Prediction of radar NCTR performance using Mutual Information

Mini AOC Conference
Simon's Town
23 May 2019

Presented by:
Jacques Cilliers
Overview

• Interaction of Target and Electromagnetic (EM) Wave
• Information theory
  – Introductory concepts
  – Mutual information (MI)
  – MI for radar
  – Fano’s Inequality (Relate P(error) to MI)
  – MI derivation for discrete input continuous output channel
• Example Results
  – F14, F15, F16 Comparison
  – F18, F35 Polarization Comparison
  – F14, F15, F16 Comparison: coherent vs. non-coherent
  – Boeing 707 (1:25) Measurements, Polarization Comparison
• Conclusions
Introduction

• Need a technique to “design” a completely new NCTR recognition function
  – Take into account the performance of and Interaction between:
    • Radar sensor, Target, Signal processing algorithms
  – Need new theory that allows for the analysis and comparison
    of disparate radar system conceptual design
  – Bound on absolute maximum performance achievable to test NCTR

• Information Theory based approach
  – Information Theory deals with fundamental limits on performance
  – Mainly used in communications - little prior art for NCTR

• Demonstrate that information theory, and specifically the concept of mutual information (MI) can improve insight during NCTR designs
• Gain scientific insight into information theory methods as applied to radar target recognition
• Gain insight into interpretation of information theoretic results for radar NCTR problems
• Standard approach
  – Use Feature Extraction followed by Classifier
  – Single confusion matrix, badly defined SNR

<table>
<thead>
<tr>
<th>Aircraft</th>
<th>F-15</th>
<th>T-38A</th>
<th>Falcon-20</th>
<th>Falcon-100</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-15</td>
<td>51</td>
<td>41</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>T-38A</td>
<td>40</td>
<td>53</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Falcon-20</td>
<td>0</td>
<td>0</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>Falcon-100</td>
<td>6</td>
<td>3</td>
<td>12</td>
<td>79</td>
</tr>
</tbody>
</table>

  – Adversarial Examples (Next slide)
  – Often researchers try to make everything an image and then use standard image classification techniques

• Information Theory approach
  – No Feature extraction
  – No Classifier
  – Estimates the best possible recognition performance given
    • Targets and materials
    • Target geometries
    • Radar waveforms
  – Graph of P(error) vs SNR
Adversarial Examples

- Attack on a classifier
  - What is the least interference which can be added to change the output to force an error

\[ x + \epsilon \text{sign}(\nabla_x J(\theta, x, y)) = \]

- \( x \)  
  - “panda”  
  - 57.7% confidence

- \( \text{sign}(\nabla_x J(\theta, x, y)) \)  
  - “nematode”  
  - 8.2% confidence

- \( x + \epsilon \text{sign}(\nabla_x J(\theta, x, y)) \)  
  - “gibbon”  
  - 99.3% confidence
Adversarial Examples

- Examples: Noise, single pixel, 3-D printed
Information theory mathematically formalizes the relatively vague concepts of a “message” and the amount of “information” the message contains.

Shannon 1948
- “The mathematical theory of communication” - Predicts
  - Maximum compression of a source
  - Maximum data rate over a channel
- Unfortunately he doesn’t say how to achieve it

**Communication:**
- **Source** → **Channel** → **Receiver**

**Radar:**
- Important distinction – mind-set change
  - The radar TX only controls the SNR and the illumination of the target via its waveform
  - Not the information content

**Information content is:**
- Function of geometry of the target
- And the interaction of the target with EM energy
Scattering mechanisms

EM Theorems
- Uniqueness – only one solution
- Superposition – consider sources in isolation
- Linearity - Maxwell’s equations are linear (linear medium)
  - Linear system theory (time and frequency response) is valid
- Scaling of targets
  - Reducing target requires increased conductivity

Target – EM Field Interaction
The radar problem can thus be analysed as a non-optimal communication system.

