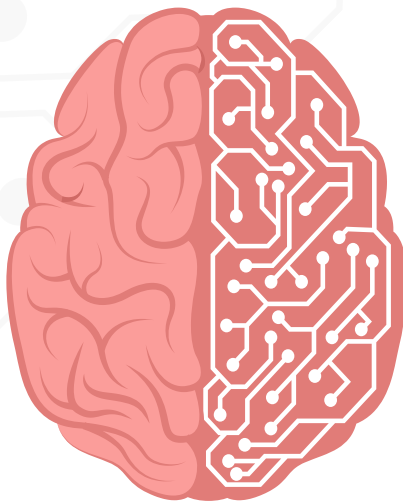




Uczenie maszynowe w wykrywaniu schizofrenii i autyzmu na podstawie wypowiedzi tekstowych

Aleksander Wawer, Zespół Inżynierii Lingwistycznej, IPI PAN

Izabela Chojnicka, AutismLabUW Research Team, Wydział Psychologii UW

Justyna Sarzyńska-Wawer, Pracownia Psycholingwistyki i Psychologii Poznawczej, IP PAN

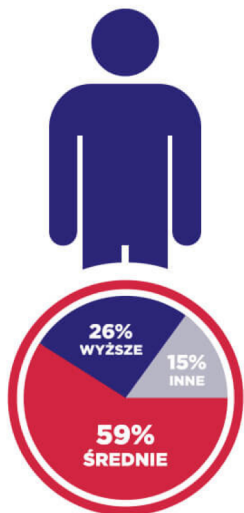
Agenda

1. Wprowadzenie do obu zaburzeń i diagnostyka
 - a. Schizofrenia, metoda **TLC**
 - b. Autyzm, metoda **ADOS-2**
2. Metody słownikowe w autyzmie
3. Uczenie maszynowe: pojedyncze zaburzenie
4. Uczenie maszynowe: oba zaburzenia
 - a. Zero-Shot
 - b. Transfer learning
5. Few-Shot
6. Konkluzje
7. Bibliografia

Schizofrenia (SCZ)

- Zaburzenie z grupy psychoz
- Jedna z najcięższych i najbardziej stygmatyzujących chorób psychicznych
- 1% populacji ogólnej

Przeciętny pacjent
ze schizofrenią w Polsce



Zaburzenia mowy

- Nieprawidłowości obserwowane na każdym poziomie:
 - fonologicznym
 - leksykalnym
 - morfologicznym
- niespójność i rozluźnienie skojarzeń (DSM IV)



Formalne zaburzenia myślenia
(Formal Thought Disorder, FDT)

Diagnostyka

- Brak badań laboratoryjnych
- Oparta o wywiad i ocenę zachowania pacjenta
- Niezgodność ekspertów
- Skale diagnostyczne (np. Skala Objawów Pozytywnych i Negatywnych PANSS)
 - Bardzo istotna dla długofalowych rokowań
 - Rola wczesnej diagnozy i badań przesiewowych

Diagnostyka FDT

- Assessment of Thought, Language, and Communication, TLC – skala schizofazji obejmująca 18 wskaźników np. neologizmy, persewercje, echolalie, uskokowość, rozkojarzenie, nielogiczność, ubóstwo mowy czy ubóstwo treści
- Ocena dokonywana na skali od 0 do 3 lub 4 punktów
- Jej stosowanie wymaga wcześniejszego treningu

Zbiór danych

- 35 pacjentów z Instytutu Psychiatrii i Neurologii; pacjenci w stanie remisji; brak zmian leków w ostatnich miesiącach; brak innych chorób
- 35 osób stanowiących grupę kontrolną (dobranych pod względem płci, wieku i wykształcenia)
- Transkrypcje wywiadów klinicznych (6 pytań: dzieciństwo, rodzina, hobby, najbliższa osoba, choroby, wiara w Boga) z usuniętymi fragmentami, w których pacjenci przyznają się do choroby

Spektrum autyzmu (ASD)

- 1,85% (CDC, 2020)
- ~1% (ASDEU, 2020)
- 1 kobieta : 4 mężczyzn
- Diagnostyka – wskaźniki behawioralne, zespół wielospecjalistyczny



Designed by pch.vector / Freepik



Spektrum autyzmu — język i komunikacja

Duża heterogeniczność

Trudności:

- wykorzystywanie mowy w celach społecznych
- komunikacja werbalna i niewerbalna
- umiejętności narracyjne



- ↓ liczba słów, w tym służących opisowi stanów poznawczych i emocjonalnych
- ↓ różnorodność leksykalna
- narracje mniej powiązane przyczynowo-skutkowo i mniej spójne

- Nietypowe intonacja, głośność, rytm, tempo mowy
- Echolalie bezpośrednie i odroczone
- Mowa stereotypowa
- Język idiosynkratyczny
- Neologizmy
- Mowa pedantyczna
- Rytuały werbalne

Zbiór danych: Osoby z ASD

Table 1. Demographics of the sample.

