IDC ANALYZE THE FUTURE

IDC EXECUTIVE SUMMARY

Sponsored by: Amazon Web Services

Developing AI-enabled applications that use machine learning and deep learning is increasingly popular across both the enterprise software market and the application development market.

Streamlining the Development of Deep Learning Applications in an AI-Enabled World

January 2019

Written by: David Schubmehl, Research Director, Cognitive/Artificial Intelligence Systems

Adapted from the Whitepaper: *Streamlining the Development of Deep Learning Applications in an AI-Enabled World* (February 2018, IDC #US43529118)

Introduction

Developing artificial intelligence (AI)-enabled applications across machine learning (ML) and deep learning (DL) has become increasingly popular across both the enterprise software market and the application development market. Organizations are seeing AI/ML/DL algorithms as being better ways to get closer to their customers, improve business processes, and reduce costs. Organizations, data scientists and developers are feeling the pressure from their management teams, boards and even customers to grab on to this new wave of computing.

However, as with many emerging technologies, critical success factors around ML/DL implementations are related to the people, process and technologies involved. Emerging technical solutions have traditionally required sharp and motivated developers who like to live on the cutting-edge of technology. This takes significant time and energy, as well as experienced developers to really build and deploy meaningful ML/DL solutions.

This IDC Executive Summary looks at some of the trends around this nascent technology and how vendors like Amazon Web Services (AWS) have developed services and tools for DL, which allow data scientists and developers to deploy DL models more quickly and easily.

Situation Overview

The market for ML and DL–based AI applications has grown rapidly and continues to surge. IDC estimates that worldwide spending on AI, ML, and DL solutions will exceed US$77 billion by 2022. And by 2026, IDC predicts 75% of all enterprise software will include some aspect of ML/DL for predictions, recommendations or advice.

Interest in AI has not escaped the Asia/Pacific region. IDC believes that Asian businesses are ready to invest in the technology to stay ahead of their competition. And in doing so, many are planning to adopt open source standards.

- In Asia, IDC estimates that spending on ML and DL solutions will rise from US$3.6 billion in 2018 to over US$15.5 billion in 2022 with a compound annual growth of over 49%.
- In a recent IDC Asia survey, 40% of respondents indicated that they are using or planning to use open source cognitive/AI or ML technology.

However, there are significant challenges faced by enterprises wanting to adopt these new technologies:

- Disparate tools and technologies are freely available but knowing where to start and which tools or technologies to use can be confusing to organizations and their data scientists and developers.
- Lack of integrated development environments for ML slows down the cycle of experimentation, development, testing and production.
- Absence of suitable developer skills in ML and data science to use these tools makes this more difficult for organizations.
- Application programming interface (API) and/or template-based solutions designed for use with prebuilt domains do exist but locating them and making use of them for a project can be problematic at best. The question for enterprises is how best to develop these ML models while minimizing the amount of effort and time needed to develop accurate predictive and prescriptive models.

Considering Amazon SageMaker: A Fully Managed Service for Machine Learning

To ease and expedite the deployment of these models, AWS created Amazon SageMaker, a fully managed ML platform that enables data scientists and developers to quickly and easily build, train and deploy ML models at any scale, removing all the barriers that typically slow down developers who want to use ML today.

Amazon SageMaker provides high-performance and scalable ML algorithms optimized for speed, scale and accuracy. Data scientists and developers can choose from supervised algorithms where the correct answers are known during training and the model can be corrected where it made mistakes. Amazon SageMaker also includes support for unsupervised learning (i.e., the algorithms must discover the correct answers on their own), such as with k-means clustering and principal component analysis (PCA), to solve problems like identifying customer groupings based on purchasing behavior. For organizations trying to build sophisticated ML-enabled solutions, Amazon SageMaker also provides a fully managed environment for reinforcement learning.

SageMaker can be used to solve a wide range of challenges difficult to work out with traditional coding methods ranging from traditional algorithm development to using other forms of ML that may not be as flexible or dynamic as DL. SageMaker allows data scientists and developers to quickly set up, test and iterate DL/ML solutions at a speed that is not possible with standard ML/DL frameworks and languages and then deploy those solutions into production quickly and easily.

