

Google が提供する機械学習プロダクトの最新情報と事例

Google Cloud Japan
AI/ML Specialist
Solution & Technology
Toshio Kodama, Dr. Eng.



Introduction



児玉 敏男 (Toshio Kodama)

グーグル・クラウド・ジャパン 合同会社

AI/ML スペシャリスト、博士(工学)

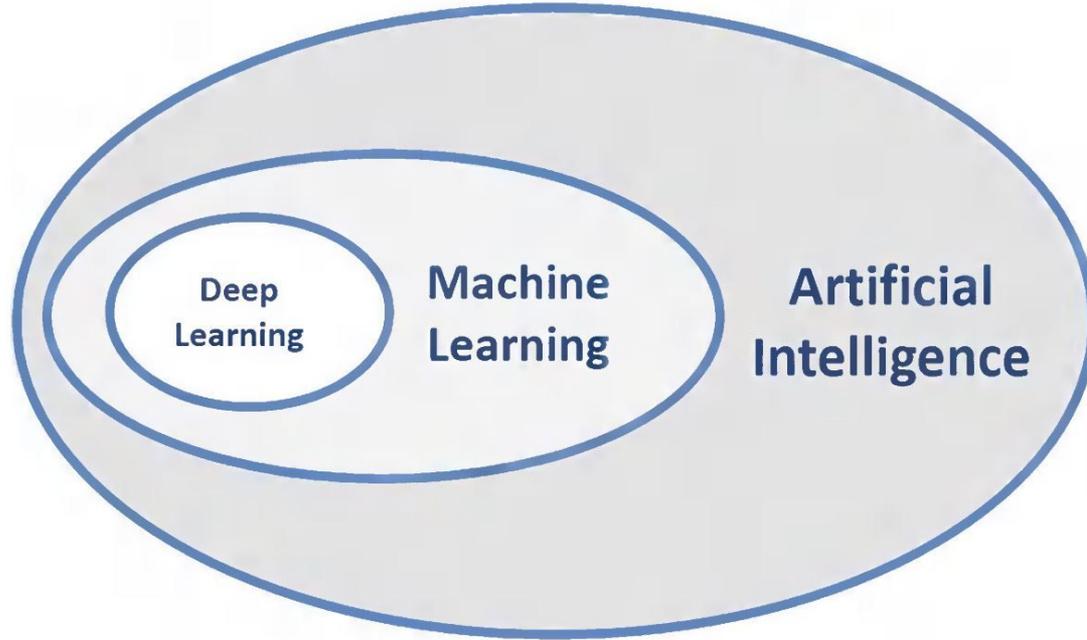
データ工学リサーチャー, 製造・小売の ML コンサルタントを経て、現在は AI/ML スペシャリストとして各インダストリーの Google Cloud AI の導入支援を担当

