Deep Physiological Affect Network for the Recognition of Human Emotions
- *Analisi Teorica* -

Nicolò Cavedoni

December 12, 2020

1 Introduzione

Lo studio di Byung Hyung Kim e Sungho Jo si basa sul riconoscimento delle emozioni tramite l'uso di un'intelligenza artificiale chiamata DPAN. Questa analizza i segnali provenienti da un elettroencefalogramma e da un fotopletismogramma, classificandoli poi in uno degli stati emotivi descritti nella teoria dimensionale di James Russell e Lisa Feldman Barrett. La DPAN è una rete neurale convoluzionale a memoria a breve termine, e viene addestrata sfruttando il dataset DEAP e applicando una nuova funzione di costo basata sul margine temporale. L'esperimento si basa sul concetto di *lateralizzazione emotiva*, e studia le possibili correlazioni tra l'attivazione asimmetrica di alcune aree del cervello e il manifestarsi di precise emozioni nel soggetto. I risultati di questo modello vengono quindi comparati con quelli di Koelstra nello studio che ha portato alla creazione dei DEAP e con quelli di un altro esperimento che però utilizzava una rete neurale molto simile, mostrando dei progressi effettivi nell'accuratezza della valutazione. Al contempo, però, lo studio sulla lateralizzazione non ha portato novità nell'intricato tema delle emozioni, limitandosi a confermare teorie precedenti ma contrastanti fra di loro.

L'obiettivo di questo trattato è quello di esaminare e spiegare tutti i modelli, le teorie e gli argomenti necessari alla piena comprensione del paper, cercando di mantenere il filo del discorso impostato dai due scienziati coreani.

2 Lateralizzazione

Il concetto di *lateralizzazione* denota un'asimmetria nell'attivazione del cervello per svolgere una determinata funzione. In pratica, gli emisferi cerebrali destro e sinistro non si attivano simmetricamente per qualsiasi cosa, ma in molti casi soltanto uno dei due è impegnato per coordinare una qualche funzionalità. La prima ipotesi di lateralizzazione risale al 1865, quando Broca studia un paziente con una lesione al lobo frontale sinistro che aveva perso quasi completamente l'uso del linguaggio [1]. Circa un secolo dopo, i premi Nobel Sperry e Gazzaniga studiano gli effetti della callosotomia nei pazienti effetti da grave epilessia, nominandoli "split brain", scoprendo che le due metà cerebrali sono in grado di lavorare autonomamente (seppure con forti impedimenti) e fortificando l'ipotesi di lateralizzazione [2]. Senza entrare troppo nel dettaglio, si può affermare che l'emisfero sinistro è specializzato nell'area logica-linguistica, mentre quello destro è più attivo quando c'è una sollecitazione emotiva. Questi risultati hanno condotto alla prima teoria di lateralizzazione emotiva, che prevede appunto una dominanza netta dell'emisfero destro nell'espressione e nell'individuazione di emozioni. Alla suddetta "ipotesi dell'emisfero destro" [3] si contrappone però una seconda teoria che prende piede pochi anni dopo, detta "ipotesi della valenza" [4], che sostiene che l'emisfero sinistro abbia un centro per le emozioni positive, mente quello destro per le emozioni negative. Sebbene ci siano diversi studi sulla lateralizzazione emotiva che conducono a risultati diversi, si tende a pensare che ognuno offra una piccola parte di verità e che in realtà siano complementari fra loro. La scoperta dell'asimmetria emisferica nei processi emotivi ha condotto diversi scienziati a sperimentare con elettrodi simmetrici, ottenendo buoni risultati. La ricerca dei due coreani punta quindi a migliorare questi risultati proponendo un approccio Deep Learning e cercando di risolvere (o almeno semplificare) il problema della *soggettività* nell'analisi emotiva. È infatti vero che ognuno di noi reagisce agli eventi in maniera diversa, per via delle esperienze pregresse e del carattere. Inoltre, la stessa emozione può manifestarsi in forme differenti da persona a persona, e tutta questa soggettività è di intralcio nel momento in cui si cercano dei pattern ricorrenti e delle verità generali.

3 Arousal e Valence

Nella sezione precedente abbiamo visto come Davidson abbia usato il termine "valenza"per distinguere le emozioni positive, originanti nell'emisfero sinistro, e quelle negative, dominate dall'emisfero destro. La parola proviene in realtà da uno dei modelli teorici più famosi e ancora oggi utilizzati per cercare di spiegare le emozioni, ossia la teoria del Circomplesso di Russell [5]. Russel propone un modello dimensionale che suddivide le emozioni in una circonferenza ipotetica detta circomplesso, centrata nell'origine di un piano cartesiano con due assi molto particolari: L'asse orizzontale lo chiama *valence* (non sarà più tradotto in "valenza" da qui in poi), e rappresenta la positività di un'emozione; un'emozione con alta valence fa stare bene e a proprio agio il soggetto, laddove una valence

negativa indica uno stato più o meno grande di disagio e malessere. Sull'asse verticale Russell individua l'*arousal*, ovvero l'intensità che si prova durante una determinata emozione: Un valore alto di arousal corrisponde a emozioni agitate (positive o negative che siano) e preponderanti, mentre un valore negativo di arousal indica uno stato di calma e di bassa intensità emotiva.