	ASD Group (<i>n</i> = 25)	TD Group (<i>n</i> = 25)	p-value
Males <i>n</i>	22	22	
Chronological age in years <i>M</i> (<i>SD</i>)	14.55 (5.46)	14.38 (5.83)	.68
Chronological age range	7.1–24.6	7.3–25.3	
Nonverbal IQ <i>M</i> (<i>SD</i>)	109.08 (13.04)	114.64 (12.50)	.98
Nonverbal IQ range	80–140	80–134	
Verbal IQ <i>M</i> (<i>SD</i>)	108.28 (18.30)	113.12 (14.33)	.24
Verbal IQ range	72–131	85–141	

ASD–autism spectrum disorder; TD–typical development; *M*–Mean; *SD*–standard deviation; IQ–Intelligence Quotient.

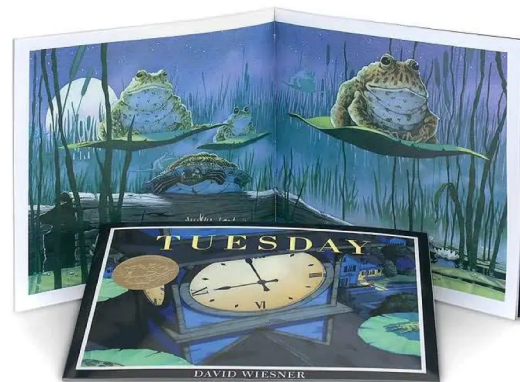
Table 1 Clinical Characteristics of the ASD Group

	ASD Group	TD Group
<i>N</i>	37	37
Males %	89.19	70.27
Chronological age in years <i>M</i> (<i>SD</i>)	15.31 (6.38)	14.81 (8.46)
Chronological age range	6.83–35.42	5.67–50.67
Nonverbal IQ <i>M</i> (<i>SD</i>)	106.41 (12.80)	115.11 (12.64)
Verbal IQ <i>M</i> (<i>SD</i>)	106.06 (17.14)	114.00 (14.42)
ADOS-2 Overall Score <i>M</i> (<i>SD</i>)	13.62 (4.81)	2.51 (3.55)
SCQ Overall Score <i>M</i> (<i>SD</i>)	21.91 (6.91)	3.64 (6.03)

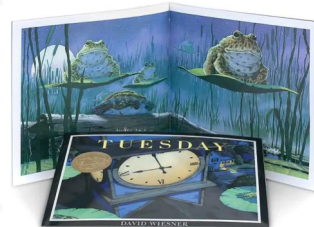
Zbiór danych: Transkrypcje narracji

Zadania ADOS-2:

- Opowiadanie historii z książki
- Opis obrazka



Metody słownikowe: autyzm



- Brak różnic: czas trwania zadania, liczba tokenów
- Krótsze wypowiedzi ASD ($p < .001$)
- Brak różnic: analiza morfoskładniowa
- Osoby z ASD mniej słów nacechowanych pozytywnie ($p = .008$, Cohen's $d = .83$) i negatywnie ($p = .054$, Cohen's $d = .58$)
- Niższy poziom abstrakcji wypowiedzi ($p = .020$, Cohen's $d = .70$),
mniej SVs ($p = .037$, Cohen's $d = .62$),
mniej DAVs ($p = .025$, Cohen's $d = .67$)

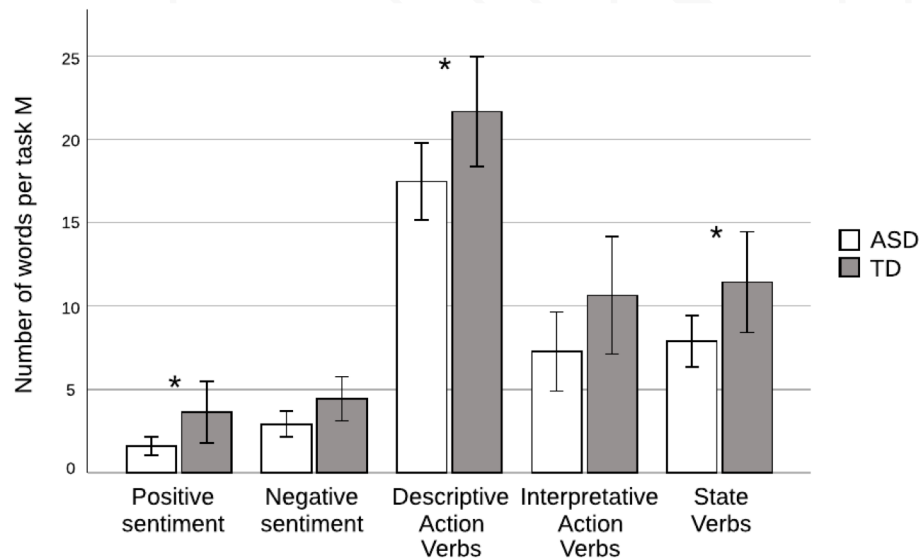


Fig 2. Group differences in the number of words used in sentiment and linguistic abstraction analyses for participants with autism spectrum disorder (ASD) and controls with typical development (TD); * $p < .05$.

