

STRUMENTI DI ORCHESTRAZIONE E ANALISI DI WORKFLOW NEL MACHINE LEARNING

Analisi Case Study e illustrazione Framework di Orchestrazione di
pipeline di lavoro

Simone Boldrini

Alma Mater Studiorum - Università di Bologna
Facoltà di Scienze

13 Ottobre 2021

Apprendimento Automatico

*L'**apprendimento automatico** é un ramo dell'Intelligenza Artificiale, che raccoglie metodi in grado di migliorare la performance di un algoritmo, autonomamente, nell'identificare pattern di dati.*

Gli algoritmi li suddividiamo in 3 categorie:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Modello

Un Modello di ML é l'output generato in seguito all'addestramento dell'algoritmo.

I modelli di ML assimilano i dati di training, con l'obiettivo di individuare potenziali relazioni tra i dati di input e quelli di output.

1 Casi di Studio

- MLN
- Insider Threat Detection

2 Analisi

3 Framework

Apprendimento automatico nel Campo delle Reti

L'**apprendimento automatico** permette ai sistemi di imparare automaticamente a prendere decisioni o predizioni basati sull'esperienza. Con la sviluppo di ML in questo campo, ricercatori e operatori di rete possono affrontare vari tipi di reti e applicazioni; i quali possono cambiare a seconda delle performance e dei requisiti.

Supervised Learning

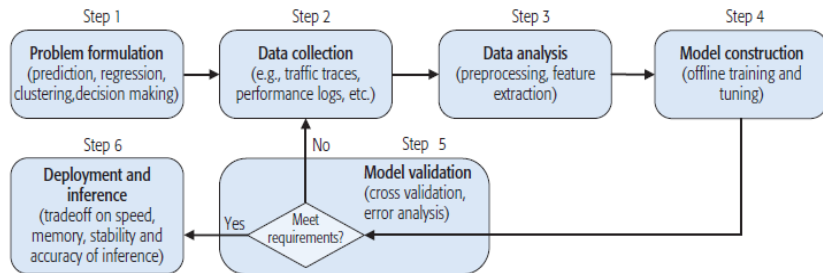
- Traffic Prediction
- Traffic Classification
- Routing Strategy
- Throughput prediction

Unsupervised Learning

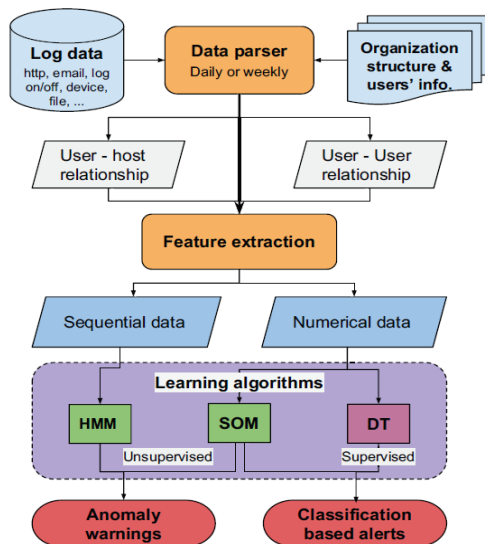
- Optimizing QoE

Reinforcement Learning

- Resource Manager{Job Scheduler}
- TCP Congestion Control

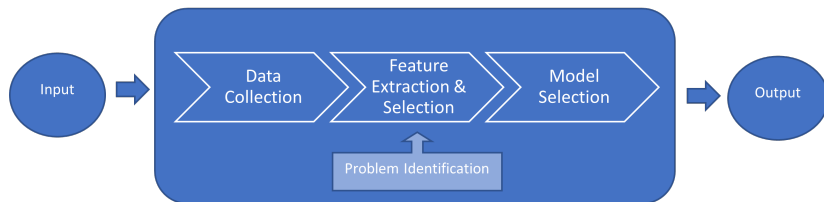


Insider Threat Detection



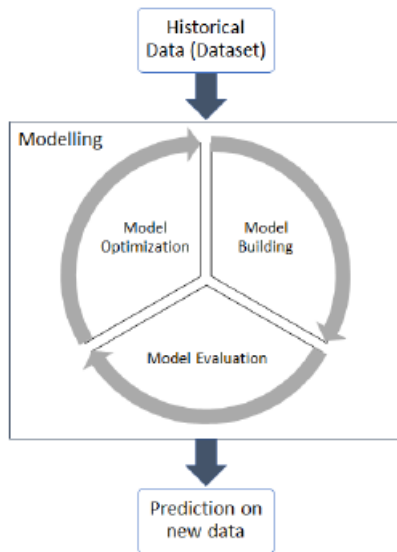
Invariante di Processo

Analizzando i diversi casi di studio, abbiamo cercato di astrarre i problemi evidenziando un pattern generico di invariante.



La raccolta dati é fondamentale in ogni processo di apprendimento automatico. Viene diversificata a seconda del problema che andremo a trattare. Di solito i dati vengono suddivisi a seconda della **metodologia di raccolta**.

Elaborazione del Modello



Risultato

In quest'ultima fase definiamo l'output desiderato:

- Classificazione
- Regressione
- Clustering

Classification

- Traffic Classification
- Image Recognition [Radiology]
- Speech Recognition
- Face2Face Traslation
- Classification based Alerts

Regression

- Traffic Prediction
- Prediction/Forecasting

Clustering

- Resourse Management [Networking]
- Anomaly warning

Strumenti di Orchestrazione

Questi strumenti che permettono l'orchestrazione di pipeline di Machine Learning hanno l'obiettivo di semplificare il processo di gestione e automatizzare l'implementazione dei modelli di ML.

Gli strumenti che andremo a mostrare di seguito sono tutti *open-source*, e si focalizzano su 3 punti chiave:

- raccolta dati
- creazione e implementazione del modello
- distribuzione(permettendo inoltre la riproduzione ed il monitoraggio)

