



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Empowering AI through Enriched Data: Combining Knowledge Graphs and Machine Learning

2526-114R-2-03

Titolo

Arricchire i Dati per Migliorare l'IA: Come Combinare Grafi di Conoscenza e Apprendimento Automatico

Docente(i)

Matteo Palmonari

Lingua

Inglese

Breve descrizione

Le applicazioni di Intelligenza Artificiale (AI) si basano in larga misura sui dati, che spesso si presentano in formati eterogenei, semi-strutturati e non strutturati. Il corso offre una panoramica delle metodologie progettate per migliorare la qualità e l'utilità dei dati che alimentano i sistemi di AI attraverso l'uso dei Knowledge Graphs (KG).

Il corso inizia con un'introduzione ai KG come modelli di dati per la rappresentazione e il ragionamento della conoscenza.

Successivamente, verrà analizzato come questi modelli di dati, combinati con tecniche di Machine Learning (ML), possano fornire soluzioni efficaci per l'arricchimento dei dati nelle applicazioni di AI. Verranno considerati due principali paradigmi di arricchimento dei dati basati sulla combinazione di KG e ML: (1) integrazione semantica dei dati, e (2) KG embeddings e predizione di collegamenti (link prediction).

(1) integrazione semantica dei dati

L'integrazione semantica dei dati si riferisce al processo di costruzione di repository di conoscenza centrati sulle entità, in grado di integrare informazioni provenienti da fonti eterogenee. Il corso si concentrerà principalmente su questo paradigma, presentando diverse tecniche di ML per interpretare, annotare, interconnettere e arricchire dati tabulari e testuali. Queste tecniche includono metodi di NLP basati su KG, come l'interpretazione e l'annotazione semantica di tabelle, la riconciliazione e il collegamento di entità denominate, e il clustering di entità.

Sulla base di questi fondamenti, verrà mostrato come applicare il paradigma dell'integrazione semantica in importanti applicazioni downstream, tra cui i sistemi di Question Answering e le applicazioni di Retrieval-Augmented Generation (RAG).

(2) KG embeddings e predizione di collegamenti (link prediction)

Infine, il corso introdurrà brevemente il secondo paradigma, che si concentra sull'apprendimento di embedding delle rappresentazioni dei KG e sul loro utilizzo per la predizione di collegamenti. In questo contesto, i collegamenti predetti sono interpretati come triple inferite a partire da caratteristiche latenti, piuttosto che da regole di ragionamento esplicite.

CFU / Ore

2CFU / 16 Ore

Periodo di erogazione

Giugno-Luglio 2026

Sustainable Development Goals

IMPRESE, INNOVAZIONE E INFRASTRUTTURE
