

Centro di
Competenze
5G



MAMMUT:

Metodi di apprendimento automatico su dataset multi-tipo

MARCO ROMITO (UNIFI)

ALBERTO CILLI

NEVIO DUBBINI (MININGFUL)



Regione Toscana



INDICE

1. INTRODUZIONE

2. APPROCCIO PROPOSTO


3. RISULTATI

4. CONCLUSIONI

Machine Learning Multimodale

Nel mondo odierno, ricco di dati e interconnesso, le tecniche Machine Learning multimodale stanno emergendo come soluzioni per integrare informazioni da diverse fonti, come testi, immagini e dati di sensori. Questo approccio promette una comprensione più completa dei fenomeni sottostanti e offre vantaggi significativi in settori come la sentiment analysis, la diagnosi medica più accurata e la guida autonoma. Inoltre tali tecniche stanno rivoluzionando la produzione industriale, ottimizzando i processi e riducendo i costi attraverso la fusione di dati da diverse fonti.





«Gli approcci multimodali che si concentrano su immagini e dati tabulari, due delle principali modalità in cui i dati sono disponibili, sono ancora rari.

La maggior parte delle tecniche comuni si basa su un processo di elaborazione dei dati non olistico e soffre di diversi problemi, come sbilanciamento delle modalità e mancanza di integrazione dei dati.»

Un modello per immagini e dati numerici

Abbiamo sviluppato un algoritmo di machine learning per il trattamento congiunto di immagini e dati tabulari, che va oltre i metodi standard. La nostra metodologia infatti è stata progettata per catturare le relazioni nascoste tra diverse modalità, favorendo la loro interazione e per estrarre il massimo delle informazioni da ciascuna, migliorando così le performance rispetto all'utilizzo delle singole modalità o di tecniche standard di fusione dei dati.

Nella pratica il modello consiste in un classificatore che ha in input immagini opportunamente modificate. Ogni immagine originale è sottoposta a una trasformazione, codificata attraverso una rete neurale, che dipende dal dataset tabulare. In questo modo i dettagli contenuti nelle variabili numeriche vengono inglobati nell'immagine modificata e tutta l'informazione è disponibile per la previsione del classificatore.

Applicazione su due dataset

Il modello è stato applicato su due diversi dataset:

- il primo, chiamato BIKED, consiste in dati riguardanti modelli di biciclette provenienti da file CAD e contiene alcune migliaia di immagini di biciclette e diverse variabili numeriche ad esse associate.
- il secondo, DeepFashion, contiene immagini di vestiti con alcune features tabulari che le descrivono ed è composto da molte più osservazioni.

In entrambi i casi il task che il modello deve eseguire è la classificazione dell'osservazione in una delle diverse classi disponibili. I test condotti hanno mostrato che la questa metodologia, in alcune circostanze, ha prestazioni migliori rispetto alle altre soluzioni considerate per questi problemi.

4. CONCLUSIONI

Limiti e possibili sviluppi



Il modello presenta alcune limitazioni, come tempi di addestramento lunghi e richiesta elevata di risorse. Bisogna valutare se i benefici prestazionali compensano lo sforzo rispetto a soluzioni più semplici.



Il machine learning multimodale appare come un area di grande potenziale e che vale la pena esplorare maggiormente, con possibilità di trovare soluzioni innovative ed efficaci a problemi complessi.

THANK
you



GRAZIE!

Grazie!

