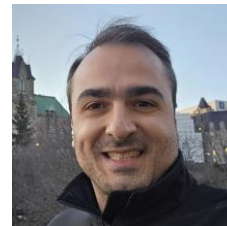


Głębokie, wielopostaciowe uczenie się diagnozowania zaburzeń psychicznych z zapisów wywiadów klinicznych

Stanisław Matwin
Instytut Podstaw Informatyki PAN
Professor Emeritus, Dalhousie University, Kanada

Work with Rudolf Uher, Habibeh Naderi, Behrouz Hajsoleimani



Zarys

- Mowa i tekst w diagnozie psychiatrycznej
- Złożona interakcja
- Modele używające pretrenowanych embeddingów nie wystarczają do modelowania zadań w których język „wielopostaciowy” zmienia się dynamicznie
- Uczenie się reprezentacji wielopostaciowej
- Eksperyment i wyniki

Dane

- 363 OB = 149 MDD, 66 BD, 19 SCHI, 129 O
- Google Speech API
- 17,000 segmentów audio
 - Subiektywne
 - Obiektywne
- 6 ekspertów anodowało te dane
 - Obiektywność/subiektywność
 - Emocje (złość, lek, radość, smutek, brak)
 - Spójność, rozwlekłość, troska, nadopiekuńczość, krytyka
 - Ciepło, nadopiekuńczość, spójność, krytyka: dla całego wywiadu

Multimodal Dataset

- Multimodal corpus of audio speech samples from clinical interviews
 - Source: FORBOW research project, 369 subjects
 - 5 minutes audio samples from Parents
 - 7 minutes (3+2+2) audio samples from Offsprings
- Data Processing:
 - Samples transcribed and broken into multiple segments based on emotional/linguistic cues (emotion, sentiment, objectivity/subjectivity, etc.).
 - Average segment length: ~17 words, ~6.47 seconds audio.
- Labeling:
 - Segments labeled as subjective (opinion, beliefs, thoughts) or objective (facts, observations).
 - 6 multidisciplinary researchers rated segments for various tasks (sentiment, obj/subj, emotion, cohesion, rumination, over-inclusiveness, worry, criticism). High inter-rater agreement in labeling.

Uczenie się reprezentacji tekstu

- Concatenation of
 - Segment-level features: BERT, XLNet, +FastText dla słów spoza słownika
 - Emotion-specific representation:
 - Doc2VecC dla uchwycenia semantyki wywiadu dla słów znaczących i/lub rzadkich
 - Doc2VecC jest trenowany na 21M tweetów

- Ta reprezentacja jest podana sieci LSTM, która uczy się diagnozy
- Ostatnia warstwa tej sieci jest *de facto* reprezentacją segmentu audio **właściwą dla całego wywiadu**
- Używamy jej do uczenia się diagnozy za pomocą RF, SVM, kNN, LDA, QDA, NB
- Jest to „jednopoziomowa” reprezentacja tekstu

„Emocjonalna” reprezentacja tekstu

- Używamy SEmEval-2018 AIT DISC: 100M tweetów
- Uczucia wg Ekmana: 4 z nich

Reprezentacja Audio

- Podobnie do tekstu, ekstrakcja atrybutów dla segmentów (WaveNet)
„emocjonalna” reprezentacja segmentów, konkatencja
- Jak dla tekstu, reprezentacja jest użyta przez sieć do predykcji
i ostatnia warstwa stanowi reprezentację wyjściową
- Jednopoziomowa reprezentacja audio