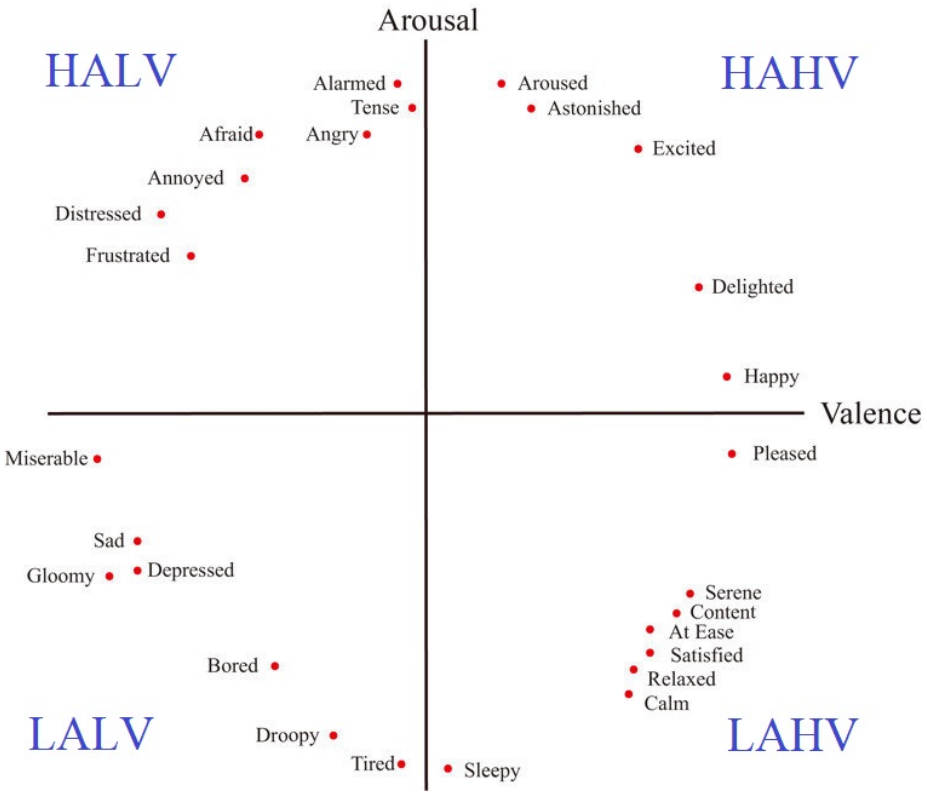


Figure 1: Circomplesso emotivo di Russell, con la denominazione dei quadranti

La teoria di Russell sarà poi raffinata una ventina d'anni dopo in una ricerca con Lisa Feldman Barrett, in cui introdurrà il concetto di episodio emozionale e tutti gli elementi che lo generano; tra questi, il più innovativo è il "Core Affect", lo stato emotivo di base, che rappresenta quella sensazione neurofisiologica comune a tutte le emozioni che percepiamo, difficile da descrivere. Il suo circomplesso si evolve quindi per includere questa nozione e rappresentare al meglio gli stati affettivi di base [6].

4 Segnali Fisiologici, EEG e PPG

I segnali fisiologici rappresentano uno strumento di analisi fondamentale per l'interpretazione delle emozioni, e il machine learning è una tecnica sempre più utilizzata per elaborare questo genere di dati. I segnali più utilizzati sono l'*elettroencefalografia (EEG)* e la *fotopletismografia (PPG)*, e sono anche i due che ci servirà conoscere per capire questo trattato.

L'EEG rileva l'attività elettrica del cervello applicando degli elettrodi sullo scalpo. La tecnologia è incredibilmente affidabile, al punto di essere utilizzata in campo biomedico in molti ambiti, come le interfacce neurali (protesi artificialli o elettrodi controllati soltanto dalle onde cerebrali), il Visual Evoked Potential (diagnostica problemi alla vista misurando il tempo di percorrenza di uno stimolo visivo) o l'immaginazione motoria (consiste nell'immaginare determinati movimenti per avere effetti benefici nella neuro-riabilitazione e nella coordinazione motoria). Come si può notare dalla Figura 2, il segnale EEG è

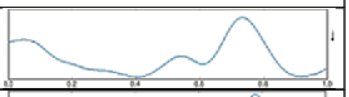
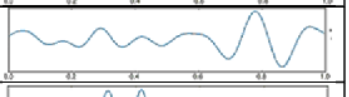
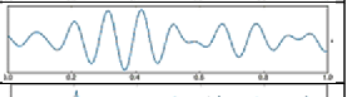
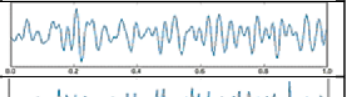
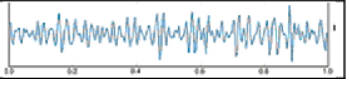
Waves	Frequency bands (Hz)	Behaviour Trait	Signal Waveform
Delta	0.3 – 4	Deep sleep	
Theta	4 – 8	Deep Meditation	
Alpha	8 – 13	Eyes closed, awake	
Beta	13 – 30	Eyes opened, thinking	
Gamma	30 and above	Unifying consciousness	

Figure 2: Le 5 bande di frequenza del segnale EEG

suddiviso in cinque bande di frequenza, generalmente distinte tra frequenze di sonno e di veglia. Poichè le onde Delta sono riscontrabili nei neonati e nelle fasi di sonno, non fanno parte del dataset utilizzato per l'esperimento, ma tutte le altre frequenze sono state analizzate separatamente alla ricerca di correlazioni significative.

