A Quick Look At Image Processing With Deep Learning

Snappy Ubuntu Core For Embedded And IoT Devices

₹120
ISSN-2456-4885

OpenSource

THE COMPLETE MAGAZINE ON OPEN SOURCE

ForYou

Volume: 06 | Issue: 01 | Pages: 116 | October 2017

EFY GROUP Publication

Data Science And Machine Learning: Making Machines Smarter

An Introduction To A Distributed Deep Learning Library

Who Said Machine Learning Is Rocket Science?

HOW DO MACHINES LEARN?

The Smart Cube

Success Story: The Smart Cube Uses Open Source To Deliver Custom Research And Analytics Services

An Interview With Chip Childers, Co-founder, Cloud Foundry Foundation

5 of the 5 Top Websites

Are Powered by MySQL™

The #1 Open Source Database

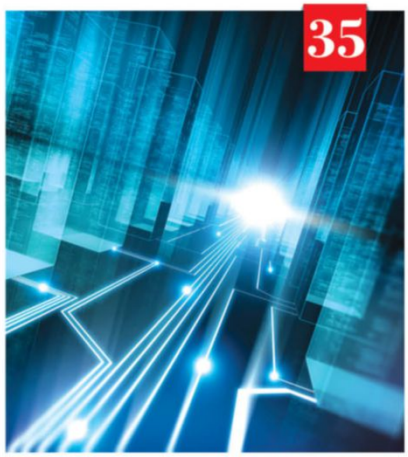
ORACLE®

oracle.com/mysql
or call 0008001005870 email: mysql-sales_in_grp@oracle.com

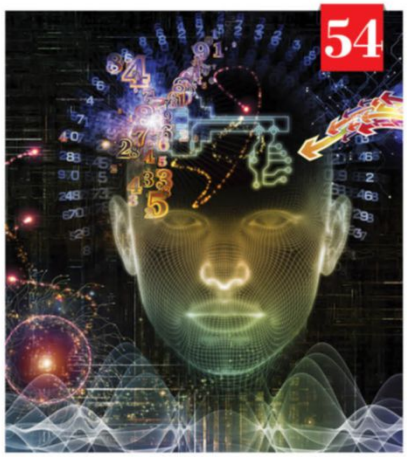
CONTENTS

OCTOBER 2017 | ISSN-2456-4885

Understanding How a **Neural Network Works Using R**



Surfing the Digital Wave with **ADDC**



A Quick Look at **Image Processing** with Deep Learning

REGULAR FEATURES

98

Who Said **Machine Learning** is **Rocket Science?**

42

"Cloud adoption is increasing but there are a lot of security concerns"

Sandip Kumar Panda, CEO, InstaSafe

22

"There are very few roadblocks for developers who use Cloud Foundry"

Chip Childers,
co-founder, Cloud Foundry Foundation

66

"Things aren't getting easier for most coders"

Janani Ravi,
former Googler and Stanford alumna with her son

DVD OF THE MONTH

Linux operating systems and programs that you can use every day.

Linux Mint 18.2 Cinnamon (64-bit live)
PCLinuxOS 64 MATE 2017.07 Desktop
Linux Lite 3.6 32-bit
Clonezilla 2.5.2.17
GParted 0.29.0

EDITOR
RAHUL CHOPRA

EDITORIAL, SUBSCRIPTIONS & ADVERTISING
DELHI (HQ)
D-87/1, Okhla Industrial Area, Phase I, New Delhi 110020
Ph: (011) 26810602, 26810603; Fax: 26817563
E-mail: info@efy.in

MISSING ISSUES
E-mail: support@efy.in

BACK ISSUES
Kits 'n' Spares
New Delhi 110020
Ph: (011) 26371661, 26371662
E-mail: info@kitsnspares.com

NEWSSTAND DISTRIBUTION
Ph: 011-40596600
E-mail: efycirc@efy.in

ADVERTISEMENTS
MUMBAI
Ph: (022) 24950047, 24928520
E-mail: efymum@efy.in

BENGALURU
Ph: (080) 25260394, 25260023
E-mail: efyblr@efy.in

PUNE
Ph: 08800295610/ 09870682995
E-mail: efypune@efy.in

GUJARAT
Ph: (079) 61344948
E-mail: efyahd@efy.in

JAPAN
Tandem Inc., Ph: 81-3-3541-4166
E-mail: japan@efy.in

SINGAPORE
Publicitas Singapore Pte Ltd
Ph: +65-6836 2272
E-mail: singapore@efy.in

TAIWAN
J.K. Media, Ph: 886-2-87726780 ext. 10
E-mail: taiwan@efy.in

UNITED STATES
E & Tech Media
Ph: +1 860 536 6677
E-mail: usa@efy.in

Printed, published and owned by Ramesh Chopra. Printed at Tara Art Printers Pvt Ltd, A-46,47, Sec-5, Noida, on 28th of the previous month, and published from D-87/1, Okhla Industrial Area, Phase I, New Delhi 110020.

SUBSCRIPTION RATES			
Year	Newstand Price (₹)	You Pay (₹)	Overseas
Five	7200	4320	—
Three	4320	3030	—
One	1440	1150	US$ 120

Kindly add ₹ 50/- for outside Delhi cheques.
Please send payments only in favour of EFY Enterprises Pvt Ltd.
Non-receipt of copies may be reported to **support@efy.in**—do mention your subscription number.

Microsoft's Project Brainwave offers real-time AI

Expanding its footprint in the artificial intelligence (AI) world, Microsoft has unveiled a new deep learning acceleration platform called Project Brainwave. The new project uses a set of field-programmable gate arrays (FPGA) deployed in the Azure cloud to enable a real-time AI experience at a faster pace.

The system under Microsoft's Project Brainwave is built on three main layers: a high-performance distributed system architecture, a DNN engine synthesised on FPGAs, and a compiler and runtime for low-friction deployments of trained models. The extensive work on FPGAs by the Redmond giant enables high performance through Project Brainwave. Additionally, the system architecture assures low latency and high throughput.

One of the biggest advantages of the new Microsoft project is the speed. FPGAs on the system are attached directly with the network fabric to ensure the highest possible speed. The high throughput design makes it easier to create deep learning applications that can run in real-time.

"Our system, designed for real-time AI, can handle complex, memory-intensive models such as LSTMs, without using batching to juice throughput," Microsoft's distinguished engineer Doug Burger wrote in a blog post.

It is worth noting that Project Brainwave is quite similar to Google's Tensor Processing Unit. However, Microsoft's hardware supports all the major deep learning systems. There is native support for Microsoft's Cognitive Toolkit as well as Google's TensorFlow. Brainwave can speed up the predictions from machine learning models.

Apache Kafka gets SQL support

Apache Kafka, the key component in many data pipeline architectures, is getting SQL support. San Francisco-based Confluent has released an open source streaming SQL engine called KSQL that enables developers with continuous, interactive queries on Kafka.

The latest announcement is quite important for businesses that need to respond to SQL queries on Apache Kafka. The same functionality was earlier limited to Java or Python APIs.

Google releases Android 8.0 Oreo with new developer tweaks

Google has released Android 8.0 Oreo as the next iteration of its open source mobile platform. The latest update has a list of tweaks for developers to let them build an enhanced user experience.

"In Android 8.0 Oreo, we focused on creating fluid experiences that make Android even more powerful and easy to use," said Android's VP of engineering, Dave Burke in a blog post.

Android 8.0 is a result of months of testing by developers and early adopters who installed and tested its preview build on their devices. Also, it is designed to make the Android ecosystem more competitive with Apple's iOS.

Android 8.0 Oreo comes with the picture-in-picture mode that enables developers to provide an advanced multi-tasking experience on their apps. The feature was originally available on Android TV but is now on mobile devices, enabling users to simultaneously run two apps on the screen. Google has added a new object to enable the picture-in-picture mode. The object, called PictureInPictureParams, specifies properties such as the active app's preferred aspect ratio.

Android Oreo features other consistent notifications too. There are changes such as notification channels, dots and timeout. You just need to use a specific method to make notifications through your apps better on Android 8.0. Google has also added features such as downloadable fonts, and adaptive icons to upgrade the interface of existing apps. Likewise, the platform has WebView APIs and support for Java 8 language features. There are also the ICU4J Android Framework APIs that reduce the APK footprint of third-party apps by not compiling the ICU4J libraries in the app package.

KSQL provides an easier way to leverage the real-time data on Kafka. Any developer who is familiar with SQL can readily use KSQL on Kafka to build solutions. The platform has a familiar syntax structure and does not require mastery of any complex infrastructure or a programming language. Moreover, KSQL coupled with Kafka's scalable and reliable environment is expected to add a lot of value to Kafka users.

"Until now, stream processing has required complex infrastructure, sophisticated developers and a serious investment. With KSQL, stream processing on Apache Kafka is available through a familiar SQL-like interface, rather than only to developers who are familiar with Java or Python. It is the first completely interactive, distributed streaming SQL engine for Apache Kafka," said Neha Narkhede, co-founder and CTO, Confluent, in a statement.

KSQL for streaming data is quite different from traditional rational SQL databases. The data is unbound, whereas the queries are continuously running and producing results. Confluent believes that it is easier to learn additional concepts and constructs while using a familiar language and tools.

Confluent has made major progress with Kafka. The platform has become the top choice for real-time enterprise application development. It has also become more than just data ingestion in recent years.

Oracle shifts Java EE 8 to open source

After years of speculation, Oracle has finally disclosed its plans of open sourcing Java EE 8. The company is shifting the latest Java Enterprise Edition to an open source foundation at the time of launching the v8.0.

Oracle has maintained the open source Java project for years, but there were recently some complaints that the company was shifting the Java EE engineering team on to other projects. Oracle had eventually restated its commitment to support Java EE last year. However, the Java community has so far been demanding that the company run the project independently.

David Delabassee, a software evangelist at Oracle, published a blog post announcing the company's decision. "Although Java EE is developed in open source with the participation of the Java EE community, often the process is not seen as being agile, flexible or open enough, particularly when compared to other open source communities," he said.

Moving Java EE core technologies, reference implementations and its test compatibility kit to an open source foundation will help the company to adopt more agile processes and implement flexible licensing. The change in the governance process is certainly quite important for a widely adopted project like Java EE.

In the official blog post, Delabassee said that Oracle will encourage innovation

Apache Software Foundation develops library for scalable in-database analytics

The Apache Software Foundation has released Apache MADlib as a new top-level project that helps deliver scalable in-database analytics. The new release is a result of discussions between database engine developers, data scientists, IT architects and academics who were looking for advanced skills in the field of data analysis.

Apache MADlib provides parallel implementations of machine learning, graphs, mathematical and statistical methods for structured and unstructured data. It was initially a part of the Apache Incubator. "During the incubation process, the MADlib community worked very hard to develop high-quality software for in-database analytics, in an open and inclusive manner in accordance with the Apache Way," said Aaron Feng, vice president of Apache MADlib.

Starting from automotive and consumer goods to finance and government, MADlib has been deployed by various industry verticals. It helps to deliver detailed analytics on both structured and unstructured data using SQL. This ability makes the open source solution an important offering for various machine learning projects.

"We have seen our customers successfully deploy MADlib on large-scale data science projects across a wide variety of industry verticals," said Elisabeth Hendrickson, vice president of R&D for data, Pivotal. Apache MADlib is available with Apache License 2.0. A project management committee (PMC) helps its daily operations and in community development.

CoreOS Tectonic 1.7 comes with support for Microsoft Azure

CoreOS, the container management vendor, has released a new version of its enterprise-ready Tectonic platform. The new release brings Kubernetes to Microsoft's Azure. Debuted as CoreOS Tectonic 1.7.1, the new platform is based on Kubernetes v1.7. The latest Kubernetes integration arrived in May. But the new version has expanded that release with stable Microsoft Azure support. This makes Tectonic a good solution for multi-cloud environments.

"Tectonic on Azure is an exciting advancement, enabling customers to use CoreOS' enterprise-ready container management platform to easily manage and scale workloads, to build and manage these applications on Azure," said Gabriel Monroy, lead product manager for containers, Azure, Microsoft. The new Azure support comes as an extension to the previous Tectonic version's compatibility with Amazon Web Services and bare metal servers. Also, since CoreOS focuses exclusively on Linux containers, there is no support for Windows containers on Azure in the latest release.

In addition to Azure, Tectonic 1.7.1 supports pre-configured monitoring alerts via Prometheus. There is also alpha support for Kubernetes network policies to help control inbound traffic and provide better security. Besides, the open source solution has fixes for common issues like latency of customer applications.

You can download the latest Tectonic version from the official CoreOS website. Users who are operating Tectonic 1.6.7-tectonic.2 with Operators can enable the new release using one-click automated updates.

by taking a hands-off approach for Java EE. This positive move will benefit the entire Enterprise Java community.

Java EE 8 is yet to launch; the version was originally supposed to be debuted at JavaOne in 2016. However, it has now been suggested that the new Java EE will be released sometime later this year. Oracle's Delabassee has hinted that specifications of Java EE 8 have almost been completed.

Microsoft aims to expand in the 'big computing' space with new acquisition

Microsoft has acquired cloud-focused Cycle Computing. The new acquisition will help the company expand its presence in the world of 'big computing', which includes high-performance computing (HPC), to cater to the growing demands of enterprises.

Utilising the resources from Cycle Computing, Microsoft is set to upgrade Azure to compete strongly with Amazon Web Services and Google Compute Engine. The Greenwich, Connecticut-based company has its flagship orchestration suite CycleCloud, which will enable Azure to more deeply support Linux workloads and provide easier switching from Linux and Windows on-premise workloads to the cloud.

"As customers continue to look for faster, more efficient ways to run their workloads, Cycle Computing's depth and expertise around massively scalable applications make it a great fit to join our Microsoft team," said Microsoft Azure corporate vice president Jason Zander, in a blog post.

As a software provider for orchestration computing, Cycle Computing has so far been supporting Amazon Web Services and Google Compute Engine. However, the company will now largely favour Azure against the other leading cloud offerings. "We see amazing opportunities in joining forces with Microsoft — its global cloud footprint and unique hybrid offering is built with enterprises in mind," stated Jason Stowe, founder and CEO, Cycle Computing.

Founded in 2015, Cycle Computing started its operations with the open source high-throughput framework HTCondor. But with the emergence of cloud computing, the company started developing solutions for cloud environments.

Raspberry Pi gets a fix for Broadpwn Wi-Fi exploit

Days after the release of Debian 9, the Raspberry Foundation has brought out a new Raspbian OS version. The new update, codenamed Stretch, includes a list of optimisations and fixes a vulnerability that had impacted several mobile devices and single-board computers in the past.

Called Broadpwn, the bug was discovered in the firmware of the BCM43xx wireless chipset in July this year. It affected a wide range of hardware, including Raspberry Pi 3 and Pi Zero W, as well as various iPhone and iPad models. Potentially, the zero-day vulnerability lets an attacker take over the wireless chip and execute a malicious code on it. The Stretch release comes with a patch for the loophole to avoid instances of any hacks and attacks on Raspberry Pi.

While the Jessie build had PulseAudio to enable audio support over Bluetooth, the new Raspbian release has the bluez-alsa package that works with the popular ALSA architecture. You can use a plugin to continue to use PulseAudio.

The latest version also has better handling of usernames other than the default 'pi' account. Similarly, desktop applications that were previously assuming the 'pi' user with passwordless sudo access will now prompt for the password.

Raspbian Scratch has additionally received an offline version of the Scratch 2 with Sense HAT support. Besides, there is an improved Sonic Pi and an updated Chromium Web browser.

The Raspberry Pi Foundation recommends that users update their single-board computers using a clean image. You can download the same from its official site. Alternatively, you can update your Raspberry Pi by modifying the sources.list and raspi.list files. The manual process also requires renaming of the word 'jessie' to 'stretch'.

Docker Enterprise Edition now provides multi-architecture orchestration

Docker has upgraded its Enterprise Edition to version 17.06. The new update is designed to offer an advanced application development and application modernisation environment across both on-premises and cloud environments.

One of the major changes in the new Docker Enterprise Edition is the support for multi-architecture orchestration. The solution modernises .NET, Java and mainframe applications by packaging them in a standard format that does not require any changes in the code. Similarly, enterprises can containerise their traditional apps and microservices and deploy them in the same cluster, either on-premises or in the cloud, irrespective of operating systems. This means that you can run applications designed for Windows, Linux and IBM System Z platforms side by side in the same cluster, using the latest mechanism.

"Docker EE unites all of these applications into a single platform, complete with customisable and flexible access control, support for a broad range of applications and infrastructure and a highly automated software supply chain," Docker product manager, Vivek Saraswat, said in a blog post.

In addition to modernising applications, the new enterprise-centric Docker version has secure multi-tenancy. It allows enterprises to customise role-based access control and define physical as well as logical boundaries for different teams sharing the same container environment. This enables an advanced security layer and helps complex organisational structures adopt Docker containers.

The new Docker Enterprise Edition also comes with the ability to assign grants for resource collections, which can be services, containers, volumes and networks. Similarly, there is an option to even automate the controls and management using the APIs provided.

Docker is offering policy-based automation to enterprises to help them create some predefined policies to maintain compliance and prevent human error. For instance, IT teams can automate image promotion using predefined policies and move images from one repository to another within the same registry. They can also make their existing repositories immutable to prevent image tags from being modified or deleted.

Google develops TensorFlow Serving library

Google has released a stable version of TensorFlow Serving. The new open source library is designed to serve machine-learned models in a production environment by offering out-of-the-box integration with TensorFlow models.

First released in beta this February, TensorFlow Serving is aimed at facilitating the deployment of algorithms and experiments while maintaining the same server architecture and APIs. The library can help developers push multiple versions of machine learning models and even roll them back.

Developers can use TensorFlow Serving to integrate with other model types along with TensorFlow learning models. You need to use a Docker container to install the server binary on non-Linux systems. Notably, the complete TensorFlow package comes bundled with a pre-built binary of TensorFlow Serving.

TensorFlow Serving 1.0 comes with servables, loaders, sources and managers. Servables are basically underlying objects used for central abstraction and computation in TensorFlow Serving. Loaders, on the other hand, are used for managing a servable life cycle. Sources include plugin modules that work with servables, while managers are designed to handle the life cycle of servables.

The major benefit of TensorFlow Serving is the set of C++ libraries that offer standards for support, for learning and serving TensorFlow models. The generic core platform is not linked with TensorFlow. However, you can use the library as a hosted service too, with the Google Cloud ML platform.

Atom 1.19 text editor gets official with enhanced responsiveness

Atom has announced the release of the next version of its text editor. Debuted as Atom 1.19, the new open source text editor update comes with an upgrade to Electron 1.6.9. The notable change in Atom 1.19 is the improved responsiveness and memory usage. The integration of a native C++ text buffer has helped to smoothen the overall performance and operations of the text editor. Also, the key feature of Git and GitHub integration, which was introduced in Atom 1.18, has been improved with new tweaks in version 1.19.

Ian Olsen, the developer behind Atom, said that the improvements in Atom 1.19 are the new steps in the 'continued drive' to deliver a fluent experience for large and small files. Large files consume less memory in Atom 1.19. In the same way, file saving in the latest Atom version happens asynchronously without blocking the UI.

Atom 1.19 comes with a full rewrite of the text editor's rendering layer. This version has restored the ability to return focus to the centre. There is also an optimised native buffer search implementation that removes trailing whitespaces. The new text editor version also comes with the 'showLineNumbers' option set to false, by default. Atom follows the tradition of pushing the stable release along with the next beta version, and has released Atom 1.20 beta for public testing. The beta release offers better support for Git integration. Olsen has added a new API that can be used for observing dock visibility, along with fixes for PHP grammar support.

RaspAnd OS now brings Google Play support to Raspberry Pi 3

RaspAnd, the popular distribution for Raspberry Pi devices, has received a new build. Debuted as the RaspAnd Build 170805, the new version comes with Android 7.1.2 Nougat and includes Google Play support.

RaspAnd developer Arne Exton has released the new version. Exton has ported Google Play Services to enable easy app installations, as well as provided users with a pre-installed Google Apps package that comes with apps such as Chrome, Google Play Games, Gmail and YouTube. The team has also worked on improving the video performance in this version.

Along with providing extensive Google Play integration, the new RaspAnd OS has addressed the screen flickering issue that was reported in the previous versions. The latest release also includes Kodi 17.3 media centre, and apps such as Spotify TV, ES File Explorer and Aptoid TV.

RaspAnd Nougat build 170805 is available for existing users as a free update. New users need to purchase an image for US$ 9 and install it on their machines using an SD card. You can use the Win32 disk manager utility or the GNU/Linux operating system.

The new RaspAnd build is specifically designed for Raspberry Pi 3 systems. Due to some higher resource requirements, the distribution is not compatible with previous Raspberry Pi models.

Google's Deeplearn.js brings machine learning to the Chrome browser

Google has developed an open source library called Deeplearn.js to enable an integrated machine learning experience on Chrome. The library helps to train neural networks without requiring any app installations. It exploits WebGL to perform computations on a GPU level.

"There are many reasons to bring machine learning (ML) into the browser. A client-side ML library can be a platform for interactive explanations, rapid prototyping and visualisation, and even for offline computation," Google's Big Picture team, comprising software engineers Nikhil Thorat and Daniel Smilkov, wrote in a blog post.

Google claims that the library gets past the speed limits of JavaScript. The structure of Deeplearn.js is similar to the TensorFlow library and NumPy. Both these Python-based scientific computing packages are widely used in various machine learning applications. Deeplearn.js comes with options for exporting weights from TensorFlow checkpoints. Authors can even import TensorFlow components into the Deeplearn.js interface. Additionally, developers have the option to use the library with JavaScript.

You can find the initial list of Deeplearn.js demo projects on its official website. The Deeplearn.js code is available for access in a GitHub repository.

Microsoft brings Linux to Windows Server

Microsoft has released its second Insider preview build for Windows Server 2016. The new version, debuted as Windows Server Insider Build 16257, enables Windows Subsystem for Linux (WSL) to offer distributions such as Ubuntu and OpenSUSE to the proprietary server platform.

Mozilla launches a ₹ 10 million fund to support open source projects in India

Mozilla has announced 'Global Mission Partners: India', an award programme that focuses on support for open source. The initiative is a part of the company's existing 'Mission Partners' program, and is aimed at supporting open source and free software projects in India with a total funding of ₹ 10 million.

The program is accepting applications from all over India. Also, Mozilla has agreed to support every project that furthers the company's mission. It has identified plenty of software projects in the country that need active backing. "Our mission, as embodied in our Manifesto, is to ensure the Internet is a global public resource, open and accessible to all; an Internet that truly puts people first, where individuals can shape their own experience and are empowered, safe and independent," the company wrote in a blog post.

The minimum incentive for each successful applicant in the 'Global Mission Partners: India' initiative is ₹ 125,000. However, applicants can win support of up to ₹ 5 million. The last date for applications (in English or Hindi) from the first batch of applicants for the award programme was September 30.

Participating projects need to have an OSI open source licence or an FSF free software licence. Also, the applicants must be based in India. You can read all the explicit conditions at Mozilla's wiki page.

The WSL is a compatibility layer to natively run Linux binary executives (in EFL format) natively on Windows. Microsoft originally introduced its WSL functionality with the Windows 10 Anniversary Update back in August 2016. And now, it is bringing the same experience to Windows Server. The new move also allows you to run open source developments such as Node.js, Ruby, Python, Perl and Bash scripts.

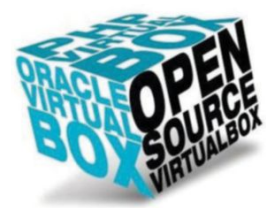
However, Microsoft has not provided native support for persistent Linux services like daemons and jobs as background tasks. You also need to enable the WSL and install a Linux distribution to begin with the advanced operations on Windows Server.

The new Windows Server test build comes with Remote Server Administration Tools (RSAT) packages. Users can install Windows 10 builds greater than 16250, to manage and administer insider builds using GUI tools with the help of RSAT packages.

You can additionally find new container images, optimised Nano Server base images, latest previews of .NET Core 2.0 and PowerShell 6.0, and a tweaked Server Core version. Also, the new release comes with various networking enhancements for Kubernetes integration and pipe mapping support. You need to register for the Windows Insiders for Business Program or Windows Insider Program to get your hands on the latest build of Windows Server. It includes various bug fixes and performance enhancements over the first preview build that was released earlier.

Oracle releases first beta of VirtualBox 5.2

Oracle has announced the first beta release of its upcoming VirtualBox 5.2. The new build comes with a feature to help users export VMs to the Oracle Public Cloud.



The new release of VirtualBox 5.2 eliminates all the hassle of exporting VMs to external drives and again importing to another VirtualBox installation. The company has also improved the handling of Virtual Machine Tools and Global Tools.

The first beta gives a glimpse of all the features that you will get to see in the stable release, and has a number of noteworthy improvements. The accessibility support in the GUI and EFI support have been enhanced in the new build. On the audio front, Oracle has added asynchronous data processing for HDA audio emulation. The audio support has also received host device callbacks, which will kick in while adding or removing an audio device.

In addition to the features limited to the beta version, Oracle is set to provide automatic, unattended guest OS installation in the next VirtualBox release. The fresh feature will be similar to the 'Easy Install' feature that was debuted on the commercial VMware Workstation 6.5 and 7 virtualisation software. The stable build will also improve the VM selector GUI. Similarly, users are expecting the upcoming releases to completely revamp the GUI on all supported platforms.

Ahead of the final release, you can download the VirtualBox 5.2 Beta 1 from the Oracle website to get a glimpse of the new additions. Users should note that this is a pre-release version and all its features may not be stable on supported systems.

For more news, visit *www.opensourceforu.com*

CODE SPORT

Sandya Mannarswamy

In this month's column, we discuss some of the basic questions in machine learning and text mining.

As we have been doing over the last couple of months, we will continue to discuss computer science interview questions, focusing on topics in machine learning and text analytics. While it is not necessary that one should know the mathematical details of the state-of-art algorithms for different NLP techniques, it is assumed that readers are familiar with the basic concepts and ideas in text analytics and NLP. For example, no one ever needs to implement back propagation code for a deep layered neural network, since this is provided as a utility function by the neural network libraries for different cost functions. Yet, one should be able to explain the concepts and derive the basic back propagation equations on a simple neural network for different loss functions, such as cross-entropy loss function or root mean square loss function, etc.

It is also important to note that many of the questions are typically oriented towards practical implementation or deployment issues, rather than just concepts or theory. So it is important for the interview candidates to make sure that they get adequate implementation experience with machine learning/NLP projects before their interviews. For instance, while most textbooks teach the basics of neural networks using a 'sigmoid' or 'hyperbolic tangent' (tanh) function as the activation function, hardly anyone uses the 'sigmoid' or 'tanh' functions in real-life implementations. In practice, the most commonly used activation function is the RELU (rectified linear) function in inner layers and, typically, softmax classifier is used in the final output layer.

Very often, interviewers weed out folks who are not hands-on, by asking them about the activation functions they would choose and the reason for their choices. (Sigmoid and hyperbolic tangent functions take a long time to learn and hence are not preferred in practice since they slow down the training considerably.)

Another popular question among interviewers is about mini-batch sizes in neural network training. Typically, training sets are broken into mini-batches and then cost function gradients are computed on each mini-batch, before the neural network weight parameters are updated using the computed gradients. The question often posed is: Why do we need to break down the training set into mini-batches instead of computing the gradient over the entire training set? Computing the gradients over the entire training set before doing the update will be extremely slow, as you need to go over thousands of samples before doing even a single update to the network parameters, and hence the learning process is very slow. On the other hand, stochastic gradient descent employs a mini-batch size of one (the gradients are updated after processing each single training sample); so the learning process is extremely rapid in this case.

Now comes the tricky part. If stochastic gradient descent is so fast, why do we employ mini-batch sizes that are greater than one? Typical mini-batch sizes can be 32, 64 or 128. This question will stump most interviewees unless they have hands-on implementation experience. The reason is that most neural networks run on GPUs or CPUs with multiple cores. These machines can do multiple operations in parallel. Hence, computing gradients for one training sample at a time leads to non-optimal use of the available computing resources. Therefore, mini-batch sizes are typically chosen based on the available parallelism of the computing GPU/CPU servers.

Another practical implementation question that gets asked is related to applying dropout techniques. While most of you would be familiar with the theoretical concept of drop-out, here is a trick question which interviewers frequently ask. Let us assume that you have employed a uniform dropout rate of 0.7 for each inner layer during training on a 4-layer feed forward neural network. After training the network,

you are given a held-out test set (which has not been seen before by the network), on which you have to report the predicted output. What is the drop-out rate that you would employ on the inner layers for the test set predictions? The answer, of course, is that one does not employ any drop-out on the test set.

Many of the interviewees fumble at this question. The key point to remember is that drop-out is employed basically to enable the network to generalise better by preventing the over-dependence on any particular set of units being active, during training. During test set prediction, we do not want to miss out on any of the features getting dropped out (which would happen if we use drop-out and prevent the corresponding neural network units from activating on the test data signal), and hence we do not use drop-out. An additional question that typically gets asked is: What is the inverted drop-out technique? I will leave it for our readers to find out the answer to that question.

Another question that frequently gets asked is on splitting the data set into train, validation and test sets. Most of you would be familiar with the nomenclature of train, validation and test data sets. So I am not going to explain that here. In classical machine learning, where we use classifiers such as SVMs, decision trees or random forests, when we split the available data set into train, validation and test, we typically use a split, 60-70 per cent training, 10-20 per cent validation and 10 per cent test data. While these percentages can vary by a few percentage points, the idea is to have to validate and test data sizes that are 10-20 per cent of the overall data set size. In classical machine learning, the data set sizes are typically of the order of thousands, and hence these sizes make sense.

Now consider a deep learning problem for which we have huge data sets of hundreds of thousands. What should be the approximate split of such data sets for training, validation and testing? In the Big Data sets used in supervised deep learning networks, the validation and test data sets are set to be in the order of 1-4 per cent of the total data set size, typically (not in tens of percentage as in the classical machine learning world).

Another question could be to justify why such a split makes sense in the deep learning world, and this typically leads to a discussion on hyper-parameter learning for neural networks. Given that there are quite a few hyper-parameters in training deep neural networks, another typical question would be the order in which you would tune for the different hyper-parameters. For example, let us consider three different hyper-parameters such as the mini-batch size, choice of activation function and learning rate. Since these three hyper-parameters are quite inter-related, how would you go about tuning them during training?

We have discussed quite a few machine learning questions till now; so let us turn to text analytics.

Given a simple sentence 'S' such as, "The dog chased the young girl in the park," what are the different types of text analyses that can be applied on this sentence in an increasing order of complexity? The first and foremost thing to do is basic lexical analysis of the sentence, whereby you identify the lexemes (the basic lexical analysis unit) and their associated part of the speech tags. For instance, you would tag 'dog' as a noun, 'park' as a noun, and 'chase' as a verb. Then you can do syntactic analysis, by which you combine words into associated phrases and create a parse tree for the sentence. For instance, 'the dog' becomes a noun phrase where 'the' is a determiner and 'dog' is a noun. Both lexical and syntactic analysis is done at the linguistic level, without the requirement for any knowledge of the external world.

Next, to understand the meaning of the sentence (semantic analysis), we need to identify the entities and relations in the text. In this simple sentence, we have three entities, namely 'dog', 'girl', and 'park'. After identifying the entities, we also identify the classes to which they belong. For example, 'girl' belongs to the 'Person' class, 'dog' belongs to the 'Animal' class and 'park' belongs to the 'Location' class. The relation 'chase' exists between the entities 'dog' and 'girl'. Knowing the entity classes allows us to postulate the relationship between the classes of the entities. In this case, it is possible for us to infer that the 'Animal' class entity can 'chase' the 'Person' class entity. However, semantic analysis involving determining entities and the relations between them, as well as inferring new relations, is very complex and requires deep NLP. This is in contrast to lexical and syntactic analysis which can be done with shallow NLP.

Deep NLP requires common sense and a knowledge of the world as well. The major open challenge in text processing with deep NLP is how best we can represent world knowledge, so that the context can be appropriately inferred. Let us consider the sentence, "India lost for the first time in a cricket test match to Bangladesh." Apart from the literal meaning of the sentence, it can be inferred that India has played with Bangladesh before, that India has beaten Bangladesh in previous matches, etc. While such inferences are very easy for humans due to our contextual or world knowledge, machines cannot draw these inferences easily as they lack contextual knowledge. Hence, any efficient NLP system requires representation of world knowledge. We will discuss this topic in greater detail in next month's column.

If you have any favourite programming questions/ software topics that you would like to discuss on this forum, please send them to me, along with your solutions and feedback, at *sandyasm_AT_yahoo_DOT_com*. Till we meet again next month, wishing all our readers a wonderful and productive year! END

By: Sandya Mannarswamy

The author is an expert in systems software and is currently working as a research scientist at Conduent Labs India (formerly Xerox India Research Centre). Her interests include compilers, programming languages, file systems and natural language processing. If you are preparing for systems software interviews, you may find it useful to visit Sandya's LinkedIn group 'Computer Science Interview Training India' at *http://www.linkedin.com/groups?home=&gid=2339182*

EXPERTISE IN OPEN SOURCE TELEPHONY SOLUTIONS

*astTECS
THE TRUE IP PBX

- CRM SOLUTION
- IP PBX SYSTEM
- CALL CENTER DIALER
- VoIP SOLUTIONS
- VIDEO CONFERENCE SOLUTION
- IVR SOLUTIONS
- VOICE LOGGER

MEET US TO BUILD YOUR CUSTOMIZED SOLUTIONS

14th Edition

OPEN SOURCE INDIA

13-14 October 2017

NIMHANS Convention Center | BENGALURU

BOOTH NO.A-1A

ISO 9001:2008

30+ COUNTRIES PRESENCE

3000+ DEPLOYMENTS & STILL COUNTING

FC COMPLIANT HARDWARES

RED HERRING GLOBAL 100 WINNER

080-6640 6640 | sales@asttecs.com | www.asttecs.com

A Quick Start to WebAssembly

Wikipedia defines WebAssembly as a portable stack machine designed to be faster to parse than JavaScript, as well as faster to execute. In this article, the author explores WebAssembly, covering its installation and its relationship with Rust.

Web applications these days are amazing. The first time I became aware of the potential power of Web applications was when I first encountered Gmail. Although Ajax calls had been in use, this was the first application that I knew of that had used them very effectively.

Still, more complex interactions and using local resources needed a plugin. Flash player is both the most used plugin as well as the best example of the problems with plugins. Security issues with Flash never seem to end.

Google tried to overcome some of the issues with the NPAPI plugins with the introduction of NaCl, the native clients. The NaCl clients run in a sandbox, minimising security risks. Google introduced PNaCl, or Portable NaCl, which is an architecture-independent version of NaCl.

Mozilla did not follow Google's lead with a native client, but instead decided to take a different approach, dropping NPAPI from current versions of Firefox.

The solution proposed by Mozilla was *asm.js*, a subset of the JavaScript language, which could run an ahead-of-its-time compiling and optimising engine. A related concept was that you could program in C/C++ and compile the code to *asm.js* using a tool like Emscripten. The advantage is that any application written for *asm.js* would run in any browser supporting JavaScript. However, it would run significantly faster if the browser used optimisation for *asm.js*.

The next step has been the introduction of a byte-code standard for Web browsers, called WebAssembly. The initial implementation targets the *asm.js* feature set and is being developed by all major browsers, including Mozilla, Google, Microsoft and Apple.

As in the case of *asm.js*, you may write the application in C/C++ and use a compiler like Emscripten to create a WebAssembly module.

Installation

The development tools for WebAssembly and Emscripten are not yet available in the official repositories. You can follow the instructions from *http://webassembly.org/getting-started/developers-guide/* for the installation.

For Linux, you need to build Empscripten from the source. It takes a substantial amount of time for downloading and building it, though.

You can test your installation by trying a simple C program, *hello.c*, as follows:

```
#include <stdio.h>
int main(int argc, char **argv){
    printf("hello, world\n");
}
```

Compile the program as follows to get a JavaScript output:

```
$ emcc hello.c -o hello.js
```

Now, test it and get the expected result.

```
$ node hello.js
hello, world
```

You can check the size of the *hello.js* file and it will be about 320K!

Now, compile the program as follows to get the WebAssembly result:

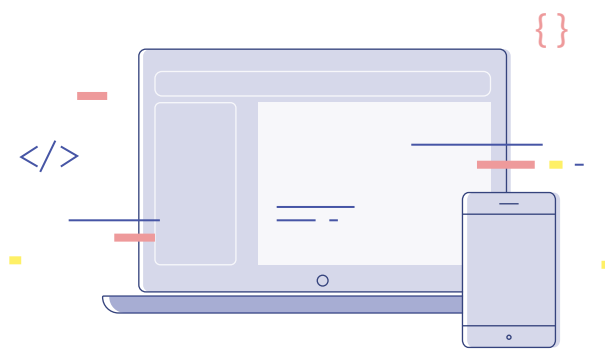
```
$ emcc hello.c -s WASM=1 -o hello.html
$ ls -lh
-rw-rw-r-- 1 anil anil   85 Aug 24 21:49 hello.c
-rw-rw-r-- 1 anil anil 101K Aug 24 21:50 hello.html
-rw-rw-r-- 1 anil anil 100K Aug 24 21:50 hello.js
-rw-rw-r-- 1 anil anil  45K Aug 24 21:50 hello.wasm
```

You will notice that it creates an *HTML* file, a *js* file and a *wasm* file. The overall size is smaller. You need the *HTML* file, as the *js* file will not execute with the *node*

ZOHO | Creator Developer Program

PARTNER WITH US

Your expertise, our platform. A profitable match.



Zoho Creator Solution Provider

- Industry-focussed projects.
- Access to cutting-edge technology.
- 15% payout on recurring revenue for lifetime.

Zoho Creator Certified Developer

- Work at your convenience.
- Explore a global client base.
- Free lifetime access to Zoho Creator developer account.

zoho.com/creator | partnership@zohocreator.com

command. For testing, run the following code:

```
$ emrun –no_browser –port 8080 .

Web server root directory: <your current directory>
Now listening at http://localhost:8080/
```

Open the browser *http://localhost:8080/hello.html* and you should see 'hello world' printed.

WebAssembly and Rust

Rust is a programming language sponsored by Mozilla Research. It's used for creating highly concurrent and safe systems. The syntax is similar to that of C/C++.

Since Firefox uses Rust, it seemed natural that it should be possible to program in Rust and compile to WebAssembly. You may follow the steps given at *https://goo.gl/LPIL8B* to install and test compiling Rust code to WebAssembly.

Rust is available in many repositories; however, you will need to use the *rustup* installer from *https://www.rustup.rs/* to install the compiler in your local environment and then add the modules needed for WebAssembly as follows:

```
$ curl https://sh.rustup.rs -sSf | sh
$ source ~/.cargo/env
$ rustup target add asmjs-unknown-emscripten
$ rustup target add wasm32-unknown-emscripten
```

You may now write your first Rust program, *hello.rs*, as follows:

```
fn main(){
    println!("Hello, Emscripten!");
}
```

Compile and run the program and verify that you get the expected output:

```
$ rustc hello.rs
$ hello
Hello, Emscripten!
```

Now, you may compile to JavaScript and test it as follows:

```
$ rustc --target=asmjs-unknown-emscripten hello.rs
$ node hello.js
Hello, Emscripten!
```

You can create the *wasm* target and an *HTML* front:

```
$ rustc --target=wasm32-unknown-emscripten hello.rs -o
hello.html
```

You can test it as with the C example, as follows:

```
$ emrun –no_browser –port 8080 .
Web server root directory: <your current directory>
Now listening at http://localhost:8080/
```

Why bother?

The importance of these projects cannot be underestimated. JavaScript has become very popular and, with Node.js, on the server side as well.

There is still a need to be able to write secure and reliable Web applications even though the growth of mobile apps has been explosive.

It would appear that mobile devices and 'apps' are taking over; however, there are very few instances in which the utility of an app is justified. In most cases, there is no reason that the same result cannot be achieved using the browser. For example, I do not find a need for the Facebook app. Browsing *m.facebook.com* is a perfectly fine experience.

When an e-commerce site offers me a better price if I use its app on a mobile device, it makes me very suspicious. The suspicions seem all too often to be justified by the permissions sought by many of the apps at the time of installation. Since it is hard to know which app publisher to trust, I prefer finding privacy-friendly apps, e.g., *https://goo.gl/aUmns3*.

Coding complex applications in JavaScript is hard. FirefoxOS may not have succeeded, but given the support by all the major browser developers, the future of WebAssembly should be bright. You can be sure that tools like Emscripten will emerge for even more languages, and you can expect apps to lose their importance in favour of the far safer and more trustworthy WebAssembly code. END

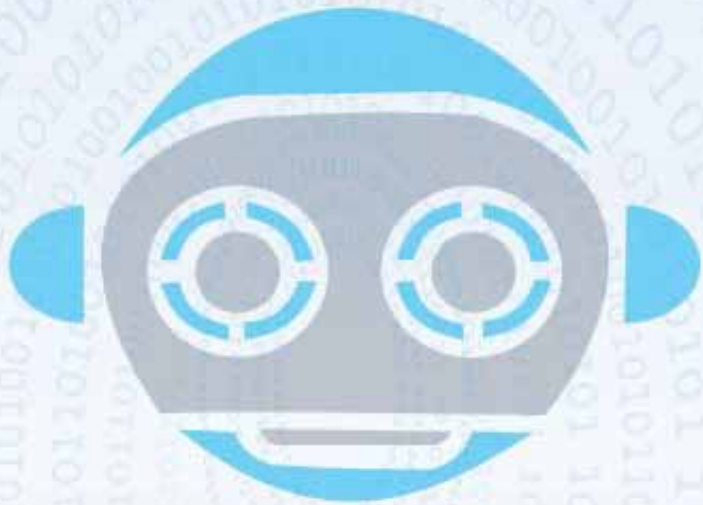
By: Dr Anil Seth

The author has earned the right to do what interests him. You can find him online at *http://sethanil.com, http://sethanil.blogspot.com*, and reach him via email at *anil@sethanil.com*.

odoo

14th Edition
OPEN SOURCE INDIA
NIMHANS Convention Center
13-14 October 2017
BENGALURU

The Bot Friend ...



Odoobot

Meet you soon !!!

Axcensa Technologies
Delivering Excellence through Quality

contactus@axcensa.com

www.axcensa.com

NEW PRODUCTS

Affordable Bluetooth headphones from Pebble

A pioneer in power banks and branded mobile accessories, Pebble has recently introduced its Bluetooth wireless headphones, the Pebble Sport. Designed exclusively for sports enthusiasts, the headphones offer comfort and performance during training and outdoor activities.

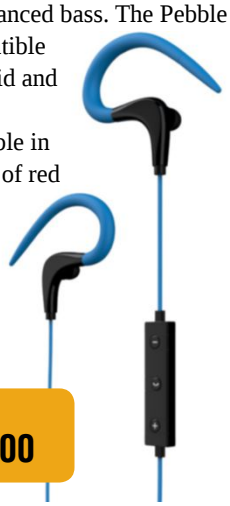
The Pebble Sport comes with premium quality sound drivers and is easy to wear. Bluetooth 4.0 provides good signal strength, and ensures high fidelity stereo music and clear tones.

The excellent grip of the device doesn't hamper rigorous movement during sports activities. It is a minimalist, lightweight design with unique ear hooks enabling all-day comfort.

On the Pebble Sport, users can listen to music for three to five hours at a stretch. It has a power capability of 55mAh and up to 10m of Bluetooth range.

The 20Hz-22kHz frequency response range offers crystal clear sound and enhanced bass. The Pebble Sport is compatible with all Android and iOS devices.

It is available in vibrant shades of red and blue via online retail stores.



Price: ₹ 1,200

Address: Pebble India, SRK Powertech Private Limited, G-135, Second Floor, Sector 63, Noida, UP – 201307, India

Powerful and compact Bluetooth speaker from Portronics

Price: ₹ 1,999

Portronics, a provider of innovative, digital and portable solutions, has recently launched an affordable, yet powerful Bluetooth speaker, the Sound Bun. The device has compact dimensions of 102mm x 102mm x 40mm and weighs less than 128 grams, which makes it easy to carry it in a pocket, pouch or laptop bag.

The high quality plastic it is made with gives the Sound Bun a premium look; and it comes with Bluetooth 4.1 and a 6W speaker, resulting in great sound. The device is backed with 5V DC – 1A power input and a frequency response range that's between 90kHz and 20kHz. The S/N ratio of 80Db allows the device to deliver good quality audio output.

