

特集 「AI 計算資源」

# 人工知能・機械学習の応用・研究に適したパブリッククラウドサービス

## Utilizing Public Cloud Services for Artificial Intelligence / Machine Learning Research and Applications

中田 秀基 産業技術総合研究所  
 Hidemoto Nakada National Institute of Advanced Industrial Science and Technology.  
 hide-nakada@aist.go.jp, http://staff.aist.go.jp/hide-nakada/

**Keywords:** machine learning, artificial intelligence, public cloud.

### 1. はじめに

急速に普及しつつあるディープラーニングに代表される近代的な機械学習・人工知能の研究・応用には、膨大な量の計算が必要となる。特に学習時のハイパーパラメータチューニングには広大な空間を探索する必要があり、計算負荷が非常に高い。これを個々のユーザが手元に準備した計算機資源でまかなうことは一般に困難である。この問題を解決する一つの方法として、パブリッククラウドを用いる方法がある。パブリッククラウドとは、クラウド事業者が運用し、一般に公開している計算資源やソフトウェアスタックをインターネットを介して利用することを意味する\*1。

パブリッククラウドは2000年代後半から存在し、特にWebサービスの領域では広く利用されている。機械学習の領域でも、単純な計算機資源としてのパブリッククラウドの利用は新しいものではなく、すでに広く行われている。2017年にはNIPSの締切り直前の時期にパブリッククラウドサービスのGPU資源が尽きたという噂もある [Register 17]。

しかし、昨今では単なる計算機資源ではない、より高度なサービスがさまざまなクラウド事業者から提供されている。これらをうまく活用することで、より効率的な研究・応用が可能になる可能性がある。本稿では機械学習・人工知能の研究・応用に適したパブリッククラウドサービスを紹介し、読者の一助としたい。

次章ではクラウドのサービス形態による分類を行う。その後の節では各分類について現在利用できるサービスを示し、その得失を議論する。

### 2. クラウドの分類

まず、一般論としてクラウドの種別を整理しておく。[NistCloud1 11]によれば、クラウドのサービス形態には三つの形態がある。IaaS (Infrastructure as a Service), PaaS (Platform as a Service), SaaS (Software as a Service) の三つである。これらは、クラウド事業者が提供するレイヤに着目した分類である。図1にこの三つを示す。

IaaSはいわゆる仮想計算機であり、仮想化したハードウェアそのものを提供する、ある意味で最も原始的な形態のクラウドである。代表的な例としてはAmazon Web ServicesのEC2 (Elastic Compute Cloud) がある。利用者がOSをインストールし、アプリケーションプラットフォーム、アプリケーションですべてを実装しなければならないため、利用者の負荷が大きい。基本的には普通の計算機でできることであればなんでもできるため自由度が高い。一方で、ソフトウェアスタックはすべてユーザが提供するため、クラウドベンダ固有の高機能なイン

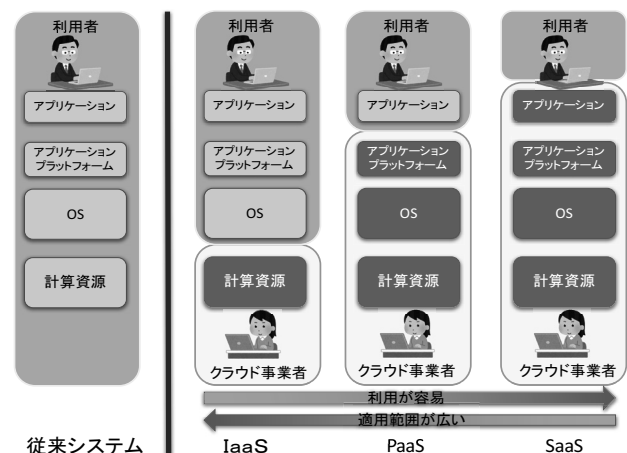


図1 クラウドの分類

\*1 パブリッククラウドに対置される語として、企業内での計算資源共有を行うプライベートクラウドがあるが、本稿では扱わない。

表1 機械学習に適した GPU をもつ IaaS

サービス名	タイプ名	GPU 機種	GPU 数
Amazon Web Services	p2	Tesla K80	1/2~8
	p3	Tesla V100	1~8
Goggle Cloud Platform	K80	Tesla K80	1/2~4
	P100	Tesla P100	1~4
Microsoft Azure	NC	Tesla K80	1~4
	NCv2	Tesla P100	1~4
さくらインターネット	Quad GPU Pascal	Titan X (Pascal)	4
	Tesla P100	Tesla P100	1

