

Eksperyment 4 - Late fusion

Hipoteza

Sieć rekurencyjna LSTM dobrze sobie poradziła w eksperymencie 3, gdzie uczyła się na tylko na części danych (osobno EEG i dane peryferyjne), więc zakładamy, że jak połączymy odpowiedź obu modeli wytrenowanych na różnych danych, to powinniśmy uzyskać lepszy wynik.

Sposób trenowania

Do trenowania wykorzystaliśmy rekurencyjną (LSTM) sieć głęboką. W tym eksperymencie dzielimy zbiór uczący na dane EEG i z urządzeń peryferyjnych, a następnie uczymy każdy z tych dwóch zbiorów danych osobno. W ostatnim kroku łączymy odpowiedź obu modeli, aby uzyskać ostateczną odpowiedź. Przykładowo model wytrenowany na danych EEG dla zadanej próbki danych da wynik 0.55 (>0.5 - emocja wykryta), a drugi model da wynik 0.4 (<0.5 - emocja niewykryta). Ostateczny wynik to $(0.55 + 0.4) / 2 = 0.475$, czyli emocja niewykryta.

Zaimplementowana sieć (każdy z modeli) składa się z 80 neuronów LSTM oraz jednego neuronu wyjściowego, która odpowiada na pytanie czy uczona emocja została wykryta.

Trenowanie każdego modelu następuje w batchach o wielkości 128 próbek.

Sieć została zaimplementowana w Kerasie i trenowana przy użyciu GPU.

Sposób testowania

Testowane jest 6 klasyfikatorów (po jednym dla każdej pary zbiorów uczący - uczucie). Każdy klasyfikator jest testowany za pomocą walidacji skrośnej z podziałem na 5 podzbiorów. Każdy podzbiór pełni raz rolę zbioru testującego, a reszta zbiorów pełni w tym czasie rolę zbiorów trenujących. Dla każdego takiego etapu wyliczamy accuracy i F1-score. Następnie liczy średnią i odchylenie standardowe ze wszystkich 5 etapów walidacji skrośnej. Z racji tego, że uczymy tutaj tak naprawdę dwie sieci rekurencyjne musimy mieć zgodność pomiędzy testowanymi skrośnie parami.

Porównanie wyników

Pochodzenie danych	Arousal		Valence		Liking	
	ACC	F1	ACC	F1	ACC	F1

Artykuł	-	0.618	-	0.608	-	0.634
Early fusion	52.89% (+/- 4.27%)	60.07% (+/- 5.83%)	53.59% (+/- 3.71%)	63.82% (+/- 3.74%)	60.55% (+/-5.26%)	72.69% (+/- 4.72%)
Late fusion	52.03% (+/- 2.46%)	52.62% (+/- 6.52%)	51.95% (+/- 3.49%)	55.76% (+/- 6.48%)	60.78% (+/-6.56%)	71.07% (+/- 8.10%)

Wnioski

Porównanie wyników jest nie do końca miarodajne, ponieważ autorzy artykułu nie podali miary "accuracy" dla wyników końcowych. Dodatkowo użyli klasyfikatora wyuczonego bezpośrednio na pokazywanych filmikach, czy my nie robiliśmy. Wyniki są wyraźnie gorsze od metody "Early fusion" i nie licząc Liking gorsze od wyników z artykułu.