Uczenie maszynowe w diagnostyce

- **Cel:** nauczyć się rozpoznawać zaburzenia na podstawie samej tylko tekstowej zawartości, jaką dostajemy z transkrypcji wywiadów – narzędzi diagnostycznych.
- **Problem:** rozmiary danych obu zbiorów są dwa rzędy wielkości za małe na wymagania współczesnego głębokiego uczenia maszynowego.
 - Trenowanie (fine-tuning) najpopularniejszego ostatnio modelu BERT na naszych danych nie przynosi żadnego efektu.
 - Nie da się wytrenować ~110 mln parametrów na <100 obserwacjach!
- **Rozwiązania:**
 - Wypróbowanie metod starszego typu (wektory słownikowe, bag-of-words itd).
 - Zmniejszenie liczby parametrów modelu – ale najlepiej korzystając nadal z głębokiego uczenia.
 - Podejścia typu few-shot learning.

Jak zmniejszyć liczbę parametrów: wektory zanurzeń zdań

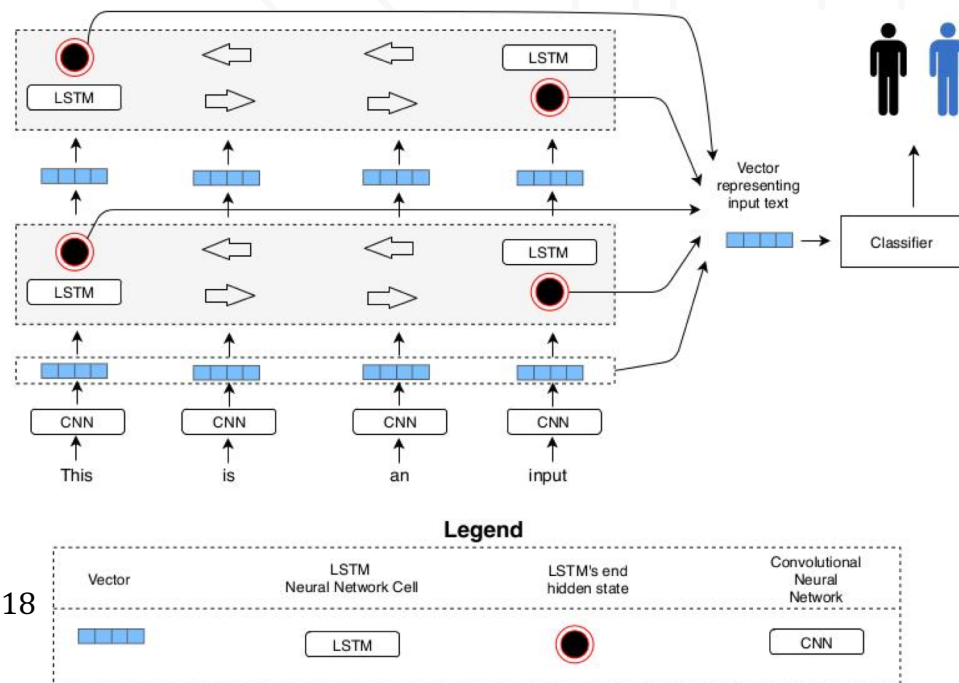
- Niektóre z pre-trenowanych głębokich sieci neuronowych to uniwersalne narzędzia, generujące z wejściowego tekstu wektory zanurzeń (utterance/sentence embeddings).
- Sieci te są pre-trenowane, zazwyczaj na zadaniu modelowania języka.
- Wykorzystujemy je tylko w celu uzyskania reprezentacji tekstu – czyli wektorów zanurzeń zdań.

Nie trenujemy ich na danych naszego problemu.

- Korzystając z wektorowych reprezentacji zdań, możemy zastosować jeden z wybranych modeli “klasycznego” uczenia maszynowego, sprawdzony w przypadku niedużych zbiorów danych.

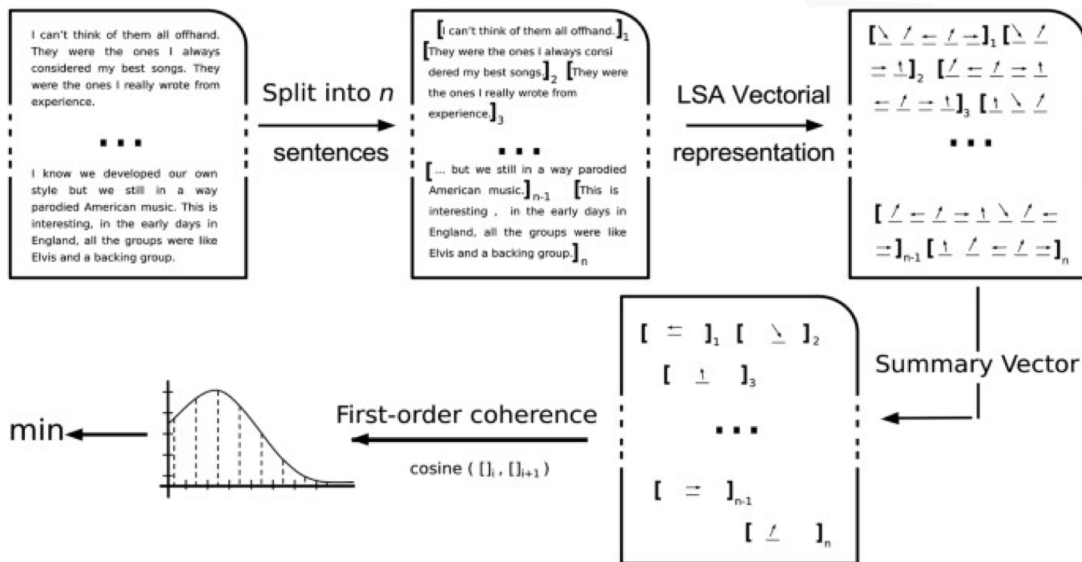
Wektory zanurzeń zdań ELMo w diagnozie SCZ

