ROBOTYKA AFEKTYWNA

Wykład 6: Architektury systemów afektywnych i etyka

Automatyka i Robotyka, studia II stopnia

dr inż. Mateusz Pomianek • Politechnika Rzeszowska

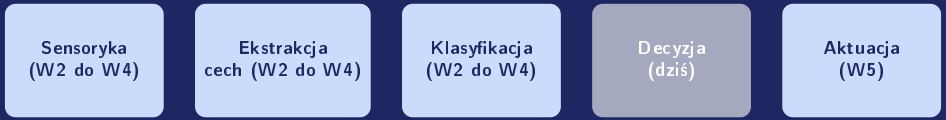
Po tym wykładzie będziesz umiał/a:

- Opisać pętlę percepcja, ocena, reakcja jako całościową architekturę systemu afektywnego
- Porównać warstwy decyzyjne: FSM, behavior tree oraz architekturę BDI
- Wskazać, jak moduły percepcji afektywnej integrują się w systemie produkcyjnym typu RoboMVP
- Rozróżnić pomoc od manipulacji emocjonalnej przy pomocy kryterium zgody i przejrzystości
- Zaprojektować podstawowe zasady ochrony prywatności danych afektywnych



SZYBKIE PRZYWOŁANIE Z CAŁEGO KURSU

Które wykłady odpowiadały za które etapy pipeline'u z W1?



Dziś domykamy architekturę i pytamy, czy to, co zbudowaliśmy, jest odpowiedzialne.

Pomoc czy manipulacja

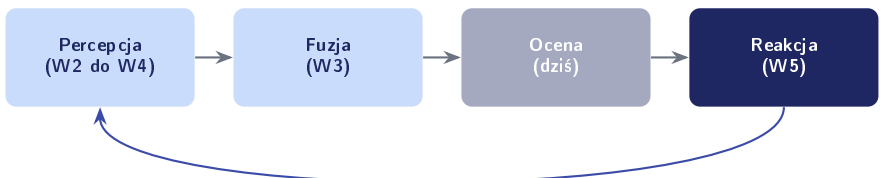
Robot obsługi klienta wykrywa frustrację rozmówcy i celowo przechodzi na cieplejszy, bardziej empatyczny ton głosu, dokładnie tak, jak uczyliśmy się na W5.

Ten sam mechanizm wykorzystuje jednak, by w chwili osłabionej czujności klienta zaproponować dodatkową, drogą usługę.

Ta sama technika inżynierska służy raz opiece, raz manipulacji. Różnicę robi intencja i przejrzystość, nie kod.



Pełna architektura: pętla percepcja, ocena, reakcja



Strzałka powrotna to pętla sprzężenia zwrotnego: reakcja robota zmienia stan operatora, który znów trafia do percepcji. System obserwuje skutki własnych działań.

Warstwa percepcji: rekapitulacja

- **Wizja (W2)**: landmarki, Action Units, CNN, wektor (valence, arousal, confidence)
- **Audio (W3)**: MFCC, pitch, jitter i shimmer, klasyfikacja przez SVM albo wav2vec2
- **Fizjologia (W4)**: EDA, HRV, EEG konsumenckie, kalibracja bazowa per operator
- **Fuzja (W3)**: ważona kombinacja kanałów na podstawie pewności klasyfikatora, degradacja łagodna przy braku danych

Warstwa decyzyjna: FSM, behavior tree, BDI

Architektura	Zasada działania	Ograniczenie
FSM (W1)	Skończona liczba stanów, proste reguły if oraz progi	Trudna do rozbudowy przy wielu warunkach
Behavior tree	Drzewo węzłów: sekwencja, selektor, warunek, akcja, wykonywane cyklicznie	Wymaga starannego projektu struktury drzewa
BDI (Belief-Desire-Intention)	Model przekonań, celów i intencji, robot rozumuje o tym, dlaczego działa	Wysoka złożoność implementacyjna

Behavior tree jest dziś standardem w robotyce mobilnej i manipulacyjnej, bo modularnie łączy proste bloki w złożone zachowania, bez eksplozji liczby stanów typowej dla FSM.

