

Progetti di ricerca applicata basati su IA del Politecnico di Torino in ambito “Media and Entertainment”

Elena Baralis, Andrea Bottino, Luca Cagliero, Juan Carlos De Martin, Paolo Garza, Enrico Magli, Antonio Santangelo, Antonio Vetrò

Politecnico di Torino
{nome.cognome}@polito.it

Abstract

Presentiamo le attività del Politecnico di Torino inerenti all’applicazione di tecniche di Intelligenza Artificiale nell’ambito “Media and Entertainment”. Le attività presentate s’incardinano in progetti di ricerca e collaborazioni con aziende provenienti da diversi settori applicativi.

1 Introduzione

All’interno del Politecnico di Torino, diversi gruppi di ricerca hanno approcciato l’analisi di dati multimediali. Nell’ambito “Media and Entertainment”, le attività progettuali e di ricerca inerenti all’uso dell’Intelligenza Artificiale includono l’uso di tecniche di Deep Learning per l’analisi di immagini e video, l’analisi semantica di contenuti multimediali, l’elaborazione e la sintesi di informazioni visuali e la mitigazione dei problemi di bias nell’apprendimento e di discriminazione nelle decisioni automatiche.

2 Applicazione di tecniche di Deep Learning

Presso il Politecnico di Torino si utilizzano tecniche di analisi di immagini e video basate su reti neurali profonde per estrarre informazioni da immagini, video e altri tipi di “media”. In particolare, le attività dell’*Image Processing Lab* (responsabile: Prof. Enrico Magli) si incentrano sui seguenti temi: monitoraggio di ambienti indoor e outdoor, reti neurali di tipo generativo, e riduzione del rumore nelle immagini.

Nel contesto del monitoraggio sono state svolte attività di ricerca in collaborazione con il Joint Open Lab *VISIBLE* di TIM. Tali attività hanno riguardato lo sviluppo di un sistema automatico di rilevazione e lettura delle targhe ai fini dell’analisi del traffico, basato su un’analisi muti-scala tramite due reti convoluzionali. Tale sistema viene addestrato su un dataset di immagini sintetiche che può essere generato per uno o più paesi, rendendo possibile il funzionamento contemporaneo su molteplici tipi di targhe. È stato inoltre sviluppato un sistema di riconoscimento di marca e modello di veicoli, sempre basato su reti convoluzionali, che utilizzano *attention window* a scale diverse. Infine è stato affrontato il problema della segmentazione di immagini satellitari; in questo problema i dati etichettati sono particolarmente scarsi, e sono state sviluppate un’architettura e la relativa procedura di training

per massimizzare l’accuratezza della segmentazione quando le immagini di test hanno caratteristiche diverse da quelle di addestramento. Per quanto riguarda le reti di tipo generativo, sono state sviluppate reti neurali di tipo “adversarial”. Tali reti sono state applicate a problemi di autenticazione biometrica fornendo prestazioni più robuste rispetto ai sistemi tradizionali. La rete neurale di tipo “adversarial” in questo caso impara un mapping delle biometrie in uno spazio latente regolarizzato, in cui le feature delle biometrie dell’utente da autenticare sono tutte vicine tra di loro, e lontane da quelle degli altri utenti. Infine vengono impiegate reti neurali su grafi per descrivere segnali che sono privi di una rappresentazione naturale su una griglia regolare, quali per esempio le nuvole di punti. In particolare, è stata sviluppata la prima rete neurale su grafi di tipo generativo, applicata alla generazione di nuvole di punti [Valsesia *et al.*, 2018]. Le reti generative sono state anche applicate a problemi di tipo biomedico quali la generazione di nuove proteine con caratteristiche note. Le reti generative su grafo sono anche alla base delle più recenti attività nel campo del *denoising* di immagini. Si è infatti visto che, tramite il grafo, queste reti neurali sono in grado di imparare modelli *non locali* delle immagini, ovvero modelli in cui un pixel viene ritenuto simile ad alcuni suoi vicini, e anche ad altri pixel meno vicini, sfruttando le caratteristiche di regolarità delle immagini naturali. Queste reti hanno recentemente battuto il miglior algoritmo di denoising attualmente esistente per rumore Gaussiano.