- Pierwsze testowanie rozwiązanie to ELMo (Peters et al. 2018).
- To 2-warstwowa, dwukierunkowa sieć LSTM, z warstwą konwolucyjną (CNN) na poziomie znaków.
- Sieć ta została pre-trenowana na zadaniu modelowania języka.
 - 20 mln słów; Wikipedia + Common Crawl.
- Generuje ona wektory o rozmiarze 1024.
- Wykorzystaliśmy polski model dostępny w ElmoForManyLangs.
- Następnie dokładamy model klasyfikacyjny:
 - SVM z jądrem RBF



Poprzednie prace nad automatycznym wykrywaniem SCZ: inżynieria cech

- Latent Semantic Analysis jako miara spójności semantycznej plus składniowe markery złożoności mowy: maksymalna długość frazy i użycie wyznaczników (np. 'który').



Wektory zanurzeń ELMo w diagnozie SCZ

- Porównaliśmy implementację metody Bedi et al. na naszych danych SCZ z przestrzenią cech z enkodera ELMo.
- Dla odpowiedzi na każde z 6 pytań TLC był to wektor ELMo o rozmiarze 1024.

Method	Questions	Accuracy
Coherence (Bedi et al.) + SVM	1-6	0.70 (± 0.05)
ELMo + SVM	1-6	0.80 (± 0.05)
	2-6	0.84 (± 0.05)

Dokładność (accuracy) to średnia w walidacji krzyżowej leave-one-out ($\pm$ błąd standardowy)

TLC	N/A	0.74
-----	-----	------

ELMo w diagnozie SCZ: objaśnialność modeli

LIME Sentiment Analysis Example

The movie is mediocre, maybe even bad.

Negative 99.8%

The movie is mediocre, maybe even ~~bad~~.

Negative 98.0%

The movie is ~~mediocre~~, maybe even bad.

Negative 98.7%

The movie is ~~mediocre~~, maybe even ~~bad~~.

Positive 63.4%

The movie is ~~mediocre~~, maybe even bad.

Positive 74.5%

The ~~movie~~ is mediocre, maybe even ~~bad~~.

Negative 97.9%

The movie is **mediocre**, maybe even **bad**.

ELMo w diagnozie SCZ: objaśnialność modeli

Chorują przez grzech. Więcej. Mmm no źródłem wszystkich chorób jest szatan. Psychicznych, no bo fizyczne sami nabywamy, nie? Czyli jak się przeziębimy, to jest choroba fizyczna, tak samo jak złamiemy rękę, czy nogę, no to też jest przez nas, nie? Dlaczego ludzie wierzą w Boga, (przerwa) ciężkie pytanie (przerwa). Yyy nie wiem, dlaczego ludzie wierzą w Boga, ale wiem dlaczego ja wierzę w Boga. Dlatego, że mmm odczułem jego obecność. Nie raz. Yyy Moi bliscy tak, z tego samego względu myślę znaczy, nie mogę mówić za nich dokładnie, ale myślę yyy że też odczuwają, odczuwali i i i i widzieli co się dzieje, yyy w pewnych sytuacjach, i myślę że to jest główny powód.

