Machine Intelligence at Google

Pınar Uğurlu Kirazcı
Cloud Customer Engineer
@pinaru
“When I look ahead to where computing is heading, it’s clear to me that we are evolving from a mobile-first to an AI-first world.”

– Sundar Pichai, Google CEO
The basics of ML
Artificial Intelligence
The science of making things smart

Machine Learning
Building machines that can learn

Neural Network
A type of algorithm in machine learning
Deep learning: applying ML to unstructured data

- **AI**
- **Machine learning**: making prediction for unseen data, using past data

**Deep learning**: handling unstructured data (image, video, audio, natural language, etc.) for machine learning
Deep Neural Networks Step 1: Training

(CAT) 
(Labeled Photos) 
(DOG)
Deep Neural Networks Step 2: Testing

Is this a Cat or Dog?

Deep neural network diagram showing layers of activated neurons and output layer.
Keys to Successful Machine Learning

- Large Datasets
- Good Models
- Lots Of Computation
ML in Products
1Billion Users each
Google app

20% of mobile searches are via voice
12% of all responses sent on mobile
Recent Translate improvements

Translation quality

Translation model

Perfect translation

Human
Neural (GNMT)
Phrase-based (PBMT)

Translation model

old: PBMT
new: GNMT

ML for Everyone
Machine Learning services on Google Cloud

Custom ML models

TensorFlow

Machine Learning Engine

Pre-trained ML models

Vision API

Speech API

Jobs API

Translation API

Natural Language API

Video Intelligence API
What is TensorFlow?

Google's open source library for machine intelligence

tensorflow.org launched in Nov 2015
Used by many production ML projects
TensorFlow Supports Many Platforms

CPU
GPU
TPU
Cloud TPU

Android
iOS
<table>
<thead>
<tr>
<th>Framework</th>
<th>GitHub Star Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>TensorFlow</td>
<td>44508</td>
</tr>
<tr>
<td>scikit-learn</td>
<td>16191</td>
</tr>
<tr>
<td>Caffe</td>
<td>15690</td>
</tr>
<tr>
<td>CNTK</td>
<td>9383</td>
</tr>
<tr>
<td>MXNet</td>
<td>7896</td>
</tr>
<tr>
<td>Torch</td>
<td>6285</td>
</tr>
<tr>
<td>Theano</td>
<td>5568</td>
</tr>
</tbody>
</table>
Cloud Machine Learning Engine

Train Your Own Models

Fully Managed ML Infrastructure
Flow to build a custom ML model

1. Identify business problem
2. Develop hypothesis
3. Acquire + explore data
4. Build a model
5. Train the model
6. Apply and scale
READY TO USE MODELS

Cloud Natural Language API

Cloud Vision API

Cloud Speech API

Cloud Translate API

Cloud Video Intelligence API
CLOUD VIDEO INTELLIGENCE API

Label Detection
Detect objects, such as dog, flower, human, in the video.

Face Detection
Detect faces throughout the length of the video.

Shot Change Detection
Detect scene changes within the video.

Video Segmentation
Segment long-running videos to provide annotations for specified time segments.

Integrated REST API
Request one or more annotation types per image.

Regionalization
Specify a region where processing will take place (for regulatory compliance).
CLOUD NATURAL LANGUAGE API

Entity Recognition
Identify entities and label by types such as person, organization, location, events, products and media.

Sentiment Analysis
Understand the overall sentiment expressed in a block of text.

Multi-Language Support
Enables you to easily analyze text in multiple languages including English, Spanish and Japanese.

Syntax analysis
Extract tokens and sentences, identify parts of speech (PoS) and create dependency parse trees for each sentence.
CLOUD SPEECH API

Speech Recognition
Powered by deep learning neural networking to power your applications.

Word Hints
Can provide context hints for improved accuracy. Especially useful for device and app use cases.

Noise Robustness
No need for signal processing or noise cancellation before calling API. Can handle noisy audio from a variety of environments.

Real-time results
Can stream text results, returning partial recognition results as they become available. Can also be run on buffered or archived audio files.

Over 80 languages
Recognizes over 80 languages & variants. Can also filter inappropriate content in text results.
CLOUD TRANSLATION API

Text Translation
Supports more than 100 languages and thousands of language pairs.

Automatic language detection
Sometimes you don’t know your source text language in advance. Can automatically identify languages with high accuracy.

Continuous Updates
Behind the scenes, Translation API is learning from logs analysis and human translation examples. Existing language pairs improve and new language pairs come online at no additional cost.

