

STRUMENTI DI ORCHESTRAZIONE E ANALISI DI WORKFLOW NEL MACHINE LEARNING

Analisi case study e illustrazione framework di orchestrazione di pipeline

Simone Boldrini

Alma Mater Studiorum - Università di Bologna
Facoltà di Scienze

13 Ottobre 2021

Obiettivo: ricerca di un'astrazione di un'invariante di processo per i processi di Machine Learning

- Casi di studio

Obiettivo: ricerca di un'astrazione di un'invariante di processo per i processi di Machine Learning

- Casi di studio
- Analisi e confronto tra CS

Obiettivo: ricerca di un'astrazione di un'invariante di processo per i processi di Machine Learning

- Casi di studio
- Analisi e confronto tra CS
- Workflow Generico

Obiettivo: ricerca di un'astrazione di un'invariante di processo per i processi di Machine Learning

- Casi di studio
- Analisi e confronto tra CS
- Workflow Generico
- Framework

Apprendimento Automatico

*L'**apprendimento automatico** 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

Apprendimento automatico nel Campo delle Reti

Qui di seguito mostriamo un'insieme di problemi nel campo del Networking che permette l'integrazione con l'apprendimento automatico

Supervised Learning

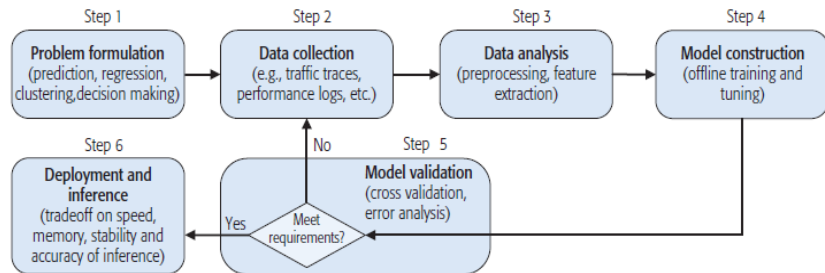
- Traffic Prediction
- Traffic Classification
- Routing Strategy
- Throughput prediction

Unsupervised Learning

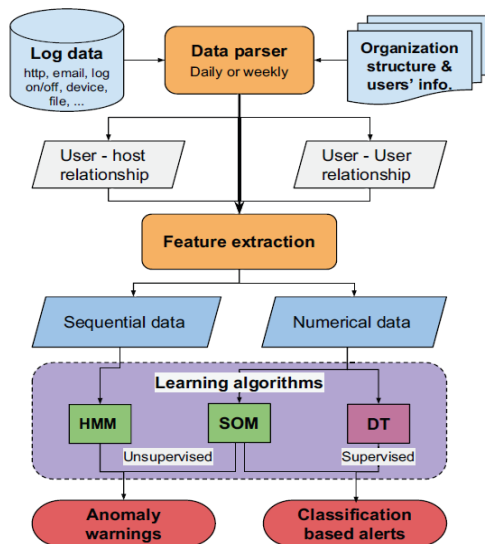
- Optimizing QoE

Reinforcement Learning

- Resource Manager{Job Scheduler}
- TCP Congestion Control

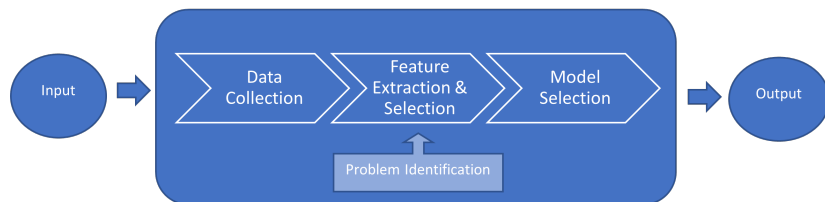


Insider Threat Detection



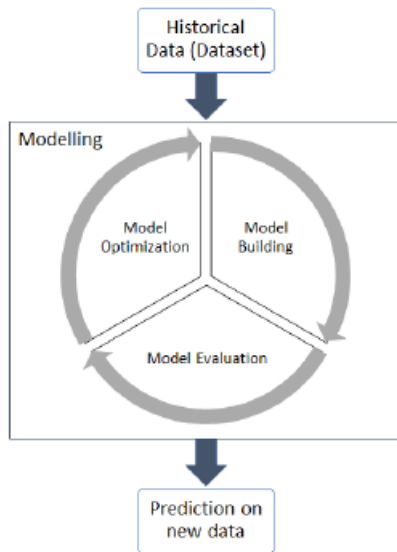
Invariante di Processo

Analizzato 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 Managment
[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)

ZenML

- Libreria Python
- Permette affiancamento ad altro strumento di orchestrazione
- Garantisce riproducibilità degli addestramenti

Flyte

- Basato su python e K8s
- Estendibile attraverso diversi plug-in
- gestisce più di 10mila Workflow

Kedro

- GUI
- Modulare
- Favorisce il versioning

MLRun

- GUI
- Servizio Server-less [architettura permette di convertire codice in microservizi]
- Vasto reperto di plug-in

ZenML

- Libreria Python
- Permette affiancamento ad altro strumento di orchestrazione
- Garantisce riproducibilità degli addestramenti
- Permette di memorizzare gli stati della pipeline nella cache

- GUI
- Modulare
- Favorisce il versioning

- Basato su python e K8s
- Estendibile attraverso diversi plug-in
- Già molto diffuso: gestisce più di 10mila Workflow

- GUI
- Servizio Server-less [architettura permette di convertire codice in microservizi]
- Vasto reperto di plug-in