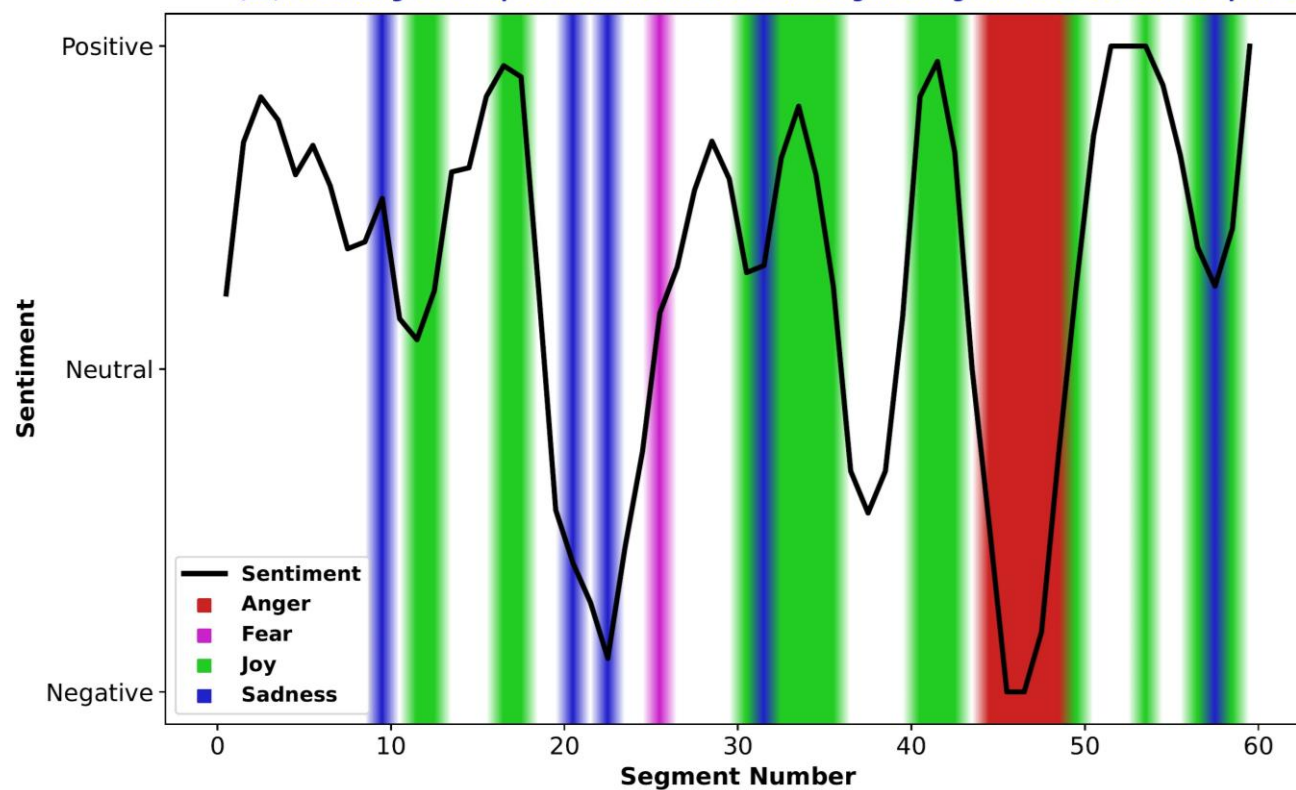
Uczenie się reprezentacji audio

- WaveNet autoencoder
- Reprezentacja spektrogramowa
- Wykorzystanie formatu Mel

Fuzja

- Dla segmentu
- Dla wywiadu:
 - LSTM na obu reprezentacjach ustala wagi dla różnych embeddingów tekstowych i audio w zależności od informacji w używanych/wymawianych słowach

(10) She would prefer if I didn't get involved with things at school because she likes to do her own thing
 (13) she loves to be doing something all the time.
 (26) I have to sort of pull it out of her, she doesn't volunteer it
 (36) She's in French immersion, and she's pretty much bilingual now.
 (45) She can be moody,
 (48) so it gets a little frustrating
 (58) She does get a bit upset when she doesn't make the goal though and it does affect her quite a bit.



[redacted] is a very complex child.
She has a wonderful personality.
She's very honest. She's generous. She has lots of talents.
She likes to bake [redacted]
and she's very artistic, very respectful.
[redacted] completed grade 12 last year.
What we're having our struggles right now with her trying to find a job.
A lot of it is because of mental illness.
So we're struggling with that.
[redacted] has five anxiety disorders
and it has been challenging right from oh, about as far as I can remember from maybe age 2.
[redacted] started with selective mutism going to school and she didn't speak until Grade 7 and then she has social anxiety,
So it's not easy dealing with [redacted] because I always found that trying to get her to do something was like pulling teeth.
I I've had to do everything with her step by step in step ladders and we're still working and moving forward.
It's really difficult trying to find her a job somewhere because
[redacted] also has a math disability.
So her learning of math I don't think exceeded the grade four level.
So she wouldn't be able to work somewhere on cash.
And the other thing is is it's hard for her to communicate with people.
So she'd have to have a job behind the scenes.
So we're gonna, we're gonna have to work on that.
Yeah, I've had a lot of struggles with [redacted].
And I just take it day by day.
Mhm.
[redacted] has a wonderful love of animals.
And [redacted] receives a lot of relaxation and peace through animals.
And she gets along fairly well with her sisters.
They do have their squabbles but for the most part they do get along.
Is there, how much time is left?

