Tesi di Laurea Magistrale in Logic & Artificial Intelligence e Big Data Analytics

Prof. Tommaso Di Noia
Politecnico di Bari

Evaluation on RecSys

Walter Anelli, Claudio Pomo (https://calendly.com/claudio_pomo)

Impact of hyperparameters on evaluation offline: Nello studio offline delle performance di un sistema di raccomandazione è utile comprendere come uno specifico protocollo influenzi la valutazione o come un particolare set di hyperparametri (per il modello in analisi) affligga la valutazione stessa. Nel corso degli ultimi anni gran parte della ricerca si è concentrata sulla fruizione delle diverse tipologie di protocolli di valutazione ma poco si è approfondito su che impatto questi avessero, in combinazione con un particolare set di hyperparametri, sulle diverse famiglie di metriche utili ad analizzare i diversi aspetti di un sistema di raccomandazione.

Beyond accuracy metrics: Basare la valutazione dei sistemi di raccomandazione esclusivamente sulle metriche di precisione o accuratezza non è più sufficiente per valutare compiutamente un sistema di raccomandazione. L'esperienza utente è diventata centrale nella valutazione di questi sistemi: come l'utente percepisce i suggerimenti proposti? La diversificazione dei suggerimenti porta un beneficio apprezzabile? Come è possibile misurare il grado di sorpresa dell'utente rispetto ad un particolare suggerimento proposto? Queste sono solo alcune domande che oggi muovono questo particolare campo di ricerca per i sistemi di raccomandazione. Si indagherà su come le metriche di novelty, diversity, covarege e serendipity possono migliorare o possono supportare le performance di questa classe di agenti software intelligenti.

Link Prediction on Knowledge Graph

Claudio Pomo (https://calendly.com/claudio_pomo)

Machine Learning on Knowledge Graph: La grande mole di dati oggi giorno a disposizione ha reso inevitabile il superamento dei classici sistemi per la memorizzazione degli stessi. Una delle strutture sviluppate più promettenti sono i Graph DataBase dai quali prendono vita i Knowledge Graph che hanno visto negli ultimi anni un'enorme diffusione grazie anche al loro impiego nell'ecosistema di Google. Un filone di ricerca molto promettente è quello che si

impegna ad investigare come le tecniche e i task consolidati di machine learning siano applicabili direttamente su questo tipo di strutture. Tuttavia non risulta semplice metchare i classici approcci e modelli di neural networks sulle strutture a grafo: lo studio degli embeddings, l'applicazione delle diverse tecniche graph specific per (deep) neural network, nuove soluzioni per il training dei modelli sono solo alcuni degli hot topic di quest'area applicativa e di ricerca.

Link prediction techniques: Con l'avvento dei Knowledge Graph, delle moderne tecniche di machine learning e la necessità di un accesso personalizzato ai dati, hanno sempre più preso piede studi di nuovi approcci che fornissero una misura di quanto buone fossero le performance dei nostri modelli di machine learning sui grafi: individuare e collegare tra loro nodi del grafo potenzialmente connessi. L'approccio classico per questo task si basa sui concetti di similarità (Common Neighbour, Jaccard's Index, Adamic/Adar Index, Katz, ecc); tuttavia a noi interessa sfruttare approcci di machine learning sia classici sia deep per completare i potenziali collegamenti mancanti della nostra graph sturcture. Allo stesso tempo, però, siamo anche interessati a valutare l'evoluzione di questi collegamenti: così come i link si vanno a formare allo stesso modo potrebbero dissolversi fornendoci così un'evoluzione temporale del nostro sistema. Tale approccio è molto diffuso nello scenario di continuation task come ad esempio playlist musicali.

Robustness of Cross-Domain Recommender Models to Adversarial Perturbations

Felice Antonio Merra (https://calendly.com/felice-merra)

L'applicazione delle tecniche di Adversarial Machine Learning nell'ambito dei sistemi di raccomandazione include l'analisi di robustezza dei sistemi di raccomandazione ad adversarial examples. Tale linea di ricerca è stata avviata con la pubblicazione del modello Adversarial Personalized Ranking for Recommendation. Tale tecnica è stata successivamente valutata su altri modelli di raccomandazione in seguito elencati.

Adversarial Personalized Ranking for Recommendation.
Adversarial Collaborative Neural Network for Robust Recommendation
Adversarial Collaborative Auto-encoder for Top-N Recommendation
Adversarial Training Towards Robust Multimedia Recommender System.
Enhancing the Robustness of Neural Collaborative Filtering Systems Under Malicious Attacks
Adversarial Sampling and Training for Semi-Supervised Information Retrieval.
Adversarial tensor factorization for context-aware recommendation.

Lo scopo di tale lavoro di tesi è quello di analizzare, implementare e valutare l'impatto di tale tecnica nell'ambito di altri modelli di racommandazione, in particolare su quelli cross-domain.

