

# 2018年度 JUAS研究成果発表 AI研究会

2019年4月18日、13:45-14:15

AI研究会

# 2018年度AI研究会

## AI研究会の発足背景と目的

計算コストの指数的な低減を背景として、AI(機械学習及び統計処理)技術がビジネスモデルを大きく変えていく傾向が顕著になっています。

これまではAIの役割が補助的であったビジネスにおいても、その活用の成否が事業に大きな影響を与えるようになりつつあります。

本研究会では自社事業にAIを一層効果的に適用できるよう、次の三つに取り組みます。

**A**

AI技術の適用によって、価値創出や業務の効率化などが期待できるビジネスモデルやプロセスを、事例研究などを通じて精査する

**B**

AIの技術的な理解を深めつつ、実践のための手法と知見を明示する

**C**

AI適用のための組織的課題や、AI運用、適用効果の評価方法などを整理する

# 運営方針

## 目的： 自社事業に資する成果物を持ち帰って頂くこと

研究会に参加してAIに関する知見を強化して頂きたいのですが、それにとどまらず、皆様の業務に実際に役立つものを持ち帰って頂きたいです

## 成果物： 領域別の研究結果と事業適用シナリオ

得意領域を深めて頂くとともに、事業全体の中での各領域の位置づけをそれぞれの立場から考察し、事業適用シナリオを作成頂く方向です  
ただし、成果物の範囲は上記に限りません

## 企業横断： AIに関与する方の企業の境界を越えた連携

JUASという組織を活用して、通常の業務では得難い、他企業でAIに関連する業務に携わる方との関係性を構築頂き、より多面的な貢献につなげて頂くことを期待しております  
GAFA等が促す事業の変化への対応にもつながれば何よりです

# 組織的な知見を、自社事業に還元

「組織的に＝分業して」 大きな問題を分割して、チームで取り組む

8月～11月： 前期分科会で領域別の研究を実施

ビジネス

実装

アルゴリズム

組織・運用



分科会のメンバーを組み替えて、新たに分科会を形成

11月～3月： 後期分科会でデータ種別のAI適用について研究

テキスト

画像・映像

音声

数値

基礎となる知識を整理の上、AI適用を試みた

# 前期分科会成果

---

## 領域別の研究

# A. ビジネス

---

## 問題意識：

各企業でAI技術の適用・導入を進めようと考えられているものの、「どのような手順で」「何をすればよいのか」が分からないケースが多い

## アプローチ：

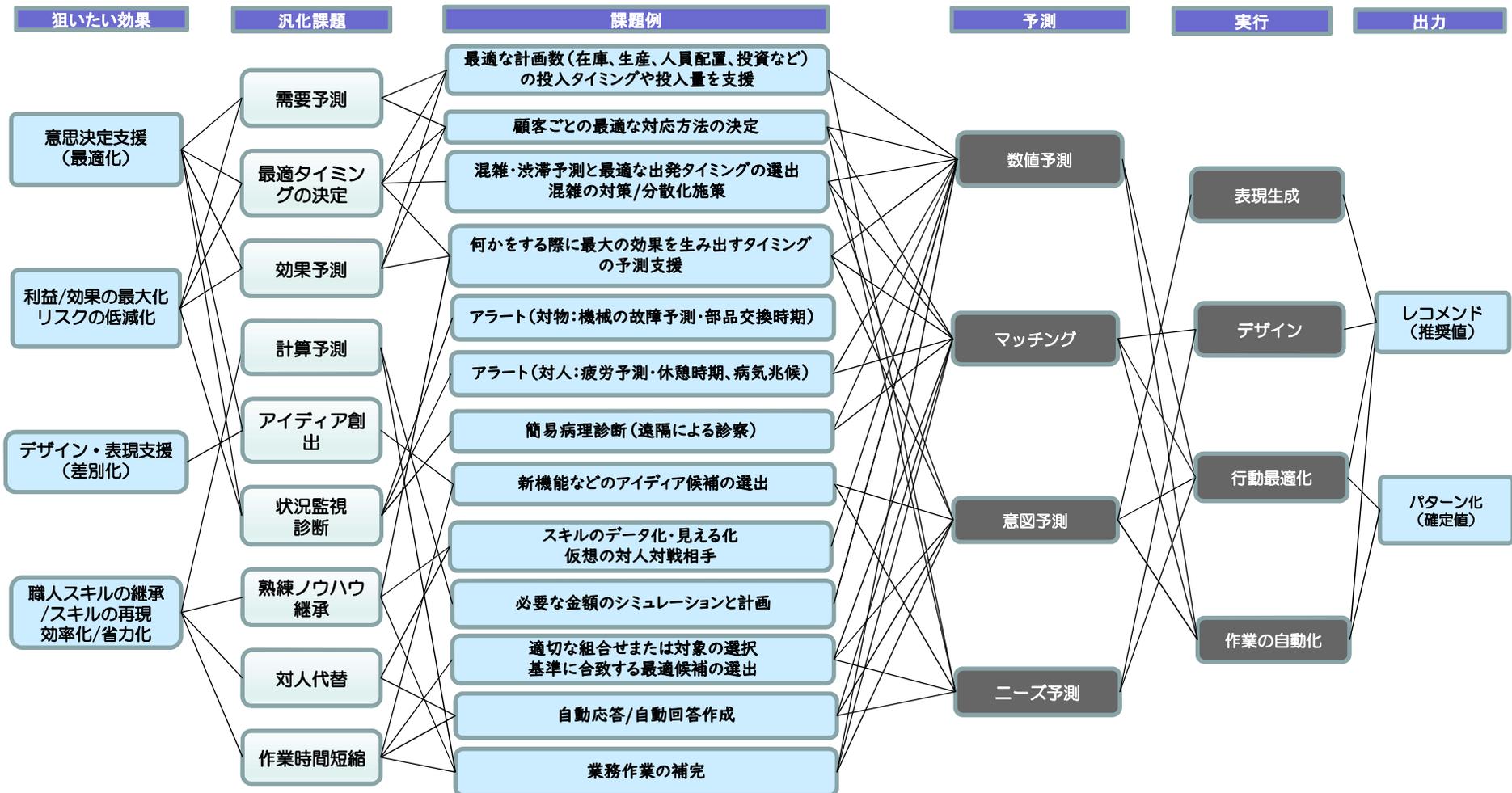
AIに何を求めるか、に関して、支援（意思決定・判断・予測）、作業（効率アップ・技術継承）、確認（状況把握）、検索（マッチング）、AI活用で何が得られるか、に関して、最適化（意思決定支援）、リスク軽減（利益/効果の最大化）、差別化（デザイン・表現支援）効率化/省力化（職人スキルの継承/スキル再現）、の各4つを整理する

