Studi sul riconoscimento delle emozioni FACS per scelta del modello da utilizzare

Essendo l’ammontare di studi che trattano l’analisi delle emozioni FACS maggiore rispetto a quelle che cercano di creare sistemi di riconoscimento automatico per gli stati d’animo, che possono direttamente aiutare a identificare i problemi nell’apprendimento delle conoscenze, ho ritenuto corretto studiare e scegliere fra i modelli da loro proposti per l’elaborazione delle informazioni per il mio caso di studio.

Il paper con i migliori risultati che ho trovato è stato [12], in questo studio viene utilizzata una random forest a due livelli:

* il primo livello è utilizzato per determinare le action units nelle sequenze di espressioni che vengono analizzate.
* Il secondo livello invece prende in input le action units estratte ed effettua la classificazione delle espressioni finali.

Il modello riesce ad ottenere una media di riconoscimento del 100% per le action units e del 96.38% per le espressioni facciali (emozioni) che risulta essere più alto dei prossimi studi riportati.

Prima ancora di arrivare al primo livello della random forest viene utilizzato il modello AAM che identifica i punti di riferimento sulla faccia del soggetto ripreso nel video analizzato (o nella sequenza di immagini analizzate) e successivamente viene utilizzato il Lucas-Kanade optical flow tracker per tracciare i punti di riferimento nei frame successivi.

Viene poi calcolato il vettore di spostamento fra i punti di riferimento trovati sul frame con un’espressione neutrale e i punti di riferimento trovati nel frame dove l’espressione facciale è al suo picco.

Questo vettore di spostamento viene utilizzato nel primo layer della random forest per la prelevazione delle Action Units.

Le espressioni facciali sono processi dinamici che derivano, come conseguenza, dell'attività muscolare facciale di varie parti di questa:

* fronte,
* occhi,
* naso,
* bocca,
* mascella,
* contorno

Le caratteristiche dinamiche possono essere rappresentate dalla differenza tra fotogrammi in termini di punti. Le sei espressioni facciali di base:

* felicità,
* tristezza,
* rabbia,
* disgusto,
* paura,
* sorpresa

possono essere descritte utilizzando le AUs, dove ogni UA è codificata in base ai coinvolgimenti muscolari facciali.

Il processo di estrazione delle caratteristiche di movimento di espressione facciale è illustrato in figura

Immagine che contiene Sito Web

Descrizione generata automaticamente

Processo di estrazione dinamica delle feature delle espressioni facciali

Random Forest è un algoritmo di machine learning che combina molti alberi decisionali, dove ogni albero dipende dai valori di un vettore casuale campionato in modo indipendente e con la stessa distribuzione.

Questo algoritmo è in grado di gestire grandi spazi di caratteristiche grazie alle due fonti di casualità, ovvero input e feature casuali, che lo rendono robusto.

Random Forest è diventato molto potente per l'estimazione della posa, la rilevazione di oggetti e altre aree in computer vision grazie alla sua bassa complessità di calcolo, facilità di implementazione, accuratezza nella classificazione e capacità di gestire grandi set di dati di formazione.

Il Random Forest è composto da molti alberi decisionali, dove ogni albero viene costruito ricorsivamente assegnando un test binario ad ogni nodo non foglia in base ai campioni di formazione.

Per la classificazione, il Random Forest combina i risultati degli alberi decisionali per votare per la classe più popolare.

Questo algoritmo ha la capacità di gestire errori di bilanciamento nella popolazione delle classi in dati sbilanciati.

Per validare i risultati i ricercatori hanno degli esperimenti sui datasets esteso Cohn-Kanade e Oulu-CAISA VIS utilizzando solo le sequenze di immagini che contengono una vista frontale o girata a 30 gradi dei soggetti.

Nel primo dataset sono categorizzate 7 espressioni (anger, contempt, disgust, fear, happy, sadness and surprise) e le persone riprese hanno un’eta compresa fra i 18 e i 50 anni.

Il secondo dataset contiene samples ripresi in diverse condizioni di luci ma sono state analizzate solo le immagini con condizioni di illuminazione normali (luce forte o buona illuminazione).

In entrambi i dataset i video riprendono delle persone con, inizialmente, delle espressioni neturali e alla fine le espressioni che più esaltano le emozioni che vogliono rappresentare.

Nella fase finale del metodo proposto, viene utilizzato l’ultimo layer della random forest per identificare l'etichetta finale dell'espressione facciale in base ai risultati della rilevazione delle AU.

Viene utilizata una rete bayesiana come confronto al metodo proposto; questo modello predittivo otteiene un tasso di riconoscimento delle caratteristiche facciali di solo l'86,3%.

Il metodo proposto dal paper, basato sulle AUs, può invece aumentare il tasso di riconoscimento medio al 89,37%.

I risultati mostrati nella tabella riportata nell’immagine qui sotto sono il tasso di riconoscimento per la rilevazione dell'espressione facciale basata sulle AU utilizzando il random forest.

Immagine che contiene tavolo

Descrizione generata automaticamente Confrontando i risultati della Tabella riportata nell’immagine qui sotto, le prestazioni sono migliorano significativamente.

Immagine che contiene tavolo

Descrizione generata automaticamente Il classificatore con rete bayesiana ha un tasso di riconoscimento relativamente basso per le emozioni di tristezza e paura (rispettivamente 68,75% e 88,89%), mentre il random forest migliora queste prestazioni al 93,75% e al 100%.

Una possibile spiegazione è che la tristezza e la paura sono di solito poco evidenti e possono essere facilmente confuse con altre emozioni. Tuttavia, il random forest ha maggiore capacità discriminativa e può trovare la differenza tra di esse.

Inoltre, i tassi di successo di emozioni come la rabbia, la sorpresa e il disgusto sono stati significativamente migliorati utilizzando il framework proposto.

Per confermare l'efficacia del metodo, sono stati selezionati casualmente set di training e testing dal database e l'esperimento è stato ripetuto per 9 volte. Tutti i risultati sono mostrati nella Tabella qui sotto.

Immagine che contiene tavolo

Descrizione generata automaticamente È stato dimostrato che il sistema ha una prestazione stabile utilizzando set di training e testing diversi.

Il tasso di riconoscimento medio è significativamente più alto del metodo di confronto [15].