

PeterPen

Dario Montagnini, Antonio Musolino,
Manuel Prandini, Giovanni Varricchione

Sommario

Lo scopo che ci siamo posti con questo progetto è stato la creazione di una penna che permettesse il riconoscimento biometrico di un utente tramite la sua scrittura. A tal fine, è stata progettata e realizzata una penna che permettesse la cattura dei dati di scrittura degli utenti, inviandoli ad un server tramite una connessione Wi-Fi. La penna è stata realizzata con chip economici e la scocca tramite una stampante 3D. Sono stati implementati due diversi modelli per il riconoscimento biometrico, uno basato su un algoritmo di deep learning ed uno su distanze fra segnali. Sul dataset disponibile è stata osservata una maggiore precisione nel riconoscimento da parte del primo modello.

Scrittura

La scrittura è uno dei tanti tratti che può essere usato in un sistema biometrico. È categorizzato come un tratto biometrico *comportamentale*, in quanto è basato su un'azione appresa dall'utente, oltre al suo stato umorale nel momento della cattura dei dati (altri esempi di questi tratti sono il modo in cui si cammina e la battitura dei tasti). Generalmente, ha una bassa *accuracy* ma un'elevata *acceptability*.

Il tratto è principalmente influenzato da caratteristiche quali la calligrafia, la pressione esercitata e l'inclinazione della penna durante la scrittura.

della scrittura: il primo, detto "*statico*" o "*off-line*", si basa su tecniche di *image processing*. La firma viene catturata e poi trasformata in un'immagine, che successivamente sarà analizzata dal modello (esempi sono [1], [2]). Il secondo approccio, detto "*dinamico*" o "*on-line*", si basa invece sui segnali che vengono catturati tramite dei sensori posti su una penna. Tipici segnali analizzati sono la pressione, l'accelerazione e la rotazione, in quanto caratterizzano la scrittura. In genere, quest'ultimi sono più performanti rispetto ai primi.

Stato dell'arte

Tipicamente i sistemi di riconoscimento della scrittura si basano sul riconoscimento della firma dell'utente. Tali sistemi sono anche detti di "*signature recognition*", l'idea alla base di quest'approccio è che la firma varia notevolmente da individuo a individuo e quindi può essere sfruttata come tratto identificativo. Naturalmente, esistono anche sistemi che permettono il riconoscimento utilizzando qualsiasi parola, in questo caso si parla di "*handwriting recognition*". Mentre i primi sono vulnerabili ad attacchi di *spoofing*, i secondi sono più resistenti ma potrebbero avere performance più basse.

Ci sono due approcci principali al riconoscimento

Capitolo 1

Architettura del Sistema

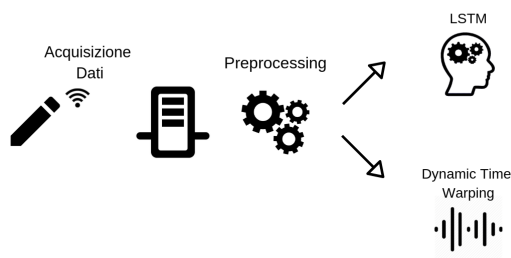


Figura 1.1: Schema dell'architettura

L'architettura del sistema è formata da 4 componenti: la penna, un server per l'acquisizione dei dati, e i moduli di *preprocessing* dei dati e di classificazione. La penna raccoglie i dati tramite i sensori e li invia al server attraverso il modulo Wi-Fi del chip. I dati vengono salvati in appositi file JSON sul server (ad ogni utilizzo della penna è associato un file diverso) e possono essere quindi usati dal modulo di *preprocessing*. Quest'ultimo si occupa principalmente di normalizzare i dati, in modo da renderli utilizzabili dai modelli di classificazione. Un modulo di classificazione è composto da una rete neurale che utilizza delle LSTM, mentre l'altro è basato sull'algoritmo *Dynamic Time Warping*.

