L'I.A. APPLICATA AL MONITORAGGIO E AL TEST AUTOMATION DELLE RETI DI NUOVA GENERAZIONE. - D.A.I.MON
PROG N. F/350156/04/X60 - CUP: B39J23002440005 - COR: 16106924

PROGETTO DI RICERCA

L'attività dell'assegnista consisterà nello studio e individuazione di un insieme di algoritmi di Intelligenza Artificiale da utilizzare nell'ambito del monitoraggio del traffico di rete TelCo in particolare su reti 5G Stand-Alone.
Gli algoritmi ricercati dovranno essere in grado di rilevare singoli eventi o andamenti anomali (AI Detection), determinare l'andamento futuro degli indicatori di interesse (AI Prediction), definire metriche in grado di valutare e contribuire a migliorare la Quality of Experience degli utenti (AI QoE).
Si valuteranno inizialmente algoritmi noti in letteratura (conformemente a quanto descritto in precedenza), verificandone l'applicabilità al contesto e verificando la necessità di eventuali adeguamenti; in alternativa si definiranno modifiche o affinamenti a tali algoritmi per renderli meglio rispondenti allo scopo. Le direzioni di ricerca che si ritengono di maggiore interesse in questa fase di indagine sono:

i) Analisi della rete usando graph neural networks. Graph neural networks sono un paradigma emergente che permette sia di inferire informazioni riguardo ad una rete di interesse e alle caratteristiche dei collegamenti, sia di classificare fenomeni inaspettati sulla rete stessa. Una direzione interessante è quella di analizzare l'utilizzabilità di queste tecniche per tutti gli scopi del progetto, ossia sia per **anomaly detection**, sia per **predizione**, sia per il **miglioramento della QoE** degli utenti.

ii) Reinforcement learning per intelligent monitoring. Reinforcement learning può essere utilizzato in diversi aspetti del processo di **anomaly detection**, specialmente qualora non sia possibile utilizzare tecniche di supervised learning.

iii) Causal Inference. Un'altra area di forte interesse è quella di causal inference, cioè estrazione di relazioni di tipo causale su eventi. Proponiamo di esplorare approcci di diverso tipo al problema di causalità sia utilizzando i metodi proposti da Rubin, sia quelli di Pearl, sia potenziali combinazioni.

iv) Una delle direzioni di ricerca più interessanti è quella di considerare metodi che combinano i due approcci menzionati sopra e quelli basati su machine learning. Metodi di causal inference possono essere utilizzati anche per counterfactual analysis per studiare potenziali what-if scenarios, per planning e per delineare risposte real-time a situazione di criticità. Tutto questo a supporto degli obiettivi di **predizione** e **miglioramento della QoE**.

v) Infine, si prenderanno in considerazione metodi di continuous e active learning. Queste tecniche sono proposte in letteratura per affrontare in modo efficiente il problema della generalizzazione dei modelli: è dimostrato che modelli addestrati su determinati dataset, caratterizzati da uno specifico contenuto in termini di eventi normali e anomali, difficilmente riescono a classificare in modo efficace tracciati che si discostano da tale schema. L'approccio più comunemente proposto per risolvere il problema consiste nel fondere quanti più dataset possibile, ma ciò non affronta (o peggiora) alcuni aspetti pratici fondamentali:

   a. l'obsolescenza dei dati utilizzati e acquisiti dal modello, che col passare del tempo (o anche improvvisamente) possono diventare non rappresentativi dello scenario reale
   b. l'onere del procedimento di etichettatura, che è il fattore di costo dominante nella costruzione e manutenzione del modello: non è pensabile di poter raccogliere centinaia di migliaia di esempi etichettati da parte di un esperto ogni volta che il modello deve essere adeguato allo scenario corrente

   Metodologie di interesse per la risoluzione strutturale di questi problemi, che consentano di realizzare una pipeline di analisi dei flussi di rete in quasi-tempo-reale, tollerando cambi anche repentini del modello, evidenziando la presenza di eventi non classificabili entro le categorie usate per l'addestramento, e richiedendo la quantità minore possibile di dati etichettati, sono:

   c. Continuous Learning (CL): si occupa dell'apprendimento dai flussi di dati. Nello specifico CL comprende insiemi di tecniche per l'addestramento di modelli ML da flussi di dati tali che i) il modello è in grado di estendere la sua conoscenza da nuovi dati e ii) il modello conserva la conoscenza dai dati passati, cioè non si limita a sovrascrivere la vecchia conoscenza con le nuove conoscenze.
   d. Active Learning (AL): si occupa di incrementare in modo intelligente il dataset di addestramento. In particolare, le pipeline AL consentono a un modello ML di scegliere (ovvero, interrogare) quali punti dati etichettare (simile a uno studente che fa domande a un insegnante). Ad alto livello, AL consente di selezionare in modo intelligente i datapoint più informativi da apprendere per il modello ML.

