Università degli Studi di Milano - Bicocca
Dipartimento di Informatica, Sistemistica e Comunicazione

Corso di Laurea in
Teoria e Tecnologia della Comunicazione

Comprensione, analisi e previsione dell'attrattività dei video nei social network

Relatore:
Prof.ssa Francesca Gasparini

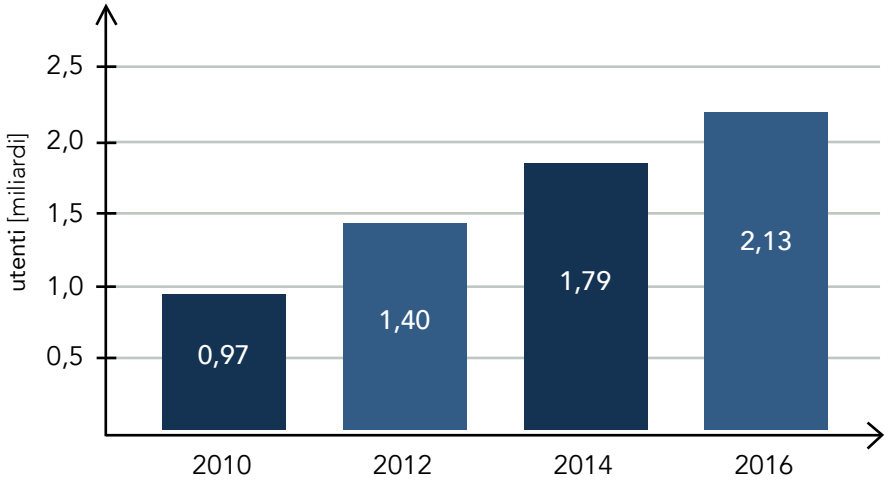
Co-relatori:
Prof.ssa Silvia Elena Corchs
Prof. Gianluigi Ciocca

Prova finale di Laurea di:
Manuela Veratti

Matricola:
808832

Anno Accademico 2016/2017

L'Attrattività di un video può essere **misurata**?



La quantità di utenti sul Web ha avuto una crescita vistosa ed esponenziale, che ha portato a un conseguente aumento dei video.

Mentre alcuni video riscuotono un **alto livello di interesse**, una buona parte risulta essere monotona, noiosa e priva di coinvolgimento e attenzione.

Tre sono gli elementi chiave che determinano il ciclo di vita[1] di un video.
Sono emersi brevi video online o **MICRO-VIDEO**, della durata di 6 secondi, che hanno un vincolo di tempo che ispira l'"attrattività".

[1] *Redi, M., O'Hare, N., Schifanella, R., Trevisiol, M., & Jaimes, A. (2014).6 seconds of sound and vision: Creativity in micro-videos.*

Perchè misurare fattori di Attrattività, Interesse e Creatività di un video?

La misura dell'interesse dei video è utile per migliorare la soddisfazione dell'utente in molte applicazioni, nella ricerca sul Web e social.

Un modello per prevedere automaticamente l'attrattività e l'interesse presenti nei video è assai complesso.

La quantità di dati generati è enorme e sorge la necessità di fornire strumenti per un'elaborazione automatica dei contenuti.

Scopo di questo lavoro è comprendere e valutare l'attrattività, l'interesse e la creatività che un video può suscitare a partire dalle sue caratteristiche visuali.

È stato utilizzato a tal scopo un dataset di micro-video[1] precedentemente etichettati come creativi e non creativi, provenienti dai social. Su cui sono state valutate varie misure di caratteristiche del segnale video, grazie alle quali è stato possibile creare un classificatore.

[1] *Redi, M., O'Hare, N., Schifanella, R., Trevisiol, M., & Jaimes, A. (2014).6 seconds of sound and vision: Creativity in micro-videos.*

Cosa contraddistingue un video "creativo" da un video "non creativo"?[1]

Video
"creativo"

Video
"non creativo"

[1] *Redi, M., O'Hare, N., Schifanella, R., Trevisiol, M., & Jaimes, A. (2014).6 seconds of sound and vision: Creativity in micro-videos.*