1.1 PeterPen

La penna è composta dal seguente hardware:

- un chip NODE MCU ESP8266;
- un accelerometro;
- un giroscopio;

- un sensore di pressione;
- un LED verde;
- un condensatore;
- due resistenze;
- una batteria da 9V;
- un involucro realizzato in PLA.

L'ESP8266, il chip principale della penna, è formato da un micro-controllore e da un modulo Wi-Fi utilizzato per la comunicazione con il server. A questo sono collegati accelerometro e giroscopio (contenuti all'interno di un unico chip), il sensore di pressione e il LED.

Il sensore di pressione misura la forza applicata sulla punta della penna grazie all'ausilio di una molla. Subito dopo il sensore di pressione è posizionato il chip dell'accelerometro e del giroscopio e, dopo questo, si trova il microcontrollore.

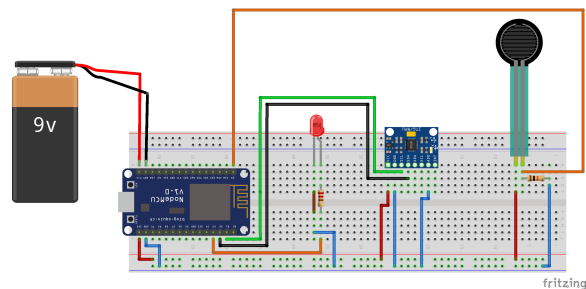


Figura 1.2: Circuito della PeterPen



Figura 1.3: Design della PeterPen, si ringraziano Cristian Cianfarani e Tiziano Grossi

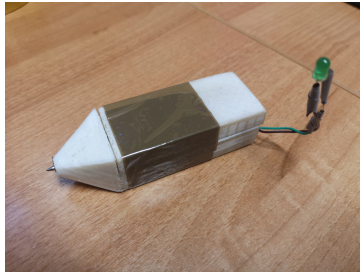


Figura 1.4: PeterPen realizzata

1.1.1 Protocollo di acquisizione

È stato ideato un protocollo per semplificare l'acquisizione e l'invio dei dati. All'accensione della penna (che può essere alimentata o tramite un cavo USB o tramite batteria, usandone o una da 9V, come in figura, o una da 3.3V o 3.6V, alimentandola però tramite il pin 3v3) il LED lampeggia durante la connessione alla rete e al server, o in caso di errore. Una volta connessa, il LED smette di lampeggiare e rimane acceso, questo indica all'utente che può iniziare a scrivere. Durante la scrittura la luce rimane spenta e, se non viene rilevata alcuna pressione per più di 2 secondi, si riaccende indicando la fine della scrittura della parola precedente e l'invio dei dati. È stato inoltre impostato come tempo massimo di scrittura 10 secondi, nel caso in cui l'utente superi tale limite il LED comincerà a lampeggiare, per segnalare l'errore e nessun dato sarà inviato.

1.1.2 Codice sorgente della penna

Il microcontrollore ESP8266 è programmato con un linguaggio derivato dal C, lo stesso che viene usato nella programmazione dei chip Arduino. Tale linguaggio richiede necessariamente l'implementazione di due funzioni in ogni programma.

La prima, `void setup()`, è eseguita un'unica volta all'inizio dell'esecuzione, ed è utilizzata per inizializzare le variabili. La seconda, `void loop()`, rappresenta il kernel del programma ed è eseguito continuamente finché il chip è attivo, nel nostro caso l'esecuzione avviene ogni 10 millisecondi.

Nel nostro caso, nella funzione `setup`, vengono invocate altre due funzioni: `mpu6050Begin` che inizializza il chip del giroscopio e dell'accelerometro e `connect` che si occupa di effettuare la connessione al server. La funzione `loop` ad ogni iterazione legge il valore della pressione e, in base allo stato (che è uno fra `READY`, `WRITING` e `ERROR`) e alla variabile `counter_writing` si comporterà diversamente.