o_explained: K47

Prediction probabilities

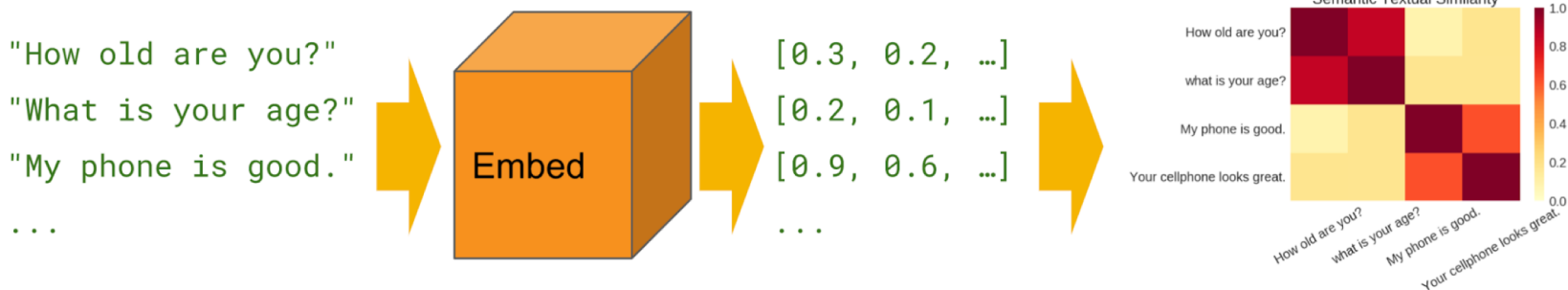


ELMo w diagnozie SCZ: objaśnialność modeli

- Często powtarzane słowa zawierające niewielką zawartość (np. „jakoś”, „trochę”, „taki”, „dobrze”) wskazują na klasyfikację FTD przez ELMo+SVM. To odkrycie jest zgodne z rolą niskiej gęstości semantycznej opisanej przez (Rezaii 2019).
- Słowa “uduchowione” czasami wskazują na FTD.
- Praca i słownictwo zawodowe znajdujące się w niektórych tekstach wskazują na zdrowe osoby / grupy kontrolne.
- Rodzinne odniesienia są zwykle ważne, ale co ciekawe, odgrywają mieszaną rolę. W większości przypadków wskazują one na zdrowe osoby / grupy kontrolne, ale można również znaleźć wiele przeciwnych przykładów.

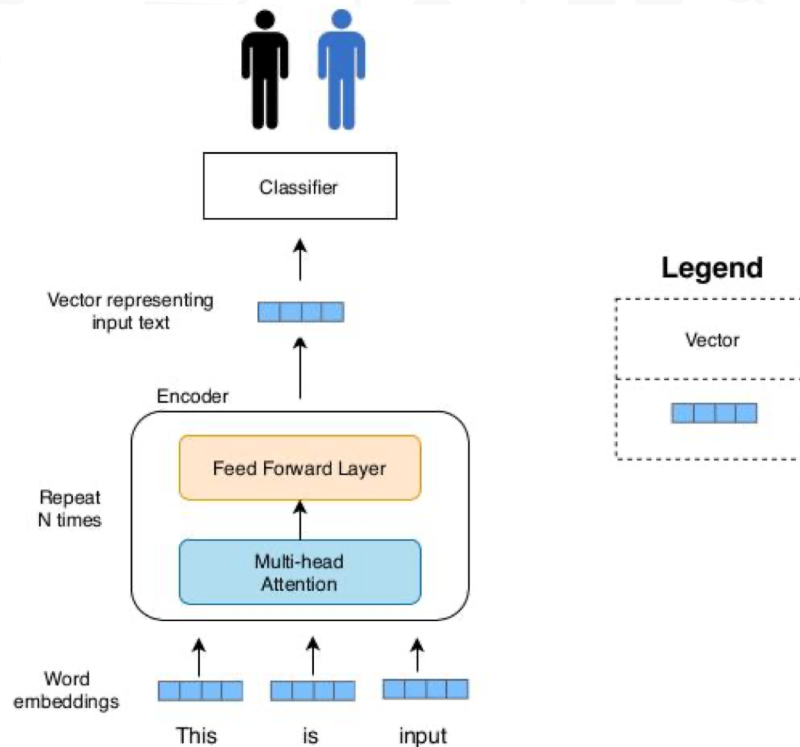
Wektory zanurzeń zdań: Universal Sentence Encoder (USE)

- USE koduje tekst w wielowymiarowe wektory, które mogą być używane do klasyfikacji tekstu, podobieństwa semantycznego, grupowania i innych zadań języka naturalnego.
- Jest trenowany w zakresie różnych źródeł danych i różnorodnych zadań w celu dostosowania do szerokiego zakresu zadań związanych ze zrozumieniem języka naturalnego.
- Dane wejściowe to tekst o zmiennej długości, a dane wyjściowe to 512-wymiarowy wektor.
- Model jest uczony i optymalizowany pod kątem tekstu dłuższego niż słowo, takiego jak zdania, frazy lub krótkie akapity.
- Architektura to Deep Averaging Network (Iyyer et al. 2015) lub transformer (Vaswani et al. 2018).



Universal Sentence Encoder w diagnozie

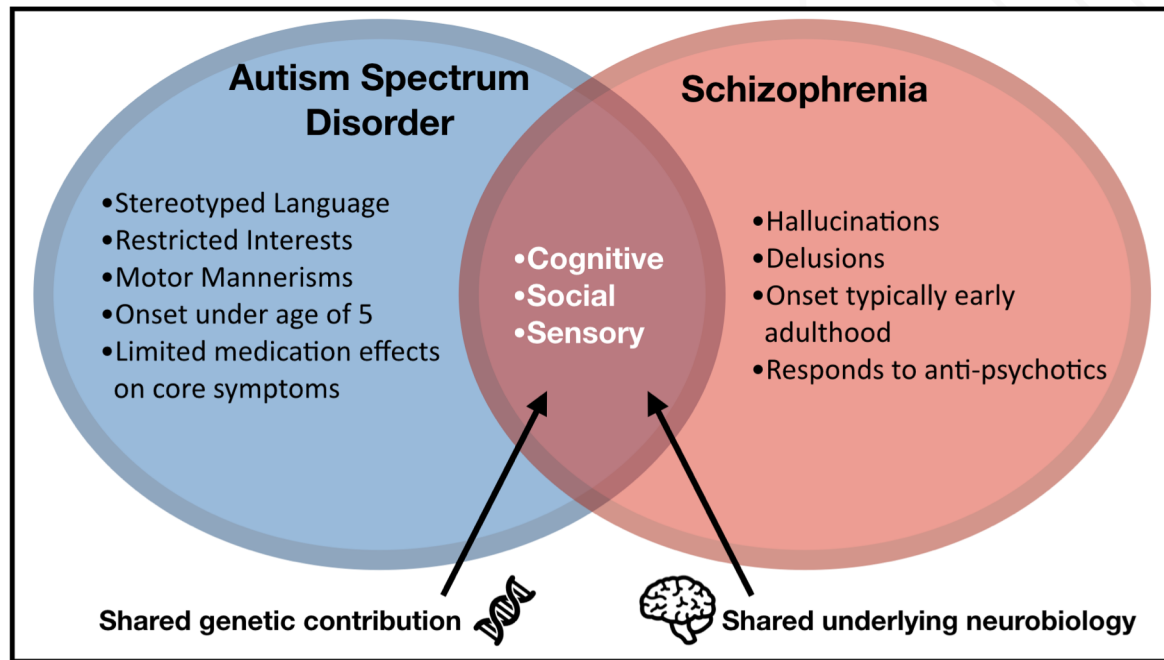
- Uzyskanie całościowego rozwiązania wymaga dołożenia modelu uczenia maszynowego (Classifier).
- Przetestowane zostały popularne algorytmy uczenia maszynowego takie jak:
 - SVM z jądrami liniowym oraz RBF,
 - XG Boost.
- Na obrazku: USE z architekturą transformer.
- Dostępna jest wersja wielojęzyczna, wspierająca m.in. język polski (multilingual USE).



