Event detection
Noise sources & signal quality

Centre for Doctoral Training in Healthcare Innovation

Dr. Gari D. Clifford,
University Lecturer & Director,
Centre for Doctoral Training in Healthcare Innovation,
Institute of Biomedical Engineering,
University of Oxford
Overview

- Univariate event detection in time series (e.g. ECG)
  - Amplitude thresholding
  - Slope thresholding
  - Energy thresholding / CUMSUM
  - Moving Average
  - Exponentially weighted moving average
  - Matched filters & correlation
  - Frequency thresholding
  - Statistical techniques:
    - Control Chart (Shewhart 1931)
    - Regression
    - Box Jenkins Models
    - Kalman Filter (dynamic parameter tracking)
    - Hidden Markov Models

- Noise in biomedical signals
  - ... and their effects on event detection

- Signal quality metrics & ‘trust factors’
  - Absolute
  - Relative
How is event detection different from:

- **Supervised Learning:**
  - In event detection, abnormal events are extremely rare, normal events are plentiful

- **Clustering:**
  - Clustering = partitioning data into groups
  - Not the same as finding statistically anomalous groups

- **Outlier Detection:**
  - Events of interest are usually not individual outliers
  - The event typically affects a subgroup of the data rather than a single data point
Amplitude thresholding

- Simple!
  - If HR > 140 BPM ... tachycardia

- Adaptive
  - If HR > 140 BPM more than 4 times in minute, increase threshold by 10%

Dumb – sensitive to outliers and noise
Simple!
- If $\Delta HR/\Delta t > 50/10$ BPM/s ... alarm

Less dumb – but even more sensitive to outliers and noise – see example on differentiation later in lecture
Control chart (from Statistical Quality Control)

- Estimate $\hat{\mu}$ and $\hat{\sigma}$ from data up to current time
  
  $\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} X_i$
  
  $\hat{\sigma} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (X_i - \hat{\mu})^2}$

- Upper control limit = $\hat{\mu} + 3\hat{\sigma}$
- Raise alarm if upper control limit exceeded
Let $W$ be the window size

A moving average window is formed as follows:

$$X_{t+1} = \frac{1}{W} (X_t + X_{t-1} + \ldots + X_{t-W+1})$$

Setting the alarm value:

- Fit a Gaussian to the $W$ observations within the window i.e. estimate $\hat{\mu}$ and $\hat{\sigma}$
- Calculate the alarm level using

$$\text{Alarm level} = \Phi\left( \max(0, \frac{X_i - \hat{\mu}}{\hat{\sigma}}) \right)$$
Exponentially Weighted Moving Average (EWMA) - a variation on the moving average

Let $Z_i$ be the EWMA statistic (which is monitored):

$$Z_i = \lambda X_i + (1 - \lambda)Z_{i-1} \quad \text{where} \quad 0 < \lambda \leq 1$$

Observations in the past receive a decreasing amount of weight.
CUmulative SUM Statistics

- CUmulative SUM (CUMSUM) Statistics – or energy thresholding
- Good at detecting shifts from the mean more quickly than control chart
- Keep a running sum of “surprises”: a sum of excesses each day over the mean
- When this sum exceeds threshold $H$, signal alarm and reset sum
CUSUM

- $r =$ reference value eg. mean
- $X_i =$ $i^{th}$ observation
- $S_i =$ $i^{th}$ cumulative sum

$$S_1 = X_1 - r$$
$$S_2 = (X_2 - r) + (X_1 - r) = (X_2 - r) + S_1$$
$$\vdots$$

$$S_k = \sum_{i=1}^{k} (X_k - r) + S_{k-1}$$

When a shift from the mean occurs, $S_i$ will start to increase
If we are only tracking increases, we can do the following:

\[ S_k = \max(0, (X_k - r) + S_{k-1}) \]

Ensures we don’t go below 0

We can also add a tolerance or a slack \( K \)

\[ S_k = \max(0, X_k - (r + K) + S_{k-1}) \]
Advanced event detection

- Box-Jenkins models (e.g. ARMA, ARIMA)
- Kalman filters (look at prediction errors/innovation)
- Change-point detection (extreme value stats)
- Hidden Markov models (threshold on $p(\text{state})$)
- Wavelets (ridges, $\partial/\partial t=0$)
An example – slope detection of the ECG

- So how do we calculate the slope?
- What’s the first difference of this signal?
An example – slope detection and the 1st difference filter

- Very noisy!
What is the impulse response of a differentiator?

(Hint: Think of what coefficients you would need in a filter to sequentially subtract one sample from the next.)
Impulse response for first difference filter
Smoothen with more taps on the filter
Smooother differentials

Differential filters applied to an ECG
Watch out for the delays

Differential filters applied to an ECG

amplitude (mV)

sample number
Impulse and frequency response
Performance evaluation statistics

- How do we measure performance?

