

EEG-based Emotion Recognition

Grazia Margarella
0522501448

Curriculum Data Science & Machine Learning

Nicola Pio Santorsa
0522501434

Curriculum Data Science & Machine Learning

Index Terms—EEG, PSD, DE, ASM, DASM, Emotion Recognition, SEED-IV, Subject-Dependent, Subject-Independent, Subject-Biased

Abstract—Nella seguente relazione è trattato l'ambito dell'emotion recognition tramite l'uso di segnali EEG. Il dataset di riferimento è SEED-IV il quale contiene registrazioni di segnali EEG e movimenti oculari di 15 soggetti sottoposti alla visione di 72 video suddivisi in 3 sessioni. Lo scopo di questo lavoro è di predire le 4 emozioni classificate dal dataset di riferimento (Felicità, Neutralità, Tristezza e Paura) utilizzando tecniche di Machine Learning e Deep Learning.

I. INTRODUZIONE

L'ambito della biometria cognitiva è un settore emergente della biometria che analizza i segnali cerebrali con lo scopo di effettuare il riconoscimento di individui oppure il riconoscimento degli stati cognitivi ed emotivi di un soggetto. In questo report tratteremo del secondo tipo di riconoscimento, in particolare per quanto riguarda le emozioni. I segnali cerebrali sono misurati tramite elettroencefalogramma (riferito in seguito con la sigla EEG) che acquisisce tali segnali tramite elettrodi posizionati sul cuoio capelluto secondo diversi standard internazionali. Il dataset di riferimento è SEED-IV, descritto da Zheng et al. in [1], ideato con lo scopo di permettere la classificazione di 4 emozioni: felicità, neutralità, tristezza e paura. Esso è stato generato tramite 3 sessioni di acquisizione su 15 soggetti ai quali sono stati applicati 62 elettrodi secondo il Sistema Internazionale 10-20. A questi soggetti sono stati sottoposti gli stimoli dati da 24 video in ogni sessione ai quali hanno assegnato una delle 4 emozioni una volta terminata la visione. Come altro dato a supporto sono stati raccolti anche i movimenti oculari tramite degli occhiali per l'eye tracking. Il processo di analisi dei dati raccolti si sviluppa in diverse fasi:

- *Preprocessing*: in particolare per quanto riguarda l'artifact filtering e la noise reduction;
- *Feature Extraction*: ossia la selezione delle feature rilevanti per la classificazione;
- *Task di classificazione*: tramite l'utilizzo di modelli di machine learning e deep learning.

Le tecniche utilizzate in letteratura per quanto riguarda queste fasi sono molteplici e sono state analizzate nella sezione successiva relativa allo stato dell'arte. Nelle sezioni successive di questo report sono descritte inoltre la metodologia applicata per analizzare i segnali EEG e i protocolli di split dei dati in train e test, i risultati ottenuti, la loro analisi e le dovute conclusioni sugli eventuali miglioramenti ottenuti.

II. STATO DELL'ARTE

A. Dataset EEG

In letteratura, sono presenti vari dataset contenenti segnali EEG utilizzati nell'ambito dell'emotion recognition. I due più importanti sono *DEAP* [2] e *SEED-IV* [1].

Il Dataset *DEAP* è composto non solo da segnali EEG, bensì anche da segnali ECG, ENG e registrazioni del volto di 32 soggetti. Per ogni soggetto sono presenti i dati registrati durante la visione di 40 video musicali. Il dataset dispone sia dei dati grezzi (*raw*) che di quelli pre-processati tramite la feature *Differential-Entropy* (*DE*), con successivo smoothing tramite *Moving Average* sulle 5 bande di frequenze del segnale utilizzando un time-step di 1 secondo senza overlap.

Il Dataset *SEED-IV* è stato costruito a partire dai dati collezionati da 15 soggetti. Questi sono stati sottoposti alla visione di 72 video divisi per 3 sessioni. Per ogni video sono stati registrati i segnali EEG utilizzando un caschetto con 62 elettrodi, che acquisiscono con una frequenza di campionamento di 1000Hz, posizionati in base al Sistema Internazionale 10-20. SEED contiene anche i movimenti oculari del soggetto, catturati tramite un occhiale adibito all'eye tracking. Il dataset dispone sia dei dati raw che dei dati pre-processati, per entrambe le tipologie. Per i segnali EEG, il pre-processing avviene tramite l'estrazione delle feature *Differential-Entropy* (*DE*) e *PSD* (*Power Spectral Density*), su cui è stato poi effettuato lo smoothing o con la *Moving Average* oppure con *Linear Dynamic System* (*LDS*). Per i dati oculari invece il preprocessing viene effettuato applicando una tecnica della riduzione della dimensionalità quale la *Principal Component Analysis* (*PCA*).

