Optimal Weighting (OW) Modell – Theoretische Erklärung

Grundlagen

Das **Optimal Weighting (OW) Modell** ist ein theoretisches Modell, das beschreibt, wie ein Mensch oder ein System die Hinweise von einem menschlichen Entscheider und eines automatisierten Hilfsmittels optimal kombiniert, um die beste mögliche Entscheidung zu treffen.

Grundprinzip des OW-Modells

- Annahme: Mensch und automatisches System als **stochastisch unabhängige Signaldetektoren** – enthalten jeweils eine Schätzung für die Wahrscheinlichkeit, dass ein Reiz ein Signal enthält
- beide bewerten die Reize anhand eigener Sensitivität: *d'* (d-prime)

Gewichtung der Hinweise

- **Ziel:** Die Hinweise des Menschen und des automatischen Hilfsmittels werden so gewichtet, dass die Teamleistung *maximal* ist.
- **Vorgehen:** Die Schätzung ist eine gewichtete Summe der beiden Hinweise:

$$Z = \Sigma\, a_i X_i$$

$$\rightarrow\ Z = a_{human} \times X_{human} + a_{aid} \times X_{aid}$$

➔ X_i – Einschätzung (Likelihood) für den i-Agenten

➔ a_i – Gewicht, das auf den jeweiligen Agenten gesetzt wird

Bestimmung der optimalen Gewichte

Die **optimalen Gewichte** sind proportional zu den *d'*-Werten der jeweiligen Agenten:

$$a_i \propto d'_i$$

Das bedeutet, der Agent mit höherer *d'*-Sensitivität bekommt ein größeres Gewicht, weil sein Urteil zuverlässiger ist.

Berechnung der Gesamtsensitivität (d’)

Ein Maß dafür, wie gut das Team zwischen den beiden Klassen, z. B. Signal vs. Rauschen, unterscheiden kann **(Leistungsmetrik)**.

- Die Sensitivät des Teams bei optimaler Kombination ergibt (wie in der Literatur beschrieben):

$$d'_{OW} = (d'_{operator} + d'_{aid})^{\frac{1}{2}}$$

➔ $d'_{operator}$ – Sensitivität des menschlichen Operators

➔ d'_{aid} – Sensitivität des automatisierten Hilfsmittels

- Mit anderen Worten: Die Team-*d’* ist die Quadratwurzel der Summe der Quadrate der einzelnen *d’*-Werte.
- Das bedeutet, die beste Leistung erreicht man durch **gewichtete Summe**, wobei die Gewichte proportional zu den jeweiligen *d’*-Werten sind.

Zusammenfassung

- Das **OW-Modell** beschreibt, wie man **theoretisch** die Hinweise optimal kombinieren sollte, um die maximale Genauigkeit (d’-Wert) zu erzielen.
- Es setzt voraus, dass man die **genauen Sensitivitäten** (d’-Werte) kennt und dass diese Hinweise **unabhängig** sind.
- Bei der praktischen Umsetzung bedeutet das: Die Hinweise sollten gewichtet werden, proportional zu ihrer Zuverlässigkeit, um die bestmögliche Entscheidung zu treffen.

Beispiel-Anwendung des OW-Modells im Farb-Detektionsexperiment

(basierend auf dem Paper: "Blue vs. Orange Dominanz" in Random-Dot-Bildern)

Rahmenbedingungen

- **Aufgabe:** Entscheide, ob ein Bild mehr **orange** (Signal) oder **blau** (Rauschen) dominiert.
- **Akteure:**

➔ *Menschlicher Operator*: $d'_{Operator} = 1.0$ (mäßige Sensitivität)
➔ *Automatisierte Hilfe*: $d'_{aid} = 3.0$ (hohe Sensitivität, 93% korrekt)

- **Annahme:** Beide bewerten das Bild unabhängig und liefern **Evidenzwerte** (X_{Operator}, X_{aid}), die normalverteilt sind.

Optimal Weighting (OW) Schritt-für-Schritt

1. **Evidenzwerte standardisieren**

- Die Rohwerte X_{Operator} und X_{aid} werden in *z-standardisierte Evidenzen* umgewandelt (Mittelwert = 0, Standardabweichung = 1).
- *Beispiel für ein orangenes Bild (Signal):*

➔ Operator: $X_{\text{Operator}} = +0.8$ (leicht positiv → tendiert zu "Orange").
➔ Hilfe: $X_{\text{aid}} = +2.5$ (stark positiv → klar "Orange").

2. **Gewichte berechnen**

- Die Gewichte a_i sind proportional zu d'_i

$$a_{\text{Operator}} = \frac{d'_{\text{Operator}}}{d'_{\text{Operator}} + d'_{\text{Hilfe}}} = \frac{1}{1+3} = 0.25$$

$$a_{\text{Hilfe}} = \frac{d'_{\text{Hilfe}}}{d'_{\text{Operator}} + d'_{\text{Hilfe}}} = \frac{3}{1+3} = 0.75$$

(Die Hilfe erhält 3x mehr Gewicht, da sie 3x sensitiver ist.)

3. **Gewichtete Evidenz kombinieren**

$$Z = a_{\text{Operator}} \cdot X_{\text{Operator}} + a_{\text{Hilfe}} \cdot X_{\text{Hilfe}} = 0.25 \cdot 0.8 + 0.75 \cdot 2.5 = 0.2 + 1.875 = 2.075$$

4. **Entscheidung treffen**

- **Schwellenwert:** Bei unvoreingenommener Entscheidung (kein Bias) ist der Schwellenwert $Z = 0$.

