Od ludzkiego zachowania do danych

Psychoinformatyka | Wykład 01: Podstawy psychoinformatyki

Cele wykładu i zakres omawianych zagadnień

- Rozróżnić konstrukt, wskaźnik i obserwację.
- Rozpoznać dane samoopisowe, behawioralne i psychofizjologiczne.
- Wykonać podstawową analizę syntetycznych czasów reakcji w Pythonie.

Dlaczego psychologia potrzebuje informatyki?

- Eksperyment może rejestrować każdą próbę z dokładnością znacznika czasu.
- Kod automatyzuje kontrolę jakości i pozwala powtórzyć analizę.
- Większa ilość danych nie zastępuje poprawnego projektu badania.

Czym jest psychoinformatyka? Definicje i przedmiot badań

- Obszar interdyscyplinarny: psychologia, informatyka, analiza danych i modelowanie.
- Przedmiotem są cyfrowe metody badania zachowania oraz procesów psychicznych.
- Wynik algorytmu to informacja o danych, a nie automatycznie rozpoznanie kliniczne.

Psychologia, informatyka i statystyka: trzy filary psychoinformatyki

- Psychologia formułuje konstrukty i hipotezy.
- Informatyka dostarcza aparatury pomiarowej, struktur danych i algorytmów.
- Statystyka kwantyfikuje efekty i niepewność; wszystkie trzy warstwy wymagają walidacji.

Historia psychoinformatyki: od eksperymentów komputerowych do sztucznej inteligencji

- Komputery umożliwiły precyzyjne zadania poznawcze i automatyczne rejestrowanie odpowiedzi.
- Internet i smartfony rozszerzyły badania na kontekst naturalny.
- Uczenie maszynowe ułatwia predykcję, ale zwiększa ryzyko przecieku danych i nadmiernej interpretacji.

Psychoinformatyka a psychologia obliczeniowa

- Psychoinformatyka: narzędzia cyfrowe, pomiary, przepływy danych i analiza.
- Psychologia obliczeniowa: formalne modele mechanizmów pamięci, uwagi i decyzji.
- Te dziedziny uzupełniają się, lecz nie są tożsame.

Psychoinformatyka, neuroinformatyka i affective computing

- Neuroinformatyka: analiza danych o układzie nerwowym, np. EEG.
- Affective computing: metody analizy sygnałów powiązanych z afektem.
- Psychoinformatyka obejmuje również samoopis, zachowanie i cyfrowe interakcje.

Od konstruktu psychologicznego do mierzalnego wskaźnika

- Konstrukt uwagi nie jest bezpośrednio widoczny w pliku CSV.
- Możliwy wskaźnik: czas reakcji i odsetek pomyłek w określonej procedurze.
- Operacjonalizacja wymaga uzasadnienia, kontroli zakłóceń i dowodów trafności.

Konstrukt	Wskaźnik	Obserwacja
Uwaga	RT	Czas w ms

Operacjonalizacja stresu, uwagi i impulsywności

- Stres: subiektywna ocena, kontekst i powtarzane pomiary.
- Uwaga: poprawność i czas reakcji w zadaniach poznawczych.
- Impulsywność: np. błędne reakcje w zadaniu Go/No-Go; jeden wskaźnik nie wyczerpuje konstruktu.

Dane samoopisowe: kwestionariusze i pomiary EMA

- Kwestionariusz: odpowiedzi na zestaw pozycji w ustalonym przedziale czasu.
- EMA: krótkie raporty zbierane wielokrotnie w kontekście codziennego życia.
- Odpowiedzi mogą zależeć od pamięci, kontekstu i reaktywności pomiaru.

Dane behawioralne: czas reakcji, poprawność odpowiedzi i aktywność użytkownika

- RT jest czasem od określonego zdarzenia do reakcji; jednostka: ms.
- Poprawność kodujemy jawnie, np. 1 = odpowiedź prawidłowa, 0 = błąd.
- Dokładność znacznika czasu zależy od urządzenia, systemu i implementacji zadania.