## 結論：

課題及び求める結果（出力）をキーとして、必要となるリソース（収集データ、利用アルゴリズム、継続運用に必要な内容など）を明確化し、スムーズなAI導入を検討する仕組みを構築した

# AI導入フレームワークを作成する

## AI技術を用いた課題解決において期待する効果(ユースケース)





## B. 実装

---

### 問題意識：

企業でAIの適用が思うように進まないのは何故か？  
なぜ多くの人がAIへの第一歩を踏み出せないのか？

### アプローチ：

まずは手を動かしてAIでできること・プロセスを体感してみる。  
AIのビジネス適用のためにやらなくてはならないことを、  
一連のプロセスを体験することで実感してみる。  
取り組みやすいテーマで楽しみながら手を動かしてみる。

⇒ テキスト分析と画像認識を題材として、機械学習を適用した

### 結論：

ツールやソースが充実しており、AIの実装自体は比較的容易である一方で、  
良い学習データの収集・整理には目視に頼る必要がまだまだ多く、  
究極的な効率化（完全自動化）は困難であると感じた。

# 使用したツールと学習データ（テキスト）

## ・ツール



<https://textmining.userlocal.jp>

Web上で利用できる無料のテキストマイニングツール  
文書の特徴を把握することが容易に可能になる

## ・データ

①

**SIGNATE**

<https://signate.jp/>

AI開発のコンペティションから「【練習問題】 スпамメール分類」  
⇒スパムとノンスпамに分けたテキストデータを使用

②



<https://www.open-governmentdata.org/>  
自治体の公共データを随時公開するオープンデータサイト

「ふっけい安心メール」の配信データ  
⇒地域防犯のために防犯情報などを記載したメール

# 使用したツール（画像）

サンプルプログラムの作成は、全て無料のものを利用し、以下のシステム構成で取り組んだ

アプリケーション		
Keras(2.2.0)	opencv-python (3.4.2.17)	他、必要ツール
tensorflow(1.9.0)		
Python 3.5		Jupyter notebook (4.4.0)
Anaconda 5.1 For Linux Python 3.6 version		
Ubuntu 16.04 LTS		
VMware Workstation 14 Player		
Windows 10 64bit		

【Wikipediaによる説明】

## ■ Anaconda

Anaconda（アナコンダ）は、データサイエンスおよび機械学習関連アプリケーション（大規模データ処理、予測分析（英語版）、科学計算）のためのPythonおよびRプログラミング言語のフリーで**オープンソースのディストリビューション**であり、パッケージ管理および展開の単純化を目指したものである。パッケージバージョンはパッケージ管理システム「Conda（英語版）」によって管理される。

## ■ Python

Python（パイソン）は、**汎用のプログラミング言語**である。コードがシンプルで扱いやすく設計されており、C言語などと比べて、さまざまなプログラムをわかりやすく、少ないコード行数で書けるといった特徴がある。

## ■ Tensorflow

TensorFlow（テンソルフロー）とは、**Googleが開発しオープンソースで公開している、機械学習に用いるためのソフトウェアライブラリ**である。

## ■ Keras

DeepLearning4j、TensorFlow、CNTK、Theano（英語版）の**上部で動作**することができる。ディープニューラルネットワークを用いた迅速な実験を可能にするよう設計され、**最小限、モジュール式、拡張可能であることに重点**が置かれている。プロジェクトONEIROS（Open-ended Neuro-Electronic Intelligent Robot Operating System）の研究の一部として開発された。中心的な開発者、メンテナはGoogleのエンジニアのFrancois Cholletである。

## C. アルゴリズム

---

### 問題意識：

何故、ユーザー企業側（発注者側）は、  
実現出来ないA I 開発案件をベンダー企業に依頼してくるのか？

### アプローチ：

「ゼロから作るDeepLearning」を輪読し、まとめ、  
ディープラーニングの基礎となるアルゴリズムを説明する。具体的には、  
§2 パーセプトロン、§3 ニューラルネットワーク、§4 ニューラルネットワークの学習、  
§5 誤差逆伝播法、§6 学習に関するテクニック、の5セクションを中心に、  
アルゴリズムを説明した