La PPG misura le minime variazioni di intensità di un fascio luminoso puntato contro la pelle, che corrispondono alla variazione di volume nei vasi sanguigni conseguente a una pulsazione. È possibile rilevare delle funzionalità nervose autonome solamente dal tempo e dalla frequenza delle pulsazioni, tuttavia servirebbero molte osservazioni su un lungo periodo di tempo, che non è il nostro caso. Nello studio analizzato, il PPG si occupa di rilevare la frequenza cardiaca; per quanto sia meno accurato rispetto a un elettrocardiogramma, è più comodo da indossare e quindi più adatto alla tecnologia di tutti i giorni.

5 Reti Neurali

Le intelligenze artificiali in grado di apprendere (machine learning) sono uno dei sistemi più studiati e utilizzati negli ultimi decenni, anche in campo biomedico. Il concetto di apprendimento diventa "Deep Learning" nel momento in cui la macchina è composta da diversi layer che prendono l'output del layer precedente come proprio input, trasformandolo e raffinando l'estrazione delle caratteristiche a ogni passaggio. Nel paper vengono presentati un po' di studi che hanno ottenuto buoni risultati classificando le emozioni a partire dai segnali fisiologici, tramite machine learning; il problema evidenziato è che le potenzialità del machine learning sono limitate quando si tratta di classificare (ovvero produrre un output discreto) qualcosa di tanto variabile tra soggetti diversi, ecco perchè ci si affida ai più potenti sistemi di deep learning. La DPAN è una rete convoluzionale a memoria a lungo termine (LSTM), ma per capire che tipo di architettura sia abbiamo prima bisogno di parlare dei concetti fondamentali di *rete neurale artificiale*, *rete convoluzionale* e *rete LSTM*.

Partendo dalle basi, una rete neurale artificiale è una funzione che traduce un vettore di tanti input chiamati *features*, o misure, in un output. È costituita da un insieme di unità di base dette neuroni connesse fra loro da sinapsi, prendendo spunto dal funzionamento del sistema nervoso umano. Ogni sinapsi è rappresentata da un peso w che influenza il valore di ogni feature x che passa attraverso di essa. La somma pesata di tutte le misure che arrivano a un neurone determina se quel neurone si attiverà o meno, passando quindi (o non) la misura ai neuroni dello strato successivo. L'apprendimento della rete neurale avviene quindi modificando i pesi delle sinapsi per ottenere i risultati desiderati. Per farlo, ci si affida a una *funzione di perdita*, che in poche parole calcola la distanza dell'output prodotto rispetto a quello ottimale. L'idea è quella di minimizzare la funzione di perdita, in modo da avere una macchina che "sbaglia poco in generale" piuttosto che una in grado di dare una risposta perfetta solo con determinati input, per poi magari produrre pessimi risultati in condizioni di input imperfette.

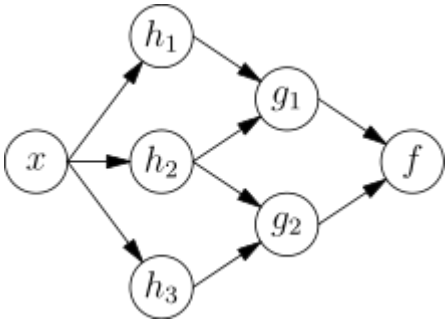


Figure 3: Schema di una generica rete neurale artificiale

Una rete neurale è detta convoluzionale (CNN) quando ha uno o più strati con una funzione precisa: I layer di convoluzione si occupano infatti di estrarre delle caratteristiche dall'input tramite l'uso di filtri, o "maschere". La maschera è una matrice che viene moltiplicata per l'input tramite il prodotto di Hadamard, ovvero una moltiplicazione cella a cella che mantiene inalterate le dimensioni. Il processo di *estrazione di features* consiste quindi nella riduzione del volume dei dati processati tramite un layer di pooling situato alla fine di ogni layer convoluzionale, che spesso e volentieri riduce anche il rumore dei dati (nel caso di Max Pooling) prima di passarli al layer successivo. L'ultimo strato di una CNN, quello di output, è completamente connesso e prende in ingresso l'immagine compressa nei layer precedenti, per compiere una classificazione o per evidenziare caratteristiche ricorrenti.

Infine, una rete LSTM (Long Short-Term Memory) è un particolare tipo di rete ricorrente, ovvero una topologia con connessioni anche all'indietro nella quale i neuroni processano sia l'output del livello precedente sia il loro stesso output all'istante di tempo precedente. Questa caratteristica che consente all'automa di prendere decisioni basate sulla sua storia passata è detta "memoria a breve termine". Tra le reti ricorrenti c'è quindi la LSTM, la cui unità base ha un'architettura come mostrato in Figura 4: La cella ha un vettore per lo stato di memoria a lungo termine (c_t) oltre a quello classico per la memoria a breve termine (h_t); inoltre può possedere tutti o alcuni dei *regolatori* che addestrano la cella su ciò che deve dimenticare dello stato passato (forget gate), sulle caratteristiche e da estrarre e mantenere dell'input corrente (input gate) e su come impostare i valori di uscita combinando l'input con informazioni provenienti dalla memoria a lungo termine (output gate).