The Sound Bun offers four hours of playback time with its 600mAh battery capacity. It is compatible with nearly all devices via Bluetooth, auxiliary cable or microSD card.

Any smartphone, laptop, tablet, phablet or smart TV can be connected to it via Bluetooth. The speaker is available in classic beige and black, via online and retail stores.

Address: Portronics Digital Private Limited, 4E/14, Azad Bhavan, Jhandewalan, New Delhi – 110055;
Ph: 09555245245

Wi-Fi system with built-in antivirus protection from TP-Link

Price: ₹ 14,090 approximately

Global consumer and business networking products firm, TP-Link, has unveiled its Deco M5 home Wi-Fi system with inbuilt protection provided by Trend Micro. The device is a mesh networking solution, which provides seamless wireless Internet coverage and security via TP-Link HomeCare.

The Wi-Fi system is powered by a quad-core processor and comes with a dual-band AC 1300 system capable of throughput speeds of 400Mbps on the 2.4GHz band and 867Mbps on the 5GHz band. It also supports MU-MIMO (Multiple-Input, Multiple-Output) data streaming, which divides bandwidth among your devices evenly. The system comes with three units that can be customised to provide continuous Wi-Fi coverage up to 418sqm (4,500 square feet). It also comes with strong parental controls from TP-Link HomeCare. The ART (Adaptive Routing Technology) enables the network to run as fast as possible by selecting the best path for device connections.

The TP-Link Deco M5 home Wi-Fi system is available online and at retail stores.

Address: TP-Link Technologies Co. Ltd, D-22/1, Okhla Phase 2, Near Maruti Suzuki Service Centre, New Delhi – 110020;
Ph: 9768012285

SSD with V-NAND technology from Samsung

Samsung has launched a portable SSD T5 with its latest 64-layer V-NAND technology, which enables it to deliver what the company claims are industry-leading transfer speeds of up to 540Mbps with encrypted data security. The company also claims that the pocket-sized SSD offers 4.9 times faster speeds than external HDD products.

Designed with solid metal, the lightweight SSD enables easy access to data, making it useful for content creators, as well as business and IT professionals. The solid state drive is smaller than an average business card (74mm x 57.3mm x 10.5mm) and weighs as little as 51 grams.

The T5 SSD can withstand accidental drops of up to two metres (6.6 feet) as it has no moving parts and has a shock-resistant internal frame. The device also features optional 256-bit AES encryption and comes with a USB Type-C port, along with a USB Type-C to Type-A cable.

The Samsung portable SSD T5 is available in deep black (1TB and 2TB models) and 'Alluring Blue' (250GB and 500GB) models via major retail stores.

Address: Samsung India, 20th to 24th Floors, Two Horizon Centre, Golf Course Road, Sector-43, DLF Phase 4, Gurugram, Haryana – 122202; **Ph:** 180030008282

A feature loaded tablet from Micromax

Indian mobile manufacturer, Micromax, has recently launched a tablet in the Indian market, called the Canvas Plex Tab. The device has a 20.32cm (8 inch) HD display with a resolution of 1024 x 600 pixels, and DTS sound for an immersive video and gaming experience.

Powered by a 1.3GHz quad-core MediaTek MT8382W/M processor, the device runs Android 5.1. It packs 32GB of internal storage and is backed with a 3000mAh non-removable battery. The device comes with a 5 megapixel primary camera on the rear and a 2 megapixel front shooter for selfies.

The tablet is a single SIM (GSM) device with a microSIM port. The connectivity options of the device include Wi-Fi, GPS, Bluetooth, USB, OTG, 3G and 4G along with a proximity sensor and accelerometer.

The Micromax Canvas Plex Tab comes bundled with one-year unlimited access to a content library on Eros Now, and is available at retail stores.

Address: Motorola Solutions India, 415/2, Mehrauli-Gurugram Road, Sector 14, Near Maharana Pratap Chowk, Gurugram, Haryana – 122001;
Ph: 0124-4192000;
Website: *www.motorola.in*

The prices, features and specifications are based on information provided to us, or as available on various websites and portals. OSFY cannot vouch for their accuracy.

Compiled by: Aashima Sharma

"THERE ARE VERY FEW ROADBLOCKS FOR DEVELOPERS WHO USE CLOUD FOUNDRY"

In the list of available options to ease cloud developments for developers and DevOps, Cloud Foundry comes out on top. The platform helps organisations advance their presence without transforming their existing infrastructure. But what has influenced the community to form a non-profit organisational model called the Cloud Foundry Foundation, which includes members like Cisco, Dell EMC, IBM, Google and Microsoft, among various other IT giants? ***Jagmeet Singh*** of ***OSFY*** speaks with ***Chip Childers, co-founder, Cloud Foundry Foundation,*** to find an answer to this question. Childers is also the chief technology officer of the Cloud Foundry platform and is an active member of the Apache Software Foundation. Edited excerpts...

Chip Childers,
co-founder,
Cloud Foundry Foundation

Q What is the ultimate aim of the Cloud Foundry Foundation?

The Cloud Foundry Foundation exists to steward the massive open source development efforts that have built up Cloud Foundry open source software, as well as to enable its adoption globally. We don't do this for the sake of the software itself, but with the goal of helping organisations around the world become much more effective and strategic in their use of technology. The Cloud Foundry platform is the foundational technology upon which over half of the Fortune 500 firms are digitally transforming themselves.

Q How is the Cloud Foundry platform different from OpenStack?

Cloud Foundry and OpenStack solve completely different problems. OpenStack projects are primarily about infrastructure automation, while Cloud Foundry is an application platform that can deploy itself onto any infrastructure, including OpenStack itself. Other infrastructure options on top of which one can run Cloud Foundry include Amazon Web Services, IBM Cloud, Google Cloud Platform, Microsoft Azure, RackHD, VMware vSphere, VMware Photon Platform and other options supported by the community.

Cloud Foundry does not just assume that its underlying infrastructure can be provisioned and managed by an API. It actually relies on that fact, so that the Cloud Foundry development community can focus on what application developers need out of an application-centric, multi-cloud platform.

Q In what way does Cloud Foundry ease working with cloud applications for DevOps?

The Cloud Foundry architecture is actually two different 'platforms'. At the lowest level is Cloud Foundry BOSH, which is responsible for infrastructure abstraction/automation, distributed system release management and platform health management. Above that is the Cloud Foundry Runtime, which is focused on serving the application developers' needs. The two layers work together to provide a highly automated operational experience, very frequently achieving operator-to-application ratios of 1:1000.

Q How does the container-based platform make application development easy for developers?

The design and evolution of the Cloud Foundry Runtime platform is highly focused on the DX (developer experience). While the Cloud Foundry Runtime does make use of containers within the architecture (in fact, Cloud Foundry's use of container technology predates Docker by years), these are not the focus of a developer's experience with the platform. What makes the Cloud Foundry Runtime so powerful for a developer is its ease of use. Simply 'cf push' your code into the system and let it handle the details of creating, managing and maintaining containers. Similarly, the access to various backing services — like the database, message queues, cache clusters and legacy system APIs — is designed to be exceptionally easy for developers. Overall, Cloud Foundry makes application development easier by eliminating a massive amount of the friction that is typically generated when shipping the code to production.

Q What are the major roadblocks currently faced when developing container-based applications using Cloud Foundry?

There are very few roadblocks for developers who use Cloud Foundry, but there are certainly areas where developers need to adjust older ways of thinking about how to best design the architecture of an application. The best architecture for an application being deployed to Cloud Foundry can be described as 'microservices', including choices like each service being independently versioned and deployed. While the microservices architecture may be new for a developer, it is certainly not a roadblock. In fact, even without fully embracing the microservices architecture, a developer can get significant value from deploying to the Cloud Foundry Runtime.

> *"The Cloud Foundry Foundation exists to steward the massive open source development efforts that have built the Cloud Foundry as open source software as well as to enable its adoption globally."*

Q Microsoft recently joined the Cloud Foundry Foundation, while Google has been on board since a long time. By when can you expect Amazon to become a key member of the community?

We think that the community and Amazon can benefit greatly by the latter becoming a part of Cloud Foundry. That said, it is important to note that Amazon Web Services (AWS) is already very well integrated into the Cloud Foundry platform, and is frequently being used as the underlying Infrastructure-as-a-Service (IaaS) that Cloud Foundry is deployed on.

Q How do you view Microsoft's decision on joining the non-profit organisation?

Microsoft has long been a member of the Cloud Foundry community, so the decision to join the Cloud Foundry Foundation represents a formalisation of its corporate support for the project. We are very happy that the company has chosen to take this step, and we are already starting to see the impact of this move on the project through increased engagement.

Q Is there any specific plan to encourage IT decision makers at enterprises to deploy Microsoft's Azure?

The Cloud Foundry Foundation is a vendor-neutral industry association. Therefore, we do not recommend any specific vendor over another. Our goal is to help all vendors integrate well into the Cloud Foundry software, community and market for the purpose of ensuring that the users and customers have a wide range of options for any particular service they may need, including infrastructure, databases, professional services and training.

Q As VMware originally conceived the Cloud Foundry platform back in 2009, how actively does the company now participate in the community?

Cloud Foundry was initially created at VMware, but the platform was transferred to Pivotal Software when it was spun out of VMware and EMC. When the Cloud Foundry Foundation was formed to support the expansion of the ecosystem and contributing community, VMware was a founding Platinum member. VMware remains heavily engaged in the Cloud Foundry Foundation in many ways, from providing engineering talent within the projects to supporting many of our other initiatives. It is a key member of the community.

Q What are the key points an enterprise needs to consider before opting for a cloud solution?

There are two key areas for consideration, based on how I categorise the various services offered by each of the leading cloud vendors, including

AWS, Google Cloud and Microsoft. These are commodity infrastructure services and differentiating services.

The infrastructure services include virtual machines, storage volumes, network capabilities and even undifferentiated database services. These are the services that are relatively similar across cloud providers. Therefore, you should evaluate them on the basis of a straightforward price versus performance trade-off. Performance criteria are not limited to actual computational performance but also include geographic location (when it matters for latency or regulatory reasons), availability guarantees, billing granularity and other relevant attributes.

The harder decision is how, when and where to make use of the differentiating service capabilities. These are the services that are unique to each cloud provider, including differentiated machine learning, IoT (Internet of Things) device management, Big Data and other more specific functionality. Selecting to use these types of services can significantly speed up the development of your overall application architecture, but they come with the potential downside of forcing a long-term cloud provider selection based on capability.

Enterprise customers need to find the right balance between these considerations. But they first need to look at what their actual needs are—if you are deploying a modern application or container platform on top of the cloud providers' infrastructure services, you are likely to want to focus on the price versus performance balance as a primary decision point. Then you can cautiously decide to use the differentiating services.

Also, it is not necessarily a decision of which single cloud provider you will use. If your organisation has either a sufficiently advanced operations team or a sufficiently complex set of requirements, you can choose to use multiple providers for what they are best fit for.

Q Do you believe a containerised solution like Cloud Foundry is vital for enterprises moving towards digital transformation?

I believe that digital transformation is fundamentally about changing the nature of an organisation to more readily embrace the use of software (and technology, in general) as a strategic asset. It's a fundamental shift in thinking from IT as a 'cost centre' to IT as a business driver. What matters the most is how an organisation structures its efforts, and how it makes the shift to a product-centric mindset to manage its technology projects.

> *"As with all community driven open source projects, our plans emerge from the collective ideas and actions of our technical community."*

That said, Cloud Foundry is increasingly playing a major role as the platform on which organisations restructure their efforts. In many ways, it serves as a 'forcing-function' to help inspire the changes required in an organisation outside of the platform itself. When you take away the technical obstacles to delivering software quickly, it becomes very obvious where the more systemic issues are in your organisation. This is an opportunity.

Q What's your opinion on the concept of serverless computing in the enterprise space?

I prefer the term 'event driven' or 'functions-as-a-service' because the notion of 'serverless' is either completely false or not descriptive enough. There is always a server, or more likely many servers, as a part of a compute service. Capabilities like AWS Lamba are better described as 'event driven' platforms.

We are early in the evolution and adoption of this developer abstraction. All the large-scale cloud providers are offering event-driven services, like AWS's Lambda. Nevertheless, any new abstraction that is going to bloom in the market needs to have a very active period of discovery by early adopters to drive the creation of the programming frameworks and best practices that are necessary, before it can truly take off within the enterprise context. I believe we are in the early stages of that necessary 'Cambrian explosion'.

Q Is there any plan to expand Cloud Foundry into that model of 'event-driven' computing?

It is quite likely. As with all community driven open source projects, our plans emerge from the collective ideas and actions of our technical community. This makes it impossible to say with certainty what will emerge, outside of the documented and agreed upon roadmaps. However, there have been several proof-of-concepts that have demonstrated how the Cloud Foundry Runtime is well prepared to extend itself into that abstraction type.

Q Lastly, how do you foresee the mix of containers and the cloud?

Containers are, without a doubt, growing in usage. They are being used locally on developer machines, within corporate data centres and within public cloud providers.

The public cloud, on the other hand, is undergoing a massive wave of adoption at the moment. This is not just 'infrastructure services'. Public clouds are offering services that span infrastructure offerings, Platform-as-a-Service, Functions-as-a-Service and Software-as-a-Service. END

RailsFactory
Ruby on Rails Solutions

Hire our expert team for your Web and Mobile application development

- ✓ 500+ Applications
- ✓ 200+ Global Startups
- ✓ 120+ Passionate Engineers
- ✓ 10+ Years of Experience

BUSINESS VALUE THROUGH CUTTING EDGE TECHNOLOGIES

Industry Focus

- ✓ e-commerce
- ✓ Retail
- ✓ Manufacturing
- ✓ Tourism
- ✓ Licensing
- ✓ Insurance & Healthcare
- ✓ Education & Media

Our Clients include

GE · Mercedes-Benz · CARATLANE · Jifflenow · apigee

Chennai | Bangalore | California | Zurich

Call us @ +91 994 044 6586 or email us < info@railsfactory.com >
www.railsfactory.com

The Smart Cube Uses Open Source to Deliver Custom Research and Analytics Services

The Smart Cube offers a range of custom research and analytics services to its clients, and relies greatly on open source to do so. The UK-headquartered company has a global presence with major bases in India, Romania, and the US, and employs more than 650 analysts around the world.

"The whole analytics ecosystem is shifting towards open source," says Nitin Aggarwal, vice president of data analytics, The Smart Cube. Aggarwal is leading a team of over 150 developers in India, of which 100 are working specifically on open source deployments. The company has customers ranging from big corporations to financial services institutions and management consulting firms. And it primarily leverages open source when offering its services.

Aggarwal tells *Open Source For You* that open source has helped analytics developments to be more agile in a collaborative environment. "We work as a true extension of our clients' teams, and open source allows us to implement quite a high degree of collaboration. Open source solutions also make it easy to operationalise analytics, to meet the daily requirements of our clients," Aggarwal states.

Three major trends in the analytics market that are being powered by open source

- **Advanced analytics ecosystem:** Whether it is an in-house analytics team or a specialised external partner, open source is today the first choice to enable an advanced analytics solution.
- **Big Data analytics:** Data is the key fuel for a business, and harnessing it requires powerful platforms and solutions that are increasingly being driven by open source software.
- **Artificial intelligence:** Open source is a key enabler for R&D in artificial intelligence.

Apart from helping increase collaboration and deliver operationalised results, open source reduces the overall cost of analytics for The Smart Cube, and provides higher returns on investments for its clients. The company does have some proprietary solutions, but it uses an optimal mix of open and closed source software to cater to a wide variety of industries, business problems and technologies.

"Our clients often have an existing stack that they want us to use. But certain problems create large-scale complex analytical workloads that can only be managed using open source technologies. Similarly, a number of problems are best solved using algorithms that are better researched and developed in open source, while many descriptive or predictive problems are easily solved using proprietary solutions like Tableau, QlikView or SAS," says Aggarwal.

The Smart Cube team also monitors market trends and seeks customer inputs at various levels to evaluate new technologies and tools, adjusting the mix of open and closed source software as per requirements.

The challenges with analytics

Performing data analysis involves overcoming some hurdles. In addition to the intrinsic art of problem solving that analytics professionals need to have, there are some technical challenges that service providers need to resolve to examine data. Aggarwal says that standardising data from structured and unstructured information has become challenging. Likewise, obtaining a substantial amount of good training sets is also hard, and determining the right technology stack to balance cost and performance is equally difficult.

Community solutions to help extract data

Aggarwal divulges various community-backed solutions that jointly power the data extraction process and help to resolve the technical challenges involved in the data analysis process. To serve hundreds of clients in a short span of time, The Smart Cube has built a custom framework. This framework offers data collection and management solutions that use open source. There is Apache Nutch and Kylo to enable data lake management, and Apache Beam to design the whole data collection process.

The Smart Cube leverages open source offerings, including Apache Spark and Hadoop, to analyse the bulk of extracted structured and unstructured data. "We deal with data at the terabyte scale, and analysis of such massive data sets is beyond the capability of a single commodity hardware. Traditional RDBMS (relational database management systems) also cannot manage many types of unstructured data like images and videos. Thus, we leverage Apache Spark and Hadoop," Aggarwal says.

Nitin Aggarwal, vice president of data analytics, The Smart Cube

Predictive analytics using open source support

The Smart Cube is one of the leading service providers in the nascent field of predictive analytics. This type of analytics has become vital for companies operating in a tough competitive environment. Making predictions isn't easy. But open source helps on that front as well.

"A wide variety of predictive analytics problems can be solved using open source. We take support from open source solutions to work on areas like churn prediction, predictive maintenance, recommendation systems and video analytics," says Aggarwal. The company uses Scikit-learning with Python, Keras and Google's TensorFlow, to enable predictive analysis and deep learning solutions for major prediction issues.

Additionally, in September 2017, The Smart Cube launched 'Concept Lab' that allows the firm to experiment at a faster pace, and develop and test solution frameworks for client problems. "This approach, enabled by opting for open source, has gained us a lot of traction with our corporate clients, because we are able to provide the flexibility and agility that they cannot achieve internally," Aggarwal affirms.

The bright future of data analytics

Open source is projected to help data analytics companies in the future, too. "We expect open source to dominate the future of the analytics industry," says Aggarwal.

The Smart Cube is foreseeing good growth with open source deployments. Aggarwal states that open source will continue to become more mainstream for data analytics companies and will gradually replace proprietary solutions. "Most of the new R&D in analytics will continue to be on open source frameworks. The market for open source solutions will also consolidate over time as there is a huge base of small players at present, which sometimes confuses customers," Aggarwal states.

According to NASSCOM, India will become one of the top three markets in the data analytics space, in the next three years. The IT trade body also predicts that the Big Data analytics sector in the country will witness eight-fold growth by 2025, from the current US$ 2 billion to a whopping US$ 16 billion.

Companies like The Smart Cube are an important part of India's growth journey in the analytics market, and will influence more businesses to opt for open source in the future. END

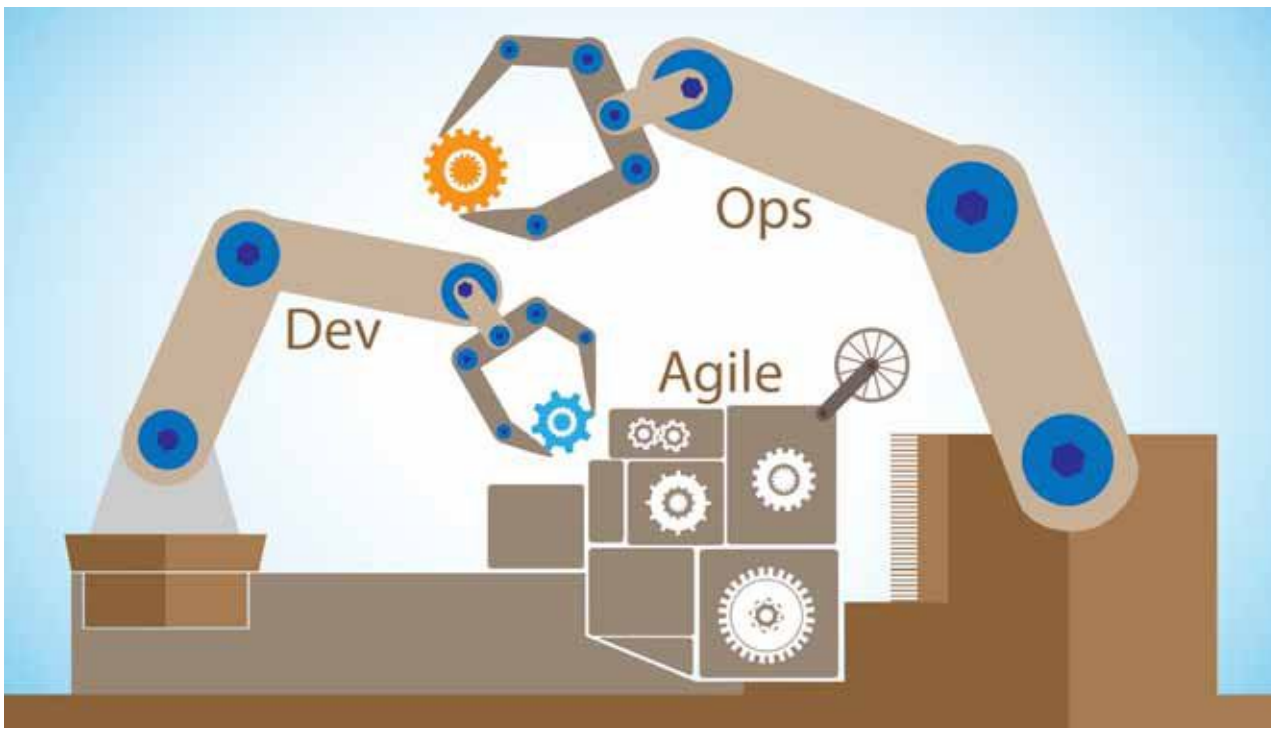
Significant open source solutions that drive The Smart Cube

- Internal set-up of Apache Hadoop and Spark infrastructure to help teams perform R&D activities prior to building Big Data solutions.
- Open source tools enable real-time social media analysis for key clients.
- A native text analytics engine uses open source solutions to power a variety of research projects.
- Concept Lab to rapidly experiment and test solution frameworks for clients.

By: Jagmeet Singh
The author is an assistant editor at *EFY*.

The DevOps Series
Using Docker with Ansible

This article is the eighth in the DevOps series. This month, we shall learn to set up Docker in the host system and use it with Ansible.



Docker provides operating system level virtualisation in the form of containers. These containers allow you to run standalone applications in an isolated environment. The three important features of Docker containers are isolation, portability and repeatability. All along we have used Parabola GNU/Linux-libre as the host system, and executed Ansible scripts on target virtual machines (VM) such as CentOS and Ubuntu.

Docker containers are extremely lightweight and fast to launch. You can also specify the amount of resources that you need such as the CPU, memory and network. The Docker technology was launched in 2013, and released under the Apache 2.0 licence. It is implemented using the Go programming language. A number of frameworks have been built on top of Docker for managing these clusters of servers. The Apache Mesos project, Google's Kubernetes, and the Docker Swarm project are popular examples. These are ideal for running stateless applications and help you to easily scale horizontally.

Setting it up

The Ansible version used on the host system (Parabola GNU/Linux-libre x86_64) is 2.3.0.0. Internet access should be available on the host system. The *ansible/* folder contains the following file:

```
ansible/playbooks/configuration/docker.yml
```

Installation

The following playbook is used to install Docker on the host system:

```
---
- name: Setup Docker
  hosts: localhost
  gather_facts: true
  become: true
  tags: [setup]

  tasks:
    - name: Update the software package repository
      pacman:
        update_cache: yes

    - name: Install dependencies
      package:
        name: "{{ item }}"
        state: latest
```

```
    with_items:
      - python2-docker
      - docker

  - service:
      name: docker
      state: started
  - name: Run the hello-world container
    docker_container:
      name: hello-world
      image: library/hello-world
```

The Parabola package repository is updated before proceeding to install the dependencies. The *python2-docker* package is required for use with Ansible. Hence, it is installed along with the *docker* package. The Docker daemon service is then started and the *library/hello-world* container is fetched and executed. A sample invocation and execution of the above playbook is shown below:

```
$ ansible-playbook playbooks/configuration/docker.yml -K
--tags=setup
SUDO password:

PLAY [Setup Docker] ************************************

TASK [Gathering Facts] *********************************
ok: [localhost]

TASK [Update the software package repository] **********
changed: [localhost]

TASK [Install dependencies] ****************************
ok: [localhost] => (item=python2-docker)
ok: [localhost] => (item=docker)

TASK [service] *****************************************
ok: [localhost]

TASK [Run the hello-world container] *******************
changed: [localhost]

PLAY RECAP *********************************************
localhost                  : ok=5    changed=2
unreachable=0    failed=0
```

With the verbose '-v' option to ansible-playbook, you will see an entry for LogPath, such as */var/lib/docker/containers/<container-id>/<container-id>-json.log*. In this log file, you will see the output of the execution of the *hello-world* container. This output is the same when you run the container manually as shown below:

```
$ sudo docker run hello-world

Hello from Docker!
```

This message shows that your installation appears to be working correctly.

To generate this message, Docker took the following steps:

1. The Docker client contacted the Docker daemon.
2. The Docker daemon pulled the *hello-world* image from the Docker Hub.
3. The Docker daemon created a new container from that image, which runs the executable that produces the output you are currently reading.
4. The Docker daemon streamed that output to the Docker client, which sent it to your terminal.

To try something more ambitious, you can run an Ubuntu container with:

```
$ docker run -it ubuntu bash
```

You can share images, automate workflows, and more with a free Docker ID at *https://cloud.docker.com/*.

For more examples and ideas, do visit *https://docs.docker.com/engine/userguide/*.

An example

A deep learning (DL) Docker project is available *(https://github.com/floydhub/dl-docker)* with support for frameworks, libraries and software tools. We can use Ansible to build the entire DL container from the source code of the tools. The base OS of the container is Ubuntu 14.04, and will include the following software packages:

- TensorFlow
- Caffe
- Theano
- Keras
- Lasagne
- Torch
- iPython/Jupyter Notebook
- Numpy
- SciPy
- Pandas
- Scikit Learn
- Matplotlib
- OpenCV

The playbook to build the DL Docker image is given below:

```
- name: Build the dl-docker image
  hosts: localhost
  gather_facts: true
  become: true
  tags: [deep-learning]

  vars:
    DL_BUILD_DIR: "/tmp/dl-docker"
    DL_DOCKER_NAME: "floydhub/dl-docker"
```

```
tasks:
  - name: Download dl-docker
    git:
      repo: https://github.com/saiprashanths/dl-docker.git
      dest: "{{ DL_BUILD_DIR }}"

  - name: Build image and with buildargs
    docker_image:
       path: "{{ DL_BUILD_DIR }}"
       name: "{{ DL_DOCKER_NAME }}"
       dockerfile: Dockerfile.cpu
       buildargs:
         tag: "{{ DL_DOCKER_NAME }}:cpu"
```

We first clone the deep learning Docker project sources. The *docker_image* module in Ansible helps us to build, load and pull images. We then use the *Dockerfile.cpu* file to build a Docker image targeting the CPU. If you have a GPU in your system, you can use the *Dockerfile.gpu* file. The above playbook can be invoked using the following command:

```
$ ansible-playbook playbooks/configuration/docker.yml -K
--tags=deep-learning
```

Depending on the CPU and RAM you have, it will take a considerable amount of time to build the image with all the software. So be patient!

Jupyter Notebook

The built *dl-docker* image contains Jupyter Notebook, which can be launched when you start the container. An Ansible playbook for the same is provided below:

```
- name: Start Jupyter notebook
  hosts: localhost
  gather_facts: true
  become: true
  tags: [notebook]

  vars:
    DL_DOCKER_NAME: "floydhub/dl-docker"

  tasks:
    - name: Run container for Jupyter notebook
      docker_container:
        name: "dl-docker-notebook"
        image: "{{ DL_DOCKER_NAME }}:cpu"
        state: started
        command: sh run_jupyter.sh
```

You can invoke the playbook using the following command:

```
$ ansible-playbook playbooks/configuration/docker.yml -K
--tags=notebook
```

The Dockerfile already exposes the port 8888, and hence you do not need to specify the same in the above *docker_container* configuration. After you run the playbook, using the 'docker ps' command on the host system, you can obtain the container ID as indicated below:

```
$ sudo docker ps
CONTAINER ID        IMAGE               COMMAND             CREATED             STATUS
PORTS               NAMES
a876ad5af751        floydhub/dl-docker:cpu   "sh run_jupyter.
sh"   11 minutes ago      Up 4 minutes        6006/tcp, 8888/
tcp   dl-docker-notebook
```

You can now log in to the running container using the following command:

```
$ sudo docker exec -it a876 /bin/bash
```

You can then run an 'ifconfig' command to find the local IP address ('172.17.0.2' in this case), and then open *http://172.17.0.2:8888* in a browser on your host system to see the Jupyter Notebook. A screenshot is shown in Figure 1.

Figure 1: Jupyter Notebook

Some of our recent courses:

On Udemy:

- GCP: Complete Google Data Engineer and Cloud Architect Guide
- Time Capsule: Trends in Tech, Product Strategy
- Show And Tell: Sikuli - Pattern-Matching and Automation

On Pluralsight:

- Understanding the Foundations of TensorFlow
- Working with Graph Algorithms in Python
- Building Regression Models in TensorFlow

Loonycorn

Our Content:

- The Careers in Computer Science Bundle
  9 courses | 139 hours
- The Complete Machine Learning Bundle
  10 courses | 63 hours
- The Complete Computer Science Bundle
  8 courses | 78 hours
- The Big Data Bundle
  9 courses | 64 hours
- The Complete Web Programming Bundle
  8 courses | 61 hours
- The Complete Finance & Economics Bundle
  9 courses | 56 hours
- The Scientific Essentials Bundle
  7 courses | 41 hours
- ~30 courses on Pluralsight
  ~80 on StackSocial
  ~75 on Udemy

About Us:

- ex-Google | Stanford | INSEAD
- 80,000+ students

Figure 2: TensorBoard

TensorBoard

TensorBoard consists of a suite of visualisation tools to understand the TensorFlow programs. It is installed and available inside the Docker container. After you log in to the Docker container, at the root prompt, you can start TensorBoard by passing it a log directory as shown below:

```
# tensorboard --logdir=./log
```

You can then open *http://172.17.0.2:6006/* in a browser on your host system to see the TensorBoard dashboard as shown in Figure 2.

Docker image facts

The *docker_image_facts* Ansible module provides useful information about a Docker image. We can use it to obtain the image facts for our *dl-docker* container as shown below:

```
- name: Get Docker image facts
  hosts: localhost
  gather_facts: true
  become: true
  tags: [facts]

  vars:
    DL_DOCKER_NAME: "floydhub/dl-docker"

  tasks:
    - name: Get image facts
      docker_image_facts:
        name: "{{ DL_DOCKER_NAME }}:cpu"
```

The above playbook can be invoked as follows:

```
$ ANSIBLE_STDOUT_CALLBACK=json ansible-playbook playbooks/
configuration/docker.yml -K --tags=facts
```

The ANSIBLE_STDOUT_CALLBACK environment variable is set to 'json' to produce a JSON output for readability. Some important image facts from the invocation of the above playbook are shown below:

```
"Architecture": "amd64",
"Author": "Sai Soundararaj <saip@outlook.com>",

"Config": {

"Cmd": [
  "/bin/bash"
],

"Env": [
  "PATH=/root/torch/install/bin:/root/caffe/build/tools:/
root/caffe/python:/usr/local/sbin:/usr/local/bin:/usr/sbin:/
usr/bin:/sbin:/bin",
  "CAFFE_ROOT=/root/caffe",
  "PYCAFFE_ROOT=/root/caffe/python",
  "PYTHONPATH=/root/caffe/python:",
  "LUA_PATH=/root/.luarocks/share/lua/5.1/?.lua;/root/.
luarocks/share/lua/5.1/?/init.lua;/root/torch/install/
share/lua/5.1/?.lua;/root/torch/install/share/lua/5.1/?/
init.lua;./?.lua;/root/torch/install/share/luajit-2.1.0-
beta1/?.lua;/usr/local/share/lua/5.1/?.lua;/usr/local/share/
lua/5.1/?/init.lua",
  "LUA_CPATH=/root/torch/install/lib/?.so;/root/.luarocks/
lib/lua/5.1/?.so;/root/torch/install/lib/lua/5.1/?.so;./?.
so;/usr/local/lib/lua/5.1/?.so;/usr/local/lib/lua/5.1/
loadall.so",
  "LD_LIBRARY_PATH=/root/torch/install/lib:",
  "DYLD_LIBRARY_PATH=/root/torch/install/lib:"
],

"ExposedPorts": {
  "6006/tcp": {},
  "8888/tcp": {}
},
"Created": "2016-06-13T18:13:17.247218209Z",
"DockerVersion": "1.11.1",
"Os": "linux",

"task": { "name": "Get image facts" }
```

You are encouraged to read the 'Getting Started with Docker' user guide available at *http://docs.ansible.com/ansible/latest/guide_docker.html* to know more about using Docker with Ansible. END

By: Shakthi Kannan

The author is a free software enthusiast and blogs at *shakthimaan.com*.

Surfing the Digital Wave with ADDC

The application driven data centre (ADDC) is a design whereby all the components of the data centre can communicate directly with an application layer. As a result, applications can directly control the data centre components for better performance and availability in a cost-optimised way. ADDC redefines the roles and skillsets that are needed to manage the IT infrastructure in this digital age.

Worldwide, the IT industry is undergoing a tectonic shift. To the Indian IT service providers, this shift offers both new opportunities and challenges. For long, the Indian IT industry has enjoyed the privilege of being a supplier of an English-speaking, intelligent workforce that meets the global demand for IT professionals. Till now, India could leverage the people cost arbitrage between the developed and developing countries. The basic premise was that IT management will always require skilled professional people. Therefore, the operating model of the Indian IT industry has so far been headcount based.

Today, that fundamental premise has given way to automation and artificial intelligence (AI). This has resulted in more demand for automation solutions and a reduction in headcount—challenging the traditional operating model. The new solutions in demand require different skillsets. The Indian IT workforce is now struggling to meet this new skillset criteria.

Earlier, the industry's dependence on people also meant time-consuming manual labour and delays caused by manual errors. The new solutions instead offer the benefits of automation, such as speeding up IT operations by replacing people. This is similar to the time when computers started replacing mathematicians.

But just as computers replaced mathematicians yet created new jobs in the IT sector, this new wave of automation is also creating jobs for a new generation with new skillsets. In today's world, infrastructure management and process management professionals are being replaced by developers writing code for automation.

These new coding languages manage infrastructure in a radically different way. Traditionally, infrastructure was managed by the operations teams and developers never got involved. But now, the new management principles talk about managing infrastructure through automation code. This changes the role of sysadmins and developers.

The developers need to understand infrastructure operations and use these languages to control the data centre. Therefore, they can now potentially start getting into the infrastructure management space. This is a threat to the existing infrastructure operations workforce, unless they themselves skill up as infrastructure developers.

So does it mean that by learning to code, one can secure jobs in this turbulent job market? The answer is both 'Yes' and 'No'. 'Yes', because in the coming days everyone needs to be a developer. And it's also a 'No' because in order to get into the infrastructure management space, one needs to master new infrastructure coding languages even if one is an expert developer in other languages.

New trends in IT infrastructure

The new age infrastructure is built to be managed by code. Developers can benefit from this new architecture by controlling infrastructure from the applications layer. In this

new model, an application can interact with the infrastructure and shape it the way required. It is not about designing the infrastructure with the application's requirement as the central theme (application-centric infrastructure); rather, it is about designing the infrastructure in a way that the application can drive it (application-driven infrastructure). We are not going to build infrastructure to host a group of applications but rather, we will create applications that can control various items of the infrastructure. Some of the prominent use cases involve applications being able to automatically recover from infrastructure failures. Also, scaling to achieve the best performance-to-cost ratio is achieved by embedding business logic in the application code that drives infrastructure consumption.

In today's competitive world, these benefits can provide a winning edge to a business against its competitors. While IT leaders such as Google, Amazon, Facebook and Apple are already operating in these ways, traditional enterprises are only starting to think and move into these areas. They are embarking on a journey to reach the ADDC nirvana state by taking small steps towards it. Each of these small steps is transforming the traditional enterprise data centres, block by block, to be more compatible for an application-driven data centre design.

The building blocks of ADDC

For applications to be able to control anything, they require the data centre components to be available with an application programming interface (API). So the first thing enterprises need to do with their infrastructure is to convert every component's control interface into an API. Also, sometimes, traditional programming languages do not have the right structural support for controlling these infrastructure components and, hence, some new programming languages need to be used that have infrastructure domain-specific structural support. These languages should be able to understand the infrastructure components such as the CPU, disk, memory, file, package, service, etc. If we are tasked with transforming a traditional data centre into an ADDC, we have to first understand the building blocks of the latter, which we have to achieve, one by one. Let's take a look at how each traditional management building block of an enterprise data centre will map into an ADDC set-up.

1. The Bare-metal-as-a-Service API

The bare metal physical hardware has traditionally been managed by the vendor-specific firmware interfaces. Nowadays, open standard firmware interfaces have emerged, which allow one to write code in any of the application coding languages to interact through the HTTP REST API. One example of an open standard Bare-metal-as-a-Service API is Redfish. Most of the popular hardware vendors are now allowing their firmware to be controlled through Redfish API implementation. The Redfish specifications-compatible hardware can be directly controlled through a general

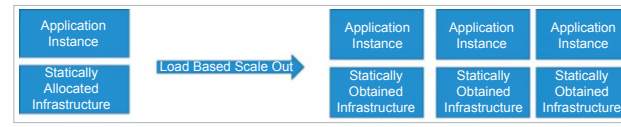


Figure 1: Application-centric infrastructure

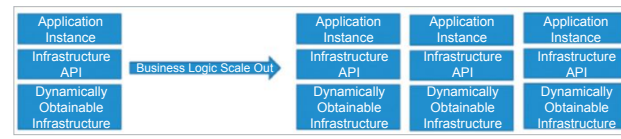


Figure 2: Application-driven infrastructure

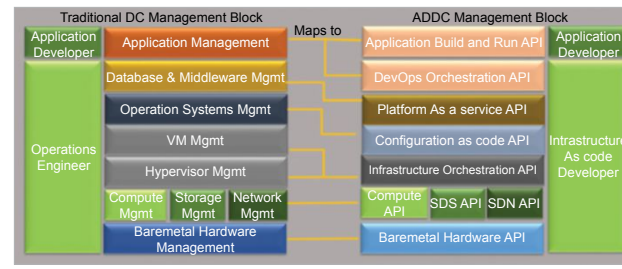


Figure 3: Traditional data centre mapped to ADDC

application over HTTP, and without necessarily going through any operating system interpreted layer.

2. The software defined networking API

A traditional network layer uses specialised appliances such as switches, firewalls and load balancers. Such appliances have built-in control and data planes. Now, the network layer is transforming into a software defined solution, which separates the control plane from the data plane.

In software defined solutions for networking, there are mainly two approaches. The first one is called a software defined network (SDN). Here, the central software control layer installed on a computer will control several of the network's physical hardware components to provide the specific network functionality such as routing, firewall and load balancers. The second one is the virtual network function (VNF). Here, the approach is to replace hardware components on a real network with software solutions on the virtual network. The process of creating virtual network functions is called network function virtualisation (NFV). The software control layers are exposed as APIs, which can be used by the software/application codes. This provides the ability to control networking components from the application layer.

3. The software defined storage API

Traditional storages such as SAN and NAS have now transformed into software defined storage solutions, which can offer both block and file system capabilities. These software defined storage solutions are purpose-built operating systems that can make a standard physical server exhibit the properties of a storage device. We can format a standard x86 server with these specialised operating systems, to create a storage solution

out of this general-purpose server. Depending on the software, the storage solution can exhibit the behaviour of SAN block storage, NAS file storage or even object storage. Ceph, for example, can create all three types of storage out of the same server. In these cases, the disk devices attached to the servers operate as the storage blocks. The disks can be standard direct attached storage (like the one in your laptop) or a number of disks daisy-chained to your server system.

The software defined solutions can be extended and controlled through the software libraries and APIs that they expose. Typically available on a REST API and with UNIX/Linux based operating systems, these are easy to integrate with other orchestration solutions. For example, OpenStack exposes Cinder for block storage, Manila for file storage and Swift for object storage. An application can either run management commands on the natively supported CLI shell or the native/orchestration APIs.

4. The Compute-as-a-Service API

Compute-as-a-Service is the ability to serve the bare metal, the virtual machine or the containers in an on-demand basis over API endpoints or through self-service portals. It is built mostly on top of virtualisation or containerisation platforms. A Compute-as-a-Service model may or may not be a cloud solution. Hypervisors that can be managed through a self-service portal and API endpoint can be considered as Compute-as-a-Service. For example, a VMware vSphere implementation with a self-service portal and API endpoint is such a solution. Similarly, on the containerisation front, the container orchestration tools like Kubernetes are not a cloud solution but a good example of Compute-as-a-Service with an API and self-service GUI. Typical cloud solutions that allow one to provision virtual machines (like AWS EC2), containers (like AWS ECS) and in some cases even physical machines (like Softlayer), are examples of compute power provided as a service.

5. The infrastructure orchestration API

Infrastructure orchestration is the Infrastructure-as-a-Service cloud solution that can offer infrastructure components on demand, as a service, over an API. In case of infrastructure orchestration, it is not only about VM provisioning. It is about orchestrating various infrastructure components in storage, networking and compute, in an optimised manner. This helps provisioning and de-provisioning of components as per the demands of business. The cloud solutions typically offer control over such orchestration through some programming language to configure orchestration logics. For example, AWS provides cloud formation and OpenStack provides the Heat language for this. However, nowadays, in a multi-cloud strategy, new languages have come up for hybrid cloud orchestration. Terraform and Cloudify are two prime examples.

6. Configuration management as code and API

In IT, change and configuration management are the traditional ITIL processes that track every change in the configuration of systems. Typically, the process is reactive, whereby change is performed on the systems and then recorded in a central configuration management database.

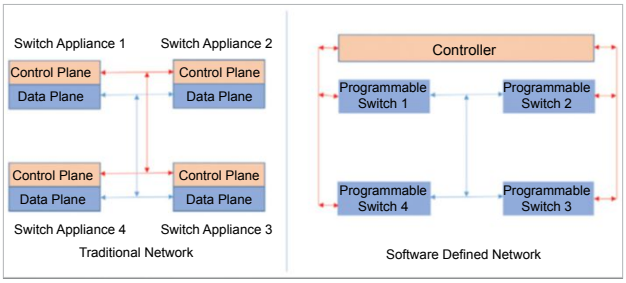


Figure 4: A traditional network vs a software defined network

However, currently, changes are first recorded in a database as per the need. Then these changes are applied to systems using automation tools to bring them to the desired state, as recorded in the database. This new-age model is known as the desired state of configuration management. cfEngine, Puppet, Chef, etc, are well known configuration management tools in the market.

These tools configure the target systems as per the desired configuration mentioned in the files. Since this is done by writing text files with a syntax and some logical constructs, these files are known to be infrastructure configuration codes. Using such code to manage infrastructure is known as 'configuration management as code' or 'infrastructure as code'. These tools typically expose an API endpoint to create the desired configuration on target servers.

7. The Platform-as-a-Service API

Platform-as-a-Service (PaaS) solutions provide the platform components such as application, middleware or database, on demand. These solutions hide the complexity of the infrastructure at the backend. At the frontend, they expose a simple GUI or API to provision, de-provision or scale platforms for the application to run.

So instead of saying, "I need a Linux server for installing MySQL," the developer will just have to say, "I need a MySQL instance." In a PaaS solution, deploying a database means it will deploy a new VM, install the required software, open up firewall ports and also provision the other dependencies needed to access the database. It does all of this at the backend, abstracting the complexities from the developers, who only need to ask for the database instance, to get the details. Hence developers can focus on building applications without worrying about the underlying complexities.

The APIs of a PaaS solution can be used by the application to scale itself. Most of the PaaS solutions are based on containers which can run on any VM, be it within the data centre or in the public cloud. So the PaaS solutions can stretch across private and public cloud environments.

Continued on page...40

Managing Log Files with the Logrotate Utility

Log files, though useful to troubleshoot and to track usage, tend to use up valuable disk space. Over time, they become large and unwieldy, so pinpointing an event becomes difficult. Logrotate performs the function of archiving a log file and starting a new one, thereby 'rotating' it.