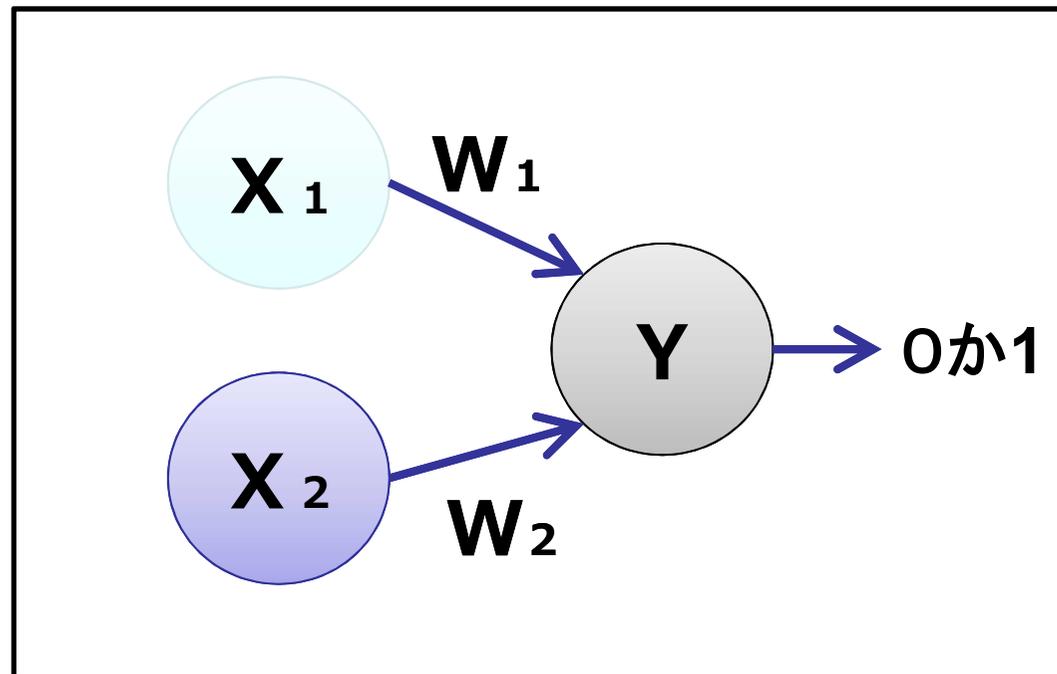
### 結論：

ユーザー企業はA I の学習アルゴリズムを理解していないことが多い。  
「教師データの無い画像解析を依頼してくる」などの事例が散見され、  
これはアルゴリズムの知見があれば生じないためである

# パーセプトロンとは・・・ (NNの原始)

「複数の信号」を入力として受け取り、  
「1つの信号」を出力する

信号は、「流す(0)」 or 「流さない(1)」の二値の値。



# パーセプトロンの数式

閾値を「 $\Theta$ 」とすると、パーセプトロンの数式は下記の通り。

$$Y = \begin{cases} 0 & (W_1 X_1 + W_2 X_2 \leq \Theta) \\ 1 & (W_1 X_1 + W_2 X_2 > \Theta) \end{cases}$$

「 $\Theta$ 」を「 $-b(=バイアス)$ 」に置き換えると、下記のように記載出来る。

$$Y = \begin{cases} 0 & (b + W_1 X_1 + W_2 X_2 \leq 0) \\ 1 & (b + W_1 X_1 + W_2 X_2 > 0) \end{cases}$$

重み( $W$ )  $\Rightarrow$  入力信号への重要度をコントロール  
バイアス( $b$ )  $\Rightarrow$  発火のしやすさを調整する

## D. 組織・運用

---

### 問題意識：

事業会社によるAI導入が、想定通りに進まない原因を調査し、PoCをシステム化へつなげるための成功要因を考察したい

### アプローチ：

成功/失敗事例を、各種文献、Webサイト、各社情報を用い調査し、調査結果を分科会内でディスカッションし、成功/失敗理由を抜き出す。そして、成功に必要な要因を、AI導入までの段階（PoCの前・中・後）と、要素（進め方・データ・組織）に分類し、必要な要因を満たすための準備、タスク例を列挙する

### 結論：

PoCを乗り越える目的に絞り、調査・ディスカッションを実施したが、結果的にPoC以降まで含めた網羅的な調査となった。また、複数の事例や参加メンバーの固有事例からアプローチしたため、実業務に沿う内容となった。

# 組織 - PoC前 -

## 成功/失敗事例

### [ 失敗例 ]

- ✓ 社内のAI推進部門が、**現場の声を聴かずに**AIシステムを開発・社内リリースしたが、現場が求めるものではなく、3か月後には利用者がいない状況に（A社）
- ✓ PoC開始後のヒアリングで、**PoCのテーマと現場ニーズが違う**ことが発覚。PoCが中止に（B社）

## AI導入に必要な要因

- ✓ 経営層および現場のコミットを獲得できる人材がいる
  - 期間、金額を設定し、経営層のコミットを獲得する
  - 経営層への定期的な報告体制と援護体制を確立する
- ✓ AI技術動向を把握し、提案する人材がいる
  - AIに関する、導入事例等の知識を普及する
  - PoCで証明する解決策の策定ができる
- ✓ 既存業務を把握し、将来像をイメージする人材がいる
  - 現場とのブリッジとなる
  - 現場の課題・要望を拾える



# 組織 - PoC中 -

## 成功/失敗事例

### [ 成功例 ]

- ✓ PoC中、**現場と試行錯誤**する体制を構築し、ノウハウを引き出すことで実用に耐えうるAIの検証ができた。また、現場の働き方からAIを検討できる技術者も誕生（C社）
- ✓ 「このように業務を変えれば100%の精度がでなくてもOK」という**提案を現場からもらえ**、無事成功した（某企業）