The factor which limits the performance of a communication system is the amount of mutual information between the transmitted signal set and the received signal set.

To analyse the radar problem the amount of mutual information between all possible target waveforms (responses) and the waveform(s) received at the radar by its receiver has to be estimated.


**Information Theory – Self Information**

- **Self information**
  - Measure of “surprise” when observing the output of a random variable (RV)
  - Message length required to transmit the specific outcome of a RV

  \[ I(x) = -\log p(x), \quad p(x) \in [0,1] \]

  - Base of log() is usually 2
  - Then information is measured in bits

- **Examples:**

<table>
<thead>
<tr>
<th>Example</th>
<th>Probability</th>
<th>Self Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coin</td>
<td>p = (\frac{1}{2})</td>
<td>(H(x) = 1) bit</td>
</tr>
<tr>
<td>Dice</td>
<td>p = (\frac{1}{6})</td>
<td>(H(x) = 2.59) bits</td>
</tr>
<tr>
<td>Lottery</td>
<td>p = (\frac{1}{10,068,347,520})</td>
<td>(H(x) = 33.23) bits</td>
</tr>
<tr>
<td></td>
<td>p = 0.99999999999900679</td>
<td>(H(x) = 1.433\times10^{-10}) bits</td>
</tr>
</tbody>
</table>

- **What happens when p = 0?**

  Standard information theory approach:

  \[ \lim_{p \to 0^+} p \log(p) = 0 \]
Information Theory - Entropy

- Average “self information” over all outputs of RV

\[ H(X) = E\{I(x)\} = - \sum_{x \in R_x} p(x) \log(p(x)) \]

\[ \lim_{p \to 0^+} p \log(p) = 0 \]

- Entropy is the average number of bits required to transmit the result of the output of the RV

- Properties:
  - \( H(X) \geq 0 \)
  - \( H(X) \leq \log_2(N) \)
  - \( H(X) = \log_2(N) \) when \( p_k(x) = \frac{1}{N} \) for all \( k = 1..N \)

- Example
  - RV with two outcomes: \( H(X) = -\frac{1}{4} \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right) \)
  - Eg. Biased coin

\[ = -\frac{1}{4}(-2) - \frac{3}{4}(-0.415) \]
\[ = 0.5 + 0.3113 = 0.8113 \text{ bits} \]
Information Theory - Entropy

- Definition:
  \[ H(X) = E\{I(x)\} = - \sum_{x \in \mathcal{X}} p(x) \log(p(x)) \]
  \[ \lim_{p \to 0^+} p \log(p) = 0 \]

- Example – 2 output Random Variable:

  \[ p_1 = p \quad \quad p_2 = 1 - p \]

  \[ H(X) = -p_1 \log_2(p_1) - p_2 \log_2(p_2) \]
  \[ = -p \log_2(p) - (1 - p) \log_2(1 - p) \]

  \[ H(X) = -\frac{1}{4} \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right) \]
  \[ = -\frac{1}{4}(-2) - \frac{3}{4}(-0.415) \]
  \[ = 0.5 + 0.3113 = 0.8113 \text{ bits} \]
Information Theory – Joint & Conditional Entropy

- **Joint Entropy**
  \[ H(X,Y) = E\{I(X,Y)\} = - \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log(p(x, y)) \]

- Measure of the uncertainty in a set of RV’s.
  - \( \leq \) sum individual entropies
  - \( \geq \) max \( (H(X), H(Y)) \)

- **Conditional Entropy**
  \[ H(X | Y) = E\{H(X | Y = y)\} = - \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log(p(x | y)) \]

- Amount of information needed to describe \( X \), given \( Y \) has occurred
  - OR: Additional bits needed to communicate \( X \) given that both parties know \( Y \).
  - Note: conditioning always reduces entropy
• Mutual Information

\[ I(X;Y) = H(X) - H(X|Y) = I(Y;X) \]

\[ = \sum_{x \in R_x} \sum_{y \in R_y} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \]