Logrotate has been designed to ease the administration of systems that generate large numbers of log files in any format. It allows automatic rotation, compression, removal and mailing of log files. Each log file may be handled daily, every week, every month, or when it grows too large (rotation on the basis of a file's size).

The application and the servers generate too many logs, making the task of troubleshooting or gaining business insights from these logs, a difficult one. Many a time, there's the issue of servers running on low disk space because of the very large log files on them.

Servers with huge log files create problems when the resizing of virtual machines needs to be done. Troubleshooting based on large files may take up a lot of time and valuable memory. The logrotate utility is extremely useful to solve all such problems. It helps in taking backups of log files on an hourly, daily, weekly, monthly or yearly basis with additional choice of log backup with compression. Also, file backups can be taken by setting a limit on the file size, like 100MB, for instance. So, after the log file reaches a size of 100MB, the file will be rotated.

The synopsis is as follows:

```
logrotate [-dv] [-f|--force] [-s|--state file] config_file
```

Any number of configuration files can be given on the command line, and one file can include another config file. A simple logrotate configuration looks like what's shown below:

```
/var/log/messages {
    rotate 5
    weekly
    compress
    olddir /var/log/backup/messages/
    missingok
}
```

Here, every week, the */var/log/messages* file will be compressed and backed up to the */var/log/backup/messages/* folder, and only five rotated log files will be kept around in the system.

Installing logrotate

Log rotation is a utility that comes preinstalled in Linux servers like Ubuntu, CentOS, Red Hat, etc. Check the folder at path */etc/logrotate.d*. If it is not installed, then you can install it manually by using the following commands.

For Ubuntu, type:

```
sudo apt-get install logrotate
```

For CentOS, type:

```
sudo yum install logrotate
```

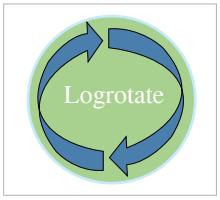


Figure 1: The logrotate utility

Configuring logrotate

When logrotate runs, it reads its configuration files to decide where to find the log files that it needs to rotate, how often the files should be rotated and how many archived logs to keep. There are primarily two ways to write a logrotate script and configure it to run every day, every week, every month, and so on.

1. Configuration can be done in the default global configuration file */etc/logrotate.conf;* or
2. By creating separate configuration files in the directory/*etc/logrotate.d/* for each service/application.

Personally, I think the latter option is a better way to write logrotate configurations, as each configuration is separate from the other. Some distributions use a variation and scripts that run logrotate daily can be found at any of the following paths:

- */etc/cron.daily/logrotate*
- */etc/cron.daily/logrotate.cron*
- */etc/logrotate.d/*

One logrotate configuration (filename: Tomcat) file given below will be used to compress and take daily backups of all Tomcat *.log* files and *catalina.out files* and after rotation, the Tomcat service will get restarted. With this configuration it is clear that multiple log file backups can be taken in one go. Multiple log files should be delimited with space.

```
/home/tomcat/logs/*.log /home/tomcat/logs/catalina.out {
        missingok
        copytruncate
        daily
        compress
        rotate 10
        olddir /var/log/backup/tomcat/
        sharedscripts
        postrotate
                /home/tomcat/bin/catalins.sh restart > /
dev/null

        endscript
}
```

To check if the configuration is functioning properly, the command given below with the *–v* option can be used. Option *-v* means 'verbose' so that we can view the progress made by the logrotate utility.

```
logrotate -dv /etc/logrotate.d/tomcat
```

Logrotate options

-d, --debug	In debug mode, no changes will be made to the logs or to the logrotate state file.
-f, --force	This instructs logrotate to force the rotation, which is necessary as per logrotate: this is useful after adding new entries to a *config* file.
-s, --state <statefile>	Tells logrotate to use an alternate state file. This is useful if logrotate is being run by a different user for various sets of log files. The default state file is */var/lib/logrotate.status.*
-m, --mail <command>	Tells logrotate which command to use when mailing logs. This command should accept two arguments: 1) the subject of the message, and 2) the recipient. The command must then read a message on standard input and mail it to the recipient. The default mail command is */bin/mail -s.*
v, --verbose	Turns on verbose mode.

The types of directives

Given below are some useful directives that can be included in the logrotate configuration file.

Missingok: Continues executing the next configuration in the file even if the log file is missing, instead of throwing an error.

nomissingok: Throws an error if the log file is missing.

compress: Compresses the log file in the *.tar.gz* format. The file can compress in another format using the *compresscmd* directive.

compresscmd: Specifies the command to use for log file compression.

compressext: Specifies the extension to use on the compressed log file. Only applicable if the compress option is enabled during configuration.

copy: Makes a copy of the log file but it does not make any modification in the original file. It is just like taking a snapshot of the log file.

copytruncate: Copies the original file content and then truncates it. This is useful when some processes are writing to the log file and can't be stopped.

dateext: Adds a date extension (default YYYYMMDD), to back up the log file. Also see *nodateext.*

dateformat format_string: Specifies the extension for *dateext.* Only *%Y %m %d* and *%s* specifiers are allowed.

Ifempty: Rotates the log file even if it is empty. Also see *notifempty.*

olddir <directory>: Rotated log files get moved in the specified directory. Overrides *noolddir*.

sharedscripts: This says that *postscript* will run once for multiple configuration files having the same log directory. For example, the directory structure */home/tomcat/logs/*.log* is the same for all log files placed in the *logs* folder, and in this case, *postscript* will run only once.

postscripts: This runs whenever a log is rotated in the configuration file specified block. The number of *postscript* executions for logs placed in the same directory can be overridden with *sharedscripts* directives.

Directives are also related to the intervals at which log files are rotated. They tell logrotate how often the log files should be rotated. The available options are:

1. Hourly (copy the file /etc/cron.daily/logrotate into the /etc/cron.hourly/ directory)
2. Daily
3. Weekly
4. Monthly
5. Yearly

Log files may also be rotated on the basis of file size. We can instruct logrotate to rotate files when the size of the file is greater than, let's say, 100KB, 100MB, 10GB, etc.

Some directives tell logrotate what number of rotated files to keep before deleting the old ones. In the following example, it will keep four rotated log files.

```
rotate 4
```

You can also use directives to remove rotated logs that are older than X number of days. The age is only checked if the log file is to be rotated. The files are mailed, instead of being deleted, to the configured address if *maillast* and *mail* are configured.

One can get the full list of commands used in logrotate configuration files by checking the man page:

```
man logrotate
```

Logrotate is one of the best utilities available in the Linux OS. It is ideal to take backups of applications, servers or any logs. By writing a script in the *postscript* section, we can move or copy backups of log files in Amazon s3 buckets as well. END

By: Manish Sharma

The author has a master's in computer applications and is currently working as a technology architect at Infosys, Chandigarh. He can be reached at *cloudtechgig@gmail.com*

Continued from page...37

Therefore, in the case of PaaS, cloudbursting is much easier than in IaaS. (Cloudbursting is the process of scaling out from private cloud to public cloud resources as per the load/demand on the application.)

8. DevOps orchestration and the API

DevOps can be defined in two ways:

1. It is a new name for automating the release management process that makes developers and the operations team work together.
2. The operations team manages operations by writing code, just like developers.

In DevOps, the application release management and application's resource demand management is of primary importance.

The traditional workflow tools like Jenkins have a new role of becoming orchestrators of all data centre components in an automated workflow. In this age of DevOps and ADDC, every product vendor releases the Jenkins plugins for their products as soon as it releases the product or its updates. This enables all of these ADDC components and the API endpoints to be orchestrated through a tool like Jenkins.

Apart from Jenkins, open source configuration management automation tools like Puppet and Chef can also easily integrate with other layers of ADDC to create a set of programmatic orchestration jobs exposed over API calls to run these jobs. These jobs can be run from API invocation, to orchestrate the data centre through the orchestration of all other API layers.

ADDC is therefore an approach to combining various independent technology solutions to create API endpoints for everything in a data centre. The benefit is the programmability of the entire data centre. Theoretically, a program can be written to do all the jobs that are done by people in a traditional data centre. That is the automation nirvana which will be absolutely free of human errors and the most optimised process, because it will remove human elements from the data centre management completely. However, such a holistic app has not arrived yet. Various new age tools are coming up every day to take advantage of these APIs for specific use cases. So, once the data centre has been converted into an ADDC, it is only left to the developers' imagination as to how much can be automated – there is nothing that cannot be done.

Coming back to what we started with – the move towards architectures like ADDC is surely going to impact jobs as humans will be replaced by automation. However, there is the opportunity to become automation experts instead of sticking to manual labour profiles. Hence, in order to combat the new automation job role demands in the market, one needs to specialise in one or some of these ADDC building blocks to stay relevant in this transforming market. Hopefully, this article will help you build a mind map of all the domains you can try to skill up for. END

By: Abhradip Mukherjee, Jayasundar Sankaran and Venkatachalam Subramanian

Abhradip Mukherjee is a solutions architect at Global Infrastructure Services, Wipro Technologies. He can be reached at *abhradipm@gmail.com.*

Jayasundar Sankaran is a principle architect at Global Infrastructure Services, Wipro Technologies. He can be reached at *jayasuntar@gmail.com.*

Venkatachalam Subramanian is a principal consultant, Talent Transformation, Wipro Technologies. He can be reached at *venkats3dworld@gmail.com.*

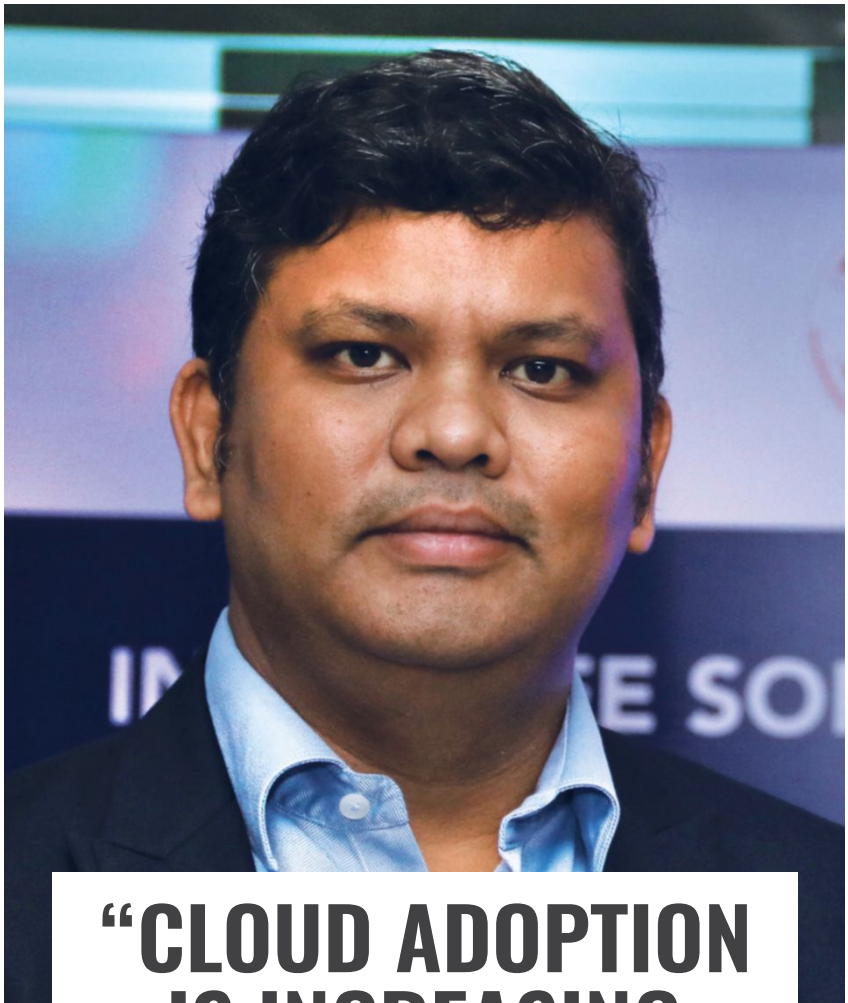
Downtime is money.

Planned or unplanned,IT downtime is expensive—especially for mission critical applications and services. Prevent it and prosper with SUSE technology. SUSE prevents hardware downtime, maximizes service availability, and minimizes human mistakes.

Discover the power of zero.
suse.com/zerodowntime

For more details, contact:
karmendra.trivedi@suse.com/+91 9711177057
prasanth.ms@suse.com/+91 9880924548
akshat.mehrotra@suse.com/+91 9833760622

SUSE
We adapt. You succeed.



"CLOUD ADOPTION IS INCREASING BUT THERE ARE A LOT OF SECURITY CONCERNS"

The cyber world is evolving at a fast pace. But alongside the growth, ensuring security is becoming a major challenge for companies offering cyber solutions. ***Sandip Kumar Panda, CEO, InstaSafe,*** highlights the key trends in the cybersecurity space while speaking with ***Jagmeet Singh*** of ***OSFY.*** Edited excerpts...

Sandip Kumar Panda,
CEO, InstaSafe

Q How has cyber security evolved to meet the sudden increase in Web attacks?

Attacks have evolved over the years from being mere pranks to clearly having profit as the key motive. Many attacks earlier were also state-sponsored, targeting the computers and networks of a rival state. The focus of hackers has now shifted to targeting the networks and data of consumers and businesses, with profit as a key aim, which was evident in the recent ransomware attacks. Cyber security research bodies like the Cloud Security Alliance (CSA) are continuously promoting the use of best practices for providing security assurance, while startups are creating innovative solutions and disrupting the market by delivering cyber security as a service to reduce the risks of adopting new technology.

Q What are the major challenges associated with offering enterprise-level security in developing markets like India?

The primary challenge in selling security solutions in developing markets like India is that users are more inclined to buy from traditional, legacy vendors. They are also not easily convinced that newer technologies such as cloud-based security can work more efficiently and cost less than the legacy systems. But that mindset is changing as the market goes through a tectonic shift from desktops to smartphones as the device of choice. Customers are realising that innovative products built in India effectively address their usage requirements.

To address developing markets, we must look at technologies that are clearly innovative, easy to use, require minimal investment in IT infrastructure and technical manpower, are priced cost-effectively and come packaged with extensive knowledge-based support. Businesses are looking to save costs in a competitive environment, which is the reason why we think the InstaSafe model of delivering trust and security on a subscription plan will be attractive to users.

Q How does cloud adoption help InstaSafe to resolve those challenges?

Cloud adoption is increasing but there are a lot of security concerns. Typically, consultants and solutions service providers assemble a patchy set of security point products to address those concerns, which is not necessarily the best way to go on a long-term basis. At InstaSafe, we have made our solution attractive to businesses as it gives organisations the agility to quickly deploy and scale their application infrastructure while closely integrating the security. Further, we have ensured that our solution is easy to deploy, manage and monitor by the IT staff, and that the end users too, find it easy to use.

Unlike a hardware-based product that takes weeks to have the box delivered and then integrated with the existing infrastructure, our Security-as-a-Service offering can be quickly deployed on any existing hardware. Organisations can also right-size their services and scale as they grow, rather than invest in infrastructure that they may never fully use and, therefore, realise faster ROI.

Q Why would an enterprise need to start relying on a Security-as-a-Service (SaaS) model instead of deploying its own team of IT security experts?

The SaaS model clearly stands out because of the agility it offers. It promises quick deployments, and a pricing that is subscription based, and hence does not require upfront capital investments in servers and other infrastructure that quickly depreciate in value. Hiring a large number of security experts and developing, running and debugging IT security software in-house is not feasible any more, because of the acute shortage of top quality security professionals. Legacy methods also have the management overhead of regularly rolling out patches, in order to make sure that all systems have been correctly configured and upgraded to the new software. Therefore, it makes far more technical and business sense to instead partner with a provider who monitors the threat landscape for you, provides patches quickly and is able to roll them out automatically over the cloud.

> *"A national cyber security policy is certainly vital for the country."*

Q How does InstaSafe Secure Access enable an advanced security layer on top of a hybrid cloud infrastructure?

The solution offered by InstaSafe provides on-demand, scalable, secure access for all users connecting with their corporate issued device or a BYOD, to access applications, located anywhere – the public cloud, the private cloud or on-premise. This is based on the 'need-to-know' access model used in military networks. Our solution creates an 'Enterprise Black Cloud', which is essentially a completely invisible network, and is accessible only after the user and the device that is being used to connect is first verified, and then a level of trust is established. As part of the seven layers of security, InstaSafe Secure Access binds the users to the device(s), ensuring their credentials don't work on any other device, and it only allows access to specific applications based on the 'need-to-know' access model.

Q Is it difficult for hackers to gain backdoor access to a hybrid environment?

All environments, whether hybrid or not, are vulnerable to backdoor attacks, and this is due to the fact that the users and the devices used by them are the weakest links in the enterprise security landscape. Despite the best defences in place, user endpoints can be easily compromised and so the hacker can gain backdoor entry to the enterprise network, including a hybrid one. Once inside, the hacker is able to move laterally with minimum effort. This kind of backdoor entry is feasible due to the trust placed on endpoints once they are inside the network. Google and some other large corporations have started to tackle such attacks. Google's BeyondCorp project provides a very good case study about the benefits of not trusting user endpoints, and provides context sensitive access.

InstaSafe Secure Access is based on the very same principles defined in SDP (software defined perimeter), which ensure that the enterprise network is a 'Black Cloud' and access is granted to the user and device only after a certain trust level is established.

Q What are the big obstacles faced when securing hybrid data centres, and how does your solution save costs for enterprises?

Hybrid data centres require solutions that are flexible, and even better, programmable. Maintaining security across these set-ups (whether a public or private cloud) is different due to the network visibility, access and scalability. Hence, the ideal security solutions to protect hybrid data centres are programmable and scalable, yet easy to deploy, maintain and monitor. InstaSafe Secure Access is a 'software only' solution delivered as a service.

It can scale and adapt along with the access to the hybrid infrastructure. This 'software only' solution that is delivered as a service positively impacts costs by significantly reducing TCO and delivering a faster ROI.

Q What strategies should a company adopt to secure cloud deployments in today's market?

Corporates need to gain a clear understanding of the shared model of security while doing cloud deployments. Typically, the cloud provider secures the hardware infrastructure, while the company needs to ensure that the network access, the operating system security, and the application security are handled effectively. As such, the cloud providers secure the underlying physical infrastructure, ensuring the logical data isolation between different customers, and so on.

Also, corporates need to invest in the skills improvement of their workforce so that they understand these changes and keep an open mind in terms of looking out for innovative security solutions — be it from startups or established vendors.

Q How does InstaSafe help to educate the market about cyber security?

We have partnered with leading organisations like the Data Security Council of India (DSCI), the Cloud Security Alliance (CSA) and the Cloud Computing Innovation Council of India (CCICI) to promote awareness of security at multiple levels, starting with CIOs.

Q Have the recent partnerships with CSA and CCICI enabled InstaSafe to enhance awareness of cloud-based security solutions in the Indian market?

Along with our partnerships with cloud security organisations, last year we published a pioneering survey on cloud and security adoption in India. We are continuing this study in 2017 and will do so in the years to come, aiming to provide an authoritative benchmark for how the country and local organisations have evolved in cloud deployments and cloud security.

Q How do you view VPNs (virtual private networks) in the security landscape?

VPNs have been around for 20 years with minimum innovation. They have limited utility by themselves, as they do not have much flexibility in deployments for hybrid set-ups. In many cases, they do not even allow for fine-tuned, multiple-level access to resources on the organisation's network, as they operate on an 'all-or-nothing' principle for access to the assigned network. SDP solutions offer the flexibility and functionality required for today's set-up, vastly improving the security posture of any enterprise, both of which VPNs cannot fulfil.

> *"The cloud security industry is moving to open source because of a large number of crowd-tested solutions out there that are open source."*

Q What are your views on India's national cyber security policy? Do you think such legal developments are vital for the country?

A national cyber security policy is certainly vital for the country. Till date, we do not have a formal, legally enforceable cyber security policy, but there has been some talk of it being under consideration. We clearly need a legal framework that would address, for example, the basic requirement that data should reside within the country. This would make it possible to prosecute the people responsible for data breaches, within India.

Q Where does open source sit in the world of cloud-based security solutions?

The security industry has to look at using open standards and the concept of sharing as key strategies. Going forward, open source will inevitably become a key element of security as people have to turn to a code base that is easily reusable, and more importantly, has been worked on, tweaked and tested for bugs by a large installed base of users.

Q Do you believe in the philosophy of releasing the code to the public?

Certainly. The cloud security industry is moving to open source because of a large number of crowd-tested solutions out there that are open source. Hypervisors and Apache Web servers are open source, while security protocols like SSL are open. They have been time-tested and crowd-tested, so they have become better than closed source software.

Q Is it lucrative to opt for a career around cyber security?

Most certainly. As a growing shortage of security professionals exists in the market today, there is clearly an opportunity.

Q Lastly, where do you see the world of cloud security moving in the next five years?

What we are going to be seeing is that as network speeds improve, IoT devices such as small little sensors located in industrial and consumer infrastructure will proliferate faster than smartphones and other end user devices. All of these IoT and IIoT sensors will be monitored and managed using cloud set-ups. This will create an environment where cyber security will become ubiquitous, since it will then directly impact the safety and well-being of humanity. END

Low-Code Platforms
The New-born Children of Abstraction

The driving force for most new inventions is need – the need to simplify things. People say technology is evolving at a rapid pace. It is this need to simplify things that is fuelling the evolution of technology.

We are finding it difficult to keep up with our growing needs, though the growth in information technology has helped us a lot to stay in the race. Today, there is a software application that can meet, and at times even exceed, every need. But what happens when the need changes?

This question has led to the inception of low-code development platforms. These combine the simplicity of interface-driven application software with the power of an integrated development environment. They allow you to modify and reconfigure a live application without any system outages.

Scope of app-building platforms

Right now, there are three types of app-building platforms in use:

- No-code platforms
- Low-code platforms
- IDEs (integrated development environments)

Each of these platforms is focused on a different target audience and serves different purposes. Website builders, blogging websites and process automation tools are some examples of a no-code development platform. The solutions built out of these platforms are mainly utility software that satisfy an organisation's internal requirements, primarily substituting forms and spreadsheets.

IDEs, on the other hand, cater to enterprise applications that work around complex business logic. Platforms like Visual Studio, Eclipse, etc, are classic examples of an integrated development environment. IDEs require in-depth programming expertise to build consumable applications.

The evolution of low-code platforms

Traditionally, an application-building process has three core layers. The first layer is the user interface layer via which a user or a customer interacts with the application. The second layer is the underlying business logic, which is the spinal cord that controls how an application works. Finally, there is the backend database, which holds all the data.

The infrastructure comes in only once the core layers are in order. We need to decide on where and how the application is going to be deployed. How secure is the data? Is the application scalable enough to meet the growing needs of the client? The compatibility of the application is also a major concern, like the supported operating systems, secure network connections, set-up and maintenance of servers, etc.

Evolving from high-level programming languages, low-code platforms have significantly abstracted the amount of code needed to create application software, with the infrastructure layer intact. Deploying an application is as simple as clicking a button when built using low-code platforms. Server hosting and maintenance, the underlying database structure and the backend framework are completely managed by low-code platform vendors.

This model of rapid application development has given users a simple interface to build powerful applications. It has brought together two distant cousins — IT and the business stakeholders, enabling faster development of a software product with improved efficiency and streamlined process management.

The future

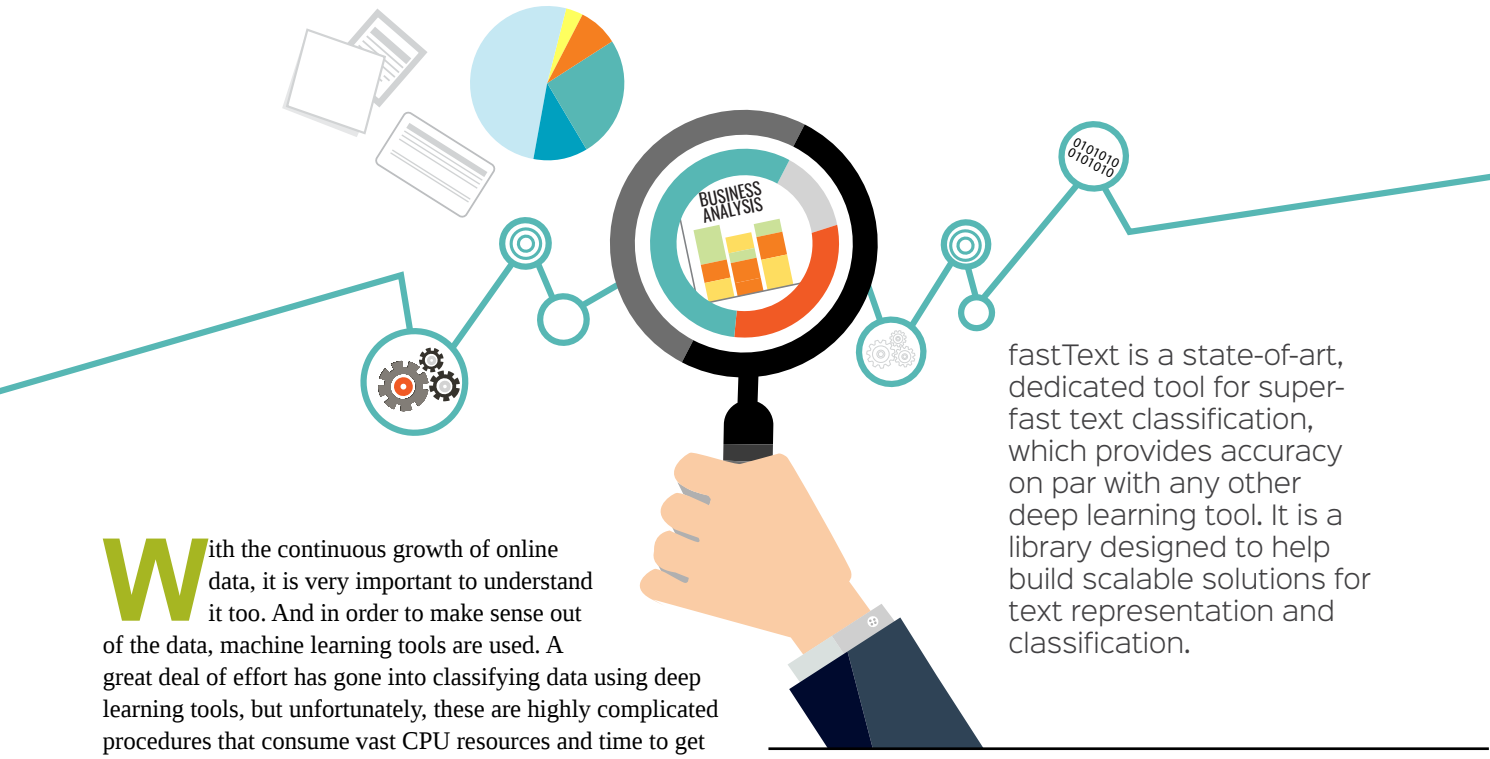
Low-code platforms are now evolving towards greater power and simplicity. While offering users the power of a high-level scripting language to develop enterprise grade solutions, they also maintain the simplicity of a no-code platform so that even people with less coding exposure can build amazing business apps. The best part is that these low-code platforms are platform-independent, which means you can build and configure an application using the Web browser interface and deploy the application across the Web, as well as on Android, iOS or Windows platforms, with minimal effort.

Gartner Inc. has predicted that the market of low-code development platforms will earn revenues of over US$ 10 billion by 2019. If the current trends in this field are anything to go by, the interfaces of these platforms will empower users with increasing simplicity. The low-code platforms of today are becoming more intuitive with each release, because the machine learning field itself is growing. The evolution is so rapid that in less than five years, we can expect to see platforms that build apps on voice commands. Imagine Siri or Google Assistant assembling modules to make an app for you! END

By Adith Mathialagan

The author is a marketing analyst at ZOHO. He can be reached at *adith.m@zohocorp.com*

Fast Text:
Incredibly Fast Text Classification



fastText is a state-of-art, dedicated tool for super-fast text classification, which provides accuracy on par with any other deep learning tool. It is a library designed to help build scalable solutions for text representation and classification.

With the continuous growth of online data, it is very important to understand it too. And in order to make sense out of the data, machine learning tools are used. A great deal of effort has gone into classifying data using deep learning tools, but unfortunately, these are highly complicated procedures that consume vast CPU resources and time to get us results. fastText is the best available text classification library that can be used for blazing fast model training and for fairly accurate classification results.

Text classification is a significant task in natural language processing (NLP) as it can help us solve essential problems like filtering spam, searching the Web, page ranking, document classification, tagging and even something like sentiment analysis. Let us explore fastText in detail.

Why fastText?

fastText is an open source tool developed by the Facebook AI Research (FAIR) lab. It is a library that is dedicated to representing and classifying text in a scalable environment, and has a faster and superior performance compared to any of the other available tools. It is written in C++ but also has interfaces for other languages like Python and Node.js.

According to Facebook, "We can train fastText on more than one billion words in less than 10 minutes using a standard multi-core CPU, and classify half a million sentences among 312K classes in less than a minute." That kind of CPU-intensive classification would generally take hours to achieve using any other machine learning tool. Deep learning tools perform well on small data sets, but tend to be very slow in case of large data sets, which limits their use in production environments.

At its core, fastText uses the 'bag of words' approach, disregarding the order of words. Also, it uses a hierarchical classifier instead of a linear one to reduce the linear time complexity to logarithmic, and to be much more efficient on large data sets with a higher category count.

Comparison and statistics

To test the fastText predictions, we used an already trained model with 9000 Web articles of more than 300 words each and eight class labels. This we looped into the Python API created using the Asyncio framework, which works in an asynchronous fashion similar to *Node.js*. We performed a test using an Apache benchmarking tool to evaluate the response time. The input was *lorem ipsum* text of about 500 lines as a single document for text classification. No caching was used in any of the modules to keep the test results sane. We performed 1000 requests, with 10 concurrent requests each time, and got the results shown in Figure 1.

The result states that the average response time was 8 milliseconds and the maximum response time was 11 milliseconds. Table 1 shows the training time required and accuracy achieved by fastText when compared to other popular deep learning tools, as per the data presented by Facebook in one of its case studies.

Table 1: Comparison between fastText and other deep learning based methods

	Yahoo		Amazon Full		Amazon Polarity	
	Accuracy (per cent)	*Time*	*Accuracy (per cent)*	*Time*	*Accuracy (per cent)*	*Time*
char-CNN	71.2	1 day	59.5	5 days	94.5	5 days
VDCNN	73.4	2 hours	63	7 hours	95.7	7 hours
fastText	72.3	5 seconds	60.2	9 seconds	94.6	10 seconds

```
ubuntu@localhost:~$ ab -n 1000 -c 10 -P testdata 127.0.0.1:3999/categorize
This is ApacheBench, Version 2.3 <$Revision: 1706008 $>
Copyright 1996 Adam Twiss, Zeus Technology Ltd, http://www.zeustech.net/
Licensed to The Apache Software Foundation, http://www.apache.org/

Benchmarking 127.0.0.1 (be patient)
Completed 100 requests
Completed 200 requests
Completed 300 requests
Completed 400 requests
Completed 500 requests
Completed 600 requests
Completed 700 requests
Completed 800 requests
Completed 900 requests
Completed 1000 requests
Finished 1000 requests


Server Software:        Python/3.5
Server Hostname:        127.0.0.1
Server Port:            3999

Document Path:          /categorize
Document Length:        23 bytes

Concurrency Level:      10
Time taken for tests:   0.810 seconds
Complete requests:      1000
Failed requests:        0
Non-2xx responses:      1000
Total transferred:      203000 bytes
HTML transferred:       23000 bytes
Requests per second:    1235.08 [#/sec] (mean)
Time per request:       8.097 [ms] (mean)
Time per request:       0.810 [ms] (mean, across all concurrent requests)
Transfer rate:          244.84 [Kbytes/sec] received

Connection Times (ms)
              min  mean[+/-sd] median   max
Connect:        0    0   0.0      0       0
Processing:     2    8   0.7      8      11
Waiting:        2    8   0.7      7      10
Total:          3    8   0.7      8      11
WARNING: The median and mean for the waiting time are not within a normal deviation
        These results are probably not that reliable.

Percentage of the requests served within a certain time (ms)
  50%      8
  66%      8
  75%      9
  80%      9
  90%      9
  95%      9
  98%      9
  99%      9
 100%     11 (longest request)
ubuntu@localhost:~$
```

Figure 1: Benchmarking with fastText

With a new update in the fastText library, FAIR has introduced compressed text classification models which enable us to use the library even on small memory devices like mobiles and Raspberry Pi. This technique allows models using gigabytes of memory to come down to only a few hundred kilobytes, while maintaining the same performance and accuracy levels.

Now that we know how well fastText can perform, let's set it up.

Configuration and usage

It is quite simple to set up fastText. There are two ways to do this – either get the source and build it yourself, or install the Python interface for it and get started. Let's look at both methods.

Building from the source code: You will just need to get the source code from the Git repository, *https://github.com/facebookresearch/fastText.git*. Then go to the directory and enter *make*, which should compile the code and generate the executable fastText library for you. The output should be as shown in Figure 2.

```
ubuntu@ubuntu:~$ git clone https://github.com/facebookresearch/fastText.git
Cloning into 'fastText'...
remote: Counting objects: 944, done.
remote: Total 944 (delta 0), reused 0 (delta 0), pack-reused 944
Receiving objects: 100% (944/944), 316.03 KiB | 317.00 KiB/s, done.
Resolving deltas: 100% (670/670), done.
Checking connectivity... done.
ubuntu@ubuntu:~$ cd fastText/
ubuntu@ubuntu:~/fastText$ make
c++ -pthread -std=c++0x -O3 -funroll-loops -c src/args.cc
c++ -pthread -std=c++0x -O3 -funroll-loops -c src/dictionary.cc
c++ -pthread -std=c++0x -O3 -funroll-loops -c src/productquantizer.cc
c++ -pthread -std=c++0x -O3 -funroll-loops -c src/matrix.cc
c++ -pthread -std=c++0x -O3 -funroll-loops -c src/qmatrix.cc
c++ -pthread -std=c++0x -O3 -funroll-loops -c src/vector.cc
c++ -pthread -std=c++0x -O3 -funroll-loops -c src/model.cc
c++ -pthread -std=c++0x -O3 -funroll-loops -c src/utils.cc
c++ -pthread -std=c++0x -O3 -funroll-loops -c src/fasttext.cc
c++ -pthread -std=c++0x -O3 -funroll-loops args.o dictionary.o productquantizer.o mat
```

Figure 2: Output

Installation using the Python interface: This is the recommended method, as you can use it later for training and prediction purposes in the same Python script.

The Python module for fastText requires Cython to be installed. Execute the following commands to install Cython and fastText:

```
pip install cython
pip install fasttext
```

And you are done! Just import fastText and use the pretrained model to start predicting the classes.

```
Import fasttext
model = fasttext.load_model('model.bin')
texts = ['fastText is really amazing', 'I love fastText']
labels = classifier.predict(texts)
print labels
```

You can also refer to the Python fastText module documentation at *https://pypi.python.org/pypi/fasttext* for more details.

With the latest machine learning tools like fastText for text classification, you can certainly expect amazing products that utilise these capabilities, particularly in the field of artificial intelligence. END

By: Krishna Modi

The author has a B. Tech degree in computer engineering from NMIMS University, Mumbai and an M. Tech in cloud computing from VIT University, Chennai. He has rich and varied experience at various reputed IT organisations in India. He can be reached at *krish512@hotmail.com*.

TRACKS

- Open Source and You (Success Stories)
- Application Development Day
- Cloud and Big Data
- Database Day
- Open Stack India
- Cyber Security Day
- Open Source in IoT
- Container Day

STAR SPEAKERS

JOERG SIMON
Maintainer,
Fedora Security Lab

PAVAN DEOLASEE,
PostgreSQL Consultant

TONY WASHERMAN,
Professor,
at CMU - Silicon Valley

K. RAJASEKHAR,
Dy Director General, NIC,
MeitY, Govt of India

DR. MICHAEL MESKES,
CEO, credativ
international GmbH

GANESH SAMARTHYAM
Co-Founder,
CodeOps Technologies

SANJAY MANWANI
MySQL India Director,
Oracle

KRISHNA M KUMAR,
Chief Architect in Cloud R&D,
Huawei India

GAGAN MEHRA,
Senior Enterprise
Architect, MongoDB

NEEPENDRA KHARE
Founder,
CloudYuga Technologies

RAVI TRIVEDI,
Founder of PushEngage.com

VIVEK SRIDHAR
Developer Advocate,
DigitalOcean

DEBABRATA NAYAK
Project Director,
NeGD (MeitY)

RAJDEEP DUA
Head, Developer Relations,
Salesforce India

SANIL KUMAR
Chief Architect,
Cloud Team, Huawei India

Register Now!
http://register.opensourceindia.in

www.opensourceindia.in

An EFYGROUP EVENT

Asia's #1 Conference on Open Source

14th Edition

OPEN SOURCE INDIA

13-14 October 2017

NIMHANS Convention Center

BENGALURU

FREEZE YOUR CALENDER NOW!

Stall Booking & Partnering Opportunities Open Now

For more details, call Omar on +91 995 8855 993 or write to info@osidays.com.

PLATINUM PARTNER

Microsoft

GOLD PARTNERS

2ndQuadrant PostgreSQL

NetApp Data Driven

SILVER PARTNERS

mongoDB

DigitalOcean

salesforce

LANYARD PARTNER

wipro

SECURITY PARTNER

InstaSafe

ASSOCIATE PARTNERS

*astTECS THE TRUE IP PBX

FCOOS

LYRA

SUSE

ZOHO

fossology

MySQL

Axcensa Technologies

ORACLE

Mydbops

eXzaTech We Go that Extra Mile

Factory

OUR EXHIBITORS

A Quick Look at Data Mining with Weka

With an abundance of data from different sources, data mining for various purposes is the rage these days. Weka is a collection of machine learning algorithms that can be used for data mining tasks. It is open source software and can be used via a GUI, Java API and command line interfaces, which makes it very versatile.

Waikato Environment for Knowledge Analysis (Weka) is free software licensed under the GNU General Public License. It has been developed by the Department of Computer Science, University of Waikato, New Zealand. Weka has a collection of machine learning algorithms including data preprocessing tools, classification/regression algorithms, clustering algorithms, algorithms for finding association rules, and algorithms for feature selection. It is written in Java and runs on almost any platform.

Let's look at the various options of machine learning and data mining available in Weka and discover how the Weka GUI can be used by a newbie to learn various data mining techniques. Weka can be used in three different ways – via the GUI, a Java API and a command line interface. The GUI has three components—Explorer, Experimenter and Knowledge Flow, apart from a simple command line interface.

The components of Explorer

Explorer has the following components.

Preprocess: The first component of Explorer provides an option for data preprocessing. Various formats of data like ARFF, CSV, C4.5, binary, etc, can be imported. ARFF stands for attribute-relation file format, and it was developed for use with the Weka machine learning software. Figure 1 explains various components of the ARFF format. This is an example of the Iris data set which comes along with Weka. The first part is the relation name. The 'attribute' section contains the names of the attributes and their data types, as well as all the actual instances. Data can also be imported from a URL or from a SQL database (using JDBC). The Explorer component provides an option to edit the data set, if required. Weka has specific tools for data preprocessing, called filters.

The filter has two properties: supervised or unsupervised. Each supervised and unsupervised filter has two categories, attribute filters and instances filters. These filters are used to remove certain attributes or instances that meet a certain condition. They can be used for discretisation, normalisation, resampling, attribute selection, transforming and combining attributes. Data discretisation is a data reduction technique, which is used to convert a large domain of numerical values to categorical values.

Be Up to Date with Technology
Strengthen Reading Habits

And, enjoy these FREE Gifts too...

You & Me

YOGG Wrist Band

Your Personal Fitness and Activity Tracker

Worth ₹2999

Worth ₹1599

Sound Pot

Portable, Aesthetically Designed and Rubberised Premium Finish Bluetooth Speaker

Wireless Headphone Worth ₹1999 + **USB Car Charger** Worth ₹499

5200mAh Power Bank Worth ₹1999

Techies

Raspberry Pi 3

Latest version of Raspberry Pi is a low cost, credit-card sized computer

Worth ₹3300

IoT Starter Kit

Self Learning Kit for Hobbyist, Enthusiast and Students

Worth ₹2000

Arduino Uno Worth ₹890 + **Smart WiFi Module** Worth ₹725

ORDER FORM

Magazine	Duration	Cover Price (₹)	You Pay (₹)	Gifts For You & Me (Please Tick Any One)	Gift for Techies
Open Source For You with DVD	1 Year	1440	1440	☐ 5200mAh Power Bank - worth Rs 1999	☐ Arduino Uno worth Rs 890 + Smart WiFi Module –worth Rs 725
	2 Years	2880	2880	☐ Wireless Headphone - worth Rs 1999 + USB Car Charger – worth Rs 499	☐ IoT Starter Kit – worth Rs 2000
	5 Years	7200	7200	☐ Yogg WristBand - worth Rs 2999 + Sound Pot – worth Rs 1599	☐ Raspberry PI 3 – worth Rs 3300

Free e-zine Access With Every Subscription | **To Subscribe Online, visit: http://subscribe.efyindia.com**

Name_____ Designation_____ Organisation_____

Mailing Address_____

City_____ Pin Code_____ State_____ Phone/Mobile_____ Email_____

Subscription No. (for existing subscribers only_____. **I would like to subscribe Open Source For You starting with the next issue.** Please find enclosed a sum of

Rs_____ by DD/MO/crossed cheque*bearing the No._____ dt._____ in favour of **EFY Enterprises Pvt Ltd,** payable at Delhi. *(*Please add Rs 50 on non-metro cheque)*

Send this filled-in form or its photocopy to: **EFY Enterprises Pvt Ltd** D-87/1, Okhla Industrial Area, Phase 1, New Delhi 110 020 | Ph: **011-26810601-03** | Fax: **011-26817563** | e-mail: **info@efy.in** | **www.efy.in**

EFYGROUP Technology Drives Us

Terms & conditions: # This offer is applicable for new subscriptions and renewal by existing subscribers. # The rates are valid for subscribers within India only. # Please allow 4-6 weeks for processing your subscription and 6-8 weeks to send your gifts # Replacement copies will be sent only if intimation of damaged / non-receipt of copies is received within 15 days of publication # If you are not satisfied with the magazine or our services, we shall refund the balance amount after three months. # Disputes, if any, are subject to exclusive jurisdiction of competent courts and forums in Delhi/New Delhi only.

```
@relation iris-weka.filters.unsupervised.attribute.Remove-R1-2

@attribute petallength numeric
@attribute petalwidth numeric
@attribute class {Iris-setosa,Iris-versicolor,Iris-virginica}

@data
1.4,0.2,Iris-setosa
1.4,0.2,Iris-setosa
1.3,0.2,Iris-setosa
1.5,0.2,Iris-setosa
1.4,0.2,Iris-setosa
1.7,0.4,Iris-setosa
1.4,0.3,Iris-setosa
1.5,0.2,Iris-setosa
1.4,0.2,Iris-setosa
1.5,0.1,Iris-setosa
1.5,0.2,Iris-setosa
1.6,0.2,Iris-setosa
```

Figure 1: ARFF file format

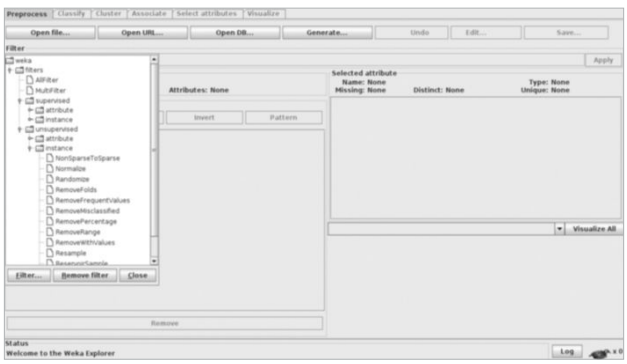


Figure 2: Weka - Explorer (preprocess)

Classify: The next option in Weka Explorer is the Classifier, which is a model for predicting nominal or numeric quantities and includes various machine learning techniques like decision trees and lists, instance-based classifiers, support vector machines, multi-layer perceptrons, logistic regression, Bayes' networks, etc. Figure 3 shows an example of a decision tree using the J4.8 algorithm to classify the IRIS data set into different types of IRIS plants, depending upon some attributes' information like sepal length and width, petal length and width, etc. It provides an option to use a training set and supplied test sets from existing files, as well as cross validate or split the data into training and testing data based on the percentage provided. The classifier output gives a detailed summary of correctly/incorrectly classified instances, mean absolute error, root mean square error, etc.