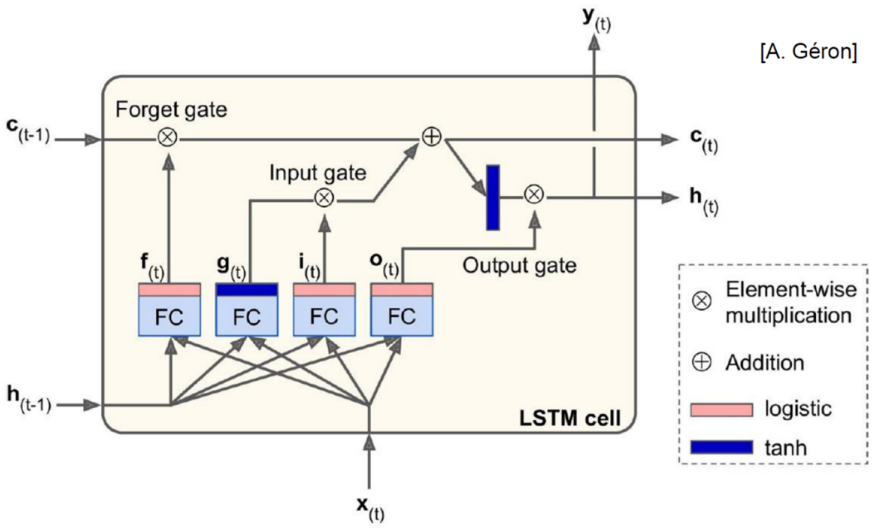


Figure 4: Architettura di una cella LSTM

6 Elaborazione dei Dati

La DPAN prende in input i segnali dell'EEG e del PPG e li analizza in sequenza temporale. L'input è dunque un tensore χ (fondamentalmente, una matrice n-dimensionale) definito sullo spazio vettoriale $(frequenza, tempo, modalita')$, dove la frequenza e il tempo sono le due dimensioni dei segnali e le modalità sono appunto elettroencefalogramma e fotopletismogramma. La sequenza temporale di questi tensori rappresenta le osservazioni che portano alla classificazione di un'emozione, in termini di coppia $(valence, arousal)$. Il processo di estrazione delle features separa quindi le due modalità, chiamando B_t le misure estratte dai due segnali EEG (uno per ogni emisfero) e H_t quelle estratte dal segnale PPG.

6.1 Estrazione del Battito Cardiaco

Analizziamo per prima l'estrazione delle features H_t, che hanno un'importanza relativa e sono più brevi da spiegare: Per calcolare il battito cardiaco si utilizzano i picchi di frequenza, presi non direttamente dal segnale ma dalla sua PSD (densità spettrale di potere). La densità spettrale di potere S_{xx} è definita come la trasformata di Fourier della funzione di autocorrelazione, quindi $\hat{R}_{xx}$. L'autocorrelazione è una funzione che mette in correlazione un segnale con una copia leggermente ritardata di sè stesso, ed è utilizzata per trovare pattern ricorrenti in un segnale e per analizzare i segnali nel dominio temporale (come quelli nell'esperimento). La trasformata di Fourier è, molto brevemente, una tecnica che scompone una funzione nelle frequenze che la costituiscono, ad esempio si possono ricavare le singole frequenze di tutte le note di un accordo. F-trasformando la funzione di autocorrelazione del segnale PPG si ottiene

$$S_{xx} = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-i2\pi f \tau} d\tau \tag{1}$$

con R_{xx} che rappresenta l'autocorrelazione e τ che rappresenta il ritardo in essa. I picchi di frequenza del segnale risultante sono quindi considerati come rappresentativi dei battiti cardiaci.

6.2 Calcolo della Lateralizzazione

Le feature estratte con l'elettroencefalografia sono calcolate con una formula apparentemente semplice, ma che utilizza fattori molto complicati che necessitano di essere analizzati. Le caratteristiche di lateralizzazione sono infatti il prodotto di Hadamard tra i due valori di *asimmetria causale* e *asimmetria spettrale*.

$$B_t = \xi_{rl} \circ \frac{\zeta_l - \zeta_r}{\zeta_l + \zeta_r} \tag{2}$$

6.2.1 Asimmetria Causale

L'asimmetria causale ξ_{rl} è un valore che indica il grado di relazione tra i due segnali r e l, misurando la quantità di influenza del primo nei confronti del secondo, in forma normalizzata. Un valore di asimmetria causale pari a zero indica che il primo segnale non ha influenza diretta sul secondo (può comunque influenzarlo indirettamente), e tanto più si avvicina a 1 quanto più r è considerato causa di l. Usiamo il termine "causa" come lo intende Clive Granger nella sua formulazione del 1969: *Dati due segnali apparentemente disgiunti, se la predizione di uno di questi migliora utilizzando le informazioni sul passato dell'altro segnale, e non solo quelle del proprio passato, allora diciamo che il secondo segnale è causa del primo* [7].

