

Progetto sull'analisi dello stress mediante tecniche di machine learning

Patrick Jusic, Loris Rossi

Campionamento e Dataset

Il dataset utilizzato per il progetto è stato gentilmente fornito dalla professoressa Federica Verdini in collaborazione con il dottorando Luca Romeo, i quali hanno svolto una profonda analisi preliminare del problema, ed un'accurata ricerca basata sulla letteratura accademica attualmente disponibile.

La fase di campionamento è suddivisa in tre fasi, nelle quali vengono monitorati i parametri del paziente. Le tre fasi hanno lo scopo di mettere in risalto diversi stati di stress cognitivo e correlarli alla variazione dei parametri fisiologici, temperatura corporea, battito cardiaco e sudorazione della pelle, sfruttando una strumentazione non invasiva.

La prima fase consiste in una fase di relax nella quale viene preparata la strumentazione per il campionamento, ed il paziente è invitato a rilassarsi in una fase precedente a quella di stimolo.

La seconda fase è invece la fase in cui si cerca di stimolare lo stress cognitivo, sottoponendo il paziente al test della Torre di Hanoi (detta anche Torre di Brahma), un rompicapo inventato nel 1883 dal matematico Edouard Lucas. Il funzionamento è semplice, inizialmente un certo numero di dischi, da 3 a 8 solitamente, di diametro crescente vengono infilati in modo da formare una specie di piramide, sul primo di 3 perni disposti uno di fianco all'altro.



Lo scopo del gioco è quello di ricomporre la "piramide" o "cono" di partenza sul terzo paletto tenendo presente che si può spostare un disco per volta e che mai e poi mai un disco di diametro più piccolo può stare sotto di uno più grande.

Il test si conclude con la replica della fase iniziale in cui il paziente è ancora invitato a rilassarsi per riportare il suo stato di test cognitivo ai parametri standard.

Alla fine del test al paziente è fornito un questionario da compilare nel quale è tenuto a valutare l'esperienza sul lato del suo stato di stress cognitivo legato alle tre fasi. Proprio la valutazione numerica fornita dal paziente in questo questionario è utilizzata come label negli algoritmi di analisi dello stress.

Il dataset fornito conta un numero di pazienti sottoposti al test pari a 17, con una variazione piuttosto importante della quantità dei dati raccolti legata al singolo paziente.

La strumentazione utilizzata durante il test è così composta:

- un bracciale posto sul polso in grado di monitorare GSR (galvanic skin response), cioè la sudorazione della pelle e ST (skin temperature), cioè la temperatura corporea
- un cardio-frequenzimetro posto a livello toracico in grado di raccogliere diversi dati riguardando il cuore tra cui, HR (heart rate), cioè la quantità di battiti cardiaci su un minuto di misura e RR (RR interval detto anche HRV, heart rate variability), cioè la distanza temporale tra un battito cardiaco e un altro

Il trattamento di questi parametri principali, con un'analisi svolta sia nel dominio temporale che in quello della frequenza, e l'utilizzo di una serie di dati statistici quali deviazione standard e varianza riescono a fornire un elevato numero di features, esattamente 32, impiegate negli algoritmi di analisi successivamente sviluppati.

Letteratura e Analisi del problema

L'approccio iniziale al problema è stato lo studio di alcuni articoli esistenti, documentazione fornita da Luca Romeo, riguardante la lettura esistente riguardante l'argomento.

Il primo importante articolo è quello a cura delle docenti Jennifer A. Healey and Rosalind W. Picard dell'MIT, intitolato *Detecting Stress During Real-World Driving Tasks Using Physiological Sensors*, comparso sulla pubblicazione *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, VOL. 6, NO. 2, di giugno del 2005. In tale pubblicazione è spiegata in maniera dettagliata l'analisi del medesimo problema in un contesto diverso, cioè quello della guida. L'approccio è molto simile, anche se la strumentazione impiegata è molto più complessa ed invasiva. Non è spiegata in maniera dettagliata la fase di pre processamento dei dati e gli algoritmi di machine learning impiegati. I valori di accuratezza riportati sono elevati, del grado del 97%. Questo è sicuramente un articolo cardine nella ricerca in questo campo.

In secondo luogo è stata presa visione di una ricerca svolta all'Università Sant'Anna di Pisa, intitolata *Stress detecting wearable system-Towards a methodology for stress recognition through the physiological analysis*, a cura di Giorgia Acerbi e diverse collaborazioni. L'approccio è molto simile a quello spiegato in precedenza, con la variazione di un strumentazione sicuramente più costosa e complessa, ma in questo caso non invasiva, e con una diversa trattazione della fase di stimolo, con la somministrazione di diversi test volti alla stimolazione dello stress cognitivo, tutti coerentemente spiegati nella pubblicazione.