Se la pressione registrata è sufficiente e lo stato è `READY`, allora la penna passa allo stato `WRITING`, acquisisce i dati e inizializza il contatore `counter_writing`. Invece, se la pressione è maggiore di un certo threshold, lo stato è `WRITING` e sono passati meno di 10 secondi dall'inizio della scrittura allora acquisiti i dati dai sensori; nel caso in cui sono passati 10 o più secondi allora la penna si sposterà nello stato `ERROR`. Infine, se lo stato è `WRITING` e sono passati almeno 2 secondi da quando non si rileva abbastanza pressione, vengono inviati i dati al server e viene impostato lo stato `READY`.

1.2 Server

Il server è programmato in Node.js, in ascolto sulla porta 8080 mediante una socket TCP. Ad ogni nuova connessione crea un nuovo file JSON all'interno della quale andrà a salvare tutti i dati ricevuti dalla PeterPen, effettuandone precedentemente il parsing. Nel momento in cui la sessione viene chiusa, il server si occupa di gestire la chiusura del file relativo alla stessa.

1.2.1 File JSON

Il file contiene un array di parole, ed ogni parola è composta da una sequenza di lunghezza variabile (ma lunga al più 1000) di vettori composti dalle 7 feature.

Le feature raccolte sono le seguenti:

1. componenti x, y e z dell'accelerometro;
2. componenti x, y e z del giroscopio;
3. valore della pressione.

Capitolo 2

Classificazione

2.1 Preprocessing

Nella fase di preprocessing i dati sono stati normalizzati, effettuando il rescaling delle singole feature, ed è stato aggiunto del padding per rendere i vettori delle singole parole lunghi esattamente 1000.

2.1.1 Normalizzazione

I dati sono stati normalizzati utilizzando la funzione di rescaling:

$$rescaling(x) = \frac{x - min}{max - min}$$

A tal fine, come valori minimi e massimi abbiamo usato:

- ± 250 per i dati dell'accelerometro;
- ± 2 per i dati del giroscopio;
- 0 e 1024 per i dati del sensore di pressione.

2.2 DTW

L'algoritmo DTW (*Dynamic Time Warping*) è un algoritmo di programmazione dinamica che permette di misurare la similarità fra due sequenze temporali; la particolarità dell'algoritmo è che riconosce anche sequenze che variano in velocità. Nel nostro caso, un esempio immediato è dato da un individuo che, scrivendo la stessa parola, impiega meno tempo a scriverla nel primo caso che nel secondo.

Il DTW calcola l'*optimal matching* fra due sequenze con le seguenti regole e restrizioni:

1. ogni elemento di una sequenza deve essere associato ad almeno un elemento dell'altra (ma non necessariamente ad uno solo);

2. il primo elemento della prima sequenza deve essere associato al primo della seconda;
3. l'ultimo elemento della prima sequenza deve essere associato all'ultimo della seconda;
4. il matching deve essere *monotono* e *crescente*, ossia, dati due indici della prima sequenza i, j tali che $i < j$, allora non possono esistere due indici della sequenza k, q tali che $k < q$, q è stato associato a i e k è stato associato a j .

Ad ogni *matching* è associato un costo (tipicamente la somma delle distanze euclidee fra ogni coppia di indici associati, che abbiamo usato nella nostra implementazione), e quello ottimale è quello il cui costo è minimo.

Algorithm 1: DTW

Data: $s_1 : [1 \dots n], s_2 : [1 \dots m]$

Result: distanza minima fra le due sequenze

DTW := $[0 \dots n, 0 \dots m]$

for $i := 0$ **to** n **do**

 | DTW[$i, 0$] := ∞

end

for $j := 0$ **to** m **do**

 | DTW[$0, j$] := ∞

end

DTW[$0, 0$] = 0

for $i := 1$ **to** n **do**

for $j := 1$ **to** m **do**

 cost := $d(s_1[i], s_2[j])$

 DTW[i, j] = cost +

 min(DTW[$i - 1, j$],

 DTW[$i, j - 1$], DTW[$i - 1, j - 1$])

end

end

return DTW[n, m]

L'algoritmo costruisce una matrice $n \times m$, dove n ed m sono rispettivamente le lunghezze delle due sequenze, e riempie ogni cella i, j della matrice con la distanza minima per raggiungerla dalla cella $[0, 0]$; l'output è dato quindi dal valore nella cella $[n, m]$.