Warstwa aktuacji: rekapitulacja

- Mapowanie afektu na parametry ruchu i prozodii, zawsze przez funkcję clip (W5)
- Ekspresja przez ciało (Laban Movement Analysis) tam, gdzie robot nie ma twarzy
- Spójność multimodalna: ruch i głos czytają ten sam temat /affect/state
- Bezpieczny zakres pracy robota jest nadrzędny wobec intensywności wykrytej emocji

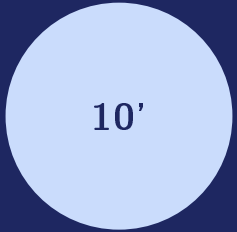
Integracja z ROS2: RoboMVP jako system produkcyjny

- Moduły percepcji afektywnej to osobne węzły ROS2, dokładnie jak w architekturze z W1
- Orchestrator decyzyjny subskrybuje /affect/state i koordynuje zadania manipulacyjne robota
- Telemetria i logowanie: dane z sesji trafiają do dalszego treningu i walidacji modeli
- Wersjonowanie modeli klasyfikacji pozwala wycofać zmianę, jeśli pogorszy jakość decyzji

SZYBKIE PRZYWOŁANIE

Czym różni się behavior tree od zwykłego FSM z W1?

30 sekund indywidualnie, potem porównaj odpowiedź z sąsiadem



PRZERWA

10 minut, wracamy o godzinie [uzupełnić]

Szybkie przypomnienie

- Mamy już pełną, zamkniętą pętlę: percepcja, fuzja, ocena, reakcja, sprzężenie zwrotne
- RoboMVP pokazuje, jak ta architektura skaluje się do systemu produkcyjnego

Teraz zajmiemy się pytaniem, którego dotąd unikaliśmy: czy zawsze powinniśmy to budować.

Dlaczego etyka robotyki afektywnej jest pilna

- Roboty czytające emocje wchodzą do fabryk, opieki zdrowotnej i domów, nie tylko do laboratoriów badawczych
- Dane afektywne ujawniają stan psychiczny, którego osoba mogła nie chcieć ujawniać
- Te same mechanizmy, które budują zaufanie, mogą je świadomie nadużywać
- Odpowiedzialność projektowa zaczyna się na etapie architektury, nie dopiero przy wdrożeniu

Manipulacja emocjonalna: gdzie leży granica

- Manipulacja emocjonalna to celowe wykorzystanie wykrytego stanu afektywnego wbrew interesowi osoby, dla korzyści operatora systemu
- Kryterium rozróżniające: czyj interes realizuje reakcja robota, osoby, z którą wchodzi w interakcję, czy właściciela systemu
- Zwodnicze wzorce projektowe, ang. dark patterns, przeniesione do interakcji afektywnej są trudniejsze do zauważenia niż w interfejsie graficznym
- Sama technika z W5, dopasowanie tonu głosu do emocji, jest etycznie neutralna, kontekst użycia decyduje

Macierz zgody i przejrzystości

	Przejrzystość obecna	**Przejrzystość brak**
Zgoda jawna	Akceptowalne: osoba wie i zgadza się na analizę emocji	Ryzykowne: zgoda niepełna, bo osoba nie wie, jak dane są używane
Zgoda domniemana	Wątpliwe: brak wyraźnej zgody osłabia legitymizację	Nieakceptowalne: ukryta analiza i ukryte użycie, definicja manipulacji

Jedyne bezpieczne pole to jawna zgoda połączona z przejrzystością. Projektując system z W1, warto już na etapie architektury zaplanować, gdzie i jak osoba jest informowana o analizie afektywnej.

