Friedrich-Alexander-Universität
Erlangen-Nürnberg

Evaluative Assessment of Digital Evidence

Establishing Bayesian Reasoning in Digital Forensic Science

Jan Gruber

Motivation

Evaluative Analysis in *Traditional* Forensic Science

- Reliability and uncertainty estimates are of *great importance*
- *Statistical* measurements, *Bayesian* reasoning, and *evaluative* reporting are standards
- Use of likelihood ratio approach is ***advised by ENFSI***

. . . in *Digital* Forensic Science

- *Rarely* structured uncertainty estimates!
- Likelihood ratio approach *only* applied in niche fields

→ *How to adhere* to the **high standards** of forensic science?

Background & Related Work

A Primer on Likelihood Ratios

- Likelihood Ratios (LRs) are numerical measures of the evidential value

$$\mathrm{LR}_{h_p,h_d}(E) = \frac{P(E \mid h_p)}{P(E \mid h_d)}$$

- The trier of fact makes a decision based on Bayes' theorem by updating:

$$\text{prior odds} \times \mathrm{LR} = \text{posterior odds}$$

Usage of Likelihood Ratios in DF

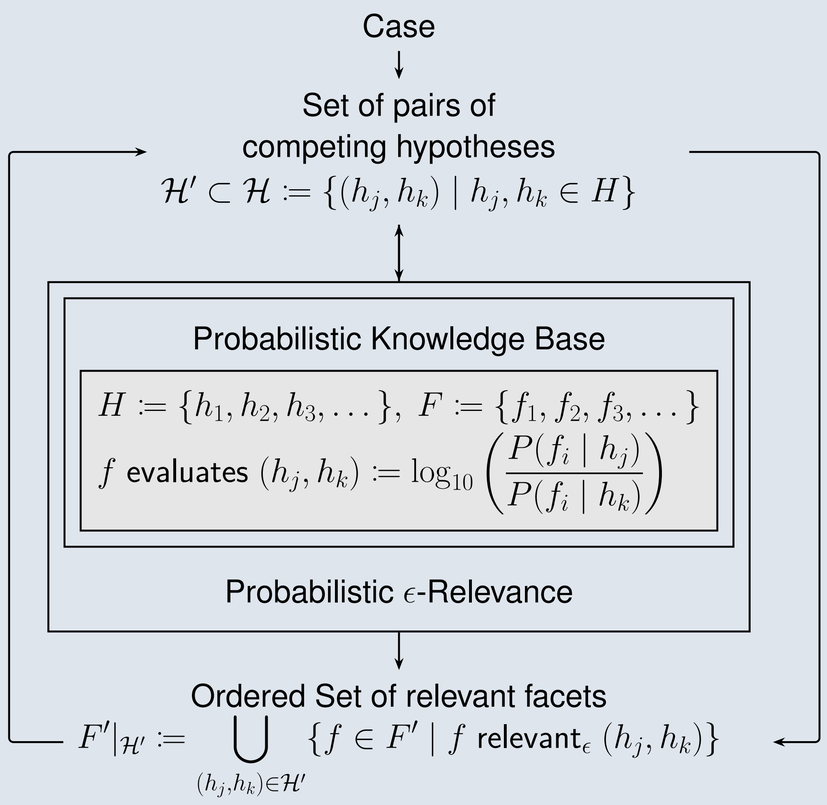
- Mostly used for **"measurements" from analog sensors**, e.g.,
  - noise patterns for camera identification (Nordgaard/Höglund, 2011; Van Houten et al., 2011)
  - geolocation data stored on smartphones (Spichiger, 2023)
- Also used when dealing with **similarity measures** for. . .
  - biometric face recognition (Macarulla et al., 2020)
  - authorship attribution of text (Ishihara, 2021)

The Probabilistic Cyber-traceological Model

Overview

Based on previous publications, we propose a **formalized method to identify relevant traces** in DF, which is able to deal *with uncertainty*:

- The *investigative knowledge base*:
  - a set of facets F
  - a set of hypotheses H
  - a function evaluates $\coloneqq F \times (H \times H) \to \mathbb{R}$
- Computation of *relevant digital evidence* for a pair of hypotheses (h_j, h_k)



- Uses Bayesian reasoning based on *observable likelihoods* $P(f_i \mid h_k)$
- Creates *awareness of uncertainty*
- *Relevance assessment* based on the weight of evidence $\lessgtr \epsilon$

Outlook

Future directions?

The big question is *how to instantiate* the evaluates function +

- *Which factors* influence uncertainty?
- How to *build probabilistic models* for digital traces using the identified factors?
- Are there types of evidence for which the LR method is especially *easy or hard to apply*?

Conclusion

The *Cyber-traceological Model* is a promising basis to solidify DF by using Bayesian reasoning:

- Uncertainty estimates are vital
- First forays for specific and confined classes of digital traces
- Formal description is a starting point for further explorations

→ **Any ideas? Collaborate with us!**

jan.gruber@fau.de