Cluster: The Cluster panel is similar to the Classify panel. Many techniques like k-Means, EM, Cobweb, X-means and Farthest First are implemented. The output in this tab contains the confusion matrix, which shows how many errors there would be if the clusters were used instead of the true class.

Associate: To find the association on the given set of input data, 'Associate' can be used. It contains an implementation of the Apriori algorithm for learning association rules. These algorithms can identify statistical dependencies between groups of attributes, and compute

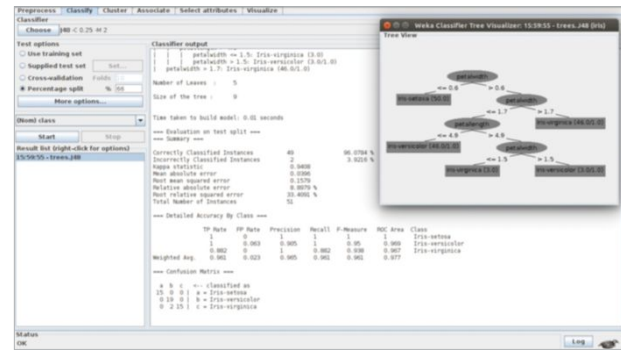


Figure 3: Classification using Weka

all the rules that have a given minimum support as well as exceed a given confidence level. Here, association means how one set of attributes determines another set of attributes and after defining minimum support, it shows only those rules that contain the set of items out of the total transaction. Confidence indicates the number of times the condition has been found true.

Select Attributes: This tab can be used to identify the important attributes. It has two parts — one is to select an attribute using search methods like best-first, forward selection, random, exhaustive, genetic algorithm and ranking, while the other is an evaluation method like correlation-based, wrapper, information gain, chi-squared, etc.

Visualize: This tab can be used to visualise the result. It displays a scatter plot for every attribute.

The components of Experimenter

The Experimenter option available in Weka enables the user to perform some experiments on the data set by choosing different algorithms and analysing the output. It has the following components.

Setup: The first one is to set up the data sets, algorithms output destination, etc. Figure 4 shows an example of comparing the J4.8 decision tree with ZeroR on the IRIS data set. We can add more data sets and compare the outcome using more algorithms, if required.

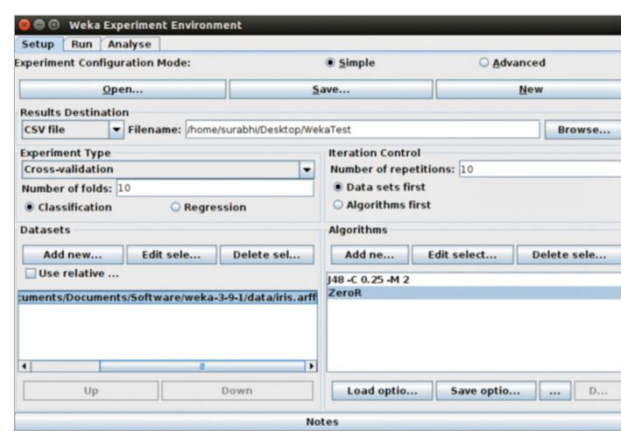


Figure 4: Weka Experimental environment

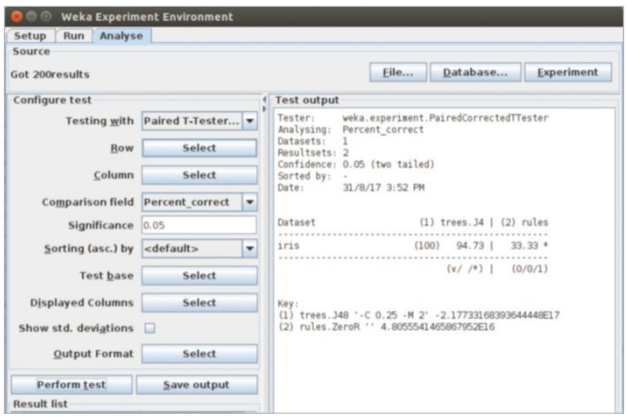


Figure 5: Analysis of the experiment

Run: You can use this tab to run the experiment.

Analyse: This tab can be used to analyse the result. The output of the analysis of the previous experiment is shown in Figure 5. It shows that accuracy using J4.8 is 94.7 per cent and using Zero R is 33.33 per cent. At the bottom of the columns, there is a comparison of the number of times the second algorithm performed better than, the same as, or worse than the first algorithm. In this case, Algorithm 2 is worse than Algorithm 1, on one occasion. The output of the analyser can also be saved.

Weka can be seamlessly used with Java applications as well, just by calling the Java APIs, without writing the machine learning code. Weka for Big Data is still in the evolving phase. The latest distribution of Weka -- 3.9.1 -- sometimes gives a heap size error in the standard settings. That can be rectified by increasing the memory settings of Weka.

Weka provides a very interactive interface for building and testing various machine learning based models. Although there are many machine learning tools available, Weka facilitates quick learning when its powerful GUI is used. Its Experimental tab helps to choose the appropriate machine learning technique. END

Reference

[1] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016) -The WEKA Workbench. Online Appendix for 'Data Mining: Practical Machine Learning Tools and Techniques', Morgan Kaufmann, Fourth Edition, 2016.
[2] *https://weka.wikispaces.com/Use+WEKA+in+your+Java+code*
[3] *www.se.cuhk.edu.hk/~hcheng/WEKA.ppt*
[4] *http://www.cs.waikato.ac.nz/ml/weka/arff.html*

By: Surabhi Dwivedi

The author works at a premier R&D organisation. Her interests include data mining, Big Data, database technologies, data modelling, Linux systems administration and open source technologies. You can contact her at *dwivedisurabhi7@gmail.com*.

A Quick Look at Image Processing with Deep Learning

Social media networks like Facebook have a large user base and an even larger accumulation of data, both visual and otherwise. Face recognition is an important feature of such sites, and has been made possible by deep learning. Let's find out how.

Deep learning is a type of machine learning in which a model learns to perform classification tasks directly from images, text or sound. Deep learning is usually implemented using neural network architecture. The term *deep* refers to the number of layers in the network—the more the layers, the deeper the network. Traditional neural networks contain only two or three layers, while deep networks can have hundreds.

In Figure 1, you can see that we are training the system to classify three animals – a dog, a cat and a honey badger.

A few applications of deep learning

Here are just a few examples of deep learning at work:

- A self-driving vehicle slows down as it approaches a pedestrian crossing.
- An ATM rejects a counterfeit bank note.
- A smartphone app gives an instant translation of a street sign in a foreign language.

Deep learning is especially well-suited to identification applications such as face recognition, text translation, voice recognition, and advanced driver assistance systems, including lane classification and traffic sign recognition.

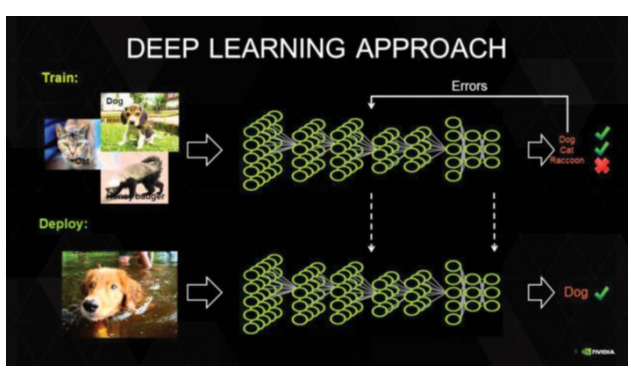


Figure 1: The deep learning approach

The learning process of deep neural networks

A deep neural network combines multiple non-linear processing layers, using simple elements operating in parallel. It is inspired by the biological nervous system, and consists of an input layer, several hidden layers, and an output layer. The layers are interconnected via nodes, or neurons, with each hidden layer using the output of the previous layer as its input.

How a deep neural network learns

Let's say we have a set of images, where each image contains one of four different categories of objects, and we want the deep learning network to automatically recognise which

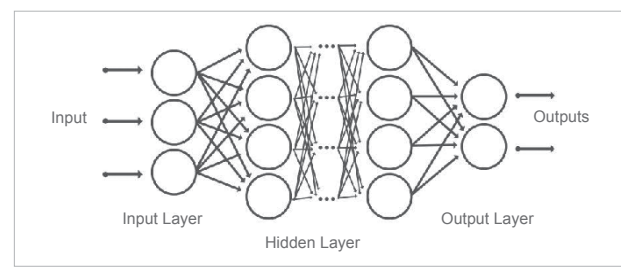


Figure 2: Deep learning neural networks

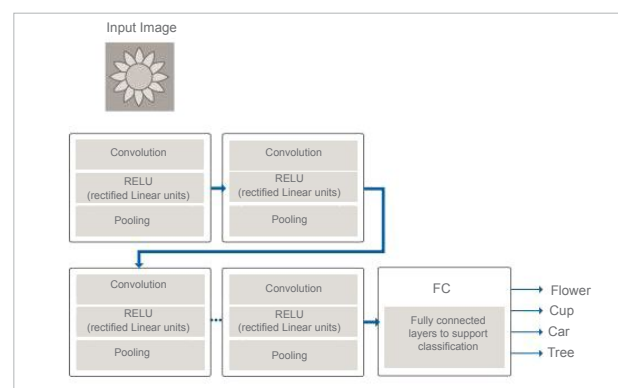


Figure 3: Neural network data training approach

Gift Yourself Knowledge
And get more gifts from us

OFFERS FOR YOU & ME

Portronics

YOGG Wrist Band

Your personal fitness and activity tracker

10W LED Bulb
Branded 10W LED bulb with a two-year warranty

USB Car Charger
Three USB ports with micro USB cable

10W LED Bulb + Three Night Lamps
Set of one 10W LED bulb with three 0.5W night lamps

ORDER FORM

Magazine	Duration	Cover Price (₹)	You Pay (₹)	You Save (₹)	Gifts For You & Me
Electronics For You	1 Year	720	625	95	☐ 10W LED Bulb
	2 Years	1440	1150	290	☐ USB Car Charger
	3 Years	2160	1725	435	☐ 10W LED Bulb+3 Night Lamps
	5 Years	3600	2880	720	☐ Yogg Smart Wristband

To Subscribe Online, visit:
http://subscribe.efyindia.com

Technology Advisers for Businesses and Techies

SUBSCRIBE TO ELECTRONICS FOR YOU, SAVE MONEY & GET GIFTS

OFFERS FOR TECHIES

Mobile Operated Robot – DIY

Robots are always a cool toy for DIYs, hobbyist and engineers. Wouldn't it be fun if the same robot is controlled by your mobile phone. Sounds interesting right?. So here is the a robot kit which works on the principle of "DUAL TONE MULTI FREQUENCY"(DTMF) to help control the robots motion.

Arduino Nano
Arduino Nano is a compact and breadboard-friendly version board based on ATmega328 processor.

ESP8266 NodeMcu WiFi Development Board
NodeMCU is an open-source firmware and development kit that helps you to prototype your IOT product within a few Lua script lines. The Development Kit based on ESP8266, integrates GPIO, PWM, IIC, 1-Wire and ADC all in one board. Power your development in the fastest way combinating with NodeMCU Firmware!

Smart Wi-Fi Module + Arduino Uno R3
SmartWiFi is a wifi module development kit that can help you to prototype your wifi / IOT products

The Arduino Uno R3 is a microcontroller board based ATmega328 AVR microcontroller.

ORDER FORM

Magazine	Duration	Cover Price (₹)	You Pay (₹)	Gifts For Techies
Electronics For You	1 Year	720	650	☐ Arduino Nano
	2 Years	1440	1320	☐ ESP8266 NodeMcu WiFi Development Board
	3 Years	2880	2450	☐ Smart Wi-Fi Module + Arduino Uno R3
	5 Years	3600	3200	☐ Mobile Operated Robot – DIY

Free ezine
Access With Every Subscription

Name _____ Designation _____ Organisation _____

Mailing Address _____

City _____ Pin Code _____ State _____ Phone/Mobile _____ Email _____

Subscription No. (for existing subscribers only) _____ **I would like to subscribe to Electronics For You starting with the next issue.** Please find enclosed a sum of

Rs _____ by DD/MO/crossed cheque*bearing the No. _____ dt. _____ in favour of **EFY Enterprises Pvt Ltd,** payable at Delhi. *(*Please add Rs 50 on non-metro cheque)*

Send this filled-in form or its photocopy to: **EFY Enterprises Pvt Ltd** D-87/1, Okhla Industrial Area, Phase 1, New Delhi 110 020 Ph: **011-26810601-03** Fax: **011-26817563** e-mail: **info@efy.in** **www.efy.in**

EFYGROUP
Technology Times Ltd.

Terms & conditions: # This offer is applicable for new subscriptions and renewal by existing subscribers. # The rates are valid for subscribers within India only. # Please allow 4-6 weeks for processing your subscription and 2 weeks to send your gifts # Replacement copies will be sent only if intimation of damaged / non-receipt of copies is received within 15 days of publication # If you are not satisfied with the magazine or our services, we shall refund the balance amount after three months. # Disputes, if any, are subject to exclusive jurisdiction of competent courts and forums in Delhi/New Delhi only.

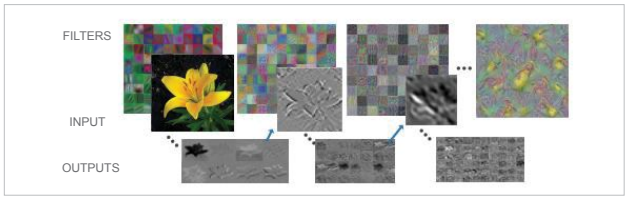


Figure 4: Image processing using deep learning

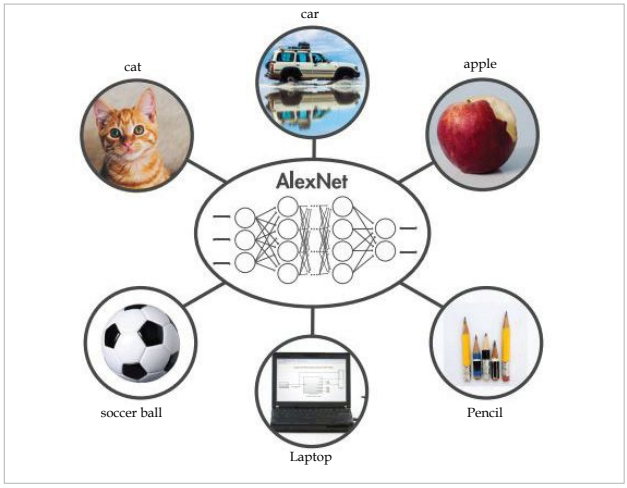


Figure 5: AlexNet

object is in each image. We label the images in order to have training data for the network.

Using this training data, the network can then start to understand the object's specific features and associate them with the corresponding category. Each layer in the network takes in data from the previous layer, transforms it, and passes it on. The network increases the complexity and detail of what it is learning from layer to layer. Notice that the network learns directly from the data—we have no influence over what features are being learned.

Implementation: An example using AlexNet

If you're new to deep learning, a quick and easy way to get started is to use an existing network, such as AlexNet, which is a CNN (convolutional neural network) trained on more than a million images. AlexNet is commonly used for image classification. It can classify images into 1000 different categories, including keyboards, computer mice, pencils, and other office equipment, as well as various breeds of dogs, cats, horses and other animals.

You will require the following software:

1. MATLAB 2016b or a higher version
2. Neural Network Toolbox
3. The support package for using Web cams in MATLAB (*https://in.mathworks.com/matlabcentral/fileexchange/45182-matlab-support-package-for-usb-webcams*)
4. The support package for using AlexNet (*https://in.mathworks.com/matlabcentral/fileexchange/59133-neural-network-toolbox-tm--model-for-alexnet-network*)

Figure 6: The Web cam's output

After loading AlexNet, connect to the Web cam and capture a live image.

Step 1: First, give the following commands:

```
camera = webcam; % Connect to the camera
nnet = AlexNet; % Load the neural net
picture = camera.snapshot; % Take a picture
```

Step 2: Next, resize the image to 227 x 227 pixels, the size required by AlexNet:

```
picture = imresize(picture,[227,227]); % Resize the picture
```

Step 3: AlexNet can now classify our image:

```
label = classify(nnet, picture); % Classify the picture
image(picture); % Show the picture
title(char(label)); % Show the label
```

Step 4: Give the following command to get the output:

```
output:
```

The table below lists a few differences between deep learning and machine learning

Machine learning	Deep learning
Good results with small data sets	Requires very large data sets
Quick to train a model	Computationally intensive
Needs to try different features and classifiers to achieve best results	Learns features and classifiers automatically
Accuracy plateaus	Accuracy is extremely high

END

References

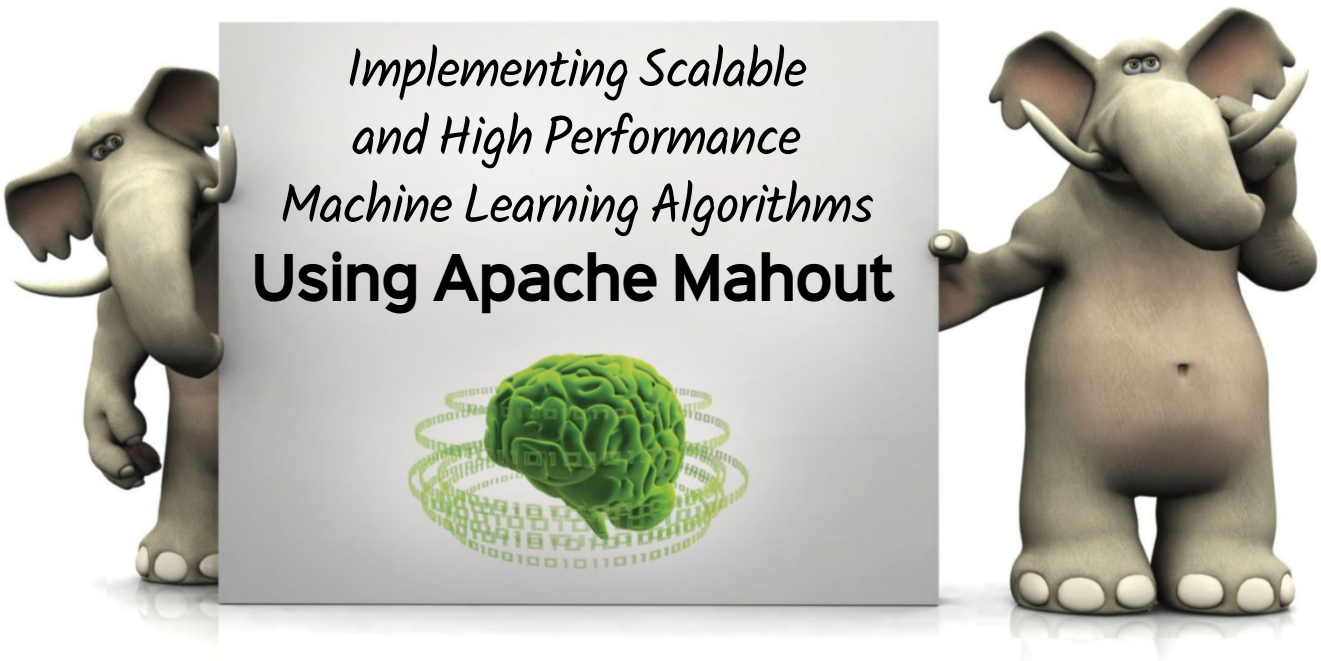
[1] *file:///D:/deep%20learning/80879v00_Deep_Learning_ebook.pdf*
[2] *https://in.mathworks.com/help/nnet/examples/create-simple-deep-learning-network-for-classification.html*
[3] *https://in.mathworks.com/campaigns/products/offer/deep-learning.html*
[4] *https://in.mathworks.com/help/nnet/examples/transfer-learning-and-fine-tuning-of-convolutional-neural-networks.html*

By: Prof. Prakash Patel and Prof. Dulari Bhatt

Prof. Prakash Patel works as an assistant professor in the IT department at Gandhinagar Institute of Technology. You can contact him at *prakashpatelit@gmail.com.*

Prof. Dulari Bhatt also works as an assistant professor in the IT department at Gandhinagar Institute of Technology. You can contact her at *dulari.bhatt@git.org.in.*

Implementing Scalable and High Performance Machine Learning Algorithms Using Apache Mahout



Apache Mahout aims at building an environment for quickly creating scalable and performant machine learning applications.

Machine learning refers to the intelligent and dynamic response of software or embedded hardware programs to input data. Machine learning is the specialised domain that operates in association with artificial intelligence to make strong predictions and analyses. Using this approach, there is no need to explicitly program computers for specific applications; rather, the computing modules evaluate the data set with their inherent reactions so that real-time fuzzy based analysis can be done. The programs developed with machine learning paradigms focus on the dynamic input and data set, so that the custom and related output can be presented to the end user.

There are a number of applications for which machine learning approaches are widely used. These include fingerprint analysis, multi-dimensional biometric evaluation, image forensics, pattern recognition, criminal investigation, bioinformatics, biomedical informatics, computer vision, customer relationship management, data mining, email filtering, natural language processing, automatic summarisation, and automatic taxonomy construction. Machine learning also applies to robotics, dialogue systems, grammar checkers, language recognition, handwriting recognition, optical character recognition, speech recognition, machine translation, question answering, speech synthesis, text simplification, pattern recognition, facial recognition systems, image recognition, search engine analytics, recommendation systems, etc.

There are a number of approaches to machine learning, though traditionally, supervised and unsupervised learning are the models widely used. In supervised learning, the program is trained with a specific type of data set with the target value. After learning and deep evaluation of the input data and the corresponding target, the machine starts making predictions. The common examples of supervised learning algorithms include artificial neural networks, support vector machines and classifiers. In the case of unsupervised learning, the target is not assigned with the input data. In this approach, dynamic evaluation of data is done with high performance algorithms, including k-means, self-organising maps (SOM) and clustering techniques. Other prominent approaches and algorithms associated with machine learning include dimensionality reduction, the decision tree algorithm, ensemble learning, the regularisation algorithm, supervised learning, artificial neural networks, and deep learning. Besides these, there are also the instance-based algorithms, regression analyses, classifiers, Bayesian statistics, linear classifiers, unsupervised learning, association rule learning, hierarchical clustering, deep

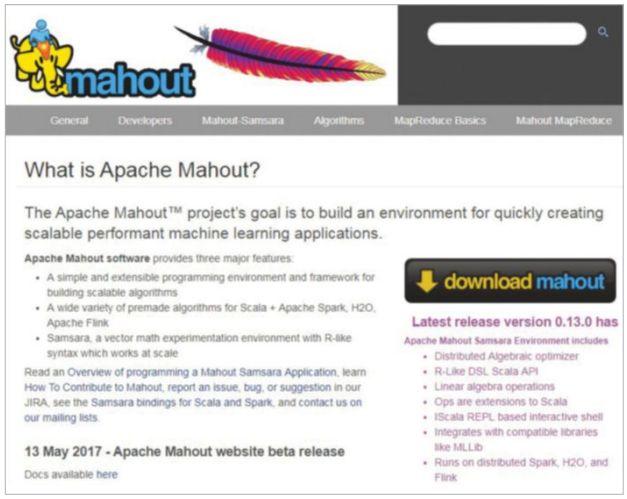


Figure 1: The official portal of Apache Mahout

cluster evaluation, anomaly detection, semi-supervised learning, reinforcement learning and many others.

Free and open source tools for machine learning are Apache Mahout, Scikit-Learn, OpenAI, TensorFlow, Char-RNN, PaddlePaddle, CNTX, Apache Singa, DeepLearning4J, H2O, etc.

Apache Mahout, a scalable high performance machine learning framework

Apache Mahout *(mahout.apache.org)* is a powerful and high performance machine learning framework for the implementation of machine learning algorithms. It is traditionally used to integrate supervised machine learning algorithms with the target value assigned to each input data set. Apache Mahout can be used for assorted research based applications including social media extraction and sentiment mining, user belief analytics, YouTube analytics and many related real-time applications.

In Apache Mahout, a 'mahout' refers to whatever drives or operates the elephant. The mahout acts as the master of the elephant in association with Apache Hadoop and is represented in the logo of the elephant. Apache Mahout runs with the base installation of Apache Hadoop, and then the machine learning algorithms are implemented with the features to develop and deploy scalable machine learning algorithms. The prime approaches, like recommender engines, classification problems and clustering, can be effectively solved using Mahout.

Corporate users of Mahout include Adobe, Facebook, LinkedIn, FourSquare, Twitter and Yahoo.

Installing Apache Mahout

To start with the Mahout installation, Apache Hadoop has to be set up on a Linux distribution. To get ready with Hadoop, the installation is required to be updated as follows, in Ubuntu Linux:

```
$ sudo apt-get update
$ sudo addgroup hadoop
$ sudo adduser --ingroup hadoop hadoopuser1
$ sudo adduser hadoopuser1 sudo
$ sudo apt-get install ssh
$ su hadoopuser1
$ ssh-keygen -t rsa
$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
$ chmod 0600 ~/.ssh/authorized_keys
$ ssh localhost
```

Installing the latest version of Hadoop

Use the following code to install the latest version of Hadoop:

```
$ wget http://www-us.apache.org/dist/hadoop/common/hadoop-
HadoopVersion/hadoop-HadoopVersion.tar.gz
$ tar xvzf hadoop-HadoopVersion.tar.gz
$ sudo mkdir -p /usr/local/hadoop
$ cd hadoop-HadoopVersion/
$ sudo mv * /usr/local/hadoop
$ sudo chown -R hadoopuser1:hadoop /usr/local/hadoop
$  hadoop namenode –format
$  cd /usr/local/hadoop/sbin
$ start-all.sh
```

The following files are required to be updated next:

- *~/.bashrc*
- *core-site.xml*
- *hadoop-env.sh*
- *hdfs-site.xml*
- *mapred-site.xml*
- *yarn-site.xml*

Web interfaces of Hadoop

Listed below are some of the Web interfaces of Hadoop.
MapReduce: *http://localhost:8042/*
NameNode daemon: *http://localhost:50070/*
Resource Manager: *http://localhost:8088/*
SecondaryNameNode:: *http://localhost:50090/status.html*

The default port to access Hadoop is 50070 and *http://localhost:50070/* is used on a Web browser.

After installng Hadoop, the setting up of Mahout requires the following code:

```
$ wget http://mirror.nexcess.net/apache/mahout/0.9/mahout-
Distribution.tar.gz
$ tar zxvf mahout-Distribution.tar.gz
```

Implementing the recommender engine algorithm

Nowadays, when we shop at online platforms like Amazon, eBay, SnapDeal, FlipKart and many others, we notice that most of these online shopping platforms give us suggestions or recommendations about the products that we like or had purchased earlier. This type of implementation or

Figure 2: Simple Logging Facade for Java

Figure 3: Stable JAR files from SLF54J portal

suggestive modelling is known as a recommender engine or recommendation system. Even on YouTube, we get a number of suggestions related to videos that we viewed earlier. Such online platforms integrate the approaches of recommendation engines, as a result of which the related best fit or most viewed items are presented to the user as recommendations.

Apache Mahout provides the platform to program and implement recommender systems. For example, the Twitter hashtag popularity can be evaluated and ranked based on the visitor count, popularity or simply the hits by the users. In YouTube, the number of viewers is the key value that determines the actual popularity of that particular video. Such algorithms can be implemented using Apache Mahout, which are covered under high performance real-time machine learning.

For example, a data table that presents the popularity of products after online shopping by consumers is recorded by the companies, so that the overall analysis of the popularity of these products can be done. The user ratings from 0-5 are logged so that the overall preference for the product can be evaluated. This data set can be evaluated using Apache Mahout in Eclipse IDE.

To integrate Java Code with Apache Mahout Libraries on Eclipse IDE, there are specific JAR files that are required to be added from Simple Logging Facade for Java (SLF4J).

The following is the Java Code module, with methods that can be executed using Eclipse IDE with the JAR files of Mahout to implement the recommender algorithm:

```
DataModel dm = new FileDataModel(new File("inputdata"));
UserSimilarity us = new PearsonCorrelationSimilarity(dm);
UserNeighborhood un = new ThresholdUserNeighborhood(Threshold
Value), us, dm);
UserBasedRecommender r=new GenericUserBasedRecommender(dm,
un, us);
List<RecommendedItem> rs=recommender.recommend(UserID,
Recommendations);
for (RecommendedItem rc : rs) {
System.out.println(rc);
```

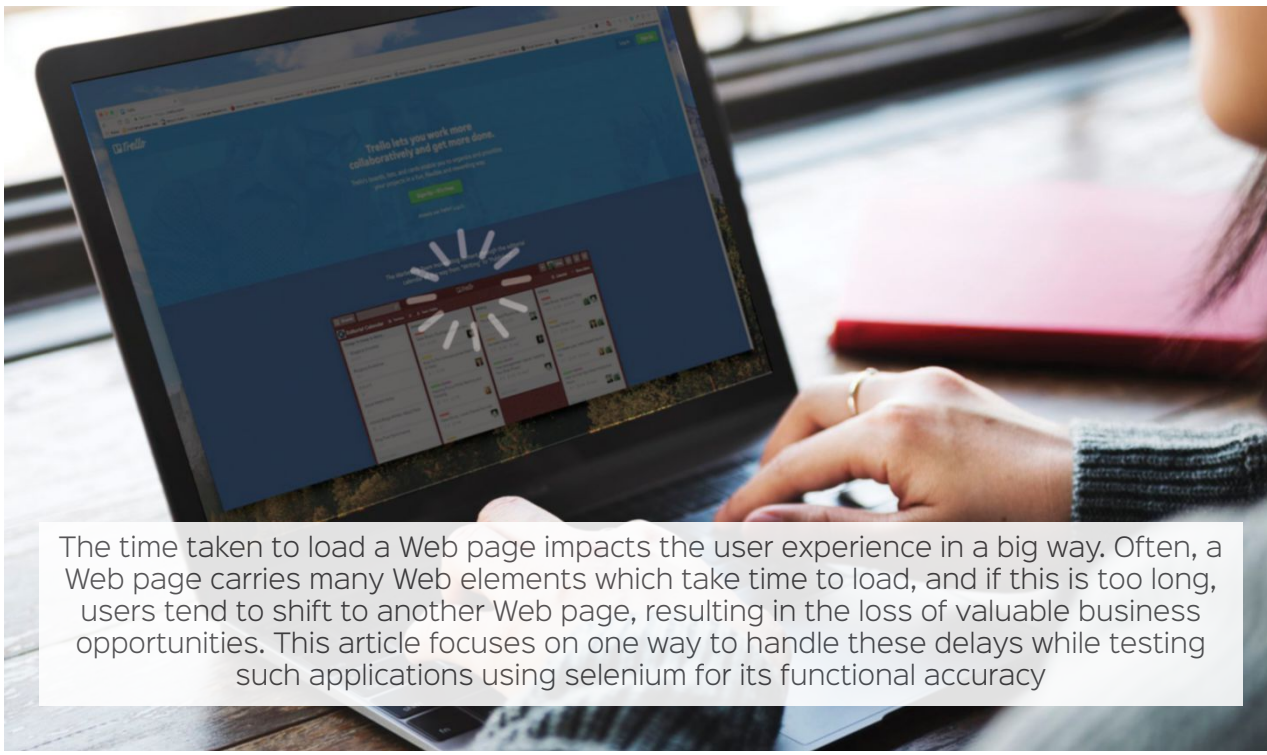
Apache Mahout and R&D

Research problems can be solved effectively using Apache Mahout with customised algorithms for multiple applications including malware predictive analytics, user sentiment mining, rainfall predictions, network forensics and network routing with deep analytics. Nowadays, the integration of deep learning approaches can be embedded in the existing algorithms so that a higher degree of accuracy and optimisation can be achieved in the results. END

By: Dr Gaurav Kumar

The author is the MD of Magma Research and Consultancy Pvt Ltd, Ambala. He is associated with various academic and research institutes, where he delivers expert lectures and conducts technical workshops on the latest technologies and tools. You can contact him at *kumargaurav.in@gmail.com*. Website: *www.gauravkumarindia.com*.

The Long and Short of Waits in Web Applications Using Selenium



The time taken to load a Web page impacts the user experience in a big way. Often, a Web page carries many Web elements which take time to load, and if this is too long, users tend to shift to another Web page, resulting in the loss of valuable business opportunities. This article focuses on one way to handle these delays while testing such applications using selenium for its functional accuracy

Most Web applications we deal with are developed using Ajax and JavaScript. In such cases, when a page is loaded by the browser, the Web elements will take some time to load and the *loading time* can be different in each run. This leads to different exceptions, resulting in a report with the wrong results even though the Web app is working correctly.

Handling the delays with accuracy

On noticing a delay in the appearance of a UI element on the page, let's suppose we use *Thread.sleep()* to handle it. The next time we run the test, the UI element loads instantly. But what about *Thread.sleep()*, which we introduced into our code? It still halts the execution, thereby inducing the extra wait time even though it's not required. Finally, executing our entire test suite takes its own sweet time to finish. Though the results are accurate and reliable, in a practical world, this approach makes no sense when manual intervention testing of the application gets the job done in one minute, compared to our automated script, which takes three minutes due to the extra overhead delay of *sleep()* that we introduced.

So to make automation robust and accurate, as well as to synchronise our test tool with the application under test by handling these delays dynamically, we use *waits* provided by Selenium.

Types of waits

- **Implicit wait:** An implicit wait is to instruct WebDriver to poll the DOM for a certain amount of time when trying to find an element or elements, if these are not immediately available. The default setting is 0. Once set, the implicit wait is set for the life of the WebDriver object instance.

Declaration:

```
driver.manage().timeouts().implicitlyWait(10, TimeUnit.
SECONDS);
```

When to use: Not recommended, but better than using *Thread.sleep()*.

- **PageLoadTimeout wait:** This wait is focused on the time a Web page needs to be loaded – the pageLoadTimeout limits the time that the script allots for a Web page to be displayed before proceeding with other operations on the Web elements present on the page. If the page loads within the time, then the script continues. If the page does not load within the timeout, the script will be stopped by a TimeoutException.

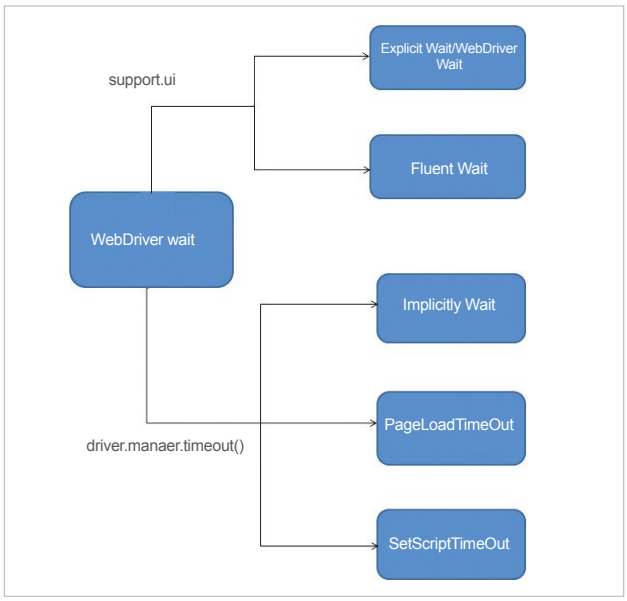


Figure 1: Types of waits

Declaration:

```
driver.manage().timeouts().pageLoadTimeout(50, TimeUnit.
SECONDS);
```

When to use: Whenever you need to ensure a page is loaded in the optimum time before proceeding to perform an operation on WebElements.

- **SetScriptTimeout:** This wait is applicable only for JavascriptExecutor, which provides us two methods for script execution — executeScript and executeAsyncScript.

The *executeScript* method will block the execution of the next line of code till the execution of this method is completed, and then it moves to the next line of code. The *executeAsyncScript* method does not block the execution of the next line of code till the execution of this method is completed.

Since the *executeAsyncScript* method doesn't block the execution of the next line of code, we use *driver.manage(). timeouts().setScriptTimeout(5,SECONDS);* to specify the amount of time for an asynchronous script to finish the execution before throwing an error.

Declaration:

```
driver.manage().timeouts().setScriptTimeout(5, TimeUnit.
SECONDS);
```

When to use: Used to specify the amount of time for an asynchronous script to finish execution, or else it will take its own time to execute and may allow a bug to slip away.

- **FluentWait:** Each FluentWait instance defines the maximum amount of time to wait for a condition, as well as the frequency with which to check the condition. The user may configure the wait to ignore specific types of exceptions while waiting, such as *NoSuchElementExceptions* when searching for an element on the page, and to ignore the exception to continue with script execution in case the element is not found.

Declaration:

```
Wait<WebDriver> wait = new FluentWait<WebDriver>(driver)
.withTimeout(30, TimeUnit.SECONDS).pollingEvery(5,
TimeUnit.SECONDS).ignoring(NoSuchElementException.class);

WebElement clickseleniumlink = wait.until(new Function<WebDriv
er,                          WebElement>() {public WebElement
apply(WebDriver driver) {
     return driver.findElement(By.xpath("//*[@id='login-
button']"));
    }
       });
```

When to use: Used when the element takes a different loading time to appear each time we run the test.

- **Explicit wait:** This allows the user to direct the driver to wait for certain conditions to occur before proceeding further in the code. *WebDriverWait* by default calls the *ExpectedCondition* every 500 milliseconds until it returns successfully.

Declaration:

```
WebDriverWait wait = new WebDriverWait(driver, 10);
wait.until(ExpectedConditions.elementToBeClickable(By.
xpath("//*[@id='login-         button']"))).click();
```

When to use: When the element takes a long time to load and also when there is a need to check the property of an element (e.g., element is clickable).

Effective use of waits in a real world

Using a combination of pageLoadTimeout + SetScriptTimeout and ExplicitWait in your framework ensures that the page is loaded at the optimal time, and that all asynchronous scripts are executed in a stipulated time. ExplicitWait helps to check the various properties of different elements based on the type of operation that needs to be performed. If there are any failures, it's easy to find the root cause for them. Using this combination of waits in your framework can deliver the most robust and accurate results. END

Reference

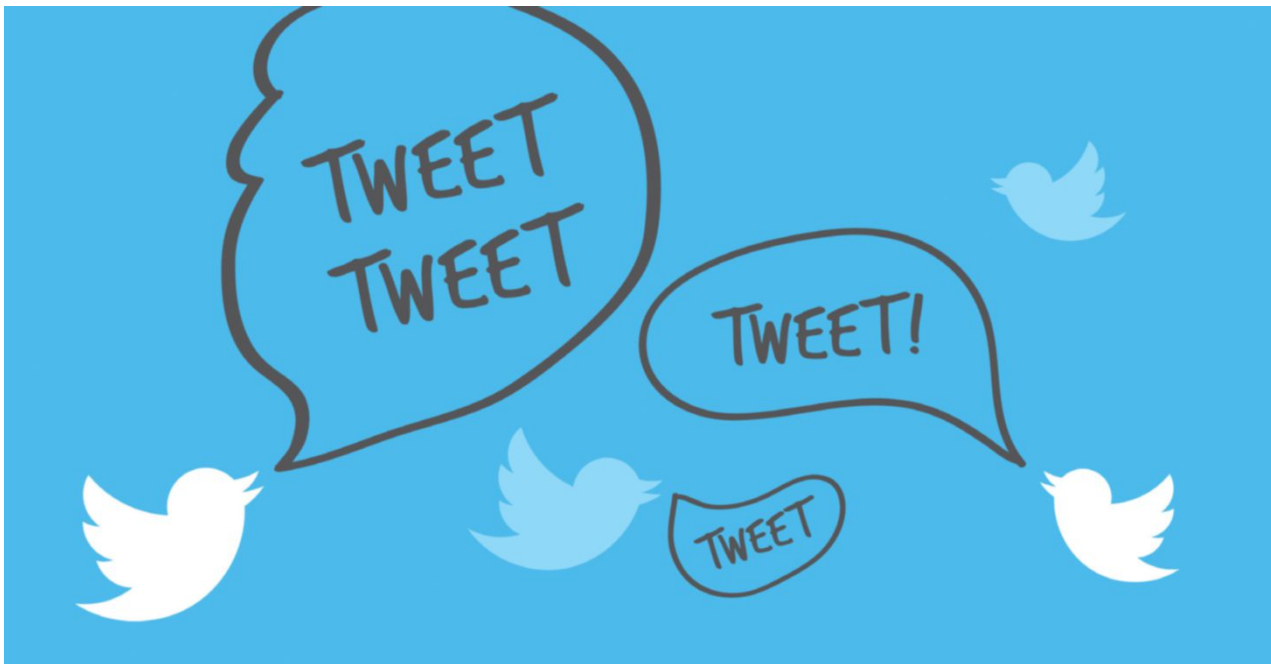
http://www.seleniumhq.org/docs/

By: Athmeeya N.R.

The author works as a Selenium automation developer at DellEmc (EMC2) and has six years of experience in automation testing.

Using Python to Automate Retweets

"By 2022, most Platform-as-a-Service (PaaS) offerings will evolve to a fundamentally serverless model, rendering the cloud platform architectures that are dominant in 2017 as legacy architectures." - Gartner



In the September 2017 issue of *OSFY*, we demystified serverless computing. We also discussed its pros and cons, the technologies involved, use cases, the challenges, etc. So how about using serverless computing to simplify or automate your tasks? For example, wouldn't it be nice to automatically retweet the tweets that have your favourite hashtags (such as #serverless or #opensource). That's the topic we'll explore in this article. We are going to implement a bot (let's call it TweetBot), in which you can specify the hashtags to retweet at specified time intervals.

We will start with writing an auto-retweeting code in Python and get it working on our machine. Later, we will deploy it in OpenWhisk on IBM Bluemix. For this article, we assume you already know the fundamentals of Python programming. You need not know anything about OpenWhisk or Bluemix — we'll introduce these to you in this article.

Developing the TweetBot

In order to write the Python code for auto-retweeting, the prerequisite is that Python 3 must be installed in your machine.

The Twitter API provides programmatic access to read and write Twitter data, create a new tweet, read user profile and follower data, retweet, and more. For this, you must first create a Twitter application (see *https://apps.twitter.com/*) and note down the OAuth credentials for the bot configuration.

The Twitter API in Python can be accessed using the *TwitterFollowBot* Python module. Using this module, you can do much more than just retweet – you can auto-follow and auto-like. For developing the TwitterBot, you must install the module into a folder rather than the default *bin* directory. For that, use the following command:

```
pip install --target <target_folder> TwitterFollowBot
```

Let us now create a program that uses the *TwitterFollowBot* Python module to retweet the latest 'count' number of tweets with a specific phrase ('#Serverless' in our case). For that, we need to create a *TwitterFollowBot* instance. The *TwitterFollowBot* uses the *config.txt* file in the current directory to configure the bot.

So, let us first configure the bot. To do so, you should create a *config.txt* file and fill in the following information so that the bot can connect to the Twitter API. Here are the entries for the *config.txt* file:

```
OAUTH_TOKEN:
OAUTH_SECRET:          API keys from
CONSUMER_KEY:          the twitter app
CONSUMER_SECRET:
TWITTER_HANDLE: <your twitter handle>
ALREADY_FOLLOWED_FILE:already-followed.txt
```

```
FOLLOWERS_FILE:followers.txt
FOLLOWS_FILE:following.txt
USERS_KEEP_FOLLOWING:
USERS_KEEP_UNMUTED:
USERS_KEEP_MUTED:
```

The files *already-followed.txt*, *followers.txt* and *following.txt* contain the respective twitter IDs. You must create these files — you can leave them empty, though. The rest of the fields may also be left empty.

```
# Program to retweet five latest tweets based on a hashtag
(#Serverless)
from TwitterFollowBot import TwitterBot

def retweet():
        # create an instance of the TwitterFollowBot
        # by default, the bot will look for a configuration
          file called config.txt
        # in your current directory
        my_bot = TwitterBot()
        # autoretweets the 5(count)
        latest tweets that matches the hashtag
        my_bot.auto_rt("#Serverless", count = 5)
        return {'message' : 'retweeted successfully'}

retweet()
```

That's the short and sweet code! Let's save it as *test.py* and run it. The output is shown in Figure 1.

When you execute the Python script, it connects to the Twitter API and retweets. The result throws up a warning indicating that the *followers.txt* isn't updated; you can just ignore that warning or update the files to get rid of the warning. The program displays the tweets that are retweeted, so we know it is working.