フラサービスの利用は難しい。

これに対して SaaS はインターネット越しにユーザが直接利用する完全なサービスを指す。Google の GMail や Microsoft の Office 365 などが代表的である。この場合、すべてのスタックをクラウド事業者が提供しているため、ユーザの負荷は軽く、クラウド事業者固有の高機能なインフラサービスを利用した高度なサービスを楽しむことができる。ただし、利用できるサービスはクラウド事業者が決定するため、限定されてしまう。

PaaS はこの二つの中間にあたる。アプリケーションフレームワークまでをクラウド事業者が提供し、ユーザはその上にアプリケーションを構築する。代表的な例としては Google の Web サービス構築クラウドである App Engine があげられる。このアプローチは、ユーザにある程度の自由度を与えつつ、クラウド事業者固有の高機能なインフラサービスを利用することができるという特徴をもつ。

### 3. IaaS (Infrastructure as a Service)

#### 3.1 人工知能・機械学習に有用な IaaS の使い方

IaaS は仮想計算機を提供する。したがって IaaS の利用は通常の計算機を利用するのとあまり変わりはない。違うのは、起動時方法ぐらいで、あとはリモートにあるクラスタを利用するのと同じだ。

しかし運用の考え方は大きく異なる場合がある。例えばハイパーパラメータのチューニングをする場合を考えてみよう。手元にあるサーバであれば、ハイパーパラメータの組合せを逐次的に試していくしか方法はない。

しかしクラウドでは同時に利用できる計算資源の数に制約はない（もちろん現実的にはさまざまな制約はあるが）。100 個の組合せを 100 台の仮想計算機で同時に評価することができるのだ。もちろんそれだけ使用料はかかるわけだが、ここで重要なのは、100 個の組合せを逐次で順番に実行しても並列に同時実行しても、かかる使用料は同じ、ということだ。これをうまく利用すればハイパーパラメータチューニングに掛かる時間を大幅に短縮することができる。

課金の単位時間は、1 時間となっている事業者が一般

的である。しかし最近では 1 分単位、さらには 1 秒単位の課金を導入している事業者もある。仮想計算機の起動と終了を細かく制御することで課金額を抑制することもできる。

また、特定の仮想計算機種の利用者が少ない場合に、比較的低額で利用できる機構を導入している事業者もある。ユーザは「この値段でなら使いたい」という値段を指定しておく、クラウド事業者が自動的に一番高い値段を指定したユーザを選択して、そのユーザの仮想計算機を起動する。一方、正規料金での利用が増えてきた場合には、仮想計算機を停止するのでユーザはそれに備えて適宜チェックするようシステムを構成しなければならない。

もう一つのポイントは GPU が利用できることだ。多くのアプリケーションで利用が進められているディープラーニングの学習、推論には GPU が非常に適しており、ディープラーニングの研究・応用には、GPU の利用が事実上不可欠となっている。一方で GPU を利用するためのインシヤルコストは安くなく、多数の GPU サーバを導入することは、どの組織にもできることではない。パブリッククラウドを用いることでこの問題を解決することができる。後述のように現在では多くのパブリッククラウドサービスが GPU を提供している。初期コストゼロで利用することができる。

#### 3.2 機械学習に適した IaaS

IaaS サービスは国内外多くのクラウド事業者が提供しているが、その多くは Web サーバなどの比較的計算負荷が小さいアプリケーションを主要なターゲットとしており、機械学習に適した IaaS は限定される。GPU を提供する機械学習に適した IaaS の代表的なサービスを表 1 に示す\*2。デスクトップ環境の高速化のための GPU をもつサービスもあるが、機械学習に適さないとされるものは掲載していない。

価格に関しては利用形態によって大きく異なるのであえて掲載していないが、1 時間当たり安いもので 100 円

\*2 2017 年 11 月末現在。

程度、高いものでも数百円程度である。三大クラウド事業者の Amazon Web Services [AWS], Microsoft Azure [Azure], Google Computing Platform [GCP] がそれぞれ GPU が利用可能なインスタンスを提供している。Amazon Web Services は NVIDIA の最新 GPU である V100 の提供を早くも開始した。国内の事業者では、さくらインターネットの高火力コンピューティング [Sakura] と呼ばれるインスタンスが GPU を提供している。