アジェンダ

- Google と AI
 - 第 3 次 AI ブームから最新研究まで
- Google Cloud AI の最新活用事例

Google と AI

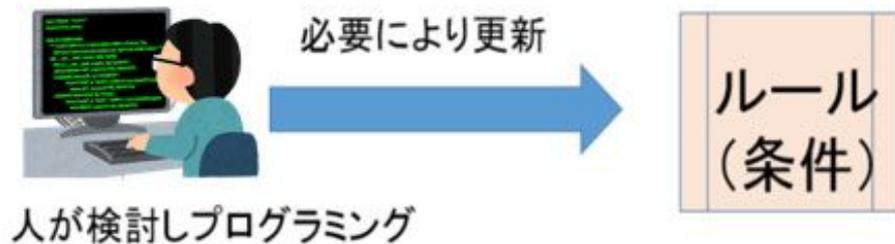
AI と ML と DL



<https://human-centered.ai/2017/11/11/difference-ai-ml/> より

MLは データ x 技術

【従来】

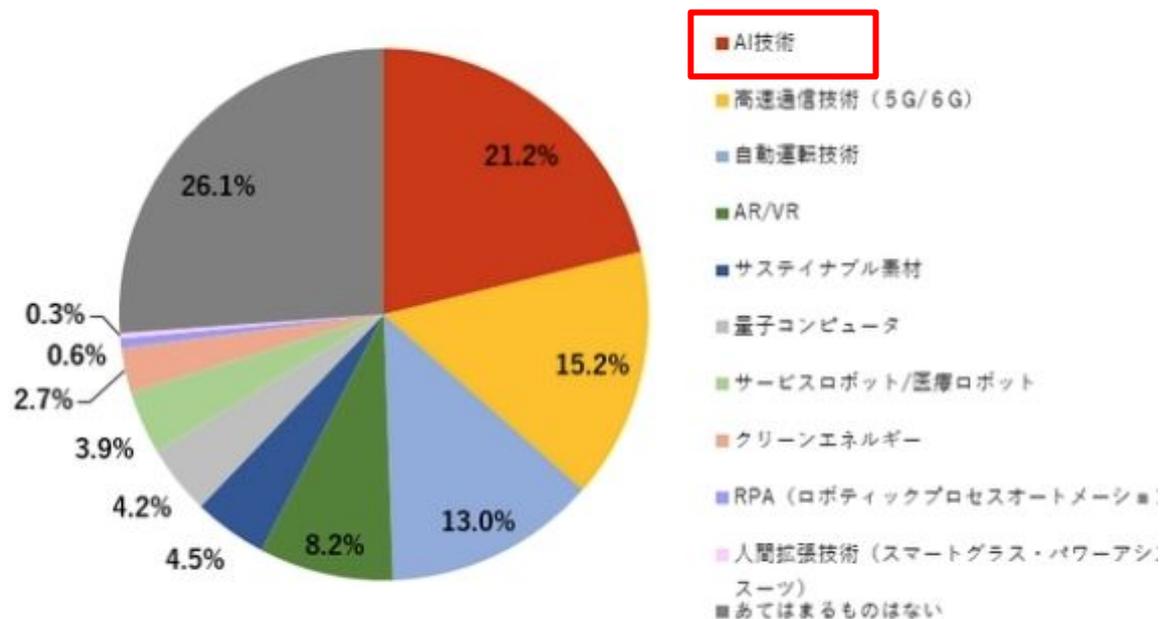


【人工知能】



2020年最も注目されたテクノロジーは1. AI技術、2. 高速通信技術【IEEE調べ】

Q1 2020年1番注目されたと感じるテクノロジーはなんだと思いますか？ (n=330)



出典 : <https://www.atpress.ne.jp/news/246425>

2020年のAI研究ランキング (AINOW)

ICML 2020でAI研究を先導しているのはどこか？

1. **グーグル(アメリカ) - 92.2**
2. スタンフォード大学(アメリカ) - 39.2
3. MIT(アメリカ) - 38.5
4. UCバークレー(アメリカ) - 34.2
5. カーネギーメロン大学(アメリカ) - 24.0
6. マイクロソフト(アメリカ) - 22.6

ICML 2020とICML 2019における世界の上位50機関のアプリケーション・インデックスの変化

1. **グーグル: +19.4**
2. スタンフォード大学(アメリカ) +14.7
3. MIT(アメリカ) +15.4
4. UCバークレー(アメリカ) +10.0
5. カーネギーメロン大学(アメリカ) +4.8
6. マイクロソフト(アメリカ) +5.9

Gleb Chuvpilo 氏 (Thundermark Capital マネージング パートナー、AINOW 編集者)
“グーグルは、機関ランキングおよび企業ランキングにおいて圧倒的な1位を堅守”

20年度のGAFAMの経営戦略キーワード

日経プレジデント(2021.1)より作成



- AI
- スマートホームデバイス

Facebook社

- フェイスブック・グループ
- 仮想通貨リブラ

Apple社

- ブランド価値
- プライバシー重視
- 医療・健康系分野

Amazon社

- オンラインとオフラインの完全統合
- アマゾン・ゴー

Microsoft社

- 政府・自治体 DX
- 製造・物流・小売 DX

「Google の猫」(2012.6)

Google は、2012 年に「人が教えることなく、AI が自発的に猫を認識することに成功した」と発表

- **You Tube** にアップロードされている動画から、ランダムに取り出した200x200ピクセルサイズの画像を1000万枚用意
- **1000台のコンピュータ**でディープラーニングにより3日間かけて学習を行った
- その結果、人間の顔、猫の顔、人間の体の写真に反応するニューロンができた



Using large-scale brain simulations for machine learning and A.I.

Posted: Tuesday, June 26, 2012

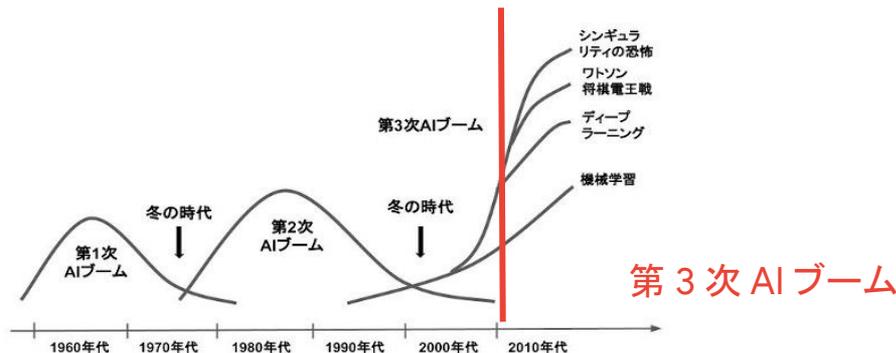
You probably use machine learning technology dozens of times a day without knowing it—it's a way of training computers on real-world data, and it enables high-quality speech recognition, practical computer vision, email spam blocking and even self-driving cars. But it's far from perfect—you've probably chafed at poorly transcribed text, a bad translation or a misgendered image. We believe machine learning could be far more accurate, and that smarter computers could make everyday tasks much easier. So our research team has been working on some new approaches to large-scale machine learning.

Today's machine learning technology takes significant work to adapt to new uses. For example, say we're trying to build a system that can distinguish between pictures of cars and motorcycles. In the standard machine learning approach, we first have to collect tens of thousands of pictures that have already been labeled as "car" or "motorcycle"—that we call labeled data—to train the system. But labeling takes a lot of work, and there's comparatively little labeled data out there.