Dane psychofizjologiczne: EEG, ECG, EDA i eye tracking

- EEG: różnice potencjałów elektrycznych rejestrowane na skórze głowy.
- ECG i EDA: pomiary związane z pracą serca i przewodnictwem skóry.
- Eye tracking: położenie spojrzenia, sakkady i fiksacje; interpretacja zależy od zadania.

Cyfrowe ślady zachowania: smartfony, wearables i aktywność internetowa

- Aktywność urządzenia może dostarczać miar częstotliwości i regularności zachowania.
- Dane lokalizacyjne i czasowe mogą być silnie identyfikujące.
- Ślady cyfrowe są wskaźnikami pośrednimi, a nie bezpośrednimi miarami zdrowia psychicznego.

Jak reprezentujemy dane psychologiczne? Rekordy, zmienne i szeregi czasowe

- Jedna próba to rekord; uczestnik, warunek, RT i poprawność to przykładowe kolumny.
- Powtarzane obserwacje jednej osoby nie są niezależne.
- W analizie czasowej konieczne są jednoznaczne znaczniki czasu i jednostki.

Studium przypadku: pomiar uwagi i czasu reakcji w eksperymencie komputerowym

- 24 uczestników, po 40 prób, dwa warunki: łatwy i trudny.
- Porównujemy rozkłady RT oraz poprawność; efekt trudności jest narzucony symulacją.
- Prawidłowa jednostka porównań uczestników to statystyka liczona osobno dla każdej osoby.

5 minut przerwy

Pytania do samodzielnego przemyślenia

1. Czym psychoinformatyka różni się od neuroinformatyki?
2. Jakie wskaźniki można wykorzystać do badania uwagi?
3. Czy czas reakcji stanowi bezpośrednią miarę poziomu uwagi?

Od pytania badawczego do eksperymentu psychoinformatycznego

- Pytanie: czy trudność zadania wiąże się z czasem reakcji?
- Manipulacja: warunek łatwy versus trudny; wyniki: RT i poprawność.
- Warunki należy równoważyć i mierzyć w zdefiniowanym protokole.

Cykl przetwarzania danych: pomiar, analiza, modelowanie i interpretacja

- Pomiar → kontrola jakości → dane analityczne → estymacja → interpretacja.
- Każde przekształcenie danych wymaga jawnego uzasadnienia.
- Oddziel dane surowe od plików po przetworzeniu.

Pomiar → Kontrola → Analiza → Interpretacja.

Badania eksperymentalne a obserwacyjne: możliwości i ograniczenia

- Eksperyment: kontrolowana manipulacja warunku, potencjalnie silniejsze wnioskowanie przyczynowe.
- Obserwacja: rejestracja zachowań bez przypisywania warunków eksperymentalnych.
- W obu przypadkach interpretacja zależy od trafności pomiaru i czynników zakłócających.

Jakość danych: błędy pomiarowe, artefakty i brakujące obserwacje

- Brak odpowiedzi zapisujemy jako wartość brakującą, a nie 0 ms.
- Czasy poniżej 150 ms można oznaczyć do kontroli według jawnego protokołu.
- Filtrowanie po obejrzeniu wyników może zniekształcać analizę.

Rzetelność i trafność cyfrowego pomiaru psychologicznego

- Rzetelność dotyczy powtarzalności/zgodności pomiaru w określonych warunkach.
- Trafność dotyczy uzasadnienia interpretacji wyniku w konkretnym zastosowaniu.
- Precyzyjny zegar programu nie gwarantuje trafnego wskaźnika uwagi.

Reprodukowalność badań: dane, metadane i kod źródłowy

- Zapisz wersje bibliotek, ziarno generatora losowego i reguły filtrowania.
- Metadane powinny objaśniać jednostki oraz znaczenie każdej kolumny.
- Notebook umożliwia ponowne wykonanie wszystkich kroków od danych do wykresu.

Jupyter Notebook i Google Colab jako środowiska badawcze

- Notebook łączy kod, wyniki, wykresy i komentarz metodologiczny.
- Google Colab ułatwia uruchamianie bez konfiguracji lokalnej.
- Pliki wejściowe muszą być dostępne w bieżącej sesji; środowiska mogą różnić się wersjami bibliotek.