B. Channel Selection

Il cervello umano è suddiviso in diverse zone, ognuna delle quali ha diversi compiti, per cui è facile intuire che ognuna di queste zone generi un segnale EEG leggermente diverso dalle altre. Per poter registrare questi segnali vengono posizionati quindi degli elettrodi e vengono mappati come in figura 1. Alcuni dataset come SEED-IV decidono di registrare segnali con 62 sensori 1, mentre altri come DEAP registrano segnali con 32 sensori. Come è stato analizzato già da [3] [4], i sensori che contribuiscono maggiormente all'emotion recognition sono 6, in particolare **FT7, FT8, T7, T8, TP7, TP8**. Altri come Jin et al. [5] preferiscono non effettuare nessuna selezione dei canali.

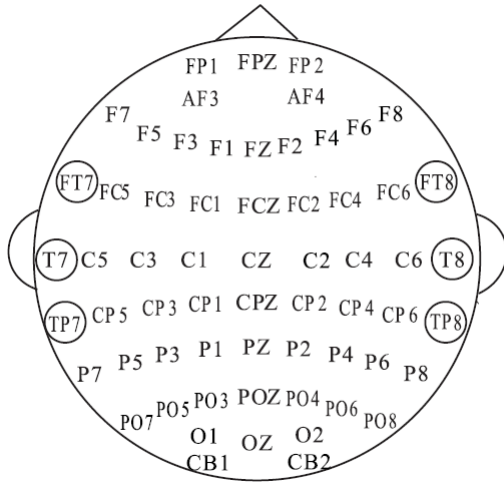


Fig. 1. Mappa Sensori SEED-IV

C. PreProcessing

Il segnale raw in sè non risulta efficace per una predizione ottima con algoritmi di intelligenza artificiale, dato che in esso è presente del rumore e degli artefatti. Per rimuovere questi sono diversi gli approcci che sono stati applicati in letteratura. In [1] i dati vengono semplicemente filtrati con un filtro passa banda tra 1 e 75Hz, in [5] invece l'approccio scelto è stato quello di filtrare con un filtro passa banda tra 0 e 75Hz e successivamente effettuare downsampling del segnale a 200Hz. Principalmente i vari approcci che sono stati utilizzati da altri autori si differenziano unicamente nella scelta delle frequenze per il filtro e il downsampling [6]. Un'altra scelta che viene effettuata in letteratura è la separazione del segnale in 5 bande, ovvero rispettivamente di *Delta*, *Theta*, *Alpha*, *Gamma* e *Beta*, approccio scelto da [5] [1] e da [7] che utilizza come dataset DEAP e valuta questo approccio su più classi di emozioni. In queste ricerche è stata utilizzata come metrica di comparazione l'accuratezza ottenuta da diversi modelli di machine learning e deep learning tramite l'utilizzo di feature diverse su diverse bande. Da essi è possibile notare che si verifica un miglioramento significativo dell'accuracy quando si opera sul segnale senza suddivisione rispetto alle singole bande.

D. EEG Features

Le feature estraibili da segnali EEG sono molteplici, ma quelle che risultano essere più utilizzate in studi di questo tipo sono la **Differential Entropy (DE)** e la **Power Spectral Density (PSD)**. Entrambe vengono utilizzate in studi come [1] [8] [9]. Ulteriori feature analizzate in questo report sono la **DASM** e la **ASM**, le quali sono feature che vanno a considerare la simmetria e l'asimmetria di segnali EEG registrati da sensori paralleli (come ad esempio FT7 e FT8) e sono stati utilizzati in ricerche come [9]. Successivamente all'estrazione delle feature alcuni studi come [1] applicano tecniche come la **Linear Dynamic System (LDS)** [10] o la **Moving Average** per ridurre

ulteriormente il rumore e gli artefatti nei segnali risultanti. In alcuni casi, come per [5], viene ritenuto opportuna l'analisi dei segnali senza estrarre nessuna feature e quindi fare predizioni solo sui dati pre-processati.

III. METODOLOGIA UTILIZZATA

A. Dataset

Il Dataset preso in analisi è **SEED-IV**. I dati utilizzati sono i segnali EEG raw contenuti nei file *.mat* forniti dagli autori del dataset [1], sui quali sono state effettuate le fasi di pre-processing e di estrazione delle feature.

B. PreProcessing

I segnali raw vengono ridotti prendendo in esame unicamente i dati provenienti dai sensori evidenziati come più rilevanti per il task di classificazione delle emozioni, ovvero **FT7, FT8, T7, T8, TP7, TP8**. Successivamente si applica un downsampling a 200Hz per ridurre il rumore presente nel segnale ed infine un filtro passa banda impostato tra 0.3 e 50Hz per ridurre gli artefatti presenti nel segnale. Sia per il downsampling che per il filtro viene utilizzata la libreria *scipy.signal*.

