



Progetto per il Corso di Deep Learning

Modello per la stima della qualità del sonno in dispositivi wearable

Ivan Carlini (549086)

Anno Accademico 2023/2024

Modello per la stima della qualità del sonno in dispositivi wearable

IVAN CARLINI

1 INTRODUZIONE

Gli activity trackers sono dispositivi indossabili che raccolgono dati dagli utenti che li utilizzano, come il numero di passi e la durata del sonno. I dati raccolti da questi dispositivi possono essere utilizzati per monitorare lo stato di salute dell'utente e fornire suggerimenti per migliorare il proprio stile di vita. In questo studio è stato preso in esame un dataset contenente dati raccolti giornalmente da dispositivi wearable nell'arco di un anno. L'obiettivo del progetto è quello di realizzare un modello in grado di stimare la qualità del sonno di un utente a partire dalle informazioni sulla durata del sonno e sull'andatura dei passi.

2 IL PROBLEMA

In input al problema si hanno quindi delle sequenze di 14 giorni di dati sul sonno e sull'andatura dei passi di un utente. In output si vuole ottenere una stima della qualità del sonno dell'utente nei giorni considerati. La lunghezza delle sequenze è stata scelta in modo da essere compatibile con la struttura del dataset e ragionevole per il tipo di problema che si vuole affrontare.

3 IL DATASET

Il dataset originale su cui si basa lo studio è composto da diversi dati aggregati raccolti su 11615 utenti giornalmente nell'arco di un anno, a partire dal 01/04/2016 fino al 31/03/2017, da dispositivi NOKIA. Tale dataset è composto da 4 colonne:

- user_id: identificatore dell'utente
- date: giorno in cui è stato raccolto il dato aggregato
- data_type: intero che identifica la tipologia di dato raccolto
- data_value: valore associato al dato raccolto

	user_id	date	data_type	data_value
0	2bc16eda651db5936cd31e735c815296fc1579d9	2016-04-01	1	10131.39
1	2bc16eda651db5936cd31e735c815296fc1579d9	2016-04-01	2	86.80
2	2bc16eda651db5936cd31e735c815296fc1579d9	2016-04-01	3	28.96
3	2bc16eda651db5936cd31e735c815296fc1579d9	2016-04-01	7	117.62
4	2bc16eda651db5936cd31e735c815296fc1579d9	2016-04-01	8	117.86

Figura 1 – Struttura del dataset originale.

La variabile `data_type` assume valori interi da 1 a 27 e identifica la tipologia del dato in base al seguente schema:

- 1: numero totale di passi effettuati
- 2: peso (kg)
- 3: BMI (kg/m^2)
- 4: pressione sanguigna sistolica (mmHg)
- 5: velocità dell'onda sfigmica arteriosa (PWV), (m/s)
- 6: PWV healthiness (1: bassa, 2: sano, 3: troppo alta)
- 7: frequenza cardiaca media (bpm)
- 8: frequenza cardiaca minima (bpm)
- 9: frequenza cardiaca massima (bpm)
- 10: durata del sonno (ore)
- 11: orario in cui l'utente si è messo a letto
- 12: orario in cui l'utente si è alzato dal letto
- 13: numero di volte in cui l'utente si è svegliato durante il sonno
- 14: tempo totale in cui l'utente si è svegliato durante il sonno (ore)
- 15: tempo impiegato dall'utente per addormentarsi (ore)
- 16: tempo impiegato dall'utente per alzarsi dal letto (ore)
- 17: durata di sonno leggero (ore)
- 18: durata di sonno REM (ore)
- 19: durata di sonno profondo (ore)
- 20: tipo di attività
- 21: durata dell'attività (secondi)
- 22: calorie consumate durante l'attività
- 23: frequenza cardiaca media durante l'attività (bpm)
- 24: frequenza cardiaca minima durante l'attività (bpm)
- 25: frequenza cardiaca massima durante l'attività (bpm)
- 26: velocità dell'andatura dei passi (passi al minuto)
- 27: velocità dell'andatura a distanza (km all'ora)

