Slide 1:

Buongiorno, sono Alessandro Carella e presenterò ora il lavoro svolto insieme ai miei relatori, il professore Stefano Ferilli e la professoressa Berardina De Carolis.

Il titolo dato al testo di tesi è Analisi espressioni facciali in contesto di e-learning e lo scopo della tesi è quello di riuscire a identificare gli stati d’animo delle persone attraverso l’analisi di immagini o video che le riprendono.

Lo sviluppo effettuato per il lavoro di tesi si è suddiviso in tre sezioni principali:

* l’identificazione di modelli predittivi adatti al task e il loro allenamento per effettuare le predizioni sui dati estrapolati dalle immagini
* la creazione di un’interfaccia grafica per effettuare delle predizioni in tempo reale e per avere un riscontro del primo lavoro effettuato
* l’estrapolazione dei dati facciali da 200 video per trattare gli stati d’animo come se fossero dei workflow per il sistema WoMan sviluppato dal professore Ferilli

Dopo una breve introduzione sullo stato dell’arte procederò ad approfondire su questi tre.

Slide 2:

Gli stati d’animo giocano un ruolo fondamentale nell’apprendimento in quanto influenzano le performance di chi sta apprendendo.

Più una persona si trova in uno stato emotivo positivo maggiore sarà la facilità con la quale riuscirà ad apprendere le informazioni che gli vengono spiegate o che sta leggendo da un libro o simil.

Gli stessi concetti sono ovviamente associabili anche ad attività lavorative, soprattutto durante il periodo iniziale.

Slide 3:

Effettuando le apposite ricerche per la stesura del capitolo sullo stato dell’arte ho notato che esistono molti studi che hanno affrontato un tema simile a quello trattato dal testo di tesi, ovvero il riconoscimento delle emozioni primarie attraverso le espressioni facciali.

Sono presenti invece in numero minore gli studi che affrontano l’analisi delle espressioni facciali per il riconoscimento dei mood (o stati d’animo).

Durante l’attività di apprendimento è difficile che gli studenti manifestino delle emozioni “forti” come la felicità, la rabbia, la tristezza, il disgusto, la paura; mentre è più probabile che vengano espressi degli stati d’animo come l’attenzione (in senso di engagement), la noia, la confusione, la frustrazione, …

Ho quindi deciso di applicare il metodo migliore fra quelli dimostrati all’interno dei paper per l’analisi delle emozioni primarie all’analisi degli stati d’animo.

Slide 4:

Gli studi per l’analisi delle emozioni utilizzano le Action Units per effettuare un’analisi matematica del volto analizzato.

Le Action Units suddividono il volto in sezioni e assegnano un valore fra 0 ed 1 in base a quanto è contratto il muscolo in quella specifica sezione del volto.

Il sistema FACS (Facial Action Coding System) è un sistema di codifica delle espressioni facciali sviluppato dallo psicologo Paul EKman e dal collega Wallace V. Friesent negli anni 70 e viene utilizzato per identificare e descrivere le eespressioni facciali in modo oggettivo e dettagliato utilizzando le action units.

Slide 5:

Negli studi analizzati per quanto riguarda l’analisi delle emozioni sono stati utilizzati diversi approcci, rappresentati qui nella slide:

Longest Common Subsequence o LCS che è un algoritmo che trova la stringa più lunga comune fra due poste in input. Per questo approccio sono state generate delle stringhe di template per le emozioni che si è deciso di analizzare ed è stato applicato il LCS con la stringa generata in output da nuove immagini proposte.

Per la generazione delle stringhe di template sono state utilizzate delle relazioni statistiche e delle tecniche di matching su dei campioni di immagini che avevano già delle label.

Negli altri studi analizzati le immagini sono invece state categorizzate attraverso la costruzione di modelli attraverso l’apprendimento superivisionato e non supervisionato.

Ho di fatti riportato sulla slide delle immagini rappresentative per il random forest e le reti neurali in deep learning.

I risultati dell’applicazione di queste 3 tecnologie hanno portato a diversi risultati, difatti l’LCS ha ottenuto uno score nelle metriche calcolate più basso rispetto al random forest che ha ottenuto dei risultati più bassi rispetto alle reti neurali.

Slide 6:

Ho optato per l’algoritmo di classificazione random forest in quanto, pur non essendo il migliore fra i tre prima citati ha comunque delle performance che si avvicinano a quelle delle reti neurali e richiede molte meno risorse computazionali a confronto.