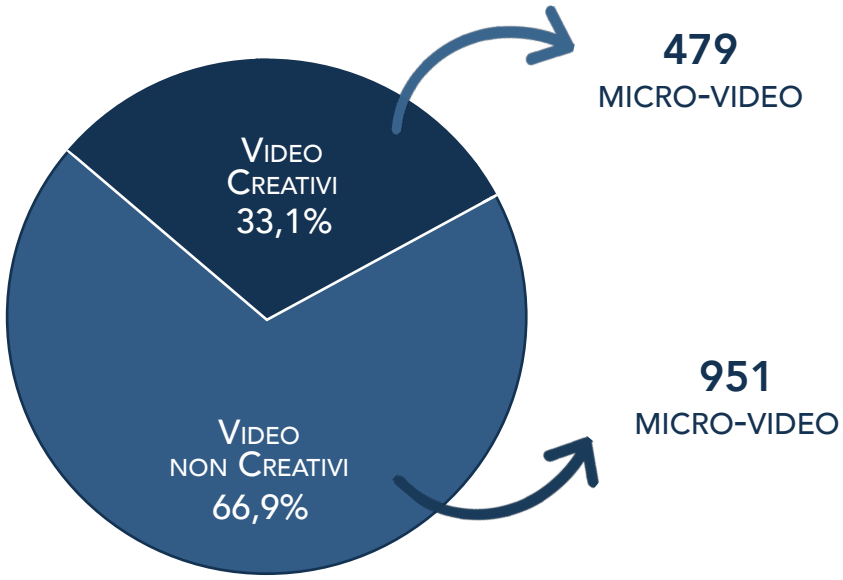
Database di Micro-Video

La CREATIVITÀ non deve essere solamente associata a qualcosa di nuovo, ma deve avere valore estetico perciò deve essere adeguata alle esigenze cognitive della situazione ed evocare emozioni.

Due sono i criteri utilizzati per etichettare i video: **NOVITÀ** e **VALORE ESTETICO**.

Valore Estetico	Sensoriale	L'audio è attraente/eclatante Le immagini sono attraenti/sorprendenti
	Emotivo/ affettivo	Il video evoca un'emozione
	Intellettuale	Il video suggerisce idee interessanti
Novità	L'audio è originale e sorprendente Le immagini sono originali/sorprendenti Il contenuto della storia è originale/sorprendente	

Attraverso un esperimento di crowdsourcing, Redi, O'Hare et al[1] hanno creato un **SET DI DATI** iniziale di 3.849 micro-video di Vine, della durata di 6 secondi.
Dei seguenti dati, riportati nel loro lavoro, solo **1.420 MICRO-VIDEO** sono disponibili: 951 non creativi e 479 creativi.



[1] *Redi, M., O'Hare, N., Schifanella, R., Trevisiol, M., & Jaimes, A. (2014).6 seconds of sound and vision: Creativity in micro-videos.*