Fortunately, recent research on self-supervised learning (SSL) and deep learning suggests we might be able to rely instead on unlabeled data—such as random images fetched off the web or out of YouTube videos. These algorithms work by building artificial neural networks, which loosely simulate neuronal (i.e., the brain's) learning processes.

Neural networks are very computationally costly, so to date, most networks used in machine learning have used only 1 to 10 million connections. But we suspected that by training much larger networks, we might achieve significantly better accuracy. So we developed a distributed computing infrastructure for training large-scale neural networks. Then, we took an artificial neural network and spread the computation across 10,000 of our GPU cores (on our data centers), and trained models with more than 1 billion connections.

We then ran experiments that asked, informally, if we think of our neural network as simulating a very small-scale "neuron brain," and show it YouTube videos for a week, what will it learn? Our hypothesis was that it would learn to recognize common objects in those videos. Instead, to our amazement, one of our artificial neurons learned to respond strongly to pictures of... cats. Remember that this network had never been told what a cat was, nor was it given even a single image labeled as a cat. Instead, it "discovered" what a cat looked like by itself from only unlabeled YouTube clips. That's what we mean by self-taught learning.



出典: <https://www.technologystories.org/ai-evolution/>

出典: <https://blog.google/technology/ai/using-large-scale-brain-simulations-for/>

「AlphaGo」(2016.3)

Google の AI 子会社 Deep Mind が囲碁対戦用 AI「AlphaGo」がプロ囲碁棋士の世界王者のイ・セドルに 4 勝 1 敗で勝利

コンピュータが人間に打ち勝つことが最も難しいと考えられてきた分野である囲碁において、人工知能が勝利を収めたことは世界に衝撃をもたらし、世界的 AI ブームを呼び起こすきっかけになった。

- **人がプレイした数百万の囲碁のデータ** と自己対戦
- **深層強化学習** による「盤面評価」と「戦術予測」

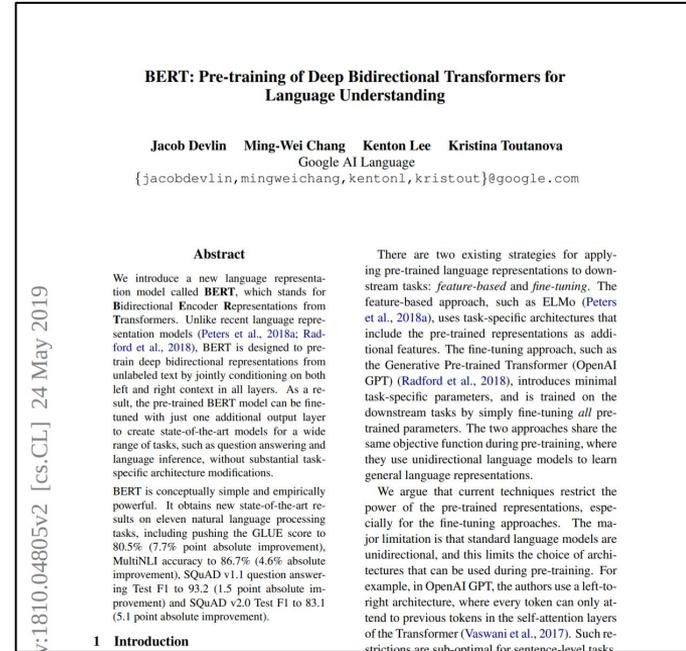


「BERT」(2018.6)

Google Research が発表した自然言語 AI。言語理解を大きく進める画期的な成果
(Bidirectional Encoder Representations from Transformers)

特徴は、文章における
「コンテキスト(文脈)を理解できること」

- Attention 機構をベースにした Transformer
- BooksCorpus の 8 億語と英語版ウィキペディアの 2 億 5000 万語を学習



出典: <https://arxiv.org/pdf/1810.04805.pdf>



parking on a hill with no curb

BEFORE

9:00 google.com



Parking on a Hill. Uphill: When headed uphill at a **curb**, turn the front wheels away from the **curb** and let your vehicle roll backwards slowly until the rear part of the front wheel rests against the **curb** using it as a block. Downhill: When you stop your car headed downhill, turn your front wheels

AFTER

9:00 google.com



For either uphill or downhill **parking**, if there is no **curb**, turn the wheels toward the side of the road so the car will roll away from the center of the road if the brakes fail. When you park on a sloping driveway, turn the wheels so that the car will not roll into the street if the brakes fail.



2019 brazil traveler to usa need a visa

BEFORE



AFTER



出典 : <https://blog.google/products/search/search-language-understanding-bert/>

日本語での事例：検索「シャバシャバじゃないカレー」で検索

The screenshot shows a Google search page with the query 'シャバシャバじゃないカレー'. The search results include a 'Did you mean' suggestion for 'シャバシャバなカレー'. The first result is from macaro-ni.jp, with the title '今日のカレー、なんか水っぽい...原因&とろみをつける...' and a snippet mentioning 'カレーが水っぽくなる原因'. The second result is from masalaryman.com, with the title 'カレーが薄い? とろみがなくて、水っぽい時の対処法と入れる...' and a snippet mentioning '改善方法①カレーの味を...'. The third result is from icedivider.com, with the title 'カレーは絶対シャバシャバ派! 東京のシャバシャバカレー36選...' and a snippet mentioning '水っぽいサラサラしたカレーの...'. The fourth result is from sbfoods.co.jp, with the title 'とろみが弱くなる原因 | スパイス・ハーブ別のQ&A | S&B...' and a snippet mentioning '原因その3ルーを加えた後の煮込みが足りない場合...'. The search interface includes the Google logo, search bar, navigation tabs (All, Images, Videos, News, Maps, More), and settings/tools links.