Il gruppo di ricerca IPL collabora con il centro di ricerca di SONY (Sony-Eutec) per quanto riguarda le applicazioni delle reti neurali all’autenticazione biometrica. Collabora inoltre con il Joint Open Lab *VISIBLE* di TIM per le attività legate all’analisi di immagini e video per applicazioni legate al traffico, e con l’Agenzia Spaziale Europea per quanto riguarda l’analisi di immagini satellitari.

3 Analisi semantica di contenuti audio-video

Il *Database And Data Mining Group* (responsabile: Prof. Elena Baralis) raccoglie competenze relative all’analisi di dati multimediali (ad es. film, trasmissioni televisive, video didattici) basata su modelli semantici. Le fonti dati considerate includono dati eterogenei quali documenti testuali, video, trascrizioni audio e i relativi metadati estratti da cataloghi multimediali. I concetti più rilevanti identificati nei segmenti video e nelle trascrizioni video sono tradotti in formato testua-

le e processati mediante tecniche di Natural Language Processing. L'arricchimento semantico dei contenuti analizzati è basato su modelli ontologici provenienti dal Web semantico. Infine, le collezioni documentali e le trascrizioni audio sono analizzate mediante tecniche di summarization [Cagliero *et al.*, 2019] in grado di estrarre i contenuti più significativi adattandosi al contesto e alla lingua di riferimento. Le attività inerenti l'uso di modelli semantici per l'analisi di dati multimedia sono svolte in collaborazione con la Radiotelevisione Italiana S.p.A. La collaborazione prevede attualmente il finanziamento di una borsa di dottorato tematica. Il gruppo di ricerca si è inoltre classificato primo al TRECVID 2017 Hyperlinking Task.

4 Elaborazione di informazioni visuali

Il Politecnico di Torino raccoglie competenze relative all'elaborazione di informazioni visuali e dell'estrazione di dati di sintesi da esse. In particolare, il gruppo *CG&VG* (responsabile: prof. Andrea Bottino) ha sviluppato diversi lavori legati all'analisi e alla ricostruzione non intrusiva del movimento dell'intero corpo umano o di sue singole parti (mani e volto). Questi approcci, basati esclusivamente sull'utilizzo di immagini RGB o RGB-D, non richiedono al soggetto di indossare alcun sensore per la cattura del movimento. Nell'ambito Media e Entertainment, queste informazioni possono essere sfruttate per lo sviluppo di interfacce uomo macchina "naturali", basate sull'utilizzo dei movimenti e dei gesti, per l'analisi di eventi sportivi (identificazione delle fasi salienti, la creazione di sintesi automatiche e la riproduzione virtuale degli eventi), per il controllo di avatar virtuali in applicazioni legate all'intrattenimento, all'e-learning e alla produzione televisiva e per lo sviluppo di sistemi di comunicazione basati sulla virtualizzazione dei partecipanti alla comunicazione.

La ricerca nell'ambito Media e Entertainment si è anche occupata dello sviluppo di algoritmi automatici per l'identificazione delle transizioni tra inquadrature diverse all'interno di un filmato. Questo processo è un componente fondamentale per lo sviluppo di diverse applicazioni, tra cui la scomposizione del video in scene, caratterizzate da sequenze di inquadrature semanticamente correlate, l'indicizzazione automatica di archivi video di grosse dimensioni. Un altro tra i più rilevanti ambiti di ricerca del gruppo riguarda l'analisi e lo sviluppo di approcci che cercano di migliorare le qualità di robustezza e di generalizzazione delle tecniche di apprendimento automatico tramite l'integrazione di informazioni di natura diversa estratte dalla stessa immagine e/o di informazioni della stessa tipologia ma calcolate in punti diversi della stessa immagine. Queste tecniche hanno dato risultati estremamente positivi in ambiti eterogenei che vanno dall'analisi di immagini mediche, allo sviluppo di sistemi di identificazione di attacchi a sistemi biometrici di riconoscimento [Toosi *et al.*, 2017] e all'analisi dei volti umani.