単一ノードに複数の GPU が搭載されているノードも多数提供されており、複数 GPU による並列学習を行うことも可能だ。さらに、Microsoft Azure は高速ネットワークが搭載されたインスタンスも提供しており、複数ノードを用いた並列学習を高速に行うこともできる。

各クラウド事業者は、これらの GPU インスタンスで利用できる仮想計算機イメージを提供している。これらのイメージには最新の GPU を利用して計算を行うためのソフトウェアスタックや、主要なディープラーニングフレームワークがインストールされている。このようなイメージを用いることによって、環境設定に要する時間を短縮することができる。これもクラウドを利用することの利点であるといえるだろう。

## 4. SaaS (Software as a Service)

### 4.1 機械学習向け SaaS

機械学習向けの SaaS とは、何らかの機械学習による予測推論を行うサービスである。多くの場合 Web サービスとして API を公開しており、ユーザはこの API を介して機械学習推論を利用することができる。

このようなサービスは新しいものではなく、Google の Prediction API は 2011 年の 5 月に公開されている。このサービスはクラス分類とリグレッションの双方をサポートしており、学習データをストリーミングもしくはバルクで与えると、新しいデータに対する予測ができる<sup>\*3</sup>。しかし、この Prediction API のモデルは比較的単純で、入力データとして低次元のデータしか想定しておらず、サービスの眼目は学習そのものよりも大量データのハンドリングにある。

これに対し、最近登場した機械学習サービスははるかに高度なアルゴリズムとなっている。また大規模なデータで学習した結果のモデルをサービスとして提供している。このようなサービスは、API を介してアプリケーションに容易に組み込むことができる。複数事業者の異なるサービスを利用したアプリケーションを構築することも容易だ。図 2 にその様子を示す。

一般に SaaS の課金は API の呼出し回数に対して行わ

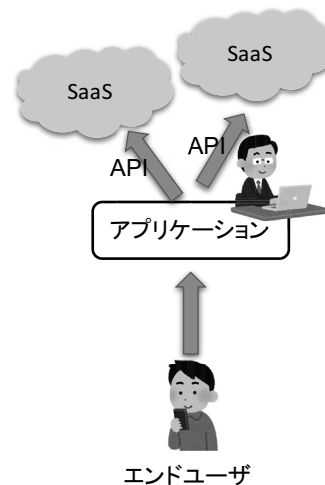


図 2 アプリケーションへの SaaS クラウドサービスの組み込み

れる。固定費用がかからないため、アプリケーションへの導入が容易である。また、アプリケーションの利用者数が増えた場合には、SaaS の側が自動的にスケールアウトしてくれるため、アプリケーション構築者は特に配慮する必要がないことも大きなメリットだ。

SaaS の問題点としてはカスタマイズの余地が少なく、要求する仕様に合致したサービスがあるとは限らないことがあげられる。しかし、合致したサービスがありさえすれば、独自に機械学習アルゴリズムを構築して学習させることなく、高度な機械学習アルゴリズムを簡単に利用することができる。以下に代表的なサービスを示す。

### 4.2 画像認識サービス

画像認識サービスはいわゆる物体認識を行うサービスである。画像データをアップロードすると、それに写っているものやそのカテゴリーをリストとして出力する。このようなサービスには多くの用途が考えられる。例えば blog サービスを運用する場合、掲載に適さないような画像が含まれていればそれを削除する必要がある。独自の機械学習モデルをトレーニングしてこれを実現することも可能だが、不適切な画像にタグ付けしてくれるサービスを利用すればはるかに安価、高速にシステムを構築することができる。