Entrambi gli articoli approfondiscono molto, in una fase iniziale, una trattazione sul concetto di stress cognitivo ed entrano nello specifico nella trattazione di molti termini medici e scientifici molto utili per poter inquadrare il problema.

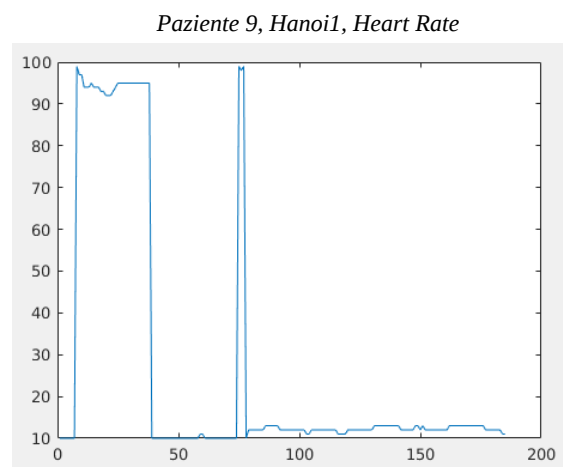
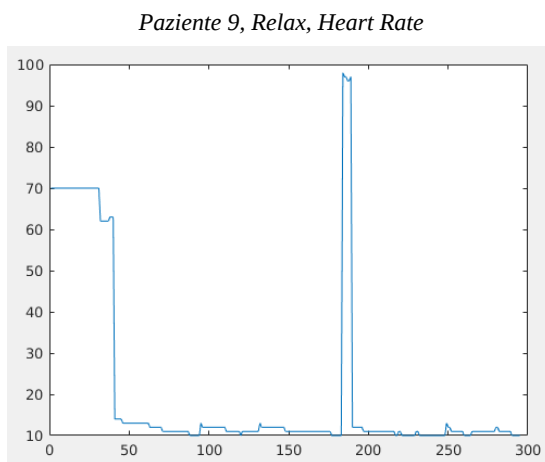
Infine è stata letta anche la tesi scritta da una laureata magistrale della nostra università sviluppata proprio sull'analisi dello stress cognitivo, tesi intitolata *Rilevamento di differenti livelli di stress tramite l'utilizzo di un sensore indossabile non intrusivo*, nella quale è spiegato in maniera riassuntiva tutto l'approccio al problema e tutte le tecniche derivanti dalla letteratura disponibile, con una serie di citazioni e referenze tra cui quelle delle pubblicazioni sopra descritte.

Elaborazione dei dati

Siamo partiti dal file “All.mat”, in cui sono presenti i dati di 17 pazienti, estrapolati secondo le modalità sopra indicate.

In particolare, i dati che si hanno sono Heart Rate(HR), Galvanic Skin Response(GSR), Skin Temperature (ST), Inter-beat Interval (RR), con i relativi istanti di campionamento, e il valore di stress associato a questi, in una scala da 1 a 10.

Visualizzando i grafici dei dati di ogni paziente, ci siamo accorti che i dati del paziente 9 sono poco plausibili, ad esempio l'heart rate assume valori vicino a 10 piuttosto frequentemente, come mostrano le immagini qui sotto:



Perciò abbiamo deciso di escludere i dati del paziente 9.

Per i dati di HR, abbiamo considerato i valori minori di 50 come degli outlier, che quindi abbiamo rimosso. Poi è stata applicata un'interpolazione spline per rimpiazzare gli outlier. Lo stesso ragionamento è stato fatto per il GSR, in cui abbiamo considerato come outlier i valori che distano dal precedente per più di 300.

I dati di HR, ST e RR sono stati sovracampionati rispetto ai dati del GSR, dato che quest'ultimo ha una frequenza maggiore di campionamento.

Infine è stato applicato uno smooth ai dati utilizzando un filtro a media mobile.

Per quanto riguarda i valori di stress, visto che l'obiettivo è effettuare una clusterizzazione su 3 classi, abbiamo associato alla scala da 1 a 10 i seguenti valori:

Questi valori derivano dalla compilazione del questionario somministrato ai pazienti alla fine delle fasi di stimolo dello stress.

Valore di stress	Label
1-4	0
5-7	1
8-10	2

Successivamente abbiamo calcolato la media di ogni feature su finestre temporali sovrapposte di 20 secondi (che nel nostro caso corrisponde a una quantità di 100 dati, visto che il GSR è campionato a 5 Hz), e con questi dati abbiamo creato una nuova tabella destinata agli algoritmi di machine learning.

Ogni riga della tabella contiene quindi HR, GSR, ST, RR e il valore di stress associato a questi dati.