Estrazione Key-Frame

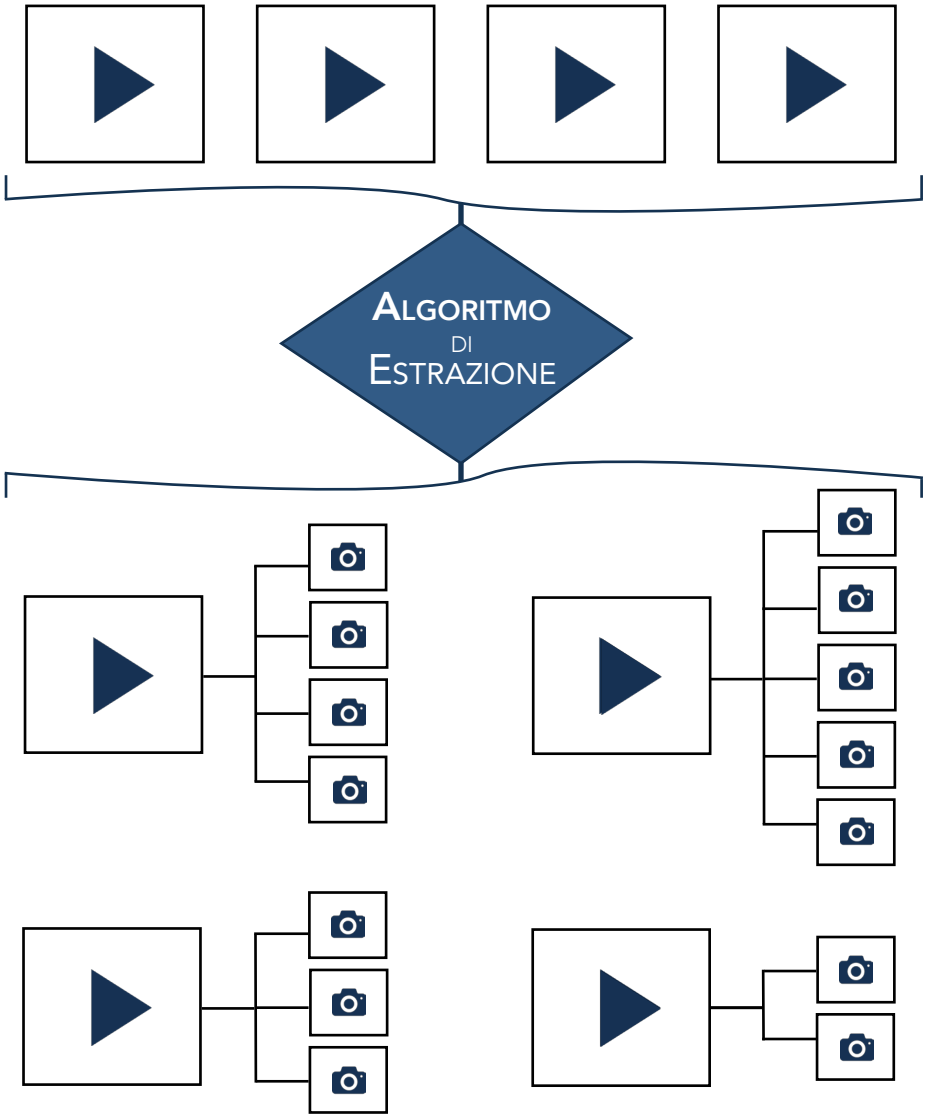
PRE-PROCESSING:

- Micro-video: 6 secondi
- Frame-Rate: da 22 a 30

6 x 26 = 156

Come ridurre la quantità di dati esaminati, nel recupero di una determinata informazione? Attraverso un'**analisi limitata ai frame salienti**.

Si è ricorso a un criterio innovativo per la selezione di frame salienti, propri di una sequenza video: **VIDEO SUMMARIZATION**[2]. L'estrazione è basata sui contenuti e combina sia proprietà del colore sia proprietà strutturali.

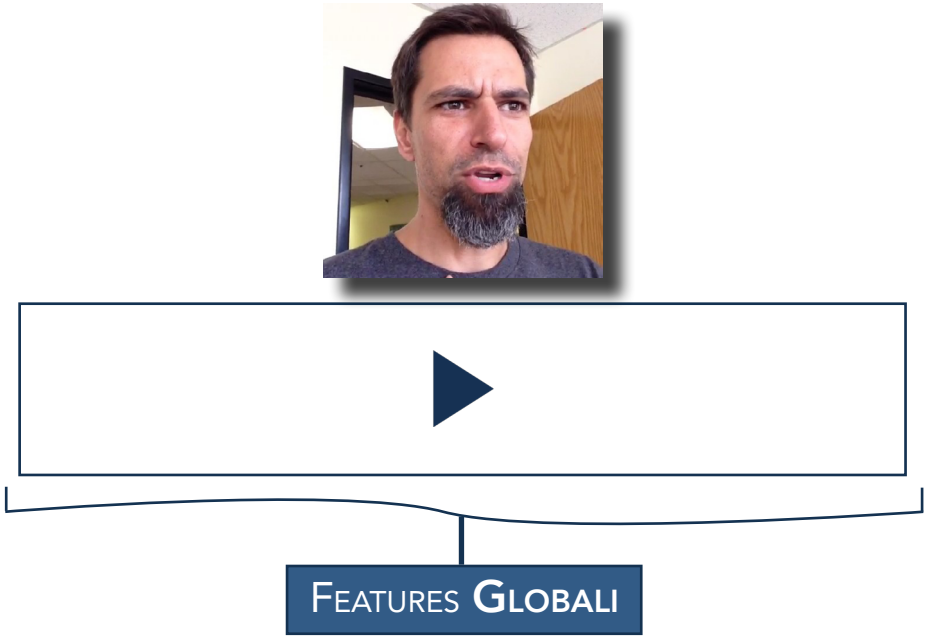


[2] *Gianluigi, C., & Raimondo, S. (2006). An innovative algorithm for key frame extraction in video summarization.*