## AI導入に必要な要因

- ✓ 現場を巻き込む体制の構築をしている
  - AI導入後の現場の業務を具体的に検討できる
  - PoC中の作業中、適宜現場の意見を拾ったり、現場から逆に提案を受けたりできる
- ✓ PoC推進のマネージャーを擁立している
  - PoCをリードできる
  - 社内のAI人材育成も含め、アサインを検討できる
- ✓ プロトタイプ開発力がある人材がいる
  - RPAやchatbotなどのAIツールを使いこなせる
  - GCPなどを組み合わせたプロトタイプを構築できる



# 組織 - PoC後 -

## 成功/失敗事例

### [ 成功例 ]

- ✓ 採用業務の効率化に対し、業務担当者が過去データを使い**試行錯誤**できる環境を整備した結果、75%削減ができた (A社)
- ✓ 空調設備の熱負荷予測のシステム化を、**Azure**を用いて実現。一からの開発に比べ、省コストで実施できた (D社)

## AI導入に必要な要因

- ✓ PoCからシステム化を判断する体制を構築している
  - 経営層、現場、IT部門を巻き込み判断できる
  - 費用対効果、KPI、業務適用可能性の観点で判断できる
- ✓ AIを実システムに落とせる人材がいる
  - AI特有の誤答時のフォローも含めシステム化できる
  - GCPやAzureなど、既にある仕組みも利用できる
- ✓ AIを育て水平展開する組織・スキームを構築している
  - 導入後の業務部門からフィードバックを受けられる
  - 部署・系統横断的な社内AI成功事例を共有できる
  - 業務部門でもAIをチューニングできる仕組みがある
  - 社内のAI普及の啓蒙活動を実施できる



## データ種別のAI適用の研究

## a. テキスト

---

### 問題意識：

業務改善に自然言語処理がどのように役に立つのか

### アプローチ：

自然言語処理に関して興味を持った内容をメンバーで分類する。  
その分類結果に従って、

- ・世の中で業務の役に立っている自然言語処理の事例調査を実施し、
- ・特に興味深かった事例について、実際に手を動かし、試す。

### 結論：

分析： 製品サービス改善のための顧客の要望分析  
検索： 受付業務のレスポンス改善（ヘルプデスクなど）  
翻訳： 外国人向けコミュニケーションサービス（通訳デバイスなど）  
は実用レベルで業務改善に結びつけられる

# 「分析」の事例調査

## 市販ツールが豊富でリアルタイムで**大量のデータ**を活用・分析している事例が多い

No.	【社名・組織】 背景・目標	データソース	ツール・ソリューション	概要	検証内容	業務改善効果
1	【CTC】 サービス利用者に必要な情報を迅速に目撃適切に利用者に提示することが課題	ニュース・広告・投稿・電子書籍など	クラウド型の言語解析サービス「コトロジ」 Jetrunkテクノロジー社 言語解析エンジン「TrueText（ツール・テキスト）」		パッケージソリューションの無し	ネットサービス事業者のテキスト分析を効率化
2	【パナソニック システムネットワークス】 お客様の声（Voice Of Customer: VOC）を製品・サービスの改善に活かすことが課題	マーケティングや営業、サポートの日々の活動で集められる膨大なVOC	野村総合研究所 テキストマイニング・システム「TRAINA」	『顧客の声ポータル』を活用して全社向けに「VOCポータルサイト」を立ち上げ、自動的に各種データソースから収集したVOCを分析して結果を配信	パッケージソリューションの無し	お客様の声に含まれるリスクキーワードの見落としや主観による判断ミス、社内連絡の遅れがなくなり、リスクに対して迅速、確実に対処
3	【国立研究開発法人情報通信研究機構（NICT）】 大規模災害の発生時に膨大な被災報告が出力され、被災状況の概要を一目で把握することが困難	Twitter上の災害関連報告	災害状況要約システム「D-SUMM（ディーサム）」（Disaster-information SUMMarizer）	指定エリア内の被災報告を瞬時に要約し、そのエリアの被災状況の概要が一目でわかるように、コンパクトかつ、わかりやすく提示し、各種救援、避難等を支援	不明	地図での表示も含めて、より直感的で分かりやすい被災状況の提示が可能。効率的な救援、避難の支援に貢献。
4	【テクノデータサイエンス・エンジニアリング(TDSE)】 参院選で選挙権年齢が「18歳以上」に引下げられ、若者有権者の動向を注目	Twitter等のソーシャルメディア	テクノデータサイエンス・エンジニアリング社 SNS分析人工知能サービス「NetBase」	ソーシャルメディアでの情報が、政党・政策の支持、投票、得票数にどのように影響するか検証	不明	ネットユーザーの意識や世代間ギャップもリアルタイムで解析し、傾向分析を直感的に見せる

# テキスト分析の実施

---

調査事例より以下の業務に活用できるのではないかと考えた。

- ・ソーシャルメディアの活用
- ・顧客の声の分析・分類

よって、インターネット上のデータを収集（スクレイピング）し、それを分析する取組を以下のようないくつかの方法を用いて実施した。



**実装 1** : TwitterAPIによるTweetデータ収集  
地図上への可視化

**実装 2** : kakaku.comからwebスクレイピングでデータ収集  
口コミ情報のポジネガ分析

**実装 3** : RPAによるAmazonレビューのデータ収集  
レビュー情報の係り受け解析

# Pythonを用いたポジネガ分析の実装

ポジネガ分析：MeCabにより形態素解析を行い、分解した単語を単語感情極性対応表※に当てはめてその文章の感情を数値化[-1(ネガティブ)~1(ポジティブ)]する。

