Title

Optimal and Heuristic Techniques for Fault Detection (MAS 8)

Authors
Kannika Sikangwan
Miodrag Potkonjak

Publication Date
2006

Center for Embedded Networked Sensing

Optimal and Heuristic Techniques for Fault Detection

Kannika Sikangwan, Farinaz Koushanfar, Miodrag Potkonjak
Computer Science Department, CENS

Introduction: Useful information may not be extracted from the sensor networks due to some faulty readings

Integrity of collected data through fault detection

- **Identify the source of faulty and related missing data**
  - Lossy wireless network links
  - Unpredictable impact of environment
  - Hardware and software faults

Faulty Data Detection and Correction Problem

- **NP-complete problems**

Problem Description: Detect faulty readings of the sensor data stream

Detection and correction of faulty sensor data

- **Primary principle: *Separation of concerns***
- **Identify and address two modular phases**
  - Inter-sensor model building and validation
  - Fault detection using local and global consistency checking

Two phase process

- **Inter-sensor prediction models for each sensor**
  - Built using non-parametric statistical modeling methods
    - **No assumptions about the underlying distribution of the variables**
  - Creation of consistency graph
- **Consistency graph utilized for fault-detection**
  - Combinatorial optimization problem
    - **Solved using both optimal and heuristic approaches**

Proposed Solution: Use consistency among sensors to identify faults

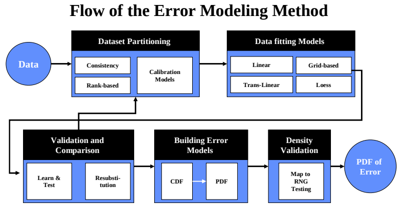
Learning and Testing

- **use p% of data to create a model (probabilistic, statistical)**
- **use the rest of data to test the model fit**

The error model enable fault detection

Model: **Markov Chain-based, Statistical, Univariate**

Flow of the Error Modeling Method



Basic Markov Chain Model

- Markov property: conditional independence of future on the past.
- Set of time dependent random vars (states).
- Markov chain model:
  - State q_t at time t is one of a finite number of states in the range $\{s_1,\ldots,s_M\}$
  - State sequence vect. $Q=(q_0,\ldots,q_N), t=1,\ldots,N$
  - N^{th} order Markov chain: Probability of state q_t at t conditioned on all N previous states up to $t-1$

$P(q_t=s_i \mid q_0=s_{j0}, q_1=s_{j1}, \ldots, q_{t-1}=s_{j(t-1)})$
$= P(q_t=s_i \mid q_{t-n}=s_{j(t-n)}, \ldots, q_{t-1}=s_{j(t-1)})$

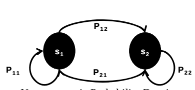
Semi Markov Chain Model

- *Semi-Markov models* ensure lagged autocorrelation are properly captured w/o a significant overhead on size/complexity of models
- Semi-Markov models are a hybrid combination of *state-space* model and *probability density function (PDF)* for each state
- The continuous PDF at each discrete state quantifies the conditional probability of staying in the same state

States of data: Correct, Faulty, Missing

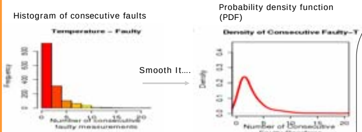
Sensor data streams tend to stay in the same state for more than one epoch

1st order Markov Model

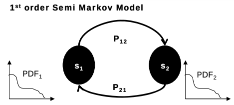


Nonparametric Probability Density Functions (PDFs)

- For simplicity, assume a two state system: (s_1) Correct, (s_2) Faulty
- Build a histograms for consecutive faulty and consecutive correct data
- Apply nonparametric statistical kernel smoothing techniques to smooth the histogram
  - Kernel Smoothing: *Silverman and Green, 1986*
- Example: 1st order Semi Markov model:



1st order Semi Markov Model

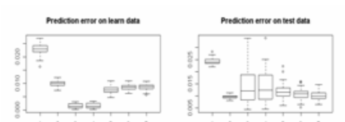


Nonparametric PDF avoid strong prior assumptions about distribution of data

Test in RNG, if no prediction of any correlated sensor using the measurement of sensor -> Faulty

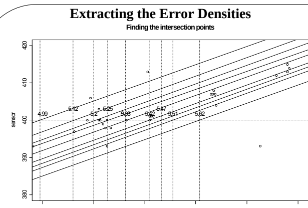
Statistical Model

Comparison of Different Models



Boxplots of the prediction errors: (1) Linear model, (2) linear model and logarithmic transform, (3) non parametric $-L_1$, (4) non parametric $-L_2$, (5) loess – span=0.2, (6) loess – span=0.3, (7) loess – span=0.4.