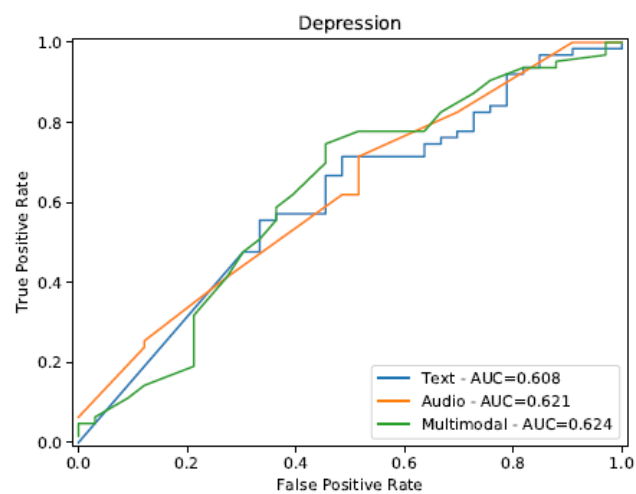
Segmenty,
w których rodzic
mówi o niepokoju
dziecka i jego
trudnościach
komunikacyjnych
mają wyższe wagi

6

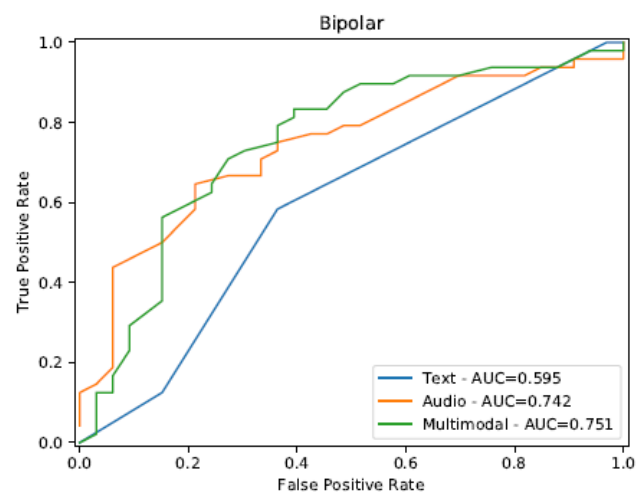
Figure 4: A random sample from subjects with depression. Each line shows a segment and they are colored based on the attention weights learned in our attention-based LSTM model. Darker colors mean the model is paying more attention to those segments for the final recognition (patient's name is replaced with blank for anonymity).

Experimental Setup

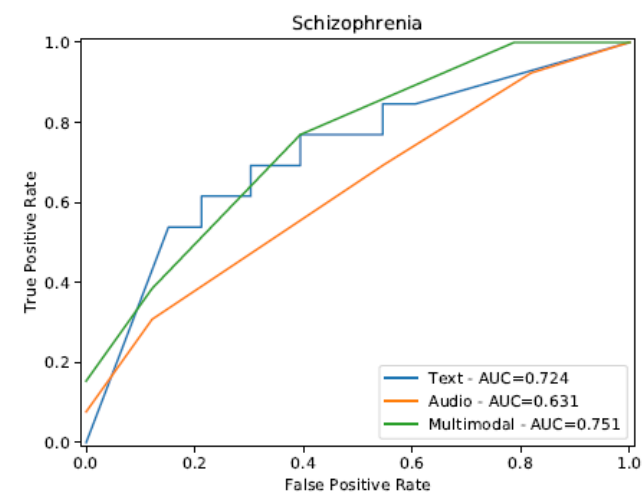
- Kwestia LOBO
- Dane niezrównoważone – GK 129, MD 149, BD 66, SCHI 19;
random oversampling



(a) Depression



(b) Bipolar



(c) Schizophrenia

Conclusion & Future Works

- Skuteczne zastosowanie jako POMOC diagnostyczna
- Dalsze prace: dane od dzieci
- Contrastive learning w miejsce transfer learning

Thank you



Results

- Sztuczna Inteligencja: na serio i osobiście
- SGH Aula I; 13:10, 9:00