Metrics in VoicePrivacy & ASVspoof Challenges

Andreas Nautsch

vitas.ai

2021-08-02

Disclaimer: as always, you will see the individual view of an opinionated researcher; not the view(s) of associated institutions :)
Outline

- Voice biometrics in a nutshell
- Security & privacy focus
- ASVspoof challenge: “t-DCF” metric $\Rightarrow$ security in voice biometrics
- VoicePrivacy challenge: “ZEBRA” metric $\Rightarrow$ privacy as ANTI voice biometrics
Biometrics with voice: WHO is speaking?

Same person?
Biometrics with voice: WHO is speaking?

Audio from before

Enrolment

Database of voice references

⇒ allows to link users over time/acts

- User knows
- Unbeknownst to user(s)

Claim

Audio of a user

Verification

Voice comparison

Score

“Automatic speaker verification” (ASV)

same person
different person
decision with consequence(s)
What’s actually going on here? … think forensic sciences

Evidence collection ➔ Evidence reporting ➔ Decision making ➔ Impact

Do we observe truth? Are we fooled?
Do we help? Is this useful?

This can be formalised: Bayesian decision theory

Can another one make better decisions, actually?!
Smells like costs?!
Security & privacy — voice biometrics two-ways

Focus: common evaluation methodology to the assessment of …
   a) Security
   b) Privacy

Evidence collection → Evidence reporting → Decision making

Spoofing detection
ASVspoof challenge
- Physical access
- Logical access
- Speech deepfakes

Zero evidence
VoicePrivacy challenge
- What was said? ✓
- Who spoke? ☒
⇒ Modify raw audio

Impact
model to …
… evaluate

Focus here:
HOW to evaluate :)

What was said?
Who spoke?
Modify raw audio
Why security?

E.g. Fraud detection in online banking

- HSBC refers to £249 mio saved through voice biometrics
- Attacking voice biometrics is possible
- Needs to be prevented

Why privacy?

E.g. surveillance through speech data

- Enabling human rights for individuals
- Bad example: GDR Ministry for State Security German (Stasi)
- Needs to be prevented
Automatic Speaker Verification
Anti-Spoofing
(ASVspoof)


DOI: 10.1109/TASLP.2020.3009494

ASVspoof metric: tandem detection cost function (t-DCF)

- Cascaded system design
  - ASV is given
  - Countermeasure (CM) ⇒ add-on security

- ASV classification task
  - target vs. nontarget

- CM classification task
  - non-/target vs. spoof

- Evaluation: overall expected operational cost from employing ASV & CM
Looking glass: Bayesian decision theory

How often do these happen?

90% ?
8% ?
2% ?

Prior probabilities
⇒ subjective quantification

“policy” is trading-off beliefs

Risk costs
⇒ subjective quantification

“policy” is trading-off beliefs

This can be formalised: Bayesian decision theory
t-DCF: at a glance

\[
t-DCF = C_{\text{miss}} \cdot \pi_{\text{tar}} \cdot P_a + C_{\text{fa}} \cdot \pi_{\text{non}} \cdot P_b + C_{\text{fa,spoof}} \cdot \pi_{\text{spoof}} \cdot P_c + C_{\text{miss}} \cdot \pi_{\text{tar}} \cdot P_d
\]

**Tandem action**
- (CM ACCEPT, ASV REJECT)
- (CM ACCEPT, ASV ACCEPT)
- (CM ACCEPT, ASV ACCEPT)
- (CM REJECT)

**Actual class**
- Target
- Nontarget
- Spoof
- Target

\[
P_c(\tau_{\text{cm}}, \tau_{\text{asv}}) = P_{\text{fa}}^{\text{cm}}(\tau_{\text{cm}}) \times P_{\text{fa,spoof}}^{\text{asv}}(\tau_{\text{asv}})
\]
// fa: score ≥ threshold → P(...)

\[
P_b(\tau_{\text{cm}}, \tau_{\text{asv}}) = (1 - P_{\text{miss}}^{\text{cm}}(\tau_{\text{cm}})) \times P_{\text{fa}}^{\text{asv}}(\tau_{\text{asv}})
\]
// miss: score < threshold → P(...)

\[
P_a(\tau_{\text{cm}}, \tau_{\text{asv}}) = (1 - P_{\text{miss}}^{\text{cm}}(\tau_{\text{cm}})) \times P_{\text{miss}}^{\text{asv}}(\tau_{\text{asv}})
\]

\[
P_d(\tau_{\text{cm}}, \tau_{\text{asv}}) = P_{\text{miss}}^{\text{cm}}(\tau_{\text{cm}})
\]
// miss: score < threshold → P(...)