シャバシャバじゃない
=とろみ
と理解している!

国内研究機関による BERT をベースにした日本語の事前学習済みモデルの構築の動き

産総研 2019.12-

https://www.aist.go.jp/aist_j/press_release/pr2019/pr20191210/pr20191210.html より

情報通信機構 2020.3-

<https://alaginrc.nict.go.jp/nict-bert/index.html> より

発表・掲載日：2019/12/10

ツイート いいね！ 5

AIの動画認識やテキスト理解の基盤となる事前学習済みモデルを構築・公開

—実世界のデータを活用する次世代AI技術の開発と社会実装の促進に期待—

NEDOと産業技術総合研究所は、実世界のデータを活用する次世代人工知能（AI）技術のソフトウェアモジュール構築の一環として、AIによる動画認識とバイオ分野に関する自然言語テキストの理解のための転移学習の基盤となる事前学習済みモデルを構築し、本日公開しました。

今回構築した事前学習済みモデルには、実世界の大量の動画やテキストデータをあらかじめ学習させているため、AI開発に用いることで、少量の学習用データでも次世代AIのソフトウェアモジュールを構築・利用できるようになります。これにより、例えば少量の動画データによる医療動画診断支援向けAIなど、実世界のデータを活用する次世代AI技術の開発と社会実装の促進が期待できます。

NEDOと産総研は、今後も各要素技術の性能を高める研究開発を継続するとともに、新たな要素技術のモジュールを開発し、公開していきます。

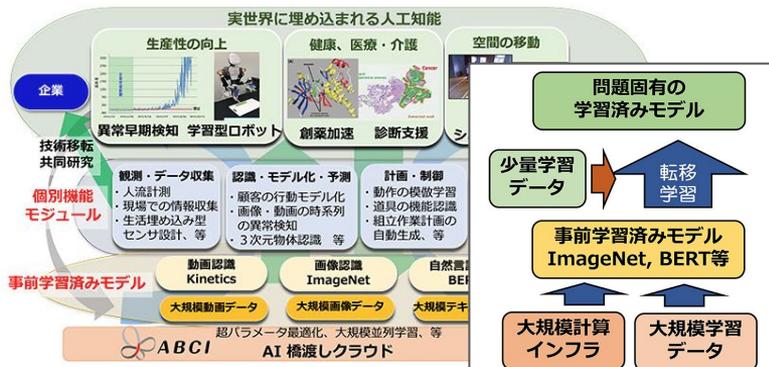


図1 次世代人工知能中核技術の研究開発の概要と、事前学習済みモデル

NICT BERT 日本語 Pre-trained モデル

概要

このページでは、日本語Wikipediaを対象に情報通信機構 データ駆動知能システム研究センターで事前学習を行ったBERTモデルをクリエイティブ・コモンズ 表示 4.0 国際ライセンス (CC BY 4.0)のもとで公開しています。

BERTモデル [1] の学習では、バッチサイズやステップ数を増やすことで性能が向上することが報告されています [2]。そこで本ページのモデルでは、[2] のRoBERTa-500Kと同等の計算量となるよう、ステップ数を [1] と同様に100万に保ちつつ、バッチサイズを [1] の16倍の4,096にして学習を行いました。

作成したBERTモデルの評価として、NICTで作成した (fine-tuning用) 学習データと評価データ (これらのデータは大規模Web情報分析システムWISDOM X、対災害SNS情報分析システムDISAANA、災害状況要約システムD-SUMM、次世代音声対話システムWEKDAの研究開発でも利用)、さらには一般に公開されている学習・評価データを用いて評価を行いました。具体的には、(1) 因果関係認識 (Hashimoto et al., ACL 2014) [3]、(2) ファクトイド質問応答のランキング (開ら, NLP 2020) [6]、(3) バイナリパタン間含意関係認識 (Kloetzer et al., EMNLP 2015) [7]、(4) 東北大 乾・鈴木研究室が公開している解答可能性付き読解 (鈴木ら, NLP 2018) [9] の実験を行いました。いずれのタスクにおいても、すでに公開済みのBERTモデルよりも高い性能が得られており、本ページで公開するBERTモデルを利用することで他の日本語の言語処理のタスクでも性能向上を見込めると考えております (詳しい性能差については評価実験の実験結果を参照)。

ダウンロード

BERTモデルはbyte pair encoding (BPE) [10] を使用したバージョンと使用しないバージョンの2種類を公開します。BPEを使用しないバージョンは、モデルの語彙数を10万語としています。

- BERT_{BASE}モデル (12-layer, 768-hidden, 12-heads)
 - BPEなし、100,000語: [NICT_BERT-base_JapaneseWikipedia_100K.zip](#) (781M)

図2 事前学習済みモデルとそれを用いた転移学習の仕組み

BERTからT5、PEGASUS、Meena(LamDA、2021 I/O)へ

(<https://ai.googleblog.com/>から引用)