Caratteristiche analizzate

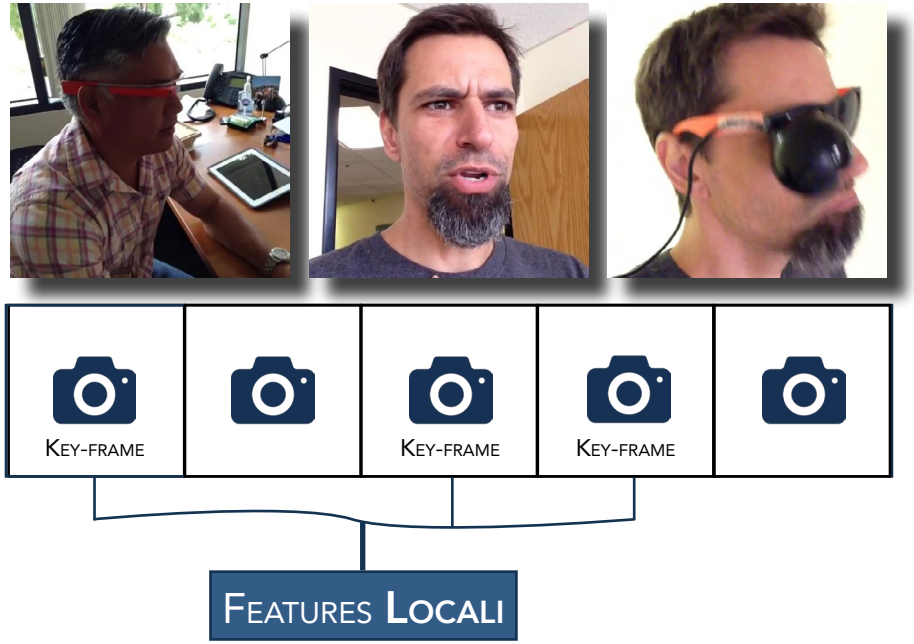
Features Globali (8)

Corrispondenti a un vettore di 8 elementi; descrivono l'intero video, inteso come sequenza di frame che varia nel tempo. (Frame Rate, numero di key-frame, ecc).



Features Locali (27)

Corrispondenti a un vettore di 327 elementi; descrivono i singoli frame, non tenendo conto dell'aspetto temporale del video. Applicate ai key-frame.



Features Locali

Caratteristiche di Basso Livello

1. a livello di grigio

Descrittori: coarseness, contrast, directionality, linelikeness, roughness, ecc.

2. a livello di colore

Descrittori: chroma variance, colorfullness, number of regions, ecc.

Esempi di applicazione di una metrica a livello di colore: *Colorfullness*[3]; a sinistra = 0.62, a destra = 0.17

Caratteristiche di Qualità

1. a singola distorsione

Misure (5) che valutano la qualità dell'immagine a singola distorsione.

2. "general purpose"

Misure (2) che valutano la qualità indipendentemente dalla distorsione.

Esempi di applicazione di una metrica a singola distorsione: *Blur*[4]; a sinistra = 3.15, a destra = 29.25

[3] *Hasler, D., & Suesstrunk, S. E. (2003, June). Measuring colorfulness in natural images.*
[4] *Marziliano, P., Dufaux, F., Winkler, S., & Ebrahimi, T. (2002). A no-reference perceptual blur metric.*

Features Locali

Caratteristiche di Qualità Fotografica

1. estetiche

2. messa a fuoco

Caratteristiche di Visual Clutter

Misure che valutano la congestione del disordine visivo, permettendo di migliorare usabilità ed estetica.