Per l'implementazione di quest'algoritmo abbiamo usato la libreria disponibile su GitHub al seguente link: github.com/pierre-rouanet/dtw.

2.3 LSTM

Le LSTM (*Long Short Term Memory*) sono un'architettura di RNN (*Recurrent Neural Network*). Le RNN sono una classe di reti neurali in cui l'output ad un certo istante t è passato anche in input alla rete nell'istante successivo $t + 1$. Uno degli strati al suo interno è utilizzato come memoria, e questo permette alla rete di riconoscere sequenze di valori, utilizzando sia il valore di input all'istante t sia lo stato interno (che è ottenuto dalla sequenza di input fino all'istante $t - 1$).

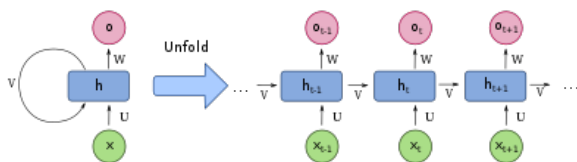


Figura 2.1: Una RNN "srotolata"

Il problema delle RNN tuttavia è che, utilizzando la *back-propagation* per effettuare l'aggiornamento dei pesi, i gradienti possono svanire fino a raggiungere 0 o esplodere. Nel primo caso, possiamo immaginarci che la RNN si scordi le informazioni che ha visto "troppo tempo fa". Per risolvere il primo problema, le LSTM permettono ai gradienti di non essere modificati quando vengono passati in input alla rete nell'istante successivo. In particolare, la rete utilizza uno strato per decidere quanto "ricordare" una determinata informazione, moltiplicando il valore per un numero compreso fra 0 e 1 (si noti in figura la moltiplicazione in alto a sinistra nella cella).

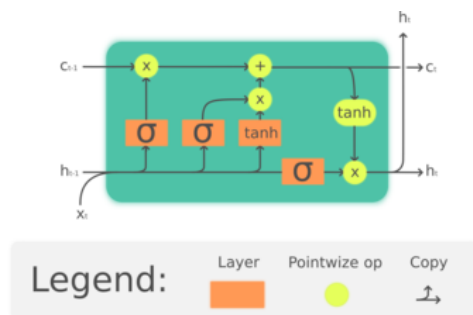


Figura 2.2: Architettura di una LSTM

Come possiamo vedere dalla figura, una cella LSTM in realtà è formata da più strati diversi, 3 con attivazione a sigmoide e 1 con attivazione a tangente iperbolica. In particolare:

- lo strato più a sinistra è detto "forget gate", e l'output regola la modifica del gradiente precedente;
- i due strati centrali sono detti "input gate", quello attivato dalla sigmoide decide quali valori andranno aggiunti alla memoria della rete, mentre quello attivato dalla tangente iperbolica dà in output i nuovi valori che potrebbero essere aggiunti. La moltiplicazione *point-wise* successiva darà il vettore di valori che saranno aggiunti alla memoria;
- lo strato più a destra è detto "output gate", e modifica l'output finale della rete.

2.3.1 Architettura del Modello

Il modello di *deep learning* implementato è formato da 5 strati, di cui 2 che sono LSTM:

1. un primo strato **Masking**;
2. due strati **LSTM**;
3. uno strato **Dropout**;
4. uno strato finale **Dense**.

Masking

Lo strato **Masking** ha come obiettivo quello di eliminare il padding, in modo da rendere meno costoso il learning.

LSTM

I due strati **LSTM** sono consecutivi, il primo ha l'opzione `return_sequences` attivata, in modo da dare in output allo strato successivo tutti i vettori della sequenza, e non solo l'ultimo. Inizialmente utilizzavamo soltanto uno strato **LSTM**, tuttavia abbiamo osservato che questa versione con due strati offre performance migliori.; inoltre, l'aggiunta del secondo strato, ci ha permesso anche di poter avere uno strato che restituisse in output tutta la sequenza elaborata, in modo da poterla analizzare in un altro strato **LSTM**.