---

**Diagram Image**: A tandem detection system with CM and ASV nodes, illustrating the decision process.
How to compare t-DCFs of different priors/costs?

- Default: simulate coin tossing performance!
- Playing through the extrema…
  - CM & ASV: all-pass
    \[ C_{fa} \cdot \pi_{non} \cdot 1 + C_{fa,\text{spoof}} \cdot \pi_{\text{spoof}} \cdot 1 \]
  - CM: no-pass
    \[ C_{\text{miss}} \cdot \pi_{\text{tar}} \cdot 1 \]
  - CM: all-pass & ASV: no-pass
    \[ C_{\text{miss}} \cdot \pi_{\text{tar}} \cdot 1 \]

\[
t-\text{DCF}'(\tau_{cm}, \tau_{asv}) = \frac{t-\text{DCF}(\tau_{cm}, \tau_{asv})}{t-\text{DCF}_{\text{default}}} \\
t-\text{DCF}_{\text{min}} = \frac{t-\text{DCF}_{\text{min}}}{t-\text{DCF}_{\text{default}}} \leq \frac{t-\text{DCF}_{\text{min}}}{t-\text{DCF}_{\text{min}}} = 1
\]

\[
t-\text{DCF}_{\text{default}} = \min \{ C_{fa} \cdot \pi_{non} + C_{fa,\text{spoof}} \cdot \pi_{\text{spoof}}; C_{\text{miss}} \cdot \pi_{\text{tar}} \}
\]
Synthetic scores; parameters as of ASVspoof 2019/21

Unconstrained t-DCF

![Graph showing the unconstrained t-DCF with different values of \( \tau_{\text{asv}} \).]

ASV-constrained t-DCF

![Graph showing the ASV-constrained t-DCF with different values of \( \tau_{\text{asv}} \).]

\[
\text{t-DCF}(\tau_{\text{cm}}) = C_0 + C_1 P_{\text{cm}}^{\text{miss}}(\tau_{\text{cm}}) + C_2 P_{\text{fa}}^{\text{cm}}(\tau_{\text{cm}})
\]

\[
C_0 = \pi_{\text{tar}} C_{\text{miss}} P_{\text{miss}}^{\text{asv}} + \pi_{\text{non}} C_{\text{fa}} P_{\text{fa}}^{\text{asv}}
\]

\[
C_1 = \pi_{\text{tar}} C_{\text{miss}} - (\pi_{\text{tar}} C_{\text{miss}} P_{\text{miss}}^{\text{asv}} + \pi_{\text{non}} C_{\text{fa}} P_{\text{fa}}^{\text{asv}})
\]

\[
C_2 = \pi_{\text{spoo}} C_{\text{fa,spoo}} P_{\text{fa,spoo}}^{\text{asv}}
\]
VoicePrivacy


DOI: 10.21437/Interspeech.2020-1815


Picture taken in Heidelberg Zoo, 2020
Motivation: evidence in court & decoupled provinces

Evidence collection → Evidence reporting → Decision making → Impact

Forensic practitioner → Judge / jury

Modeling impact to decision making → Cost policy remains unknowable

Evaluation: relative information gain

Realm of privacy preservation

Safeguard → Protected data → Classifier → Scores → Decision policy → Action

Realm of the adversary

Error trade-offs → Inference information → Impact
Zero evidence “ZEBRA” framework: two metrics

- **Expected privacy disclosure**
  - Population level
  - Minimise empirical cross-entropy (ECE); regardless of prior probability

- **Worst-case privacy disclosure**
  - Individual level
  - Minimise strength-of-evidence; across data records
Textbook: empirical cross-entropy (ECE) — step by step