Typically:
- Take a set of labelled data
  - (different from the one on which the algorithm was ‘tuned’)
- Run your algorithm across it
  - Vary your algorithm’s parameters
- Report how well it performed

So what statistics do we use?
Performance evaluation statistics

TP (True Positive) = correctly labelled events
FN (False Negative) = missed real events
FP (False Positive) = event that was detected that is not there
TN (True Negative) = correctly labelled non-events – NOT relevant in QRS detection!

- Sensitivity (Recall): $Se = \frac{TP}{(TP+FN)}$
- Positive Predictive Value (Precision): $PPV = \frac{TP}{(TP+FP)}$
- Specificity: $Sp = \frac{TN}{(TN + FP)}$.
- Negative Predictive Value: $NPV = \frac{TN}{(TN+FN)}$
  (proportion of patients with negative test results who are correctly diagnosed as not having the condition)

- Accuracy: $Acc = \frac{TP+TN}{(TP+FN+TN)} = (Se)(Pr) + (Sp)(1 – Pr)$
  \[Pr=Prevalence=Proportion=a/(a+b)\]

- Accuracy in Information retrieval is sometimes measures in terms of Precision & Recall by the F-measure

$$F = 2 \cdot \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

- The F-measure more appropriate in situations with very skewed class imbalance or when the minority class is more important
How do we evaluate event detection algorithms?

- Can’t use prediction accuracy for “event” vs “non-event”
  - Class imbalance: many more “non-events” than “events”
  - Guessing “non-event” (TN) all the time results in very good accuracy

Use only:

- TP (True Positive) = correctly labelled events
- FN (False Negative) = missed real events
- FP (False Positive) = event that was detected that is not there

i.e.

- Sensitivity (Recall): $Se = \frac{TP}{TP+FN}$
- Positive Predictive Value / precision: $PPV= \frac{TP}{TP+FP}$

- But you always trade-off Se for PPV – e.g. lots of false alarms in the ICU to save missing lives!

- Most event detection algorithms have a tunable threshold for when an alarm is raised
  - Trades off specificity and sensitivity (true and false alarm) rate
  - Need performance over multiple thresholds

How do we represent this trade-off and compare accuracies of different detectors over all ranges?
How to evaluate event detection algorithms

To evaluate accuracy, usually use a Receiver Operating Characteristic (ROC) curve (usually Sensitivity vs [1-Specificity])

Calculate **AUC**: Area Under Curve

To evaluate timeliness of detection, use an Activity Monitoring Operating Characteristic (AMOC) curve (Fawcett and Provost 1999)
- **Accuracy** \(\frac{TP+TN}{TP+FN+TN}\): the degree of closeness of measurements of a quantity to that quantity's actual (true) value.
  - Low mean, high variance / std

- **Precision (reproducibility~repeatability)** \(\frac{PPV=TP}{TP+FP}\): the degree to which repeated measurements, under unchanged conditions, gives the same results.
  - High abs(mean) or **bias**, low variance/std
Accuracy/precision trade-off

- Generalised F measure: \[ F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}} \]

- \[ F_\beta = \frac{(1 + \beta^2) \cdot \text{true positive}}{(1 + \beta^2) \cdot \text{true positive} + \beta^2 \cdot \text{false negative} + \text{false positive}} \]

- \( \beta > 1 \) reflects a weighting towards Se/recall (e.g. \( F_2 \))
- \( \beta \) measures the effectiveness of ‘retrieval’ (detection) with respect to a user who attaches \( \beta \) times as much importance to recall as precision

- Equal weighting \( \beta = 1 \):

\[ F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \]

- (Harmonic mean of precision and recall)
Measuring significant differences in AUC

- Q: How do I know if I have made a better detector?
- A: Test to see if AUCs is larger and the increase is significantly different

- Two algorithms, or classifiers, C₁ and C₂
- Two AUROCs, AUC₁ and AUC₂
- How to test whether AUC₁ is significantly different than AUC₂?
Analysis of AUROC

- First guess: t-test!
- Close, not quite..

- What are the assumptions of the t-test?
  - Data are continuous
  - Data are normally distributed
  - Data are random samples
  - Two populations have equal variance
  - Two samples are independent
Analysis of AUROC

- First guess: t-test!
- Close, not quite..
- What are the assumptions of the t-test?
First guess: t-test!
Close, not quite..

What are the assumptions of the t-test?

- Data are continuous
- Data are normally distributed
- Data are random samples
- Two populations have equal variance
- Two samples are independent
Our samples are not independent... They are identical!