例文：

「友人に、すすめられて、購入しました。購入はAmazonアマゾンにて、定価よりも安く買えました。耐久性もあり、丈夫で安心して使えそうです。」

単語:友人	数値:-0.119046
単語:購入	数値:-0.24144899999999997
単語:購入	数値:-0.24144899999999997
単語:定価	数値:-0.391315
単語:耐久	数値:-0.227229
単語:性	数値:-0.388637
単語:安心	数値:0.981397



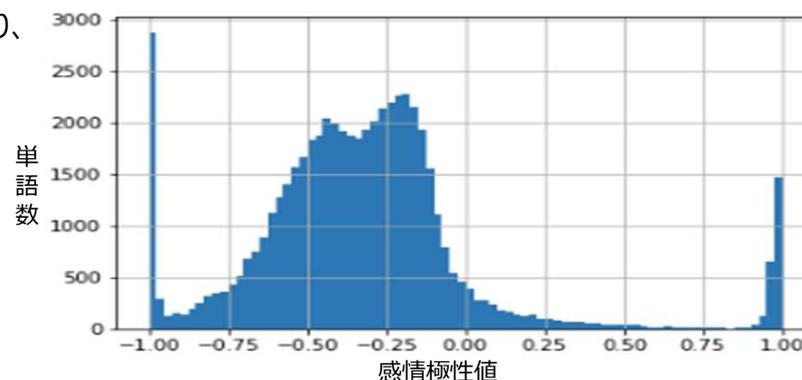
文章としての感情値は単純に平均で算出  
SUM(各数値) / 単語数  
**-0.089 ←微妙にネガティブ**

結果考察：利用した単語感情極性対応表は以下の通り全体的にマイナス（ネガティブ）の数値に振れていて、上記のポジティブな文章でもマイナスの数値（ネガティブ）となったと思われる。

※単語感情極性対応表：東工大の高村教授が公開しており、単語単位でイメージ（極性）を数値化（-1~1）している

## 【単語感情極性対応表の基本統計量】

count	55125	min	-1.000000
mean	-0.319764	25%	-0.522353
std	0.382738	50%	-0.339964
		75%	-0.176277
		max	1.000000