Dropout

Lo strato **Dropout** modifica alcuni input in modo casuale settandoli a 0, per tentare di prevenire l'*overfitting*. La probabilità di impostare un input a 0 è 0.5.

Dense

Lo strato **Dense** ha come scopo quello di restituire la classificazione finale, usando come funzione di attivazione la sigmoide.

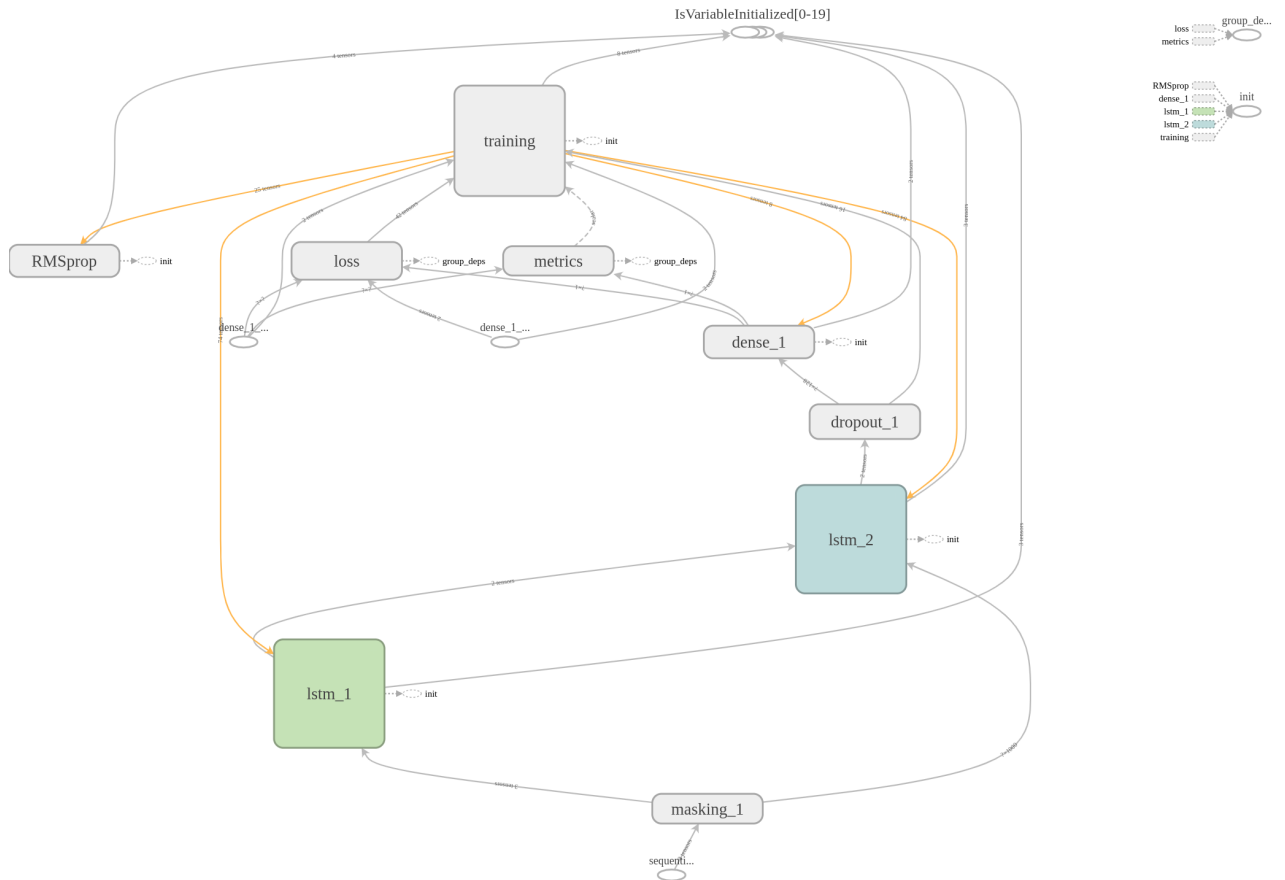


Figura 2.3: Modello usato, rappresentazione da TensorFlow

Capitolo 3

Test

Capitolo 4

Lavori futuri

Innanzitutto, vorremmo effettuare una stampa della scocca più precisa. Abbiamo incontrato varie difficoltà nel montare la penna, dovute principalmente alle imprecisioni della stampante.

Inoltre, andrebbe sviluppato anche un chip apposito, in quanto la maggior parte dei pin sull'E-SP8266 è inutilizzato.

Per quanto riguarda il lato software, ci sono due possibili sviluppi.

Dal lato DTW, si potrebbe provare a dare dei pesi alle singole feature nel calcolo del costo. Questa possibile modifica nasce dall'osservazione che, con molta probabilità, nel calcolo della distanza fra due parole scritte, ci sono alcune componenti dei dati che potrebbero essere più importanti di altre.

Infine, abbiamo osservato che il modello di deep learning, allenato soltanto per riconoscere la parola "Computer" scritta da un solo utente, ha "erroneamente" riconosciuto anche dei "Ciao" scritti dallo stesso utente. Questo ci fa pensare che, con alte probabilità, il modello si potrebbe prestare anche al riconoscimento della scrittura dell'utente, allenandolo su varie parole diverse scritte dallo stesso.

```
In [51]: with open('../dati/dati_con_penna/concatenati/pad/dario_1_concat_dario_2_pad.json', 'r') as f:
         attack = json.load(f)

         attack = np.array(attack)
         #print(attack.shape)
         model_dario_reloaded.evaluate(attack, np.zeros(attack.shape[0]))
