Appunti per Relazione Interazione Naturale/Computazione Affettiva

**Mappaggio di Action Units estratte da video in probabilità di emozioni**

Francesco Garavaglia, Saverio Ruggieri

**Abstract** (Sommario di quello che è stato fatto)

**- Contesto**

Il riconoscimento facciale per le emozioni è importante perchè permette di risalire allo stato emotivo di una persona.

* **Cosa è stato fatto?**

In questo articolo si propone di trovare una correlazione fra le action units (pattern di espressioni facciali) e valori di Arousal e Valence annotati da operatori umani durante un esperimento e i valori di Arousal e Valence associati

* **Come è stato fatto?**

Abbiamo utilizzato due tecniche per l’analisi di serie temporali per andare a vedere il contributo dato dalle Action units nella predizione di valori di Arousal e Valence. Abbiamo utilizzato due approcci: in uno si è tenuto conto dell’impatto dei valori passati della serie( modello ARIMA dove le action units sono state usate come variabili esogene) mentre nell’altro solo delle a

Come?

Analizzando le serie temporali con due modelli:

ARD - ridge regression

ARIMA - autoregressor moving average

verranno descritte le due tecniche

- Risultato (devo già dire qualcosa qui?)

- Dataset recola

In questo progetto ci siamo occupati di ricercare una correlazione tra valori di arousal/valence e le action units (pattern di espressioni facciali), di un dataset.\\(Inizialo con nell’ambito delle espressioni facciali)Arousal indica l'indica l'intensità di una emozione, Valence indica la sua positivià. Attraverso studi si è estratto un dataset (spiegare le annotazioni). ARIMA, forecast di serie temporali, quanto è simile la predizione con action unit rispetto alla serie originale.\\

# INTRODUZIONE

Cosa scriverci:

* IMPORTANZA DEL PROBLEMA:

Qui spiegare perchè il problema´e importante nell’ambito dell’Interazione Naturale Utilizzo di espressioni facciali per rilevare stato d’animo delle persone (anche qua una ref a qualche articolo)

Perché ci serve trovare una correlazione? Perché le annotazioni sono manuali e sulla base di dati come la pressione sanguigna (vedere se visivamente possiamo avere degli aiuti che non siano quelli annotati dall’occhio umano

* APPROCCIO SEGUITO:

Annotare con EWE e approccio che tiene conto del tempo e che non tiene conto

Qui delineare brevemente quello che è stato fatto nel progetto (ANALISI e PREDIZIONE con ARD e ARIMA)

* Contributi: qui indicare i principali contributi del progetto:
  + Concatenazione di serie temporali
  + Estrazione #lags diversa dal modello PACF applicando ARIMA a serie concatenate (trovo il p migliore)
  + Test del modello su validation test
  + Analisi del risultato: c’è una correlazione fra le action unit e Valence/Arousal?
  + Fit di un modello ARD (Automatic Relevance Determination) per serie temporali senza considerare i valori precedenti

Note per il mio progetto:

* Nella sezione 2: studi riguardo al riconoscimento e prelievo di dati
* Nella sezione 4 viene descritto il modello matematico ARIMA e ARD utilizzato per la predizione(forecasting) di serie temporali

# ANALISI DELLO STATO DELL’ARTE

Rassegna brevissima dei principali lavori in letteratura che affrontano questo problema (usa Google Scholar: <http://scholar.google.it/>)

* Face unit (action unit)
* AMHUSE: A Multimodal dataset for HUmour Sensing (guarda il goal)

AMHUSE con DANTE per le annotazioni:

video con 36 soggetti in presenza di 3 video divertenti più uno neutrale. 4 annotatori hanno preso i video e rilevato i valori in modo manuale di arousal e valence con DANTE, un tool per l’anotazione delle emozioni). Chiamiamo Rate la valutazione di un Annotatore

I dati raccolti (gathered data) includono

* RGB video
* Sequenze di responsi fisiologici (conduttanza cutanea, battito e volume del sangue, temperatura)

**Action Units Gathered Data:**

We also extracted Action Units (AU) activations, relying on a freely available AU detector [2]. The AU detector provides, at each frame, the activation level of the following AUs AUk , k = 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 45, plus the binary presence of AU 28.