Python w psychoinformatyce: wczytywanie danych eksperymentalnych

- Wczytaj pi_w01_data.csv przy użyciu pandas.read_csv.
- Sprawdź kolumny participant, condition, trial, rt_ms i correct.
- Przykład korzysta z tego samego pliku co pi_w01.ipynb.

Przykład: ten sam kod jest dostępny w pi_w01.ipynb.

```
import pandas as pd
df = pd.read_csv('pi_w01_data.csv')
print(df[['participant','condition','rt_ms','correct']].head())
```

Python w psychoinformatyce: obliczanie podstawowych statystyk

- Do analizy RT użyj poprawnych prób w zadanym zakresie 150–2000 ms.
- Podsumuj liczbę obserwacji, średnią, medianę i odchylenie standardowe.
- Poprawność oblicz osobno dla warunków.

Przykład: ten sam kod jest dostępny w pi_w01.ipynb.

```
valid = df.loc[df.correct.eq(1) & df.rt_ms.between(150, 2000)]
summary = valid.groupby('condition').rt_ms.agg(['count','mean','median','std'])
print(summary.round(1))
```

Python w psychoinformatyce: wizualizacja rozkładu czasów reakcji

- Histogram pokazuje liczbę prób w przedziałach czasu reakcji.
- Osie: RT [ms] i liczba poprawnych prób; rozróżnij oba warunki.
- Wizualizacja rozkładu poprzedza testowanie hipotez.

Przykład: ten sam kod jest dostępny w pi_w01.ipynb.

```
import matplotlib.pyplot as plt
for name, group in valid.groupby('condition'):
    plt.hist(group.rt_ms, bins=15, alpha=.5, label=name)
plt.xlabel('Czas reakcji [ms]'); plt.ylabel('Liczba prób')
plt.legend(); plt.show()
```

Od wykresu do interpretacji psychologicznej: przykład analizy wyników

- Zestaw mediany RT dla każdej osoby w obu warunkach.
- Różnica: mediana trudny minus mediana łatwy, w milisekundach.
- Ponieważ to symulacja, różnica wynika z założonego modelu generowania danych.

Przykład: ten sam kod jest dostępny w pi_w01.ipynb.

```
pairs = valid.groupby(['participant','condition']).rt_ms.median().unstack()
pairs['difference_ms'] = pairs['trudny'] - pairs['łatwy']
print(pairs['difference_ms'].describe())
```

Korelacja a przyczynowość: pułapki wnioskowania na podstawie danych

- Współwystępowanie dwóch zmiennych nie identyfikuje automatycznie kierunku przyczynowego.
- Zmienne pominięte i selekcja próby mogą wytworzyć pozorny związek.
- Analiza danych nie naprawia błędnego projektu badania.

Prywatność, świadoma zgoda i etyczne granice analizy zachowania

- Zbieraj tylko dane niezbędne do określonego celu i stosuj adekwatną podstawę prawną.
- Identyfikatory pseudonimowe nadal mogą umożliwiać powiązanie danych z osobą.
- Uczestnik powinien rozumieć, jakie zachowania są rejestrowane i jak długo przechowuje się dane.

Co potrafi psychoinformatyka, a czego nie? Podsumowanie wykładu

- Pozwala powtarzalnie rejestrować i analizować obserwowalne dane.
- Nie gwarantuje automatycznie diagnozy ani wyjaśnienia mechanizmu psychologicznego.
- Poprawny wniosek wymaga konstruktu, procedury, jakości danych i walidacji.

Literatura naukowa, narzędzia i materiały do samodzielnej pracy

- Yarkoni (2012), Psychoinformatics: New Horizons at the Interface of the Psychological and Computing Sciences, DOI: 10.1177/0963721412457362.
- PsychoPy: psychopy.org; Jupyter: jupyter.org; pandas: pandas.pydata.org.
- Notebook: pi_w01.ipynb; dane: pi_w01_data.csv.