

CCSE 2019

Conference, on computer science for enterprise

2019.07.13@東大

CCES 2019について

CCESは企業研修に特化したカンファレンスで、最近のいわゆるWEB系企業の中でも比較的イケてる[1]企業が主体となって運営している。このようなイケてる企業では企業内に研究所を持っていることが多く企業のイノベーションを加速させる目的および、企業価値の向上（リクルーティング）のために研究活動を実施しているところが多くある。

CCESはこれらの企業研究の発表の場として学会のような形式でのカンファレンスである。CCESのテーマは主に最近のイケてる企業はまあAIをやっているので、AI分野が中心で機械学習、コンピュータビジョン、IoTなどについて発表がなされた。

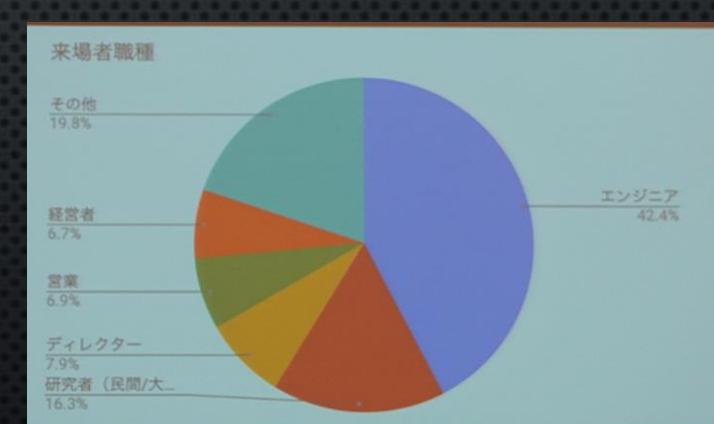
CCESは昨年度から実施されており、今回が2回目である。登壇者は参加者は以下のようにになっている。

●参加者 800人 (一般446人,学生189人)

●参加者属性

学生29.1[%] 社会人: 70.9[%]

エンジニア42[%],研究者16.3[%]..



CCES 2019について

協賛企業 Nがいない＼(^o^)／、GAFAがない。

[1] 大むねWEB系のイケてる企業が主体、イケてるとはある程度自力がついてきたエンジニアが転職したいと思う企業郡である。

協賛企業一覧



CyberAgent.

Rakuten
Institute of Technology



mercari



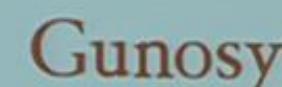
JapanTaxi



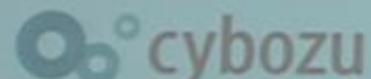
LIFULL



EXAWIZARDS



LayerX



Pixie Dust Technologies, Inc.

RICOH

LEAPMIND

:DeNA

LINE

ABEJA CA ABEJA

iSiD
IT Solution Innovator

Panasonic

発表について

昨年度に比較して登壇している人の数が増えたので、公演は3会場で並列に行われた。発表10分、質疑応答5分の15分の短い発表だったので、結構忙しく聞きたいテーマの発表が会場が別れていたりしたので、結構慌ただしい感じであった。（すぐに場所移動）