La causalità secondo Granger può essere calcolata esprimendola secondo un *Modello Autoregressivo Multivariato (MVAR)*:

$$X(t) = \sum_{k=1}^{p} A(k)X(t-k) + E(t) \tag{3}$$

dove $X(t)$ è un campione di dati al tempo t appartenente a un insieme di n segnali. Viene espresso come una somma di p valori precedenti dei campioni, pesata da una matrice $n \times n$ di coefficienti A e da un valore casuale $E(t)$ che, nel caso delle frequenze, rappresenta il rumore bianco aggiunto. Il termine $A(k)$ si riferisce alla matrice dei coefficienti relativi al k-esimo ritardo, quindi

$$\begin{bmatrix} a_{11}(k) & \dots & a_{1n}(k) \\ \vdots & \ddots & \vdots \\ a_{n1}(k) & \dots & a_{nn}(k) \end{bmatrix}$$

La relazione che intercorre tra il MVAR e la causalità di Granger è che questi coefficienti possono essere sfruttati, interpretando $a_{ij}(k)$ come *l'influenza che il segnale $x(j)$ esercita su $x(i)$ al ritardo k.*

Dall'applicazione del MVAR alla causalità di Granger sono nati diversi metodi per calcolarla, come il Granger Causality Index (GCI) o la Directed Transfer Function (DTF). Nel paper, lo strumento utilizzato per calcolare l'asimmetria causale ξ_{rl} è invece la *Partial Directed Coherence (PDC)*, la scoperta più recente a livello di stimatori di connettività cerebrale [8]. La formula riportata è proprio quella della PDC originale, dunque

$$\xi_{rl}(f) = \frac{|A_{rl}(f)|}{\sqrt{a_k^H(f)a_k(f)}} \tag{4}$$

dove f indica il dominio delle M frequenze nel tempo N. Tutto ciò che c'è da sapere di questa formula riguarda la matrice $A(f)$, che vedremo a breve, e l'operazione a_k^H chiamata "trasposta Hermitiana", che non è altro che la trasposta di a_k in cui ogni valore è poi cambiato con il suo complesso coniugato . A_{rl} è l'elemento rl-esimo di $A(f)$, mentre a_k è la k-esima colonna della

stessa matrice, che è costruita come

$$A(f) = I - \sum_{d=1}^{p} A_d(n) e^{-j2\pi f d} \tag{5}$$

che non è altro che la costruzione originale di Baccalà e Sameshima per il caso $i = j$, compresso nell'elemento d della diagonale principale della matrice dei coefficienti. Nella formula, la n rappresenta sempre l'istante di tempo selezionato, mentre la j è l'indice di colonna della matrice dei coefficienti. Questo risultato piuttosto complicato si ottiene F-trasformando la matrice dei coefficienti del MVAR, che in questa formula è indicata come A_d .

6.2.2 Asimmetria Spettrale

L'asimmetria spettrale $\frac{\zeta_l - \zeta_r}{\zeta_l + \zeta_r}$ coinvolge ancora una volta la PSD, e, stando a quanto riportato sul paper, le variabili ζ_l e ζ_r rappresentano il "logaritmo naturale della PSD" di tutte e quattro le bande di frequenza considerate, rispettivamente negli emisferi sinistro e destro. Anche in questo caso l'asimmetria è direzionata, da sinistra verso destra, e indica che all'aumentare di questo valore corrisponde una maggiore attivazione nell'emisfero sinistro rispetto a quello destro. La formula utilizzata assomiglia molto a quella dell'indice SASI [9], ma non è chiaro se sia effettivamente la stessa o sia stata rivisitata per l'esperimento. Per calcolare il SASI si utilizzano sì gli spettri di potenza, ma non v'è menzione di logaritmi, in quanto vengono impiegate le sommatorie di tutte le PSD nelle bande di frequenza basse (Theta e Alpha) e alte (Beta e Gamma), per ciascun elettrodo utilizzato e ciascun partecipante. Dopotutto, però, l'indice SASI non fa distinzione tra emisferi destro e sinistro, ma calcola l'asimmetria tra le bande alte e quelle basse dell'EEG, dunque è lecito pensare che la formula utilizzata sia più adatta per la lateralizzazione emisferica.

7 Le Novità della DPAN

Abbiamo visto che la DPAN è un modello di rete neurale sia convoluzionale, sia LSTM. È costruita sulla base della Conv-LSTM di Shi e compagni, la cui ricerca è stata fondamentale per comprendere l'efficacia di questo tipo di rete nell'elaborazione di serie temporali di dati [10]. Le LSTM si comportano meglio delle reti ricorrenti classiche nella gestione di lunghe sequenze di dati, perché non sono affette dallo spinoso problema della scomparsa/esplosione del gradiente, "intrappolandolo" all'interno dello stato della cella c_t. Tuttavia, gli strati completamente connessi sono ridondanti nel processare informazioni spettrali, dunque la scelta degli strati convoluzionali serve ad alleggerire il carico.

$$
\begin{aligned}
i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ c_{t-1} + b_i) \\
f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ c_{t-1} + b_f) \\
o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ c_t + b_o) \\
C_t &= f_t \circ C_{t-1} + i_t \circ tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \\
H_t &= o_t \circ tanh(C_t)
\end{aligned}
$$