         #attack

49/49 [=====] - 1s 21ms/step

Out[51]: [0.584262435472979, 0.897959189755576]

In [54]: with open('../dati/dati_con_penna/concatenati/pad/antonio_1_concat_antonio_2_pad.json', 'r') as f:
         attack = json.load(f)

         attack = np.array(attack)
         #print(attack.shape)
         model_dario_reloaded.evaluate(attack, np.zeros(attack.shape[0]))
         #attack

49/49 [=====] - 1s 21ms/step

Out[54]: [2.0834212467384796e-05, 1.0]

In [52]: with open('../dati/dati_con_penna/concatenati/pad/manuel_1_concat_manuel_2_pad.json', 'r') as f:
         attack = json.load(f)

         attack = np.array(attack)
         #print(attack.shape)
         model_dario_reloaded.evaluate(attack, np.zeros(attack.shape[0]))
         #attack

49/49 [=====] - 1s 21ms/step

Out[52]: [1.4550829922770892e-05, 1.0]

In [53]: with open('../dati/dati_con_penna/concatenati/pad/giovanni_1_concat_giovanni_2_pad.json', 'r') as f:
         attack = json.load(f)

         attack = np.array(attack)
         #print(attack.shape)
         model_dario_reloaded.evaluate(attack, np.zeros(attack.shape[0]))
         #attack

49/49 [=====] - 1s 20ms/step

Out[53]: [1.4182651817626118e-06, 1.0]
```

Figura 4.1: Come si può osservare, il modello rifiuta, correttamente, tutti gli altri utenti, mentre circa nel 10% dei casi riesce a riconoscere correttamente l'utente, nonostante non abbia mai visto quella parola

Bibliografia

- [1] M. Sharif, M. Khan, M. Faisal, M. Yasmin, and S. Lawrence Fernandes, “A framework for offline signature verification system: Best features selection approach,” *Pattern Recognition Letters*, 02 2018.
- [2] V. L. F. Souza, A. L. I. Oliveira, and R. Sabourin, “A writer-independent approach for offline signature verification using deep convolutional neural networks features,” *CoRR*, vol. abs/1807.10755, 2018.