# MODELLO TEORICO

Descrizione teorica degli aspetti principali sui quali si è lavorato. Etichettare equazioni, immagini e tabelle

* Arima, Ridge Regression
* Far vedere funzionamento di ARIMA (come ci si è arrivati) e modello matematico
* Che cos’è AIC?s 3 righe (ref alla spiegazione)
* PACF
* Abbiamo usato ARIMA(p,1,0) dove d=1 per rendere stazionaria la serie in analisi (abbiamo fatto un controllo pe la stazionarietà della serie, è presente nell’algoritmo che abbiamo scritto, è quindi possibile anche analizzare le serie una ad una ed effettuare il forecasting anche di una serie sola)
* ADF (usato come test di stazionarietà delle serie)

# SIMULAZIONE ED ESPERIMENTI

Descrizione sintetica delle simulazioni effettuate.

## 4.1 Dataset

RECOLA database (multimodale perché si usano diversi approcci con cui i dati vengono rilevati, in questo caso, annotazione di Arousal/Valence/ e pattern di espressioni facciali)

Cosa contiene RECOLA? Video di partecipanti che completano un task che richiede collaborazione, in remoto a coppie (comunicazione in remoto una coppia 🡪 dyadic teams).

* 6 annotatori francesi (3 donne, 3 uomini)
* Sono stati annotati solo i primi 5 minuti

<https://diuf.unifr.ch/main/diva/recola/data/emoSPACE13_RECOLA_Database.pdf>

In this paper, we introduce a new multimodal corpus (database i cui dati sono ottentuti riloevando od osservando i segnali del corpo) of spontaneous

interactions in French called RECOLA, for REmote COLlaborative and Affective interactions.

Usare solo Valence e Arousal porterebbe a perdita di informazione, perché l’’annotatore approssima qualcosa che vede. A noi non interessa, stiamo cercando di vedere se c’è qualche correlazione tra quello che viene fatto con Arousal e con Valence

ECG: elettrocardiogramma

EDA: conduttanza cutanea

<https://diuf.unifr.ch/main/diva/recola/data/Kantharaju18-ARA.pdf>

ANNEMO was used to perform emotion ratings and the annotations were done separately for each emotional dimension, using a slider with values ranging from -1 to +1

* Emotional behaviour contiene annotazioni su segnali AUDIOVISIVI di Arousal e Valence fatti con ANNEMO (slider +-10)
* Le action unit sono state ricavate a parte con FACS.

Esiste una correlazione fra le due cose?

Nelle action units non teniamo contodella presenza (se è visibile un pattern sulla faccia) ma solo del valore della sua intensità

Arousal valence [-10,10] normalizzato in [-1,1]

* CSV per Arousal, Valence, Action Units. 🡪 14 sample
* Scrivere come sono stati rilevati i dati e sottolineare che ci sono stati dati pronti
* Scrivere lo split nostro