画像認識サービスを提供する主なクラウド事業者とそのサービスを表 2 に列挙する。

クラウド事業者によっては、単なる物体認識だけでなくより高度な認識サービスを提供している。

- 画像中の顔抽出と感情推定
- 画像中の著名人の同定
- 製品ロゴの検出
- 風景写真からのランドマーク認識
- 画像中の文字列認識

● ユーザデータを用いた物体認識のファインチューン  
これらの高度な機能を独自に実装するのは容易なことではない。クラウドサービスを用いることで、このよう

\*3 ちなみに、このサービスは Google の他のサービスに置き換えられ、2018 年 4 月に早くもシャットダウンされる予定となっている。

表2 機械学習のSaaS

サービス種	事業者名	サービス名	特徴
画像認識サービス	Amazon Web Services	Amazon Rekognition	著名人認識
	IBM	Visual Recognition	著名人認識
	Google Cloud Platform	Vision API	画像中の文字列抽出・認識が強力
	Microsoft Azure	Computer Vision API	ユーザーデータを用いたカスタマイズ
音声認識 発話サービス	Amazon Web Services	Polly	発話サービス
	IBM Watson	Speech to Text	認識サービス
		Text to Speech	発話サービス
	Google Cloud Platform	Cloud Speech API	認識・発話双方
	Microsoft Azure	Bing Speech API	認識・発話双方
対話サービス	Amazon Web Services	Amazon Lex	独立したサービスだが 2016年にGoogleにジョイン
	IBM	Conversation	
	Google	api.ai	
	Microsoft Azure	Bot Framework	
翻訳サービス	IBM Watson	Language Translator	音声翻訳
	Google Cloud Platform	Cloud Translation API	
	Microsoft Azure	Translator Speech API	
		Translator Text API	
動画処理サービス	Google Cloud Platform	Cloud Video Intelligence	音声も用いた高度なサービス
	Microsoft Azure	Video Indexer	

な高度な技術を誰でも容易に利用できるようになる。このことの産業的なインパクトは非常に大きい。

特筆すべきサービスとして、Microsoft AzureのCustom Vision Serviceがあげられる。このサービスでは、システムが提供する学習済みのモデルに対して、ユーザが提供するデータを用いたファインチューンを行うことができる。システムが提供するモデルが著名人の識別しかできない場合に、特定の人物を追加で識別できるように学習させることができる。

#### 4.3 音声認識・発話サービス

音声認識、音声合成は、音声を用いたユーザインタフェースを構築するうえで、非常に重要な要素技術である。多くのクラウド事業者がこれらのサービスを構築している。

音声認識サービスは音声データを受け取って、テキストを返す。この際言語を指定する必要がある。対応している言語の数はサービスによって異なるが、例えばGoogleのサービスでは110もの言語とその方言をサポートしている。英語だけでも13の方言がサポートされている。

音声合成サービスはテキストから音声データを合成する。エフエム和歌山がAWSの音声合成サービスを用いた「AIアナウンサー」を運用し始めたというニュースは記憶に新しい[ITmedia 17]。年間費用がわずか1,000円とのこと。この驚くべき低費用は、この種のパブリッククラウドサービスの優位性を端的に表しているといえるだろう。音声合成は古くからある技術だが、自然な抑揚での合成は容易ではなく、昨今の人工知能研究の成果

が生かされている。

変わり種としてはAzureのSpeaker Recognition APIがある。このサービスは音声の内容を認識するのではなく、音声を発話している話者を認識する。認証に使えるとのことなのでそれなりの精度が出ているものと思われる。

また、Azure Custom Speech Serviceでは、ユーザがアプリケーションに固有の言語モデル、音響モデルを与えることが可能となっている。

#### 4.4 対話サービス

何らかの情報を対話的にエンドユーザから取得するアプリケーションを構築するためのサービスである。例えば、レストランの予約を受け付けるアプリケーションを考えてみよう。アプリケーションはエンドユーザから、予約日時、人数、名前、連絡先を所得する必要がある(図3)。このような情報を自然な対話で引き出すアプリケーションを構築するサービスが対話サービスである。

ユーザ(アプリケーション構築者)は、対話で取得したい情報のスキーマと、その情報がエンドユーザからの入力のどの部分に現れるかを指定する。サービスは、エンドユーザからの入力を解析して必要な情報を取り出し、情報に不足があれば不足情報の入力を促すような出力を生成する。

対話サービスの一例として、MicrosoftのBot Frameworkがあげられる。このサービスは、テキストベースのチャットボットを作成するもので、電子メール、Facebook、Skype、Slack、SMSなどの多彩なメッセージチャンネルに対応したチャットボットを作成することが