After we get the code working in our machine, it is time to deploy and execute it in the cloud using the serverless approach. We are going to use Apache OpenWhisk as the serverless platform. We will use IBM Bluemix as the cloud platform to deploy the OpenWhisk action(s).

Apache OpenWhisk

Apache OpenWhisk is an open source serverless cloud platform that executes functions in response to events at any scale.

OpenWhisk was started by IBM and is now incubated by Apache. Adobe is an important contributor to the project and has contributed the API Gateway. OpenWhisk executes functions (called actions) in response to events (called triggers). Since it is serverless, you just need to provide your code to be executed; specifically, you don't need to concern yourself with how to manage the life cycle or operations of the underlying containers that execute the code.

Figure 1: Output for *test.py*

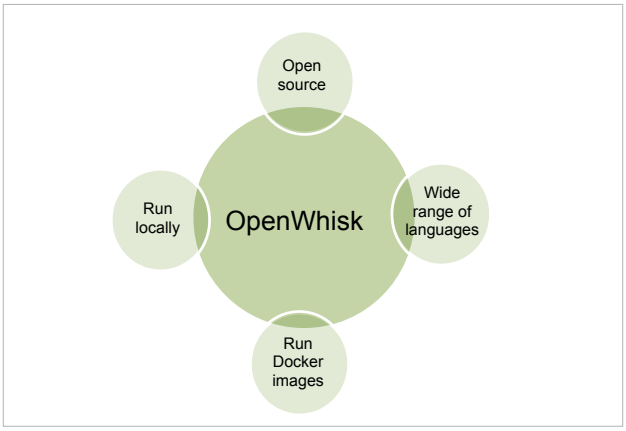


Figure 2: Key aspects of OpenWhisk

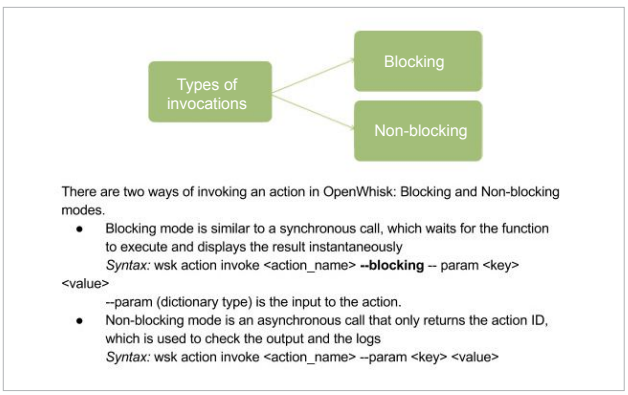


Figure 3: Types of action invocations

You can run OpenWhisk on your laptop, on-premise, or in the cloud. When running in the cloud, you could use a Platform-as-a-Service (PaaS) version of the OpenWhisk provided by IBM Bluemix or you can provision it yourself into Infrastructure-as-a-Service (IaaS) clouds, such as Bluemix, Amazon EC2, Microsoft Azure and Google Cloud Platform.

Here are some key aspects of OpenWhisk:

- It's open source. If you want, you can explore the code and tinker with it, and change it according to your requirements. Support for a wide range of programming languages including Node.js 6, Python 3, PHP 7.1 and Swift 3.1.1 is available.
- Actions (serverless functions) can also be custom executable programs packaged in a Docker container, i.e., you can run Docker images in OpenWhisk.
- You can run it locally! None of the major serverless platforms provide this feature. You can build and run actions locally on your own hardware behind your own firewall, and then deploy it to execute in the cloud.

TweetBot using OpenWhisk

In order to use TweetBot in a serverless approach, you need to install OpenWhisk in your machine and have a BlueMix account. For installing OpenWhisk, refer to *https://console.bluemix.net/openwhisk/cli* for the set-up and configuration information. The beauty of serverless technology is that you don't have to rewrite your entire application; just tweak the plain code that runs in your machine and you'll be fine! Surprisingly, our *TweetBot* requires only one change — it should have a main function with an input parameter (dictionary type) and is to be saved as a *__main__.py* file.

Here is the modified Python code:

```
# Program to retweet five latest tweets based on a hashtag
(#Serverless)
from TwitterFollowBot import TwitterBot

def retweet(dict):
     # create an instance of the TwitterFollowBot
     # by default, the bot will look for a configuration file
     called config.txt
     # in your current directory
     my_bot = TwitterBot()
     # autoretweets the 5(count)
     latest tweets that matches the hashtag
     my_bot.auto_rt("#Serverless", count = 5)

def main(dict):
     retweet(dict)
     return {'message' : 'retweeted successfully'}
```

Now, save this Python code in a file named *__main__.py*. Create the *config.txt*, *already-followed.txt*, *followers.txt* and *following.txt* (as earlier), and zip them all with the *__main__.py* file and the *TwitterFollowBot* dependency module files.

Invocation process

Once the *wsk* CLI (command line interface) is installed and the zip file is ready, follow the steps given below.

Step 1. Create and update an action: Log in to the IBM Bluemix account (*https://console.bluemix.net/openwhisk/*) and create a new action. You can upload the zip files with dependencies only via the CLI. The syntax is:

```
wsk action create <action-name> --kind <language:version>
<file_name>
     Sample Command:
 wsk action create tweetBot --kind python:3 OpenWhisk.zip
```

Step 2. Invoke the function (non-blocking mode): The syntax is:

```
wsk action invoke <action-name>
     Sample Command: wsk action invoke tweetBot
```

Figure 4: Invocation process

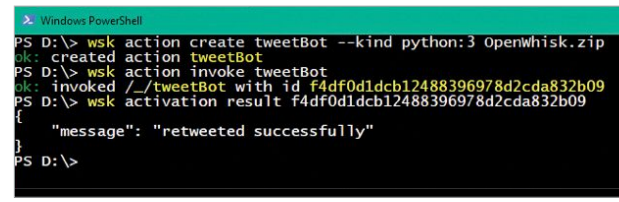


Figure 5: Result of the command

Step 3. Check for the result: Since we invoked the function in a non-blocking mode (because we haven't added the '--blocking' parameter), the command returned immediately, but it is executing in the background. The syntax is:

```
wsk activation result <action ID>
Sample Command:
wsk activation result f4df0d1dcb12488396978d2cda832b09
```

Step 4. Check out the logs: The syntax is:

```
wsk activation logs <action ID>
Sample Command:
wsk activation logs f4df0d1dcb12488396978d2cda832b09
```

Join us at the India Serverless Summit 2017

These are the best of times, and these are the worst of times! There are so many awesome new technologies to catch up on. But, we simply can't. We have seen a progression of computing models — from virtualisation, IaaS, PaaS, containers, and now, serverless — all in a matter of a few years. You certainly don't want to be left behind. So join us at the Serverless Summit, India's first confluence on serverless technologies, being held on October 27, 2017 at Bengaluru. It is the best place to hear from industry experts, network with technology enthusiasts, as well as learn about how to adopt serverless architecture. The keynote speaker is John Willis, director of ecosystem development at Docker and a DevOps guru (widely known for the book 'The DevOps Handbook' that he co-authored). *Open Source For You* is the media partner and the Cloud Native Computing Foundation is the community partner for this summit. For more details, please visit the website *www.inserverless.com.*

Automate the invocation

The moment the action was invoked, your Twitter account would have retweeted the five latest tweets that have the hashtag '#Serverless' in it. However, this is still a manual invocation. For maximum impact, it would be better to

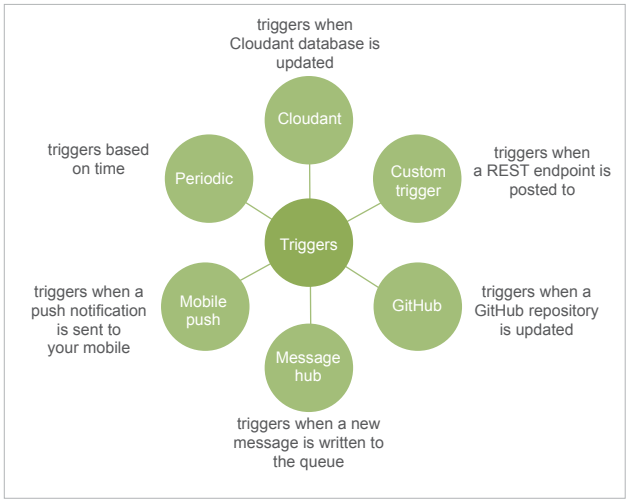


Figure 6: Types of triggers

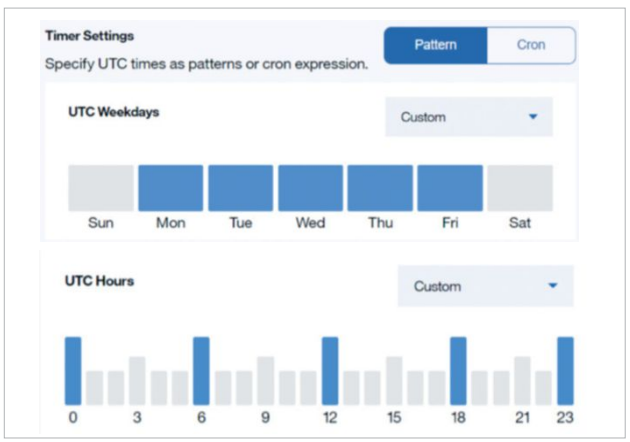


Figure 7: Periodic trigger time settings

automate the invocation process as well, so that you can configure the action and forget it once and for all.

A periodic trigger would be the best option. It triggers the action based on a specific time, and will retweet the latest tweets with '#Serverless' in it. One can either choose a pattern or write a cron expression.

The pattern shown in Figure 7 will invoke the TweetBot action every week day (Monday to Friday) every six hours, keeping your Twitter account live without your intervention!

Words of caution

Before we end this article, here are a couple of things you should be careful about:

1. OpenWhisk is open source and free, but to deploy it, we use IBM Bluemix which isn't free. The TweetBot action will seamlessly run in the free tier, but for any other application, please estimate the monthly costs. The pricing calculator *(https://console.bluemix.net/openwhisk/learn/pricing*) could be handy.
2. This action uses the Twitter API in the background via the *TwitterFollowBot* module. Misusing this may lead to the banning of your Twitter app, so please make sure you clearly read the Twitter automation rules (*https://support.twitter.com/articles/76915*).

In this article, we discussed a TweetBot that can be written to automatically retweet the latest tweets with certain given hashtags. We wrote the actions (serverless functions) in Python, and deployed them in OpenWhisk on IBM Bluemix. We also used triggers to automatically invoke the actions at specified time intervals. With this introduction, we invite you to further explore OpenWhisk as well as other exciting serverless technologies. END

References

[1] Apache OpenWhisk: *https://openwhisk.incubator.apache.org/*
[2] IBM Bluemix Functions: *https://console.bluemix.net/openwhisk/*
[3] Twitter API: *https://dev.twitter.com/rest/public*

By: Ganesh Samarthyam and Srushith Repakula

The authors work at CodeOps Technologies, which is a software technology, consulting and training company based in Bengaluru.

OSFY Magazine Attractions During 2017-18

MONTH	THEME
March 2017	Open Source Firewall, Network security and Monitoring
April 2017	Databases management and Optimisation
May 2017	Open Source Programming (Languages and tools)
June 2017	Open Source and IoT
July 2017	Mobile App Development and Optimisation
August 2017	Docker and Containers
September 2017	Web and desktop app Development
October 2017	Artificial Intelligence, Deep learning and Machine Learning
November 2017	Open Source on Windows
December 2017	BigData, Hadoop, PaaS, SaaS, Iaas and Cloud
January 2018	Data Security, Storage and Backup
February 2018	Best in the world of Open Source (Tools and Services)

Janani Ravi, former Googler and Stanford alumna with her son

"THINGS AREN'T GETTING EASIER FOR MOST CODERS"

Even though many new programming languages and frameworks have emerged lately, many developers still need a guide — either online or offline — to enhance their coding skills. ***Janani Ravi,*** a former Googler and Stanford alumna, closely observed this behaviour before co-founding an online content studio called Loonycorn. Launched in October 2015, the studio produces developer-focused online courses for platforms such as Pluralsight, StackSkills and Udemy. ***Jagmeet Singh*** of ***OSFY*** speaks with Ravi to understand the need for an enhanced training model for developers. Edited excerpts...

Q How did the idea to launch Loonycorn come about?

The seed of the idea came in 2013, when my husband and I were both working with Google in Singapore. We encountered a video by RSA (the Royal Society of Arts) about what motivates people to work. The video has over 10 million hits on YouTube. We were blown away by how vividly we were able to recall the video even years after we watched it. After that, we started experimenting with a video-making tool called Videoscribe, and then started to build its automated version. By 2015, we had built a preliminary version of a video auto-generation engine. We filed for a patent for it in the US as well as in India, but then we realised that we enjoyed making content more than building an engine to help others build content. So we quit our jobs and started working on building niche content full time. By doing this on a regular basis, our startup morphed from a tech startup into a content startup.

Q So what's different about Loonycorn when compared to the others in the market?

We are different in a good way because we are super-small with a team of eight members, have no external funding and are focused on just sitting down in one place for hours on end and doing work that we can be proud of. In about 18 months since we started Loonycorn, we have built about 100 courses and enrolled more than 100,000 students. We have also made lots of mistakes, but they were honest mistakes. I think we have largely avoided making the same mistake twice.

Q How do you research developer-centric content?

Until quite recently, we only made content on subjects that we already knew very well and had used in real life. But now we have started learning topics and technologies, and making content based on them. For this, we plunge in and learn the advancements as systematically and thoroughly as possible. We play around with a lot of ideas, particularly in the area of financial trading; so we try and find use cases for new technologies related to financial trading.

Q What are the challenges you've faced while searching for the most appropriate content for developers?

The biggest challenge for me is getting out of my comfort zone and learning new technologies. There is not much of a challenge in searching for new technologies; there are some amazing technologies to learn from. The challenge is to learn those technologies and become adept at them — it is the challenge of overcoming our own inertia to master new stuff.

Q What, according to you, are the main issues the world of computer science faces today?

One big challenge is that the world is becoming a winner-takes-all world. This is not only in technology — it is happening in every walk of life. In the US, for instance, social mobility is far less than it used to be in the past. Those who have money, skills and talent keep getting more. The have-nots are getting squeezed. I think all this is largely driven by super-specialisation and by the speed at which technology changes. There is no alternative to learning and innovating. If you don't learn, you are no longer relevant and slowly fade away.

Q Do you believe the Internet is making things easier for 21st-century coders?

Things aren't getting easier for most coders. The Internet makes talent very mobile; so, now we all have to keep pace with the best in the world to stay relevant. It is not very different from what is happening in the other spaces such as the restaurant market. A decade ago, a restaurant needed to be good enough or better than other restaurants in a one-kilometre radius to survive. But now, with apps like Zomato and Yelp, and taxi aggregators such as Uber and Ola, customers can even travel for a distance of 15 kilometres to go to a great place to eat rather than sitting and tasting food at a mediocre one.

> *"In about 18 months since we started Loonycorn, we have built about 100 courses and enrolled more than 100,000 students."*

Q How important, do you think, is it for developers to have learning habits to enrich their careers?

Learning habits are getting more and more important in my opinion. As computer science professor Cal Newport writes in his book *Deep Work*, developers need to be prepared to sit down and do 'deep work'. It is satisfying and professionally very rewarding.

Q Do Indian developers lag behind in skills, when compared with their counterparts in developed regions like the US?

There is a grit gap and not a skills gap between Indian and American developers. The best coders here are obviously as good as the best anywhere. In fact, we have so much to be proud of in tech. But at the same time, we need more coders who are like active batsmen. The willingness to stick with a difficult problem, sit in one place for hundreds of hours, code up thousands of lines of code — we are way behind in the perseverance needed to do all these tasks. This grit problem is the root cause of the skills gap — technology changes so fast that everyone without exception is going to find their skills becoming obsolete. The problem is that we are missing the grit to learn new technologies, and prefer to continue to survive in the market with older learnings.

Q You've indicated that financial trading is one of the focus areas of Loonycorn. What are the major offerings in the programming world that help coders to enhance their skills on that front?

My husband (Vitthal Srinivasan) has been a quant trader for quite a while in the past, and his take is that machine learning is the way to distil good trading signals. These days, you can rely on systems such as TensorFlow to ease the experience. It is notably Python that makes systems smarter for financial trading.

Q How does Python help developers build sophisticated financial models?

There is a ton of powerful stuff that developers can leverage to build new solutions for financial trading. Python has Pandas, Numpy, Scikit and, of course, TensorFlow, for supporting deep learning and machine learning in traditional models.

Q Lastly, what's the advice you would like to give developers who are about to step into the coding space?

My brief advice for aspiring coders is to learn to sit down and work long hours — hundreds of hours on a regular basis. They don't have to worry about money or any rewards at the initial stage. These fruits will follow if they are skilled enough. Here, I will use the term *karmayogi* — work for the sake of the work itself. END

Regular Expressions in Programming Languages: PHP and the Web

This is the fourth article in the series on regular expressions. In the past three articles, we have discussed regular expression styles in Python, Perl and C++. Now, we will explore regular expressions in PHP.

Let me first introduce the PHP environment before discussing regular expressions in it. This basic introduction of PHP will be sufficient even for non-practitioners to try out the regular expressions discussed here. Even if you're not interested in PHP, the regular expressions discussed here will definitely interest you.

So let us start with the expansion of PHP. Earlier, the expansion of PHP was 'Personal Home Page'. But now it has been replaced with the recursive backronym 'PHP: Hypertext Preprocessor'. PHP was developed by Rasmus Lerdorf in 1994, and now the PHP development team is responsible for producing the PHP reference implementation. The standard PHP interpreter is free software released under the PHP License. PHP can be called a general-purpose programming language but it is mostly used for Web development as a server-side scripting language. The latest version is PHP 7.1, which was released in December 2016.

Standalone PHP scripts

PHP is mostly used for server-side scripting. Hence, most of the time you will see PHP scripts embedded inside HTML. I am sure all of you have heard about HTML (Hypertext Markup Language), which is used as a markup language for creating Web pages. Even if you are an absolute beginner in HTML, there's no need to worry. You won't need any specific HTML skills to understand the regular expressions

in this article. Even though PHP is almost always paired with HTML, it doesn't mean that you can't have standalone PHP scripts running offline in your machine, but it is a bit weird to use PHP to develop an application that only works in the offline mode. You may find some other programming languages that work better than PHP for such purposes.

The first PHP script we are going to run is a standalone PHP script called *first.php* shown below.

```
<?php

     echo 'I don\'t depend on HTML always';

?>
```

Execute the command *php -f first.php* in a terminal to run the script *first.php*. This and all the other PHP scripts and HTML files discussed in this article can be downloaded from *opensourceforu.com/article_source_code/October17PHP.zip*. It is also possible to make PHP scripts executable. Consider the slightly modified PHP script called *second.php* shown below.

```
#!/usr/bin/php
<?php

     echo 'I don\'t depend on HTML always';

?>
```

Execute the command *./second.php* in a terminal to run the script *second.php*. But before doing this, make sure that you have a PHP executable in your system. Sometimes this executable named 'php' may not be present in the directory */usr/bin*. In that case, find the path to the 'php' executable and replace the line of code *#!/usr/bin/php* with the line of code *#!/YOUR_PATH_TO_PHP/php*. Also, make sure that you have the execute permission for the file *second.php*. Figure 1 shows the outputs of the two PHP scripts *first.php* and *second.php*.

The 'Hello World' script in PHP

In each of the articles in this series, I have discussed a different programming language but I never had the chance to discuss a 'Hello World' program. So here it is — the 'Hello World' script in PHP embedded inside HTML, called *hello.php*, is shown below:

```
<html>
     <head>
       <title>Hello World Script PHP</title>
     </head>
     <body>
       <?php
           echo '<b> Hello World </b>';
       ?>
     </body>
</html>
```

But to run this PHP script, you need a Web server like Apache. My system has XAMPP, which is a free and open source Web server solution stack that provides Apache HTTP Server and MariaDB, a database. XAMPP can also interpret PHP and Perl scripts on its own. Make sure you have Apache HTTP Server available in your system by using XAMPP or a similar LAMP based Web server solution stack. From this point onwards, I assume all of you have XAMPP in your system. Even if you are using a different Web server, it will not affect the output of the PHP scripts in this article. Just make sure that you know how to run PHP scripts with your Web server.

Now if you have XAMPP, use the command *sudo /opt/lampp/lampp start* in a terminal to start the XAMPP service. Of course, you will need root privileges to do this. After this, open a Web browser and type 'localhost' on the address bar. If the XAMPP service is running, you will see the welcome page of XAMPP. To run the PHP script *hello.php*, copy it into the directory */opt/lampp/htdocs*. All the PHP scripts discussed in this article, except *first.php* and *second.php*, should be copied into this directory because we need a Web server to process them. But in the case of *first.php* and *second.php*, this is not necessary because they are standalone PHP scripts and can be executed from anywhere. Now, on the address bar of the Web browser, type *localhost/hello.php*. You will see the Web browser displaying the message 'Hello World' in bold. Figure 2 shows the output of the PHP script *hello.php* in the Mozilla Firefox Web browser.

Now let us examine the script *hello.php* in detail. Most of the HTML tags used in the script, like *<html>*, *<head>*, *<title>*, *<body>*, etc, are self-explanatory; so let us not waste time worrying about them. The PHP interpreter parses the PHP part of the script starting with the opening tag *<?php*

```
# php -f first.php
I don't depend on HTML always
# ./second.php
I don't depend on HTML always
```

Figure 1: Output of standalone PHP scripts

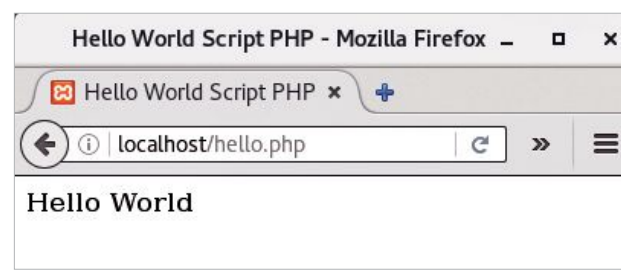


Figure 2: 'Hello World' in PHP

and ending with the closing tag ?> inside which you can have PHP statements separated by semi-colons. The line of PHP code 'echo ' Hello World ';' passes the output ' Hello World ' to the body of the HTML script. Now, a Web browser will process this further by interpreting the HTML tag <*b*> which specifies bold text. This is why bold text is displayed on the Web browser as shown in Figure 2.

Regular expressions in PHP

Now that we know how to set up a server and run PHP scripts, it is time for us to discuss regular expressions in PHP. There are three sets of regular expression functions in PHP to choose from. These are the *preg* functions, *mb_ereg* functions and *ereg* functions. Out of these three, we will be discussing just one set of functions used for regular expression processing, the *preg* functions.

There are some good reasons to choose *preg* functions over the other two. First of all, *preg* is PCRE based. We have already discussed PCRE (Perl Compatible Regular Expressions) regular expression style in detail in the first two articles in this series. Those articles covered Python and Perl, both of which use PCRE style regular expressions. So, it is wise to use this style because then it is not necessary to discuss the syntax of the regular expressions used in PHP. All you have to do is just refresh the syntax you have learned while learning regular expressions in Python and Perl. This is one point in favour of *preg* functions, while there are some faults with the other two sets of regular expressions.

The *mb_ereg* functions are more complicated and are useful only if we are processing multi-byte character sets. We will come across multi-byte character sets only when processing languages like Korean, Japanese or Chinese that have a huge number of characters. As an aside, let me add, unlike most other languages which use an alphabet with a fixed number of characters, these languages have tens of thousands of logograms to represent different words.

Now, let us come back to our business; it would be unnecessary to burden learners by discussing the *mb_ereg* set of functions with no real benefit in sight. And what disqualifies the *ereg* set of functions? Well, they are the oldest set of functions but they were officially deprecated from PHP 5.3.0 onwards. Since we have decided to stick with the *preg* set of functions in PHP to handle regular expressions, we don't need any further discussion regarding the syntax, because we are already familiar with the PCRE syntax.

The main functions offered by the *preg* regular expression engine include *preg_match( ), preg_match_all ( ), preg_replace( ), preg_replace_all( ), preg_split( )*, and *preg_quote( )*. The function *preg_match( )* can give different results based on the number of parameters used in it. In its simplest form, the function can be used with just two parameters as *preg_match($pat, $str)*. Here, the regular expression pattern is stored in the variable *$pat* and the string to be searched is stored in the variable *$str.* This function returns true if the given pattern is present in the string and returns false if no match is found.

A simple PHP script using regular expressions

Now that we have some idea about the regular expression syntax and the working of one function in the *preg* set of functions, let us consider the simple PHP script called *regex1.php* shown below:

```
<html>
    <body>
      <?php
          $pat = '/You/';
          $str = 'Open Source For You';
          if(preg_match($pat,$str))
          {
              echo '<b> Match Found </b>';
          }
          else
          {
              echo 'No Match Found';
          }
      ?>
    </body>
</html>
```

To view the output of this script, open a Web browser and type *localhost/regex1.php* on the address bar. The message 'Match Found' will be displayed on the Web browser in bold text. This script also tells us how the function *preg_match( )* searches for a match. The function searches the entire string to find a match. Let us analyse the script *regex1.php* line by line. The HTML part of the code is straightforward and doesn't need any explanation. In the PHP part of the script, we have used two variables *$pat* and *$str*. The pattern to be matched is stored in the variable *$pat* by the line of code '$pat = '/You/';'. Here we are going for a direct match for the word 'You'. As you might have observed, the delimiters of the regular expression are a pair of forward slashes (/). The variable *$str* contains the string which is searched for a possible match, and this is done by the line of code '$str = 'Open Source For You';'. The next few lines of code have an *if-else* block to print some messages depending on the condition of the *if* statement.

In the line of code 'if(preg_match($pat,$str))' the function *preg_match( )* returns true if there is a match and returns false if there is no match. In case of a match, the line of code 'echo ' Match Found ';' inside the *if* block will print the message 'Match Found' in bold text. In case there is no match, the line of code 'echo 'No Match Found';' in the *else* block will print the message 'No Match Found'.

It is also possible to call the function *preg_match()* with three parameters as *preg_match($pat, $str, $val)* where the array variable *$val* contains the matched string. Consider the PHP script *regex2.php* shown below:

```
<?php
    $pat = '/b+/';
    $str = 'aaaabbbbaaaa';
    if(preg_match($pat,$str,$val))
    {
        $temp = $val[0];
        echo "<b> Matched string is $temp </b>";
    }
    else
    {
        echo 'No Match Found';
    }
?>
```

To view the output of this script, open a Web browser and type 'localhost/regex2.php' on the address bar. The message 'Matched string is bbbb' will be displayed on the Web browser in bold text. This also tells us that the function *preg_match( )* goes for a greedy match, which results in the longest possible match. Thus, the function does not match strings *b*, *bb*, or *bbb*; instead *bbbb* is the matched string. The variable *$val[0]* contains the entire text matched by the regular expression pattern. At this point, I should also mention the difference between strings inside single quotes and double quotes in PHP. The former are treated literally, whereas for the strings inside double quotes, the content of the variable is printed instead of just printing their names.

Other functions in *preg*

There are many other useful functions offered by the *preg* class of functions in PHP for regular expression processing other than the function *preg_match()*. But we will only discuss a very useful function called *preg_replace()* which replaces the matched string with another string. The function can be used with three parameters as follows: *preg_replace($pat, $rep, $str)* where *$pat* contains the regular expression pattern, *$rep* contains the replacement string, and *$str* contains the string to be searched for a pattern. Consider the PHP script *regex3.php* shown below:

```
<?php
    $pat = '/World/';
    $rep = 'Friends';
    $str = 'Hello World';
    if(preg_match($pat,$str))
    {
        $str = preg_replace($pat,$rep,$str);
        echo "<b> The modified string: $str </b>";
    }
    else
    {
        echo 'No Match Found';
    }
?>
```

The function *preg_replace()* will not modify the contents of the variable *$str* as such. Instead the function will only return the modified string. In this example, the line of code '$str = preg_replace($pat,$rep,$str);' replaces the word 'World' with the word 'Friends', and this modified string is explicitly stored in the variable *$str*. To view the output of this script, open a Web browser and type *localhost/regex3.php* on the address bar. The message 'The modified string: Hello Friends' will be displayed on the Web browser in bold text. In case of both *regex2.php* and *regex3.php*, I have only shown the PHP portion of the scripts for want of space, but the complete scripts are available for download.

A regular expression for validating numbers

Now we are going to look at how our knowledge of regular expressions will help us validate numbers using PHP. The aim is to check whether the given number entered through a text box in an HTML page is an integer or a real number, and print the same on the Web page in bold text. If the input text is neither an integer nor a real number, then the message 'Not a number' is displayed on the Web page in bold text. But remember, this statement is factually incorrect as mathematicians will be eager to point out that the input text could still be a number by being an irrational number like Π (Pi) or a complex number like 5 + 10i. It could even be a quaternion or an octonion, even more bizarre number systems. But I think as far as practising computer science people are concerned, integers and real numbers are sufficient most of the times. To achieve this, we have two scripts called *number.html* and *number.php*. The script *number.html* is shown below:

```
<html>
    <body>
      <form action="number.php" method="post">
          Enter a Number:
          <input type="text" name="number">
          <input type="submit" value="CLICK">
      </form>
    </body>
<html>
```

The script *number.html* reads the number in a text field, and when the *Submit* button is pressed the script *number.php* is invoked. The input data is then passed to the script *number.php* by using the POST method for further processing. The script

number.php is shown below. At this point, also remember the naming convention of HTML files. If the HTML file contains embedded PHP script, then the extension of the HTML file is *.php*, and if there is no embedded PHP script inside an HTML script, then the extension of the file is *.html*.

```
<html>
    <body>
      <?php
          $pat1 = '/(^[+-]?\d+$)/';
          $pat2 = '/(^[+-]?\d*\.\d+$)/';
          $str = $_POST["number"];
          if(preg_match($pat1,$str))
          {
              echo '<b> Integer </b>';
          }
          elseif(preg_match($pat2,$str))
          {
              echo '<b> Real Number </b>';
          }
          else
          {
              echo '<b> Not a number </b>';
          }
      ?>
    </body>
</html>
```

The HTML section of the file only contains the tags *<html>* and *<body>* and their meaning is obvious. But the PHP script in the file requires some explaining. There are two regular expression patterns defined by the PHP script stored in the variables *$pat1* and *$pat2*. If you examine the two regular expression patterns carefully, you will understand the benefits of using *preg* which is based on PCRE. I have reused the same regular expression patterns we have discussed in the earlier article dealing with Perl. The line of code '$pat1 = '/(^[+-]?\d+$)/';' defines a regular expression pattern that matches any integer. Even integers like +111, -222, etc, will be matched by this regular expression.

The next line of code '$pat2 = '/(^[+-]?\d*\.\d+$)/';' defines a regular expression pattern that matches real numbers. Here again, we are only identifying a subset of real numbers called rational numbers. But then again, let us not be too mathematical. For a detailed discussion of these regular expressions, refer to the earlier article on Perl, in this series. The best part is that any regular expression that we have developed there can be used in PHP without making changes. I have made a slight change in the second regular expression pattern /(^[+-]?\d*\.\d+$)/ to accommodate real numbers of the form .333 also. The original Perl regular expression was /(^[+-]?\d+\.\d+$)/ which will only validate real numbers like 0.333 and not .333.

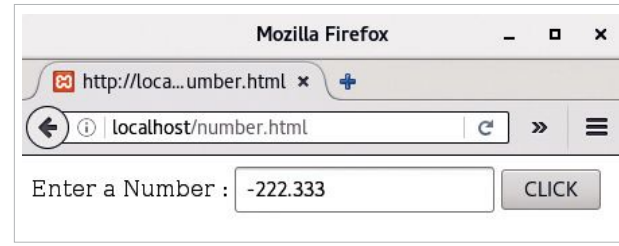


Figure 3: HTML page from *number.html*

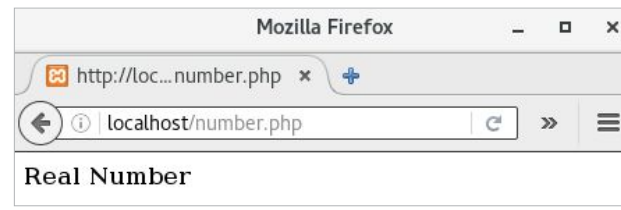


Figure 4: Output of *number.php*

The next line of code '$str = $_POST["number"];' reads the input data from the HTML file *number.html* and stores it in the variable *$str*. The next few lines of code contain an *if-else* block which matches the input text with the two regular expression patterns. The function *preg_match( )* is used in the *if* statement and the *elseif* statement to search for a match. Depending on the results of these matches, the PHP script prints the suitable message in bold text in the Web browser. To view the output of the HTML script, open a Web browser and on the address bar, type *localhost/number.html*. The resulting HTML page is shown in Figure 3. Enter a number in the text field and press the *Submit* button. You will see one of the three possible output messages on the Web page — 'Integer', 'Real Number', or 'Not a number'. Figure 4 shows the output obtained when the number -222.333 is given as input.

Now that we have discussed a useful regular expression, it is time to wind up the article. Here, I have discussed the programming language PHP almost as much as the regular expressions in it. I believe the whole point of this series is to explore how regular expressions work in different programming languages by analysing the features of those programming languages rather than discussing regular expressions in a language-agnostic way. And now that we have covered PHP regular expressions, I am sure you will have some idea about using regular expressions on the server side. But what about regular expressions on the client side? In the last example, the validation could have been done on the client side itself rather than sending the data all the way to the server. So, in the next article in this series, we will discuss the use of regular expressions in JavaScript – a client-side scripting language. END

By: Deepu Benson

The author is a free software enthusiast whose area of interest is theoretical computer science. He maintains a technical blog at *www.computingforbeginners.blogspot.in* and can be reached at *deepumb@hotmail.com*.

An Introduction to Deeplearning4j, the Distributed Deep Learning Library

There are many deep learning libraries that are becoming popular among the developer community such as Theano, Torch, Caffe, etc. Deeplearning4J is an open source and distributed deep learning library targeted at Java Virtual Machine (JVM). This article provides an introduction to its capabilities and features.

Machine learning has led to a tremendous shift in the way we create algorithms to solve problems. It has made developers and researchers shift away from the traditional step-by-step solutions approach to a holistic model, which in many aspects mimics the methods that biological organisms employ while solving problems or acquiring a new skill. These days, machine learning models are employed in all those sectors that were originally difficult to deal with by the traditional algorithmic approaches. Real-time object detection from videos, classification and prediction systems are some of the tasks performed with the help of machine learning. The popular quote by Prof. Tom Mitchell, who is a pioneer in shaping the domain, clearly defines machine learning as follows: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

Deep learning

Deep learning is a recent improvement in the machine learning domain. A basic requirement of machine learning is to identify and build a feature set. This task is generally carried out by experts in that domain, manually. So, for each problem domain, customised features need to be built by expert humans. This often creates a bottleneck in the overall process flow.

Deep learning differs from traditional machine learning in the way features are built. It attempts to build the features automatically from the given *large* data set. Note the emphasis on the word 'large'. For deep learning to work at a decent accuracy level, we need considerably large sized data sets. This is significant, because the machine requires a fairly large amount of data in order to detect the discriminative features of a set. Feature engineering is an important component for the successful implementation of any machine learning related project.

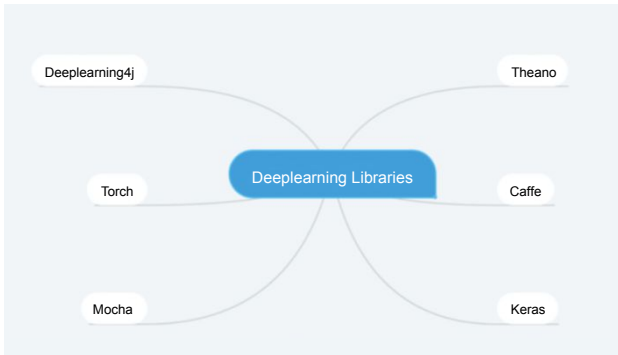


Figure 1: Popular deep learning libraries

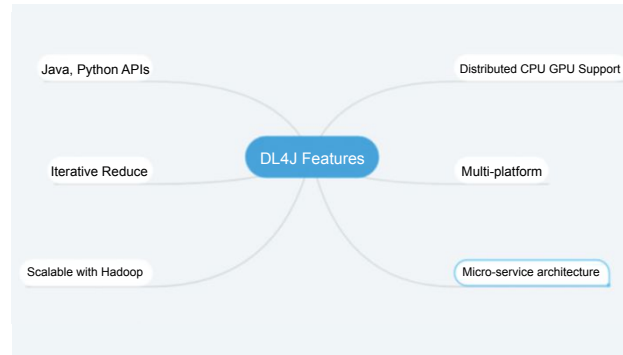


Figure 2: DL4J features

Deep learning frameworks/libraries

There are many popular libraries that can be used to perform deep learning tasks. Some of them are listed below:

- Caffe
- Theano
- Torch
- Deeplearning4J
- Mocha

A detailed comparison of deep learning frameworks is available at *https://deeplearning4j.org/compare-dl4j-torch7-pylearn*.

Deeplearning4j (DL4J)

This article explores the Deeplearning4J (DL4J) library. DL4J has been developed in Java and is targeted at Java Virtual Machine (JVM). An interesting feature of Deeplearning4J is the ability to build fast prototypes. The attributes of DL4J are listed below.

- *Distributed:* The training process can be accelerated because DL4J is distributed by nature. It can harness multiple GPUs, and the performance is on par with other major libraries such as Caffe.
- *Open Source:* DL4J is a production quality open source library available for performing deep learning tasks. The active community of developers keeps DL4J fresh.
- *Interoperability:* DL4J is written in Java, and hence all the JVM based languages such as Scala, Clojure and Kotlin are compatible. Interoperability with other major frameworks such as Caffe and Torch can be achieved through Keras.

DL4J features

The major features of DL4J are:

- Java, Scala and Python APIs
- Parallel training through iterative reduce
- Scalable with Hadoop
- Distributed CPU and GPU support

DL4J incorporates both a distributed, multi-threaded deep learning framework and a single-threaded deep learning framework. Another important feature of DL4J is that it is the first deep learning framework adopted for a microservice architecture.

Prerequisites

The prerequisites to start development with DL4J are listed below:

- Java 1.7 or above (DL4J supports only the 64-bit versions).
- *Apache Maven:* This is a dependency management and build tool. The Maven version in your system can be checked with the following command:

```
mvn --version
```

If you are new to Maven, an excellent 'getting started' guide (Maven in Five Minutes) is available at *http://maven.apache.org/guides/getting-started/maven-in-five-minutes.html*

- The official document recommends the use of InelliJ IDE or Eclipse. The community edition of InelliJ IDE can be downloaded from the official website.
- *Git:* Get the latest version with the following command:

```
$ git clone git://git.kernel.org/pub/scm/git/git.git
```

Using the DL4J examples

To download and use the examples from DL4J, use the following commands:

```
$ git clone https://github.com/deeplearning4j/dl4j-examples.git
$ cd dl4j-examples/
$ mvn clean install
```

1. Run IntelliJ. Choose the *Import Project* option. Select the folder 'dl4j-example'.
2. Select 'Import Project from external model' and make sure that Maven is chosen.
3. Follow the wizard's instructions. Click *Finish* to complete. It might take some time to get all the dependencies in the local system.

4. Select a particular example and right-click the file to run it.

The deep neural networks are made up of multiple layers. The MultiLayerConfiguration with parameters is customised to suit the requirements of the current problem. The hyperparameter variable decides the learning behaviour of the network. These parameters include the following:

- Number of times to update the weights of the model
- Initialisation mechanism for these weights
- The type of activation function to link with the nodes
- The specific optimisation algorithm to be used
- The speed with which the network should learn

A sample configuration is shown in the following code snippet:

```
MultiLayerConfiguration conf = new NeuralNetConfiguration.
Builder()
        .iterations(1)
        .weightInit(WeightInit.XAVIER)
        .activation("relu")
        .optimizationAlgo(OptimizationAlgorithm.STOCHASTIC_
GRADIENT_DESCENT)
        .learningRate(0.05)
        // ... other hyperparameters
        .list()
        .backprop(true)
        .build();
```

A new layer can be added by invoking *layer()* on *NeuralNetConfiguration.Builder()*. In this process, we need to specify the following:

- Location (order) where the layer has to be added
- The number of input and output nodes (nIn and nOut)
- The type of layer

```
layer(0, new DenseLayer.Builder().nIn(784).nOut(250)
                .build())
```

After completing the configuration process, the training of the model can be carried out with the following command:

```
model.fit
```

DL4J's neural networks

DL4J supports many powerful neural networks. Some of them are listed below:

- Restricted Boltzmann machines
- Convolutional nets
- Recurrent nets
- Deep-belief networks
- Stacked denoising autoencoders

The reason for choosing JVM

Most of the deep learning/machine learning libraries are in Python. But DL4J is based on JVM. One may ask why this deliberate choice was made. The official documentation lists the major reasons for choosing the Java Platform:

- Many large government organisations and companies have Java as their major platform. There are millions of Java developers and Java is the largest programming language, as of now.
- Java is the basis of many powerful techniques/tools such as Hadoop, ElasticSearch, Lucene and Hive.
- One prime complaint against Python is its slow speed. Java is definitely quicker than Python. Hence, when handling projects with massive data sets, Java may be a better choice.
- Java is inherently secure and cross-platform. It can be easily used on all major platforms such as Linux, Windows, OSX and Android.

DL4J on Android

Deeplearning4j can be used with Android mobile phones as well. The prerequisites for running it on Android are:

- Emulator (with API level 21 or above) or a device
- Android Studio 2.2 or later

The following dependencies must be added to the *build.gradle* file:

```
compile 'org.deeplearning4j:deeplearning4j-core:0.8.0'
compile 'org.nd4j:nd4j-native:0.8.0'
compile 'org.nd4j:nd4j-native:0.8.0:android-x86'
compile 'org.nd4j:nd4j-native:0.8.0:android-arm'
compile 'org.bytedeco.javacpp-presets:openblas:0.2.19-
1.3:android-x86'

compile 'org.bytedeco.javacpp-presets:openblas:0.2.19-
1.3:android-arm'
```

A clean and easy-to-follow tutorial is available at *https://deeplearning4j.org/android.*

This article has provided an introduction to the Deeplearning4J library. If you are interested in exploring more about deep learning, there are many interesting online courses and resources. A compiled list of links is available at *https://deeplearning4j.org/deeplearningforbeginners.html.* END

References

[1] Deeplearning4j official homepage: *https://deeplearning4j.org/*
[2] Maven: *http://maven.apache.org/guides/getting-started/maven-in-five-minutes.html*

By: Dr K.S. Kuppusamy

The author is an assistant professor of computer science, School of Engineering and Technology, Pondicherry Central University. He has more than 12 years of experience in academia and industry. He can be contacted at *kskuppu@gmail.com.*

Converting a Django App to a REST API

This article demonstrates how easy it is to convert a Django app into a REST API using the REST framework. The underlying assumption here is that the reader's machine has a working version of Python and Python-pip.

Django is a well-known Web framework, which allows developers to develop ORM-based apps at a fast pace. If you don't have Django installed on your system, use *pip install Django* on a terminal. It will install Django on your system.

Once Django is installed, let's create a project, within which we will create a *contacts* app. Open a terminal or command prompt and execute the following command [refer to Figure 2]:

```
cd myproject
django-admin startapp contacts
```

Open the *myproject* directory and you can see the structure shown in Figure 3.

