

ME DETECTION : DOCUMENTAZIONE CODICE (WORK IN PROGRESS)

Marioemanuele Ghianni, University of Florence

28/05/2020

Introduzione

L'idea di base del progetto è di sviluppare un programma per la detection automatica di microespressioni facciali in sequenze video basata sull'istogramma del flusso ottico e la classificazione di queste features tramite SVM.

Il dataset di riferimento è CASME2 scaricabile al link segnalato nel readme della repository github.

Il progetto segue parzialmente tecniche discusse in due paper (<https://arxiv.org/abs/1903.10765> e <https://arxiv.org/abs/1812.10306>) di cui è consigliata la lettura per una maggiore comprensione.

Preprocessing

Le sequenze video sono dapprima divise in finestre da 100 frames che, in un'acquisizione a 200fps (come lo è quella del dataset CASME2), corrispondono a una durata temporale di mezzo secondo che è anche un upper bound delle micro espressioni facciali.

Si ha dunque una sliding window che scorre i vari frames video. Vi è inizialmente una fase di preprocessing dove fondamentalmente si acquisiscono i frame e si allineano, tramite individuazione dei landmarks, relativamente al primo frame della finestra che chiameremo reference frame.

Altro step importante è l'individuazione dei ROIs (Region of interest): vengono individuate le parti del viso relative alle sopracciglia e alla bocca e viene fatto un "crop" di queste dal viso (ROI1 bocca, ROI2 e ROI3 sopracciglio sx e dx).

Sono state scelte queste 3 parti per rendere più specifica l'individuazione in quanto sono le parti del viso dove si concentrano gran parte delle micro espressioni facciali.

Istogramma del flusso ottico

Per ogni ROI viene calcolato il flusso ottico denso tramite Farneback rispetto al corrispondente ROI del reference frame.

Ottenuti questi flussi ottici, si costruisce, per ogni ROI, un istogramma del flusso ottico da 8 bin; ogni bin contiene la somma delle magnitudine dei flussi ottici aventi una determinata orientazione.

Vengono infine concatenati tutti gli istogrammi relativi a un ROI dei vari frames della finestra formando così un descrittore per ogni ROI.

SVM

Una volta che è stata processata tutta la finestra ed abbiamo ottenuto i 3 descrittori relativi agli istogrammi delle 3 parti del viso, vengono dati in pasto a un SVM lineare implementata tramite LibSVM.

Le funzioni presenti in svm.py utilizzano un modello preaddestrato per determinare se la finestra processata contiene una micro espressione o meno.

Si considera presente una micro espressione se la finestra contiene almeno l' 80% dei frames compresi tra l'Onset e l' Offset frame indicati nel file excel 'CASME2-coding' allegato al dataset. E' inoltre definita micro espressione rilevata se almeno uno dei 3 classificatori relativi ai 3 ROI, rileva una microespressione indipendentemente se essa sia o meno relativa a quella parte del volto.

Le finestre contenenti microespressioni sono indicate sia nel file di test che in quello di training con la label "+1" mentre le finestre con microespressione assente con la label "+2".

Ovviamente i 3 modelli sono addestrati con dati estratti da soggetti del dataset differenti da quelli del testset e sono considerate microespressioni le finestre che hanno una microespressione relativa a quella determinata parte del volto.

Ciò è consultabile sempre presso il file excel allegato che indica nello specifico l'Action Unit e quindi la parte del volto a cui si riferisce la microespressione.

Il programma inoltre, processando i vari frame, genera 3 file .txt relativi ai 3 ROI dove vengono salvati i descrittori in formato SVM con label di default a +2.

Questi file possono essere usati come dati di training eliminando i caratteri di separazione delle finestre e delle folder, impostando opportunamente la label a +1 nei casi di microespressione indicati sempre dal file excel.

Risultati e sviluppi futuri

Per quanto riguarda i risultati, si ottiene la matrice di confusione di seguito raffigurata e si ottiene un valore di **recall** pari a circa 0.41 con una **precision** pari a 0.56 e un **F1-score** di 0.47.

Risultati migliori dei risultati esposti dagli articoli citati nell'introduzione ma che evidenziano comunque una difficoltà nel problema di identificazione di micro espressioni.

Gran parte degli errori di classificazione, nello specifico i falsi positivi, derivano dai classificatori

delle sopracciglia mentre alcuni sono errori di "collocazione temporale" nel senso che viene rilevata una microespressione nella finestra precedente o successiva a quella dove vi è effettivamente e sono considerabili "meno gravi".

Si potrebbe ovviare a questi errori rivedendo un attimo la definizione matematica di finestra contenente una micro espressione adottando un'altra formulazione che magari tenga conto maggiormente del 'apex frame' invece che basarsi su una percentuale di frames.

Ci sono inoltre articoli che dimostrano un'efficacia maggiore di un descrittore basato su LBP rispetto a HOOF (https://www.researchgate.net/publication/319769812_Reading_Hidden_Emotions_Spontaneous_Micro-expression_Spotting_and_Recognition).

Inoltre molti studi si stanno occupando dell'uso di CNN per questo tipo di problemi e ci sono vari risultati positivi in questo ambito.

Altro fattore che potrebbe aver influenzato l'accuratezza del classificatore è senza dubbio la scarsità di dati di training in quanto, nonostante CASME2 sia uno dei dataset del settore più grandi e completi disponibili online, comprende solamente una ventina di soggetti con poche microespressioni registrate a testa.

Questo è dovuto sicuramente a una difficoltà nel costruire dataset di microespressioni spontanee e non artificiali.

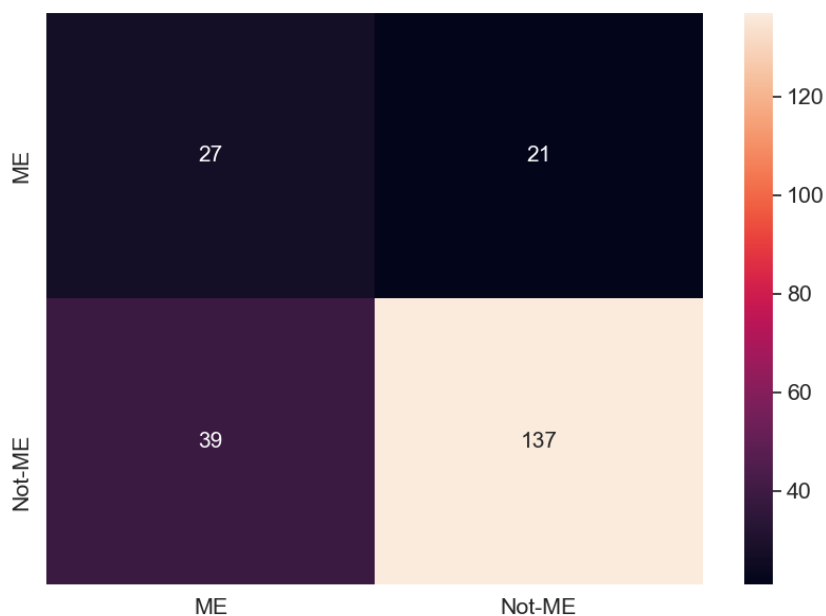


Figure 1: Matrice di confusione dei risultati; sull' "asse y" le label predette e sull' "asse x" le label reali.