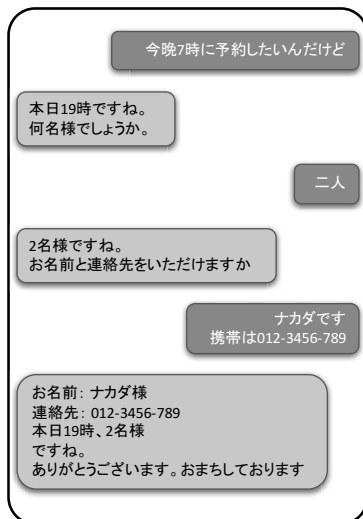


図3 対話アプリケーションのイメージ

できる。

Amazon Lex は音声にも対応したチャットボット作成サービスである。Lex 自体は API からの呼出ししかできず、メッセージアプリケーションなどから直接呼び出すことはできないが、AWS の他のサービスや AWS Lambda (任意のプログラムをサーバを用いずに軽量に実行する枠組み) を組み合わせることで、任意のメッセージチャネルに対応することができるように設計されている。

Google 系では Cloud Platform に直接には含まれていないが api.ai [APIAI] と呼ばれるサービスでこれを実現している。

#### 4.5 翻訳サービス

翻訳サービスは、以前から多くの事業者が Web 上で提供してきたが、SaaS としての翻訳サービスはこれを API として提供することで、容易にアプリケーションに組み込めるようにしたものである。翻訳サービスの品質はディープラーニングを用いた技術が導入されたことによって昨今飛躍的に向上したといわれている。サポートされる言語の数も増大しており、日常的に目にする機会のある言語はほぼすべてカバーされているといつてよい。

多くの翻訳サービスはテキストベースとなっているが、Azure では、Translator Speech API と呼ばれる音声から音声への直接翻訳を行うインタフェースが提供されている。

#### 4.6 動画サービス

静止画ではなく動画を対象として解析を行うサービスも提供されている。動画を自動的にシーンに分割し各シーンごとに、そのシーンの特徴や登場するエンティティを用いてラベル付けする。このような前処理を行うことで、動画に対する検索が可能になる。

このようなサービスとしては、Google の Cloud Video Intelligence や、Microsoft Azure の Video Indexer があ

る。Video Indexer は、動画像だけではなく音声も同時に処理し、人物の認識やトラッキング、動画中のテキストの認識、話者の特定、翻訳まで一括で行う総合的なサービスとなっている。

#### 4.7 その他のサービス

上に述べたカテゴリーに属さないようなサービスも多く提供されており、各クラウド事業者を特徴づけている。

Azure の Linguistic Analysis API は、テキストを構文解析しそのツリーを返すサービスである。現在のところ日本語には対応していないようだ。Google の Cloud Natural Language API も類似のサービスである。

IBM Watson [Watson] の Personality Insights は、個人の性格分析をするサービスである。ある個人が記述した blog や電子メールなどの文章を入力として、個人のパーソナリティを特性 5 因子<sup>\*4</sup>、価値、ニーズの三つに分けて出力するという。Watson には Tone Analyzer というテキストの感情分析を行うサービスも提供されている。これも現在のところ、残念ながら英語のみのサービスとなっている。

Amazon Machine Learning は簡単なクラス分類や回帰を行う機械学習サービスを提供する。似たような機能をもつ後述の Azure Machine Learning Studio と比較すると、プログラミングの自由度がなく、与えられたデータに対する単純なクラス分類と回帰を行うことしかできない。このためここでは SaaS として分類している。このサービスには、データの前処理などは含まれていない。前処理は AWS Lambda などの別サービスで行うことが前提とされているためだ。このため、Amazon Machine Learning を利用するには比較的高い技術が要求される。Azure Machine Learning Studio がプログラムを書けないエンドユーザを対象にしているのとは対照的である。

#### 4.8 使用例

SaaS 型クラウドサービスの使用例として Google Cloud Platform の Vision API を見てみよう。API は Web サービスとして提供されている。つまり、特定の URL に対して HTTP を用いてアクセスすることでサービスを利用することができる。HTTP 上のプロトコルとしては REST と gRPC がサポートされている。Google は C#, Go, Java, JavaScript, PHP, Python, Ruby 用のクライアントライブラリを提供しており、これらを用いると容易に利用することができる。

Python のクライアントライブラリを利用して、Web 上の画像を認識するためのサンプルプログラムを図 4 に示す。わずかなコードで利用できることがわかるだろ

\*4 神経症傾向、外向性、解放性、協調性、誠実性の五つをいうらしい。

```

from google.cloud import vision
from google.cloud.vision import types

url="http://example.com/xxx.jpg"

client = vision.ImageAnnotatorClient()
imageSrc = types.ImageSource(image_uri=url)
image = types.Image(source=imageSrc)

response = client.label_detection(image=image)
for label in response.label_annotations:
    print("{:.2f} : {}".format(label.score,
                               label.description))