Schizofrenia a autyzm – teoretyczne podstawy dla transfer learning

- Zaburzenia postrzegane jako dwa krańce tego samego kontinuum w zakresie funkcji/procesów poznawczych (wiele cech tych zaburzeń ma przeciwne fenotypy – w schizofrenii mamy nadmiar, a w autyzmie niedobór):
 - Poznanie społeczne: kontakt wzrokowy, sprawczość, mentalizowanie
 - Komunikacja werbalna i język
- Język – objawy FDT:
 - W schizofrenii dominują objawy pozytywne (natłok mowy, rozkojarzenie, brak spójności itd.), ale negatywne również mogą się pojawić
 - W autyzmie dominują objawy negatywne (ubóstwo mowy, ubóstwo treści, persewercje), natężenie objawów pozytywnych jest takie samo jak w przypadku osób zdrowych
- Język – inne objawy:
 - Wydźwięk (sentyment)
 - Mniej spójne i powiązane przyczynowo narracje (ale z różnych powodów)

Schizofrenia a autyzm – teoretyczne podstawy dla transfer learning



Transfer learning: realizacja

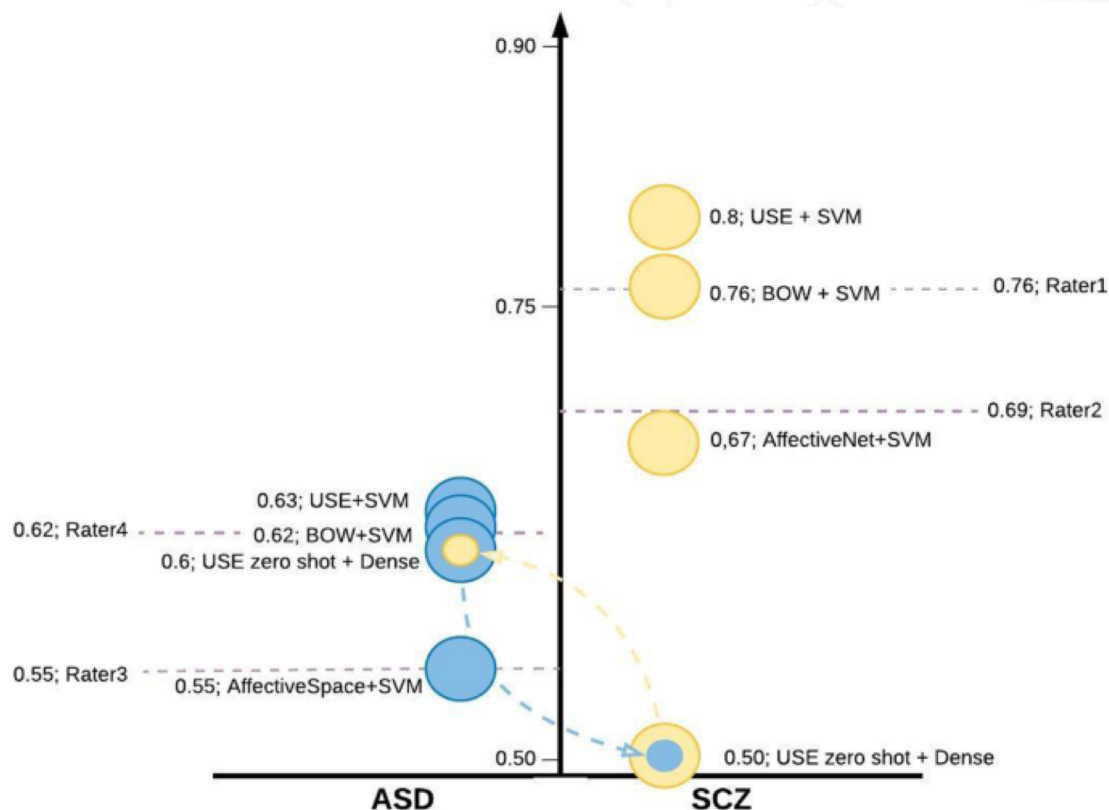
- Wykorzystaliśmy sieć USE, dokładając do niego warstwę gęstą (Dense, Fully Connected) liczącą 64 neuronów, po niej warstwy dropout i klasyfikacyjna (sigmoid).
- **Zero-shot ([0-shot])**. Trenujemy model na jednym zbiorze danych i testujemy go na drugim.
- **Transfer learning ([pre-train])**:
 - Trenujemy na całym zbiorze źródłowym (np. SCZ), a następnie losowo dzielimy dane docelowe (np. ASD) na dwa podzbiory, treningowy (70%) i testowy (30%).
 - Trenujemy na danych źródłowych tak długo, jak tylko strata (loss) zmniejsza się na docelowym podzbiorze treningowym. Celem jest uniknięcie nadmiernego dopasowania do danych źródłowych.
 - Kontynuujemy uczenie na docelowym podzbiorze treningowym przez 10 epok.
 - Obliczamy jakość predykcji na podzbiorze testowym docelowym.
 - Powtarzamy 10 razy i uśredniamy wyniki.

