, we are motivated to use such prediction residuals as features for discriminating various types of faults from normal operating conditions. It is easy to see that there is a clear separation between the prediction residuals under three operating conditions if we use this feature alone with a threshold. We would like to learn a classifier that combines the strength of several such simple and weak classifiers (accuracy > 0.5) to identify different fault types. Boosting is a natural choice for a such a classifier [2].

We train three boosting classifiers using the prediction errors as features. The first classifier is for discriminating normal operation from overcharged and undercharged operations. The second classifier is for discriminating overcharged operation from normal and undercharged oper-
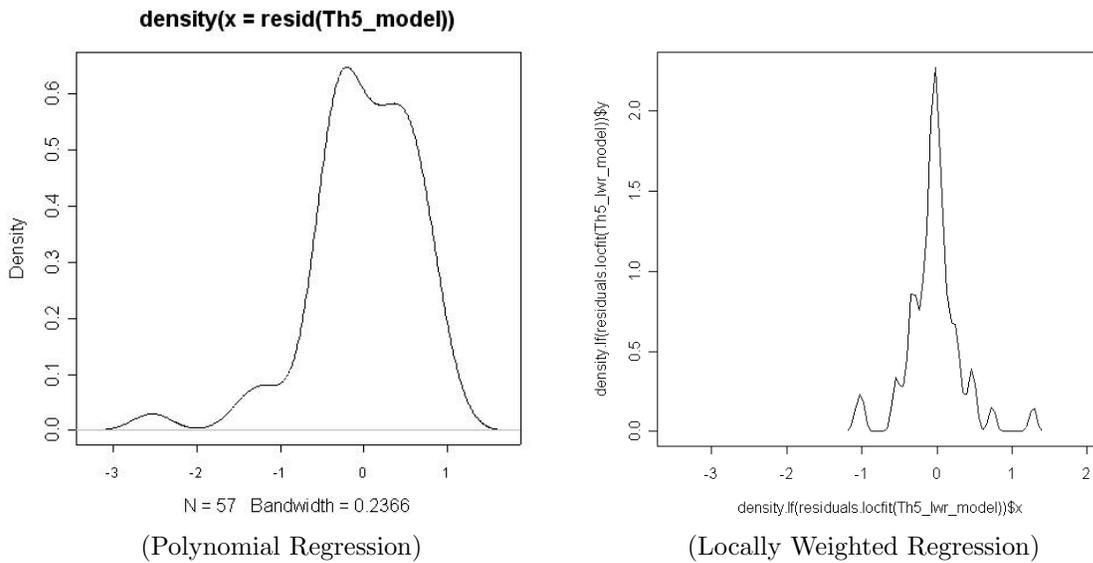
density(x = resid(Th5_model))

(Polynomial Regression)   (Locally Weighted Regression)

Fig. 4. Density of prediction residuals with normal data, for the state variable condensing temperature.
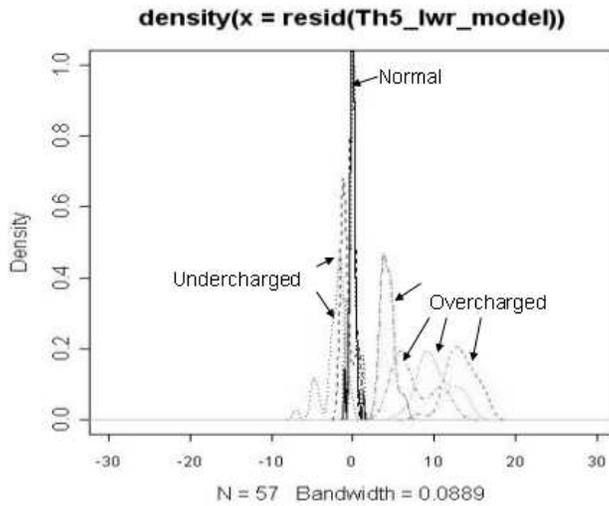

density(x = resid(Th5_lwr_model))

Fig. 5. Density of prediction residuals under faulty operation of HVAC equipment for the state variable condensing temperature.

ations. The third classifier is for discriminating under-charged operation from overcharged and normal opera-tions. In the following subsections, we present the cross validation performances of these three classifiers on vari-ous splits of the data for training and testing.