所感：実装は至って容易に可能であるが、実用化に向けては単語単位の感情数値だけを利用するのではなく、構文解析・意図解析なども行う必要があると思われる。

## b. 画像

---

### 問題意識：

AIに関するどのような知見と技術を得る事で画像認識への理解が深まるか？

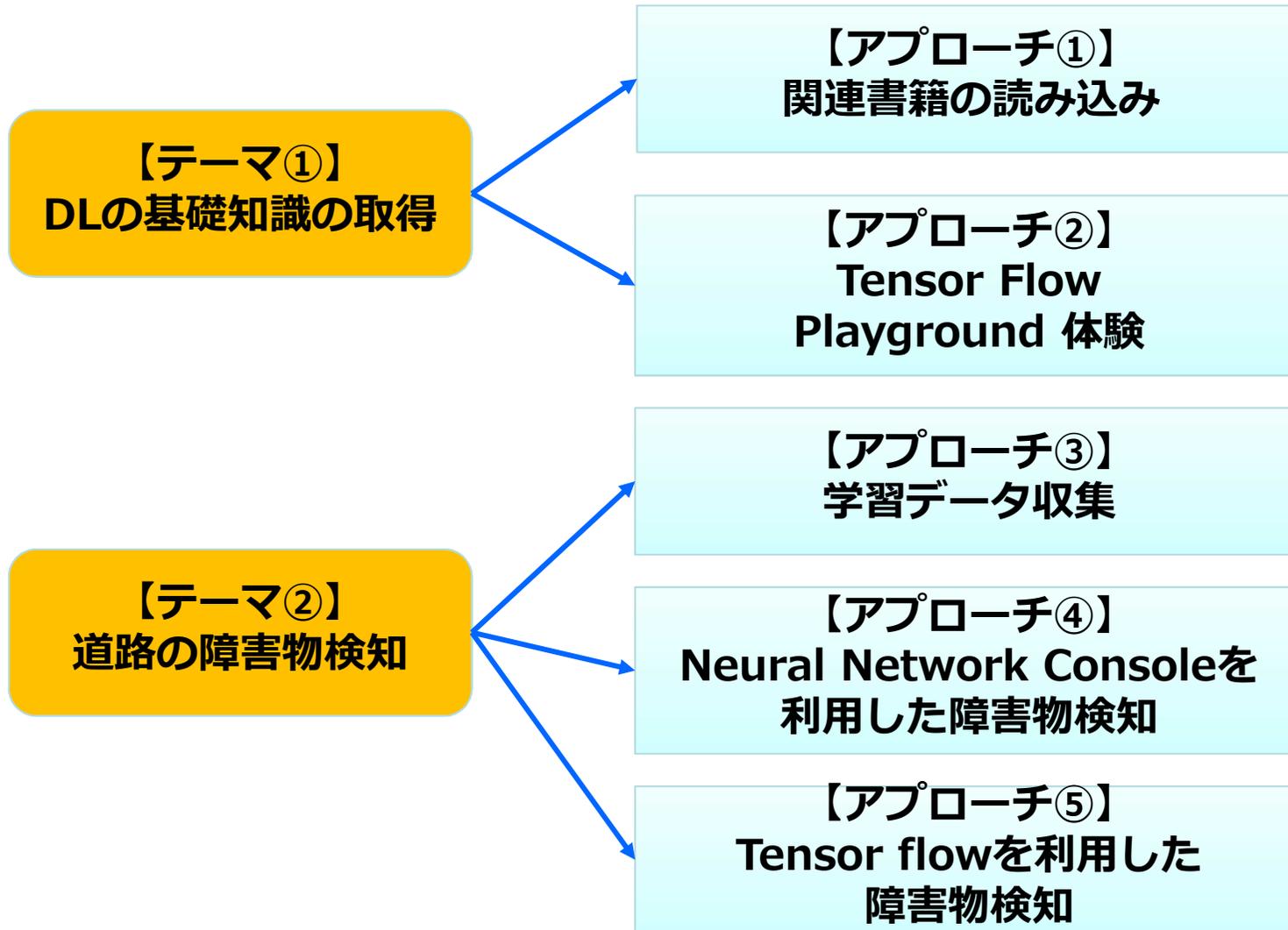
### アプローチ：

- ・各人のスキルセットに合わせて、Deep Learningの基礎知識を得る
- ・道路の障害物などを画像認識で検知できるか試してみる

### 結論（感想）：

- ・画像認識については、あまり業務でも行った事がない内容で新鮮で、セグメンテーションや画像集めなどデータ準備から解析までの流れを理解できた
- ・いきなり99%という正答率が出て、とても焦りました
- ・あまり分科会に参加できませんでしたが、同じ課題意識のある方で意見交換ができて大変良い時間となりました

# アプローチ一覧



# TensorFlowを利用した障害物検知

画像分類モデルInceptionResnetV2をTensorFlow + Kerasで実装し評価

## 教師データ例

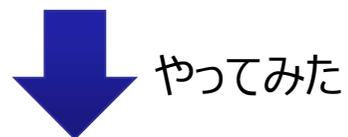
障害なし



障害あり



人は判断がつくが、この教師データで分類が可能なのか？（教師データ1000枚）



**正答率：57.5%（障害なし：40枚、障害あり：40枚）**

## C. 音声

---

### 問題意識：

音声AIは今後の進展が期待される分野であることを背景として、様々な活用における音声に関するAI技術の現状と将来はどうなっているのか

### アプローチ：

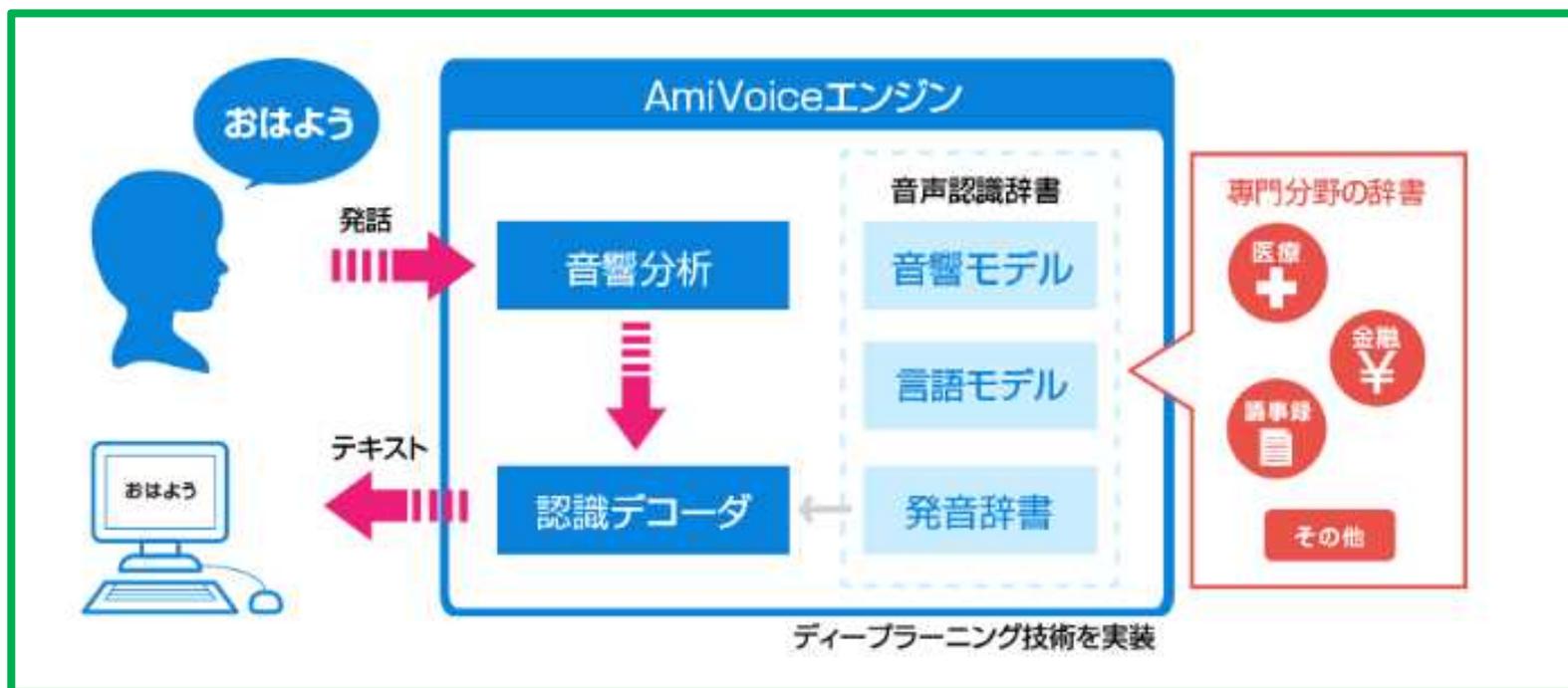
音声認識の基本的な仕組みと、音声を活用したAI技術を調査し、音声認識（話者認識）の現時点での実現度を実際に確認する。そして、音声認識の現状と今後はどうなっていくのかを考察する