Nelle equazioni che compongono la DPAN ci sono diverse cose da spiegare: innanzitutto, l'operatore $\circ$ indica il prodotto di Hadamard (prodotto cella a cella) mentre $*$ indica la convoluzione tra due elementi. Si chiama *convoluzione* tra due funzioni f e g l'operazione $(f * g)(t) := \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$, quindi l'integrazione del prodotto delle due funzioni, con una delle due traslata di un valore τ. La tangente iperbolica $tanh$ e la sigmoide $\sigma = \frac{e^t}{1+e^t}$ sono le due funzioni di attivazioni classiche più adatte per l'architettura LSTM. i_t, f_t e o_t sono le equazioni dei tre regolatori input, forget e output gate, mentre C_t rappresenta lo stato della cella al tempo t e H_{t-1} lo stato nascosto (ovvero l'output della cella) all'iterazione precedente. Tutte le W rappresentano i pesi delle sinapsi, in input (W_x) e in output (W_h), mentre i pesi W_c sono un insieme di pesi che la rete può apprendere.

La ConvLSTM calcola i valori degli stati affettivi partendo dalle features fisiologiche all'istante t e combinandole con gli stati nascosti delle celle e con la memoria dell'istante precedente. Nell'ultimo layer della DPAN, abilito alla classificazione, si utilizza la funzione *Softmax*, che è sostanzialmente una sigmoide per la classificare più di due classi, come nel nostro caso di più stati emotivi. Il vettore degli output y_t viene quindi compresso in un vettore $\sigma(z)$ della stessa dimensione e normalizzato in questo modo

$$
\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}},\ j = 1, \ldots, K \tag{6}
$$

7.1 La Funzione di Perdita

Il modello di Conv-LSTM di Shi e compagni, preso come metro di paragone, usa la più classica *cross-entropy* come funzione di perdita durante l'addestramento

$$L_{CE} = -\sum_{i=1}^{n} b_i \log \sigma(z)_i \tag{7}$$

dove $\sigma(z)$ è di nuovo la probabilità Softmax per la classe i analizzata tra le n classi disponibili, mentre b è un indicatore binario che corrisponde a una classificazione corretta (0) o meno (1). La DPAN invece adotta un altro approccio per l'addestramento, dopo che i suoi creatori notano un "difetto" tipico delle Conv-LSTM: infatti, funzioni di perdita come quella descritta sopra daranno sempre lo stesso peso allo stesso errore, ma questo è un aspetto migliorabile dal momento che conosciamo il contesto emotivo del soggetto analizzato. Se la rete ha già classificato in precedenza, nella stessa sequenza temporale, emozioni sufficienti ad avere una buona previsione della prossima, l'errore di classificazione dovrebbe essere punito maggiormente, invece che avere la stessa loss dello stesso errore compiuto all'inizio della sequenza (quando non c'erano informazioni sul contesto).
Per questo motivo, la funzione di perdita originale viene modificata con una basata sul *margine discriminativo* $m_t(y)$, calcolato come la differenza tra lo score considerato "esatto" $s_t(y)$ e lo score emotivo massimo tra tutti quelli relativi alle emozioni "sbagliate" $s_t(y)'$ computato finora.

$$m_t(y) = s_t(y) - \max s_t(y)' \tag{8}$$

Questo margine viene quindi inserito nella nuova funzione di perdita, che non è altro che una cross-entropy a cui viene sommato il margine di errore più grave tra quelli precedenti $m_{t'}(y)$ meno il margine attuale, regolabile con un coefficiente λ durante l'addestramento

$$L_{TM} = -\log s_t(y) + \lambda \max[0,\ m_{t'}(y) - m_t(y)] \tag{9}$$

L'idea di fondo è che più la macchina osserva una data emozione y , nella stessa sequenza, e più dovrebbe sentirsi sicura nell'affermare che quell'emozione è classificabile dalle immagini osservate. Nel momento in cui però lo stato emotivo del soggetto inizia a cambiare, lo score $s_t(y)'$ della nuova emozione (che finora era "sbagliata") aumenta, abbassando il margine discriminativo e quindi la soglia di sicurezza della rete nella valutazione. È in questo momento che la L_{TM} inizia ad avere delle perdite, diventando diversa da zero.

8 L'Esperimento

Tutti i dati per allenare la DPAN sono presi dal dataset pubblico *DEAP*, risultato di uno studio di Koelstra sulla classificazione delle emozioni dando in pasto i risultati dell'EEG a un classificatore bayesiano naive [11]. I soggetti dell'esperimento dovevano valutare il loro stato emotivo percepito durante l'ascolto di musica di diverso genere, utilizzando diverse misure tra cui le già note arousal e valence. Seguendo il database fornito da Koelstra, sono stati presi in analisi 8 elettrodi per emisfero cerebrale e il segnale del PPG di tutti i 32 partecipanti. I due segnali sono stati analizzati per un minuto, diviso in 6 frame da 10 secondi, e campionati a 256Hz per poterne estrarre la PSD. L'elettroencefalogramma presenta sempre un rumore dato dallo sbattere involontario delle palpebre, che è stato rimosso esattamente come nell'esperimento di Koelstra. Il segnale del PPG viene invece pulito dagli artefatti motori involontari con un algoritmo avanzato di analisi indipendente dei componenti (cICA) [12], applicabile in quanto si ha una conoscenza pregressa della fonte del segnale.
Dopodichè sono spiegate abbastanza chiaramente le scelte implementative per la configurazione della DPAN, per utilizzare quella con il valore di perdita più basso nel training set. L'unica variabile di cui non abbiamo ancora parlato è il learning rate, che misura quanto è necessario calibrare la rete neurale durante l'apprendimento, una volta che è stato stimato l'errore e sono stati aggiornati i pesi. Durante la discesa stocastica sul gradiente (l'algoritmo di apprendimento del Deep Learning) è possibile applicare il momentum nella procedura di aggiornamento dei pesi, secondo la formula