PIANO ATTIVITÀ

L'analisi sopra descritta porterà all'identificazione degli algoritmi che meglio si adattano allo scenario applicativo, confrontando l'efficacia di classificatori basati su deep learning, classificatori supervisionati delle principali famiglie (decision trees, k-nearest neighbors, support vector machines, logistic regression, linear discriminant analysis, statistici es. Bayes), classificatori non supervisionati delle principali famiglie (clustering, statistici, reti neurali come le self-organizing maps, angle-based, neighbor, density, isolation forests), valutata in particolare con riguardo alla tipologia dei dati da analizzare.

Una volta selezionati tali algoritmi saranno sottoposti ad una successiva prototipizzazione (prima), ingegnerizzazione e validazione (poi) negli interagendo con gli altri attori del progetto DAIMON, responsabili dello sviluppo sperimentale. È importante sottolineare il fatto che questo tipo di attività richiede un continuo processo di verifica e conseguente validazione proprio per la sua natura. Infatti, trattandosi di algoritmi che si basano su meccanismi di autoapprendimento (di diverso tipo e natura) e dei quali si verifica a posteriori l'efficacia, è molto probabile che sarà necessario rivedere iterativamente le soluzioni adottate in funzione dei risultati che si otterranno con la loro applicazione su sequenze di dati sempre più consistenti e prolungate nel tempo, valutati con le metriche prestazionali tipicamente impiegate per algoritmi di classificazione di questo tipo, quali il coefficiente F1, l'accuratezza, il coefficiente di correlazione di Matthews, il fattore di recall, e le matrici di confusione. Per questa ragione si prevede che queste attività si sviluppino lungo gran parte della durata del progetto biennale, prevedendo per questo la possibilità di un rinnovo dell'assegno, intendendo che nella prima fase verrà svolta l'analisi e selezione iniziale, dopodiché man mano che si avranno risultati dalla prototipazione queste scelte andranno attentamente validate, eventualmente estendendole o modificandole.

L'attività prevederà lo svolgimento dei seguenti task:

I) **(M01-M04)** *studio analitico dei nessi causali tra le principali metriche e misure di rete nei vari contesti, con particolare riferimento ai protocolli di rete 5G, e indagine su modelli e classi di algoritmi esistenti.* Lo studio sarà orientato ad individuare le informazioni di dettaglio (ovvero i parametri di rete e gli indicatori) cui applicare gli algoritmi di Detection / Prediction / misura della QoE per le finalità prefissate (ovvero per verificare il corretto funzionamento della rete, le sue performance e la sua sicurezza). Si effettuerà un censimento e un'analisi degli algoritmi di Intelligenza Artificiale e dei modelli esistenti in letteratura, valutandone la possibile applicazione per il rilevamento di comportamenti anomali sul traffico di rete, per analisi predittive e di misura della QoE.

II) **(M05-M13)** *selezione algoritmi di Anomaly Detection, di Prediction, e di misura della QoE e studio su eventuali adattamenti*; saranno selezionati gli algoritmi di Intelligenza Artificiale più promettenti per l'Anomaly Detection sul traffico di rete 5G. Si valuteranno e studieranno inoltre eventuali necessità di adattamenti / miglioramenti del modello teorico rispetto al contesto del progetto corrente.

III) **(M010-M18)**: *costruzione prototipi SW per algoritmi IA di Anomaly Detection, Prediction e QoE;*
Per ciascuno degli algoritmi IA individuato nei tre ambiti (Detection, Prediction, QoE), sarà realizzato un software prototipo. Lo sviluppo comincerà in parallelo alle attività di studio previste al punto II, man mano che i vari algoritmi saranno disponibili. Ciascun prototipo sarà realizzato attraverso un processo di back-end indipendente, in grado di leggere l'insieme dei dati da sottoporre all'analisi (da un file o da un database) e di produrre in uscita il risultato dell'algoritmo di IA (su file o su DB).

IV) **(M13-M18)**: *studio di eventuali nuovi algoritmi e loro affinamento iterativo.*
Sulla base delle analisi effettuate e dei risultati ottenuti con gli studi previsti nei task precedenti, si valuterà la possibilità di definire nuovi algoritmi di Intelligenza Artificiale da applicare al progetto.
Sulla base delle sequenze di dati sempre più consistenti via via acquisite e dei risultati delle attività derivanti dalla prototipazione, le definizioni saranno attentamente validate, eventualmente estendendole o modificandole.

V) **(M14-M24)**: *integrazione dei risultati dell'applicazione degli algoritmi di AI all'automazione tramite interazione con il piano di controllo ed i sistemi di orchestrazione.*
Sarà effettuata una valutazione complessiva comparativa degli algoritmi IA attraverso i risultati ottenuti utilizzando i prototipi. Sarà selezionato un sotto-insieme degli algoritmi più efficaci, da utilizzare nelle successive fasi di ingegnerizzazione, test e validazione