```

図4 Pythonを用いた Google Vision API の呼出し

```

0.99 : car
0.98 : land vehicle
0.91 : vehicle
0.88 : sports car
0.88 : motor vehicle
0.87 : automotive design
0.85 : lamborghini miura
0.71 : supercar
0.68 : race car
0.67 : lamborghini

```

図5 Google Vision API の呼出し結果

う。このプログラムを英語 Wikipedia の **Lamborghini Miura** の写真に対して実行した結果を図5に示す。ラベルが確信度とともに得られており、かなり高い確信度で **Miura** であると判断していることがわかる。

ここでは単なる物体認識を行ったが、ほぼ同様の手順で顔の認識も行うことができる。

### 5. PaaS (Platform as a Service)

#### 5.1 機械学習向け PaaS

PaaS は、クラウド事業者がフレームワークまで提供し、その上のアプリケーションをユーザが提供するクラウド提供形態である。任意のアプリケーションを記述できるという自由度を保ちつつクラウド事業者のもつ高度な技術を享受できる形態である。

#### 5.2 Azure Machine Learning Studio

Azure Machine Learning Studio (旧名 Azure Machine Learning, 以下 ML Studio) は、GUI を用いて機械学習パイプラインをクラウド上に構築し、訓練データの読み込み、整形から、機械学習モデルの訓練、訓練済みモデルの Web アプリケーションとしての公開までを一括して行うことのできるプラットフォームである。

基本的なデータ読み込み手法、整形手法、機械学習モデルは実装され提供されているが、ユーザが提供した既存の R や Python のコードを利用することもできる。

ML Studio は GUI に重点をおいて開発されている。Web 上でインタラクティブにグラフとして機械学習パイプラインを記述することができる。図6に簡単な学習パイプラインの様子を示す。

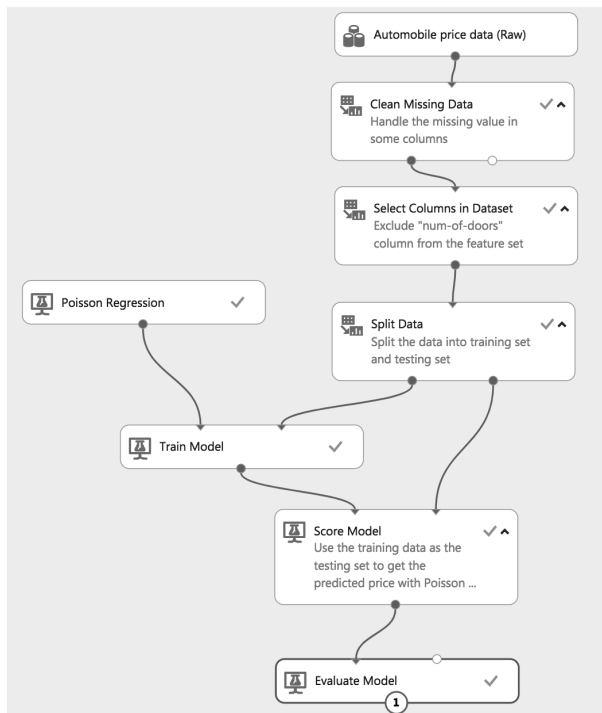


図6 Azure MLにおける機械学習パイプライン

パイプラインの様子を示す。データの流は矢印で表されている。プログラムが全く書けないユーザであっても、GUIを用いることで学習パイプラインを変更したり、モデルのハイパーパラメータの調整をしたりすることができる。

訓練データは Azure 上のさまざまなストレージサービスから読み出すことができる。入力ファイルフォーマットとしては CSV, TSV, SQL データベースなどをデフォルトでサポートしているが、直接サポートしていないファイルフォーマットであっても、Python で読み出しモジュールを記述すれば自由に取り込むことができる。