### 結論：

現状では、話者認識（認証）という点においては、まだまだ懸念が残っている。特に、録音声で認識されてしまう恐れがあるため、盗聴などには注意が必要。音声を活用した議事録作成ツール・サービスは大規模会議向け・高額であるのが現状だが、小規模向けの製品・サービスも検討されている模様

# 音声認識の基本的な仕組み

コンピュータにとって、音声<sup>①</sup>を正しく認識することは簡単ではない！



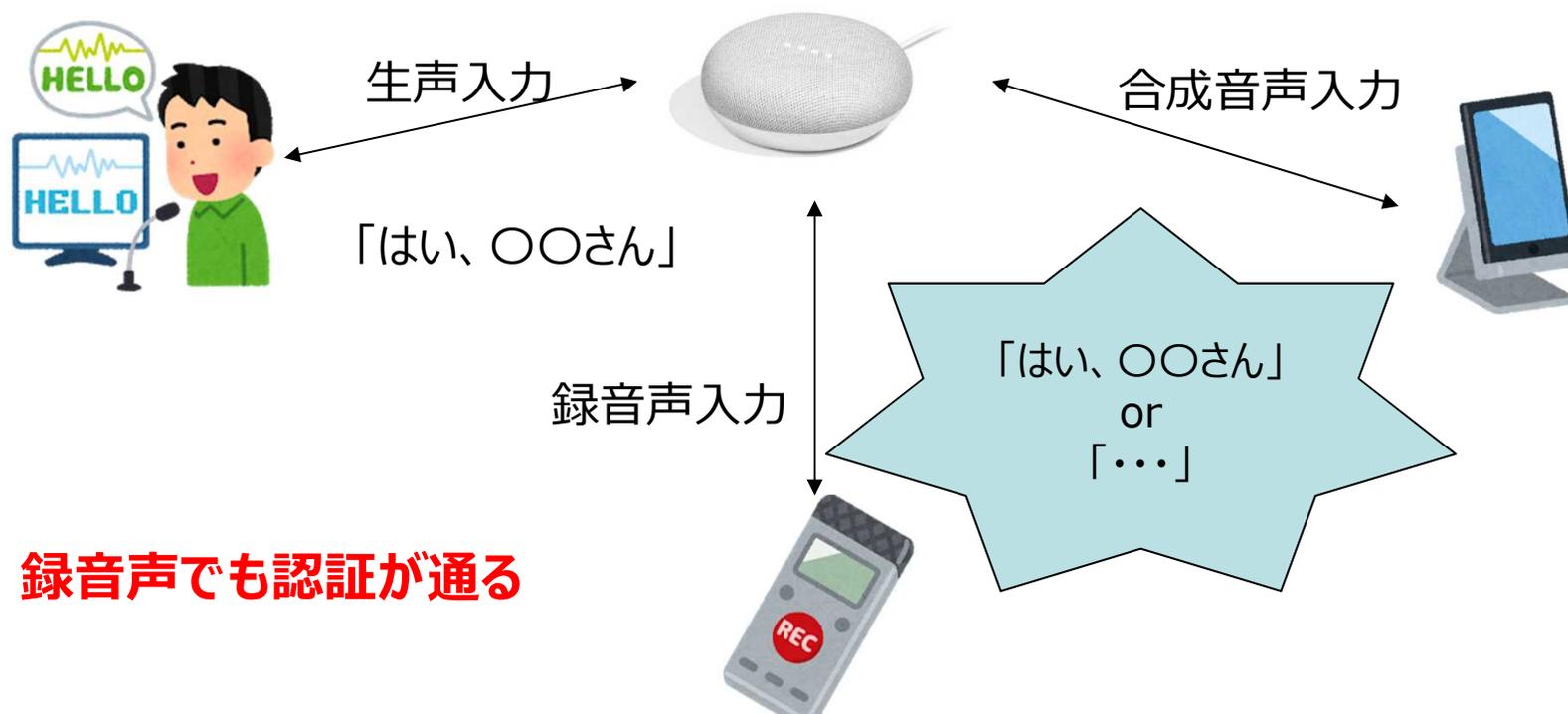
音声認識処理では、声の情報と言語の情報を密接に組み合わせながら、文字へと変換します

(出典：AdvancedMedia)

# 音声認識の検証

【話者認識を合成音声等で欺けるか？】

生声と録音声、合成音声のGoogleHomeMiniによる認識



**録音声でも認証が通る**

- 入力（「オッケー、グーグル」「わたしはだれ」）を認識して返答するか
- 入力媒体（生声、録音声、合成音声）に対して、正しく認識できるか
- 話しかけ語句毎に話者が異なる場合にどのような返答になるか