C. Feature Extraction

Dai segnali preprocessati vengono poi estratte le feature **DE, PSD, ASM, DASM**. Rispettivamente la PSD è calcolata tramite il metodo *signal.welch* di *scipy* segmentando a 4 secondi, mentre la DE è calcolata tramite il metodo *stats.differential_entropy* di *scipy*. Le altre due feature sono state determinate invece a partire dalla PSD con l'utilizzo di due formule. Per la DASM:

$$\frac{|psd_left_hemisphere - psd_right_hemisphere|}{psd_left_hemisphere + psd_right_hemisphere}$$

Per l'ASM:

$$\frac{psd_left_hemisphere - psd_right_hemisphere}{psd_left_hemisphere + psd_right_hemisphere}$$

Per tutte e 4 le feature estratte, è stato ritenuto opportuno non effettuare alcuna suddivisione in bande, come supportato da [7], per poter confrontare questo approccio con quelli che invece operano questa distinzione.

D. Modelli

Per effettuare il task di classificazione sui segnali EEG sono stati utilizzati vari modelli con prestazioni diverse. I più rilevanti sono stati l'*HistGradientBoostClassifier* e la rete neurale *Bi-Gru*.

1) *HistGradientBoost*: L'*HistGradientBoost* è un metodo di ensemble learning basato su decision tree particolarmente adatto per dataset molto grandi con la presenza di valori nulli, come nel nostro caso, per i quali propone una strategia di recovery. Infatti per quanto riguarda lo split su dati nulli va a dedurre la classe di appartenenza tramite gli altri dati presenti, seguendo una logica a maggioranza. Una sua implementazione è proposta dalla libreria *sklearn*.

2) *Bi-GRU*: La GRU è una Recurrent Neural Network utilizzata per apprendere su flussi di informazioni nel tempo mantenendo uno stato interno di memoria degli istanti precedenti. Questa rete è stata preferita ad una *LSTM*, precedentemente testata, dato che la GRU risulta più efficiente dal punto di vista computazionale e riporta risultati migliori in alcuni scenari simili a quello preso in esame. La nostra scelta però non è ricaduta sull'uso di una singola rete GRU, bensì di due (da qui il nome Bi-Gru) che vanno ad analizzare il nostro segnale **bidirezionalmente** in modo che possa cercare dei pattern sia in forward che in backward.

E. Setup Esperimento

Per quanto riguarda la suddivisione dei dati tra train e test per i modelli selezionati sono stati adottati i tre protocolli descritti in [8], e sono:

- **Subject-Dependent**: vengono selezionati casualmente 16 video dal dataset i quali sono assegnati al training set, mentre i rimanenti 8 per il test set, accertandosi che nel test ci sia una rappresentazione equa delle label (ovvero che ne siano presenti 2 di ogni tipo), ciò per ogni sessione. Il risultato finale viene determinato da una media aritmetica tra le 3 sessioni.
- **Subject-Independent**: viene adottato l'approccio *Leave-One-Subject-Out* che consiste nel considerare un solo soggetto come istanza di test e gli altri 14 come istanze di train utilizzando la **cross-validation**, ciò per ogni sessione. Il risultato finale viene determinato da una media aritmetica tra le 3 sessioni.
- **Subject-Biased**: viene effettuata una suddivisione casuale del dataset in cui l'80% del dataset viene utilizzato per il training set e il restante 20% per il test set.

IV. ANALISI DEI RISULTATI

Nel seguente capitolo andiamo ad analizzare i risultati generati dai modelli proposti nel capitolo precedente applicati ai seguenti dataset:

- Dataset Preprocessato : Dataset che è stato preprocessato utilizzando le tecniche descritte nel capitolo precedente con tutte le feature presentate;
- Dataset PreProcessato DE : Dataset che è stato preprocessato utilizzando le tecniche descritte nel capitolo precedente con la sola feature DE standardizzata;
- Dataset DE SeedIV : Dataset preprocessato da [1] con la sola feature DE a cui è stato effettuato lo smoothing con una tecnica di LDS.

I dataset presentati contengono valori nulli dati dalla lunghezza variabile dei video presi in esame. Questo tipo di problematica è da tenere in considerazione nello sviluppo dei modelli di ML e DL.

Ognuno di questi dataset è stato suddiviso utilizzando i 3 protocolli visti in [8] e nel capitolo precedente, ovvero: **Subject-Dependent**, **Subject-independent** e **Subject-Biased**.