Open *models.py* in the *contacts* directory and create a *class Contact* to represent the contact:

```
from django.db import models

# Create your models here.
```

```
[root@localhost work]# pip install django
WARNING: Running pip install with root privileges is generally not a good idea. Try `pip install --user` instead
Collecting django
  Downloading Django-1.11.3-py2.py3-none-any.whl (6.9MB)
    100% |████████████████████████████████| 7.0MB 153kB/s
Collecting pytz (from django)
  Downloading pytz-2017.2-py2.py3-none-any.whl (484kB)
    100% |████████████████████████████████| 491kB 1.4MB/s
Installing collected packages: pytz, django
Successfully installed django-1.11.3 pytz-2017.2
[root@localhost work]# cd
```

Figure 1: Installing Django using Pip

```
[work@localhost demoapp]$ django-admin startproject myproject
[work@localhost demoapp]$ cd myproject/
[work@localhost myproject]$ django-admin startapp contacts
[work@localhost myproject]$
```

Figure 2: Creating a project and app

```
class Contact(models.Model):
    firstname = models.CharField(max_
length=100)
   lastname = models.CharField(max_
length=100, null=True, blank=True)
    address = models.TextField(max_
length=100)
    city = models.CharField(max_
length=100)
    state = models.CharField(max_
length=100)
    country = models.CharField(max_
length=100)
    home_phone = models.
IntegerField(null=True, blank=True)
    work_phone = models.
IntegerField(null=True, blank=True)
    home_other = models.
IntegerField(null=True, blank=True)
    work_other = models.
IntegerField(null=True, blank=True)
    mobile_prim = models.IntegerField()
    mobile_seco = models.

IntegerField(null=True, blank=True)
```

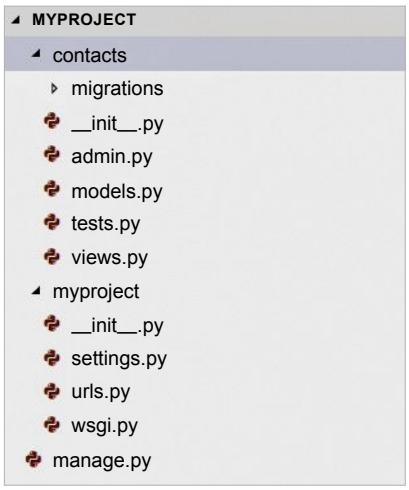


Figure 3: Project and app directory structure

```
# Application definition

INSTALLED_APPS = (
    'django.contrib.admin',
    'django.contrib.auth',
    'django.contrib.contenttypes',
    'django.contrib.sessions',
    'django.contrib.messages',
    'django.contrib.staticfiles',
    'contacts'
)
```

Figure 4: *settings.py* in the *myproject* directory

Open *settings.py* in the *myproject* directory and add *contacts* in the installed app (refer to Figure 4).

In *settings.py*, at the end you can see databases that show where the database and table will be created. For this example, we will use SQLite3 as the database. To create a *contacts* table, execute the following command:

```
python manage.py makemigrations
```

This will create the migrations script. You can see that the migrations directory is created in the *contacts* directory with *0001_initial.py*. Now, using the *migrate* command, create the table. It will create some other tables also.

```
python manage.py migrate
```

You can see the *db.sqlite3* file in the *myproject* directory. Our app is ready, but still has no view that will generate any output. We can use Django templates to create a view, but can omit that section as we want a REST based interface for contacts. The next section will explore how to convert this app to a REST API.

```
(django1.8) λ python manage.py makemigrations
Migrations for 'contacts':
  0001_initial.py:
    - Create model Contact
```

Figure 5: Output of *makemigrations* command

```
(django1.8) λ python manage.py migrate
Operations to perform:
  Synchronize unmigrated apps: staticfiles, messages
  Apply all migrations: admin, contenttypes, sessions, auth, contacts
Synchronizing apps without migrations:
  Creating tables...
    Running deferred SQL...
  Installing custom SQL...
Running migrations:
  Rendering model states... DONE
  Applying contenttypes.0001_initial... OK
  Applying auth.0001_initial... OK
  Applying admin.0001_initial... OK
  Applying contenttypes.0002_remove_content_type_name... OK
  Applying auth.0002_alter_permission_name_max_length... OK
  Applying auth.0003_alter_user_email_max_length... OK
  Applying auth.0004_alter_user_username_opts... OK
  Applying auth.0005_alter_user_last_login_null... OK
  Applying auth.0006_require_contenttypes_0002... OK
  Applying contacts.0001_initial... OK
  Applying sessions.0001_initial... OK
```

Figure 6: Running the *migrate* command

Why REST?

By using REST we ensure that our application can be accessed from anywhere and have any front-end. Let's suppose we want to create a website using the latest technology like React or Angular, and also a mobile app or hybrid mobile app. However, if the website already exists and we want to create only a mobile app or React based app, we need to modify and return JSON instead of HTML. Instead of creating two different types of back-end logic to return HTML, JSON or some other format, we can expose our application as a REST API, which is for any client to consume and display it in the way it is capable of. Any client requesting the resource will get JSON as output, and will decide how to parse and display data.

For converting the application to a REST API, we need to use the *rest_framework* and perform the following steps.

- Open a terminal or command prompt and execute the following command:

```
pip install djangorestframework
```

- Open *settings.py* in the *myproject* directory, and add *rest_framework* in *INSTALLED_APPS*.
- Create new files named *serializers.py*, *views.py* and *urls.py* in the *contacts* directory.

For any model to be exposed as REST we need to create a *serializer* class in *serializer.py*. We will create the *ContactSerializer* class with all the fields. Open *serializer.py* and add the following code to it:

```
from rest_framework import serializers
from models import Contact

class ContactSerializer(serializers.
ModelSerializer):
    class Meta:
        model = Contact
        fields = '__all__'
```

We need to specify URLs and views for the file. First, let's create a view with the *get* and *post* methods. The *get* method will return all data in JSON format, and the *post* method will save the data after validating it. This can be extended to use the *delete* and *put* logic also. Open the *views.py* file in the *contacts* directory and add the following code:

```
from rest_framework import status
from rest_framework.decorators import
api_view
from rest_framework.response import
Response
from models import Contact
from serializers import
ContactSerializer

@api_view(['GET', 'POST'])
def contact_list(request):
    """
// List all snippets, or create a new
snippet.
    """
    if request.method == 'GET':
        contacts = Contact.objects.
all()
        serializer =
ContactSerializer(contacts, many=True)
        return Response(serializer.
data)

    elif request.method == 'POST':
        serializer =
ContactSerializer(data=request.data)
        if serializer.is_valid():
            serializer.save()
            return Response(serializer.
data, status=status.HTTP_201_CREATED)
        return Response(serializer.
errors, status=status.HTTP_400_BAD_
REQUEST)
```

We need to specify URLs to reach this view. Open *urls.py* in the *myproject* directory, where we will include the *contacts* view for processing.

```
from django.conf.urls import include,
url
from django.contrib import admin

urlpatterns = [
    # Examples:
    # url(r'^$', 'myproject.views.
home', name='home'),
    # url(r'^blog/', include('blog.
urls')),

    url(r'^admin/', include(admin.site.
urls)),
    url(r'^contacts/',
include('contacts.urls')),
]
```

Open *urls.py* in the *contacts* directory, which will point to the view we created in *views.py* in this directory.

```
from django.conf.urls import include,
url
from django.contrib import admin
from contacts import views

urlpatterns = [
    # Examples:
    # url(r'^$', 'myproject.views.
home', name='home'),
    # url(r'^blog/', include('blog.
urls')),

    url(r'^api/', views.contact_list),

]
```

Now that we have done that, *contacts/api* will call our *view contact_list* and display all data from *contacts*. To test this, open a terminal and start the server using the following command:

```
python manage.py runserver
```

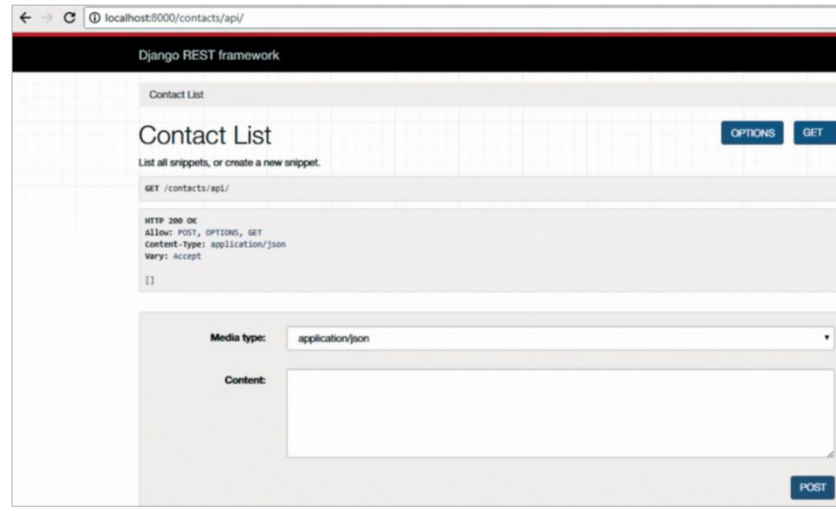


Figure 7: Output of the first rest API

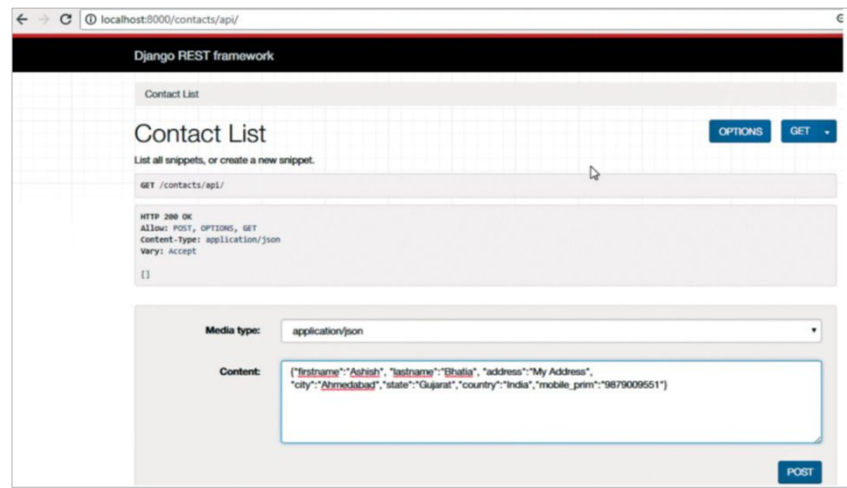


Figure 8: Posting data

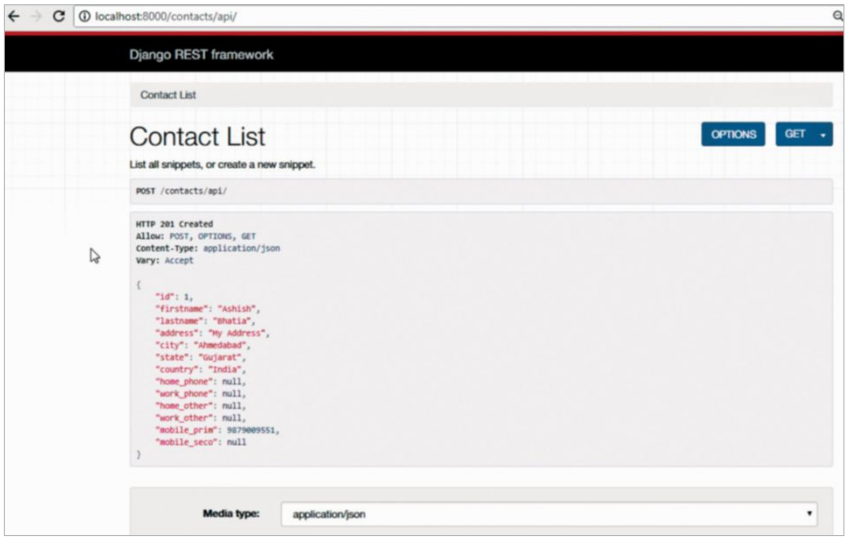


Figure 9: After posting data

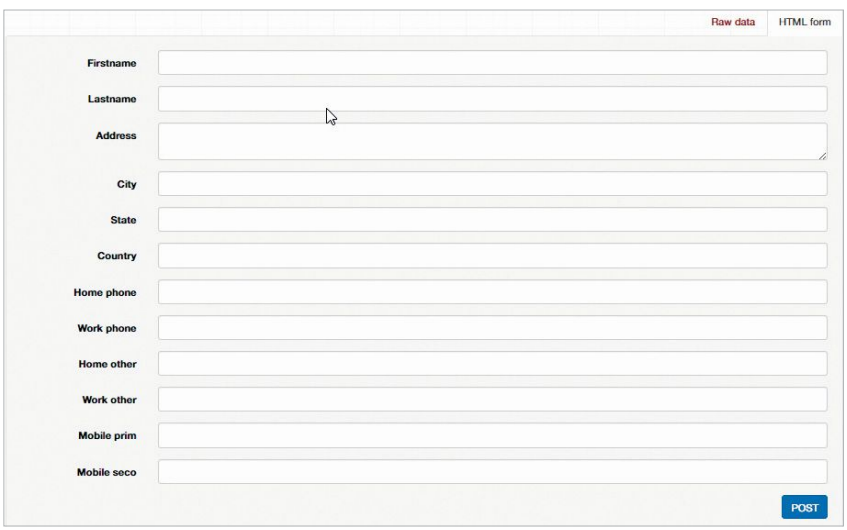


Figure 10: HTML form for *post*

Open a browser with the URL *localhost:8000/contacts/api* and it will display what is shown in Figure 7.

As we have no records, the *get* section displays no data. In the *post* section, add data in JSON format as shown in Figure 8, and press the *Post* button.

Our *views post* portion will be executed and data will be inserted in SQLite3. The page will get reloaded to display the record (Figure 9).

If we are creating a REST API, then *get*, *post*, *put* and *delete* are commonly used. Creating the views with a function for all models with all methods can be a tedious job. Instead of the *function base* view, we can create a *class base* view by inheriting the *generics* view to provide common functionality. Change *views.py* as follows:

```
from rest_framework import status
from rest_framework.decorators import
api_view
from rest_framework.response import
Response
from models import Contact
from serializers import
ContactSerializer
from rest_framework import generics

class ContactList(generics.
ListCreateAPIView):
    queryset = Contact.objects.all()
    serializer_class = ContactSerializer

class ContactDetail(generics.
RetrieveUpdateDestroyAPIView):
    queryset = Contact.objects.all()
    serializer_class = ContactSerializer
```

Change *urls.py* in the *contacts* directory to point to the *class base* view as shown below:

```
from django.conf.urls import include,
url
from django.contrib import admin
from contacts import views

urlpatterns = [
    # Examples:
    # url(r'^$', 'myproject.views.home',
name='home'),
    # url(r'^blog/', include('blog.
urls')),

    url(r'^api/', views.ContactList.
as_view()),

]
```

Refresh the page and you will see that the *post* section has a form like structure. You can use the raw format also, but a new form like structure is more elegant.

A lot more can be done in REST; this article is just to get things rolling. END

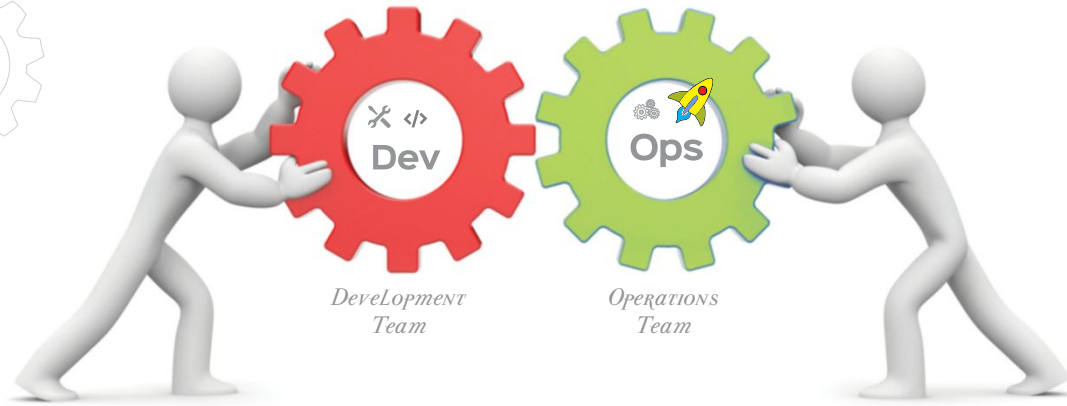
References

[1] *https://www.djangoproject.com/*
[2] *http://www.django-rest-framework.org*

By: Ashish Singh Bhatia

The author is a technology enthusiast and a FOSS fan. He loves to explore new technology, and to work on Python, Java and Android. He can be reached at *ast.bhatia@gmail.com*. He blogs at *https://openfreeidea.wordpress.com/* and *http://etutorialsworld.com/*.

Starting the DevOps Journey Using Cucumber and Selenium



DevOps is a software building process which emphasises communication and collaboration between the teams involved in product management, software development, and operations. Cucumber is a behaviour driven development tool, which when combined with Selenium, a test recording and playback tool, improves the development team's efficiency and productivity.

It is often said that continuous change is the law of the universe and the same is true in the software industry. We have seen a variety of software development models, starting from Waterfall, V and spiral models to the incremental option. All these models have different requirements and guidelines and suit different scenarios. These days, most organisations have embraced the Agile methodology for software development.

The Agile method of developing software and applications focuses on delivering high quality products frequently and consistently, which leads to an increase in business value and profits.

Table 1 lists the differences between the Waterfall and Agile software development approaches.

You can see that the Waterfall model can cause overshooting in time and resources, which can lead to huge losses to the company in terms of profits and user satisfaction. To avoid this, organisations have started adopting the Agile model. There are other reasons too, for choosing this model, some of which are listed below.

- *Client engagement:* In the Agile model, the client is engaged in the software development process at every step — before, during and after the sprint. This helps the development team to understand the client's vision clearly so that defect-free and high quality software can be developed and delivered in less time.
- *Transparency:* Since the client is actively involved in all the sprint activities, ranging from feature prioritisation and planning to reviewing and, finally, to deployment, this

Table 1

Waterfall model	Agile model
Processes are divided into different phases such as design, development, tests, etc.	Here the software development process is divided into sprints, usually spanning a few weeks.
First, the final product to be developed is defined according to the customers' needs and then the different phases begin until a 'finished' product is released.	Small targets are finalised and work is done on them, sprint by sprint. This means the final product is developed in small iterations.
It's difficult to incorporate changes in the requirements since it involves cost and time.	Due to the iterative approach being used in this model, it becomes easy to incorporate changes in the requirements.
Client participation in the development becomes negligible in this process.	Clients or product owners are actively involved in the process and constant feedback is given.

ensures transparency to the client.

- *Timely delivery:* Usually, the duration of a sprint is fixed and varies between one and four weeks, which forces the team to deliver features rapidly and frequently. This also helps product owners to predict the costs involved in the development and keep these under check.
- *Changing requirements:* The Agile methodology also allows teams to incorporate changes in the requirements at an early stage of the development cycle, which helps companies to develop high end products without overshooting their budgets.
- *User focused:* Instead of test cases, the Agile model employs user stories that have business and user focused acceptance criteria. This helps teams to understand the needs of the users and deliver products that can be beta tested in time, so that the necessary changes can be done at the earliest.

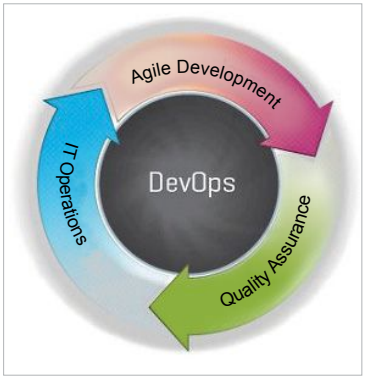


Figure 1: Agile and DevOps complement each other with the support of the QA and IT operations teams

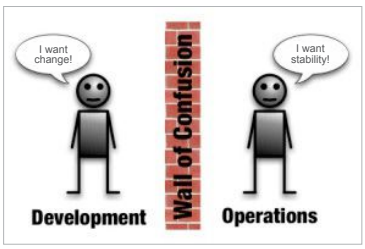


Figure 2: Wall of confusion between the development and operations teams

Steps in the Agile approach

Let's look at the steps involved in implementing the Agile methodology.

1. *Discovery:* To develop a high quality product, one needs to have a clear vision and considerable experience in the technology used in that project. Discovery sessions are significant, since they are the basis for all the upcoming activities in the sprint. During these sessions, the clients' goals, the users' expectations and the business challenges are understood deeply so that no ambiguity remains in the minds of the team, regarding the product.
2. *Product backlog:* The result of successful discovery sessions is product backlog, which contains a list of all the features that need to be developed. These features are then classified on the basis of priority by the product owner (in discussion with the client), so that high priority features can be developed, tested and delivered first.
3. *Iterations:* After the high-level product backlog is finalised along with the priority, sprints are planned and work begins on the features mentioned in the backlog.

Note: Every successive sprint in Agile is both iterative and incremental. It is iterative in the sense that it provides improvements based on the experience gained in the previous sprints, and incremental because it adds new features to the system.

4. *Cycle:* If all the features are completed and tested successfully, then the cycle stops; otherwise, additional sprints are planned to carry out the remaining work.

Agile and DevOps: The connection

Agile and DevOps – these two terms have become the buzzword these days. Though these two words are used interchangeably, there's a stark difference between them. Agile is mainly concerned with software development and the processes or steps involved in it, whereas DevOps comes into the picture after a high quality product has been developed, i.e., it is about the deployment and management of software. The term DevOps is derived from two words – development and operations. Before delving deeper into the details of DevOps, let's see how it emerged in the IT scene.

We have seen how organisations reaped benefits by implementing the Agile methodology, but this model also had some hitches, which are listed below:

- There were chances of incompatibility between old features and new features during integration.
- Often, budget goals and deadlines were missed.
- There was a lack of cooperation between the development and IT operations teams.

Usually, whenever any product is released or any service is made live by an IT organisation, two departments come together to support this release – the development and operations teams. Yet, there is a lack of coordination between development activity and operations activity. The development team feels it is being paid to bring about 'change', whereas the operations team is looking at stability and considers 'change' its enemy. This conflict in mindsets often leads to inefficiency and overhead costs for the company. DevOps is a practice employed to smoothen the IT service delivery by promoting communication between development and operations teams, which is essential to increase a company's productivity. It helps the company to continually deliver software with highly stable features, faster and more frequently.

DevOps brings more flexibility to the Agile methodology and leverages its productivity. It widens the scope of Agile principles by including operations teams in its ambit instead of stopping the Agile cycle at code check-in only. So, you can deduce that Agile principles and processes can be employed as a part of DevOps. In layman's language, we can say that by using the Agile methodology, high-end products are developed and by implementing DevOps, the developed products are deployed in a timely manner. So the Agile model and DevOps complement each other, but are totally different from one another.

The need for DevOps in IT

We have seen how DevOps helps in reducing the friction between the development and operations teams. Now let's see what effects DevOps has if it's integrated into our software development process.

- Bugs can be identified and resolved at an early stage of development due to better communication and collaboration between teams.
- In the Agile model, with DevOps, there is better management of resources since there is less free time due to the presence of cross-functional teams.
- One striking feature of DevOps is the use of version control systems that can reduce the time and effort of the coder.
- Implementation of DevOps also provides an opportunity to dedicate more time and effort to innovation and research.

Behaviour driven development (BDD)

We have just discussed the Agile model and DevOps, and the need to implement these in today's software development scenario. But since DevOps involves various teams such as developers, testers, stakeholders, etc, sometimes there can be a lack of communication between them too. Developers can misinterpret the needs of business stakeholders, and testers can misunderstand the ideas of developers. This can cause a huge negative impact on the overall productivity of the team, affecting actual deliverables. So there is a need for a common language to drive the team and bridge the communication gap. In addition, the following disadvantages were observed in Agile projects:

- Usually, user stories in the Agile model are more focused on the users and their needs rather than on the business logic of the feature. This aspect gets overlooked during sprint planning meetings and can lead to the development of unnecessary features.
- Acceptance criteria, which indicate the completion of a user story, can be understood differently by different individuals.
- Most often, Agile teams adopt the test driven development (TDD) approach, but this approach is very costly.

These weaknesses in the Agile model led to the birth of behaviour driven development (BDD). BDD is also an Agile software development process, which encourages effective communication between the business team, developers, project managers and QA by increasing the focus on business goals and business values. It was conceived by Dan North in 2003 who defines it as follows: "BDD is a second-generation, outside-in, pull-based, multiple-stakeholder, multiple-scale, high-automation, Agile methodology. It describes a cycle of interactions with well-defined outputs, resulting in the delivery of working, tested software that matters."

BDD is an extension of TDD with a few minor differences such as:

1. Tests in BDD are written in plain English.
2. Tests are more behaviour focused and deal with the functionality of the application.
3. BDD uses extensive examples.

BDD has a lot of advantages over the traditional TDD approach. A few of these have been listed below.

- Since BDD involves plain English, it encourages collaboration among different parties involved in the software development cycle. Everyone has a clear understanding of the project and can contribute to planning sessions constructively.
- BDD puts more emphasis on business values and needs, which can help developers in delivering better results because they can understand what the stakeholder wants and work accordingly.
- Due to a single language being used, there is less scope for misunderstanding and misconceptions.
- BDD also leads to better acceptance testing, since the user can also understand the user stories.

BDD testing tools

There are various tools available in the market, which support the BDD framework. Some of these are Cucumber, Specflow, Behave, JBehave, JBehave Web, Lettuce, Behat, Kahlan, etc.

Cucumber: Cucumber is the most widely used open source tool that supports behaviour driven development. It allows you to write application behaviour in a simple, English-like language known as Gherkin. It is written in Ruby but can support various languages like Java, JS, Python, .NET, C++, etc. Due to its wide language support, it can integrate with almost all testing tools and frameworks. You can read more about this tool at *https://cucumber.io/docs*.

Gherkin: Gherkin is a language that Cucumber understands. It is defined on GitHub as, "… a business readable, domain-specific language that lets you describe the software's behaviour without detailing how that behaviour is implemented." Gherkin is easy to learn and understand by non-programmers, but allows illustration of business rules in real-world domains. It has a particular syntax that contains keywords such as *scenario*, *given*, *then*, *and*, *examples*, *but*, etc. A sample Gherkin document looks like what's shown below:

```
Feature: Refund item
Scenario: Jeff returns a faulty microwave
Given Jeff has bought a microwave for $100
     And he has a receipt
     When he returns the microwave
     Then Jeff should be refunded $100
```

So you can see that we have specified the behaviour of the 'Refund item' system using various keywords underlined above in plain English text. You can study Gherkin and its various rules at *https://cucumber.io/docs/reference*.

Configuring Cucumber with Selenium

Now, let's look at how we can integrate Cucumber with Selenium for automated testing in DevOps.

The prerequisites are any IDE (I will be taking Eclipse Neon for tutorial purposes) and the latest Java installed on your system.

Jars required: The following jars/drivers need to be downloaded before starting with the configuration:

- Selenium-server-standalone (latest version)
- Selenium client for Java (latest version available)
- Cobertura-2.1.1 or above
- Cucumber-core-1.2.5 or above
- Cucumber-java-1.2.5 or above
- Cucumber-junit-1.2.5 or above
- Cucumber-jvm-deps-1.0.6 or above
- Cucumber-reporting-3.9.0 or above
- Gherkin-2.12.2
- Junit-3.4.2 or above
- Mockito-all-2.0.2-beta or above
- A driver corresponding to your browser (I will be using the Chrome driver)

After all the required files have been downloaded, follow the steps given below.

1. Launch Eclipse and create a Java project named 'CucumberSelenium' in it.
2. Create three packages under the *src* folder named *cucumberTest*, *resources.Files* and *testScripts*, which will each contain the *runner* file, *test case feature* file and the *step definition* file.

Feature file: This file will contain our test case written in Gherkin language and will have the *feature* extension.

Runner file: Our Cucumber-Selenium framework will not have the *Main* method since we will be using JUnit to run our Java class. So this Java class will be run as a JUnit test to run our script.

Note: Each feature file will have its separate runner class in this framework.

Step definition file: We have designed our *feature* file and *runner* file, but how will Cucumber get to know what code to execute for the specific test step mentioned in the *feature* file? This is taken care of by a separate Java class called *step definition* file.

3. Now create the *CucumberRunner.java* class under the *cucumberTest* package, *testCase.feature* file under the *resources.Files* package and *TestCase.java* class under the *stepDefinition* package.
4. Next, copy the following in your runner class and feature file.

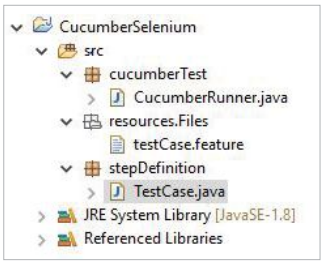


Figure 3: The file hierarchy in the project will look like this

For *CucumberRunner.java*, copy:

```
package cucumberTest;

import org.junit.runner.RunWith;
import cucumber.api.CucumberOptions;
import cucumber.api.junit.Cucumber;

@RunWith(Cucumber.class)
@CucumberOptions(
      features = "src/resources/Files/
testCase.feature",
       glue = {"stepDefinition"},
       tags = {"@cucumber"}
       )

public class CucumberRunner {

}
```

For *testCase.feature*, copy:

```
Feature:To check functionality of google search page

     @cucumber
     Scenario Outline:
       Given <required> browser is opened
          When <url> is opened
          And <keyword> is searched in search box
              Then the first link in search results should be
opened
              And browser is closed

     Examples:
     | required | url                  | keyword  |
| "chrome" | "http://www.google.com"     | "DevOps" |
```

5. Now run your *runner* file as a JUnit test, and you can see the empty methods being auto generated in the console output of Eclipse for each test step in the *feature* file. The significance of these methods is that Cucumber reads test steps from your *feature* file and searches for the corresponding method in the package mentioned in the *glue* option of the *runner* file. The variable values are also passed as arguments to the methods, to make use of them while scripting. So you can see how

```
1 Scenarios (@[33m1 undefined@[0m)
5 Steps (@[33m5 undefined@[0m)
0m0.000s


You can implement missing steps with the snippets below:

@Given("^\"([^\"]*)\" browser is opened$")
public void browser_is_opened(String arg1) throws Throwable {
    // Write code here that turns the phrase above into concrete actions
    throw new PendingException();
}
```

Figure 4: Methods' skeleton being generated by Cucumber by parsing the *feature* file

beautifully Cucumber has taken care of every minute detail while designing this framework.

6. Now copy the following code in the *step definition* file *TestCase.java*:

```
package stepDefinition;

import java.util.concurrent.TimeUnit;
import org.openqa.selenium.By;
import org.openqa.selenium.WebDriver;
import org.openqa.selenium.chrome.ChromeDriver;
import cucumber.api.java.en.Given;
import cucumber.api.java.en.Then;
import cucumber.api.java.en.When;

public class TestCase {

     WebDriver driver = null;

     @Given("^\"([^\"]*)\" browser is opened$")
     public void browser_is_opened(String arg1) throws
Throwable {
       if(arg1.equals("chrome")) {
           System.setProperty("webdriver.chrome.driver",
"D:\\Selenium\\chromedriver.exe"); // file path of driver
where it is stored.
           driver = new ChromeDriver();
           driver.manage().window().maximize();
           driver.manage().timeouts().
implicitlyWait(30,TimeUnit.SECONDS);
       }
     }
     @When("^\"([^\"]*)\" is opened$")
     public void is_opened(String arg1) throws Throwable {
       driver.get(arg1);
     }
     @When("^\"([^\"]*)\" is searched in search box$")
     public void is_searched_in_search_box(String arg1)
throws Throwable {
       driver.findElement(By.id("lst-ib")).sendKeys(arg1);
       driver.findElement(By.xpath("//*[@id='tsf']/div[2]/
div[3]/center/input[1]")).click();

     }

     @Then("^the first link in search results should be
opened$")
     public void the_first_link_in_search_results_should_be_
opened() throws Throwable {
       driver.findElement(By.xpath(".//*[@id='rso']/div[3]/
div/div[1]/div/div/h3/a")).click();
     }

     @Then("^browser is closed$")
     public void browser_is_closed() throws Throwable {
         driver.close();
     }
}
```

7. Since we have defined the code, run the *runner* file and the corresponding steps will be performed according to the defined test case.
8. Once execution is complete, you can see the execution status of the steps and scenarios in the console as shown in Figure 5.

```
Starting ChromeDriver 2.31.488763
Only local connections are allowed.
Aug 08, 2017 8:21:40 PM org.openqa.
INFO: Detected dialect: OSS

1 Scenarios (2[32m1 passed2[0m)
5 Steps (2[32m5 passed2[0m)
0m13.935s
```

Figure 5: Output of the script depicting one scenario and five steps being passed

So you can see how easily Cucumber can be configured with Selenium Web driver to implement the BDD framework. Using this framework, you can start your DevOps journey in the testing field, within your organisation.

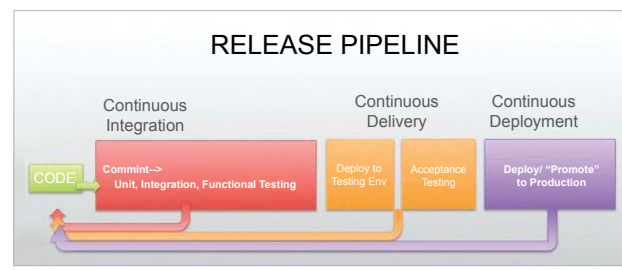


Figure 6: Complete release pipeline with CI/CD tools implemented (*Image courtesy:* Pinterest)

CI/CD tools and DevOps

This framework can also be integrated smoothly with continuous integration and continuous delivery (CI/CD) tools like Jenkins, TeamCity, Bamboo, etc, so that automated tests can be run every time developers check their code into a central repository; reports can then be published to the required stakeholders as and when required.

So we have discussed the shift from the Waterfall model to the Agile model as well as the simultaneous implementation of DevOps and the Agile methodology. Try this BDD inspired framework using Selenium to leverage your team's productivity and efficiency. END

By: Vinayak Vaid

The author works as an automation engineer at Infosys Limited, Pune. He has worked on different testing technologies and automation tools like QTP, Selenium and Coded UI. He can be contacted at *vinayakvaid91@gmail.com*.

Tracing Vehicular Mobility in ns-3

ns-3 is free simulation software that's meant for educational, research and development purposes. It has a number of high quality, tested and validated models, and is backed by a large community of developers and users.

The ns-3 simulator is an open source project meant for educational and development purposes. It is also used to enact discrete-event simulation. Practical implementation or experimental testing of new protocols on real-world networks is not always possible. ns-3 helps to abstract real-world entities and allows to simulate, as well as to compute complex calculations on huge networks of computer systems. It can be treated as a real-time network emulator, as it can simulate scenarios on realistic data. It provides a platform for developers to build simulation models and can comprehend every single step, i.e., the entire workflow of the simulation, from configuration to the collection and analysis of traces.

The ns-3 project includes and supports various real-world protocols and its implementations. It supports various routing protocols, which include OLSR, AODV and many more for IP-based applications on IP networks. It also supports non-IP based network simulation. The ns-3 simulator majorly involves models for Wi-Fi, WiMax and LTE.

Mobility provision in the ns-3 simulator

The ns-3 simulator provides support for mobility in the form of mobility models and helper classes. Mobility models are used to track and perpetuate the current position and speed of an object. Mobility helper classes are used to place nodes and set up mobility models. Users communicate with the mobility system using the mobility helper classes as they integrate a mobility model and position allocator, and can be used along with a node container to install mobility functions on a set of nodes. The ns-3 simulator provides a ns-2 mobility helper class, which allows it to interpret the ns-2 file format of the trace file. The following statements show the ns-2 trace file format:

```
$node_ (node number) set X_ x
$node_ (node number) set Y_ y
$node_ (node number) set Z_ z
$ns_ at $time $node_ (node number) setdest x2 y2 speed
$ns_ at $time $node_ (node number) setdest x3 y3 speed
$ns_ at $time $node_ (node number) setdest x4 y4 speed
$ns_ at $time $node_ (node number) setdest x5 y5 speed
```

In the above traces, *x* and *y* are initial positions while *z* specifies the speed in m/s. These are initialised using the *set* statements. The command *setdest* is used to define the new position (*x2*, *y2*) of the node at the given time (*$time*), which

thus instructs the node to propagate to its new position. If the node reaches its destination, it will pause there; or it may also happen that the simulation ends before. In that case, the node will not be able to reach the destination. Also, sometimes a new destination is set during a course to a location; in that scenario, the node will change course to the new destination in between.

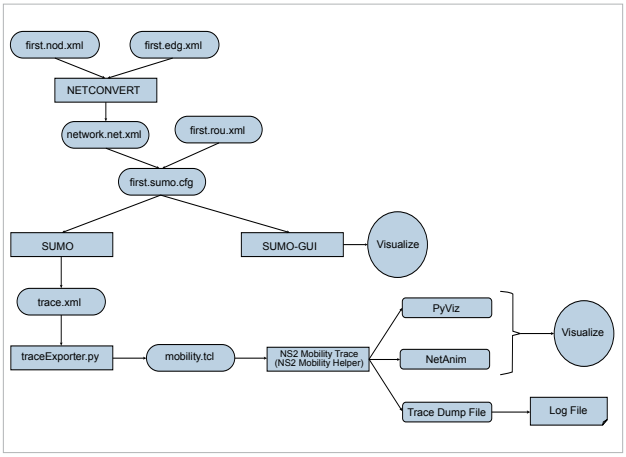


Figure 1: Hierarchy of the integration of SUMO files with ns-3

SUMO

Simulation of Urban Mobility (SUMO) is an open source traffic simulation package that includes net-import and demand modelling components. Many research topics, like the route choice and the simulation of vehicular communication or traffic light algorithms, are studied with the help of SUMO. The framework of SUMO serves in various projects like traffic management strategies or to simulate automatic driving. Simulations in SUMO comprise time-discrete and space-continuous vehicle movements, multi-lane highways and streets, traffic lights, Open GL for the graphical interface and fast execution speed. SUMO is edge-based, can operate with other applications at run time, has portability and detector-based outputs.

The flow of the integration of SUMO traces with ns-3 for mobility provisioning is shown in Figure 1.

Constructing a scenario in SUMO

In order to generate traffic with the help of the SUMO simulator, a scenario or network needs to be created. SUMO Street Network consists of nodes (junctions) and edges (streets connecting the junctions). SUMO Simulator requires *.nod.xml* and *.edg.xml* to define the junctions and streets joining them.

The *nod.xml* file contains the location (*x* and *y* coordinates) of junctions. An example is given below:

```
first.nod.xml

<nodes>
    <node id="1" x="-300.0" y="5.0" />
    <node id="2" x="+300.0" y="5.0" />
    <node id="3" x="5.0" y="-300.0" />
    <node id="4" x="5.0" y="+300.0" />
    <node id="5" x="5.0" y="5.0" />
</nodes>
```

To join the above nodes, edges are defined in the *edg.xml* file using the target node ID and source node ID. An example is given below:

```
first.edg.xml

<edges>
    <edge from="1" id="A" to="4" />
    <edge from="4" id="B" to="2" />
    <edge from="2" id="C" to="5" />
    <edge from="5" id="D" to="3" />
    <edge from="2" id="E" to="3" />
</edges>
```

To build a network, the above defined node file and edge file are required. And using the *netconvert* utility, a *net.xml* file is generated as follows:

```
$netconvert -n first.nod.xml -e first.edge.xml -o network.net.xml
```

Vehicles in the traffic are defined with route data in the *rou.xml* file. For example:

```
first.rou.xml

<routes>
<vType id="BoogyA" length="5.75" maxSpeed="90.0" sigma="0.4" />
<vType id="BoogyB" length="7.5" maxSpeed="60.0" sigma="0.7" />
  <route id="rou01" edges="A B C D"/>
  <route id="rou02" edges="A B E"/>
  <vehicle depart="0" id="v0" route="rou01" type="BoogyA"
color="1,0,0" />
  <vehicle depart="1" id="v1" route="rou02" type="BoogyA" />
  <vehicle depart="2" id="v2" route="rou01" type="BoogyA" />
  <vehicle depart="3" id="v3" route="rou02" type="BoogyA" />
  <vehicle depart="5" id="v4" route="rou01" type="BoogyA" />
  <vehicle depart="6" id="v5" route="rou02" type="BoogyA" />
  <vehicle depart="7" id="v6" route="rou01" type="BoogyA" />
  <vehicle depart="8" id="v7" route="rou02" type="BoogyA" />
  <vehicle depart="9" id="v8" route="rou01" type="BoogyA" />
  <vehicle depart="11" id="v9" route="rou02" type="BoogyA" />
  <vehicle depart="14" id="v10" route="rou01" type="BoogyA"/>
  <vehicle depart="16" id="v11" route="rou01" type="BoogyA"/>
  <vehicle depart="17" id="v12" route="rou01" type="BoogyB"
color="1,0,0" />
  <vehicle depart="18" id="v13" route="rou02" type="BoogyB" />
  <vehicle depart="19" id="v14" route="rou02" type="BoogyB" />
```

```
<vehicle depart="21" id="v15" route="rou01"
type="BoogyB" />
<vehicle depart="22" id="v16" route="rou01"
type="BoogyB" />
<vehicle depart="23" id="v17" route="rou02"
type="BoogyB" />
<vehicle depart="24" id="v18" route="rou02"
type="BoogyB" />
<vehicle depart="27" id="v19" route="rou02"
type="BoogyB" />
</routes>
```

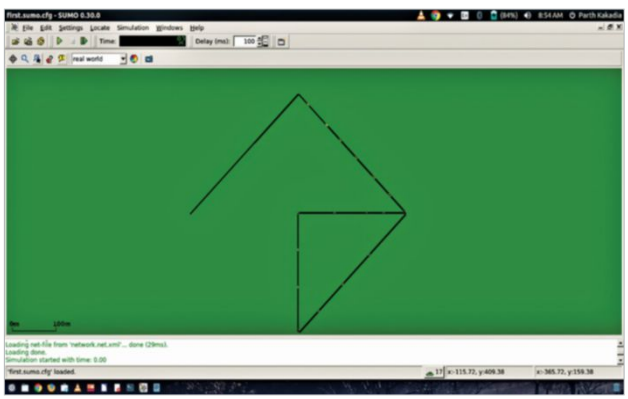


Figure 2: Simulation in the SUMO GUI

Files *first.rou.xml* and *network.net.xml* are required to generate configuration files for simulating the traffic on the network. The following configuration file is generated, which is then used to simulate the scenario:

```
first.sumo.cfg

<configuration>
    <input>
        <net-file value="network.net.xml"/>
        <route-files value="first.rou.xml"/>
    </input>
    <time>
        <begin value="0"/>
        <end value="150"/>
    </time>
    <time-to-teleport value="-1"/>
</configuration>
```

To simulate the scenario, the following command is required to be fired from the terminal:

```
$sumo -c first.sumo.cfg
```

To visualise the simulation in the SUMO GUI, use the following command:

```
$sumo-gui -c first.sumo.cfg
```

Generating traces

SUMO trace files can be generated with the help of a *.cfg* file (configuration file), as follows:

```
$sumo -c first.sumo.cfg –fcd-output trace.xml
```

Traces of the vehicles' position are dumped into the XML file. These SUMO traces can be used to provide mobility to nodes in the ns-3 simulator. The ns-3 simulator provides the ns-2 mobility helper class, which can be used to read the movements of nodes from trace files. To make use of the SUMO traces generated from the SUMO simulator, the XML file needs to be converted to a TCL file, which then is provided to the ns-2 mobility helper. SUMO provides tools like *traceExporter.py* and *traceExporter.jar* to convert SUMO trace to different file formats.

```
$python $SUMO_HOME/tools/traceExporter.py –fcd-input trace.
xml –ns2mobility-output mobility.tcl
```

mobility.tcl generated above contains the position and the velocities of the nodes (vehicles).

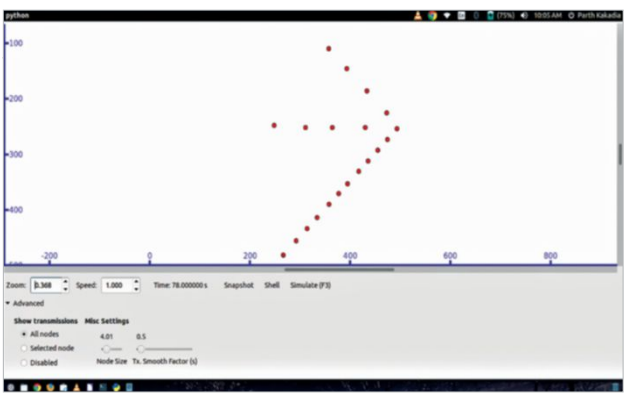


Figure 3: Simulation in pyViz

mobility.tcl as the argument along with a number of nodes in the scenario and the duration of the simulation. At the end of the simulation, it generates a log file. The command for executing the same is given below:

```
$./waf --run "ns2-mobility-trace --traceFile=path_to_
tracefile/mobility.tcl --nodeNum=18 --duration=200.0
--logFile=ns2-mob.log" --vis
```

--vis renders the simulation in Python Visualizer.

Visualising the scenario in NetAnim: A NetAnim compatible XML file can be generated by adding the following lines of code in the file *ns2-mobility-trace.cc (ns-dev/src/mobility/examples/)*.