## d. 数値

---

### 問題意識：

数値オープンデータを使って新たな相関を見出すことができるか？  
また、データ間の相関関係を発見し、メンバーがA Iを実感してみたい

### アプローチ：

家計調査等のe-Stat（政府統計データ）から取得できるデータを中心にRStudioを使った回帰分析を実施する

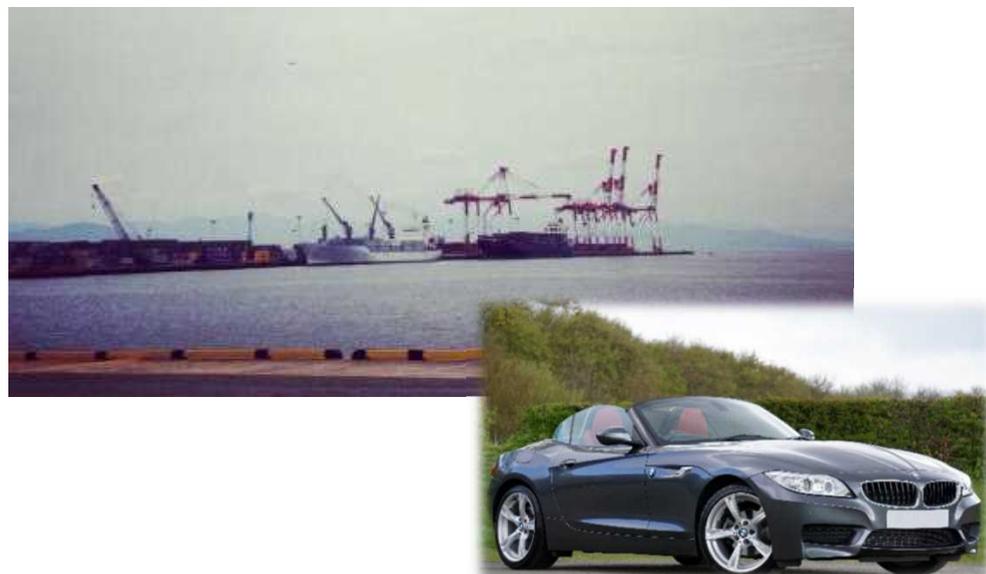
### 結論：

オープンデータの利用方法、利用価値を認識できたが、実際に使いたいデータを見つけ出すのはなかなか難しかった（データの時間粒度等）。相関関係を発見することができた一方で、因果関係に関しては必ずしも明らかではなかった。

# 家計調査と港湾調査のデータの間の特相分析

政府統計名： 港湾調査  
 使用統計データ名： 港湾統計(月報)  
 データ単位： 月次  
 データ期間： 123ヶ月(2005年1月 ~ 2015年3月)  
 使用品目： コンテナ個数

品目	1 食料	2 住居	3 光熱・水道	4 家具・家事用品	5 被服及び履物	6 保健医療	7 交通・通信	8 教育	9 教養娯楽	10 その他の消費支出
コンテナ個数	78.3%	98.8%	22.3%	99.7%	25.6%	83.2%	99.4%	50.7%	35.6%	69.3%
輸出コンテナ数	51.8%	90.6%	40.0%	99.0%	28.5%	75.7%	99.8%	45.8%	23.0%	85.1%
輸入コンテナ数	86.3%	99.4%	27.1%	99.8%	46.1%	85.5%	98.2%	60.2%	36.0%	40.6%



「住居」や「家具」に輸入コンテナ数に関わるのは分かるが、「交通・通信」になぜ関わるのか？

⇒  
 家計調査の「交通・通信」の内訳をみると自家用車が存在  
 推測であるが、輸入車や外車の自動車部品との関係が強いのではないかとチーム内で判断

画像出典: <https://www.comeluck.jp>  
<http://gahag.net/011465-vegetables-paprika-pumpkin/>

# 不動産取引情報、人口動態、賃金動態の間の相関調査

政府統計名： 不動産取引情報、人口動態、賃金動態  
 使用統計データ名： 不動産取引情報(中古マンション)、出生率、婚姻件数、離婚件数、給与水準  
 データ単位： 年次  
 データ期間： 13年(2005年 ~ 2017年)  
 説明変数： 中古マンション取引価格

説明変数	面積	建築年	最寄り駅 目白	用途地域 (第1種中高層住居専用 地域)	リフォーム 済み	最寄り駅からの 距離	調停・競売 等	合計特殊 出生率	婚姻 件数	離婚 件数	建設業 給与	製造業 給与
中古マンション 取引価格	100.00%	99.26%	91.19%	99.60%	98.83%	100.00%	99.99%	78.96%	93.16%	23.60%	54.21%	51.91%



- ・面積、建築数、一部の最寄り駅・用途地域、リフォーム 正の相関関係
- ・最寄り駅からの距離、調停・競売等 負の相関関係
- ・出生率、婚姻件数、離婚件数、建設業/製造業給与 相関関係なし

# おわりに

---

## 皆様に感謝：

熱心に取り組んで頂いた参加者の皆様、  
常にサポートして頂いた日本情報システムユーザー協会の皆様、  
一緒に運営して頂いた監事団の皆様、一年間ありがとうございました！

## AIについて：

AIをビジネスに適用する際に関連する知識やノウハウの範囲は広く、  
まだまだ改善の余地は広いと感じています。ビジネス、実装、技術、運用などの  
領域別の知識と、データ種毎の特性についての考察は有益でした

## 来年度に向けて：

今年度の研究を生かして、ビジネス適用のための研究を深めていきます。  
来年度はビジネス面からの考察の比率を高めていく予定です。  
皆様のご参加をお待ちしております！

---

**ご清聴ありがとうございました**