Rozpoznawanie zaburzeń w kilku scenariuszach

		SCZ			ASD		
SINGLE DOMAIN METHODS							
Representation	classifier	acc.	spec.	sens.	acc.	spec.	sens.
AffectNet (EN)	SVM	0.627	0.617	0.638	0.486	0.486	0.486
	XG Boost	0.617	0.574	0.660	0.419	0.459	0.378
AffectiveSpace (EN)	SVM	0.585	0.617	0.553	0.554	0.622	0.486
	XG Boost	0.532	0.532	0.532	0.310	0.378	0.243
Bag of Words	SVM	0.766	0.745	0.788	0.622	0.568	0.676
	XG Boost	0.670	0.681	0.660	0.466	0.540	0.351
POS + LCM + sentiment	SVM	0.606	0.425	0.787	0.459	0.351	0.568
	XG Boost	0.681	0.659	0.702	0.513	0.540	0.486
USE	SVM	0.808	0.830	0.787	0.635	0.649	0.622
	XG Boost	0.713	0.723	0.568	0.568	0.568	0.568
USE	Dense	0.758	0.729	0.813	0.582	0.443	0.769
TRANSFER LEARNING METHODS							
USE	Dense [0-shot]	0.521	0.085	0.957	0.608	0.730	0.486
USE	Dense [pre-train]	0.741	0.776	0.731	0.526	0.482	0.630

Miary to dokładność (acc.), swoistość (spec., $TP/(TP+FP)$) i czułość (sens., $TN/(TN+FP)$). SINGLE DOMAIN METHODS to średnie w walidacji krzyżowej leave-one-out na pojedynczym zbiorze danych. TRANSFER LEARNING METHODS jak na poprzednim slajdzie.

Rozpoznawanie zaburzeń w kilku scenariuszach



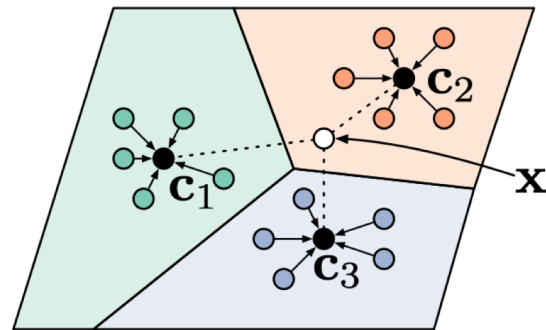
Few-Shot: Prototypical Networks

- Few-shot to nauka z zaledwie kilku przykładów treningowych.
- Dzielimy dane (próbkę) na zbiory *support* (S) oraz *query* (Q).
- *“Nasze podejście opiera się na założeniu, że istnieje zanurzenie, w którym punkty skupiają się wokół pojedynczego prototypu – reprezentacji dla każdej klasy.”*
- Prototypowe wektory zanurzeń obliczane są jako średnia:

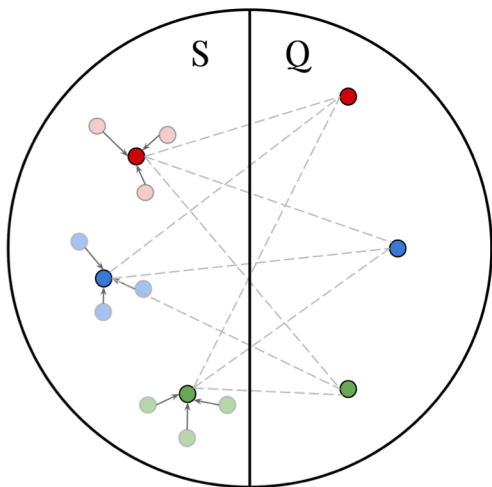
$$\mathbf{c}_k = (1/|S_k|) \cdot \sum_{(\mathbf{s}_i, y_k) \in S_k} f_{\theta}(\mathbf{s}_i)$$

gdzie $\mathbf{c}_k$ jest prototypem k -tej klasy, f_{θ} jest funkcją zanurzenia, S są obserwacjami *support*.