Dato che diversi valori del GSR sono piuttosto elevati, abbiamo applicato il logaritmo in base 10 per ottenere una compressione dei valori, poi abbiamo normalizzato tutti i dati tra 0 e 1, poichè i classificatori sono più efficienti se si utilizzano valori piccoli.

Infine abbiamo effettuato uno shuffle dei dati per evitare dei bias nella fase di apprendimento degli algoritmi di machine learning.

Poichè le 3 classi non hanno lo stesso numero di dati, abbiamo creato un altro dataset in cui le classi sono bilanciate, per poter verificare in quale caso i classificatori si comportano meglio.

Tecniche di machine learning

Le tecniche di machine learning che sono state utilizzate sono: Multi Layer Perceptron (MLP), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) e Quadratic Discriminant Analysis (QDA).

Abbiamo usato Python come linguaggio di programmazione, per il quale è stata sviluppata un'ottima libreria, SciKit Learn, fornita di algoritmi di machine learning che abbiamo utilizzato per le tecniche sopra citate.

La libreria SciKit comprende delle ottimizzazioni sui parametri dei diversi algoritmi, e gestisce autonomamente la divisione del dataset in training data e test data.

Riportiamo qui a destra un estratto del codice utilizzato.

```
# Importazione del dataset e divisione tra features e label
dataset = np.genfromtxt(dataset_path, delimiter=',')
X = dataset[:, 0:4]
y = dataset[:, 5]

# Split del dataset in training data e test data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33)

# Normalizzazione dei dati (media 0, varianza 1)
scaler = StandardScaler()
scaler.fit(X_train)
X_train = scaler.transform(X_train)
X_test = scaler.transform(X_test)

# Istanza del classificatore
mlp = MLPClassifier(hidden_layer_sizes=(100, 100, 100))

# Apprendimento del classificatore
mlp.fit(X_train, y_train)

# Test del classificatore sui test data
predictions = mlp.predict(X_test)

# Stampa dei risultati
print(confusion_matrix(y_test, predictions))
print(classification_report(y_test, predictions, digits=4))
```

Inoltre, a scopo didattico, abbiamo implementato un Multi Layer Perceptron "from scratch" seguendo il libro online <http://neuralnetworksanddeeplearning.com/>.

Risultati ottenuti

Qui di seguito sono riportati i risultati dei classificatori per ogni tipologia di algoritmo. Come parametro di accuratezza è stato utilizzato l'F1-score.

Classificatore	Classi non bilanciate	Classi bilanciate
SVM	91%	90%
LDA	52%	44%
QDA	62%	63%
MLP SciKit	99,3%	99,2%
MLP from scratch	93%	89%

Come mostrano i risultati, il classificatore MLP della libreria SciKit si è dimostrato il migliore, con un'accuratezza del 99,3%. Da ciò possiamo dedurre che ci sia una **effettiva** relazione tra il livello di stress e i dati biometrici utilizzati.

Notiamo inoltre che utilizzando il dataset con le classi bilanciate, i risultati in generale non migliorano, probabilmente perchè diminuisce il numero di dati a disposizione.

I risultati sono promettenti, ma è vero anche che non sono stati provati dati diversi dal dataset fornito, e probabilmente il classificatore MLP ha ottime prestazioni soltanto con valori simili a quelli presenti nel dataset. Questa è una conseguenza del sovracampionamento dei dati e della sovrapposizione delle **finestre** temporali, che causano una ridondanza nei dati. In questo modo non c'è una netta distinzione tra i dati di training e quelli di test, e quindi non siamo in grado di determinare se il classificatore abbia le stesse prestazioni su dati provenienti da un dataset diverso.

Panorama ricerche

Gli articoli precedentemente citati, anche se in parte datati, mettono in risalto dei buoni risultati nell'analisi dello stress cognitivo.

Negli ultimi 10 anni, partendo dal lavoro svolto da Jennifer A. Healey and Rosalind W. Picard dell'MIT e da tutte le referenze citate nella loro pubblicazione, i maggiori progressi si sono ottenuti nell'hardware impiegato nell'acquisizione dei parametrici **fisiologici**, probabilmente perchè le migliorie apportate nel software non sono state sostanziali quanto quelle ottenute sull'hardware.

A conferma di ciò ci sono i risultati pubblicati dall'Istituto Sant'Anna di Pisa sopra citati, **fino** ad arrivare al prodotto Feel della Sentio, un bracciale innovativo comparso sulla rivista Business Insider, che unisce tutte le migliorie dell'hardware a un design accattivante, così da ottenere un prodotto commercializzabile, reperibile a questo link: <http://www.myfeel.co/>.

Il prodotto mette a disposizione un'interfaccia utente da app mobile, in cui sono salvati e analizzati tutti i dati della persona in tempo reale, e presenta un sistema automatico di suggerimenti per migliorare la condizione emotiva, il che lo rende un prodotto già completo.