Il gruppo ha collaborato a progetti di ricerca industriale con SEAT Pagine Gialle, Alenia, Altec, Sysman e Leva Engineering e ha partecipato ai progetti di ricerca finanziata TRAME (Tecnologie e Ambienti di Rete per la memoria del XXI secolo, PRIN, 2005), SMAT-F1 (progetto F.E.S.R. 2011-2012),

SMAT-F2(2013-2015) e V-PoP (Virtual Point of Presence, regionale PoloICT 2014-2018).

5 Verso una IA socialmente sostenibile

Il centro Nexa su Internet e Società (responsabile: prof. Juan Carlos De Martin) studia gli sviluppi dei sistemi intelligenti dal punto di vista della loro sostenibilità sociale [Vetrò *et al.*, 2019], con un approccio antropocentrico e interdisciplinare (informatico, giuridico, filosofico morale, sociologico e semiotico). Ad esempio, il centro Nexa studia le discriminazioni prodotte da decisioni prese da sistemi autonomi, su determinati gruppi sociali. Per risolvere questi problemi, sono state immaginate etichette grafiche che aumentino la consapevolezza sui bias e la qualità nei dati da cui attingono i sistemi intelligenti, e le potenzialità delle visualizzazioni interattive.

Il Centro Nexa studia altresì i problemi giuridici che si sollevano, quando si utilizzano sistemi di Intelligenza Artificiale, nei più svariati settori. Problemi di privacy, di proprietà dei dati da cui attingono gli algoritmi, di trasparenza, di responsabilità da parte di chi prende decisioni e agisce, sulla base del funzionamento degli algoritmi stessi. I risultati di questi lavori si possono applicare anche al settore dei media. Per esempio, il centro Nexa collabora con TIM sui problemi etici e giuridici dell'utilizzo di chatbot, ma soprattutto con AGCOM, per la regolamentazione delle campagne elettorali sulle piattaforme digitali e per il contrasto alla diffusione delle Fake News. A questi tavoli di lavoro, partecipano player come Google e Facebook, Mediaset e Rai, grandi testate giornalistiche e diverse associazioni di aziende che operano in ambito pubblicitario, nonché le associazioni dei consumatori. Su questi temi, insieme al Centro per la Riforma dello Stato e Fondazione P&R, il centro Nexa ha prodotto il rapporto "Persuasori Social". Infine, il centro Nexa sta sperimentando un approccio basato su reti neurali per ricostruire la semantica da fonti di dati eterogenee, anch'esso potenzialmente utilizzabile in molti ambiti mediatici legati al web.

Riferimenti bibliografici

- [Cagliero *et al.*, 2019] Luca Cagliero, Paolo Garza, e Elena Baralis. Elsa: A multilingual document summarization algorithm based on frequent itemsets and latent semantic analysis. *ACM Trans. Inf. Syst.*, 37(2):21:1–21:33, January 2019.
- [Toosi *et al.*, 2017] A. Toosi, A. Bottino, S. Cumani, P. Negri, e P. L. Sottile. Feature fusion for fingerprint liveness detection: a comparative study. *IEEE Access*, 5:23695–23709, 2017.
- [Valsesia *et al.*, 2018] Diego Valsesia, Giulia Fracastoro, e Enrico Magli. Learning localized generative models for 3D point clouds via graph convolution. In *Proc. of International Conference on Learning Representations (ICLR)*, 2018.
- [Vetrò *et al.*, 2019] A. Vetrò, A. Santangelo, E. Beretta, e J. C. De Martin. AI: from rational agents to socially responsible agents. *Digital Policy, Regulation and Governance*, 2019.