機械学習アルゴリズムとしては大別して、異常値検出、クラス分類、クラスタリング、回帰の4種類がサポートされており、それぞれに個別のアルゴリズムが提供されている。例えばクラス分類に関しては、ロジスティック回帰、SVM、決定木、ニューラルネットワークが提供されている。ただし、ニューラルネットワークは隠れ層が1層の小規模なものしか利用できない。クラスタリング手法としては K-means がサポートされている。異常値検出には SVM ベースのものと、主成分分析を用いたものがサポートされている。

機械学習パイプラインの構築は、システムが提供するモジュールを並べて、矢印でつなぐことで行う。図6に簡単な機械学習パイプラインを示す。図中の四角はそれぞれモジュールを示している。モジュールにはそれぞれ上部に入力ポートが、下部に出力ポートがある。

一番右上にあるモジュールはデータ入力を行い、その下に続く二つのモジュールで入力されたデータの整形を行っている。整形に必要なパラメータの設定もすべて

GUIで行うことができる。右の上から4番目の **Split Data** と書かれたモジュールは、訓練データとテストデータを分離するモジュールである。下部に出力ポートが二つあり左に訓練データが、右にテストデータが出力される。

左側の **Poisson Regression** と書かれたモジュールが学習に用いるアルゴリズムを示しており、そこから線の伸びた **Train Model** と書かれたモジュールで学習を行う。下から2番めの **Score Model** と書かれたモジュールは学習したモデルの評価を行う。左の入力ポートに学習済みモデルを、右の入力ポートに **Split Data** モジュールからのテストデータを受け取る。一番下の **Evaluate Model** は、**Score Model** の結果を受け取り、最小二乗誤差で評価を行う。この例で示したパイプラインは非常に単純だが、交差検証を行ったり、複数のアルゴリズムの結果を比較するようなパイプラインを記述することも可能だ。

さらに、学習したモデルを用いた予測 **Web** サービスを構築することができる。ボタンをクリックするだけで、**Web** サービスインタフェースとそこからの入力と、学習済みのモデルを用いるパイプラインが生成される。図7に **Azure ML** が生成した **Web** インタフェースを示す。生成された **Web** サービスはもちろん **API** を用いて呼び出すこともできる。

**ML Studio** は比較的少量のデータと軽量のモデルを対象に設計されており、例えば大規模な **CNN** を大容量データを用いて学習する、といったような用途には適していないように思われる。しかし、プログラムを全く書けない人であっても学習パイプラインを設計構築し、**Web** アプリケーションとしてデプロイできることは画期的である。

Request-Response Batch

▼ input1 🔍 🗨

symboling	<input type="text" value="1"/>
normalized-losses	<input type="text" value="1"/>
make	<input type="text"/>
fuel-type	<input type="text"/>
aspiration	<input type="text"/>
num-of-doors	<input type="text"/>
body-style	<input type="text"/>
drive-wheels	<input type="text"/>
engine-location	<input type="text"/>
wheel-base	<input type="text" value="1"/>

図7 Azure ML が生成した Web サービスインタフェース

### 5.3 Google Machine Learning Enigne

**Google Machine Learning Enigne** (以下 **ML Engine**) は、**Google** の機械学習フレームワークである **Tensorflow** を実行する **PaaS** 環境である。ユーザは提供された開発環境を用いて、手元のコンピュータで任意の機械学習プログラムを **Tensorflow** で記述する。小規模データでの稼働を手元で確認したら、開発環境の機能を用いて機械学習プログラムをクラウドにサブミットする。プログラムはクラウド上のキューに格納され、順次実行される。この際、機械学習プログラムの用いる訓練データはクラウド上のストレージに置くことを仮定している。ユーザはプログラム実行の状態をモニタリングし、終了したら結果を取得する。この様子を図8に示す。

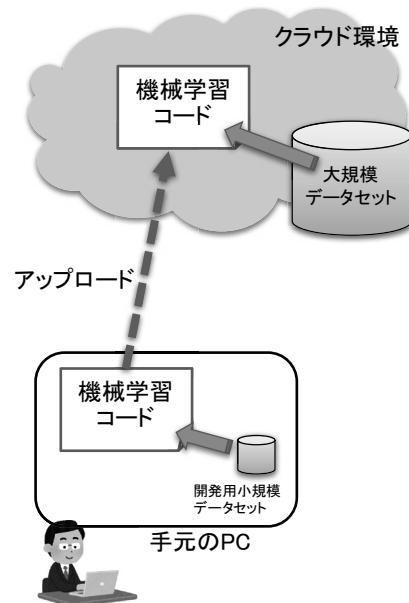


図8 ML Engine を用いた機械学習プログラムの開発と実行

**ML Engine** の利点は実行環境の管理をユーザが行う必要がない点にある。**IaaS** 環境で同様なことをするには、**Tensorflow** とバッチキューイングシステムを自分で運用する必要があるが、**ML Engine** であればすべてをクラウド事業者にまかせ、ユーザはアルゴリズムに専念することができる。また実行環境に対してチューニングされた **Tensorflow** の利用が期待できる。