Prior uncertainty

Posterior uncertainty

Evidence

// P(s | θ) is a research topic

RELATIVE information gain

Expectation across scores

ECE(Θ | S) := \[ \frac{\pi}{|S_A|} \sum_{a \in S_A} \log_2 \left( 1 + \frac{1 - \pi}{\alpha \pi} \right) + \frac{1 - \pi}{|S_B|} \sum_{b \in S_B} \log_2 \left( 1 + \frac{b \pi}{1 - \pi} \right) \]

P: reference probability space

Θ = { A: "same person", B: "different person" }

\[ H_P(\Theta) = - \sum_{\theta \in \Theta} P(\theta) \log_2 P(\theta) \]

\[ H_P(\Theta | S) = - \sum_{\theta \in \Theta} P(\theta) \int_s P(s | \theta) \log_2 P(\theta | s) ds \]

Well-calibrated scores (strength-of-evidence)

"the classifier"

Integrate \( \pi \) out

Ramos, Franco Pedroso, Lozano-Diez, Gonzalez-Rodriguez: Deconstructing Cross-Entropy for Probabilistic Binary Classifiers, Entropy 20(3), 2018
Shannon’s perfect secrecy to strength-of-evidence

(a) ZEBRA idea

(b) ECE plot
On the highest strength-of-evidence

● Basic idea
  ○ Sustain probabilistic interpretation of scores
  ○ Account for binary decision setting  // P(“yes”) = 1 - P(“no”)
  ○ Take the highest strength-of-evidence
  ○ Keep in mind the world is larger than one dataset
    ⇒ apply Laplace’s rule of succession &
    return a prediction of the worst case disclosure

● Make reporting digestible, lessons from forensic sciences
  ○ Everyone interprets numbers & ratios differently
  ○ Thus: categorical tags & scale

<table>
<thead>
<tr>
<th>Tag</th>
<th>Category</th>
<th>Posterior odds ratio (flat prior)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( l = 1 = 10^0 )</td>
<td>50 : 50 (flat posterior)</td>
</tr>
<tr>
<td>A</td>
<td>( 10^0 \leq l &lt; 10^1 )</td>
<td>more disclosure than 50 : 50</td>
</tr>
<tr>
<td>B</td>
<td>( 10^1 \leq l &lt; 10^2 )</td>
<td>one wrong in 10 to 100</td>
</tr>
<tr>
<td>C</td>
<td>( 10^2 \leq l &lt; 10^4 )</td>
<td>one wrong in 100 to 1000</td>
</tr>
<tr>
<td>D</td>
<td>( 10^4 \leq l &lt; 10^5 )</td>
<td>one wrong in 10 000 to 100 000</td>
</tr>
<tr>
<td>E</td>
<td>( 10^5 \leq l &lt; 10^6 )</td>
<td>one wrong in 100 000 to 1 000 000</td>
</tr>
<tr>
<td>F</td>
<td>( 10^6 \leq l )</td>
<td>one wrong in at least 1 000 000</td>
</tr>
</tbody>
</table>
VoicePrivacy 2020 challenge; an example

Expected privacy disclosure (population)

Worst-case privacy disclosure (individual)

Categorical tag

- Perfect privacy (0.00, 0.00, 0)
- B1 (0.11, 2.27, C)
- B2 (0.36, 3.58, C)
- Unprotected data (0.58, 3.98, C)

- … coin-tossing
- … a DNN baseline
- … signal processing (on-the-fly)
- … the x-vector DNN (SOTA)
Wrapping up …
Summary

● One framework, two application spaces
  ⇒ Bayesian decision theory

● Security focus: ASVspoof challenge
  ⇒ tandem detection cost function (t-DCF)

● Privacy focus: VoicePrivacy challenge
  ⇒ zero evidence “ZEBRA” with expectation & worst-case metrics
Take-home message(s)

- Towards a holistic approach
  - Think interdisciplinary for solutions
  - Develop multidisciplinary skills

- Expectation is not the sole metric
  - One might not know all parameters all the time — theory & models are indispensable
  - Consider the worst-case — avoid running into marginalising societies

- “Privacy as anti-biometrics”
  ⇒ we need more conversations across fields :)
    // there’s so much more in speech than biometrics alone