W05. Fenotypowanie cyfrowe

Psychoinformatyka | Wykład 05

2026-11-02

Cele i zakres wykładu

- Rozróżnić dane aktywne, pasywne i wskaźniki pochodne.
- Zaprojektować krótki protokół dziennego pomiaru EMA.
- Zrozumieć błędy, braki i ograniczenia wnioskowania z cyfrowych śladów.

Czym jest fenotypowanie cyfrowe?

- Pomiar wzorców zachowania i kontekstu z urządzeń cyfrowych.
- Fenotyp cyfrowy to opis cech i zmian oparty na zdefiniowanym protokole.
- Cyfrowy biomarker wymaga znacznie silniejszej walidacji w określonym zastosowaniu.

Fenotypowanie a tradycyjne badanie

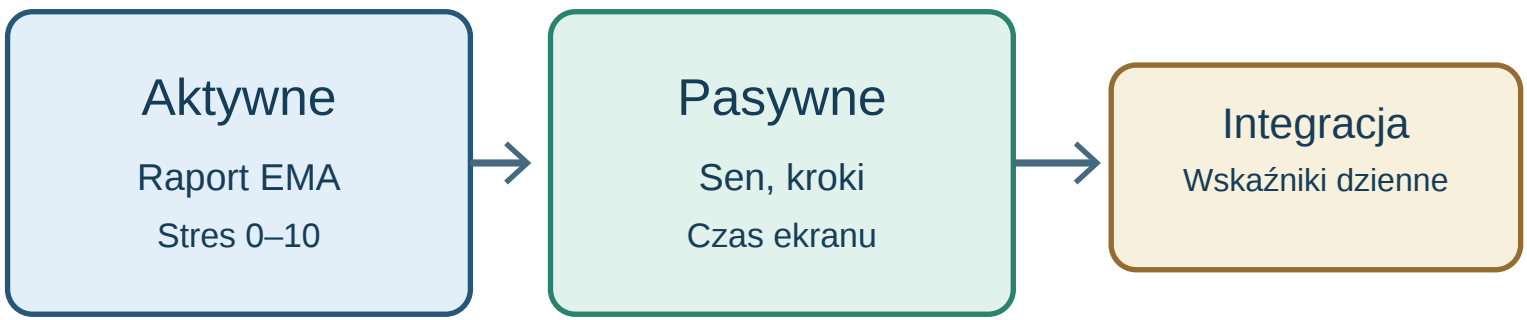
- Kwestionariusz retrospektywny pyta o doświadczenia z minionego okresu.
- EMA rejestruje krótkie raporty w codziennym kontekście.
- Pomiar pasywny obejmuje wybrane zachowania bez każdorazowej odpowiedzi.

Trafność ekologiczna i reaktywność

- Obserwacja w naturalnym środowisku zwiększa kontekstową reprezentatywność sytuacji.
- Powtarzane pytania mogą zmieniać uwagę i sposób raportowania.
- Ekologiczne nie oznacza automatycznie obiektywne ani trafne.

Pomiary aktywne i pasywne

- Aktywne: codzienny raport stresu, nastroju lub snu.
- Pasywne: liczniki kroków, zdarzenia ekranu, wybrane sygnały czujników.
- Dane pasywne wymagają zgody, interpretacji i oceny jakości czujnika.



Wyniki wymagają synchronizacji czasu i walidacji operacjonalizacji

Rysunek 1: Dane aktywne, pasywne i ich integracja

EMA: ecological momentary assessment

- Krótki pomiar powtarzany blisko doświadczenia.
- Rejestrowanie stanu, kontekstu, czasu i kompletności odpowiedzi.
- Wyniki nie są równoznaczne z diagnozą kliniczną.

Projektowanie harmonogramu EMA

- Ustal liczbę pytań, pory dnia i czas odpowiedzi.
- Losowe wezwania zmniejszają przewidywalność, lecz mogą obciążać uczestnika.
- Zapisuj czas wezwania i odpowiedzi, by rozdzielić pomiar od opóźnienia.

Sampling zależny od zdarzenia

- Event-contingent: samodzielne zgłoszenie wystąpienia zdefiniowanego zdarzenia.
- Time-contingent: wezwanie o określonej porze lub wylosowanym oknie.
- Event-contingent może faworyzować zdarzenia łatwiejsze do zauważenia.

Czujniki współczesnych smartfonów

- Akcelerometr i żyroskop mierzą ruch urządzenia, nie bezpośrednio cechę psychiczną.
- GPS i dostawcy lokalizacji szacują pozycję, często z różną dokładnością.
- Zdarzenia ekranu i aplikacji zależą od systemu, uprawnień i sprzętu.

Akcelerometr i żyroskop

- Surowy akcelerometr obejmuje wpływ grawitacji; przyspieszenie liniowe wymaga jej odfiltrowania.
- Licznik kroków jest estymacją z algorytmu urządzenia.
- Ruch telefonu ≠ zawsze ruch ciała, gdy urządzenie leży na biurku.

