

L'utilizzo dell'Intelligenza Artificiale nell'ambito della media e broadcast industry

Tommaso Cesano⁺, Anna Maraga⁺, Matteo Tomei^{*}

⁺Metaliquid, ^{*}Università degli Studi di Modena e Reggio Emilia
tommaso@meta-liquid.com, anna@meta-liquid.com, matteo.tomei@unimore.it

Abstract

Il presente documento introduce il tema dell'applicazione di tecniche di Intelligenza Artificiale (IA) e nello specifico di deep learning nell'ambito dell'industria media e broadcast, che rappresenta il mercato di riferimento per Metaliquid.

I contenuti video costituiscono un asset di primaria importanza per i player di questo settore, ma il loro valore, che risiede nel conoscere cosa accade al loro interno istante per istante, non è sfruttato a causa della mancanza di metadati descrittivi.

In questo documento verranno presentati alcuni scenari di applicazione di image recognition e video analysis per l'arricchimento dei metadati dei contenuti, che aprono l'opportunità per i broadcaster di offrire nuovi servizi e modalità innovative di fruizione dei video.

1 Introduzione

Oggi il concetto di “televisione lineare” è ormai superato: lo spettatore assume un ruolo attivo e come utente può scegliere che cosa guardare all'interno di vastissimi cataloghi di contenuti, costantemente arricchiti ed aggiornati.

Un aspetto significativo di questa evoluzione è la costante crescita del volume di contenuti che i player devono gestire lungo tutta la *content supply chain*.

In quest'ottica, diventa cruciale la strategia di gestione e metadattazione dei contenuti, siano essi d'archivio o provenienti da un *live stream*.

1.1 Il ruolo dell'IA nell'estrazione di metadati descrittivi

La valorizzazione dei contenuti video è possibile laddove si abbiano molte informazioni riguardo a cosa accade nel corso del loro svolgimento. Informazioni di interesse possono essere ad esempio l'identità delle persone che compaiono, l'ambientazione specifica in cui si svolge l'azione, la presenza di dialoghi o di musica, la presenza di prodotti o brand a scopo commerciale e molto altro.

L'applicazione dell'IA è dunque cruciale per la produzione di metadati descrittivi con *timecode* puntuale di apparizione di ogni concetto o evento.

Fino ad oggi l'annotazione manuale dei contenuti era l'unico modo possibile per estrarre metadati di questo tipo. Ovviamente questo comportava un enorme sforzo in termini di tempo e costi elevati, oltre che una ridotta standardizzazione dei metadati ottenuti.

Con lo sviluppo delle GPU di ultima generazione, è stato possibile sviluppare applicazioni di deep learning anche in ambito industriale (e non solo più di ricerca) e creare architetture di reti neurali più profonde capaci di processare i contenuti video in *real time* per estrarre metadati.

A seconda del “compito” assegnato alle reti neurali, è necessario procedere con approcci specifici nel disegno delle architetture di reti neurali, nella costruzione del dataset, nel processo di addestramento delle reti e nella loro applicazione in inferenza.

1.2 Il valore di un deep learning framework proprietario in una industry con esigenze specifiche

Ogni società nel mercato media e broadcast ha esigenze specifiche per quanto riguarda la generazione di metadati, ossia l'identificazione di concetti o eventi varia da compagnia a compagnia in base alla tipologia di contenuto e agli obiettivi di business legati all'utilizzo di quei dati. Ad esempio, la gestione di un archivio di una redazione giornalistica per la ricerca veloce dei video da parte degli editor che devono creare nuovi contenuti richiede metadati diversi da quelli necessari per l'automatizzazione del processo di estrazione di azioni salienti nei contenuti sportivi.

Questi diversi casi d'uso richiedono che l'analisi dei contenuti e la restituzione dei dati possano essere effettuate in diverse fasi del workflow, in diversi formati e che l'output abbia “forme” diverse.

Inoltre, la strategia di gestione dei media è diversa da player a player e questo implica la necessità di sviluppare una soluzione che possa rispondere in modo efficiente ad ogni esigenza: servizi cloud, on-premises o soluzioni ibride.

Le esigenze di personalizzazione dei modelli per il riconoscimento di concetti o eventi, unita alle esigenze tecniche specifiche di ogni società, rendono poco applicabili le soluzioni dei big player (AWS Rekognition, Google Cloud Vision, Microsoft Azure, ecc.). Queste soluzioni sono infatti maggiormente orientate alla descrizione del mondo in modo

generico (AGI), richiedendo da un lato competenze di data science e deep learning per l'addestramento di modelli "su misura" che i player dell'industria non hanno e dall'altro rendendo impossibile rispondere ad alcune richieste specifiche che necessitano di un pieno controllo sulla tecnologia. Metaliquid ha quindi intrapreso la strada dello sviluppo di un framework di deep learning proprietario, per poter rispondere in modo efficace ed offrire un valore aggiunto nell'industria media, dove i contenuti video rappresentano il principale asset, ma il cui valore rimane ad oggi per la maggior parte inespresso.