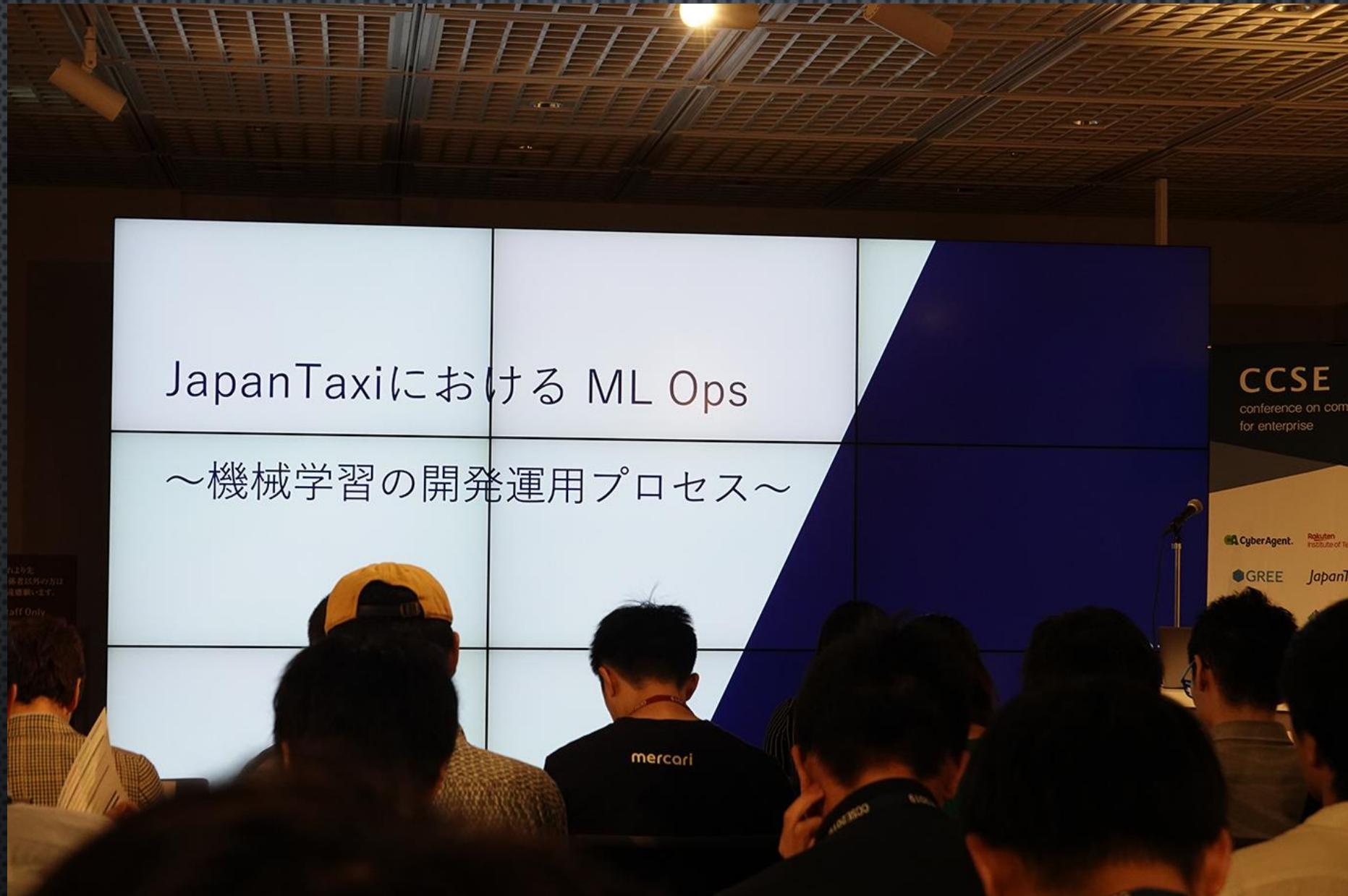
メイン会場以外の発表会場は比較的狭く発表場所によってはよく資料が見えないなどことがあったので、今後運営者様には改善をお願いしたい。

- ・発表全体の所感
 - 発表されている方の職種は大きく分けて3つで 研究開発のマネージャークラスの人、研究者、エンジニアの3種の方がいらっしゃったので、それぞれに応じた内容の内容だったように思う。
あと全般的に学生さんへのリクルーティング的な要素は強かった。
- ・研究開発のマネージャークラス
 - ⇒ 自社のサービス紹介やノウハウの紹介、宣伝、リクルーティング
- ・研究者
 - ⇒ 自分の研究発表
- ・エンジニア
 - ⇒ 運用周りのノウハウや手法の解説。（うちの会社はこんななんしている）

発表について

- ・気になったワードやOSS
基本的に、みんながつかっている、共通のツールや、イケてるOSSが気になった。あと語彙。
- ・Kubernetes
⇒ コンテナ化したアプリケーションのデプロイ、スケーリング、および管理を行うための、オープンソースのコンテナオーケストレーションシステムであって、ぶっちゃけみんなつかってる。
- ・Airflow
⇒ パイプライン処理OSS (snakemakeみたいなん)
- ・Amazon SageMaker
⇒ AWS が提供する機械学習のワークフローをカバーするマネージドサービス。
(まだあまり使われていないような気がするが、使っている人いた)
- ・ML Ops
⇒ 機械学習のモデルの学習、やモデルのデプロイなどを全体のフロー（運用）を実施する言葉。
DevOpsからの派生語。一般的に機械学習のアルゴリズムの実装者とは違い、運用を主として行なうエンジニアおよび、スキーム。
- ・社会実装
⇒ なんだ、この言葉と思ったんだが、というか誰が言い始めたん？という感じで基本的には企業研修などの成果を社会へ中へ活かしていくことだそうだ。
基本的には胡散臭い言葉だと思ったので、「社会実装」とか高らかに主張する人物は、要注意しておいたほうがよいだろう。