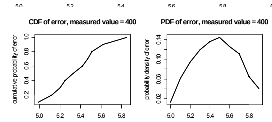
Extracting the Error Densities



Finding intersection of the line y=400 with 2K% lines

The points that are far from the calibration curve have a high error

Q: The calculated probability that sensor has measurement with specific error



CDF of the error (left) was formed from intersection points above. PDF of the error is by differentiation and smoothing CDF (right)

Using Probabilities to Identify Faulty Measurements

- Measurement x_i is faulty with a probability Q_i $(0 \le Q_i \le 1)$
- Q_{total} indicates total number of faulty measurements
- Q_i's are added to the eq's as optimization variables
- OF: $\max_p \prod_{i=1}^{n} P_{\varepsilon i} P_{\delta i1} P_{\delta i2} P_{\delta i3}$

1. $|(1-Q_i) f_i(x_i + g_\varepsilon^{-1}(P_{\varepsilon i}), y_1, y_2, \ldots, y_k)| \le g_\delta^{-1}(P_{\delta i1})$

s.t. $(1 \le i \le n)$

2. $|Q_i(1-Q_i)| \le g_\delta^{-1}(P_{\delta i2})$

3. $|Q_{total} - \sum_{i=1}^{n} |Q_i|| \le g_\delta^{-1}(P_{\delta i3})$

Method 2: Cross-Validation by Weighting the Discrepancies

$$\max_p \sqrt[r]{P_{\varepsilon j} P_{\delta j1}} \prod_{i=1, i \ne j}^{n} P_{\varepsilon i} P_{\delta i1}$$

Faults ~8% are faulty in Intel data

Univariate Modeling

- Given the time series of sensor readings, can a sensor Y be predicted from readings at sensor X (i.e., $Y^*=f(X)$) s.t. an error norm is minimized?
- Commonly used forms of error norm

$$L_p = \left(\sum_{t=1}^{T} w(t)\,|y(t)-y^*(t)|^p\right)^{1/p} \qquad 1 \le p < \infty$$

$$L_p = \max_{t=1}^{T} w(t)\,|y(t)-y^*(t)| \qquad p = \infty$$

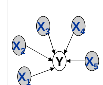
Note: the regression approach is transparent to the ensemble learning method

- Hidden covariates → isotonicity constraint
- Combinatorial isotonic regression (CIR)
- Flexibility of the combinatorial domain
  - Optimally constrain the number of level sets
  - Controlling the slopes, imposing additional constrains such as convexity, symmetry, etc.

Univariate CIR Approach

- Histogram → error matrix E, $e_{ij} = \varepsilon_p(x_i, \hat{y}_j)$
- Build the cumulative error matrix CE
- Map the problem to combinatorial domain

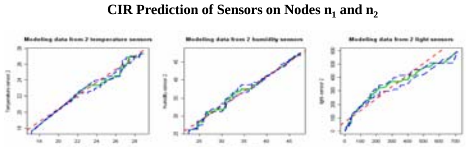
Mapping from An Additive Model to A Set of Classifiers



- At $t=t_1$, sensors $X_1,\ldots,X_M$ measure values $x_1(t_1), x_2(t_1),\ldots, x_M(t_1)$
- Define a range,
  - $R_\aleph = \aleph\{f_i(x(t_1))\}$
- For each $r_{ni} \in R_\aleph$,
  - If $(f_i(x_i(t_1)) < r_n)$
    - $G_i(x_i(t_1))=0$
  - Else
    - $G_i(x_i(t_1))=1$

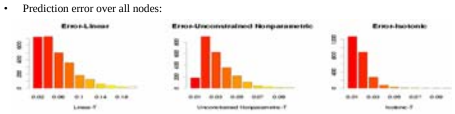
$Y^*_1=f_1(X_1)$ — $G_1(r_n)$
$Y^*_2=f_2(X_2)$ — $G_2(r_n)$
$Y^*_3=f_3(X_3)$ — $G_3(r_n)$
$Y^*_4=f_4(X_4)$ — $G_4(r_n)$
$Y^*_5=f_5(X_5)$ → $G_5(r_n)$

CIR Prediction of Sensors on Nodes n_1 and n_2

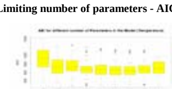


CIR Prediction Error on Temperature Sensors

- Prediction error over all nodes:



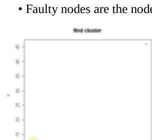
- **Limiting number of parameters - AIC criteria**



Non-parametric Histogram (L1 error)

Prediction value is the median of all look-up-table value of predicting node

- Map the distance between nodes in a new space with weight on faulty readings
- Use NLP to minimize prediction errors among all nodes
- Faulty nodes are the nodes outside of cluster



Detect faulty reading at one time snapshot

NLP of consistent model

- Use the discrepancy between the predicted value from model and measurement

$$\sum_{j=i}^{n} \sum_{i=1}^{n} (\text{prediction error}_{ij} - \text{distance}_{ij})^2$$

Distance: distance between nodes in new space

$$\sum_{j=i}^{n} \sum_{i=1}^{n} (\max((pred_{ij}-real_i), (pred_{ji}-real_j)) - \text{sqrt}((x_i-x_j)^2+(y_i-y_j)^2)^2$$

(xi, yi) reading i in new space

UCLA – UCR – Caltech – USC – CSU – JPL – UC Merced