T5(2019.10)

文章の生成

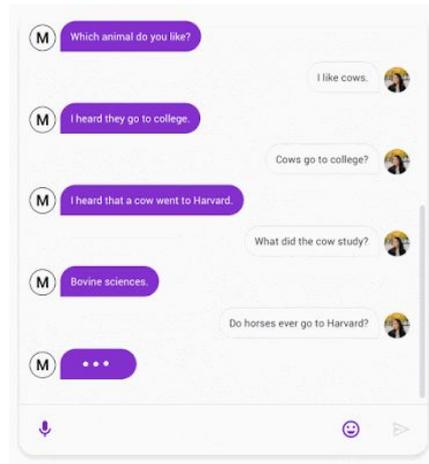


PEGASUS(2019.12) 文章の要約

TRANSFORMER

Meena(2020.1)

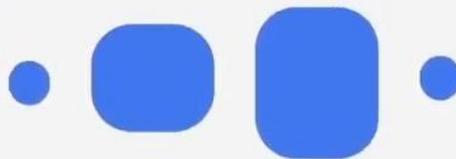
会話モデル



コンタクト センター AI



Prisha Aggarwal
Call ID: CC1-421764



こんにちは、プリシャさん、
お電話ありがとうございます。

例えば、CCAI のコア技術 ”Spec-to-Text” の主要ベンチマークのトップは、Google Research / Brainによる成果

Speech Recognition

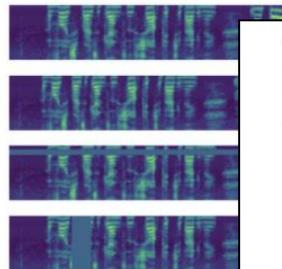
462 papers with code • 78 benchmarks • 51 datasets

Speech recognition is the task of recognising speech within audio and converting it into text.

(Image credit: [SpecAugment](#))

Benchmarks

Trend	Dataset	Best Model	Paper Title	Paper	Code	Comments
	LibriSpeech test-clean	🏆 Conformer + Wav2vec 2.0 + SpecAugment-based Noisy Student Training with Libri-Light	Pushing the Limits of Semi-Supervised Learning for Automatic Speech Recognition	Paper	Code	Comments
	LibriSpeech test-other	🏆 Conformer + Wav2vec 2.0 + SpecAugment-based Noisy Student Training with Libri-Light	Pushing the Limits of Semi-Supervised Learning for Automatic Speech Recognition	Paper	Code	Comments



Edit

Pushing the Limits of Semi-Supervised Learning for Automatic Speech Recognition

Yu Zhang* James Qin* Daniel S. Park* Wei Han Chung-Cheng Chiu
Ruoming Pang Quoc V. Le Yonghui Wu

Google Research, Brain Team
{ngyuzh, jamesqin, danielspark, weihan, chungcheng, rpang, qvl, yonghui}@google.com

Abstract

We employ a combination of recent developments in semi-supervised learning for automatic speech recognition to obtain state-of-the-art results on LibriSpeech utilizing the unlabeled audio of the Libri-Light dataset. More precisely, we carry out noisy student training with SpecAugment using giant Conformer models pre-trained using wav2vec 2.0 pre-training. By doing so, we are able to achieve word-error-rates (WERs) 1.4%/2.6% on the LibriSpeech test/test-other sets against the current state-of-the-art WERs 1.7%/3.3%.

1 Introduction

Recently, semi-supervised learning (SSL) methods have been used to drastically improve the performance of automatic speech recognition (ASR) networks. The goal of semi-supervised learning is to use a large unlabeled dataset to help with improving the performance of a supervised task defined by a labeled dataset. In this work, we combine recently proposed pre-training and self-training methods to obtain state-of-the-art (SOTA) performance on LibriSpeech [1]. We use the audio from the Libri-Light dataset [2], derived from the LibriVox database of free public domain audio books, as the unlabeled data for semi-supervised learning.

Recall that in iterative self-training, a series of models are trained where a given model in the series serves as a teacher to the succeeding model by generating labels on the unlabeled dataset. The student to this teacher model is trained on the dataset obtained by combining the supervised set with the teacher-labeled dataset. Meanwhile, the pre-training model is trained on the unlabeled dataset.

Transformer が消える！？ gMLP の提案 (2021.5)