SageMaker provides the tools and capabilities for organizations to solve challenges from pricing optimization and offer management to predictive maintenance and even intelligent process automation. Organizations like GE Healthcare are using SageMaker to improve patient outcomes via the development of DL algorithms that provide predictive and prescriptive diagnoses to their clients and customers. Digital Globe is using SageMaker to train geospatial models against petabytes of Earth observation imagery datasets so that its customers can just push a button, create a model and deploy it.

AWS has continued to develop and improve SageMaker. In 2018, it introduced over 100 new features and capabilities, including Workflows, Experiment Management and Version Control. These new features and capabilities make it even easier to develop and deploy ML and DL models more quickly. In addition, AWS also announced many new ML-based compute initiatives such as AWS Inferentia and Amazon Elastic Inference which SageMaker can take full advantage of. Finally, AWS is also offering an ML category of products offered by AWS Marketplace that includes over 150+ algorithms and model packages, which can be used and customized with SageMaker. As organizations begin to develop AI-enabled applications using ML and DL models, they should evaluate tools and services like SageMaker.

Challenges and Opportunities

The AI market is already crowded and getting more competitive with every passing day. The need and desire for better (and simpler) tools, quicker time to market and efficiency are key concerns in the AI-enabled applications market. There are numerous established and emerging vendors addressing and providing services and solutions within this space at a very wide range of capabilities. As such, AWS faces the challenge of continuing as a leader in this market and will need to maintain an aggressive pace of engineering and innovation. Although AWS is productizing open source DL/ML services as the foundation of their solutions, this approach is also not new to this market. What is new is that managed services like Amazon SageMaker and the AWS Deep Learning AMIs (Amazon Machine Image) combine numerous DL tools, frameworks and technologies into a single integrated platform that provides significant productivity enhancements for organizations, data scientists and developers. AWS needs to keep providing this level of innovation and expertise in this emerging market.

Conclusion

Great AI-enabled applications require both advanced technology and solid design judgement. Organizations should make sure the AI-enabled solution they are building will be able to help achieve the desired business outcome and/or address the issues that it is planned to be overcome by utilizing DL. Engage in-house subject matter experts, the right stakeholders and consulting partners to help develop the right use cases to align with the desired business outcome. Make sure to include past project experiences in the organization's design thinking approach and, if available, include predefined use cases that have been developed for peers within the organization's industry to help develop the optimal use cases for the desired outcome. This process should involve continuous innovation and prototyping until the right use cases have been developed.

There are a wide variety of tools and libraries available, but it is not always clear which libraries are best for the use cases or jobs that data scientists and developers should accomplish to successfully develop AI-enabled applications. Offerings such as Amazon SageMaker and the AWS Deep Learning AMIs provide the ways and means for data scientists and developers to become more productive and deploy DL models, as well as supporting services such as data curation, integration and management to solve a wide range of challenges that are difficult to work out with traditional coding methods and address the organization's business needs. Organizations should be evaluating tools and services like these as they begin to develop and deploy AI-enabled applications using DL models.

MESSAGE FROM AMAZON WEB SERVICES

Visit ***https://aws.amazon.com/events/aws-innovate/machine-learning/*** to read the full White Paper and join AWS experts as we cover a broad range of AI, ML and DL topics and provide technical depth on architecture and codes.

Whether you are a beginner or an experienced user, you can learn something new at AWS Innovate Online Conference – ML and AI Edition. You can also hear from customers who are partnering with AWS for innovative AI and learn how to build, train and deploy sophisticated models with any framework and unlock an intelligent tomorrow, today.

About the analyst:

David Schubmehl, *Research Director, Cognitive/Artificial Intelligence Systems*

Dave Schubmehl is Research Director for IDC's Cognitive/Artificial Intelligent Systems and Content Analytics research. His research covers information access and artificial intelligence technologies including content analytics, search systems, unstructured information representation, cognitive computing, deep learning, machine learning, unified access to structured and unstructured information, Big Data, visualization, and rich media search in SaaS, cloud and installed software environments.

IDC Corporate USA

5 Speen Street
Framingham, MA 01701, USA

T 508.872.8200

F 508.935.4015

Twitter @IDC

idc-insights-community.com

www.idc.com