Esempi di visual clutter[5]

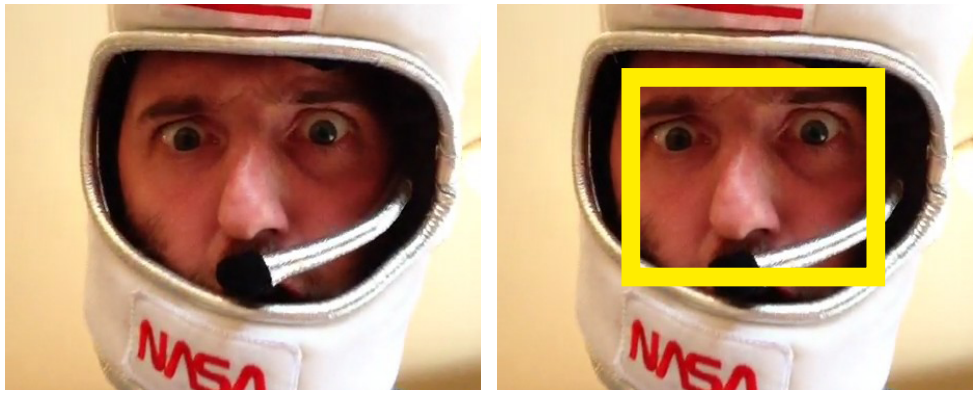
Caratteristiche Semantiche

1. face detector[6]

Applicato per estrarre il numero di facce presenti all'interno dei key-frame estratti.

2. skin detector[7]

Applicato per estrarre una percentuale di regioni di pelle delle immagini a colori.



Esempio di applicazione del *Face Detector*[6]

[5] *Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter.*
[6] *Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features.*
[7] *Gasparini, F., Corchs, S., & Schettini, R. (2008). Recall or precision-oriented strategies for binary classification of skin pixels.*

Classificatore

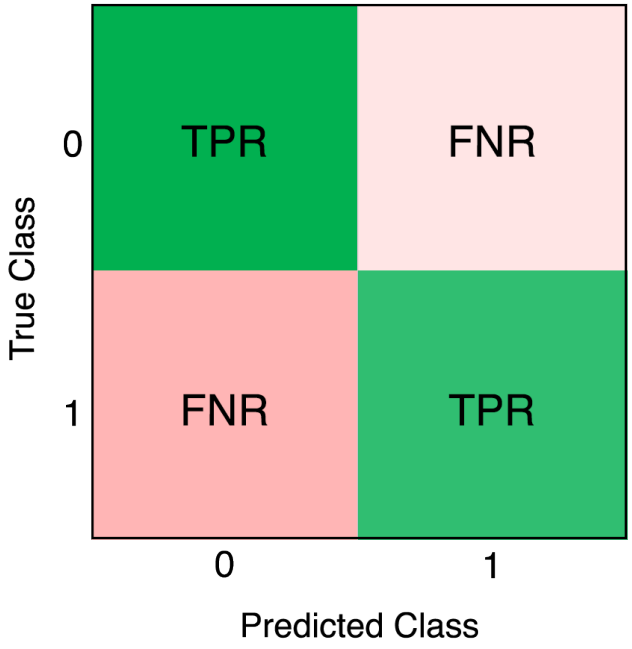
Per far fronte al problema di classificazione binaria, ho utilizzato il Classification Learner App di Matlab, che necessita di:

- Un insieme di dati di input
- Risposte note ai dati

così da adoperare i dati per addestrare un modello che genera previsioni per la risposta a nuovi dati.

Per misurare le **perfomance** di un classificatore, si utilizza la matrice di confusione, da cui è emergono valori di:

- True Positive Rate (TPR)
- False Negative Rate (FNR)

