

Object Classification...a Capodanno!

The Regressionists: Aldeghi Chiara 791835, Crippa Chiara 791876, Iacoban Iuliana Anastasia 790668

Abstract

Vi siete mai chiesti quale sia l'oggetto più fotografato e postato sui *social*? E se il periodo di interesse fosse proprio quello di Capodanno? Questa è stata la domanda che ci ha spinto a compiere la suddetta analisi, con l'aspettativa di riscontrare differenze o somiglianze culturali nei diversi Paesi analizzati. Paesi fra loro simili postano le stesse immagini? Si sono considerate le immagini provenienti dal *social network* Twitter e, dopo l'implementazione del modello di *deep learning* *vgg16* e successive analisi, con i risultati ottenuti si è data risposta al nostro quesito.

Indice

Introduzione	1
1.Dati	1
2.Preprocessing	2
3.Neural Network	2
4.Risultati	3
5.Conclusione	4
Sitografia	4

Introduzione

Diverse sono state le occasioni in cui, sia nel corso di lezioni universitarie, sia durante interessanti conferenze, vi sia stata presentata la capacità e l'abilità con cui certi algoritmi di *machine learning* catturano e identificano l'oggetto presente in una fotografia, un video o una sequenza di immagini. Questo aspetto, in continua evoluzione, ha comunemente incuriosito tutti i membri del gruppo 'The Regressionists' del corso di *Statistical Learning M* dell'anno accademico 2018-2019. In particolare, ci siamo domandate quale potesse essere l'oggetto più fotografato e postato su Twitter, nel periodo tra il 31 Dicembre 2018 e il 2 Gennaio 2019 in alcuni dei Paesi del mondo, selezionati casualmente in base alla

densità di popolazione, all'attività di questi ultimi sul *social* considerato, ma anche alla reperibilità dei dati.

1. Dati

Per adempiere a questo obiettivo si sono reperiti i dati dalla piattaforma *social* Twitter facendo *web scraping*. Si è implementato uno script in Python che, tramite la libreria *selenium*, utilizza *headless chrome* per fare *scrolling* da *html* e scaricare gli *urls*, in maniera univoca, di ciascuna immagine pubblicata nell'arco temporale di nostro interesse. Lo script è stato costruito in modo da filtrare i dati per tipologia di file (immagini jpg) e nazione. Si è scelto di utilizzare uno *scraper*, piuttosto che delle *standard API* per ottenere un campione di dati sufficientemente grande da garantire risultati robusti e consistenti ai fini dell'analisi. Infatti, il metodo delle *API*, inizialmente implementato (e successivamente abbandonato), ha portato a scarsi risultati sia in termini di numerosità del campione, che di tempistiche: le *standard API* permettono di accedere all'1% del materiale pubblicato nei 7 giorni precedenti alla data della richiesta di dati e ciascuna richiesta può essere effettuata a distanza di 15 minuti da quella precedente. Fare *web-scraping* ha dunque permesso di ottenere 1000 immagini in formato jpg per 42 paesi e solo 234 immagini per la Russia, nazione che abbiamo deciso di considerare ugualmente, nonostante la discrepanza nella numerosità delle osservazioni a nostra disposizione. Tale decisione è stata maturata col fine ultimo di offrire una visione completa, dei risultati relativi a questo progetto,

anche nel continente asiatico, impoverito dall'assenza di uno dei Paesi più popolati e influenti al mondo, vale a dire la Cina (a causa di una politica restrittiva che vieta l'utilizzo della suddetta piattaforma *social*). I Paesi da noi considerati ricoprono in maniera omogenea tutti i continenti, nella speranza di identificare delle differenze culturali significative, ma anche per rendere l'analisi più completa. In particolare, si sono scelti:

- **Africa:** *Angola, Egitto, Marocco, Nigeria, Senegal, Tanzania, Uganda.*
- **America:** *Argentina, Brasile, Canada, Cile, Colombia, Ecuador, Messico, Perù, Stati Uniti, Venezuela.*
- **Asia:** *Arabia Saudita, Corea del Nord, Filippine, Giappone, India, Indonesia, Pakistan, Russia, Thailandia.*
- **Europa:** *Belgio, Francia, Germania, Irlanda, Italia, Norvegia, Paesi Bassi, Regno Unito, Spagna, Svezia, Turchia.*
- **Oceania:** *Australia, Fiji, Indonesia, Nuova Zelanda, Papua Nuova Guinea, Samoa, Tonga.*