3.1. Creazione del dataset personalizzato

Poiché i dati utili ai fini del progetto sono unicamente quelli relativi al sonno e all'andatura dei passi, è stato necessario un processamento del dataset originale in modo da ottenere un dataset più leggibile e contenente i soli dati di interesse per lo studio. Questi dati in particolare sono:

- `sleepduration` (10): durata del sonno
- `stepsgaitspeed` (26): andatura dei passi
- `bedin` (11): orario in cui l'utente si è messo a letto
- `bedout` (12): orario in cui l'utente si è alzato dal letto
- `awakeduration` (14): tempo in cui l'utente è rimasto sveglio durante la notte
- `timetosleep` (15): tempo impiegato dall'utente per addormentarsi
- `remduration` (18): durata del sonno in fase REM
- `deepduration` (19): durata del sonno profondo

		sleepduration	bedin	bedout	awakeduration	timetosleep	remduration	deepduration	stepsgaitspeed
user_id	date								
1003e58667235e01b49008155604980b3900b00e	2016-08-01	11.31	22.11	9.86	0.53	0.15	NaN	3.99	108.81
	2016-08-10	9.15	22.34	7.66	0.28	0.02	NaN	3.58	105.02
	2016-08-15	7.87	23.93	7.95	0.27	0.02	NaN	3.19	88.27
	2016-09-05	8.34	22.70	7.16	0.28	0.10	NaN	3.11	84.91
	2016-09-06	8.08	23.55	7.72	0.20	0.00	NaN	3.60	88.60

Figura 2 – Struttura del dataset personalizzato.

3.2. Creazione degli insiemi di training e di test

Per estrarre caratteristiche significative dalle sequenze di 14 giorni di sleepduration e di stepsgaitspeed, sono state utilizzate due reti WaveNet distinte, addestrate sul task di forecasting. Dati 14 giorni di dati in input, queste reti hanno quindi il compito di prevedere il valore di sleepduration e di stepsgaitspeed al 15° giorno. È dunque stato necessario estrarre dal dataset sequenze di 15 giorni in modo da poter creare gli insiemi di training e di test. Poiché il dataset presenta molti giorni mancanti, il numero di sequenze di 15 giorni consecutivi non nulli risulta essere ridotto. Si è dunque deciso di consentire al più un giorno mancante tra i diversi valori della sequenza (un'alternativa sarebbe potuta essere quella di fare un'interpolazione tra il giorno precedente e il giorno successivo al giorno mancante). Attraverso questa strategia è stato possibile estrarre dal dataset 273472 sequenze di 15 giorni da utilizzare per il training e il test del modello.

3.3. Calcolo dell'indice di qualità del sonno

Per poter addestrare il modello finale nella stima della qualità del sonno è stato necessario definire un indice di qualità del sonno ricavabile a partire dai dati nel dataset. L'indice di qualità del sonno proposto da [Arora et al. (2020)] si presta molto bene a questo scopo. Tale indice è una versione rivisitata del Pittsburgh Sleep Quality Index (PSQI) che include però nel calcolo dell'indice anche le informazioni sulla durata del sonno profondo e sulla durata del sonno REM. L'indice di qualità del sonno utilizzato può assumere valori interi da 0 e 14 in cui valori bassi indicano una migliore qualità del sonno mentre valori più elevati evidenziano una qualità del sonno inferiore. Il calcolo di questo indice consiste nella somma di diverse componenti, ognuna delle quali si riferisce ad una diversa caratteristica del sonno. Queste caratteristiche sono:

- Sleep onset latency: tempo impiegato dall'utente per addormentarsi (corrisponde a timetosleep)
- Actual sleep duration: tempo di sonno effettivo (corrisponde a sleepduration)
- Sleep efficiency: percentuale del tempo passato a letto in cui l'utente dorme (ricavabile da sleepduration, bedin e bedout)
- Sleep disturbance: tempo totale per cui l'utente è risultato sveglio durante la notte (corrisponde ad awakeduration)
- Percentage deep sleep: percentuale di sonno deep (ricavabile da deepduration e sleepduration)
- Percentage REM: percentuale di sonno REM (ricavabile da remduration e sleepduration)

Una volta calcolati i valori di qualità del sonno per ogni giorno nel dataset, questi possono essere utilizzati per addestrare il modello nella stima della qualità del sonno di un utente su un dato periodo di input (14 giorni nel caso dello studio).