We must factor in the covariance between our two AUCs:

- Jackknife – this is a method of deriving “independent” samples from your data & estimating bias & variance by LOO
- Non-parametric test which factors in the covariance
DeLong (1988) presented the application of earlier work to the AUROC – resulting in these lovely equations (V is the Mann Whitney Stat, L a suitable contrast matrix & theta is the AUC):

\[
\begin{align*}
S_{01}^{(s)} &= \frac{1}{m-1} \sum_{i=1}^{m} [V_{10}^{r}(X_i) - \hat{\theta}^{r}][V_{10}^{s}(X_i) - \hat{\theta}^{s}] \\
S_{01}^{(s)} &= \frac{1}{n-1} \sum_{j=1}^{n} [V_{01}^{r}(Y_j) - \hat{\theta}^{r}][V_{01}^{s}(Y_j) - \hat{\theta}^{s}].
\end{align*}
\]

\[
L\hat{\theta}^{r} - L\theta^{r} \over \left[ L \left( \frac{1}{m} S_{10} + \frac{1}{n} S_{01} \right) L' \right]^{1/2}
\]

But this looks familiar!

Boils down to the same p-value!
A gentle introduction to all that is AUROC can be found in:


See also:


- ... And the code in the appendix of this presentation
What are noise, data and information?

Some definitions:

- **Data**: alpha-numeric representations of sensor or hardware outputs. It may be (actually has to be) processed and filtered in some way in the hope of increasing the SNR. E.g. binary amplitude data from the ECG, or derived blood pressure values in mmHg.

- **Information**: processed data that has been turned into a label or actionable event - e.g. hypotension or VFIB.

- **Noise**: any sources of information present in your signal which reduce your ability to extract out the information in which you are interested.
  - Highly relative concept – one person’s noise is another’s data
  - We can add noise through signal processing as we saw ...

- **Artifact**: Often used interchangeably with noise, but I prefer to define it as something that results from a (mostly) deterministic change in the signal that distorts the original information. Examples include changes in amplitude and phase of a signal due to hardware or software filters.

- **Missing data**: Data (not information) which were somehow removed or the recording of them were lost (but does not include information which was obscured due to noise or artifact):
  - **Lost Data**: Data which entered the A/D and at some point were represented in an binary form (either in memory, on a serial bus or network feed, or on some more permanent media). This type of problem includes:
    - Bus overload
    - Processor-initiated changes in the data to adjust for asynchronicities / clock drift
    - Network errors / dropout
    - Data storage corruption
  - **Undersampled data**: Data which were never recorded and perhaps should have been recorded. This includes
    - **Temporal undersampling** (such as recording the ECG at 100Hz for QT studies, or a nurse recording blood pressure with a cuff every hour)
    - **Spatial undersampling** (not recording a channel of data that is needed, such as only having one precordial lead during ischemia studies, or not recording the O2 saturation for patients who you suspect have apnea).

- ... but noise & artifacts can be information too.
Power line interference: 50 ±0.2 Hz mains noise (or 60 Hz in many data sets) with an amplitude of up to 50% of full scale deflection (FSD), the peak-to-peak ECG amplitude

Sensor pop or contact noise: Loss of contact between the sensor and the skin manifesting as sharp changes with saturation at FSD for periods of around 1 s on the ECG (usually due to an electrode being nearly or completely pulled off);

Patient–sensor motion artifacts: Movement of the electrode away from the contact area on the skin, leading to variations in the impedance between the electrode and skin causing potential variations in the ECG and usually manifesting themselves as rapid (but continuous) baseline jumps or complete saturation for up to 0.5 second;

Electromyographic (EMG) noise: Electrical activity due to muscle contractions lasting around 50 ms between dc and 10,000 Hz with an average amplitude of 10% FSD level;

Baseline drift: E.g. respiratory motion with an amplitude of ~15% FSD at frequencies drifting between 0.15-0.3 Hz;

Hardware electronics noise: Artifacts generated by the signal processing hardware, such as signal saturation;

Electrosurgical noise: Noise generated by other medical equipment present in the patient care environment a frequencies between 100 kHz and 1 MHz, lasting for approximately 1 and 10 seconds; - may include defibrillation artifact too.

Quantization noise: Steps introduced into data

Clock drift & missing data: Sampling frequency is not constant – always use a real-time OS

Aliasing: Spurious frequencies because sampling frequency is too low or data were resampled

Signal processing artifacts: (e.g., Gibbs oscillations, IIR filters, ).