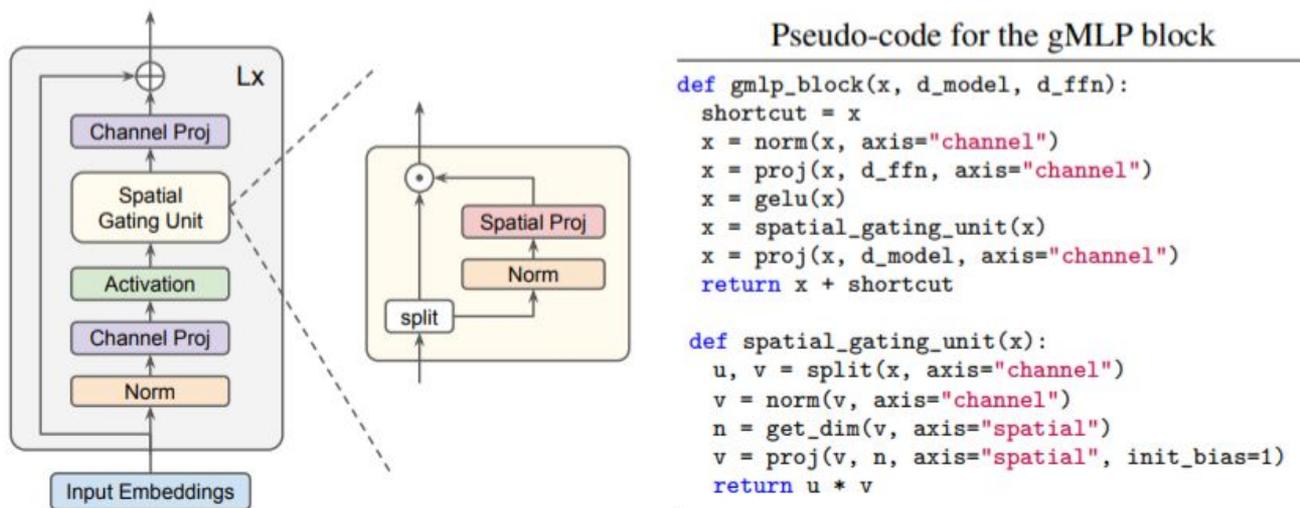
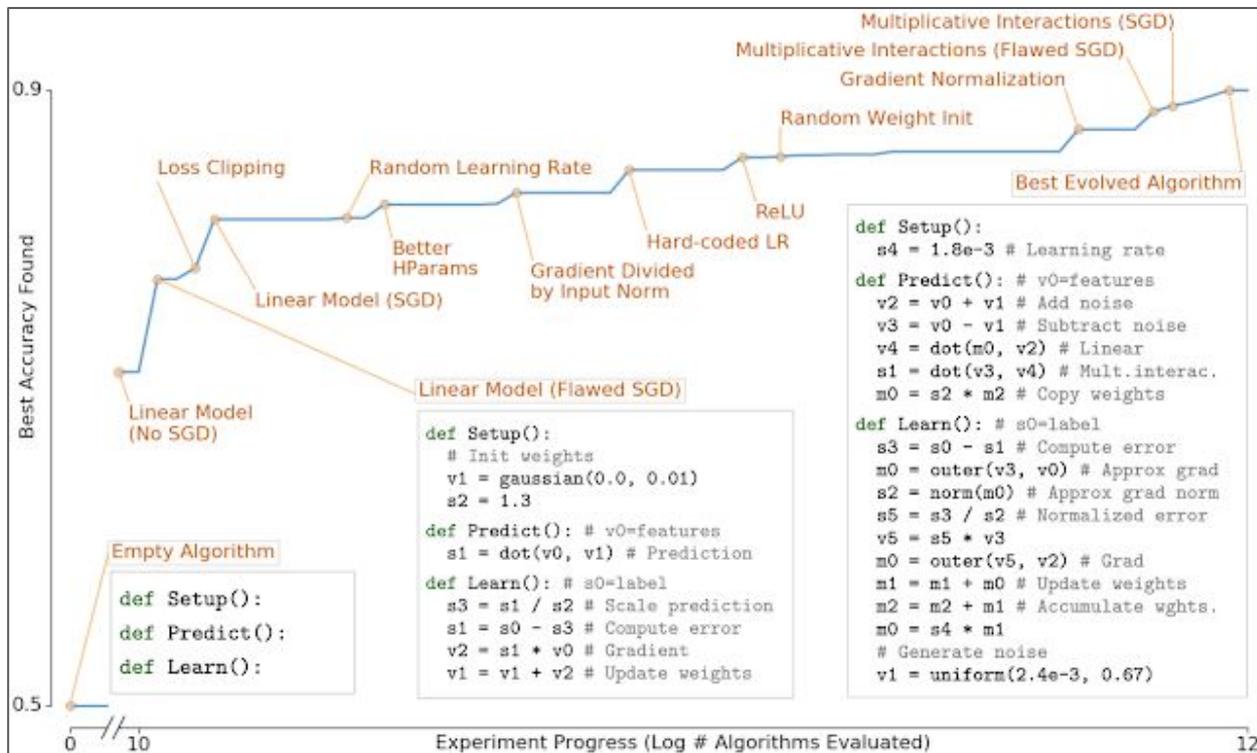


Figure 1: Overview of the gMLP architecture with Spatial Gating Unit (SGU). The model consists of a stack of L blocks with identical structure and size. All projection operations are linear and “ \odot ” refers to element-wise multiplication (linear gating). The input and output protocols follow BERT for NLP and ViT for vision. Unlike Transformers, gMLPs do not require positional encodings, nor is it necessary to mask out the paddings during NLP finetuning.

(<https://arxiv.org/abs/2105.08050>)

AutoML Zero (2020.7) の衝撃

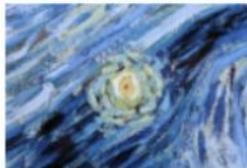
基本的な数学の組み合わせだけで ML を自動生成



マルチモデル！視覚と言語の表現学習 ALIGN(2021.6)

“Van Gogh Starry Night ...”