GPS: mobilność i regularność

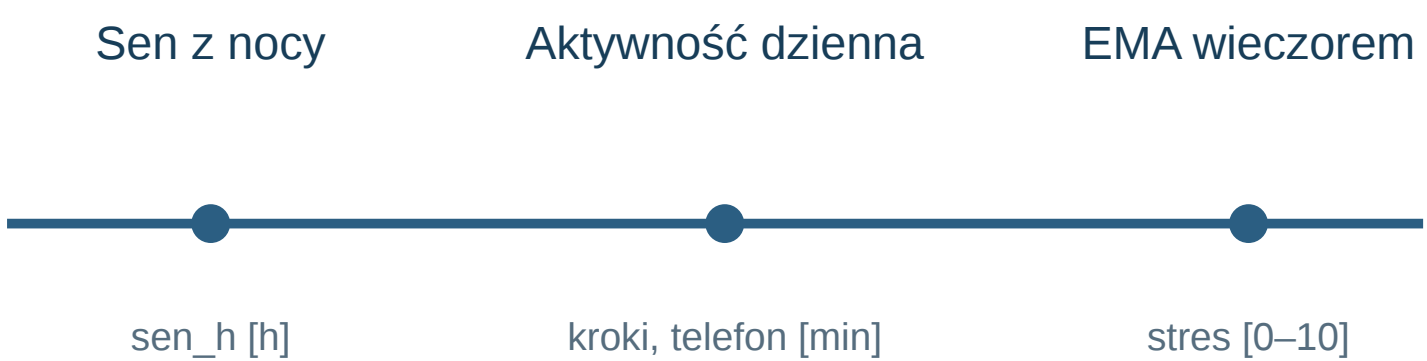
- Liczba odwiedzanych obszarów i czas spędzany w lokalizacjach to cechy pochodne.
- Brak sygnału GPS nie oznacza pozostawania w domu.
- Dokładne trajektorie są szczególnie wrażliwe ze względu na możliwość reidentyfikacji.

Czas ekranu i rytm dobowy

- Czas aktywności ekranu opisuje użytkowanie urządzenia, nie jednoznacznie zachowanie społeczne.
- Rytm dobowy wymaga oznaczania strefy i sposobu liczenia doby.
- Kontekst pracy, nauki i dostępności telefonu zmienia znaczenie pomiaru.

Wearables: sen i fizjologia

- Urządzenia mogą szacować sen, kroki, tętno i inne sygnały.
- Estymatory snu mają błędy zależne od algorytmu i użytkownika.
- Weryfikuj zakres dostępnych pomiarów i nie traktuj ich jako złotego standardu.



Rysunek 2: Oś czasu: sen, aktywność, samoopis EMA

Synchronizacja wielu źródeł

- EMA ma timestamp samoopisu, czujnik ma timestamp pomiaru.
- Należy ustalić jednostki, strefę i definicję dnia analitycznego.
- Dane należy agregować na osi czasu bez pozornego wyrównywania braków.

Przykład: 14 dni, 36 uczestników

- 36 uczestników × 14 dni = 504 rekordy osoba–dzień.
- Brakuje 34 samoopisów stresu (6,7%); zostaje 470 par sen–stres.
- Dane są syntetyczne; część braków zależy od wartości stresu.

5 minut przerwy

Pytania do samodzielnego przemyślenia

1. Czym różni się aktywny EMA od pomiaru pasywnego?
2. Co rejestruje akcelerometr, a czego nie rejestruje?
3. Dlaczego wyniki z naturalnego środowiska różnią się od laboratoryjnych?
4. Czy krótszy czas ekranu zawsze oznacza lepsze samopoczucie?

Od sygnału do cechy pochodnej

- Surowe próbki → walidacja → okna czasowe → statystyka → interpretacja.
- Liczba kroków jest cechą pochodną, a nie bezpośrednim opisem aktywności fizycznej.
- Każda transformacja może usunąć informację o kontekście.

Pipeline: czujnik → kontrola jakości → cecha dzienna → porównanie z EMA.

Wskaźniki mobilności

- Dystans, regularność przemieszczania się, liczba stref aktywności.
- Porównywalność zależy od częstości próbkowania i dokładności lokalizacji.
- Nie interpretuj ograniczonej mobilności jako objawu zaburzenia bez niezależnych danych.

Aktywność fizyczna i bezruch

- Kroki, epizody bezruchu i rozkład aktywności w czasie.
- Brak telefonu przy ciele powoduje niedoszacowanie liczby kroków.
- Porównania wymagają uwzględnienia czasu obserwacji oraz sposobu noszenia.