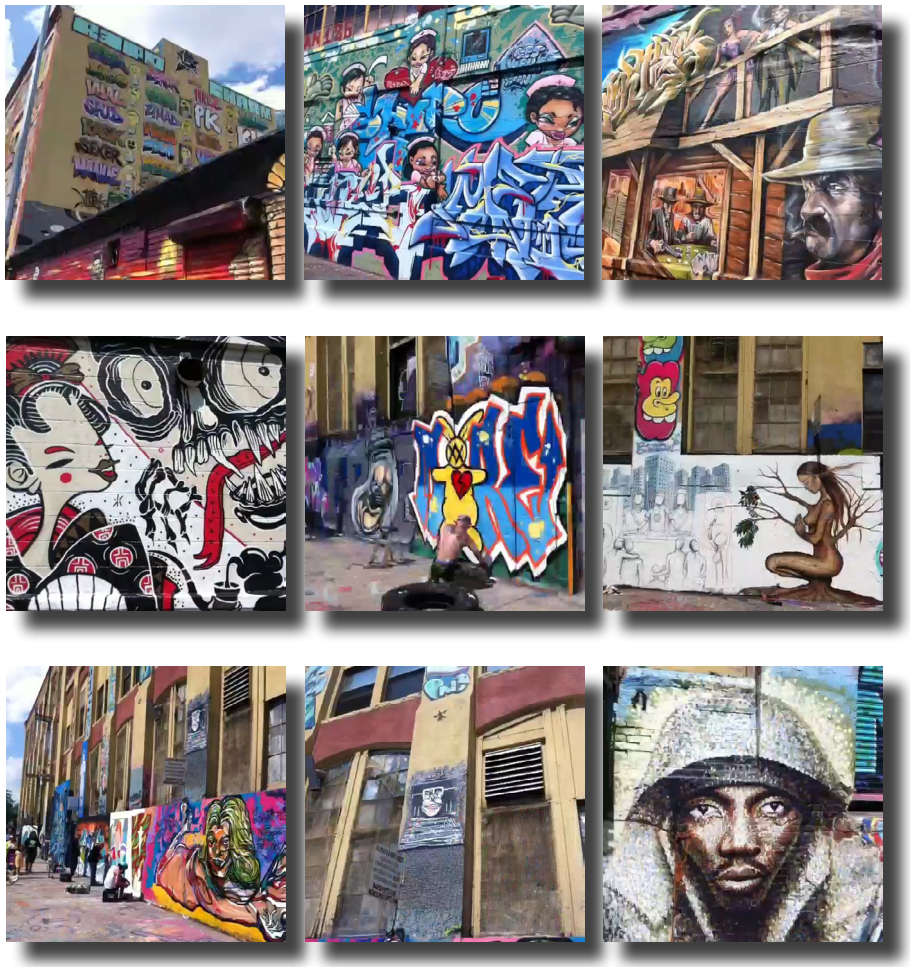


È possibile calcolare il grado di **accuratezza** del classificatore analizzato: grado di riconoscimento del classificatore.

$$\text{Accuratezza} = \frac{\text{somma elementi sulla diagonale principale}}{\text{numero totale elementi da classificare}}$$

Esempio di Video "Creativo"[1]

Key-Frame (9)



Caratteristiche

Globali
- FRAME = 177.93
- FRAME RATE = 28

Basso livello a livello di grigio
- COARSENESS = 0.81
- CONTRAST = 0.42
- DIRECTIONALITY = 0.08
- LINELIKENESS = 0.72
- ROUGHNESS = 0.50
- EDGE DENSITY = 0.08

Basso livello a livello di colore
- CHROMA VARIANCE = 8.25
- COLORFULLNESS = 0.36

di Qualità a singola distorsione
- BLUR = 4.21
- NOISE = 0.63

Visual Clutter — CLUTTER CONGESTION = 26.40

Semantiche — FACE DETECTOR = 1

[1] *Redi, M., O'Hare, N., Schifanella, R., Trevisiol, M., & Jaimes, A. (2014).6 seconds of sound and vision: Creativity in micro-videos.*

Esempio di Video "Non Creativo"[1]

Key-Frame (1)

Caratteristiche

Globali
- FRAME = 192.76
- FRAME RATE = 29

Basso livello a livello di grigio
- COARSENESS = 0.73
- CONTRAST = 0.32
- DIRECTIONALITY = 0.02
- LINELIKENESS = 0.67
- ROUGHNESS = 0.42
- EDGE DENSITY = 0.03

Basso livello a livello di colore
- CHROMA VARIANCE = 10.25
- COLORFULLNESS = 0.45

di Qualità a singola distorsione
- BLUR = 4.22
- NOISE = 0.64

Visual Clutter — CLUTTER CONGESTION = 20.14

Semantiche — FACE DETECTOR = 0

[1] *Redi, M., O'Hare, N., Schifanella, R., Trevisiol, M., & Jaimes, A. (2014).6 seconds of sound and vision: Creativity in micro-videos.*