  $\rightarrow Z = 2.075 > 0$ → **Team-Entscheidung: "Orange"**.

- *Hinweis:* Bei Bias (z. B. Vorliebe für "Blau") würde der Schwellenwert verschoben werden.

Sensitivität des Teams (d'_{OW})

$$d'_{\mathrm{OW}} = \sqrt{(d'_{\mathrm{Operator}})^2 + (d'_{\mathrm{Hilfe}})^2} = \sqrt{1^2 + 3^2} = \sqrt{10} \approx 3.16$$

Interpretation: Das Team ist besser als die Hilfe allein (d' = 3.0), weil der Operator zusätzliche (wenn auch schwache) Information liefert.

Warum ist *Z* entscheidend?

- ***Z* ist der *integrierte Evidenzwert***, der die Unsicherheiten beider Quellen optimiert berücksichtigt.
- Im Experiment nutzten die Teilnehmer *Z* **nicht optimal**:

➔ Sie gewichteten die Hilfe zu wenig (z.B. ignorierten sie die Stärke der Evidenz X_{aid})
➔ Ergebnis: Ihre reale Team-Sensitivität lag näher am *Probability Matching-Modell* (suboptimal)

Kontrast zum tatsächlichen Experiment

- **Im Paper:** Die Hilfe gab binäre Urteile ("Orange/Blau") *plus* Evidenzstärke (z. B. "Orange 2.14").
- **Ideales OW-Modell:** Hätte die Teilnehmer veranlasst, die *numerischen Evidenzwerte* zu nutzen.
- **Realität:** Teilnehmer ignorierten die Evidenzstärke und folgten der Hilfe nur mit fester Wahrscheinlichkeit (~93%), daher blieb $d'_{Team} \ll d'_{OW}$

Sensitivität des Menschen berechnen

Die Studie nutzte Methoden der Signal Detection Theory (SDT), um die Sensitivität der Teilnehmer und der automatisierten Hilfe zu messen.

$$d' = z(Hit\ Rate) - z(False - Alarm\ Rate)$$

Beispiel:

- ➔ Ein Teilnehmer hat HR = 0.82 und FAR = 0.14
- ➔ Z$(0.82) \approx 0.92,\ z(0.14) \approx -1.08$
- ➔ $d' = 0.92 - (-1.08) = 2.0$

Für die automatisierte Hilfe:

- ➔ Die Hilfe wurde als *idealisiertes SDT-Modell* programmiert
- ➔ $d'_{aid} = 3.0$ (festgelegt durch die Parameter $\mu_{Signal} = 1.5,\ \mu_{Rauschen} = -1.5,\ \sigma = 1)$

Vorbereitungsphase: Baseline-Sensitivität ermitteln

➔ **Schritt 1: Kurzes Vor-Experiment ohne Hilfe**

- **Aufgabe:** Identisch zur Hauptstudie (z. B. "Orange vs. Blau" in Random-Dot-Bildern).
- **Durchführung:**
  - 50–100 Trials *ohne* automatisierte Hilfe.
  - Messung von **Hit Rate (HR)** und **False-Alarm Rate (FAR)**

Beispiel: Bei HR=0.75, FAR=0.10:

d' = z(0.75) − z(0.10) ≈ 0.67 − (−1.28) = 1.95

➔ **Schritt 2: Dynamische Anpassung**

- Falls die Hauptstudie längere Blöcke hat, kann d'_{human} zwischendurch neu berechnet werden (z. B. alle 50 Trials), um Lern- oder Ermüdungseffekte zu berücksichtigen

Umsetzung im Experiment

Sensitivität (d') messen

1. **Design:**
    - 2AFC-Aufgabe (z. B. "Orange vs. Blau") mit klar definierten Signal- und Rauschbedingungen.
    - Ausreichend viele Trials (mindestens 50 pro Bedingung).
2. **Datenanalyse:**
    - Berechne HR und FAR für jeden Teilnehmer
    - Korrigiere für extreme Raten (z. B. mit *Hautus-Korrektur*: +0.5 zu Hits/False Alarms).
    - Transformiere zu d' und Bias (c)

Evidenzwerte kontrollieren

- **Für eine automatisierte Hilfe:**
    - Simuliere Evidenzwerte aus Normalverteilungen mit definiertem d' (z.B. $\mu = d'/2$, $\sigma = 1$)
    - Beispiel für d' = 2.0:
        - Signal (orange): $N(1.0, 1)$
        - Rauschen (lila): $N(-1.0, 1)$
- **Für menschliche Evidenzen:**
    - Nutze **Confidence Ratings** (z.B. Skala von 1–5), um subjektive Evidenzstärke zu erfassen
    - Transformiere Ratings in kontinuierliche Werte (z. B. mittels *z-Scores*)

Mögliche Designimplikationen

- **Training:** Erkläre das Konzept von d' und Z mit Beispielen (zB: "Die Hilfe ist 3x so gut wie Sie – also hören Sie stärker auf sie")
- **Vereinfachte Darstellung:** Zeige nur die **empfohlene Entscheidung** (basierend auf Z) statt des Rohwerts
- **Adaptives Design:** Passe d'_{human} dynamisch an, falls sich die Leistung ändert (z.B. durch Lernen)