1.3 Action localization e understanding

Un passo fondamentale per un'accurata metadateazione del contenuto video è certamente la comprensione delle molteplici azioni che si svolgono all'interno di esso. Metadateazioni di questo tipo, a loro volta, possono diventare input per i sistemi di *content recommendation* e *retrieval*, oltre ad abilitare applicazioni di riconoscimento specifiche sulla base dei requisiti imposti dal content producer e dal cliente. In questo contesto, Metaliquid sta studiando soluzioni state dell'arte per il riconoscimento di azioni in video con le quali sia possibile recuperare la regione di video in cui l'azione avviene, ottenendo quindi un'annotazione densa, concettualmente simile a quella fornita da reti di object detection su immagini. La stessa tipologia di task viene attualmente studiata anche da grandi player del mercato multimediale, come Facebook, e da importanti centri di ricerca internazionali, a testimonianza dell'importanza del problema e dell'interesse nei potenziali risultati. Nel caso specifico di Metaliquid, vengono studiate nuove modalità di estrazione e di encoding delle feature spazio-temporali e nuove modalità di training che tengano conto della presenza di azioni con pochi esempi. Nelle fasi di design particolare attenzione viene data ai requisiti computazionali delle reti, cercando di ridurre i tempi di esecuzione e i requisiti di memoria delle soluzioni sviluppate, permettendone così un'adozione più ampia e un minore costo di esecuzione.

2 Casi d'uso

L'utilizzo dei metadati trova diversi scenari di applicazione. Se infatti la soluzione offerta è prettamente business-to-business (B2B), l'impatto dell'utilizzo dell'IA può essere esteso lungo tutta la *content supply chain*, fino a permettere una significativa evoluzione della *user experience* (UX) dell'utente finale. Presentiamo di seguito alcuni casi d'uso dei metadati generati da Metaliquid.

2.1 Digital Asset Management

Le soluzioni di *Digital Asset Management* (DAM) ed archiviazione di contenuti sono uno degli scenari principali di arricchimento di metadati. La ricerca di contenuti di interesse all'interno di archivi di centinaia di migliaia di video (o immagini) è enormemente potenziata dall'associazione a ciascuno di metadati descrittivi, poiché è possibile ricercare ed individuare non solo contenuti che presentano concetti o

eventi di interesse, ma anche l'esatto segmento del contenuto a cui sono associati quei metadati.

2.2 Content recommendation e discovery

L'esperienza di fruizione di contenuti delle piattaforme *over-the-top* (OTT) e di *video on demand* (VOD) implica, come detto, un ruolo attivo dell'utente nel selezionare i contenuti da guardare. In quest'ottica, il valore del provider del servizio risiede non solo nell'offerta di qualità, ma anche nell'ottimizzazione della ricerca di contenuti in linea con gli interessi dell'utente e nel suggerimento di nuovi contenuti. L'integrazione dei metadati descrittivi nelle soluzioni di recommendation rappresenta un'opportunità di personalizzazione dell'esperienza senza precedenti. Se ad oggi i dati disponibili erano relativi alla similarità di comportamento con utenti dal profilo analogo, grazie all'utilizzo di metadati descrittivi è possibile da un lato suggerire contenuti simili a quelli che l'utente ha già visto, dall'altro evitare di proporre quelli che l'utente ha dimostrato di non apprezzare, ad esempio interrompendo la visione di un video in presenza di specifici eventi come episodi di violenza.

2.3 User experience innovative

Una UX piacevole è dunque oggi un elemento fondamentale nella strategia di acquisizione di utenti.

L'introduzione di nuove modalità di interazione sulle piattaforme è infatti una delle vie che i player perseguono per offrire un sempre maggiore coinvolgimento dell'utente.

I metadati descrittivi permettono di offrire esperienze interattive in tempo reale durante la fruizione stessa dei contenuti, mostrando informazioni contestuali a quello che sta accadendo nel video in quel preciso istante. È possibile ad esempio riconoscere ed accedere alla biografia di un politico che appare in un servizio giornalistico oppure vedere le statistiche relative a un calciatore che ha appena segnato e accedere ad una sua collezione di gol, precedentemente individuati e metadati dall'IA.

Le possibilità in questo scenario sono innumerevoli, basti pensare all'introduzione di esperienze di digital commerce legate al riconoscimento di prodotti all'interno di show televisivi.

2.4 Automatic actions recognition

I contenuti sportivi hanno bisogno di un editing in *real time* per la creazione di highlights, grafiche e clip che oggi giorno vengono fatti a mano. Questo tipo di attività presenta diversi limiti: il margine di errore è molto alto e si tratta di un processo lungo e costoso. L'utilizzo di servizi di IA per il riconoscimento automatico di azioni salienti presenta invece innumerevoli vantaggi, sia economici sia in termini di efficienza. Tale servizio permette inoltre di offrire servizi personalizzati agli utenti creando highlights che rispondono alle loro preferenze. Questa stessa tecnologia può essere utilizzata anche per automatizzare i teasers degli show televisivi, i trailer dei film e un'ampia varietà di contenuti.