– Reduction in a-priori uncertainty in \( X \)

– Amount of information one RV contains about another RV
Information Theory – Properties of MI

\[ I(X;Y) = \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \]

- Always positive \( I(X;Y) \geq 0 \)
- Zero if and only if \( X \) and \( Y \) are statistically independent
  \[ I(X;Y) = 0 \quad \text{iff} \quad P(X,Y) = P(X)P(Y) \]
- More reliable measure of independence than correlation
- Maximum number of partitions of \( X \) based on observation of \( Y \)
  \[ N = \left\lfloor 2^{I(X,Y)} \right\rfloor \]
- Equivalent to maximum number of classes which a classifier can discern.
Information Theory – MI vs Correlation

• Correlation coefficient
  – 2\textsuperscript{nd} Order Statistic

\[\rho_{X,Y} = \frac{1}{\sigma_X \sigma_Y} \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]\]

\[= \frac{1}{\sigma_X \sigma_Y} \sum_{x,y} (x - \mu_X)(y - \mu_Y) P(X = x, Y = y)\]
Information Theory – MI vs Correlation Example 1

- Example: 64 equally probable points on unit circle

- Correlation = 0.0
- MI = \( \log\left(\frac{1/64}{(1/32 \times 1/32)}\right) \)
  = 4 bits
- \( H(Q) = \log(32) \)
  = 5 bits
- \( H(Q|I) = H(Q) - MI \)
  = 5 - 4
  = 1 bit
- \( H(I,Q) = \log(64) \)
  = 6 bits
- \( MI = H(I) + H(Q) - H(I,Q) \)
  = 5 + 5 - 6 = 4 bits

- Meaning of 1 bit?
Example: 64 equally probable points on Lissajous curve

- Correlation = 0.0
- MI = \( \log\left(\frac{1/64}{(1/32 \times 1/16)}\right) \)
  = 3 bits
- \( H(Q) = \log(32) = 5 \) bits
- \( H(I) = \log(16) = 4 \) bits
- \( H(Q|I) = H(Q) - MI \)
  = 5 - 3 = 2 bits
- \( H(I|Q) = H(I) - MI \)
  = 4 - 3 = 1 bit
- \( H(I,Q) = \log(64) \)
  = 6 bits
- MI = \( H(I) + H(Q) - H(I,Q) \)
  = 5 + 4 - 6 = 3 bits
Information Theory – Channel Capacity

- **Channel capacity:** 
  \[
  \max_{p(x)} I(X_N; Y_N) = WT \log_2 \left( 1 + \frac{P_{av}}{WN_0} \right) \quad C = W \log_2 \left( 1 + \frac{P_{av}}{WN_0} \right)
  \]

- **Limits**
  - SNR \( \rightarrow \infty \) then \( C \rightarrow \infty \)
  - BW \( \rightarrow \infty \) then:
    \[
    C_\infty = \frac{P_{av}}{N_0} \log_2 (e) = \frac{P_{av}}{N_0 \ln(2)}
    \]
    Increasing bandwidth doesn’t continue adding information!

- **Shannon’s coding theorem:**
  As long as rate is less than channel capacity, then
  \( P(\text{error}) \) can be made **arbitrarily** small

- **NOT possible in radar!**
  - Can’t control the target’s waveform
• Fano’s inequality – lower bound on P(error)

\[ P(\hat{X}(y) \neq X) \geq \frac{H(X) - I(X;Y) - 1}{\log_2(N)} \]

• Data processing inequality

\[ X \rightarrow Y \rightarrow Z \quad I(X;Y) \geq I(X;Z) \]

– No function or algorithm \( Z = f(Y) \) can increase information content of \( Y \)

• Discrete input continuous output MI:

\[ I(X,Y) = \sum_{i=1}^{N} \int_{-\infty}^{\infty} P(x_i) p(y|x_i) \log_2 \left( \frac{p(y|x_i)}{p(y)} \right) dy \]
Information Theory – Fano’s Inequality