Per valutare l'efficacia dei nostri modelli essi sono comparati con il modello **RODAN** sviluppato da [8].¹

A. Subject-Dependent

TABLE I
ACCURACY SUBJECT DEPENDENT

	PrePro. Dataset	PrePro. DE	DE SEEDIV
HistGradientBoost	36%	32%	52%
Bi-Gru	*	20% ± 5	35%
RODAN	-	-	70% ± 9.7

Come si può vedere dalla I i nostri modelli non sono stati efficaci quanto RODAN che è riuscito a raggiungere il 70% di accuracy, nonostante ciò possiamo osservare il 52% del HistGradientBoost. Non risultano essere significative le statistiche ottenute del Dataset Preprocessato elaborato con l'approccio proposto che riesce a garantire unicamente il 36%.

B. Subject-Independent

TABLE II
ACCURACY SUBJECT INDEPENDENT

	PrePro. Dataset	PrePro. DE	DE SEEDIV
HistGradientBoost	88% ± 4	87% ± 4	94%
Bi-Gru	*	*	75% ± 15
RODAN	-	-	60.75% ± 10

Come si può vedere da II entrambi i nostri modelli si sono rivelati migliori di RODAN con 94% con l'algoritmo HistGradientBoost, e il 75% ± 15 per il Bi-GRU. Sono da notare inoltre gli score ottenuti sui dati preprocessati con l'approccio proposto che risultano essere significativamente alti.

C. Subject-Biased

TABLE III
ACCURACY SUBJECT BIASED

	PrePro. Dataset	PrePro. DE	DE SEEDIV
HistGradientBoost	71%	65%	97%
Bi-Gru	*	33%	40%
RODAN	-	-	98.15% ± 0.2

Come si può vedere da III i modelli proposti non sono efficaci quanto RODAN, nonostante sia da sottolineare quanto il modello di HistGradientBoost sia molto vicino alle performance di riferimento pari al 98% con un accuracy del 97%. Per quanto riguarda il modello Bi-Gru, esso non è riuscito a raggiungere il 50% di accuracy con nessun dataset.

¹Nelle tabelle successive dove viene posto l'asterisco (*) non è stato possibile addestrare i modelli sul dataset dal momento che la loro grande dimensionalità rende il tempo di esecuzione elevato. Prima di interrompere l'esecuzione abbiamo lasciato eseguire per circa 2 ore per poi osservare che l'accuracy non superava il 40%.

V. CONCLUSIONI

Per quanto riguarda le tecniche di preprocessing proposte, come discusso nel capitolo precedente, esse non hanno incrementato le performance dei modelli, anzi in alcuni casi come per l'approccio I hanno portato anche a prestazioni peggiori. Come aspetti positivi possiamo però notare alcuni score di accuratezza elevati in II e III, inferiori a quelle di riferimento ma comunque rilevanti. Per quanto riguarda i modelli selezionati per il confronto con RODAN, l'unico che ha avuto prestazioni comparabili e a volte anche superiori(II) è l'HistGradientBoost. Il Bi-Gru invece si è rivelato un modello non ottimale nei casi I e III.

Come sviluppi futuri a questo approccio potrebbe essere fruttuoso utilizzare anche il dataset DEAP per poter confrontare i modelli e gli approcci anche su dati diversi per verificarne l'efficacia.

REFERENCES

- [1] Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE transactions on cybernetics*, 49(3):1110–1122, March 2019.
- [2] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2012.
- [3] Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, 7:1–1, 09 2015.
- [4] Dan Nie, Xiao-Wei Wang, Li-Chen Shi, and Bao-Liang Lu. Eeg-based emotion recognition during watching movies. pages 667 – 670, 06 2011.
- [5] Longbin Jin and Eun Yi Kim. Interpretable cross-subject eeg-based emotion recognition using channel-wise features. *Sensors*, 20(23), 2020.
- [6] Badajena, J.C., Sethi, S., Dash, and S.K. A survey on eeg-based neurophysiological research for emotion recognition. 2023.
- [7] Md Zaved Iqbal Ahmed, Nidul Sinha, Souvik Phadikar, and Ebrahim Ghaderpour. Automated feature extraction on asmap for emotion classification using eeg. *Sensors*, 22(6):2346, 2022.
- [8] Wai-Cheong Lincoln Lew, Di Wang, Katsiaryna Shylouskaya, Zhuo Zhang, Joo-Hwee Lim, Kai Keng Ang, and Ah-Hwee Tan. Eeg-based emotion recognition using spatial-temporal representation via bi-gru. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 116–119, 2020.
- [9] Xiang Li, Dawei Song, Peng Zhang, Yazhou Zhang, Yuexian Hou, and Bin Hu. Exploring eeg features in cross-subject emotion recognition. *Frontiers in Neuroscience*, 12:162, 03 2018.
- [10] Li-Chen Shi and Bao-Liang Lu. Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2010:6587–90, 08 2010.
- [11] Rab Nawaz, Hwa Cheah, Humaira Nisar, and Vooi Yap. Comparison of different feature extraction methods for eeg-based emotion recognition. *Biocybernetics and Biomedical Engineering*, 40, 05 2020.