## 4.2 Architettura del sistema

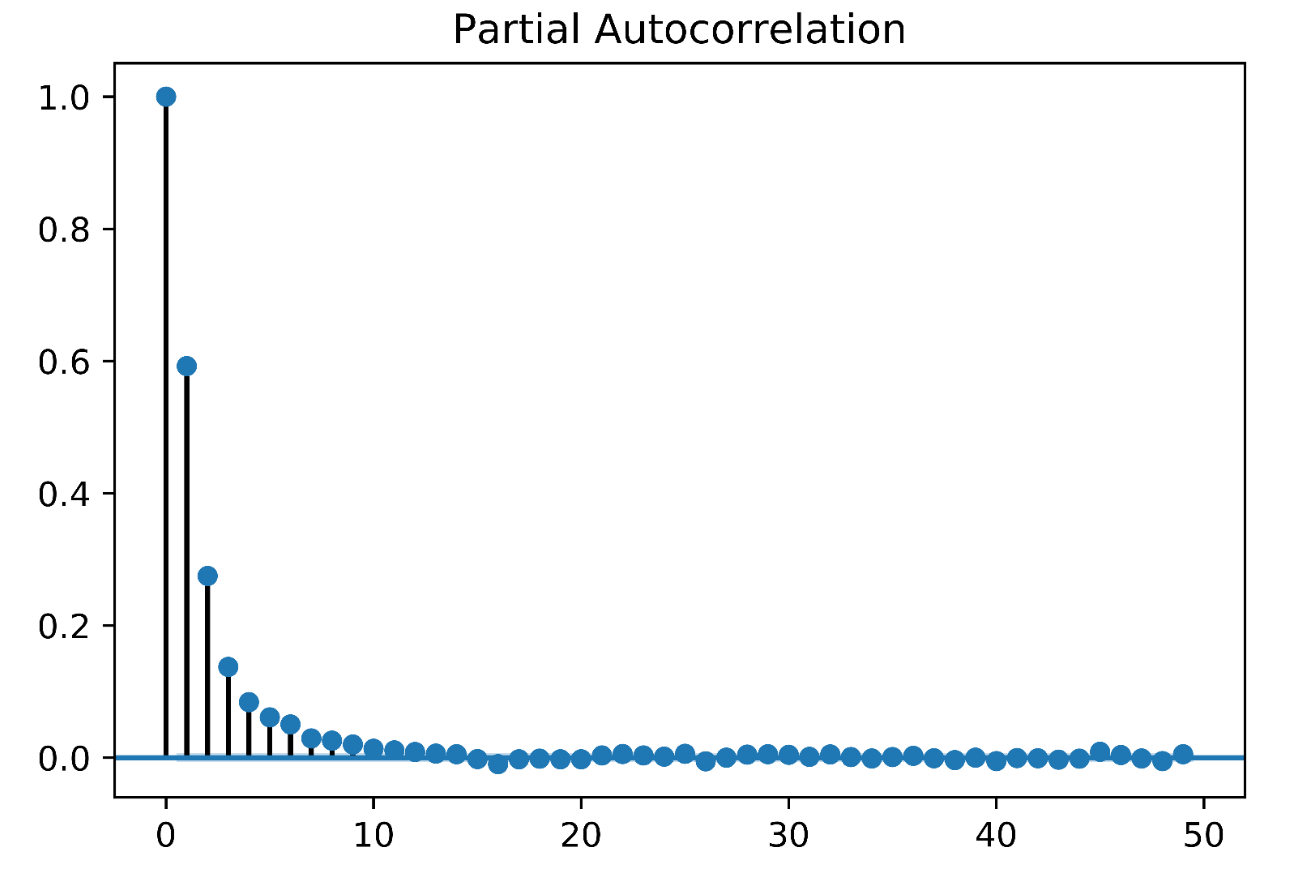
* GOLD STANDARD (implementazione della regola descritta al punto 2)
  + Gold standard con EWE deriva dalle annotazioni fatte sui vari video dei ragazzi che reagiscono agli stimuli (videoclip divertenti)
  + Spiegazione su DRIVE [NOTE PROGETTO IN CA](https://docs.google.com/document/d/1E7HqvdW1gJS-wFvzvRWMauoy25P_uOofdvrpTiZ7ZaQ/edit?usp=sharing)
* Analisi serie senza tener conto del tempo
* Analisi serie temporali
  + ARIMA (comprovata stazionarietà con ADF, concatenamento serie)
  + Validation

## 4.3 Dettagli implementativi

* Spiegare utilizzo PACF
* Abbiamo usato numpy e pandas (qui mi serve sapere esattamente per cosa) librerie python

# RISULTATI OTTENUTI

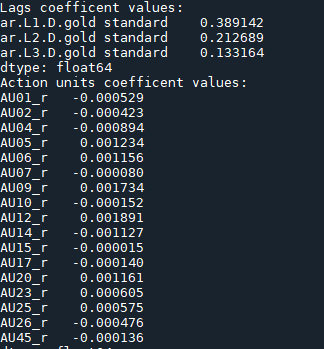
Sintesi dei principali risultati ottenuti mediante tabelle e/o grafici, immagini output del sistema.

PACF 

Da qui estraggo p = [3,4,5,6,7,9]

* Analisi dei pesi restituiti applicando ARIMA (descrivere prima da qualche parte la feature summary e cosa restituisce
* Tabella con action units significative sul training (o sul validation)
* Grafico del validation test a confronto con la predizione
* Grafico correlazione action unit SIGNIFICATIVE con Arousal e Valence (capire se serve)

Come faccio la correlazione? A me interessa veder come variano le action units significative al variare di Arousal e Valence, tendenzialmente mi interessa vedere cosa succede ai picchi e se sono situazioni simili



* Estrazione dei coefficienti stimati con ARIMA(3,1,0)
* Con p=[3,4,5,6,7,8,9] p=6 risulta essere il p migliore (min AIC), quindi usiamo ARIMA(6,1,0)

# COMMENTI CONCLUSIVI

Commenti finali su risultati ottenuti, criticità dell’approccio, problemi aperti, ecc.

* ARIMA non funziona (pesi più alti ai lag, modello sballato nel forecast, molti pesi sono anche inaffidabili).
* ARD trova pattern degli occhi evidenziate le action unit rilevanti per Arousal e per Valence, due parole sulla likelihood (riguarda il documento).

# APPENDIX

Eventuali dettagli tecnici, dimostrazioni matematiche, spezzoni di codice

# REFERENCES

[bibliografia]