Include the header file:

```
#include "ns3/netanim-module.h"
```

And add the following line in the program:

```
AnimationInterface anim ("ns2mobilityanimation.xml");
```

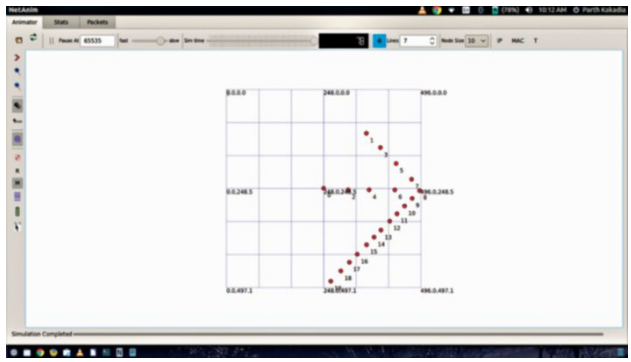


Figure 4: Visualisation in NetAnim

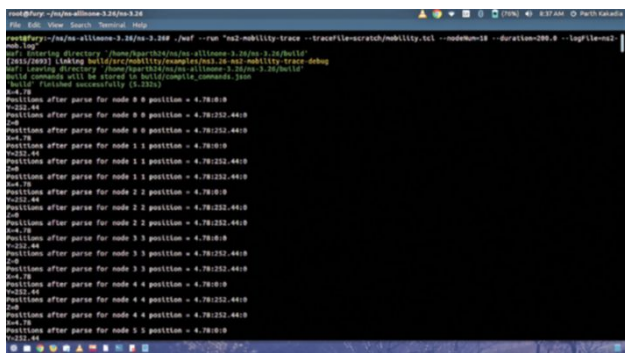


Figure 5: Trace dump in the log file

To visualise the scenario, open *ns2mobilityanimation.xml* in NetAnim.

Generating traces in the form of a file: By executing the *ns2-mobility-trace.cc* as shown above, the *ns2-mob.log* file is generated. END

References

[1] https://www.nsnam.org
[2] https://www.sumo.dlr.de
The source repository can be downloaded from:
https://bitbucket.org/yashsquare/ns3_support

Acknowledgment

We thank Dr Manish Chaturvedi, assistant professor (computer science engineering), Pandit Din Dayal University, Gandhinagar, for providing insights on certain technicalities in this article.

By: Parth Kakadia and Jitendra Bhatia

Parth Kakadia is a FOSS enthusiast. He can be reached at *parthkakadia143@gmail.com*.

Jitendra Bhatia works as assistant professor at Vishwakarma Government Engineering College. You can contact him at *jitendrabbhatia@gmail.com*.

Understanding How a Neural Network Works Using R

In the simplest of terms, a neural network is a computer system modelled on the human nervous system. It is widely used in machine learning technology. R is an open source programming language, which is mostly used by statisticians and data miners. It greatly supports machine learning, for which it has many packages.

The neural network is the most widely used machine learning technology available today. Its algorithm mimics the functioning of the human brain to train a computational network to identify the inherent patterns of the data under investigation. There are several variations of this computational network to process data, but the most common is the *feedforward-backpropagation* configuration. Many tools are available for its implementation, but most of them are expensive and proprietary. There are at least 30 different packages of open source neural network software available, and of them, R, with its rich neural network packages, is much ahead.

R provides this machine learning environment under a strong programming platform, which not only provides the supporting computation paradigm but also offers enormous flexibility on related data processing. The open source version of R and the supporting neural network packages are very easy to install and also comparatively simple to learn. In this article, I will demonstrate machine learning using a neural network to solve quadratic equation problems. I have chosen a simple problem as an example, to help you learn machine learning concepts and understand the training procedure of a neural network. Machine learning is widely used in many areas, ranging from the diagnosis of diseases to weather forecasting. You can also experiment with any novel example, which you feel can be interesting to solve using a neural network.

Quadratic equations

The example I have chosen shows how to train a neural network model to solve a set of quadratic equations. The general form of a quadratic equation is: $ax^2 + bx + c = 0$. At the outset, let us consider three sets of coefficients a, b and c and calculate the corresponding roots r1 and r2.

The coefficients are processed to eliminate the linear equations with negative values of the discriminant, i.e., when $b2 – 4ac < 0$. A neural network is then trained with these data sets. Coefficients and roots are numerical vectors, and they have been converted to the data frame for further operations.

The example of training data sets consists of three coefficients with 10 values each:

```
aa<-c(1, 1,  -3,  1, -5, 2,2,2,1,1)
bb<-c(5, -3,  -1,  10, 7, -7,1,1,-4,-25)
cc<-c(6, -10, -1,  24, 9, 3,-4,4,-21,156)
```

Data preprocessing

To discard equations with zero as coefficients of x^2, use the following code:

```
k <- which(aa != 0)
aa <-aa[k]
bb <-bb[k]
cc <-cc[k]
```

To accept only those coefficients for which the discriminant is zero or more, use the code given below:

```
disc <-(bb*bb-4*aa*cc)
k <- which(disc >= 0)
aa <-aa[k]
bb <-bb[k]
cc <-cc[k]
a <-  as.data.frame(aa)                # a,b,c vectors are
converted to data frame
b <-  as.data.frame(bb)
c <-  as.data.frame(cc)
```

Calculate the roots of valid equations using conventional formulae, for training and verification of the machine's results at a later stage.

```
r1 <- (-b + sqrt(b*b-4*a*c))/(2*a)  # r1 and r2 roots of each
equations
r2 <- (-b - sqrt(b*b-4*a*c))/(2*a)
```

After getting all the coefficients and roots of the equations, concatenate them columnwise to form the input-output data sets of a neural network.

```
trainingdata <- cbind(a,b,c,r1,r2)
```

Since this is a simple problem, the network is configured with three nodes in the input-layer, one hidden-layer with seven nodes and a two-node output-layer.

R function *neuralnet()* requires input-output data in a proper format. The format of formulation procedure is somewhat tricky and requires attention. The right hand side of the formula consists of two roots and the left side includes three coefficients a, b and c. The inclusion is represented by + signs.

```
colnames(trainingdata) <- c("a","b","c","r1","r2")
```

Equations	Coefficients			Roots		Predicted roots (neural network outputs)	
	a	b	c	r1	r2	pre-r1	pre-r2
1	1	5	6	-2	-3.0	-1.9999879917	-3.000010276
2	1	-2	-8	4	-2.0	3.2248691572	-3.467243981
3	2	1	-3	1	-1.5	0.9677422405	-1.716275669

Table 1: Outputs of *print()* in tabular form

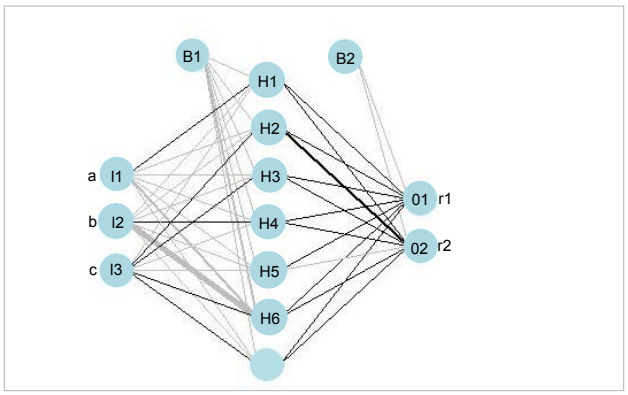


Figure 1: Neural network model of the quadratic equation solver

```
net.quadroot <- neuralnet(r1+r2~a+b+c, trainingdata,
hidden=7, threshold=0.0001)
```

An arbitrary performance threshold value 10^{-4} is taken and this can be adjusted as per requirements.

The configuration of the just-constructed model with its training data can be visualised with a plot function.

```
#Plot the neural network
plot(net.quadroot)
```

Now it is time to verify the neural net with a set of unknown data. Arbitrarily take a few sets of values of three coefficients for the corresponding quadratic equations, and arrange them as data frames as shown below.

```
x1<-c(1, 1, 1,  -2,  1, 2)
x2<-c(5, 4, -2,  -1,  9, 1)
x3<-c(6, 5, -8, -2,  22, -3)
```

Since there is no zero coefficient corresponding to x^2, we can accept only those coefficients for which the discriminant is zero or more than zero.

```
disc <-(x2*x2-4*x1*x3)
k <- which(disc >= 0)
x1 <-x1[k]
x2 <-x2[k]
x3 <-x3[k]
y1=as.data.frame(x1)
y2=as.data.frame(x2)
y3=as.data.frame(x3)
```

The values are then fed to the just-configured neural model *net.quadroot* to predict their roots. The predicted roots are collected into *net.result$net.result* and can be displayed with the *print()* function.

```
testdata <- cbind(y1,y2,y3)
net.results <- compute(net.quadroot, testdata)
#Lets see the results
print(net.results$net.result)
```

Now, how does one verify the results? To do this, let us compute the roots using the conventional root calculation formula, and verify the results by comparing the predicted values with them.

Calculate the roots and concatenate them into a data frame.

```
calr1 <- (-y2 + sqrt(y2*y2-4*y1*y3))/(2*y1)
calr2 <- (-y2 - sqrt(y2*y2-4*y1*y3))/(2*y1)
r<-cbind(calr1,calr2)     #Calculated roots using formula
```

Then combine the test data, its roots and the predicted roots into a data frame for a decent tabular display for verification (Table 1).

```
#Combine Inputs, Expected roots and predicted roots.
comboutput <- cbind(testdata,r,net.results$net.result)
#Put some appropriate column heading
colnames(comboutput) <- c("a","b","c","r1","r2","pre-
r1","pre-r2")
print(comboutput)
```

It is clear from the above outputs that our neural network has learnt appropriately and produces an almost correct result. You may need to run the neural network several times with the given parameters to achieve a correct result. But if you are lucky, you may get the right result in the first attempt itself! END

By: Dipankar Ray

The author is a member of IEEE, IET, with more than 20 years of experience in open source versions of UNIX operating systems and Sun Solaris. He is presently working on data analysis and machine learning using a neural network and different statistical tools. He has also jointly authored a textbook called 'MATLAB for Engineering and Science'. He can be reached at *dipankarray@ieee.org*.

Snappy Ubuntu Core for Embedded and IoT Devices

Ubuntu Core is a minimalistic version of Ubuntu. It is lightweight and is designed for use with embedded systems and IoT devices. Snaps are universal Linux packages, which are faster to install, easier to create, and safe to run and work on multiple distributions.

The biggest challenge in delivering Linux applications is dependency resolution. As distributions update underlying libraries frequently, it is not always possible to offer application binaries as offline archives. One can't think of downgrading an application if underlying packages are upgraded, which are in turn required by some more applications. Also, it's not possible to keep multiple versions of the same library in view of different applications. Ubuntu Core has a solution for all this complexity, whereby a single device can host multiple isolated applications with their own dependencies.

Ubuntu 'snaps' are universal packages supported on many Linux distributions and on various devices. Any Linux application can be bundled as a snap, which can be easily deployed across the distributions and across devices with isolated environments. This ensures secure execution and better transactional management with flexible upgradable or downgradable support.

Snaps are based on Ubuntu Core, which is a minimal rendition of Ubuntu, specifically designed for embedded and IoT devices. The core image consists of the kernel, a minimal set of drivers and a basic file system in the form of a core snap with basic functionality. You can't update the core by adding packages unlike other typical Linux distributions. Each snap is bundled with its own dependencies for the functionality not provided by the core. It supports transactional updates of the snaps, which can be upgraded or downgraded easily without affecting other snaps. Also, snaps can be remotely upgraded.

The snap format file system

When Ubuntu Core is installed, the OS snap gets installed initially. This snap is packed with *rootfs*, providing basic features like network services, standard libraries and daemon management. It is the platform for subsequent snaps offering application logic. This enables universal packaging as all snaps run on the basis of the OS snap provided by the core, eliminating any dependencies with the host environment.

Each snap comes with an isolated file system in *squashfs* format, which gets mounted in a sub-directory under */snap*. For example, when *foo* snap is installed, its file system is mounted under */snap/foo/xx*, where *xx* represents a revision of the snap. There can be multiple

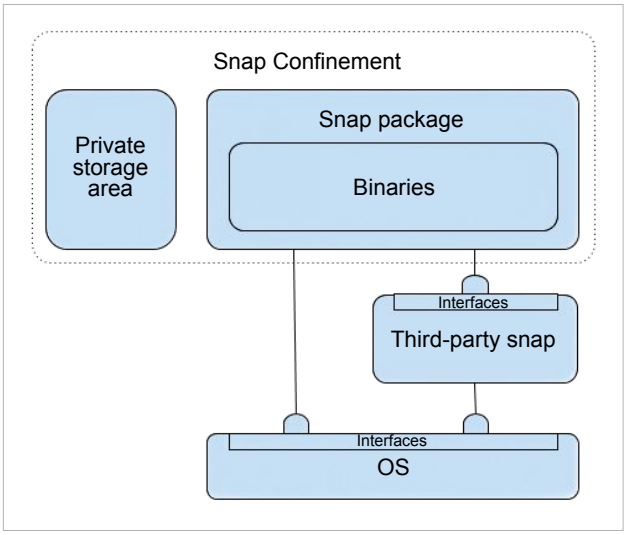


Figure 1: The snap execution environment

revisions of file systems under */snap/foo* but */snap/foo/current* always points to the desired version as per the recent upgrade or downgrade. Each user-accessible application of the snap is linked under */snap/bin*, with the <*snap-name*>.<*app-name*> convention, e.g., */snap/bin/foo.test* for an application named *test* under the snap *foo*. The meta data of the snap is available in the form of *meta/snap.yaml* under the mounted file system.

Use cases

As the underlying stack is read only and each snap has a self-contained file system, Ubuntu Core is the perfect choice for industrial solutions for which security is a critical concern. Here are some cases of Ubuntu Core powered solutions:

- IoT gateways for industrial solutions, like the Dell Edge gateway
- DeviceHive IoT toolkit
- Cloud based servers like NextCloud
- Digital signage
- Robots and drones

Some popular snaps

Here are some popular applications available from the snap store:

- Atom, a text editor
- NextCloud, a cloud based server for data storage
- VLC, a multi-media player
- Docker, a container manager
- Cassandra, a scalable NoSQL database
- Blender, a 3D creation suite
- Stellarium, a planetarium software
- *Rocket.Chat*, a Web chat server
- The Telegram desktop client
- Jenkins, a leading automation server for building, testing and deployments
- OpenHAB, a home automation middleware software
- Mosquitto, a MQTT broker
- Clients for AWS, like Azure cloud solutions

Installation

Ubuntu Core comes with a snap manager, *snapd*, which is present by default in Ubuntu Xenial (16.04 LTS) onwards; for older versions of Ubuntu and other distributions you can install the snap manager as follows:

```
apt-get install snapd
```

If you are planning to build custom snaps, install snapcraft also, as follows:

```
apt-get install snapcraft
```

Replace *apt-get* with any other package manager like *yum*, *dnf*, *zypper*, *pacman*, *poky*, etc, for non-Debian based distributions.

Ubuntu Core is available for dedicated installation for various embedded targets like Raspberry Pi Models 2 and 3, Compute Module 3, Intel Joulie, Dragon Board, Intel NUC kit, Samsung Artik, etc. Also, Ubuntu Core can be installed through KVM on any desktop as a virtual machine, for which you can download the core image for the AMD 64 architecture, extract the archive and run the following command:

```
qemu-sytem-x86_64 -smp 2 -m 1500 -netdev user, id=mynet0,
hostfwd=tcp::8022-:22,hostfwd=tcp::8090-:80 -device virtio-
net-pci,netdev=mynet0 -drive file=ubuntu-core-16-amd64.
img,format=raw
```

Now log in to the core using ssh as follows:

```
ssh username@localhost -p 8022
```

Working with snaps

Let's try working with snaps in CLI mode, assuming *osfy* as the snap's name.

To search for available snaps based on the *osfy* keyword, use the following command:

```
snap find osfy
```

To install a specific snap from the store, use the command given below:

```
snap install osfy
```

Observe the file hierarchy in */snap/osfy/current*. Run the application associated with snap by using

the following command:

```
osfy.test          #from /snap/bin
```

For listing of installed snaps, use the command given below:

```
snap list
```

To check the information about a particular installed snap, use the following command:

```
snap info osfy
```

The following command updates all installed snaps:

```
\sudo snap refresh
```

The following command updates a particular installed snap:

```
sudo snap refresh osfy
```

To roll back a snap to a previous version, use the command given below:

```
sudo snap revert  osfy
```

The next command uninstalls a specific snap:

```
sudo snap remove osfy
```

The following command removes symlinks of the apps in */snap/bin* and stops the associated services:

```
sudo snap disable osfy
```

To start the snap services and make symlinks for the apps available in */snap/bin*, use the following command:

```
sudo snap enable osfy
```

To list the environment variables associated with a snap, use the command given below:

```
sudo snap run osfy.env
```

Classic snap

Since the core is read-only and not meant for development purposes, you can install 'classic snap' to enter in classic mode on devices deployed with solely Ubuntu Core. Normal package management with apt-get is allowed in classic mode. But avoid this classic snap if secure deployment is the prime concern.

```
sudo snap install --devmode classic
sudo classic
```

Now, in the prompt enabled by classic mode, you can try any developer commands such as those shown below:

```
sudo apt-get update
sudo apt-get install vim gcc snapcraft
```

Interfaces

Interfaces allow snaps to exchange services with each other. A snap provides the service in the form of *slot* and another snap can consume it in the form of *plug*. Some important interfaces are *network*, *network-bind*, *camera*, *bluetooth-control*, *firewall-control*, *log-observe*, etc. To list out all possible interfaces and consumer snaps plugged into each of them, use the following command:

```
snap interfaces
```

To list interfaces plugged by a particular snap, the following command should be used:

```
snap interfaces osfy
```

In the absence of a *plugs* entry in the *apps* section of the manifest file, we need to connect the interfaces manually. For example, if a *foo* snap is providing a *bar* interface slot and *osfy* snap wants to plug into it, run the following command:

```
snap connect osfy:bar foo:bar
```

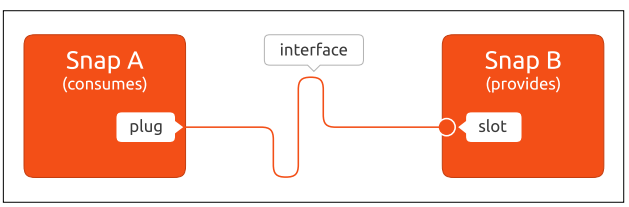


Figure 2: Snap interfaces

Building custom snaps

You can package any Linux application as a snap. For this, create an empty directory and run the following command to create the basic skeleton of metadata in the name of *snapcraft.yaml:*

```
snapcraft init
```

Now edit the *parts*, *apps* sections of *snapcraft.yaml* as per the tutorial in *https://snapcraft.io/docs/build-snaps/your-first-snap*. And, finally, run the *snapcraft* command to build the snap by retrieving all required sources, building each part, and staging and priming the outcome of all parts in stage, prime directories. Finally, the contents of the prime sub-directory will be bundled

with the snap file system.

```
snapcraft
```

The following command installs the generated local snap:

```
snap install hello_x.xx_xxx.snap
```

Plugins

Plugins help snapcraft in building parts of the snap. For listing all available plugins, use the command given below:

```
snapcraft list-plugins
```

Here are some standard plugins for building snaps:

- autotools - to build code based on *./configure, make, make install, cycle*
- *make, cmake* - to build *Makefile, CMAke*
- *qmake* - to build Qt based projects
- *maven,gradle* - to build Java code
- *python2, python3* - to build Python code or install packages via Pip
- *nodejs* - to build node.js code or install npm packages
- *dump* - this just dumps the contents from the snapcraft directory to the snap file system via staging, prime directories

To get more information on each plugin, use the following command:

```
snapcraft help xxxx
```

Some snaps are not available in stores but can be installed as custom snaps, by building with snapcraft using suitable plugins. Let's consider a few case studies on building custom snaps.

Node-RED is a visual prototyping tool for wiring services and components together. It can be easily installed through npm as a node.js package. In the *parts* section of *snapcraft.yaml* we'll specify the *node.js* plugin and specify each npm package under node-packages. Here is some code for the *apps* and *parts* of the manifest file:

```
apps:
  red:
    daemon: simple
    command: bin/nodered
    plugs:
      - network-bind
      - network
parts:
  red:
    plugin: nodejs
    node-packages:
      - node-red
```

We can specify other *parts* and *apps* for any Node-RED add-ons required in this snap (Reference: *github.com/muka/nodered.snap)*.

The InfluxData TICK Stack comes with Telegraf, InfluxDb, Chronograf and Kapacitor for data collection, storage, visualisation and analysis with time series data management support. Since binary tar balls are available for all these four components, the dump plugin can be used to build a snap. Here is the code snippet of *snapcraft.yaml* (Reference: *github.com/morphis/influxdb-snap)*:

```
apps:
influxd:
    daemon: simple
    command: usr/bin/influxd
    plugs :
        - network-bind
    parts:
      influxdb:
         plugin: dump
         source: https://dl.influxdata.com/influxdb/releases/
influxdb-<ver>_linux_<arch>.tar.gz
```

IoT gateways like Eclipse Kura as well as servers like Eclipse Leshan and Californium can be packed as snaps using the Maven plugin to build the *parts* and suitable *apps* sections to launch the generated jar files.

Please stay tuned to *github.com/rajeshsola/iot-examples/tree/master/snappy-examples for customized snapcraft.yaml* for the above snaps.

Publishing your snaps

Custom snaps built by you can be published to the store. Just create an account on *dashboard.snapcraft.io*, log in, create a namespace and publish the snaps with the following commands:

```
snapcraft login
snapcraft register osfy-demos
snapcraft push hello_x.xx_xxx.snap
```

END

References

[1] *www.ubuntu.com/core*
[2] *www.ubuntu.com/desktop/snappy*
[3] *https://developer.ubuntu.com/core*
[4] *https://snapcraft.io/docs/reference/plugins/*
[5] *https://snapcraft.io/docs/core/interfaces*

By: Rajesh Sola

The author is a faculty member of C-DAC's Advanced Computing Training School, and an evangelist in the embedded systems and IoT domains. You can reach him at *rajeshsola@gmail.com*.

37TH **GITEX** TECHNOLOGY WEEK

08 - 12 OCT 2017
DUBAI WORLD TRADE CENTRE

@gitextechweek
gitex.com

Converging technologies, disruptive business models and moonshots collide across **60 real-world case studies** from **80+ disruptive global practitioners**

EXCITE. ENGAGE. EXPERIENCE.

HOW? Eight dedicated power conferences at this year's GITEX

MON/TUE 9-10 — SMART CITIES (NEW FOR 2017)

Big data, happiness, AI, Cognitive cities

EXPERTS INCLUDE

Siim Sikkut, Government CIO (Estonia)
Taking on Bitcoin & Etherium with an official virtual currency for Estonia and the EU

MON 9 — HEALTHCARE

Robotics, 3D Printing, AR/VR, Patient Experience, Cyber-security

EXPERTS INCLUDE

Shafi Ahmed, Associate Dean, Barts & London Medical School (UK)
World's first AR/VR livestream surgeon (Google Glass, Snapchat, Twitter) / curator of NHS Twitter account

MON 9 — RETAIL

Digital transformation, customisation, AR/VR, drones, magic mirrors

EXPERTS INCLUDE

Scott Emmons, Head of the Innovation Lab, Neiman Marcus (USA)
Retail business model innovator through AI and big data

TUES 10 — FINANCE (NEW FOR 2017)

Blockchain, AI, robotics, automation, machine learning

EXPERTS INCLUDE

Raj Chowdhury, Previous Head of Innovation, ICICI Bank (India)
World's first to execute inter-bank blockchain transfer

+ TECH LEADERS

WED 11 — DIGITAL MARKETING

Gamification, data analytics, AI, digitalisation, automation

EXPERTS INCLUDE

Patrick Grove, Co-founder and Group CEO, Catcha Group/ Co-founder & Chairman, iflix (Malaysia)
Financial review Richlister overtaking Netflix in Asia

WED 11 — TRANSPORT & LOGISTICS (NEW FOR 2017)

Autonomous vehicles, robotics, AI, smart technology

EXPERTS INCLUDE

Tin Hang Liu, CEO & Founder, OSVehicle (USA)
Saved mobility sector millions of $$$ creating an open-source, modular autonomous vehicle platform

+ THURS 12 EDUCATION

Game-changing advantage from the full GITEX experience
– grow your network and source technology as you gain critical insight

Over 120 hours of high-level networking across 8 unique options

NEW FOR 2017

Invite-only executive briefings daily

HOSTED BY
SAP, Google, Oracle, Wipro, Dark Matter, Microstrategy, VMWare, F5 and Enterprise Ireland

Find out how you can apply - Email: Fahad.Khalife@dwtc.com

CO-LOCATED WITH

GITEX FUTURE STARS
Startups, Dream Makers, Rule Breakers & Game Changers.

WITH TODAY'S FAST CHANGING TECHNOLOGY, IT'S EITHER EVOLVE OR BE EXTINCT.
Get your passes before they run out at www.gitex.com or call +971 4 308 6805

Organised by

Cyber Security Innovation Partner

Exclusive ICT Transformation Partner

VIP Majlis Lounge Partner

Business Transformation Partner

Exclusive Digital Transformation Partner

Autonomous & Robotics Partner

Finance Vertical Title Partner

Smart Cities Vertical Title Partner

Grow your Business with Google

Xerox Future of Work Global Tour

Mobile Reseller Feature Sponsor

Official Publisher

Official Publication

Official Airline Partner

Official Courier Handler

Who Said Machine Learning is Rocket Science?

If you've been curious about machine learning, this article is a good place to start exploring it. Apart from introducing readers to the basics, it also encourages you to learn by pointing you in the direction of various online courses and libraries.

All of a sudden, everyone has started talking about artificial intelligence or AI. Rapid advances in this field have certainly forced people to believe that this is going to drive innovation for at least a decade.

We experience the application of AI in our daily lives in innumerable ways. This could be in the form of an advertisement of a product on a social media platform where you were planning to buy something, or it could be a recommendation of another product that pops up as you're purchasing a product from an e-commerce website. For example, a recommendation may be made to buy a mobile cover just as you're buying a mobile phone, because the site's system has predicted a higher probability of being able to sell a mobile cover along with the phone. So, knowingly or unknowingly, we are already experiencing AI.

There have been some incredible developments in AI that have led many to believe it is going to be the technology that will shape our future.

1. ***AlphaGo beats world champion at the game Go:*** In March 2016, Google's DeepMind achieved a major victory in deep learning. AlphaGo, a division of the company, mastered the ancient Chinese game Go and defeated Lee Sedol, the world champion in four out of five games. According to many, Go is considered to be the most complex professional game because of a huge number of potential moves that can be made.
2. ***AI predicted US election results:*** Many of us were surprised by the outcome of the US presidential election results, but a startup called MogIA based in Mumbai was able to predict it successfully a month before the results were declared. The company analysed social media sentiment through millions of social media data points. This was the firm's fourth successful prediction in a row.
3. ***AI improves cancer diagnosis:*** There have been some path-breaking innovations in the field of healthcare. It is believed that the healthcare industry is going to benefit the

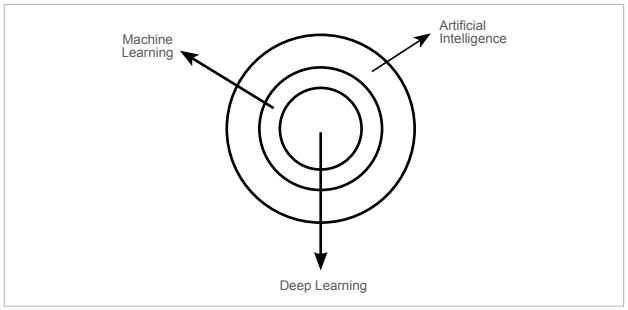


Figure 1: Artificial intelligence, machine learning and deep learning

most from AI. There are AI programs that can now predict the occurrence of cancer with 90 per cent accuracy just by analysing the symptoms of a patient, which can help a doctor to start treatment early.

Many of us often confuse the terms 'artificial intelligence', 'machine learning' and 'deep learning'. Hence, we use the terms interchangeably. But these are not the same things.

Artificial intelligence, in broader terms, can be described as a branch of computer science that can imitate human beings. It has been demonstrated that computers can be programmed to carry out very complex tasks that were earlier only performed by humans. From self-driving cars and Amazon's Alexa or Apple's Siri, to a computer program playing a game of chess, all of these are applications of artificial intelligence.

Machine learning can be referred to as a subset of AI. It is considered one of the most successful approaches to AI, but is not the only approach. For example, there are many chat bots that are rule based, i.e., they can answer only certain questions, depending on the way they were programmed. But they will not be able to learn anything new from those questions. So this can be categorised as AI as the chat bots replicate human-like behaviour, but can't be termed as machine learning. Now, the question is: can machines really 'learn'? How is it possible for a machine to learn if it doesn't have a brain and a complex nervous system like humans? According to Arthur Samuel, "Machine learning can be defined as a field of study that gives computers the ability to learn without being explicitly programmed." We can also define it as the computer's ability to learn from experience to perform a certain task, whereby the performance will improve with experience. This is akin to a computer program playing chess, which can be categorised as machine learning, if it learns from previous experiences and subsequently makes better moves to win a game.

Deep learning can be categorised as a subset of machine learning. It uses neural networks to simulate human decision-making skills. A neural network consists of many neurons and hence resembles a human nervous system. Have you ever wondered how Facebook detects *your* face amongst many, in an image? Image detection is one of the examples of deep learning, which is much more complex as it needs lots of data to train itself. For instance, a deep learning algorithm can learn to recognise a car but will have to be trained on a huge data set which consists of cars as well as other objects. If this is not done, it might make a wrong decision like identifying a bus as a car. Hence, compared to other machine learning algorithms, a deep learning algorithm requires much more data in order to observe and understand every minute detail to make the right decisions.

From Figure 1, you can see how all these terms are related to each other, yet are not the same.

Now that you have understood the differences between artificial intelligence, machine learning and deep learning, let's dig deeper into machine learning.

There are three main types of machine learning algorithms.

1. ***Supervised learning:*** The data set in supervised learning consists of input data as well as the expected output. The algorithm is a function which maps this input data to the expected outcome. The algorithm will continue to learn until the model achieves a desired level of accuracy. Then this model can be applied to new sets of data, for which the expected outcome is not available but needs to be predicted from a given set of data.
   For instance, let's look at a car manufacturing company that wants to set a price for its newest model. In order to do so, it can use this supervised learning model. The company's input data set may consist of details of previous car models — their features like the number of air bags, electronic gadgets, etc. The output (or expected outcome) would be the sale price of the car. Now an algorithm can be designed to map those input parameters (or features) to the expected outcome. Once the algorithm achieves the desired level of accuracy, this model can be applied to the firm's new car model. This can help the company predict the car price at which it should be launched.
   For better results, the company can use a data set of car models of other manufacturers and their prices. This would help the company in setting a competitive price.
   In machine learning, the best results are not achieved by using a great algorithm but by using the most data.
2. ***Unsupervised learning:*** The only difference between supervised and unsupervised learning is that the data set doesn't have the expected outcome as in the supervised learning model. The data set will only have input parameters (or features) and the algorithm will have to predict the outcome. For instance, if a shirt manufacturing company is looking to manufacture three different sizes of shirts (small, medium and large), its data comprises the shoulder, waist and chest sizes of its customers. Now, depending upon this huge data set, the company needs to group the sizes into three categories so that there can be a best fit for everyone. Here, an unsupervised learning model can be used to group the different data points in three different sizes and predict an appropriate shirt size for every customer.

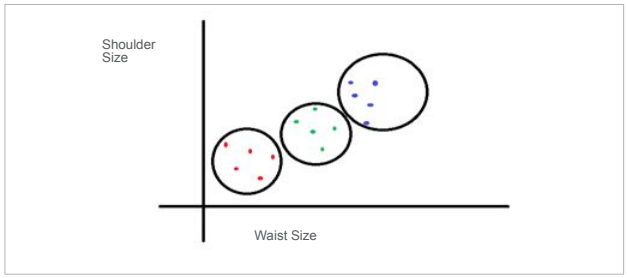


Figure 2: Unsupervised learning using clustering

As per the graph given in Figure 2, let's consider a company that has only the shoulder and waist measurements as input parameters of the data set. It will now have to categorise this data set into three groups, which can help the company predict the shirt size for every customer. This technique is known as clustering, in which the data set is clustered into the desired number of clusters. Most of the time, the data set is not like the one shown in this example. Data points that are very close to each other make it difficult to implement clustering. Also, clustering is just one of the many techniques used in unsupervised learning to predict the outcome.

3. ***Reinforcement learning*:** In reinforcement learning, a machine or an agent trains itself when exposed to a particular environment, by a process of trial and error. Let's consider a kid who wants to learn to ride a bicycle. First, she will try to learn from someone who already knows how to ride a bicycle. Then, she will try riding on her own and might fall down a number of times. Learning from her previous mistakes, she will try to ride without falling. And when she finally rides the bike without falling, it can be considered as a reward for her efforts. Now let's consider this kid as a machine or an agent who is getting punished (falling) for committing a mistake and earning a reward (not falling) for not committing any mistake. A chess-playing program can be a good example of this, whereby one wrong move will penalise the agent and it may lose a game, while a combination of one or more right moves will earn it a reward by making it win.

These are the three basic learning models of machine learning. As per the need, these models can be used in combination to generate a new model. For instance, supervised learning can sometimes be used along with unsupervised learning, depending upon the data set as well as the expected outcome.

People often feel that machine learning is only for someone who is good with mathematics or statistics, and will be impossible to learn for anyone else. This is a fallacy. Machine learning is not rocket science after all. The only thing that is required to learn it is eagerness and curiosity. The number of tools and libraries available has made it even easier to learn it. Google's TensorFlow library, which is now open source, or the numerous Python libraries like NumPy and scikit-learn, are just a few of these. Anyone can use these libraries and even contribute to them to solve problems, as they are open source. You don't need to worry about the complexities involved in your algorithm, like complex mathematical computations (such as gradient, matrix multiplication, etc) as this task can be left for these libraries to implement. Libraries make it easier for everyone so that instead of getting involved in implementing complex computations, the user can now focus on the application of the algorithm.

There are also many APIs available that can be used to implement an AI based system. One such API is IBM's Watson – a cognitive, computing based AI system. Cognitive computing is a mixture of different techniques such as machine learning, natural language processing, AI, etc. Watson is capable of doing many tasks like answering a user's queries, helping doctors to spot diseases, and a lot more.

If you are excited by the prospects that machine learning offers, our digital education era has made things easier for you. There are many massive open online courses (MOOC) offered by many companies. One such course is provided by Coursera-Machine Learning. This is taught by Andrew Ng, one of the co-founders of Coursera. This course will give you a basic understanding of the algorithms that are implemented in machine learning, and it includes both supervised learning and unsupervised learning. It's a self-paced course but designed to be completed in 12 weeks. If you want to dig deeper and study deep learning, which is a subset of machine learning, you can learn it through another course provided by *fast.ai*. This course is split into two parts: *Practical deep learning for coders (Part 1)* and *Cutting edge deep learning for coders (Part 2)*. Both have been designed for seven weeks each and provide you a great insight into deep learning. If you further want to specialise in deep learning, you can opt for a deep learning specialisation course by Coursera and *deeplearning.ai*.

We all know that theory without practice is like a body without life. So, for you to practice, there are many sources that can provide you a huge data set to test your knowledge and implement what you've learnt. One such website is *Kaggle*, which provides a diverse data set and can help you overcome your major hurdle, i.e., getting data to test your learning model.

If you sometimes feel lost in this journey of learning, when your algorithm does not work as expected or when you don't understand a complex equation, remember the famous dialogue in the movie, *The Pursuit of Happyness:* "Don't ever let someone tell you that you can't do something. Not even me. You got a dream; you gotta protect it. When people can't do something themselves, they're gonna tell you that you can't do it. You want something, go get it. Period." END

By: Puneet Batra

The author works as a software professional at Xoriant Solutions. He is always keen to learn new technologies and share his knowledge with others. He can be reached at *puneetbatra29@gmail.com*.


www.IndiaElectronicsWeek.com
DRIVING TECHNOLOGY, INNOVATION & INVESTMENTS
RESULTS
INDIA ELECTRONICS WEEK
7-9 FEBRUARY 2018
KTPO Trade Center • Whitefield • Bengaluru

Profit from IoT

India's #1 IoT show. At Electronics For You, we strongly believe that India has the potential to become a superpower in the IoT space, in the upcoming years. All that's needed are platforms for different stakeholders of the ecosystem to come together.

We've been building one such platform: **IoTshow.in**--an event for the creators, the enablers and customers of IoT. In February 2018, the third edition of IoTshow.in will bring together a B2B expo, technical and business conferences, the Start-up Zone, demo sessions of innovative products, and more.

Who should attend?

- **Creators of IoT solutions:** OEMs, design houses, CEOs, CTOs, design engineers, software developers, IT managers, etc
- **Enablers of IoT solutions:** Systems integrators, solutions providers, distributors, resellers, etc
- **Business customers:** Enterprises, SMEs, the government, defence establishments, academia, etc

Why you should attend

- Get updates on the latest technology trends that define the IoT landscape
- Get a glimpse of products and solutions that enable the development of better IoT solutions
- Connect with leading IoT brands seeking channel partners and systems integrators
- Connect with leading suppliers/service providers in the electronics, IT and telecom domain who can help you develop better IoT solutions, faster
- Network with the who's who of the IoT world and build connections with industry peers
- Find out about IoT solutions that can help you reduce costs or increase revenues
- Get updates on the latest business trends shaping the demand and supply of IoT solutions

India's Electronics Manufacturing Show

Is there a show in India that showcases the latest in electronics manufacturing such as rapid prototyping, rapid production and table top manufacturing?

Yes, there is now - EFY Expo 2018. With this show's focus on the areas mentioned and it being co-located at India Electronics Week, it has emerged as India's leading expo on the latest manufacturing technologies and electronic components.

Who should attend?

- **Manufacturers:** CEOs, MDs, and those involved in firms that manufacture electronics and technology products
- **Purchase decision makers:** CEOs, purchase managers, production managers and those involved in electronics manufacturing
- **Technology decision makers:** Design engineers, R&D heads and those involved in electronics manufacturing
- **Channel partners:** Importers, distributors, resellers of electronic components, tools and equipment
- **Investors:** Startups, entrepreneurs, investment consultants and others interested in electronics manufacturing

Why you should attend

- Get updates on the latest technology trends in rapid prototyping and production, and in table top manufacturing
- Get connected with new suppliers from across India to improve your supply chain
- Connect with OEMs, principals and brands seeking channel partners and distributors
- Connect with foreign suppliers and principals to represent them in India
- Explore new business ideas and investment opportunities in this sector

Showcasing the Technology that Powers Light

Our belief is that the LED bulb is the culmination of various advances in technology. And such a product category and its associated industry cannot grow without focusing on the latest technologies. But, while there are some good B2B shows for LED lighting in India, none has a focus on 'the technology that powers lights'. Thus, the need for LEDAsia.in.

Who should attend?

- **Tech decision makers:** CEOs, CTOs, R&D and design engineers and those developing the latest LED-based products
- **Purchase decision makers:** CEOs, purchase managers and production managers from manufacturing firms that use LEDs
- **Channel partners:** Importers, distributors, resellers of LEDs and LED lighting products
- Investors: Startups, entrepreneurs, investment consultants interested in this sector
- **Enablers:** System integrators, lighting consultants and those interested in smarter lighting solutions (thanks to the co-located IoTshow.in)

Why you should attend

- Get updates on the latest technology trends defining the LED and LED lighting sector
- Get a glimpse of the latest components, equipment and tools that help manufacture better lighting products
- Get connected with new suppliers from across India to improve your supply chain
- Connect with OEMs, principals, lighting brands seeking channel partners and systems integrators
- Connect with foreign suppliers and principals to represent them in India
- Explore new business ideas and investment opportunities in the LED and lighting sector
- Get an insider's view of 'IoT + Lighting' solutions that make lighting smarter

EFY CONFERENCES

India's Mega Tech Conference

The EFY Conference (EFYCON) started out as a tiny 900-footfall community conference in 2012, going by the name of Electronics Rocks. Within four years, it grew into 'India's largest, most exciting engineering conference,' and was ranked 'the most important IoT global event in 2016' by Postscapes.

In 2017, 11 independent conferences covering IoT, artificial intelligence, cyber security, data analytics, cloud technologies, LED lighting, SMT manufacturing, PCB manufacturing, etc, were held together over three days, as part of EFY Conferences.

Key themes of the conferences and workshops in 2018

- Profit from IoT: How suppliers can make money and customers save it by using IoT
- IT and telecom tech trends that enable IoT development
- Electronics tech trends that enable IoT development
- Artificial intelligence and IoT
- Cyber security and IoT
- The latest trends in test and measurement equipment
- What's new in desktop manufacturing
- The latest in rapid prototyping and production equipment

Who should attend

- Investors and entrepreneurs in tech
- Technical decision makers and influencers
- R&D professionals
- Design engineers
- IoT solutions developers
- Systems integrators
- IT managers

SPECIAL PACKAGES FOR
• Academicians • Defence personnel
• Bulk/Group bookings

Reasons Why You Should NOT Attend IEW 2018

We spoke to a few members of the tech community to understand why they had not attended earlier editions of India Electronics Week (IEW). Our aim was to identify the most common reasons and share them with you, so that if you too had similar reasons, you may choose not to attend IEW 2018. This is what they shared…

#1. Technologies like IoT, AI and embedded systems have no future

Frankly, I have NO interest in new technologies like Internet of Things (IoT), artificial intelligence, etc. I don't think these will ever take off, or become critical enough to affect my organisation or my career.

#2. I see no point in attending tech events

What's the point in investing energy and resources to attend such events? I would rather wait and watch—let others take the lead. Why take the initiative to understand new technologies, their impact and business models?

#3. My boss does not like me

My boss is not fond of me and doesn't really want me to grow professionally. And when she came to know that IEW 2018 is an event that can help me advance my career, she cancelled my application to attend it. Thankfully, she is attending the event! Look forward to a holiday at work.

#4. I hate innovators!

Oh my! Indian startups are planning to give LIVE demonstrations at IEW 2018? I find that hard to believe. Worse, if my boss sees these, he will expect me to create innovative stuff too. I better find a way to keep him from attending.

#5. I am way too BUSY

I am just too busy with my ongoing projects. They just don't seem to be getting over. Once I catch up, I'll invest some time in enhancing my knowledge and skills, and figure out how to meet my deadlines.

#6. I only like attending vendor events

Can you imagine an event where most of the speakers are not vendors? Where most talks will not be by people trying to sell their products? How boring! I can't imagine why anyone would want to attend such an event. I love sales talks, and I am sure everybody else does too. So IEW is a big 'no-no' for me.

#7. I don't think I need hands-on knowledge

I don't see any value in the tech workshops being organised at IEW. Why would anyone want hands-on knowledge? Isn't browsing the Net and watching YouTube videos a better alternative?

#8. I love my office!

Why do people leave the comfort of their offices and weave through that terrible traffic to attend a technical event? They must be crazy. What's the big deal in listening to experts or networking with peers? I'd rather enjoy the coffee and the cool comfort of my office, and learn everything by browsing the Net!

#9. I prefer foreign events

While IEW's IoTshow.in was voted the 'World's No.1 IoT event' on Postscapes.com, I don't see much value in attending such an event in India—and that, too, one that's being put together by an Indian organiser. Naah! I would rather attend such an event in Europe.

Hope we've managed to convince you NOT to attend IEW 2018! Frankly, we too have NO clue why 10,000-plus techies attended IEW in March 2017. Perhaps there's something about the event that we've not figured out yet. But, if we haven't been able to dissuade you from attending IEW 2018, then you may register at **http://register.efy.in.**

Conference Pass Pricing	
One day pass	INR 1999
PRO pass	INR 7999

Special privileges and packages for...
Defence and defence electronics personnel
Academicians
Group and bulk bookings

INDIA ELECTRONICS WEEK 7-9 FEBRUARY 2018

KTPO Trade Center
Whitefield • Bengaluru

Colocated shows

The themes

- Profit from IoT
- Table top manufacturing
- Rapid prototyping and production
- LEDs and LED lighting

The co-located shows

Why exhibit at IEW 2018?

More technology decision makers and influencers attend IEW than any other event

It's a technology-centric show and not just a B2B event

Over 3,000 visitors are conference delegates

Besides purchase orders, you can bag 'Design Ins' and 'Design-Wins' too

Co-located events offer cross-pollination of business and networking opportunities

India's only test and measurement show is also a part of IEW

360-degree promotions via the event, publications and online!

