IT for CEOs & CFOs is published by House of Words Media Limited.

The current issue and full text archive of this journal is available on **https://www.itceoscfos.com**

Technology and Innovation

Knowledge Graphs Push the Boundaries of Data Science

Maya Natarajan

Maya Natarajan
Senior Director Product Marketing
Neo4j

Biography

Maya Natarajan is Senior Director Product Marketing at native graph database leader Neo4j (https://www.neo4j.com).

Responsible for knowledge graphs at Neo4j, Maya is passionate about bringing different technologies together to solve complex problems. At Neo4j, Maya is championing the use of knowledge graphs to bring context to various systems. Maya has positioned technologies from Blockchain to Predictive & User-Based Analytics to Machine Learning to Deep Learning to Search to BPM and beyond in a myriad of industries at various small and large companies.

Maya started her career in the biotechnology area where she was in R&D focusing on cardiovascular drugs, and she has five patents to her name.

Keywords Knowledge Graphs, Graph databases, Artificial Intelligence (AI), Data analytics
Paper type Research

Abstract

Defined by The Turing Institute, the UK's the national institute for data science and artificial intelligence, as the best way to "encode knowledge to use at scale in open, evolving, decentralized systems," Knowledge Graphs[1] are a big trend in the advanced data science world, but they aren't as widely known as they could be. In this article, graph database expert Maya Natarajan asks: Does the real magic of knowledge graphs come into play when they are used to get the best from machine learning?

Introduction

Organizations are increasingly using Artificial Intelligence (AI) for decision-making. However, due to a lack of contextual information, such systems have not yet been able to achieve their full potential as reliable ways to find solutions for complex business and social problems.

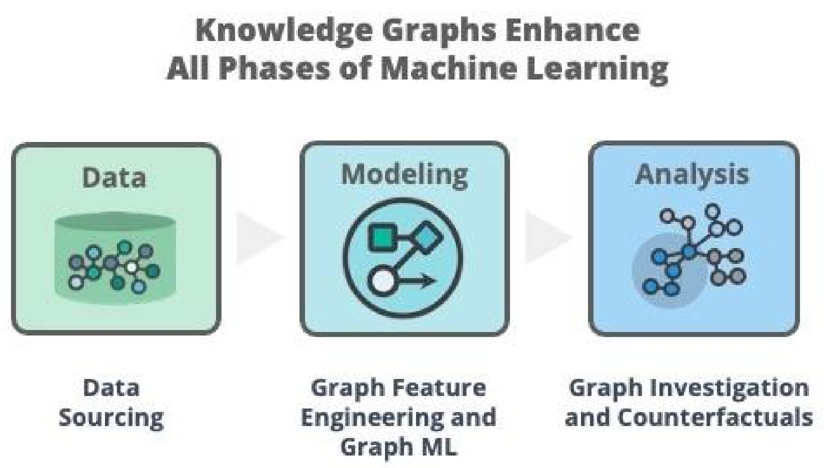
Recently, we asked 100 senior tech executives – CIOs, CTOs, and Chief Data Officers – what they think they need to solve this problem. A massive 88% said the same thing: knowledge graphs.

Given that these executives represent large organizations across verticals using graph technology for a wide array of use cases, something's clearly going on. If

IT for CEOs & CFOs is published by House of Words Media Limited.
The current issue and full text archive of this journal is available on **https://www.itceoscfos.com**

you are not familiar with the term, 'knowledge graphs' are a proven way to capture data relationships and convey their meaning. Knowledge graphs are defined by Stanford University[2] as "a compelling abstraction for organizing world's structured knowledge over the internet, and a way to integrate information extracted from multiple data sources".

Figure 1: Knowledge Graphs enhance all phases of machine learning



That's a good definition, for knowledge graphs drive intelligence into the data itself in many ways. They give AI the 'context' it needs to be more explainable, accurate, and repeatable. Knowledge is every company's most prized asset. Its value is limited, however, unless organizations can leverage that knowledge in the correct context. While neither AI/machine learning (ML) or knowledge graphs are new technologies, it's only lately that they have come of age to derive that context.

The powerful combination of the two is spurring an explosion of interest in Contextual AI. Machine learning is enhanced using knowledge graphs because of their innate ability to reveal context. Contextual information is known to increase predictive accuracy. Context also makes decisioning systems more flexible and it provides a framework for tracking data lineage.

This is where knowledge graphs come into play. A knowledge graph places data in context by establishing connections among data. A knowledge graph enriches the data's meaning and utility by adding a layer of semantics, thereby allowing software agents to reason about it. By adding relationships to data and enhancing it with semantics, knowledge graphs drive intelligence into data, making it smarter.

IT for CEOs & CFOs is published by House of Words Media Limited.

The current issue and full text archive of this journal is available on **https://www.itceoscfos.com**

That's useful, as machine learning is used in every industry. In healthcare, it's being used to better detect cancers. In supply chains, to find factors that positively and negatively impact business. In financial services, ML is used to allow investors to identify new opportunities or know when to trade.

Why data scientists give up too easily on context

Unfortunately, most current data science approaches leave out contextual information. Connections are difficult to process in standard relational databases. However, knowledge graphs capture and make this contextual information usable, as they utilize graph data structures. They can enhance every step of the machine learning process – from data sourcing and training machine learning models to analyzing predictions and applying results. Contextualized AI systems end up more reliable, robust, explainable, and trustworthy. If you can't trust the data used for ML, you can't trust the results, after all.