Table II summarizes the cross validation performance of the boosting classifier trained to discriminate between the normal operation and the faulty operation. It shows that the boosting classifier trained with residuals from Lo-cally Weighted Regression (LWR) outperforms the Boost-ing classifier trained with residuals from Polynomial Re-gression (PR). We can draw similar conclusions for the other two classifiers (Overcharged Vs Rest and Under-charged Vs Rest) from the tables III and IV. The over-all classification accuracy for residuals from LWR is 95% whereas the classification accuracy for residuals from PR is only 78%.

The results so far indicate that we can build discrimi-

Run	LWR	PR
1	0.06	0.45
2	0.07	0.45
3	0.04	0.21
4	0.38	0.44
5	0.08	0.21
6	0.08	0.22
7	0.05	0.22
8	0.05	0.22
Avg. Error	0.1012	0.3042

TABLE II

Comparison of cross-validation performances using residuals from Locally Weighted Regression (LWR) and Polynomial Regression(PR) for Normal Vs Rest

Run	LWR	PR
1	0.07	0.01
2	0.06	0.02
3	0.013	0.02
4	0.10	0.21
5	0.05	0.14
6	0.02	0.09
7	0.05	0.08
8	0.05	0.25
Avg. Error	0.0516	0.1025

TABLE III

Comparison of cross-validation performances using residuals from Locally Weighted Regression (LWR) and Polynomial Regression(PR) for Overcharged Vs Rest

Run	LWR	PR
1	0.0	0.0
2	0.0	0.0
3	0.0	0.0
4	0.0	0.29
5	0.0	0.47
6	0.0	0.48
7	0.0	0.44
8	0.012	0.41
Avg. Error	0.0015	0.2612

TABLE IV

Comparison of cross-validation performances using residuals from Locally Weighted Regression (LWR) and Polynomial Regression(PR) for Undercharged Vs Rest

native classifiers to distinguish between normal and both of the faulty operations (overcharged and undercharged). We have got similar promising results with locally weighted regression on simulator data.

IV. Conclusion

In this paper, we presented a comparison between the performance of global regression method (Polynomial Regression) and a local regression method (Locally Weighted Regression) for modelling the state variables of a HVAC equipment as a function of two driving condition variables. The prediction residuals obtained by using these two regression methods were treated as features to train three classifiers. The first classifier learns to distinguish normally charged HVAC from the overcharged and the undercharged, the second classifier to distinguish the overcharged HVAC from the normally charged and the undercharged and the third classifier to distinguish the undercharged HVAC from the normally charged and overcharged. The overall classification accuracy when using prediction residuals from locally weighted regression was found to be 95%. The same measure dropped to 78% when using prediction residuals from polynomial regression. Locally Weighted

Regression is clearly superior to Polynomial Regression for the task of modelling the state variables as a function of external driving conditions, and is hence is more useful for FDD in HVAC equipment. This can be explained by the greater flexibility of LWR that is helpful in accommodating various local dependencies, and its superior extrapolation abilities.

References

[1] Braun, J., and Li, H. Automated fault detection and diagnosis of rooftop air conditioners for California. Final Report 2.1.6a,b, Purdue University, Mechanical Engineering Dept., 2003.
[2] Freund, Y., and Schapire, R. E. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence 14*, 5 (Sep, 1999), 771 – 780.
[3] Loader, C. *Local regression and likelihood.* Springer, Heidelberg, 1999.
[4] Rossi, T., and Braun, J. A statistical, rule-based fault detection and diagnostic method for vapor compression air conditioners. *International Journal of HVAC Research 3* (1997), 19–37.
[5] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S. Fourth Edition.* New York: Springer, 2002.
[6] Zogg, D. *Fault diagnosis for heat pump systems.* PhD thesis, Swiss Federal Institute of Technology, Zurich, 2002.