2. Preprocessing

Prima di procedere con l'analisi è stato necessario sottoporre le immagini a nostra disposizione a un'operazione di *preprocessing*, col fine di agevolare il modello nell'identificazione e nella selezione della componente più rilevante dell'immagine considerata. Tra le metodologie in uso, si è deciso di applicare il metodo *standard*, come proposto dagli autori stessi del modello di *object classification*. Importante, per il corretto funzionamento del metodo, è che le immagini date in *input* al modello abbiano la stessa distribuzione in termini di *pixel*. Per questo motivo viene applicata la cosiddetta tecnica di normalizzazione, la quale prevede la scelta di tre medie e tre *standard deviation* per applicare la normalizzazione che, nel suddetto caso, sono state calcolate da un campione casuale di immagini del *training set* di *Imagenet*. Va specificato che le medie e le *standard deviation* sono state calcolate a partire dai canali RGB (*red, green, blue*) delle foto.

Per la media i valori selezionati sono: [0.485, 0.456, 0.406], mentre per la *standard deviation* sono: [0.229, 0.224, 0.225]. I valori scelti sono tre perché i colori sono solitamente indicati da triplette di valori decimali [R, G, B], per quanto detto prima. Date le medie e le *standard deviation* è stato possibile normalizzare le immagini sottraendo da ciascun *pixel* la media e rapportando per la *standard deviation*. In seguito, si sono ridimensionate le foto a 256 *pixel* e successivamente ritagliate centralmente a 224 *pixel*. Durante questa fase, circa il 3% delle immagini sono state eliminate per incompatibilità di formato (.png).

3. Neural Network

Ai fini di raggiungere lo scopo della nostra analisi si è deciso di utilizzare il modello delle reti neurali convoluzionali, appartenente alla classe dei modelli di reti neurali. Si presenta infatti come una rete neurale ordinaria con nodi, neuroni, archi e pesi sinaptici, ma ha la caratteristica di ricevere in *input* immagini. Il modello risulta essere innovativo rispetto agli altri modelli di reti neurali perché, oltre al fatto di basarsi su un percettore multistrato, caratterizzato da una fase di *preprocessing* minimale, possiede anche la proprietà dell'*invarianza di traduzione*. In altri termini l'oggetto da identificare viene riconosciuto in qualsiasi luogo o qualsiasi posa assuma nella foto. Questo aspetto ha sempre rappresentato un limite per gli altri modelli di *object classification*, che necessitano di un allenamento costante su diverse foto di uno stesso oggetto, in luoghi e orientamenti differenti. Un aspetto interessante è comprendere come lavorano le reti neurali convoluzionali *CNN* o *ConvNet*. Queste si basano sul concetto di scomposizione dell'immagine in una matrice di *pixel*, dove ogni immagine è una matrice 3D rappresentante l'altezza, la profondità e l'ampiezza. Il primo livello della rete è chiamato *livello convulzionale*, il quale ha l'obiettivo di apprendere schemi e caratteristiche delle immagini. Quando l'immagine viene "scomposta" interviene ciò che viene chiamato *filtro* o *kernel*, ovvero una matrice di dimensioni inferiori a quella di partenza (solitamente 3X3) che filtra le sottomatrici dei *pixel*. Il risultato ottenuto da tale procedimento si definisce *Feature Map* o *Activation Map*, ovvero una somma delle moltiplicazioni fra gli elementi della matrice dei *pixel* e della matrice filtro, per ciascuna posizione. Il livello