“details”



“in black and white”



“on a canvas”



“in dark wood frame”

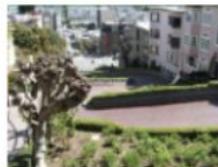


“Lombard street ...”

“view from bottom”



“view from top”



“bird's eye view”



“in heavy rain”



“seagull in front of ...”

“Golden Gate Bridge”



“London Tower Bridge”



“Sydney Harbour Bridge”



“Rialto Bridge”



Google Cloud AI の最新活用事例

Google Cloud の AI プロダクト

Services & Solutions
容易に実装

APIs
事前学習済みモデル

Vertex AI

AutoML
データから自動で
モデル開発

Custom Code
独自コードで
モデル開発

Infrastructure
AI 開発の基礎

Solutions Talent Solution Contact Center AI Document AI			Services ASL Professional Services Cloud AI Partners						
Sight Vision Video Intelligence		Language Natural Language Translation		Conversation Speech-to-Text Text-to-Speech Dialogflow Enterprise		Structured Data Recommendations AI			
Sight Vision Video		Language Natural Language Translation			Structured Data Tables				
Built-in Tools Datasets Data Labeling Pre-built Algorithms Notebook VM Images Training Predictions				On-prem Kubeflow		Integrated with Dataflow Dataproc BigQuery Dataprep データポータル			
Accelerators TPU GPU CPU				Frameworks TF K PyTorch TensorFlow Spark					

※ 2021 年 6 月現在、AutoML の Translation は Vertex AI の一部ではなく、従来の単独プロダクトのみの提供となっています

コールセンター業務のチャットボット (MUFG Bank)

- コールセンター業務でのオペレータ接続の待ち時間が課題になっていた
- Dialogflow を利用し、シナリオに基づいたチャットボットを構築

選択

テキスト入力

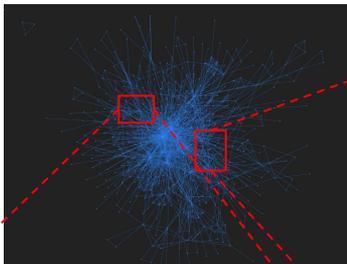
「プロジェクトの正式スタートに先駆け、2018 年春ごろから、Dialogflow も含めて、およそ 10 社 くらいのチャットボット製品を比較しています。その際重視したのはやはり精度。特にチューニング後の精度と、それにかかった時間を重要な評価対象としています。Dialogflow はその点で最も優秀でしたね。最終的には、そこに GUI のみで開発・メンテナンスができる運用のしやすさ、個人情報扱わない上でのセキュリティも加味して、Dialogflow を選びました。」
(システム企画部・多川氏)

約 27,000 件/月 (2020.1 現在、全体の 50-60%) の問い合わせをチャットボットで対応可能に。オペレーターに繋ぐ件数が減少、オペレーターの配置人数を削減予定。接続の待ち時間が減少、満足度が向上した。

災害記録文書データの可視化(化学メーカー)

経産省が公開する災害情報の記録文書データから、Natural Language API を利用してロケーションやイベントに関するワードを自動抽出し、共起ネットワークで可視化

災害情報の共起ネットワーク



「階段」と「坑内」「詰め所」「足場」が関連 = 階段に共通した事故がある

共起ネットワーク作成

分析ワードの自動抽出

罹災者は昼の勤務で構内の内部掃除作業を行っていたが、構内は50度を超え非常に暑い状況にあった。その時は夜間作業者との2交代制であった。午後の作業で手足がしびれる症状が出たので、休憩室で冷水を飲み休憩した。

RESET

See supported languages

Entities	Sentiment	Syntax	Categories
(罹災者) ₃ は(昼) ₄ の(勤務) ₅ で(構内) ₁ の(内部掃除作業) ₂ を行っていたが、(構内) ₁ は(50) ₁₃ 度を超え非常に暑い(状況) ₆ にあった。その時は(夜間作業者) ₈ との(2) ₁₄ 交代制であった。午後の(作業) ₉ で(手足) ₁₂ がしびれる(症状) ₇ が出たので、(休憩室) ₁₀ で(冷水) ₁₁ を飲み休憩した。			
1. 構内 Salience: 0.25	LOCATION		2. 内部掃除作業 Salience: 0.14 EVENT
3. 罹災者 Salience: 0.13	PERSON		4. 昼 Salience: 0.10 OTHER
5. 勤務 Salience: 0.08	EVENT		6. 状況 Salience: 0.07 OTHER
7. 症状 Salience: 0.05	OTHER		8. 夜間作業者 Salience: 0.05 PERSON