Tra i modelli da analizzare il lavoro di tesi dovrà includere i seguenti modelli:

- Collective Matrix Factorization (CMF)

- SocialMF
- DeepSoR

Adversarial Attacks on Oblivious Recommender

Felice Antonio Merra (https://calendly.com/felice-merra)

Uno dei lavori più recenti nell'ambito degli Attacchi ai sistemi di raccomandazione è Adversarial attacks on an oblivious recommender. Questo lavoro ha aperto una nuova linea di investigazione nell'ambito della sicurezza dei RS in quanto per la prima volta ha proposta la generazione di profili di attacco tramite tecniche di Machine Learning. Tale lavoro non ha previsto alcuno confronto con le tecniche di attacco esistenti.

L'obiettivo di questo lavoro di tesi è quello di implementare l'algoritmo di attacco proposto in Adversarial attacks on an oblivious recommender e confrontarlo con le tecniche di attacco che costituiscono lo stato dell'arte.

Il confronto dovrà essere eseguito sulla base di differenti modelli di machine learning comunemente usati nei RS quali User-kNN, Item-kNN e SVD.

Named Entity Recognition & Named entity Linking

Giovanni Maria Biancofiore (https://calendly.com/giovannimaria-biancofiore)

Nell'ambito del Natural Language Processing un task di rilievo è l'identificazione di Entità significative all'interno di un testo non strutturato. Tale procedimento risulta essere cruciale per identificare la semantica contenuta in una frase, un passo o un intero componimento espresso in linguaggio naturale.

Syntax Tree: si vuole esplorare la possibilità di utilizzare determinate grammatiche per la costruzione di alberi sintattici al fine di poter individuare entità e relazioni di rilevo all'interno del testo non strutturato espresso in linguaggio naturale, secondo approcci che possono essere supervisionati e non. Mentre i primi risultano alquanto diffusi e fortemente domain oriented, gli ultimi sono ancora oggetto di ricerca carenti di soluzioni pratiche funzionanti/performanti.

Searching on Linked Data: si vogliono utilizzare particolari tecniche di ricerca su Basi di Conoscenza espresse tramite Linked Data, al fine di poter individuare all'interno di un testo espresso in linguaggio naturale entità e relazioni rilevanti, etichettandoli opportunamente secondo gli standard previsti dal Knowledge Graph interrogato. Si valuteranno nel dettaglio le performance caratteristiche del sistema così implementato, in relazione alle possibili tecniche pre-esistenti.

Aggregation of Models in Federated Machine Learning

Antonio Ferrara (https://calendly.com/sciueferrara)

Nell'ambito del Machine Learning, i recenti requisiti in termini di privacy obbligano ad una revisione del modo di collezionare i dati privati degli utenti. Una delle proposte più interessanti in questa direzione, suggerita da Google, prevede l'addestramento di una deep neural network in maniera distribuita, mantenendo i dati privati su ogni dispositivo e aggregando i modelli risultanti in un unico modello globale. Questa operazione finale, deve permettere di trovare un modello che sappia descrivere i dati di tutti gli utenti della federazione. Una delle sfide già affrontate e delle quali si vuole estendere e perfezionare l'approccio è relativa alla determinazione di una misura di qualità dei modelli locali e dei relativi dispositivi, al fine di enfatizzare il contributo dei migliori per ottenere il miglior modello possibile. La ricerca andrà svolta chiedendosi: "Come contraddistinguere un buon modello locale? Come posso ottenere informazioni sulla qualità senza dover raccogliere e "rubare" dati privati? Davvero il modello complessivo ha bisogno dei migliori modelli per generalizzare?".

Il lavoro prevederà sia uno studio dello stato dell'arte, sia una profonda comprensione del lavoro già svolto, e infine una sua estensione teorica e sperimentale nelle direzioni descritte.

Federated Machine Learning & Recommender Systems

Antonio Ferrara (https://calendly.com/sciueferrara)

Nell'ambito del Machine Learning, i recenti requisiti in termini di privacy obbligano ad una revisione del modo di collezionare i dati privati degli utenti. Una delle proposte più interessanti in questa direzione, suggerita da Google, prevede l'addestramento di deep neural network in maniera distribuita, mantenendo i dati privati su ogni dispositivo e aggregando i modelli risultanti in un unico modello globale. Un'interessante estensione di questa architettura è relativa al dominio dei sistemi di raccomandazione, ovvero quella famiglia di modelli tesa a suggerire e raccomandare agli utenti articoli, oggetti o prodotti di possibile gradimento. Per far ciò, infatti, gli attuali modelli di sistemi di raccomandazione, devono conoscere il più possibile l'utente, quali sono le sue preferenze, quali articoli o oggetti ha visto, guardato, acquistato o scartato, quali posti ha visitato, etc. In questo lavoro di tesi si esploreranno possibili tentativi di ripensare gli attuali modelli di sistemi di raccomandazione in modo da mantenere privati tutti o parte di questi dati, al fine di preservare al massimo la privacy dell'utente ma continuando a godere dei vantaggi di questi sistemi.

Il lavoro prevederà sia uno studio dello stato dell'arte sul Federated Machine Learning, uno studio sui più importanti modelli di sistemi di raccomandazione, e infine una loro connessione teorica e sperimentale, secondo gli obiettivi descritti.