Esperimenti e Analisi

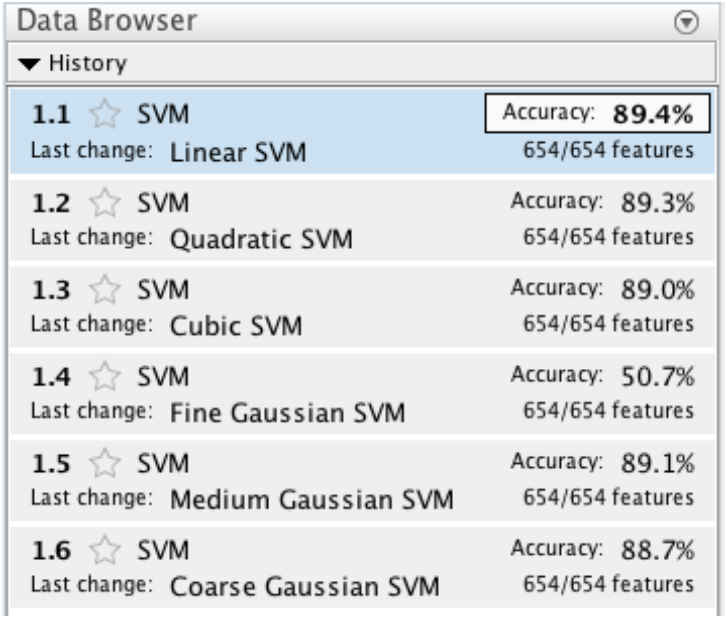
Dal momento che ciascun video è costituito da un numero variabile di key-frame estratti, per ogni video sono state calcolate diverse misure di aggregazione per vettori di caratteristiche della stessa dimensione.

Successivamente, è stata eseguita una suddivisione delle caratteristiche in:

- → **Globali**
- → **Locali**
- → **Combinazione Globali e Locali**

La stima delle prestazioni dipende dai dati utilizzati. È stata eseguita una **cross-validation**: training e testing insieme.

Sono stati adoperati i modelli di classificazione *Support Vector Machine* disponibili, tra i quali per ogni suddivisione di caratteristiche applicate è stato selezionato il miglior modello con grado di precisione e accuratezza maggiore.



Risultati

Caratteristiche	Accuratezza della Classificazione
Globali	66.4%
Locali	90.1%
Globali + Locali	**90.6%**

Analisi	Accuratezza della Classificazione	Caratteristiche	Dataset
Redi, O'Hare et al	80%	**Valore estetico + novità (audio+visiva)**	D-100
Mio lavoro di tesi	90.6%	**Globali + locali**	D-100*

** Sotto-insieme di video, ricavato dal set di dati di Redi, O'Hare et al[1]*

Workflow e Linee guida per la creazione di un Dataset "attrattivo"

1. Mole di dati

Quantitativo equiparato di video creativi e non creativi.

2. Scelta dei Social Network

Da cui scaricare i video che comporanno il dataset finale; consigliati: Flickr, Instagram, Vine e YouTube.

3. Categorie, etichette e filtri

Applicare eventuali filtri al motore di ricerca dei social selezionati, per ridurre il quantitativo di dati generati.

4. Analisi dei video filtrati

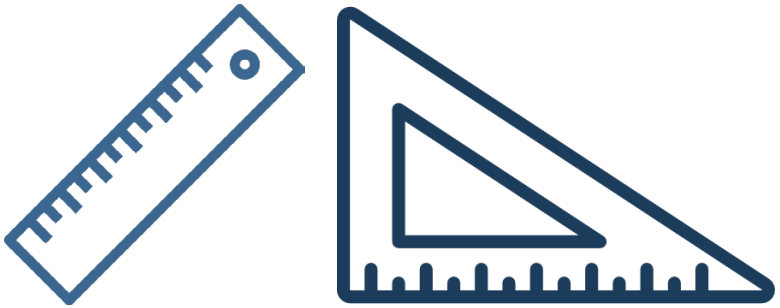
Da considerare alcuni fattori: numero di like normalizzato (in base ai follower dell'utente possessore del video), numero di follower del proprietario, visualizzazioni, commenti, preferiti e durata.

5. Validazione dei video filtrati

Attraverso un esperimento di crowdsourcing, da cui si ottiene un grado di concordanza tra gli osservatori sui dati raccolti e sottoposti ad esperimento. Il dataset finale dovrà essere quello per cui c'è accordo tra osservatori.

Incipit & finale: l'attrattività può essere misurata? Certo che sì.