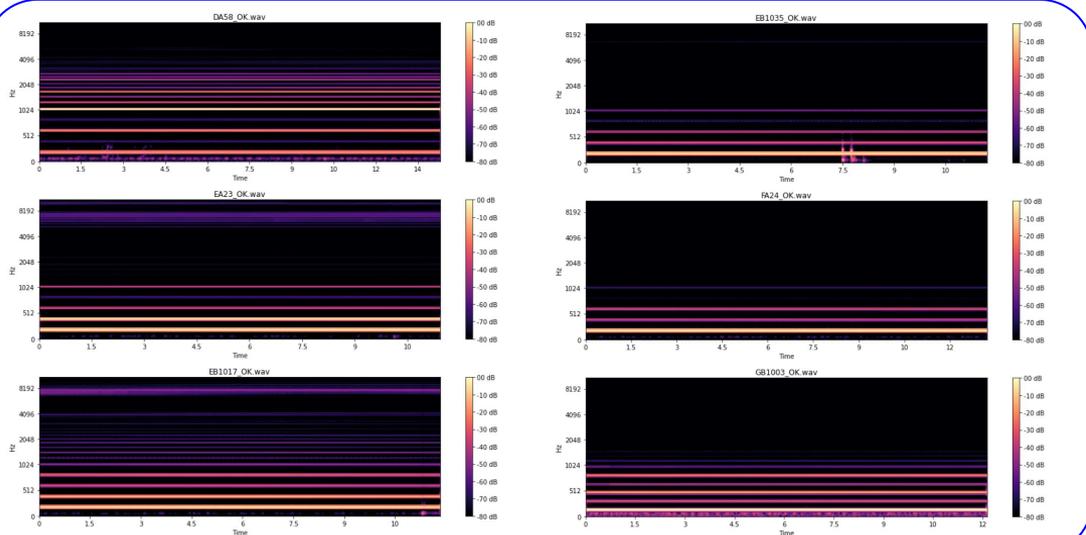
「終了」でネットワークが連結
=作業終了後の災害が多い

災害情報の可視化によりインサイトが得られ、災害事故対策に活用

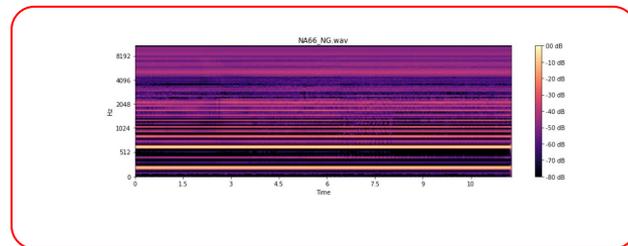
モーターの異音検知（電子部品メーカー）

- 工作機械のモーター保全業務として異音検知を検査者が行っているが、検査者の暗黙知に基づいて人によって検知の基準が異なるという課題があった
- Vertex AI Notebooks を用いて、モーターの音声データをスペクトログラムとして可視化、正常音/異音の ML による 2 値分類を行い精度を検証予定

OKデータ



NGデータ



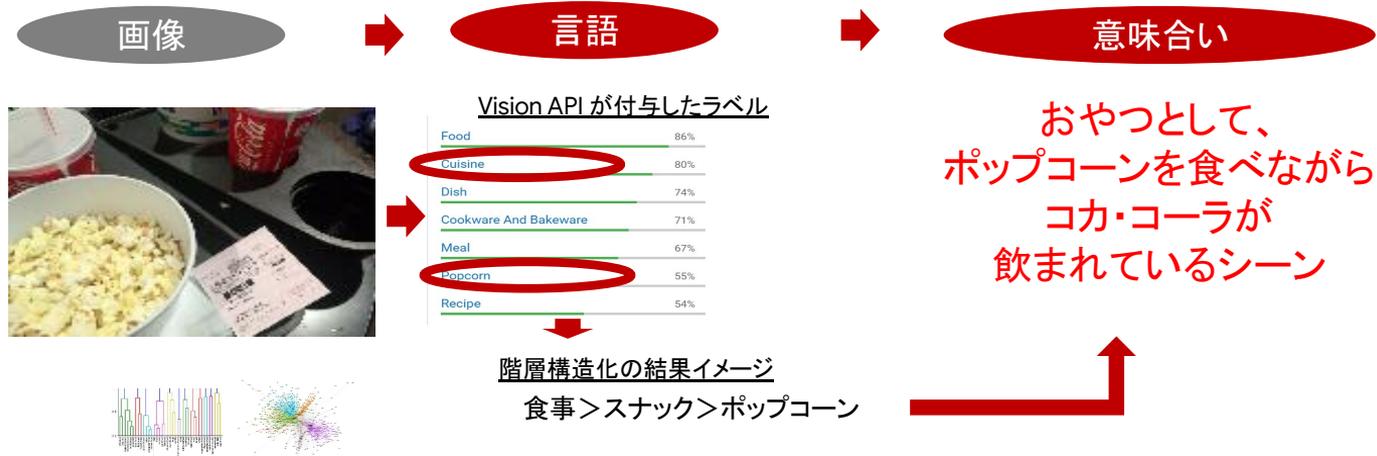
モーター異音検知基準の標準化により、保全業務の精度向上とコスト削減を見込む

SNS 投稿動画解析によるマーケティング(コカ・コーラ)

New

- SNS の投稿画像からコカ・コーラが映っているものを抽出・同時に映っているものに対して AutoML Vision、Vision API で自動ラベル付けを実施
- 「山登りの時にコカ・コーラを飲む」など、これまで把握できていなかったシーンを発見

Vision API のラベルによる考察



従来のアンケートによる消費行動調査と比較し、よりリアルな消費者分析が可能に。今後は消費者の表情や投稿ツイートの文章から感情分析まで消費者行動分析を拡大予定

Thank you!