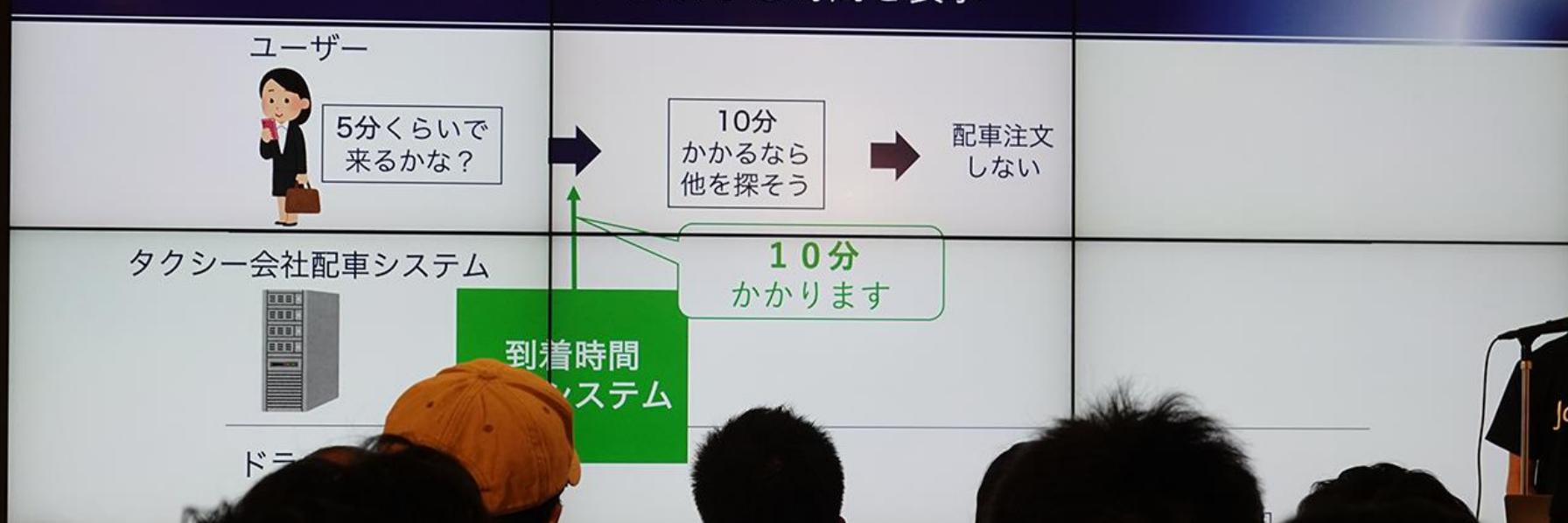
個別発表



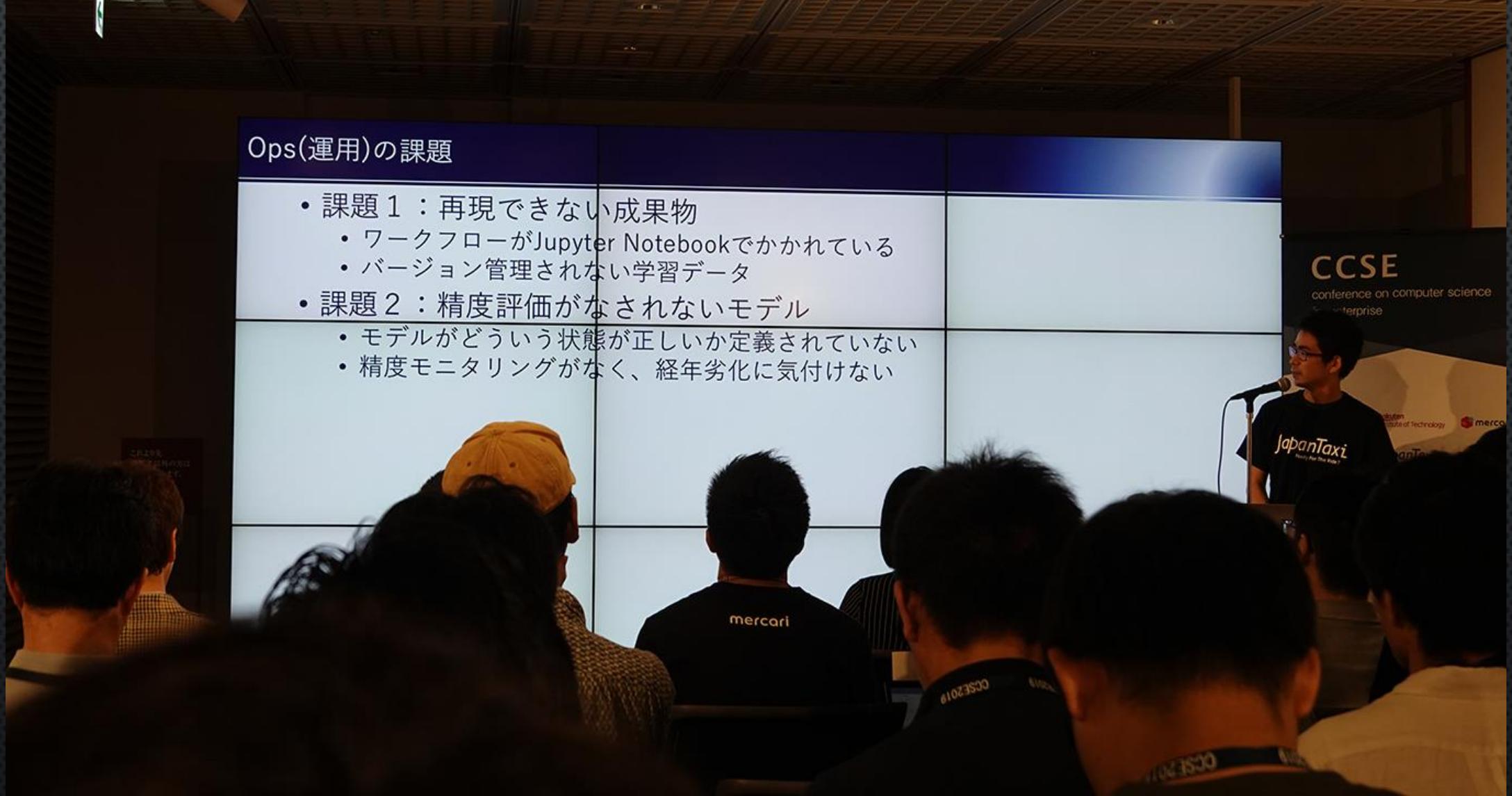
Japan taxiによる、taxiの配車（到達時間予測システム）の運用方法の共有

- Japan taxiはタクシーの配車アプリを提供しているが、ユーザのキャンセルが問題になっており、キャンセルをさせにくくさせるために、タクシーの配車時間を予測するサービスを提供している。この予測システムの構築について発表された。
発表者はML Opsの運用エンジニア。

解決策：期待値調整をするために到着予想時間を表示



アルゴリズム開発者が別人なので、Ops(運用)を行なう上でかなり大変だったとのこと。どこの世界も同じだなー。Jupyter Notebookでアルゴリズムを記述しているやつほんと死んでほしい。



頁1：再現されない成果物

- 対応策

- Jupyter Notebookに書かれたワークフローをバッチ処理化
 - 処理を再実行可能な単位に分割
 - 分割単位：データマート生成のSQL処理、学習データ生成、モデル推定