- Strictest version of Fano’s inequality – lower bound on P(error)
  \[ H(P_e) + P_e \log_2 (N - 1) \geq H(X | Y) \]

- H(Pe) – Entropy of RV of error event

- Derivation to relate \( P_{\text{error}} \) to MI
  \[ H(P_e) + P_e \log_2 (N - 1) \geq H(X | Y) \]
  \[ I(X;Y) \leq H(X) - H(P_e) - P_e \log_2 (N - 1) \]
  \[ = f(P_e), \]
  \[ P_e \leq f^{-1}(I(X;Y)) \]
MI Calculation
– Discrete input, continuous output channel

• Starting from:

\[
I(X,Y) = \sum_{i=1}^{N} \int_{-\infty}^{\infty} P(x_i) p(y|x_i) \log_2 \left( \frac{p(y|x_i)}{p(y)} \right) dy
\]

• N-Dimensional MI in Gaussian noise:

\[
I(a_k, y) = \log_2(N) - \frac{1}{N} \sum_{k=0}^{N-1} E \left\{ \log_2 \left( \sum_{i=0}^{N-1} \exp \left( -\frac{|a_k + z - a_i|^2 - |z|^2}{2\sigma^2} \right) \right) \right\}
\]

\[
\text{SNR} = \frac{E\{|a_n|^2\}}{E\{|z_n|^2\}} = E\{|a_n|^2\}/D\sigma^2
\]

• N-Dimensional MI in Gaussian noise, amplitude only:

\[
s_{ij} = \sqrt{a_{ij}^2 + a_{Qij}^2}
\]

\[
s_{kj} = \sqrt{a_{kj}^2 + a_{Qkj}^2}
\]

\[
r_{kj} = \sqrt{(a_{kj} + z_{kj})^2 + (a_{Qkj} + z_{Qkj})^2}
\]
• Why can’t we just use a N-Dimensional histogram?
  – Largest addressable element in MATLAB:
    • $2^{48} - 1 = 2.8\times10^{14}$
    • This is approximately 2 million GBytes of RAM
  – Say 20 bins per dimension
  – And 100 dimensions
  – This gives $20^{100} = 1.2677\times10^{130}$ bins in the histogram
  – This is approximately $9.44\times10^{121}$ Gbytes of RAM
  – Still out by a factor of $4.7\times10^{115}$
  – This is about 300 years by Moore’s law

  – Currently highest dimension problem was $N = 4002$

  – For now, HAVE TO do the maths!
MI Calculation - Validation

- Low dimensional comms results
- High dimensional Gaussian
MI Calculation - Validation

- MFSK, paper by NASA (Butman 1973)
  - Application: Planetary lander in dense turbulent dispersive atmosphere
  - E.g. Venus, Jupiter, Saturn
MI Calculation – F14, F15, F16

Geometry (F15, shg)

Geometry (F14, shg)
Calculation of High Range Resolution Profiles (HRRP)

Span: 22 m x 5 m
Pixels: 440 x 100
Resolution cell: 5 cm x 5 cm
Example Result: MI for 1024 Waveforms (HRR Profiles) F14, F15, F16

<table>
<thead>
<tr>
<th></th>
<th>Capacity Bound</th>
<th>Gauss 1024 waveforms</th>
<th>F16</th>
<th>F14</th>
<th>F15</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR [dB]</td>
<td>-10.5</td>
<td>-5.2</td>
<td>19.3</td>
<td>21.0</td>
<td>24.0</td>
</tr>
</tbody>
</table>

25 to 30 dB Worse than optimal
Example Result: Effect of Multiple Targets on $P(\text{error})$

Convert MI to $P(\text{error})$ using Fano’s inequality

Fano’s inequality relates MI and $P(\text{error})$ for a given scenario

As you add more HRR profiles to the set to be classified, the noise induced requires a higher SNR to achieve the same $P(\text{error})$

(Curse of Dimensionality at work !)