Prywatność danych afektywnych

- Dane afektywne, podobnie jak inne dane biometryczne, należą do kategorii wymagającej szczególnej ochrony
- Minimalizacja danych: przetwarzaj lokalnie, na edge, tak jak w architekturze z W1 i W5, zamiast wysyłać surowe dane do chmury
- Ogranicz czas przechowywania: wynik klasyfikacji potrzebny do decyzji w czasie rzeczywistym nie musi być archiwizowany bezterminowo
- Oddziel dane identyfikujące osobę od wyniku klasyfikacji emocji, tam gdzie to możliwe

Antropomorfizacja: korzyści i ryzyka

Korzyści

Łatwiejsza, bardziej intuicyjna komunikacja z robotem. Wyższa wstępna akceptacja społeczna, patrz W1, slajd 6.

Ryzyka

Nadmierne zaufanie, złudne przekonanie, że robot naprawdę rozumie. Niespójność multimodalna z W5 pogłębia dyskomfort, gdy złudzenie pęka.

Zaufanie: budowa powolna, utrata gwałtowna

- Zaufanie do robota rośnie stopniowo, dzięki wielokrotnym, spójnym i przewidywalnym interakcjom
- Jedno zdarzenie niespójności, na przykład błędna klasyfikacja emocji prowadząca do niewłaściwej reakcji, może zniszczyć zaufanie natychmiast
- Ta asymetria oznacza, że bezpieczne wartości domyślne, patrz clip z W5, są ważniejsze niż maksymalizacja trafności pojedynczej decyzji
- Przejrzystość, patrz slajd 15, jest tańszym sposobem budowania zaufania niż doskonalenie samego modelu

Checklista: etyczny review węzła afektywnego

```
def ethical_review(node_design):
    checks = {
        'zgoda': node_design.has_explicit_consent_flow(),
        'przejrzystosc': node_design.discloses_affect_sensing(),
        'minimalizacja': node_design.stores_raw_data() is False,
        'interes_osoby': node_design.optimizes_for_user_wellbeing(),
        'bezpieczne_wartosci': node_design.uses_safe_defaults(), # clip, W5
    }
    passed = all(checks.values())
    if not passed:
        raise DeploymentBlocked(failed=checks)
    return True
```

Studium przypadku: robot w opiece nad osobami starszymi

- System wykrywa u podopiecznego oznaki samotności, obniżona energia głosu, mniej inicjowanych rozmów, patrz W3
- Robot inicjuje rozmowę i proponuje kontakt wideo z rodziną
- Pytanie do dyskusji: czy to wsparcie, czy niebezpieczne zastępowanie realnego kontaktu z ludźmi wygodnym substytutem
- Zastosuj macierz ze slajdu 15 oraz kryterium interesu ze slajdu 14 do tego scenariusza

Kluczowe pojęcia dzisiejszego wykładu

Pętla percepcja, ocena, reakcja	**FSM, behavior tree, BDI**	RoboMVP
Manipulacja emocjonalna	**Macierz zgody i przejrzystości**	Zaufanie i prywatność

Quiz kontrolny

1. Jakie kryterium odróżnia empatyczną reakcję robota od manipulacji emocjonalnej?

2. Robot analizuje emocje bez informowania o tym rozmówcy, ale ma jego domniemaną zgodę na usługę. W którym polu macierzy ze slajdu 15 się to znajduje?

3. Dlaczego jedno zdarzenie niespójności potrafi zniszczyć zaufanie budowane przez wiele poprawnych interakcji?

Wykład 7: studia przypadków i trendy badawcze

- Studium przypadku pełnego wdrożenia systemu HRI, od czujnika do reakcji
- Benchmarking modeli emocji pod kątem wdrożenia embedded: kwantyzacja, ONNX, TensorRT
- Wymagania projektu zaliczeniowego oraz otwarte problemy badawcze

Praca własna przed W7:

Przygotuj jednozdaniowy opis architektury swojego projektu zaliczeniowego, zgodny z pętlą ze slajdu 5.

Dziękuję za uwagę

Pytania?