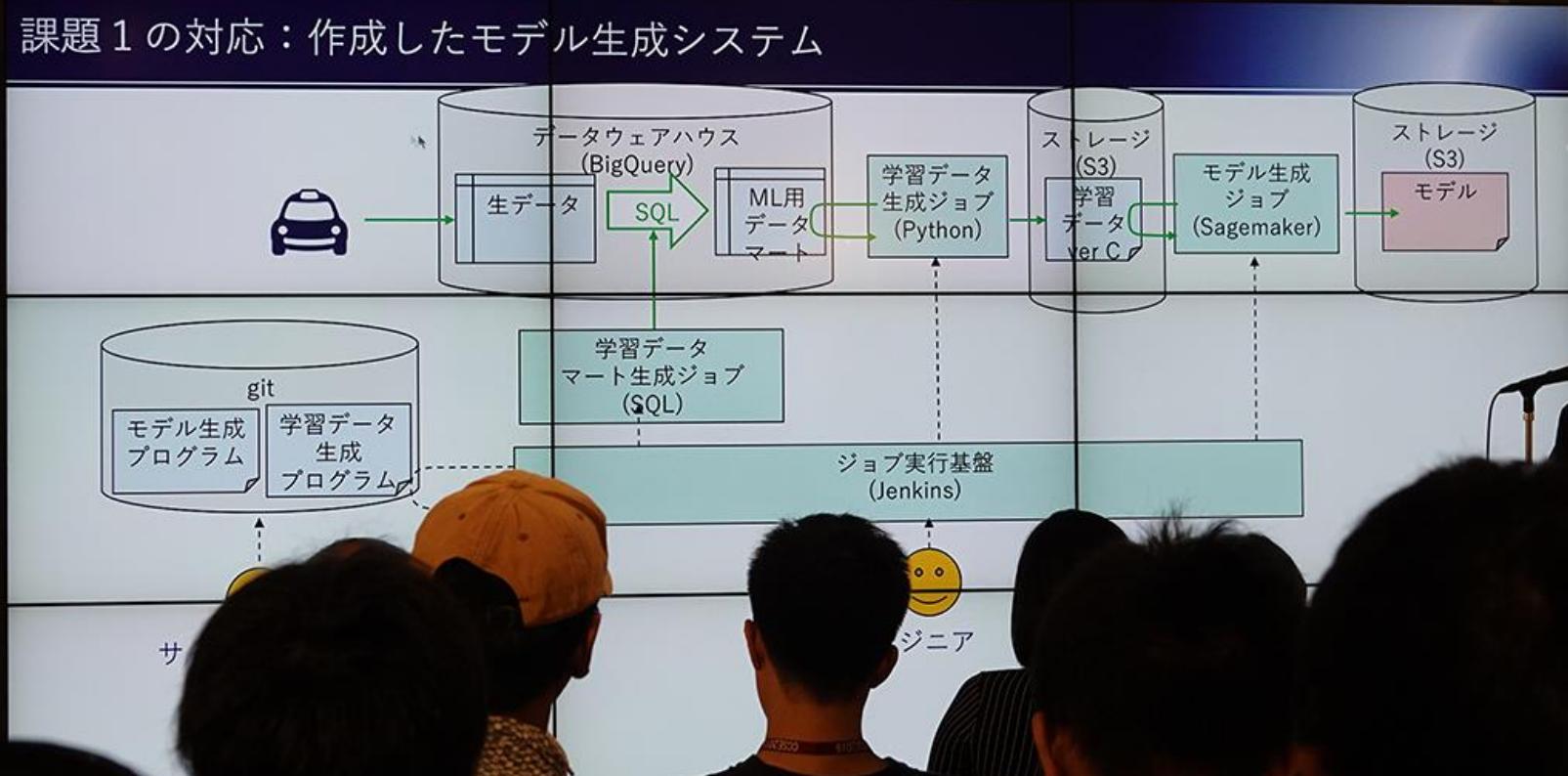
- 運用に必要な機能の追加

- ログの出力、設定をファイルに分離、エラーハンドリング、テストコード

- ML成果物のバージョン管理

- 学習データ：生成したソースコードのgitコミット番号をデータ名に付ける
 - モデル：AWS Sagemakerでバージョン管理

課題1の対応：作成したモデル生成システム



CCSE

conference on computer science
for enterprise



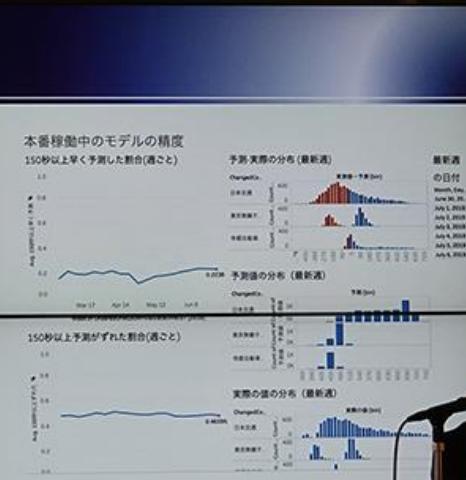
mercori
sansa

ノートブックの関数化、モデルの精度に関しては定量的なチェック機構の導入などを実施しているとのこと。

課題2：精度評価がなされないモデル

・対応策

- ・モデルの精度評価方法の定義
 - ・精度異常しきい値の定義と監視
 - ・5分以上早く予測してしまった割合
 - ・5分以上予測がズレた割合
 - ・予測分布の目視確認
 - ・タクシー会社ごとの予測分布を目視で確認し、前週と比較して大きな違いがないか