Cyfrowe wskaźniki snu

- Czas trwania i regularność są wskaźnikami zależnymi od metody pomiaru.
- Czas snu poprzedniej nocy należy przypisać do właściwego dnia EMA.
- Brak pomiaru snu nie jest równoznaczny z brakiem snu.

Zależności wewnątrz i między osobami

- Korelacja w obrębie osoby opisuje odchylenia od jej indywidualnej średniej.
- Korelacja między osobami dotyczy średnich różnych uczestników.
- Współczynniki mogą mieć inny znak i wielkość; dane powtarzane są zależne.

Dlaczego giną pomiary?

- Telefon może być rozładowany, wyłączony lub mieć odebrane uprawnienia.
- Uczestnik może pominąć raport EMA zwłaszcza w określonych stanach.
- W symulacji część braków zależy od ukrytej wartości stresu (MNAR).
- Dla rzeczywistych danych mechanizm braków wymaga dodatkowych założeń.

Python: wczytywanie zbioru EMA

- Otwórz CSV: 36 osób × 14 dni, 504 wiersze.
- Sprawdź unikalność klucza i 34 brakujące samoopisy.

```
import pandas as pd
dane = pd.read_csv('pi_w05_data.csv')
assert dane.shape == (504, 6)
assert dane.osoba.nunique() == 36
assert not dane.duplicated(['osoba', 'dzien']).any()
assert dane.stres.isna().sum() == 34
```

Python: statystyki dzienne

- Podsumuj sen, kroki i czas telefonu, zachowując jednostki.
- Porównaj średnie dla poszczególnych uczestników.
- Przypomnij różnicę między liczbą rekordów a liczbą osób.

```
stat = dane.groupby('osoba')[['sen_h','kroki','telefon_min']].mean()
print(stat.describe().round(2))
```

Python: wykresy w czasie

- Dwa panele: stres [0–10] oraz sen [h], osobne osie.
- Braki EMA pozostają przerwami, bez interpolacji.

```
import matplotlib.pyplot as plt
osoba = dane.loc[dane.osoba.eq(1)].sort_values('dzien')
fig, ax = plt.subplots(2, 1, sharex=True, figsize=(7, 4))
ax[0].plot(osoba.dzien, osoba.stres, 'o-')
ax[0].set_ylabel('Stres [0–10]')
ax[1].plot(osoba.dzien, osoba.sen_h, 's-')
ax[1].set(xlabel='Dzień', ylabel='Sen [h]')
fig.tight_layout()
plt.show()
```

Python: sen i stres

- 470 par: r wewnątrz osób ≈ −0,542.
- Średnie 36 osób: r między osobami ≈ −0,042.

```
pary = dane.dropna(subset=['sen_h', 'stres']).copy()
for v in ['sen_h', 'stres']:
    pary[v+'_c'] = pary[v] - pary.groupby('osoba')[v].transform('mean')
r_wewn = pary.sen_h_c.corr(pary.stres_c)
srednie = pary.groupby('osoba')[['sen_h', 'stres']].mean()
r_miedzy = srednie.sen_h.corr(srednie.stres)
print(round(r_wewn, 3), round(r_miedzy, 3))
```

Interpretacja bez nadmiernych wniosków

- Związek snu ze stresem nie dowodzi kierunku przyczynowego.
- Zmienne kontekstowe mogą wpływać na oba pomiary.
- Stanowisko: ostrożna interpretacja opisowa oraz jawne ograniczenia.

Zastosowania w badaniach zdrowia psychicznego

- W badaniach analizowano m.in. mobilność, sen, aktywność i zmienność samopoczucia.
- Przenoszenie modelu na inne urządzenia i populacje wymaga walidacji.
- Nie jest to samodzielna metoda rozpoznawania depresji ani innych zaburzeń.

Zgoda, bezpieczeństwo i kontrola uczestnika

- Minimalizuj dane, zwłaszcza lokalizacyjne i komunikacyjne.
- Informuj o zakresie pomiaru, retencji, usuwaniu danych i uprawnieniach.
- Ogranicz wtórne użycie danych i ryzyko reidentyfikacji.

Literatura i dokumentacja

- Onnela i Rauch (2016). Neuropsychopharmacology, 41, 1691–1696. DOI.
- Shiffman, Stone i Hufford (2008). Annual Review of Clinical Psychology, 4, 1–32. DOI.
- pandas, Matplotlib i Jupyter.

Wnioski i samodzielna praca

- Aktywne EMA i rejestracja pasywna dostarczają różnych wskaźników; żaden nie jest bezpośrednim odczytem emocji.
- Uwzględniaj kontekst, jakość danych i zależność wielokrotnych pomiarów tej samej osoby.
- Minimalizuj dane i określ zasady zgody, bezpieczeństwa, retencji i kontroli uczestnika.
- Przećwicz analizę w `pi_w05.ipynb` i porównaj z laboratorium L04.