successivo *ReLU* mappa i valori della matrice precedente convertendoli tutti in valori positivi, quindi imponendo 0 nel caso il valore di partenza sia negativo. Il livello successivo viene definito di *pooling*, dove dalla *Activation Map* si procede col cosiddetto processo di *max pooling* o *downsampling*. Questa fase permette di fare un'astrazione della matrice precedente in modo tale che la rete tratti pochi parametri significativi. Queste fasi vengono ripetute più volte all'interno del processo affinché il modello impari nei diversi livelli a riconoscere le molteplici caratteristiche dell'immagine. Da ultimo, affinché il modello possa classificare l'immagine, le informazioni vengono trasmesse al livello *fully connected*, che restituisce in *output* la classificazione dell'oggetto presente nell'immagine con la relativa probabilità associata alla etichetta. Molti sono i modelli di reti convoluzionali presenti in letteratura. Per il presente lavoro si è deciso di utilizzare il modello *VGGNet-16* (*Visual Geometry Group*). La caratteristica di tale modello è quella di possedere una matrice di filtro 3X3 e 16 strati, poiché si è dimostrato che, diminuendo il campo recettivo dei neuroni e aumentando la profondità della rete, questa performava meglio in termini di *error rate*. Si è deciso inoltre di utilizzare un modello *Vgg* pre-allenato tramite il *database* di immagini di *ImageNet*, il quale è stato sviluppato con l'obiettivo di fare *object recognition*. In particolare, il modello è stato pre-allenato su circa un milione di foto, validato su 50.000 e testato su 100.000. Questo algoritmo è in grado di classificare molti oggetti, attribuendogli una probabilità di riconoscimento, grazie alle numerose e dettagliate etichette di cui dispone. Le *labels* del modello *VGG* sono infatti 1000, motivo per cui si è deciso di aggregarle in 39 classi di natura più generale (vedi codice r), in favore di una più semplice e adeguata interpretazione.

4. Risultati

Per raggiungere l'obiettivo che tale analisi si è prefissato, dopo aver applicato il modello *vgg16* alle immagini, si è deciso di considerare gli oggetti a cui è stata associata una probabilità maggiore di 0.3, con lo scopo di aumentare i livelli di robustezza dei nostri risultati. In seguito, si è individuato l'oggetto più fotografato e postato, nel periodo di interesse, per ciascuno dei Paesi precedentemente elencati. Queste

informazioni sono sinteticamente rappresentate mediante la mappa sottostante.

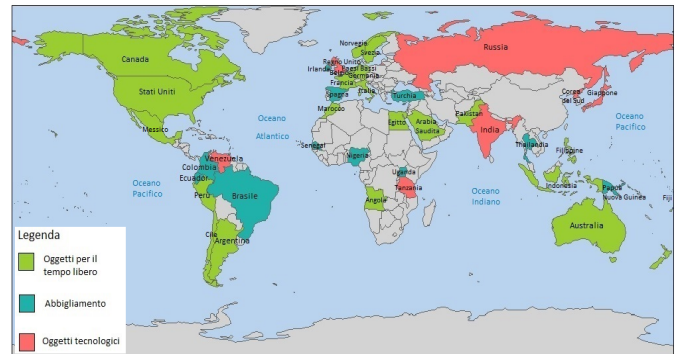


Figura 1: Mappa del Mondo colorata in base all'oggetto più fotografato e postato su Twitter dal 31 Dicembre 2018 al 2 Gennaio 2019

È evidente come gli oggetti più fotografati e postati su Twitter dal 31 Dicembre 2018 al 2 Gennaio 2019 siano rappresentati da tre principali categorie: "Oggetti per il tempo libero", "Abbigliamento" e "Oggetti tecnologici".

- **Oggetti per il tempo libero:** Stati Uniti, Canada, Perù, Cile, Argentina, Messico, Marocco, Egitto, Arabia Saudita, Angola, Pakistan, Francia, Belgio, Paesi Bassi, Germania, Norvegia, Svezia, Australia, Filippine, Nuova Zelanda, Samoa, Fiji, Italia, Indonesia
- **Abbigliamento:** Brasile, Ecuador, Colombia, Senegal, Nigeria, Uganda, Turchia, Spagna, Papua Nuova Guinea, Thailandia, Tonga, Irlanda.
- **Oggetti tecnologici:** Regno Unito, Venezuela, Russia, Tanzania, India, Corea del Nord, Giappone.

Le diverse colorazioni, attribuite a ciascuna di queste classi, mettono in luce un curioso aspetto, tale per cui, zone limitrofe risultano essere caratterizzate dalla stessa categoria. Una spiegazione di questo evento, potrebbe essere la presenza di *re-tweet*, considerato il fatto che i *post* sono pubblicati in un periodo di tempo limitato. Forti conferme di questa ipotesi si possono riscontrare nelle foto presenti in ciascuno degli stati indagati. In questo modo, risultano evidenti anche lievi differenze e analogie culturali, soprattutto quando vengono confrontati America del Nord, Asia, Europa ed Oceania. Più eterogenea è la zona geografica ricoperta dal continente africano e dall'America del Sud.