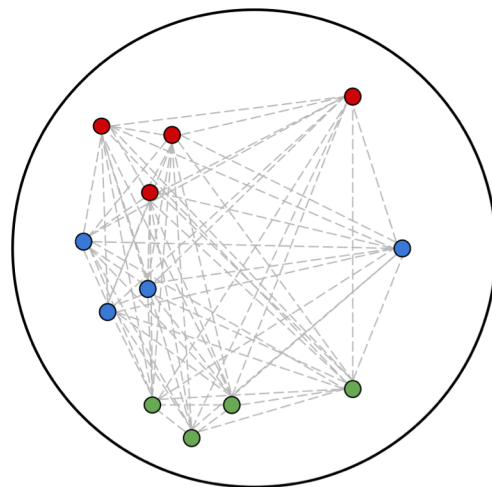
- Następnie obliczamy odległość euklidesową między każdą obserwacją zbioru *queries* (Q) a prototypami.
- Z tego wyliczany jest softmax, prawdopodobieństwo przynależności do każdej klasy.
- Uczenie modelu to minimalizacja funkcji straty (loss) ujemnej log-softmax.



Few-Shot: Prototypical Networks



Prototypical Networks (2017)



Neighbourhood Component Analysis (2005)

$$\mathcal{L}_{\text{proto}}(S, Q) = -\frac{1}{|Q|} \sum_{(\mathbf{q}_i, y_i) \in Q} \log \left(\frac{\exp -\|f_{\theta}(\mathbf{q}_i) - \mathbf{c}_{y_i}\|^2}{\sum_{k'} \exp -\|f_{\theta}(\mathbf{q}_i) - \mathbf{c}_{k'}\|^2} \right).$$

Few-Shot: Prototypical Networks

- Scenariusz to 2-way (k-way) gdzie parametr n (n-shot) określa nt w tabeli. Wyniki few-shot są uśrednione z 5 eksperymentów.
- Baseline: uśrednione 10 razy wyniki z domyślnym uczeniem sieci neuronowej Dense (wektory USE)
- Rozmiar losowych zbiorów treningowych to nt przypadków, zbiory testowe to 20 obserwacji.

	SCZ			ASD		
	Prototypical Networks		Dense baseline	Prototypical Networks		Dense baseline
nt	qt=4	qt=8	N/A	qt=4	qt=8	N/A
2	0.53	0.525	0.505	0.497	0.512	0.545
10	0.586	0.561	0.545	0.476	0.463	0.502
20	0.609	0.591	0.557	0.468	0.453	0.510
30	0.646	0.624	0.596	0.5	0.448*	0.518

Parametr nt – liczba obserwacji treningowych, qt – liczba obserwacji query.

* – problemy z trenowaniem modelu.

Podsumowanie eksperymentów

- Najlepsze metody automatyczne przewyższają jakością ludzkie oceny (psychiatrów).
- Najwyższe wyniki uzyskuje hybryda głębokiego oraz „klasycznego” uczenia maszynowego:
 - głębokie sieci (enkodery) przekształcają teksty do postaci wektorów zanurzeń wypowiedzi,
 - wektory te są potem klasyfikowane przez dobrze znane metody typu SVM.
- Podejścia oparte o transfer learning okazały się przydatne, prawdopodobnie ze względu na wspólne podłoże obu zaburzeń.
- Korzyści wykazano jednak tylko w przypadku SCZ do ASD.
- Uczenie typu few-shot wykazało obiecujące wyniki na zbiorze danych SCZ.

Uczenie maszynowe w psychiatrii

- Brak obiektywnych (pozbawionych czynnika ludzkiego) metod, potrzeba dostarczenia psychiatrom i psychologom klinicznym nowych narzędzi wspomagających w diagnostyce.
- Potrzeba badań przesiewowych na dużą skalę.
- Rola wczesnego wykrywania zaburzeń.
- Słaby dostęp do psychiatrów w Polsce.
- Możliwość monitorowania stanu pacjentów na odległość.

Bibliografia

1. Aleksander Wawer, Izabela Chojnicka, Justyna Sarzyńska-Wawer, Łukasz Okruszek. Single and Cross-Disorder Detection for Autism and Schizophrenia. Cognitive Computation <https://doi.org/10.1007/s12559-021-09834-9> 2021
2. Izabela Chojnicka, Aleksander Wawer. Social language in autism spectrum disorder: A computational analysis of sentiment and linguistic abstraction. PLoS ONE 15(3): e0229985. <https://doi.org/10.1371/journal.pone.0229985> 2020
3. Aleksander Wawer, Izabela Chojnicka. Detecting autism from picture book narratives using deep neural utterance embeddings. [International Journal of Language Disorders – w recenzji]
4. Justyna Sarzyńska-Wawer, Aleksander Wawer, Aleksandra Pawlak; Julia Szymanowska; Izabela Stefaniak, Michał Jarkiewicz, Łukasz Okruszek. Detecting formal thought disorder by deep contextualized word representations. [Psychiatry Research – 2 runda recenzji]