Dopo aver estratto i frame salienti e analizzato caratteristiche specifiche del segnale video e dei key-frame, posso affermare che l'**ATTRATTIVITÀ** può essere misurata ed è possibile creare un classificatore binario che predice l'attrattività.



Come **ASPIRAZIONI FUTURE** c'è l'intenzione di ingrandire l'insieme delle caratteristiche per modellare i fattori di attrattività, interesse e creatività presenti nei video:

- → **FEATURES DINAMICHE** (movimento e velocità degli oggetti)
- → **FEATURES DI AUDIO**

In un secondo momento, con un maggior quantitativo di materiale, sarà possibile incrementare le analisi e gli esperimenti. Grazie anche alla realizzazione di un'ANALISI DETTAGLIATA DELLO **STATO DELL'ARTE**, da cui sono evidenti argomenti, metodi, analisi, dataset ed eventuali collegamenti tra i riferimenti bibliografici utilizzati.

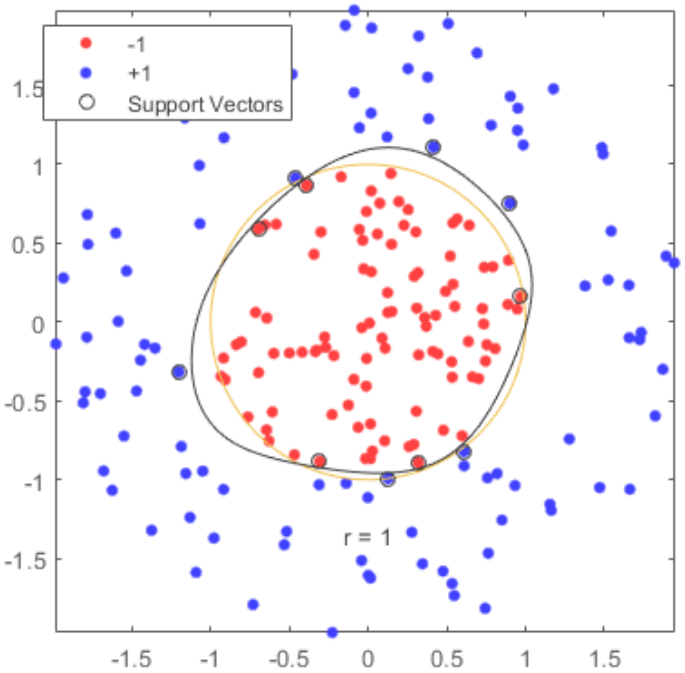
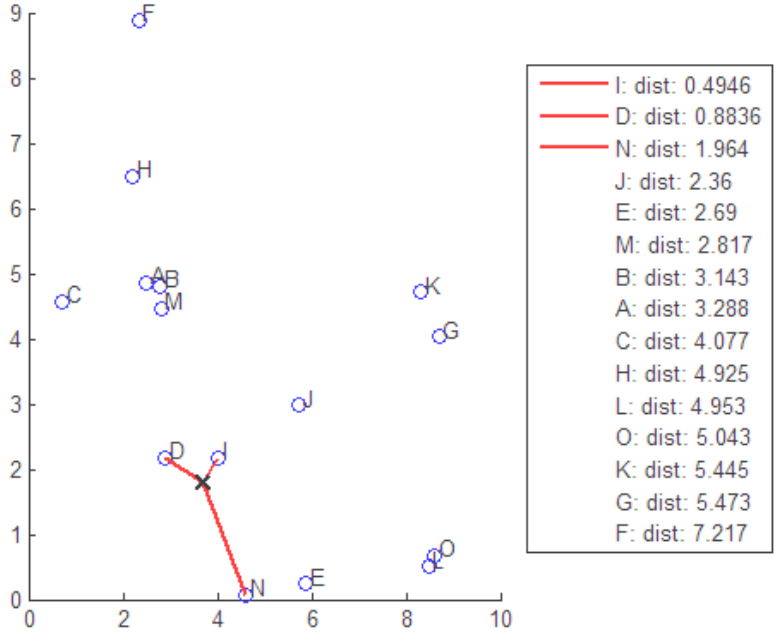
"La Creatività è l'arte di sommare 2 e 2 ottenendo 5"

Arthur Koestler

Grazie
per
l'Attenzione

Modelli di classificazione: SVM & K-NN

SVM o support vector machine



K-NN o k-nearest neighbor