Premium edition BETA
Tailored for users who need precise, long-form translation services (e.g. livestream translations, high volume emails, articles and documents)
DIALOGFLOW

Beta Q4

- Reduce commerce and support costs
- Customer experience improves with ML
- No coding required
Results Enabled by Google Cloud

- **Faster time to insights** - Customer and product analytics up to 35x faster
- **Increase customer stickiness** - Used to only have weekly playlists. Now users can get up to seven playlists customized for them per day.

“Queries that used to take 20 minutes now take seconds. Faster results means more queries can be done, resulting in more finely tuned recommendations.”

Wouter De Bie, Data Architect, Spotify
ML for Enterprises
The Challenge: Computing Power

DNN requires large training datasets

Large models doesn't fit into a GPU

Requires try-and-errors to find the best design, configs and params

Need to spend a few days or weeks to finish a training
Google Cloud is
The Datacenter as a Computer
2nd generation Tensor Processing Unit for TensorFlow

Designed by Google

Each Cloud TPU provides up to 180 teraflops of performance as well as 64 gigabytes of ultra-high-bandwidth memory.
Cloud Machine Learning Engine (ML Engine)

- Fully managed **distributed training** and **prediction**
- Scales to **tens of CPUs and GPUs**
- Supports custom **TensorFlow** graphs
- **HyperTune** for hyper-parameter tuning automation
- [cloud.google.com/ml](http://cloud.google.com/ml)
Machine Learning helped Airbus reduce error rates from 11% to 3% in the critical process of correcting satellite image maps.
AXA: finding “large loss” car accidents

About 7% - 10% of insured drivers cause a car accident every year

1% of all car accidents require insurance payment over $10K

Predict if a driver may cause a large loss case during the insurance period

-- Tensorflow gives 78% accuracy

Enables better pricing
ML solution for real-world problems

Inputs → Preprocessing → Asset Creation → Train model → Model + Assets

Distributed training, Hyper-parameter tuning

Deploy: Including Model Versioning

Prediction: load balanced, auto-scaled, scale-to-zero

Remote Clients → Cloud ML

REST API call with input variables
Building systems that know how to walk, with no reference to how humans actually walk.
**COMPUTE**
From virtual machines with proven price/performance advantages to a fully managed app development platform.
- Compute Engine
- App Engine
- Container Engine
- Container Registry
- Cloud Functions

**STORAGE**
Scalable, resilient, high performance object storage and databases for your applications.
- Cloud Storage
- Cloud Bigtable
- Cloud Datastore
- Cloud SQL
- Cloud Spanner

**NETWORKING**
State-of-the-art software-defined networking products on Google's private fiber network.
- Cloud Virtual Network
- Cloud Load Balancing
- Cloud CDN
- Cloud Interconnect
- Cloud DNS

**MANAGEMENT TOOLS**
Monitoring, logging, and diagnostics and more, all an easy to use web management console or mobile app.
- Stackdriver Overview
- Monitoring
- Logging
- Error Reporting
- Debugger
- Deployment Manager & More

**BIG DATA**
Fully managed data warehousing, batch and stream processing, data exploration, Hadoop/Spark, and reliable messaging.
- BigQuery
- Cloud Dataflow
- Cloud Dataproc
- Cloud Dataprep
- Cloud Datalab
- Cloud Pub/Sub
- Genomics

**MACHINE LEARNING**
Fast, scalable, easy to use ML services. Use our pre-trained models or train custom models on your data.
- Cloud Machine Learning Engine
- Vision API
- Video Intelligence API
- Speech API
- Translate API
- NLP API

**DEVELOPER TOOLS**
Develop and deploy your applications using our command-line interface and other developer tools.
- Cloud SDK
- Deployment Manager
- Cloud Source Repositories
- Cloud Endpoints
- Cloud Tools for Android Studio
- Cloud Tools for IntelliJ
- Google Plugin for Eclipse
- Cloud Test Lab
- Cloud Container Builder

**IDENTITY AND SECURITY**
Control access and visibility to resources running on a platform protected by Google’s security model.
- Cloud IAM
- Cloud IAP
- Cloud KMS
- Cloud Resource Manager
- Cloud Security Scanner
- Cloud Platform Security Overview
cloud.google.com
console.google.com
How will Artificial Intelligence affect crime, war, justice, jobs, society and our very sense of being human? The rise of AI has the potential to transform our future more than any other technology—and there’s nobody better qualified or situated to explore that future than Max Tegmark, an MIT professor who’s helped mainstream research on how to keep AI beneficial.
What is the difference between supervised and unsupervised machine learning?
Thank You!