$$\Delta W = \eta \cdot \frac{\partial E}{\partial W} + (\gamma \cdot \Delta W_{t-1}) \tag{10}$$

dove ΔW, l'incremento dei pesi, si ottiene moltiplicando il learning rate η al gradiente dei pesi, a cui viene aggiunto il momentum γ rispetto all'ultima modifica dei pesi, nell'iterazione precedente. L'utilità del blocco tra parentesi nella formula di apprendimento è che velocizza la procedura e aiuta a non incagliare il gradiente nei minimi locali (è stato chiamato "momentum", come la quantità di moto in fisica, proprio perchè permette di proseguire nella discesa nonostante una forza esterna - il minimo locale - freni il tragitto).

8.1 Risultati - Prestazioni della DPAN

La DPAN viene messa a confronto con il classificatore bayesiano di Koelstra usato per la creazione del DEAP e con la LSTM completamente connessa di Shi, che era applicata alle previsioni atmosferiche delle precipitazioni ed è stata riassemblata per questo esperimento. La valutazione dell'accuratezza viene fatta tramite matrici di confusione, tabelle che riportano tutte le predizioni di una macchina in rapporto a quelle corrette e a quelle totali. I risultati sono a favore della DPAN, che ottiene un'accuratezza attorno all'80% nella determinazione di arousal e valence, a fronte di 66% e 68% della Conv-LSTM nelle rispettive label e di 63% ottenuto dal classificatore di Koelstra. Il motivo principale è che

la LSTM convoluzionale è in grado di interpretare meglio i pattern spettrali-temporali generati da un'emozione, catturando pattern locali e tenendoli in memoria, e agevolando quindi il problema della variabilità tra soggetti diversi e nello stesso soggetto.
Il classificatore bayesiano ha difficoltà nell'interpretare dati complessi; basandosi sulla stima della massima verosimiglianza ($\max P(\theta|X)$) dei parametri, è possibile che vengano trascurati i massimi della probabilità a posteriori ($\max P(X|\theta)$) tra le varie classi di emozioni. L'impatto di questo difetto è più evidente nel caso in cui la valence di un'emozione è attorno a 5 e l'arousal è attorno a 1, dove il classificatore bayesiano produce risultati abbastanza deludenti, finendo per compiere errori di identificazione.
La FC-LSTM ha invece un numero eccessivo di connessioni, ed è difficile che evidenzi dei pattern ricorrenti nello stesso soggetto. La DPAN ottiene risultati migliori grazie alle penalità imposte durante l'addestramento da parte della funzione di perdita marginale.
Dalle matrici di confusione mostrate nel paper risulta quindi che tutti e tre i modelli se la cavano discretamente nella predizione di stati emotivi ad alto arousal, ma in condizioni di valence neutra e di basso arousal (nelle quali è più difficile stabilire con certezza la presenza di uno stato emotivo, anche per un essere umano) la DPAN è più efficace nell'individuare cambiamenti fisiologici, dove invece i competitor fanno fatica.

8.2 Risultati - Lateralizzazione Emotiva

La caratteristica B_t relativa alla lateralizzazione cerebrale è stata calcolata come da (2), per ogni coppia di elettrodi su emisferi opposti e sempre direzionata da sinistra verso destra. Ognuno di questi B_t viene analizzato assieme ai valori di valence e arousal autovalutati dai partecipanti all'esperimento del DEAP, per cercare una correlazione fra i due elementi, in tutte e quattro le bande di frequenza dell'EEG. Il metodo utilizzato per calcolare la correlazione prende il nome di *coefficiente di Spearman*: questo misura la correlazione tra due variabili x e y, e indica il "grado di monotonicità" della funzione che correla x e y, che equivale a (-)1 se la funzione è monotona (de)crescente, mentre i valori compresi indicano una monotonia locale o un'assenza di correlazione (0). Per tutte le coppie $(B_t, valence/arousal)$ che presentano una correlazione, con coefficiente quindi diverso da zero, è stato calcolato il *p-value* per verificare la significatività di quella correlazione. Il p-value è la probabilità di ottenere risultati equi o meno probabili rispetto a quelli ottenuti in un test, dando per vera l'ipotesi nulla; nel nostro caso l'ipotesi nulla è che non vi sia una correlazione tra un'emozione (misurata in valori soggettivi di arousal e valence) e l'attivazione asimmetrica di specifiche aree del cervello, e un p-value inferiore alla soglia è considerato significativo per confutarla.
Dati quindi i 32 partecipanti al test del DEAP, per ognuno sono stati calcolati separatamente tutti i p-value delle correlazioni in entrambe le direzioni, sulle 4 bande di frequenza e su tutte le coppie di elettrodi opposti, ed infine combinati

in un unico p-value con il metodo di Fisher, che è dato da

$$X = -2\sum_{i=1}^{k} \log(p_i) \tag{11}$$

dove tutti i p_i sono i p-value di ogni partecipante ($k = 32$). Si ottiene quindi un "p-value globale" per ogni coppia di elettrodi in entrambe le direzioni, in tutte e 4 le bande di frequenza e in entrambi i domini arousal e valence. Nella