Per validare i risultati di questo progetto si è realizzato un *test* a mano, calcolando la proporzione di corrette classificazioni su 1163 foto del *set* di immagini in nostra dotazione. È risultato che il 41% delle immagini considerate nel *test* è stato correttamente classificato, mentre il 59% ha avuto un'errata interpretazione. In questo caso, è bene specificare che quest'ultimo dato è dovuto sia alle ridotte dimensioni del campione utilizzato per il *test* (limite dovuto alla natura del *test*), ma anche da errori totalmente giustificabili. Riteniamo infatti che laddove il modello sbaglia, escludendo rare eccezioni, anche l'occhio umano potrebbe essere tratto in inganno. Essendo queste, fotografie e immagini amatoriali, talvolta poco definite e nitide, ci riteniamo ugualmente soddisfatte della *performance* del modello. Alcuni limiti sono posti soprattutto dal periodo di interesse e dalla mancanza di etichette adeguate a descrivere un avvenimento come Capodanno.

5. Conclusione

In conclusione, mediante questo progetto, oltre a rispondere alla nostra iniziale domanda, si è anche creato uno spunto per un progetto di più ampie vedute. Un esempio potrebbe essere approfondire un'analisi di *brand detection* da parte di un'azienda di medie-grandi dimensioni, incuriosita da quale potrebbe essere l'impatto dei propri prodotti sui consumatori, presenti sui *social*. Aspetto che non abbiamo potuto sviluppare in prima persona per mancanza di immagini, rappresentanti la vasta gamma di prodotti su cui allenare il modello di *object classification*. Gli autori stessi del modello ritengono fondamentale aggiornare e aggiungere, ogni anno, nuove immagini alla fase di *training* del modello, per incrementare la performance di questo innovativo algoritmo.

Sitografia

- [1] ImageNet: VGGNet, ResNet, Inception, and Xception with Keras:
<https://www.pyimagesearch.com/2017/03/20/image-net-vggnet-resnet-inception-xception-keras/>
- [2] An intuitive guide to Convolutional Neural Networks:
<https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050>
- [3] An Intuitive Explanation of Convolutional Neural Networks:
<https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>
- [4] A Beginner's Guide to Convolutional Neural Networks (CNNs):
<https://skymind.ai/wiki/convolutional-network>
- [5] Image Data Pre-Processing for Neural Networks:
<https://becominghuman.ai/image-data-pre-processing-for-neural-networks-498289068258>
- [6]
<https://pytorch.org/docs/stable/torchvision/transforms.html>
- [7] Very Deep Convolutional Networks for Large-Scale Image Recognition: <https://www.arxiv-vanity.com/papers/1409.1556/>
- [8] <https://chromedevtools.github.io/devtools-protocol/>
- [9] Stretta della Cina su Twitter: chi lo usa può finire in carcere:
<https://www.lastampa.it/2019/01/13/tecnologia/stretta-della-cina-su-twitter-chi-lo-usa-pu-finire-in-prigione-KHyCTQpKC7dAFSVh104pDL/pagina.html>

[10] How Africa Tweets 2018: <https://portland-communications.com/pdf/How-Africa-Tweets-2018.pdf>

Sitografia per i codici implementati:

[1] Using Headless Chrome with Selenium: <https://blog.miguelgrinberg.com/post/using-headless-chrome-with-selenium>

[2] <https://selenium-python.readthedocs.io/api.html#module-selenium.webdriver.chrome.webdriver>

[3] <https://docs.python.org/2.4/lib/optparse-conflicts-between-options.html>

[4] <https://docs.python.org/2/library/argparse.html>

[5] Performing Sentimental Analysis on Twitter (Part 1a — Data Extraction Twitter): <https://medium.com/@williamai/performing-sentimental-analysis-on-twitter-and-facebook-part-1-acc6c6125d21>

[6] PyTorch quick start: Classifying an image: <http://blog.outcome.io/pytorch-quick-start-classifying-an-image/>

[7] rworldmap: A New R package for Mapping Global Data: https://journal.r-project.org/archive/2011-1/RJournal_2011-1_South.pdf

[8] Computing the mode in R: <https://www.r-bloggers.com/computing-the-mode-in-r/>