CCSE
conference on computer science
in enterprise

Robotics
Institute of Technology
JapanTaxi

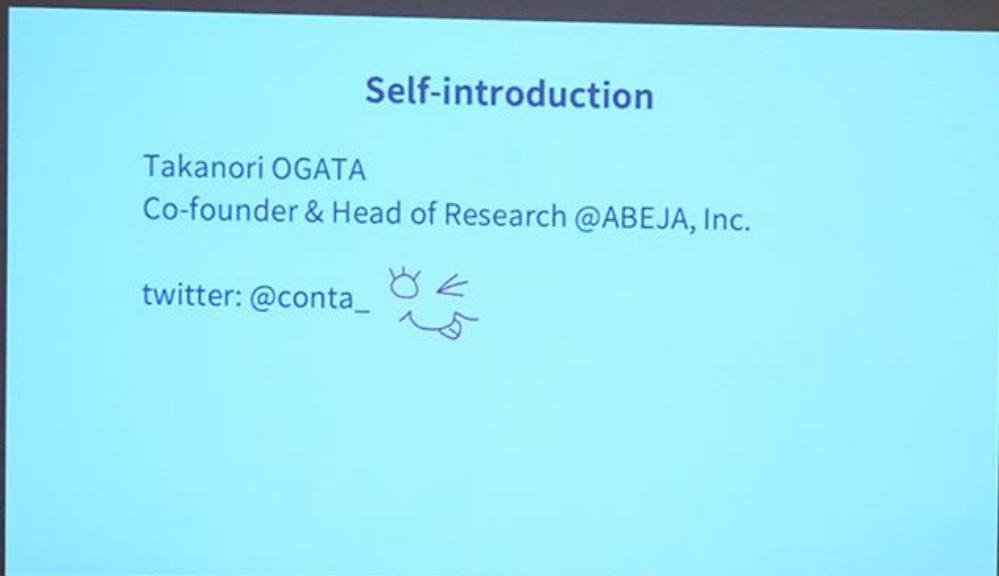


これより先
回答者以外の方は
お答えいただけません



機械学習における教師データ作成とその効率化について ABEJA

ABEJAによる機械学習における教師データの作成の効率化とその効果についての定量的発表。とりあえず、ABEJAすげーぜという宣伝だったように思う。というか、ABEJAってあまりよくわかつたけど、機械学習周りを全般的にサポートするツールや機能を提供している企業だったんすというか、ABEJAがなにやっている会社がちゃんと認識した。



基本的に、イケてる感出している会社だったのだが、そうだったのか
という感じ。



ABEJA は AI-Ops周りをサポートしますよ、共通プラットフォームでね。
という感じであった。たしかに頑張って作っている。(*'ω`*)

AI-Ops for your business

Maintenance & Operation



Model Version Management



Algorithm Development



Data Management



Resource Management



Reduce your tasks,
concentrate on your business

Algorithm Development



ABEJA PLATFORM



MLの開発・運用で起きる課題とは？

おれのかんがえた、さいきょうのかいはつかんきょうで草。

機械学習プロダクトの開発・運用のツラミ

- おれのかんきょうではうごく問題
- データサイエンティスト運用一人プレー
- 顧客: データは大量に持っている！(ただしアノテーションはない)
- データサイエンティスト: え、おれKubernetesとAWS、運用レベルで覚えないといけないの？
- データサイエンティスト: え、おれフロントエンド作るの？
- アノテーションお願いします！->(クオリティーすごい) ->全部俺のターン！(やり直し)
- アノテーション->ディレクトリ名やファイル名が紐付いてなくて紐づけできない！！！！
- 機械学習エンジニアがメンテナンスしやすいコード書いてくれない
- フレームワークのバージョンコロコロ変わる、最新使わせろ！vs 運用考えろ！問題
- いにしえの機械学習モデルはある（ただし学習コードとデータがないので二度と作れない）
- アノテーションお願いします！->:innocent: (クオリティーしゅごい)
- 顧客：データやで->(宅OO便、zipパスワード)、これ毎回やんの？