Component 1: Sleep onset latency (a)	Recorded value	Score (α)
	$a \leq 15 \text{ min}$	0
	$15 \text{ min} < a \leq 30 \text{ min}$	1
	$30 \text{ min} < a \leq 60 \text{ min}$	2
	$a > 60 \text{ min}$	3
Component 2: Actual sleep duration (b)	Recorded value	Score (β)
	$b > 7 \text{ hrs}$	0
	$6 \text{ h} < b \leq 7 \text{ h}$	1
	$5 \text{ h} < b \leq 6 \text{ h}$	2
	$b \leq 5 \text{ hrs}$	3
Component 3: Sleep efficiency (c)	Recorded value	Score (γ)
	$c > 85\%$	0
	$75\% < c \leq 85\%$	1
	$65\% < c \leq 75\%$	2
	$c \leq 65\%$	3
Component 4: Sleep disturbance: in bed awake duration (d)	Recorded value	Score (δ)
	$d \leq 20 \text{ mins}$	0
	$20 \text{ mins} < d \leq 30 \text{ mins}$	1
	$30 \text{ mins} < d \leq 40 \text{ mins}$	2
	$d > 40 \text{ mins}$	3
Component 5: Percentage deep sleep (e)	Recorded value	Score (ϵ)
	$e > 10\%$	0
	$e \leq 10\%$	1
Component 6: Percentage REM (f)	Recorded value	Score (η)
	$20\% \leq f \leq 25\%$	0
	$f < 20 \text{ OR } f > 25$	1

Figura 3 – Componenti del PSQI rivisitato.

4 WAVENET PER ESTRARRE LE CARATTERISTICHE DELLE SEQUENZE DI INPUT

Per poter estrarre delle caratteristiche rilevanti da una sequenza di 14 giorni, sono state addestrate due reti WaveNet distinte, una sulla sleepduration e una sullo stepsgaitspeed per il task di forecasting. L'architettura WaveNet risulta essere adeguata per l'analisi di time series. La rete convolutiva WaveNet impiega più layer convolutivi 1D, dove ogni layer raddoppia il dilatation factor. In questo modo la LRF cresce esponenzialmente mentre il numero dei parametri cresce linearmente. Questo consente di riconoscere pattern brevi nei primi layer e pattern più estesi nei layer successivi. Inoltre, rispetto agli approcci RNN presenta un training più rapido per via della mancanza di collegamenti ricorrenti. [van den Oord et al. (2016)]

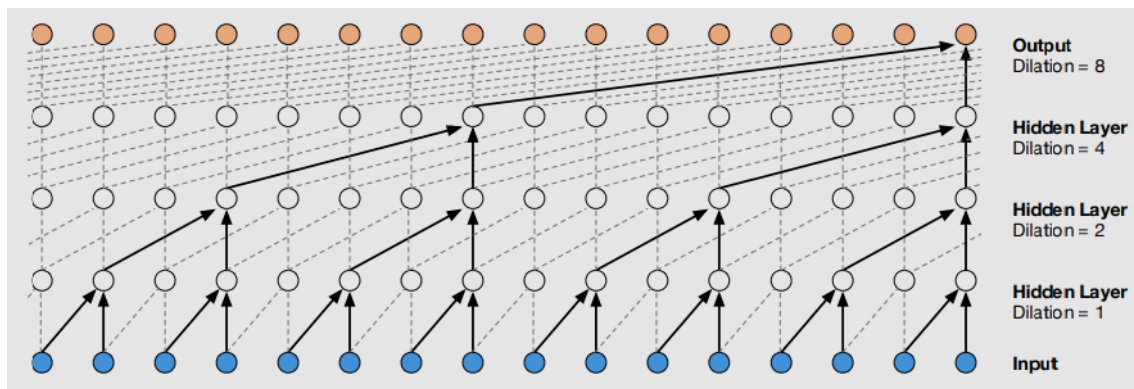


Figura 4 – Architettura WaveNet.