Emotion	Theta			Alpha			Beta			Gamma		
	Elec. pair	R^+	R^-	Elec. pair	R^+	R^-	Elec. pair	R^+	R^-	Elec. pair	R^+	R^-
Valence	(F7, F8)	0.53	−0.04	(F7, T8)	0.69	−0.07	(T7, FC6)	0.1	−0.59	(FC5, T8)	0.47	−0.11
	(F7, FC2)	0.67	−0.11	(FC5, F8)	0.61	−0.11	(T7, C4)	0.11	−0.53			
	(FC5, F8)	0.55	−0.02	(FC1, FC6)	0.39	−0.02	(FC5, FC6)	0.09	−0.48			
	(T7, FC6)	0.49	−0.17	(C3, FC6)	0.43	−0.08						
	(PO3, PO4)	0.48	−0.05	(PO3, F8)	0.56	−0.1						
Arousal	(F7, F8)	0.62	−0.03	(C3, C4)	0.41	−0.08	(F8, PO3)	0.01	−0.29	(F7, F8)	0.11	−0.29
	(T7, FC6)	0.58	−0.16	(T7, C4)	0.34	−0.03						
				(T7, T8)	0.66	−0.11						
				(CP1, T8)	0.44	−0.05						

Figure 5: Coppie di elettrodi significativamente correlate con la lateralizzazione

figura 5 sono quindi annotate tutte le coppie di elettrodi con un p-value molto significativo (¡ 0.01) nella correlazione con il rispettivo rating emotivo. Vengono anche indicati i valori massimi (R^+) e minimi (R^-) registrati dal coefficiente di Spearman tra tutti i 32 partecipanti, che ci aiutano a capire la direzione di quella correlazione. Coefficienti positivi indicano che al crescere della valence o dell'arousal corrisponde una crescita nella lateralizzazione B_t; i valori negativi invece indicano che la correlazione è opposta, e quindi che all'aumentare del rating soggettivo corrisponde un'attivazione maggiore nell'emisfero destro rispetto a quello sinistro (o una decrescita di B_t, che dir si voglia).

I risultati che si possono trarre da questi punti focali sono però contrastanti e mostrano ancora una volta che c'è più di un nodo ancora da sciogliere nello studio delle emozioni. Per quanto riguarda la valence, le bande Theta e Alfa riportano una maggiore attività nell'emisfero sinistro, con correlazioni significative per diverse coppie di elettrodi. Questo risultato sarebbe in linea con la *valence hypothesis* di Davidson, non fosse che per le bande Beta e Gamma l'attività cerebrale è maggiore nell'emisfero destro, a sostegno della *right hemisphere hypothesis* di Levy. Sebbene gli studi più recenti suggeriscano che le due teorie siano in qualche modo complementari, i dati ricavati non sono sufficientemente rilevanti per sbilanciarsi a favore di una o dell'altra.
Le coppie di elettrodi significative a livello di arousal sono ancora meno, purtroppo, fatta eccezione per la banda Alfa. Supportati da diversi studi, gli autori della ricerca lo attribuiscono al fatto che gli stati emotivi di alto arousal sono condizionati principalmente dal sistema nervoso autonomo, più che dal cervello. I dati raccolti sono comunque concordi con quelli dell'esperimento del DEAP sulle emozioni HALV (tipo la rabbia).

References

[1] Paul Broca. Sur le siège de la faculté du langage articulé (15 juin). *Bulletins de la Société Anthropologque de Paris*, 6:377–393, 1865.

[2] Roger W Sperry, Michael S Gazzaniga, and Joseph E Bogen. Interhemispheric relationships: the neocortical commissures; syndromes of hemisphere disconnection. *Handbook of clinical neurology*, 4(273-290), 1969.

[3] Jerre Levy, Wendy Heller, Marie T Banich, and Leslie A Burton. Are variations among right-handed individuals in perceptual asymmetries caused by characteristic arousal differences between hemispheres? *Journal of Experimental Psychology: Human Perception and Performance*, 9(3):329, 1983.

[4] Richard J Davidson, David Mednick, Edward Moss, Clifford Saron, and Carrie Ellen Schaffer. Ratings of emotion in faces are influenced by the visual field to which stimuli are presented. *Brain and cognition*, 6(4):403–411, 1987.

[5] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.

[6] James Russell and Lisa Barrett. Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of personality and social psychology*, 76:805–19, 06 1999.

[7] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

[8] Luiz A Baccalá and Koichi Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biological cybernetics*, 84(6):463–474, 2001.

[9] Hiie Hinrikus, Anna Suhhova, Maie Bachmann, Kaire Aadamsoo, Ülle Võhma, Jaanus Lass, and Viiu Tuulik. Electroencephalographic spectral asymmetry index for detection of depression. *Medical & biological engineering & computing*, 47(12):1291, 2009.

[10] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28:802–810, 2015.

[11] 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, 2011.

[12] Wei Lu and Jagath C Rajapakse. Approach and applications of constrained ica. *IEEE transactions on neural networks*, 16(1):203–212, 2005.