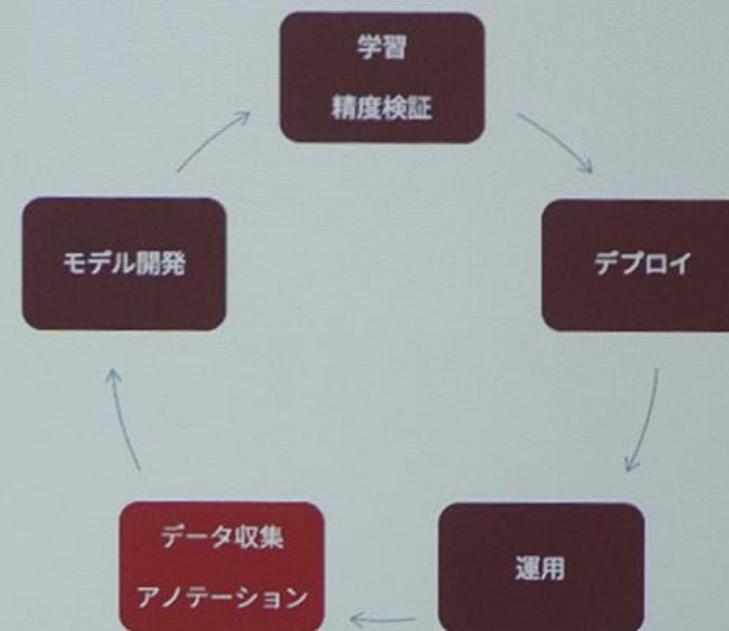
苦節7-8年のつらみ。。。

データサイエンティスト一人無双！！でも草。

CI/CD for ML

データ収集 & アノテーションは
機械学習における上流プロセス
このプロセスを失敗すると
下流のプロセスはほぼ失敗する

どの程度重要なのか？



CIFAR1000（画像のラベルを当てる問題）にランダムにゴミデータを入れた場合の精度劣化について定性評価している。アノテーションの重要性について定量評価している。
※簡単なことなのに以外とこの手の資料見たことなかった。

Effect of data quantity and quality

■データ量に関する実験(Quantity)

- Dataset: CIFAR100
- 学習データの量を変更(20%-100%)

■データ品質に関する実験(Quality)

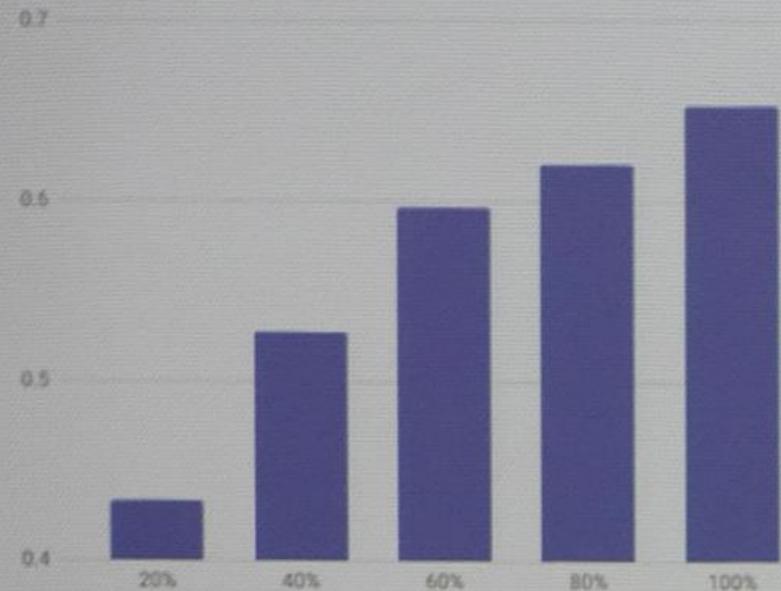
- Dataset: CIFAR100
- アノテーションを特定の割合でランダムに変更(0%-50%)