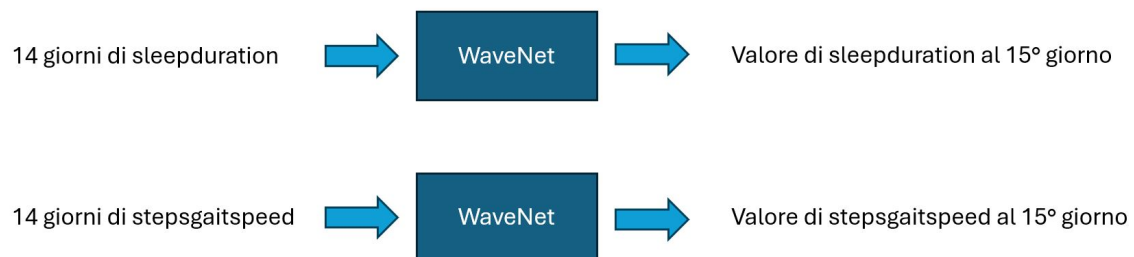


Figura 5 – Forecasting attraverso le reti WaveNet.

4.1. estrazione delle attivazioni

Le due reti WaveNet sono quindi state addestrate con l'obiettivo di estrarre caratteristiche rilevanti dalle sequenze di 14 giorni di input. Una rappresentazione di queste caratteristiche è stata poi ricavata dalle attivazioni dell'ultimo strato convolutivo delle reti WaveNet addestrate.

5 COSTRUZIONE DEL MODELLO FINALE

L'obiettivo del progetto è quello di ottenere un modello che, a partire da 14 giorni di dati sulla durata del sonno e sull'andatura dei passi di un utente, restituisca una stima della qualità del sonno dell'utente. Dunque, una volta ricavate le rappresentazioni delle sequenze dei dati di input attraverso le due reti WaveNet, queste sono state poi concatenate e utilizzate come input per una rete densa, la quale il compito di stimare la qualità del sonno dell'utente a partire da tali rappresentazioni. Una volta composto il modello finale, per poter addestrare unicamente i parametri della rete densa, è stato necessario congelare i parametri delle reti WaveNet per fare in modo che questi non venissero alterati durante la fase di addestramento.

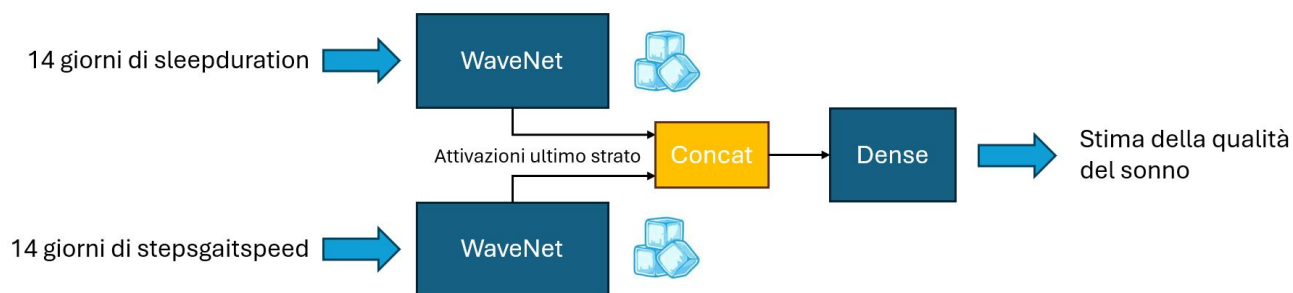


Figura 6 – Struttura della rete finale.

6 RISULTATI OTTENUTI

I risultati ottenuti dal modello finale mostrano un MSE di 1.1153 sull'insieme di test. Il modello finale riesce quindi a stimare l'indice di qualità del sonno con un RMSE di 1.0561.