The only show in Bengaluru in the FY 2017-18

Your brand and solutions will reach an audience of over 500,000 relevant and interested people

IEW connects you with customers before the event, at the event, and even after the event

Bag year-end orders; meet prospects in early February and get orders before the FY ends

The world's No.1 IoT show is a part of IEW and IoT is driving growth

It's an Electronics For You Group property

IEW is being held at a venue (KTPO) that's closer to where all the tech firms are

Special packages for 'Make in India', 'Design in India', 'Start-up India' and 'LED Lighting' exhibitors

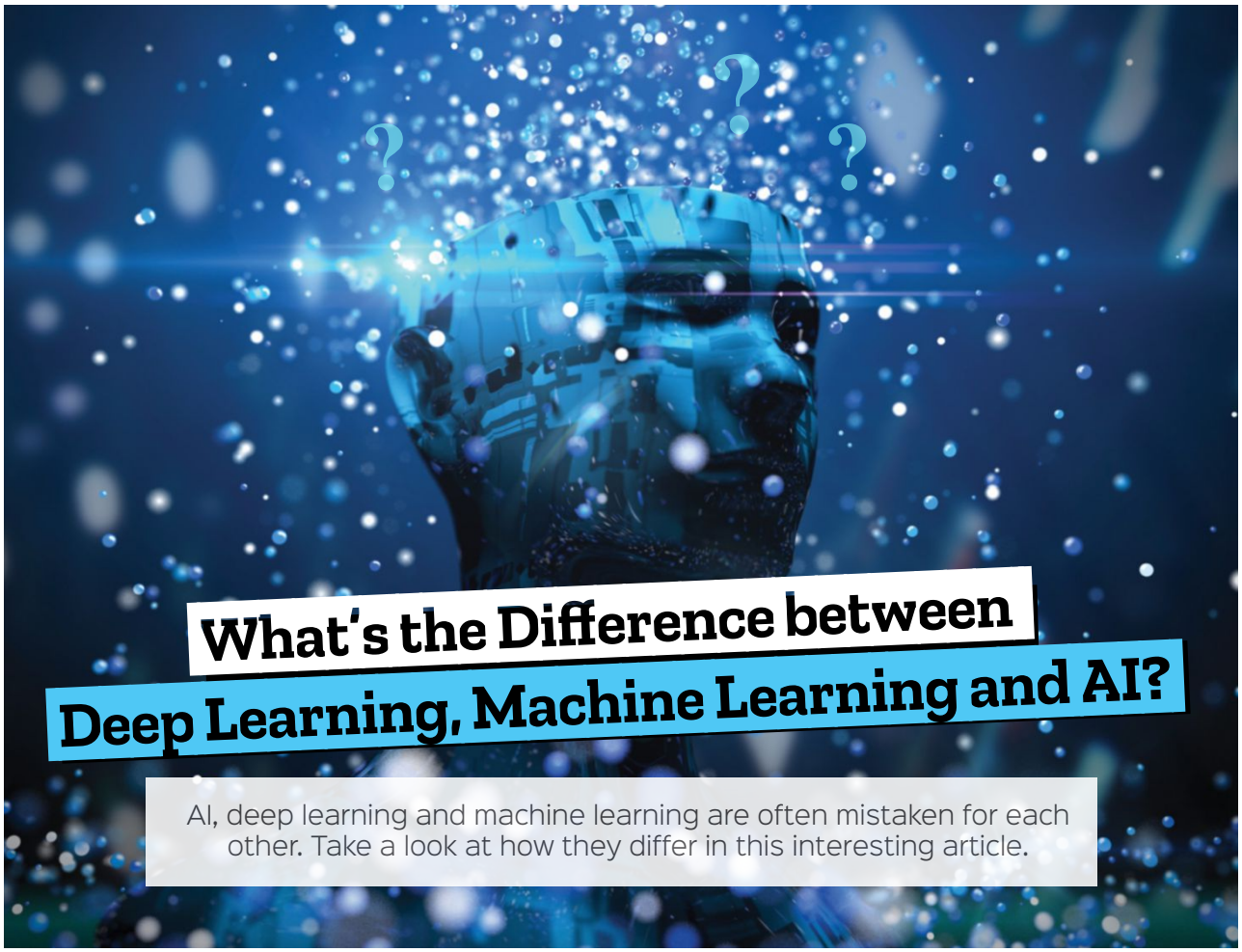
Why you should risk being an early bird

1. The best locations sell out first
2. The earlier you book—the better the rates; and the more the deliverables
3. We might just run out of space this year!

To get more details on how exhibiting at IEW 2018 can help you achieve your sales and marketing goals,

Contact us at +91-9811155335 *Or* Write to us at growmybiz@efy.in

EFY Enterprises Pvt Ltd | D-87/1, Okhla Industrial Area, Phase -1, New Delhi– 110020



What's the Difference between Deep Learning, Machine Learning and AI?

AI, deep learning and machine learning are often mistaken for each other. Take a look at how they differ in this interesting article.

From being dismissed as science fiction to becoming an integral part of multiple, wildly popular movie series, especially the one starring Arnold Schwarzenegger, artificial intelligence has been a part of our life for longer than we realise. The idea of machines that can think has widely been attributed to a British mathematician and WWII code-breaker, Alan Turing. In fact, the Turing Test, often used for benchmarking the 'intelligence' in artificial intelligence, is an interesting process in which AI has to convince a human, through a conversation, that it is not a robot. There have been a number of other tests developed to verify how evolved AI is, including Goertzel's Coffee Test and Nilsson's Employment Test that compare a robot's performance in different human tasks.

As a field, AI has probably seen the most ups and downs over the past 50 years. On the one hand it is hailed as the frontier of the next technological revolution, while on the other, it is viewed with fear, since it is believed to have the potential to surpass human intelligence and hence achieve world domination! However, most scientists agree that we are in the nascent stages of developing AI that is capable of such feats, and research continues unfettered by the fears.

Applications of AI

Back in the early days, the goal of researchers was to construct complex machines capable of exhibiting some semblance of human intelligence, a concept we now term 'general intelligence'. While it has been a popular concept in movies and in science fiction, we are a long way from developing it for real.

Specialised applications of AI, however, allow us to use image classification and facial recognition as well as smart personal assistants such as Siri and Alexa. These usually leverage multiple algorithms to provide this functionality to the end user, but may broadly be classified as AI.

Machine learning (ML)

Machine learning is a subset of practices commonly aggregated under AI techniques. The term was originally used to describe the process of leveraging algorithms to parse data, build models that could learn from it, and ultimately make predictions using these learnt parameters. It encompassed various strategies including decision trees, clustering, regression, and Bayesian approaches that didn't quite achieve the ultimate goal of 'general intelligence'.

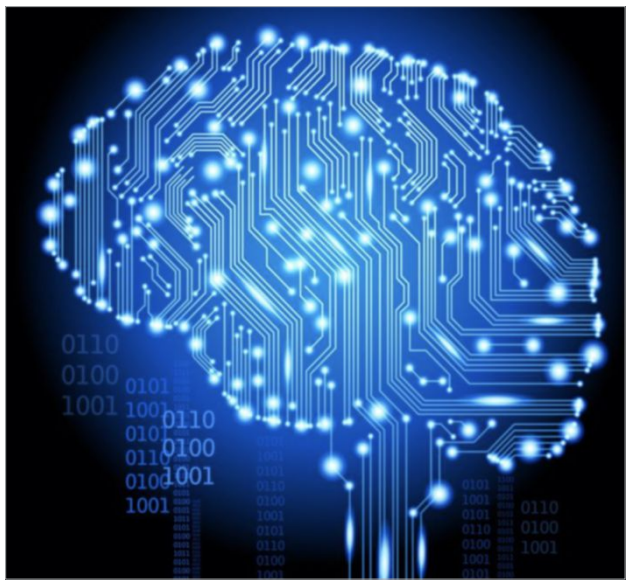


Figure 1: Conventional understanding of AI [*Image credit:* Geeky-Gadgets]

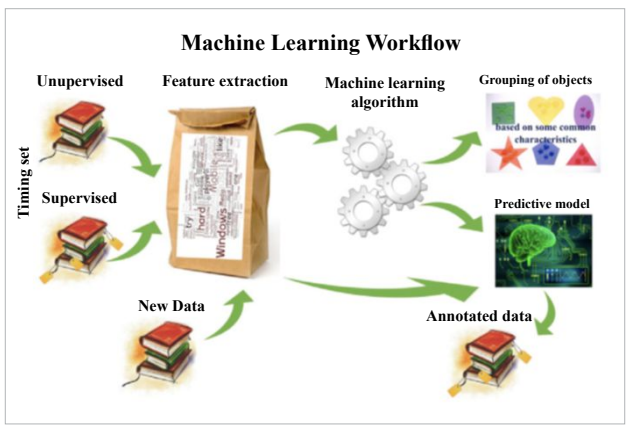


Figure 2: Machine learning workflow [*Image credit:* TomBone's Computer Vision Blog]

While it began as a small part of AI, burgeoning interest has propelled ML to the forefront of research and it is now used across domains. Growing hardware support as well as improvements in algorithms, especially pattern recognition, has led to ML being accessible for a much larger audience, leading to wider adoption.

Applications of ML

Initially, the primary applications of ML were limited to the field of computer vision and pattern recognition. This was prior to the stellar success and accuracy it enjoys today. Back then, ML seemed a pretty tame field, with its scope limited to education and academics.

Today we use ML without even being aware of how dependent we are on it for our daily activities. From Google's search team trying to replace the PageRank algorithm with an improved ML algorithm named RankBrain, to Facebook automatically suggesting friends to tag in a picture, we are surrounded by use cases for ML algorithms.

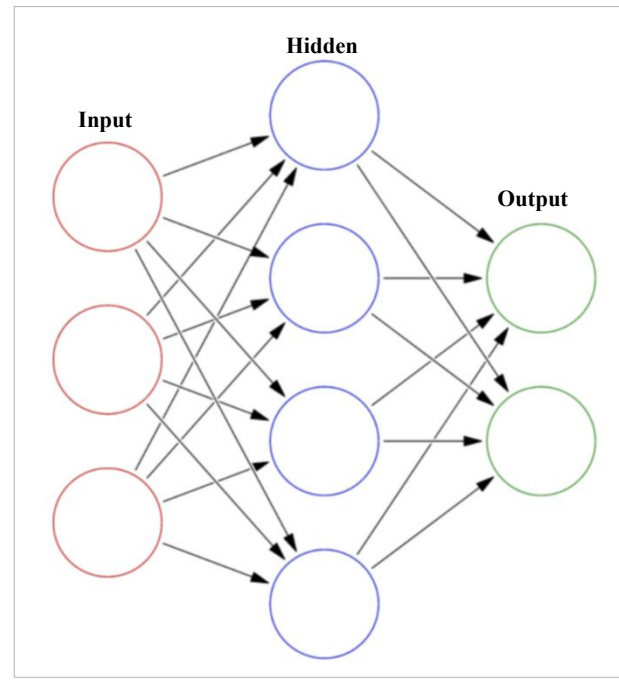


Figure 3: Artificial neural networks [*Image credit:* Shutterstock]

Deep learning (DL)

A key ML approach that remained dormant for a few decades was artificial neural networks. This eventually gained wide acceptance when improved processing capabilities became available. A neural network simulates the activity of a brain's neurons in a layered fashion, and the propagation of data occurs in a similar manner, enabling machines to learn more about a given set of observations and make accurate predictions.

These neural networks that had until recently been ignored, save for a few researchers led by Geoffrey Hinton, have today demonstrated an exceptional potential for handling large volumes of data and enhancing the practical applications of machine learning. The accuracy of these models allows reliable services to be offered to end users, since the false positives have been eliminated almost entirely.

Applications of DL

DL has large scale business applications because of its capacity to learn from millions of observations at once. Although computationally intensive, it is still the preferred alternative because of its unparalleled accuracy. This encompasses a number of image recognition applications that conventionally relied on computer vision practices until the emergence of DL. Autonomous vehicles and recommendation systems (such as those used by Netflix and Amazon) are among the most popular applications of DL algorithms.

Comparing AI, ML and DL

Comparing the techniques: The term AI was defined in the Dartmouth Conference (1956) as follows: "Every aspect of learning or any other feature of intelligence can in principle

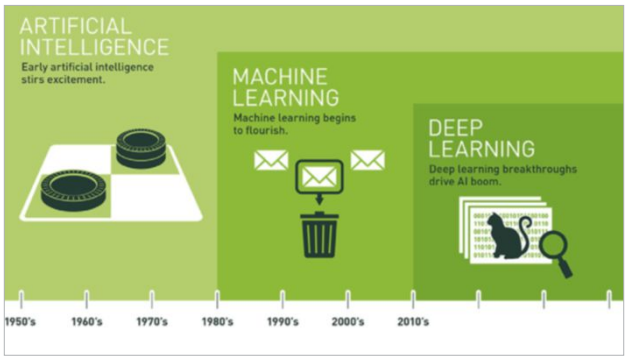


Figure 4: A comparison of techniques [*Image credit*: NVIDIA]

Figure 5: Applications of artificial intelligence [*Image credit:* The Telegraph, UK]

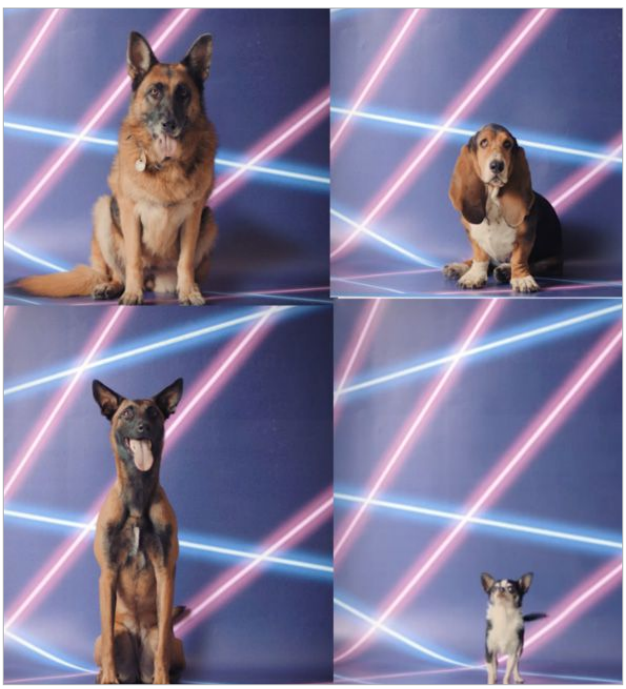


Figure 6: Deep learning for identifying dogs [*Image credit:* Datamation]

be so precisely described that a machine can be made to simulate it." It is a broad definition that covers use cases that range from a game-playing bot to a voice recognition system within Siri, as well as converting text to speech and vice versa. It is conventionally thought to have three categories:

- Narrow AI specialised for a specific task
- Artificial general intelligence (AGI) that can simulate human thinking
- Super-intelligent AI, which implies a point where AI surpasses human intelligence entirely

ML is a subset of AI that seems to represent its most successful business use cases. It entails learning from data in order to make informed decisions at a later point, and enables AI to be applied to a broad spectrum of problems. ML allows systems to make their own decisions following a learning process that trains the system towards a goal. A number of tools have emerged that allow a wider audience access to the power of ML algorithms, including Python libraries such as scikit-learn, frameworks such as MLib for Apache Spark, software such as RapidMiner, and so on.

A further sub-division and subset of AI would be DL, which harnesses the power of deep neural networks in order to train models on large data sets, and make accurate predictions in the fields of image, face and voice recognition, among others. The low trade-off between training time and computation errors makes it a lucrative option for many businesses to switch their core practices to DL or integrate these algorithms into their system.

Classifying applications: There are very fuzzy boundaries that distinguish the applications of AI, ML and DL. However, since there is a demarcation of the scope, it is possible to identify which subset a specific application belongs to. Usually, we classify personal assistants and other forms of bots that aid with specialised tasks, such as playing games, as AI due to their broader nature. These include the applications of search capabilities, filtering and short-listing, voice recognition and text-to-speech conversion bundled into an agent.

Practices that fall into a narrower category such as those involving Big Data analytics and data mining, pattern recognition and the like, are placed under the spectrum of ML algorithms. Typically, these involve systems that 'learn' from data and apply that learning to a specialised task.

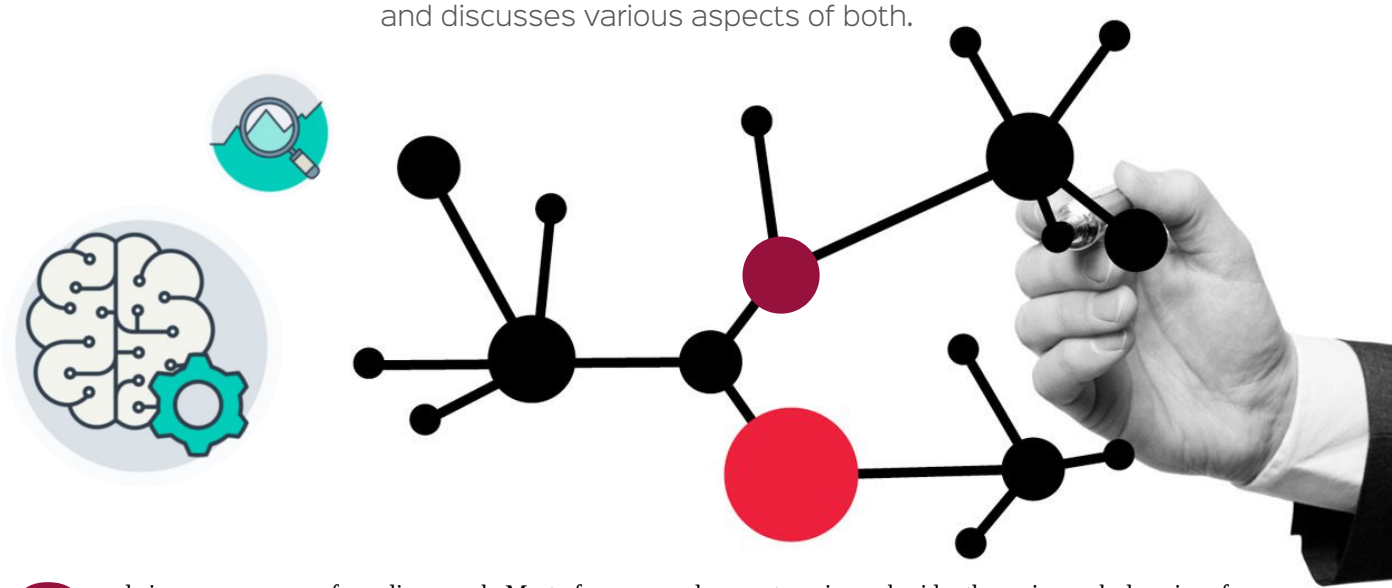
Finally, applications belonging to a niche category, which encompasses a large corpus of text or image-based data utilised to train a model on graphics processing units (GPUs) involve the use of DL algorithms. These often include specialised image and video recognition tasks applied to a broader usage, such as autonomous driving and navigation. END

By: Swapneel Mehta

The author has worked at Microsoft Research, CERN and at startups in AI and cyber security. He is an open source enthusiast who enjoys spending time organising software development workshops for school and college students. You can contact him at *https://www.linkedin.com/in/swapneelm, https://github.com/SwapneelM* or *http://www.ccdev.in.*

Data Science and Machine Learning: Working Together to Make Machines Smarter

This article clarifies the role of data science in relation to machine learning, and discusses various aspects of both.



Google is now a synonym for online search. Most of us will agree with this, because whenever we try to get information on something we don't know, we say, "Google it!" Have you ever given a thought to how Google comes up with the answers to different questions in an instant? Google and other such search engines make use of different data science algorithms and machine learning techniques to come up with the best results for all our search queries, and that too, in less than a second. Google processes more than 20 petabytes of structured and unstructured data daily and, even then, is able to instantly flash our search results. Had there been no data science and machine learning, Google would not have been able to perform all this and it would not have been the all-pervasive search engine we all depend on today. Data science is one of the roots that support the tree of the digital world.

Data science is also known as data-driven science as it deals with various scientific processes, methods and systems to extract knowledge or insights from large sets of data — either unstructured or structured. We all have access to huge amounts of data, which is about many aspects of our lives – it could be related to communication, shopping, reading the news, searching for information, expressing our opinions, etc. All this is being used to extract useful insights by employing different data science techniques. Data science is basically a concept that unifies statistics with data analysis in order to analyse and relate real world activities with data. It employs different techniques and theories that are drawn from many fields — from within the broad areas of statistics, mathematics, information science and computer science, besides the various sub-domains of machine learning, cluster analysis, classification, databases, data mining and visualisation.

According to the Turing Award winner Jim Gray, data science is the fourth paradigm of science. Gray asserts that everything about science is rapidly changing because of the impact of information technology and the data deluge. Data science plays a crucial role in transforming the information collected during datafication and adds a value to it. Datafication is nothing but the process of taking different aspects of life and turning these to data. For instance, Twitter datafies different stray thoughts, LinkedIn datafies the professional networks, and so on. We take the help of different data science techniques to extract useful parts out of the collected information during datafication.

Drew Conway is famous for his Venn diagram definition of data science. He applied it to study and analyse one of the biggest problems of the globe –terrorism. If we take a look at his Venn diagram definition, data science is the union of hacking skills, statistical and mathematical knowledge, and substantive expertise about the specific subject. According to him, data science is the civil engineering of data. It requires a practical knowledge of different tools and materials, coupled with a theoretical understanding of what's possible.

The workflow for data science

Data science comprises several sequences of processes which are followed to deduce useful insights from the raw set of

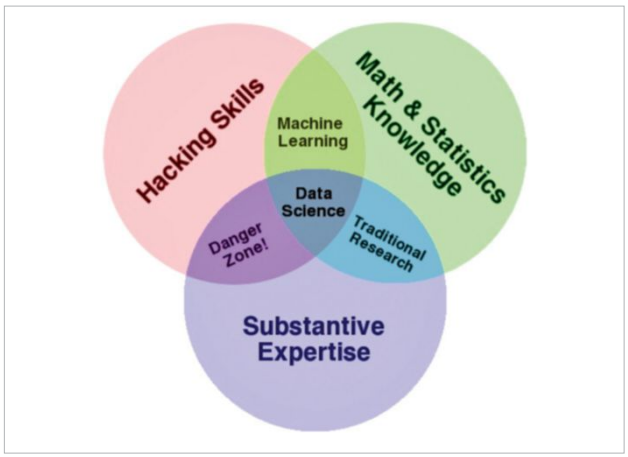


Figure 1: A Venn diagram definition of data science by Drew Conway (*Image credit:* googleimages.com)

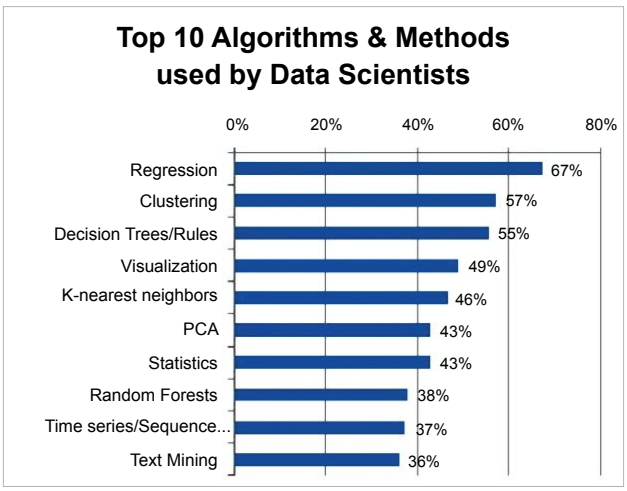


Figure 2: The top algorithms and methods used by data scientists (*Image credit:* googleimages.com)

data. This ultimately helps the system to make decisions. Let us have a look at the different processes followed in data science.

Collection of raw data: This is the first step implemented in data science and deals with the collection of actual raw data, on which different data science operations need to be performed. There are broadly two ways to do this:

1. We can pick one or many tools to collect data automatically from different data sources. This option is widely used in order to collect data from large data sources. We just need to copy-paste a small code snippet into our website and we are ready to go (e.g., Hotjar, Google Analytics, etc).
2. We can also collect the data for ourselves using a JavaScript code snippet that sends the data in a *.csv* plain text file on the server. This is a bit difficult to implement as it requires some coding skills. But if we think about the long term, this solution is more profitable.

Data processing: This refers to the refinement of raw data that has been collected during the data collection process. We all know that the raw data is unprocessed and unorganised. It needs to be arranged and organised so that it becomes easier to perform operations on it. Once the data is processed, we get an output data which is the processed, categorised and summarised version. Data processing is required in most of the experiments and surveys. The collected raw data sometimes contains too much data to analyse it sensibly. This is especially the case when we do research using computers as this may produce large sets of data. The data then needs to be organised or manipulated using the deconstruction technique.

Data set cleaning: This is the process of removing unwanted data from the processed data set and keeping only what's required for analysis. This helps to reduce the large set of data to a smaller one by removing the inconsistent or incorrect data, and makes it easier to perform different analysis tasks on it.

Exploratory data analysis: This approach is used to analyse the data sets in order to summarise their important characteristics, often with the help of visual methods. A statistical model can also be used for analysis, but primarily, exploratory data analysis is for visualising what the data can tell us beyond the formal modelling or the hypothesis testing task. This approach was promoted by John Tukey to encourage different data scientists to explore the data, and hence possibly formulate the hypotheses that could lead to new methods of data collection and experiments. This is different from the initial data analysis, which focuses mostly on checking the assumptions required for model fitting, the hypothesis testing, the handling of different missing values and making transformations of variables as required.

Models and algorithms: Once the data is cleansed, some sets of data will need exploratory analysis whereas other sets can be directly used for the selection of data models and algorithms. This phase of data science deals with the process of selecting the right and appropriate algorithm on the basis of the data set obtained after data cleaning, and also on the basis of the knowledge obtained about the data set during exploratory data analysis. The algorithm chosen is such that it's most efficient for the available data set. This process also includes the design, development and selection of the data models, which can be used to perform the required operations on the data, to obtain the required data product.

Report communication: This is the part of data science that deals with generating and developing visual reports in the form of graphs and pie-charts, which can be used by data scientists to analyse the data patterns and make the appropriate decisions. This decision is the final output, which is then utilised in different applications.

Data product: This is the final data product, which is used to continuously improve and change the application system whose data is analysed. This can be considered as the end product, which represents the whole set of operations performed on the collected raw data set.

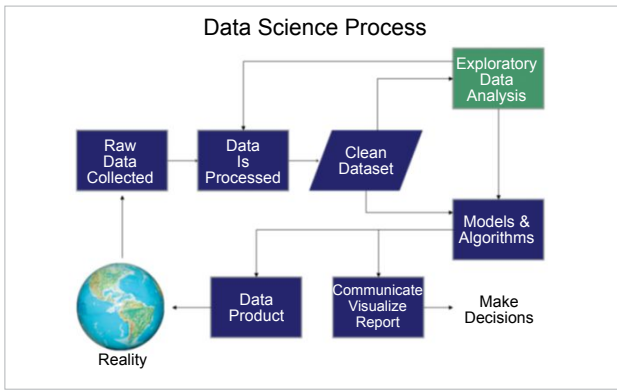


Figure 3: Data science workflow (*Image credit:* googleimages.com)

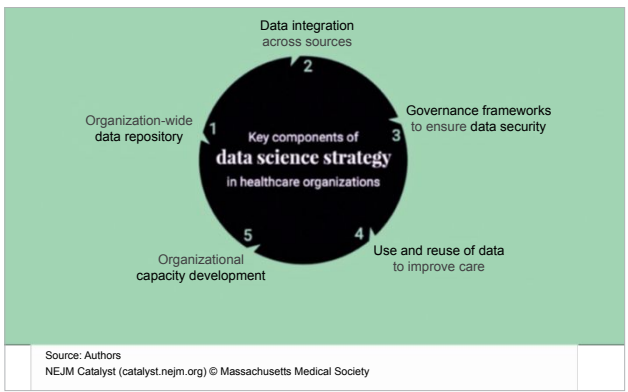


Figure 4: Different components of data science in a healthcare organisation (*Image credit:* googleimages.com)

What is machine learning?

Machine learning is a part of computer science that gives any system the ability to learn on its own without being programmed. It makes a machine learn in the same way as human beings learn by themselves. Just as we learn any system on the basis of our experience and the knowledge gained after analysing the system, even machines can analyse and study the system's behaviour or its output data and learn how to take decisions on that basis. This is the backbone of artificial intelligence. It makes machines get into a self-learning mode without any explicit programming. When the machine is fed with new data, it learns, grows and changes by itself.

Machine learning has evolved from the concept of pattern recognition and computational learning theory in artificial intelligence. It explores the study and construction of different algorithms that can learn from data and make predictions on them. These algorithms do not follow the static program instructions, but make data-driven predictions or decisions by building a model from some sample inputs.

There are three types of machine learning, differentiated on the basis of the learning signal available to any learning system.

1. **Supervised learning:** In this type of learning, the machine is presented with few example inputs and also their desired outputs, which are given by a teacher. The goal is to learn a general rule that maps inputs to outputs.
2. **Unsupervised learning:** This is a type of machine learning in which no labels are given to the learning algorithm, leaving it to find the structure in its input on its own.
3. **Reinforcement learning:** Under this learning system, a computer program actually interacts with a dynamic environment for which it must perform a specific goal (for example, driving a vehicle or playing a game against an opponent).

How is machine learning related to data science?

Machine learning is very closely related to (and sometimes overlaps with) data science or computational statistics, as both focus on making predictions with the help of machines or computers. It has strong ties with mathematical optimisation, which provides different methods and theories to optimise learning systems. Machine learning is often combined with data science, with the latter actually focusing more on exploratory data analysis, and this is known as unsupervised learning.

If we talk specifically about the field of data science, machine learning is used to devise various complex models and algorithms that lend themselves to prediction. This is also known as predictive analytics. All these analytical models allow data scientists, researchers, engineers and analysts to produce reliable and repeatable decisions in order to uncover various hidden insights by learning from the historical relationships and trends in the large sets of data.

Data analysis has been traditionally characterised by the trial and error approach, and we all know that this becomes impossible to use when there are large and heterogeneous sets of data to be analysed. The availability of large data is directly proportional to the difficulty of developing new predictive models that work accurately. All the traditional statistical solutions work for static analysis, which is limited to the analysis of samples frozen in time. Machine learning has emerged as a solution to all this chaos, proposing different clever alternatives to analyse huge volumes of data. It is able to produce accurate results and analyses by developing various efficient and fast working algorithms for the real-time processing of data.

Some applications of machine learning

Machine learning has been implemented in a number of applications. Some of them are:

1. Google's self-driving car
2. Online recommendation engines such as friend recommendation on Facebook
3. Various offer recommendations from Amazon
4. Cyber fraud detection
5. Optical character recognition (OCR)

The role of machine learning in data science

1. Machine learning helps to analyse large chunks of data easily and hence eases the work of data scientists in an automated process.

2. Machine learning has changed the way data interpretation and extraction works by involving several automatic sets of generic methods, which have replaced statistical techniques.
3. It provides insights that help to create applications that are more intelligent and data-driven, and hence improves their operation and business processes, leading to easier decision making.
4. Machine learning software systems improve their performance more and more, as people use them. This occurs because the algorithms used in them learn from the large set of data generated on the basis of the users' behaviour.
5. It helps in inventing new ways to solve some sudden and abrupt challenges in the system, on the basis of the experience gained by the machine while analysing the large data sets and behaviour of the system.
6. The increasing use of machine learning in industries acts as a catalyst to make data science increasingly relevant.

Some of the machine learning techniques used in data science

1. **Decision tree learning:** This is a machine learning technique that uses a decision tree as the predictive model, which further maps observations about an item to the conclusions about the target value of the item.
2. **Association rule learning:** This is a method used for discovering several interesting relations between the variables in large databases.
3. **Artificial neural networks:** Such learning techniques are also called neural networks. These are learning algorithms that are inspired by the structure and functional aspects of biological neural networks. Different computations are actually structured in terms of interconnected groups of artificially designed neurons, which help to process information using the connectionist approach to computation. All the modern neural networks are basically non-linear statistical tools used for data modelling. They are usually used to model several complex relationships between the inputs-outputs and to find patterns in the data.
4. **Inductive logic programming (ILP):** This approach uses logical programming as a representation for several input examples, the background knowledge and the hypotheses. If we are given an encoding of any known background knowledge with a set of examples, which represent a logical database of facts, then an ILP system will easily derive a hypothesised logic program which entails all the positive and no negative examples. This type of programming considers any type of programming language for representing the hypotheses, such as functional programs.
5. **Clustering:** Cluster analysis is a technique used for the assignment of a set of different observations into various subsets (also called clusters) so that the observations present within the same cluster are similar to some pre-designated criteria, whereas the observations drawn from all the different clusters are dissimilar. All the different clustering techniques have different assumptions on the structure of data, which is often defined by some similarity metric and is evaluated, for example, by internal compactness and the separation between different clusters. It is basically an unsupervised learning method and one of the common techniques used for statistical data analysis.
6. **Bayesian networks:** A Bayesian network is a probabilistic graphical model which represents a set of random variables and all their conditional independencies using a directed acyclic graph. For instance, a Bayesian network can represent the probabilistic relationships between different diseases and their symptoms. If we are given the symptoms, then the network can easily compute the probabilities of the presence of various diseases. Very efficient algorithms are used to perform the inference and learning.
7. **Reinforcement learning:** This is a technique related to how an agent ought to take different actions in an environment in order to maximise some notion of the long-term reward. This type of algorithm attempts to find a policy that maps different states of the world to the different actions the agent ought to take in those states. This type of learning differs from the supervised learning problem, for which correct input/output pairs are not presented, nor are the sub-optimal actions explicitly corrected. END

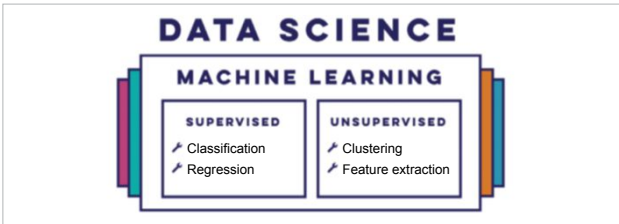


Figure 5: Machine learning and data science (*Image credit:* googleimages.com)

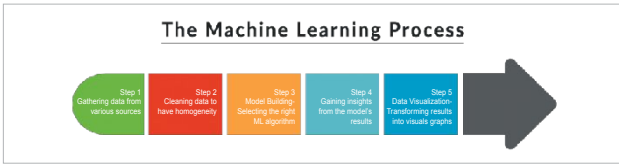


Figure 6: The machine learning process (*Image credit:* googleimages.com)

References

[1] *http://www.wikipedia.org/*
[2] *http://catalyst.nejm.org/*
[3] *https://techdecisions.co/*
[4] *https://www.simplilearn.com*

By: Vivek Ratan

The author has completed his B. Tech in electronics and instrumentation engineering, and currently works at Infosys, Pune. He can be reached at *ratanvivek14@gmail.com.*

Autofetch the last argument

So often, we copy the last command's argument and paste it for the next command. We do this task manually by selecting the argument, and then copying and pasting it.

Here is a time-saving tip that I'm sure you will love. Check the following example:

```
mkdir /home/cg/root/maulik
```

The *mkdir* command is to create a new directory. The next task is to go inside this directory. Most people will copy the path, write the *cd* command and then paste it.

Here is another simpler approach. Use '!$', which will fetch the last argument automatically. The other option is to use the 'Esc + .' command after *cd*.

Note: '!$' or 'Esc + .' will fetch the last argument for any command. If the last command has no argument, then these will fetch the last command.

```
bash-4.3$
bash-4.3$ pwd
/home/cg/root
bash-4.3$
bash-4.3$
bash-4.3$ mkdir /home/cg/root/maulik
bash-4.3$ cd !$
cd /home/cg/root/maulik
bash-4.3$ pwd
/home/cg/root/maulik
```

—Maulik Parekh. maulikparekh2@gmail.com

Retain the remote terminal using the screen

If we happen to close the remote terminal session, we can still run the task on the remote terminal without an issue by using the 'screen' tool.

We need to have *screen* installed on the remote Linux machine so that we can use it.

```
$apt-get install screen
```

Start a screen session on the remote machine, as follows:

```
$screen -S name
```

After a disconnect to the remote Linux machine, run the following command to get back to the same screen terminal that was being used:

```
$screen -dr name
```

Here, 'name' is the screen session's name. To know the list of the active screen sessions, use the following command:

```
$screen -list
```

—Vasanthakumar T, vasanth_22k@rediffmail.com

Using the long list command without *ls -l*

You can use the undocumented long list or the *ll* (double l) command as follows. If appended with the directory names, it will produce a long list of the following:

1. All the items present in appended directories.
2. The hidden files as well.
3. The space occupied by directories.
4. The directories in sequence, ranging from those with the least files to those with the most files.

On many distros, this won't work, in which case, you can create an alias *ll* of *ls –l*.

—Tejas Rawal, tejasprawal@gmail.com

Disk space usage by 'n' number of items in a file

To find out the disk usage by 'n' number of items in a

file, use the following commands:

```
#cat /tmp/test.txt
name1
name2
foldername/name3
...
nameN
foldernameN

#for i in `cat /tmp/test.txt` ;do du -sh $i ; done
```

—Remin Raphael, remin13@gmail.com

A command line hack for crontab

This small script installs itself in crontab if the script is already not there.

I had a script *[/tmp/script.sh]* that I needed to run in crontab, but as it was under development, every now and then the version got updated, and I had to recopy the script and add it.

The following script helped me a lot.

```
#crontab -l | grep -q ' bash /tmp/script.sh '  && \
echo "exists" || cat <(crontab -l)  \
<(echo "0 */2 * * * bash /tmp/script.sh ") | crontab -"
```

You can modify this simple script to suit your needs.

—Murali Sridhar, muraleesridhar@gmail.com

How to find a file larger than 1GB, 100MB or any size

We often need to get the listing of a directory and fetch files that are bigger than, say, 1GB or 100MB. Here's a tip that you can use for this. Run the following commands with root access.

* For file sizes of 1GB or more, use the command:

```
#sudo find / -xdev -type f -size +1G -exec ls -lh {} \;
```

* For file sizes of 100MB or more, give the following command:

```
#sudo find / -xdev -type f -size +100M -exec ls -lh {} \;
```

—Rupin Puthukudi, rupinmp@gmail.com

Disabling the touchscreen in Ubuntu

Many of the latest laptops come with a touchscreen. Sometimes we need to disable or enable the touch facility for a particular purpose.

Here is a command that will help you to do that, if you are using Ubuntu:

```
$xinput set-prop 10 'Device Enabled' 0
```

For enabling the touchscreen, type:

```
xinput set-prop 10 'Device Enabled' 1
```

—Sethu Ram, dycmiti@konkanrailway.com

How to change a user's password in a non-interactive way

If you want to change a password without prompting for it, say, in scripts, use the following command:

```
$ sudo echo -e "password\npassword" | passwd testuser
```

—Krishna Murthy Thimmaiah, gtk321@rediffmail.com

Increase the size of your 'history'

The 'history' command displays a list of the previous commands you used on the terminal. By default, it stores only the last 1000 commands that you used. Often, we need to store more than 1000 commands. To enable this, we have to change the following lines in the '.bashrc' file in the home folder.

To open the '.bashrc' file, use the following command in the terminal:

```
$vi ~/.bashrc
```

…and change the following two lines:

```
HISTSIZE=1000
HISTFILESIZE=2000
```

For instance, if you want to increase your history size to 40000, then these two lines should be:

```
HISTSIZE=40000
HISTFILESIZE=40000
```

Now, save the file.

—Nagaraju Dhulipalla, nagarajunice@gmail.com

END

Share Your Linux Recipes!

The joy of using Linux is in finding ways to get around problems—take them head on, defeat them! We invite you to share your tips and tricks with us for publication in *OSFY* so that they can reach a wider audience. Your tips could be related to administration, programming, troubleshooting or general tweaking. Submit them at *www.opensourceforu.com*. The sender of each published tip will get a T-shirt.

DVD OF THE MONTH

Linux operating systems and programs that you can use every day.

Linux Mint 18.2 Cinnamon (64-bit live)

Linux Mint is a modern, elegant and comfortable operating system, which is both powerful and easy to use. It is one of the most popular desktop Linux distributions and is used by millions of people. Linux Mint 18.2 is a long-term support (LTS) release, which will be supported until 2021. The version bundled with this DVD is the Cinnamon desktop edition.

PCLinuxOS 64 MATE 2017.07 Desktop

Everything you do with 'that other OS', you can do with PCLinuxOS. It comes with a complete Internet suite for surfing the Net, sending and receiving email, instant messaging, blogging, tweeting and watching online videos. The bundled ISO image can be found in the *other_isos* folder on the root of the DVD.

Linux Lite 3.6 32-bit

The desktop is clean and simple. With easy access to the menu, system settings and configuration, navigating around and doing your computing in Linux Lite is simple and intuitive. It comes with a complete set of software to simplify your daily computer usage. The bundled ISO image can be found in the *other_isos* folder on the root of the DVD.

Clonezilla 2.5.2.17

Clonezilla is a partition and disk imaging/cloning program. It helps you to do system deployment or bare metal backup and recovery. The bundled ISO image is of the live edition, and is suitable for single machine backup and restore.

GParted 0.29.0

GParted is a free partition editor to graphically manage your disk partitions. With GParted, you can resize, copy and move partitions without data loss. It enables you to grow or shrink your *C:* drive, create space for new operating systems or attempt data rescue from lost partitions.

What is a live DVD?

A live CD/DVD or live disk contains a bootable operating system, the core program of any computer, which is designed to run all your programs and manage all your hardware and software.

Live CDs/DVDs have the ability to run a complete, modern OS on a computer even without secondary storage, such as a hard disk drive. The CD/DVD directly runs the OS and other applications from the DVD drive itself. Thus, a live disk allows you to try the OS before you install it, without erasing or installing anything on your current system. Such disks are used to demonstrate features or try out a release. They are also used to test hardware functionality, before actual installation. To run a live DVD, you need to boot your computer using the disk in the ROM drive. To know how to set a boot device in BIOS, please refer to the hardware documentation for your computer/laptop.

Dell recommends Windows 10 Pro.

GST READY DELL VOSTRO PC

AVAIL DELL GST OFFERS AND BECOME A DIGI VYAPARI TODAY.

OFFERS AVAILABLE#

2 YEARS ADDITIONAL
NEXT BUSINESS DAY WARRANTY~

WORTH **Rs. 6990.00**
FOR **Rs. 1490.00**

BUSY **GST SOFTWARE**
1 YEAR LICENSE

WORTH **Rs. 7200.00**
FOR **Rs. 2499.00**

1 YEAR ACCOUNTING
'CLOUD' SERVICES

WORTH **Rs. 2499.00**
FOR **Rs. 499.00**

CANON WIFI
PRINTER

WORTH **Rs. 4495.00**
FOR **Rs. 1999.00**

Vostro 5568 Notebook

- 7th Gen Intel® Core™ i5-7200U processor
- Windows 10 Home (Dell recommends Windows 10 Pro)
- Office Home & Student Edition
- 8GB DDR4 RAM
- 1TB* HDD
- 39.6 cm (15.6) Full HD Screen

Vostro 3268 Desktop

- 7th Gen Intel® Core™ i5-7400 processor
- Windows 10 Home (Dell recommends Windows 10 Pro)
- Office Home & Student Edition
- 4GB DDR4 RAM
- 1TB* HDD
- 3-Year Onsite Service Retail warranty

Dell Vostro PC range starts at Rs. 21635.00 onwards.**

For more details, visit: www.gstreadywithdell.com

Dell.co.in/buyonline

Powered by 7th Gen Intel® Core™ i5 processor.

Intel Inside®.
Powerful
Productivity
Outside.

intel CORE i5 7th Gen

R N I No. DELENG/2012/49440, Mailed on 27/28th of Advance month Delhi Postal Regd. No. DL(S)-01/3443/2016-18
Published on 27th September 2017 Licenced to Post without Pre-Payment Licence No U(S)-56/2016-17

Loonycorn

Some of our recent courses:

On Udemy:

- GCP: Complete Google Data Engineer and Cloud Architect Guide
- Time Capsule: Trends in Tech, Product Strategy
- Show And Tell: Sikuli - Pattern-Matching and Automation

On Pluralsight:

- Understanding the Foundations of TensorFlow
- Working with Graph Algorithms in Python
- Building Regression Models in TensorFlow