Other Biological sources & sinks: (e.g., non-conductive tissues, fetal/maternal mixture, observer pulse).
How to measures noise? SQI

**Absolute:**
- Route mean square (RMS) power
  - in a given temporal region (e.g. isoelectric region on ECG)
  - in entire signal
  - in a given frequency band
  - after application of a filter
- Higher order statistical descriptors (skew, kurtosis, entropy)

**Relative:**
- **Ratio of the peak signal amplitude to the noise amplitude** in a quiescent temporal region (e.g. isoelectric)
- **Crest factor** / peak-to-RMS ratio
  (the ratio of the peak value of a signal to its RMS value);
- **Power Ratio:** $\Sigma$in-band (e.g. 5 to 40 Hz in ECG)/ $\Sigma$total power
- Percentage agreement with another signal
  - Cross correlation
  - Cross spectral coherence
  - Mutual information
SQ example – the ECG

- Inter-channel comparison
Filters allow you to pick out ‘interesting’ parts of the data.

You then threshold on the filtered data to find the events.

Need to tune thresholds using databases:
- Build a ROC Curve and pick an operating point.

Need to be careful – noise looks like your events!
Now the practical ...

- Today you will design and implement your own QRS detector. Try:
  - Building conditioning filters, then
  - Amplitude thresholding
  - Gradient thresholding
  - Energy thresholding
First ... reading assignment

- Sketch algorithm sequence
- Properties of Pan, Hamilton & Tompkins detector
How does the energy-thresholding algorithm work?
How does the energy-thresholding algorithm work?

Pan, Hamilton & Tompkins Algorithm

Graphs showing different transformations over time (s): BPF, 1st diff, square, integrate.
P&T detection example

raw ECG (blue) and zero-phase FIR filtered ECG (red)

Integrated data with scan boundaries over scaled ECG

ECG with R-peaks (black) and S-points (green) over ECG

ECG+R+S

RR intervals

RR (s)
Properties of Pan, Hamilton & Tompkins detector
Properties of Pan, Hamilton & Tompkins detector

- IIR filter – keep only QRS frequency
- Nonlinear filters: Differentiate, square, integrate to isolate QRS complexes
- Threshold on energy packets
- FIR filter for peak location and searching
- Refactory blanking
- Delays
  - Filter delays
  - Scan back
... now the lab -

- Build conditioning filters, then
- Amplitude thresholding
- Gradient thresholding
- Energy thresholding
References

- **Daniel B. Neill & Weng-Keen Wong**, Tutorial T7 on Event Detection, 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Paris, France, June 28th - July 1st 2009 (several slides based on this PPT)

- Activity Monitoring Operating Characteristic (AMOC) curve (Fawcett and Provost 1999)

- Control Chart: (Shewhart 1931)

- Exponentially Weighted Moving Average (Roberts 1959)


- QRS detector papers – (Pan & Tompkins 1985), (Friesen et al 1990).
Significance for ROC:

```python
for r=1:K % For each X/Y column pair
    % compare 0s to 1s
    for i=1:m
        phi1=sum(gt(X(i,r),Y(:,r))); % Xi>Y
        phi2=sum(eq(X(i,r),Y(:,r))); % Xi=Y
        V10(i,r)=(phi1+phi2*0.5)/n;
        theta(r)=theta(r)+phi1+phi2*0.5;
    end
    theta(r)=theta(r)/(n*m);
    for j=1:n
        phi1=gt(X(:,r),Y(j,r)); % X>Yj
        phi2=eq(X(:,r),Y(j,r)); % X=Yj
        V01(j,r)=(sum(phi1)+sum(phi2)*0.5)/m;
    end
end

% Calculate S01 and S10, covariance matrices of V01 and V10
S01 = (1/(n-1))*((V01'*V01)-n*(theta'*theta));
S10 = (1/(m-1))*((V10'*V10)-m*(theta'*theta));

% Combine for S, covariance matrix of theta
S = (1/m)*S10 + (1/n)*S01;
```
E.g. – Maternal/fetal ECG overlap in time

- Maternal ECG is much larger in amplitude
- Maternal and fetal ECG overlap in time domain
- Maternal features are broader, but
- Fetal ECG is in-band of maternal ECG (they overlap in freq domain)
- 5 second window … Maternal HR=72 bpm / Fetal HR = 156 bpm
E.g. – Maternal/fetal ECG overlap in time

Fetal / Maternal Mixture

Maternal

Noise

Fetal
E.g. – Maternal/fetal ECG overlap in frequency

MECG & FECG spectral properties

Fetal QRS power region
Fetal ECG example

Transmission properties of the Fetal ECG

Schematic representation of the media through which the current generating the fetal ECG is conducted to the surface of the maternal abdomen at approximately 30 weeks.

ME: Maternal Epidermis;
SF: Subcutaneous Fat;
FS: Fascia Superficialis;
MR: Musculus Rectus Abdominalis;
FP: Fascia Profunda; P: Peritoneum;
EG: Electrode Gel: E: Electrode;
S: Serosa; M: Myometrium;
D: Decidua; C: Chorion;
VC: Vernix Caseosa; A: Amnion;
FE: Fetal Epidermis;
MB: Maternal Bladder.