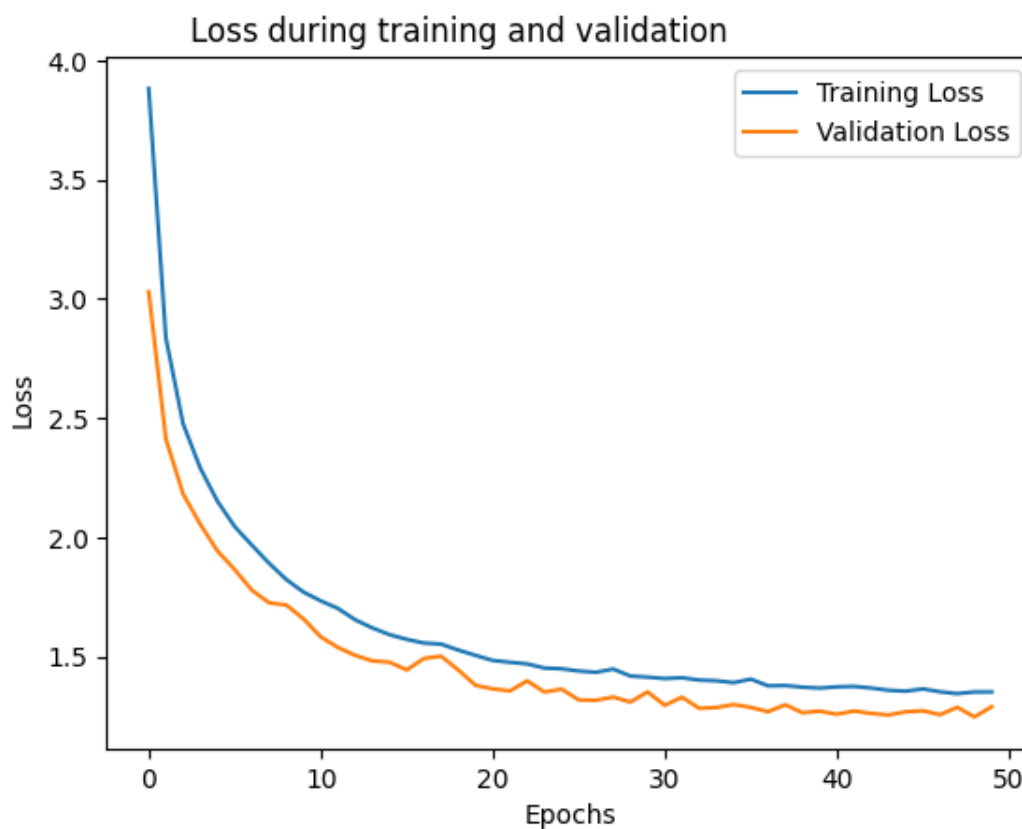


Figura 7 – Struttura della rete finale.

7 POSSIBILI APPLICAZIONI DEL MODELLO

Il modello ottenuto potrebbe essere incorporato nei dispositivi degli utenti. Una volta raccolti i dati dell'utente nell'ultima sequenza di giorni, questo potrebbe stimare la qualità del sonno dell'utente ed eventualmente fornire suggerimenti in base ai risultati ottenuti. Ad esempio, se il modello ottiene un indice di qualità del sonno al di sopra di una determinata soglia, il dispositivo potrebbe suggerire all'utente di prendersi maggior cura del proprio sonno.



Figura 8 – Esempio di possibili notifiche mostrate da dispositivi che incorporano il modello costruito.

8 CONCLUSIONI

Il lavoro svolto mostra un'applicazione originale di architetture di deep learning note con il fine di realizzare un modello in grado di stimare la qualità del sonno di utenti che utilizzano dispositivi wearable in grado di raccogliere dati sulla durata del sonno e sull'andatura dei passi.

8.1. Margini di miglioramento

Il dataset utilizzato presenta un grande sbilanciamento verso dati relativi a una buona qualità del sonno. Si hanno quindi pochi dati di utenti con una scarsa qualità del sonno. Il modello potrebbe quindi migliorare le proprie prestazioni attraverso un maggiore bilanciamento dei dati nel dataset.

Riferimenti

- Arora, A., Chakraborty, P., & Bhatia, S. (2020). Analysis of Data from Wearable Sensors for Sleep Quality Estimation and Prediction Using Deep Learning. *Arabian Journal for Science and Engineering*. <https://doi.org/10.1007/s13369-020-04877-w>
- van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio. *Arxiv*. <https://arxiv.org/abs/1609.03499>