さらに **Google** は、ディープラーニングに特化されたチップである **TPU (Tensor Processing Unit)** をクラウドで提供することを予定している。これも **ML Engine** の一部として実装されると思われる。

## 6. おわりに

本稿ではさまざまな機械学習・人工知能向けパブリッククラウドサービスを、**IaaS**、**PaaS**、**SaaS** というサービス形態に基づいて整理したうえで、さまざまなサービスを紹介した。これらのパブリッククラウドサービスを

利用することで、初期コストを大きく低減することができる。

特に昨今の SaaS に属するサービスの充実には驚くべきものがある。人工知能・機械学習をアプリケーションに組み込みたい場合には、独自に実装することを考える前に、まずはこれらのクラウドサービスが利用可能かを検討するべきだろう。

Google Machine Learning Engine のような、機械学習フレームワークと一体となった PaaS は、SaaS ではカバーしきれない機械学習アプリケーションの構築に有効だろうと思われる。今後同様のサービスが他の事業者から提供されることが予想される。

IaaS に関しては、初期コストなしで利用を開始できること、導入までの時間を短くできる点大きいメリットだが、稼働率が十分高い場合には、オンプレミスで計算機を保有したほうが一般に安価である点に注意する必要がある。ただし、導入そのものにかかるコストや管理コスト、メンテナンスコストを勘案した場合、フル稼働であっても、クラウドのほうがトータルで安価であるという判断も当然あり得る。このあたりは組織ごとに最適な解は異なるので十分検討する必要がある。

また、パブリッククラウド以外の外部利用できる計算機システムの利用も考慮するべきだろう。大学の共用計算機システムの中には、一定の条件のもとに商用利用を認めているものもある。東京工業大学の TUBAME [Tubame] がその一例である。利用形態にもよるが、これらのシステムは一般にパブリッククラウドよりもはるかに安価なので検討する価値があるだろう。

少し以前までの機械学習・人工知能は比較的小規模な計算機環境で実行されてきたが、現代的な機械学習・人工知能には大規模な計算機環境が必須となった。パブ

リッククラウドの活用がこの変化への適応の一助となることを期待する。

### ◇ 参考文献 ◇

- [APIAI] api.ai, <https://api.ai/>
- [AWS] Amazon Web Services, <https://aws.amazon.com/>
- [Azure] Microsoft Azure, <https://azure.microsoft.com/>
- [GCP] Google Cloud Platform, <https://cloud.google.com/>
- [ITmedia17] ITmedia ビジネス ONLINE 「AI アナウンサー」年間 1000 円の衝撃, <http://www.itmedia.co.jp/business/articles/1710/12/news020.html>
- [NistCloud1 11] The NIST Definition of Cloud Computing, Technical Report SP 800-145, National Institute of Standards and Technology (2011)
- [Register17] Cloud giants 'ran out' of fast GPUs for AI boffins, [https://www.theregister.co.uk/2017/05/22/cloud\\_providers\\_ai\\_researchers/](https://www.theregister.co.uk/2017/05/22/cloud_providers_ai_researchers/)
- [Sakura] 高火力コンピューティング, <https://www.sakura.ad.jp/koukaryoku/>
- [Tubame] TUBAME とは, <http://www.gsic.titech.ac.jp/tsubame>
- [Watson] Watson, <https://www.ibm.com/watson/>

2017年11月6日 受理

### 著者紹介



中田 秀基 (正会員)

1995年東京大学大学院工学系研究科情報工学専攻修了。同年、工業技術院電子技術総合研究所入所。2001年独立行政法人産業技術総合研究所に改組、現在に至る。2001～06年まで東京工業大学客員助教授を併任。2011年より筑波大学連携大学院教授を併任。並列、分散計算、機械学習、人工知能に興味をもつ。博士(工学)。ACM、情報処理学会各会員。