Ho però deciso di creare altri classificatori sempre basati sull’apprendimento supervisionato in modo da metterli a confronto con il random forest, quali:

* k-nearest neighbors
* Naive Bayes
* Support vector machine
* Support Vector Regression

Fra questi i risultati migliori, a seguire del random forest, sono stati ottenuti dal k-nearest neighbors, come vederemo meglio nella tabella riportata fra qualche slide.

Slide 7:

Prima di passare ai risultati è utile parlare del dataset sul quale sono stati allentati questi modelli:

ho utilizato 2 dataset per la costruzione del mio dataset finale ovvero il DAiSEE, contenente per lo più video di 5/6 secondi e lo Student-engagement-dataset che è invece composto da immagini.

Il primo dataset contiene molti video ai quali è stata assegnata la label enaged e la label bored.

È poi composto da altre 2 label ovvero frustrated e confused.

Il secondo dataset presenta invece 6 labels di cui 4 conincidono però con quelle del DAiSEE e vengono aggiunte le classi drowsy e looking away.

Slide 8:

Visto il match totale fra le label presenti nel DAiSEE e per quelle nello Student engagement dataset ho deciso di accorpare i file presenti nel primo all’interno del secondo.

Ovviamente, essendo i file all’interno del DAiSEE video sono stati estratti dei frame da ognuno dei video ed è stata effettuata un’analisi sulle immagini ottenute.

Il dataset è risultato comunque molto sbilanciato ed ho quindi effettuato un’operazione di resampling, ed ho ottenuto un dataset di 2000 samples per ogni label, quindi in totale 12000 samples, sul quale ho effettuato il training e il test dei modelli predittivi.

Slide 9:

Riporto qui i risultati ottenuti per le metriche di accuracy, precision, recall e balanced accuracy per ognuno dei modelli predittivi creati.

Come già anticipato, sia dalla slide 6 che dagli studi ritrovati, il random forest è l’algoritmo che ottiene i risultati migliori ma il knn ottiene comunque dei risultati che si avvicinano.

Slide 10:

Per ottenere un riscontro più diretto dei modelli predittivi ho deciso di creare un’interfaccia attraverso la quale scegliere il modello predittivo che si preferisce utilizzare per poi arrivare alla seconda ed ultima schermata dove è possibile vedere, a sinistra la webcam che sta riprendendo la persona e a destra:

il mood rilevato nel frame attualmente utilizzato

tutti i valori di affidabilità rilevati dal modello predittivo (non presente per tutti i modelli, in quanto alcuni danno direttamente il risultato)

il mood più frequente nell’ultimo minuto

il tempo passato dalla precedente predizione, in modo da ottenere una stima delle performance del programma sulla macchina

una descrizione in linguaggio naturale per quanto riguarda l’attivazione di ogni action unit, il nome di questa, il muscolo contratto relativo e il valore ottenuto nell’analisi.

Slide 11:

In ultimo ho fornito al professore dei file da dare in input al framework woman da lui sviluppato.

Woman è un framework per la gestione dei workflow basato sulla logica del primo ordine.

Questo permette l’analisi e la raffinazione automatica dei workflow basandosi su processi precedentemente analizzati.

Ho quindi trattato i singoli frame del dataset DAiSEE come se fossero degli step di un workflow per analizzarli.

Slide 12:

I frame sono stati analizzati e l’attivazione (ovvero un valore superiore o uguale a 0.5) per ogni Action Unit è stato riportato come l’inizio di un processo e la disattivazione di questa è stata riportata come la fine di questo processo.

Ovviamente le immagini all’interno dello student engagement dataset non sono state utilizzate in quanto sono per la maggior parte foto indipendenti l’una dall’altra.

Slide 13/14:

Come è possibile vedere il numero di nuove task apprese dal sistema arriva a 0 dopo pochi video (o workflow) analizzati, mentre il numero di nuove transizioni non arriva mai a 0.

È comunque possibile notare, anche solo dai grafici un trend negativo. Questo fa pensare che aggiungendo altri video il numero di nuove transizioni potrebbe arrivare a 0.

Siccome per l’analisi dei 200 video, 50 per ogni mood, sono state necessarie più o meno 36 ore ho deciso di effettuare una predizione attraverso la regressione lineare ed ho ottenuto che ci vorrebbero più o meno 100 video per ogni classe (in totale il triplo rispetto a quelli già analizzati) per arrivare a 0 nuove transizioni.