In the initial step of data sourcing, knowledge graphs are used for data lineage to track the data that feeds machine learning: knowledge graphs track where the data came from, how the data changed, where the data is used, and who used it. Data lineage and master data management using knowledge graphs also serve as an audit trail for compliance, especially in regulated industries.

The next phase is training a machine learning model. This involves providing an ML algorithm with training data and significant features to learn a function for making predictions. Machine learning models without context require exhaustive training, strictly prescriptive rules, and can only be applied to specific applications. Knowledge graphs also allow graph feature engineering using simple graph queries and/or more complex graph algorithms. We know that relationships are highly predictive of behaviour, so using these connected, contextual features maximises the predictive power of models while increasing how broadly a solution can be applied.

Once a machine learning model has been developed, it is essential to understand if it is useful and makes correct predictions. Knowledge graphs with incorporated relationship information allow for easy graph investigations and counterfactual analysis. A domain expert might test hypotheses by exploring similar communities in the knowledge graph. They might also question odd results by drilling into hierarchies and dependencies.

Knowledge graphs built on graph technologies have significant advantages as graphs naturally store, compute and analyze connections and relationships among data. Moreover, graph algorithms are specifically developed to leverage the topology of data through connections. The algorithms find communities, uncover influential components, and infer patterns and structure. Incorporating the predictive elements of context from a knowledge graph into machine learning not only increases accuracy but reduces false positives.

Leveraging knowledge graphs with machine learning

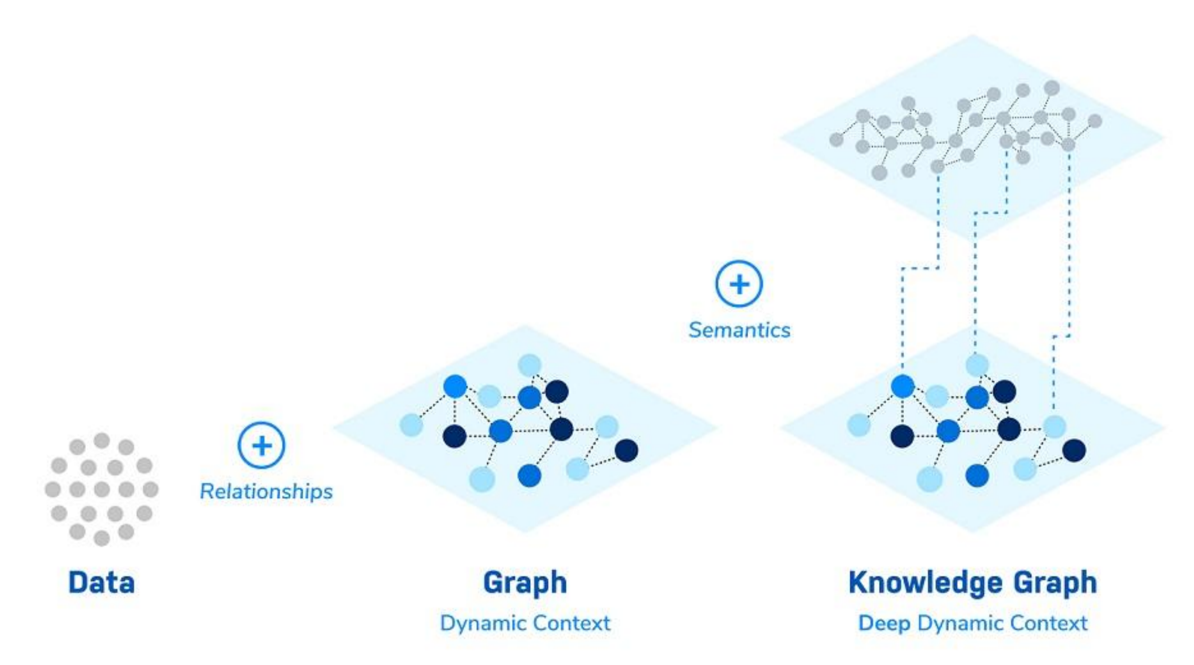
Graph-native learning involves computing machine learning tasks within a graph structure. It takes knowledge graph augmented machine learning to the next level.

IT for CEOs & CFOs is published by House of Words Media Limited.

The current issue and full text archive of this journal is available on https://www.itceoscfos.com

It provides the ability to learn generalized, predictive features directly from within the graph. You needn't know what data structures are most predictive. This is significant, as organizations don't always know which features are most important. Organizations also don't know how to represent connected data for use in machine learning models: again, knowledge graphs help.

Figure 2: Knowledge Graphs provide deep dynamic context



Neo4j users today are leveraging knowledge graphs with machine learning for many use cases. Such use cases include enhancing heuristics to more complex uses like training embeddings in a graph-native learning model. A global e-commerce leader that has created a shopping bot to add context to machine learning to make better heuristic decisions about user intent. Finally, a global pharmaceutical company that is combining a knowledge graph, graph queries and graph algorithms with traditional ML approaches to map and predict patient journeys.

AI-enhanced knowledge graphs are driving the next wave of competitive advantage for companies that use them together successfully.

Reference

[1] The Alan Turing Institute, *Knowledge graphs - How do we encode knowledge to use at scale in open, evolving, decentralized systems?* Available at https://www.turing.ac.uk/research/interest-groups/knowledge-graphs

[2] Stanford University, *Knowledge Graphs - What should AI know?* Available at: https://web.stanford.edu/class/cs520/2020/notes/What_is_a_Knowledge_Graph.html