Exchangeability Martingales for Selecting Features in Anomaly Detection

Giovanni Cherubin\textsuperscript{1}, Adrian Baldwin\textsuperscript{2}, Jonathan Griffin\textsuperscript{2}

\texttt{@gchers}

COPA18, Maastricht
13 June, 2018
Problem

(Unsupervised) Anomaly Detection on multi-variate time series.
Feature Selection
Supervised Learning

- Class A
- Class B
Feature Selection

Supervised Learning

Class A

Class B

Density of i-th feature’s values
Feature Selection

Anomaly Detection

Density of i-th feature’s values
Feature Selection

Anomaly Detection

- Normal
- Anomaly

Density of i-th feature's values
A feature has to follow the same pattern over time:
Feature Selection
Anomaly Detection

A feature has to follow the **same pattern** over time:
Feature Selection

Anomaly Detection

A feature has to follow the same pattern over time:
Feature Selection

Anomaly Detection

A feature has to follow the **same pattern** over time:
Notation

Consider a time series $x_1, x_2, \ldots, x_n$, where $x_i$ is $d$-dimensional.

We indicate the values of the $i$-th feature of an object with:

$$v_j = x_j^i$$
Notation

Consider a time series $x_1, x_2, \ldots, x_n$, where $x_i$ is $d$-dimensional.

We indicate the values of the $i$-th feature of an object with:

$$v_j = x^i_j$$

**Goal**: determine if $v_1, v_2, \ldots, v_n$ are i.i.d. (exchangeable).
Testing i.i.d. Hypothesis

[VNG’05]
Testing i.i.d. Hypothesis

[VNG’05]

1) \( \mathbf{v} = v_1, v_2, \ldots, v_n \)
Testing i.i.d. Hypothesis

[VNG’05]

1) \( V = v_1, v_2, \ldots, v_n \rightarrow \text{CP} \rightarrow \ p_1, p_2, \ldots, p_n \)
Testing i.i.d. Hypothesis

[VNG’05]

1) \( \mathbf{V} = v_1, v_2, \ldots, v_n \rightarrow \text{CP} \rightarrow p_1, p_2, \ldots, p_n \)

If \( \mathbf{V} \) are i.i.d. => p-values are \( \text{Uni}(0,1) \)

[VGS’05]
Testing i.i.d. Hypothesis

[VNG’05]

1) \(V = v_1, v_2, \ldots, v_n \rightarrow \text{CP} \rightarrow p_1, p_2, \ldots, p_n\)

If \(V\) are i.i.d. \(\Rightarrow\) p-values are Uni(0,1) \[VGS’05\]

2) \(p_1, p_2, \ldots, p_n\)
Testing i.i.d. Hypothesis

[VNG’05]

1) \( \mathbf{V} = v_1, v_2, \ldots, v_n \rightarrow \text{CP} \rightarrow p_1, p_2, \ldots, p_n \)

If \( \mathbf{V} \) are i.i.d. \( \Rightarrow \) p-values are Uni(0,1) [VGS’05]

2) \( p_1, p_2, \ldots, p_n \rightarrow \text{b} \rightarrow M_1, M_2, \ldots, M_n \)
Testing i.i.d. Hypothesis
[VNG’05]

1) \( \mathbf{V} = v_1, v_2, \ldots, v_n \rightarrow \text{CP} \rightarrow p_1, p_2, \ldots, p_n \)

If \( \mathbf{V} \) are i.i.d. \( \Rightarrow p \)-values are Uni(0,1) \[\text{VGS’05}\]

2) \( p_1, p_2, \ldots, p_n \rightarrow b \rightarrow M_1, M_2, \ldots, M_n \)

If \( p \)-values are Uni(0,1) \( \Rightarrow P(\exists t: M_t \geq \lambda) \leq 1/\lambda \) \[\text{V’39}\]
Testing i.i.d. Hypothesis

[VNG’05]

1) \( V = v_1, v_2, \ldots, v_n \) \( \rightarrow \) CP \( \rightarrow \) \( p_1, p_2, \ldots, p_n \)

If \( V \) are i.i.d. \( \Rightarrow \) p-values are Uni\((0,1)\) [VGS’05]

2) \( p_1, p_2, \ldots, p_n \) \( \rightarrow \) b \( \rightarrow \) \( M_1, M_2, \ldots, M_n \)

If p-values are Uni\((0,1)\) \( \Rightarrow \) \( P(\exists t: M_t \geq \lambda) \leq 1/\lambda \) [V’39]

Reject HP if some \( M_t > \lambda \)
Exchangeability Martingales

Martingale: a sequence of r.v. $M_1$, $M_2$, ... that keeps the conditional expectation:

$$E(M_{t+1} \mid M_1, \ldots, M_t) = M_t$$

Exchangeability martingale defined for betting function $b_i$:

$$M_t = \prod_{i=1}^{t} b_i(p_i) \quad t = 1, 2, \ldots$$

[F+’12] Plug-in martingales: $b$ is density estimate (use KDE with Gaussian kernel).
Exchangeability Martingales

They give us:

- a test for i.i.d.
- easy to interpret plots
Monitor a Windows service via the process monitoring API

Features:
- CPU and memory usage (user/privileged)
- Number and type of resources
- Loaded DLLs
- Network information
Experiments
Experiments

Train AD model and test it on test data, before and after feature selection.
Experiments

Train AD model and test it on test data, before and after feature selection.

Details:
Experiments

Train AD model and test it on test data, before and after feature selection.

Details:
• Plug-in Martingales
Experiments

Train AD model and test it on test data, before and after feature selection.

Details:
• Plug-in Martingales
• CP with k-NN (k=5) nonconformity measure
Experiments

Train AD model and test it on test data, before and after feature selection.

Details:
• Plug-in Martingales
• CP with k-NN (k=5) nonconformity measure
• One-class SVM, RBF kernel, parameters from grid search
Experiments

Train AD model and test it on test data, before and after feature selection.

Details:
• Plug-in Martingales
• CP with k-NN (k=5) nonconformity measure
• One-class SVM, RBF kernel, parameters from grid search
• Measure Precision/Recall and F1
## Results

<table>
<thead>
<tr>
<th>Feature selection</th>
<th># Features</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>88</td>
<td>0.725</td>
<td>0.999</td>
<td>0.840</td>
</tr>
<tr>
<td>Martingales (λ = 20)</td>
<td>72</td>
<td>0.858</td>
<td>0.999</td>
<td>0.923</td>
</tr>
</tbody>
</table>

Training on 4 hours of data, test on 24 hours
Results

Training set: 4 to ~24 hours of data
Test set: 24 hours of data
Future Work

• Apply/extend to other unsupervised problems (e.g. clustering)

• Other betting functions and strategies for constructing exchangeability martingales

• Application to Windows service: combine features selected across devices, measure differences between different machines
References


Exchangeability Martingales for Selecting Features in Anomaly Detection

Giovanni Cherubin¹, Adrian Baldwin², Jonathan Griffin²

@gchers

COPA18, Maastricht
13 June, 2018
Feature Selection
Robustness

Before feature selection

After feature selection

SVM hyperplane

- Normal
- Normal (training)
- Anomalies
Exchangeability Martingales for Selecting Features in Anomaly Detection

Giovanni Cherubin¹, Adrian Baldwin², Jonathan Griffin²
@gchers

COPA18, Maastricht
13 June, 2018