| Confusion matrix for straight-and-level trajectory #1 with noise |
|-------------------|--------|-----------|--------|--------|
| Aircraft          | F-15   | T-38A     | Falcon-20 | Falcon-100 |
| F-15              | 51     | 41        | 0       | 8       |
| T-38A             | 40     | 53        | 0       | 7       |
| Falcon-20         | 0      | 0         | 90      | 10      |
| Falcon-100        | 6      | 3         | 12      | 79      |
Example Result: Effect of Multiple Targets on $P(\text{error})$

Low $P(\text{error})$ Example

$P(\text{error})$
$P(\text{incorrect identification})$

Convert MI to $P(\text{error})$ using Fano’s inequality

Fano’s inequality relates MI and $P(\text{error})$ for a given scenario

Confusion matrix for straight-and-level trajectory #1 with noise

<table>
<thead>
<tr>
<th>Aircraft</th>
<th>F-15</th>
<th>T-38A</th>
<th>Falcon-20</th>
<th>Falcon-100</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-15</td>
<td>51</td>
<td>41</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>T-38A</td>
<td>40</td>
<td>53</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Falcon-20</td>
<td>0</td>
<td>0</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>Falcon-100</td>
<td>6</td>
<td>3</td>
<td>12</td>
<td>79</td>
</tr>
</tbody>
</table>
F15: Effect of oversampled HRR (in azimuth) and using knowledge (estimate) of aspect angle

- **F15**
  - Length = 19.43 m
  - Wingspan = 13.05 m
  
  Az sampling (req)
  - Broadside: 0.011°
  - Head on: 0.016°

- Az step = 0.01°
- 36,000 waveforms
- Max MI = 15.136 bits

Effect of “known” approximate heading 10 dB gain!
Effect of Carrier Frequency

- WB (480 MHz), 2.5 GHz, 10 GHz, 17.5 GHz

- Higher frequency (17.5 GHz)
  - Looking for more randomness due to shorter wavelength
  - BUT: Lower information !! (~ 1 dB loss in SNR)

- Lower frequency (2.5 GHz)
  - More information !! (~ 2-3 dB gain in SNR)

- Scattering mechanisms persist over wider angles

- Percentage BW
  - 19.2 %
  - 4.8 %
  - 2.74 %
• WB (480 MHz) vs. UWB (8 GHz and 16)
Example Result: Use of Polarisation

- Single polarization at 116 frequencies – $116 \times 2$ (IQ) = 232 Dimensional integral
- \([VV \ HH \ 2VH]\) at 116 frequencies – 696 Dimensional integral (Accuracy : 0.01 bits)
Result Non-coherent processing, F14, F15, F16

- Coherent
- Non-coherent (Envelope only)

Loss ≈ 10 dB
Result Non-coherent processing, F14, F15, F16

Coherent

Non-coherent (Envelope only)

Loss ≈ 10 dB
Boeing 707 Scale Model (1:25)

Measured frequency: 2-18 GHz
Scaled frequency: 80 – 720 MHz
Azimuth: 1799 step (0.2°)

Frequncies: 2001
Dual Pol MI: 2001 x 2 (pol) x 2 (I,Q) = 8004 Dimensional Integral
Conclusions

- Developed a method to apply Mutual Information to predict NCTR performance
- Now allows identification performance to be estimated
  - P(error) vs SNR
- Identification performance comparisons and trade offs were carried out for wide range of radar parameters and target types
- Brings new insight to allow system level decisions
  - Especially During the design phase of a radar
- Might help to avoid classification steps often used in existing NCTR techniques which might destroy potentially useful information
- New insight gained into the MI concept and interpretation of results
- Can’t add “unknown” target into this theory, yet.
- Opened a new set of questions

J.E. Cilliers, Information Theoretic Limits on Non-cooperative Airborne Target Recognition by Means of Radar Sensors
Available from: http://discovery.ucl.ac.uk/10049414/