それぞれテストデータにおいての精度を評価・比較

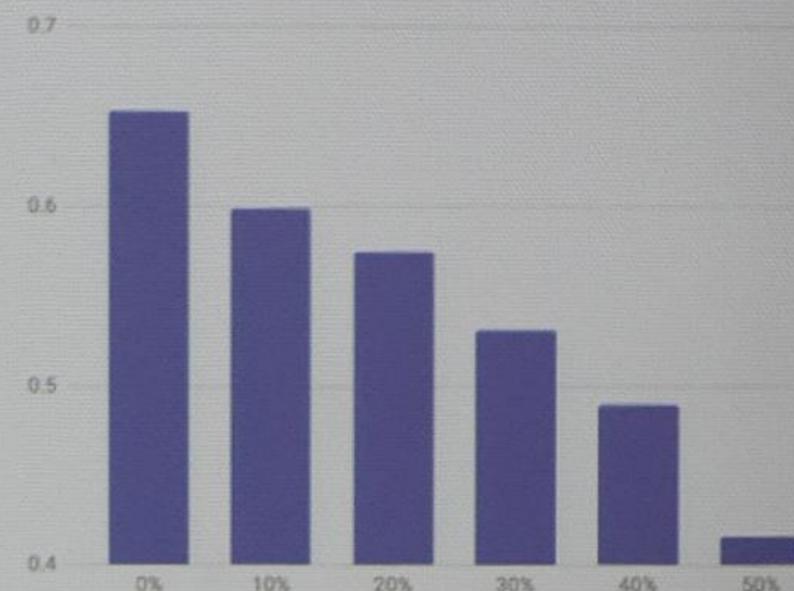
<https://tech-blog.abeja.asia/entry/ml-annotation>

データの量、品質どちらともいいにきまっているが、
品質が悪いデータが混じてくると比例して精度が低下。
データ量に関しては、結構耐久性がある。

Effect of data quantity and quality



データ量の実験結果



データ品質の実験結果

ABEJA いろんなツールあるよ。
アノテーションを楽にできるよー。という宣伝。

Templates and API Backends

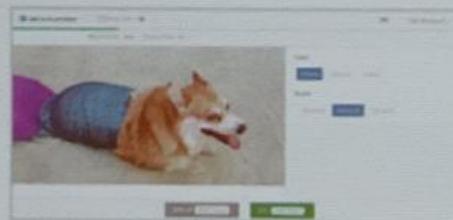
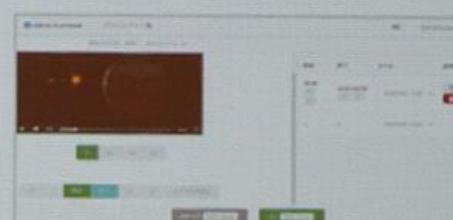


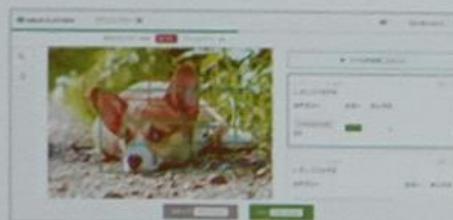
Image Classification



Image Segmentation



Video Scene Recognition



Object Detection

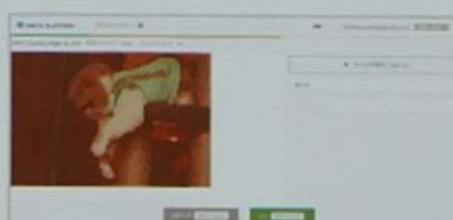
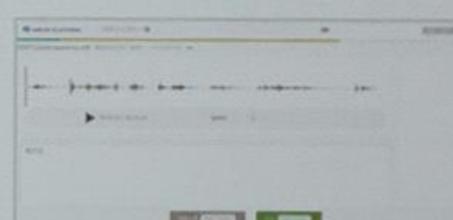


Image to Text



Audio to Text

And more

アノテーション用の論文よんで社会実装しているよーという宣伝。

Extreme clicking for efficient object annotation (CVPR 2017)

BBを直接描画するのではなく、物体の上下左右の4点をクリックするだけでアノテーションを行うExtreme Clickingの提案

35s > 7sでアノテーション可能で
(既存手法と比べてx5) Qualityは既存手法と
同等

また、アノテーションがGrabCutに適用でき、
BBでSegmentationしたものよりも精度良い

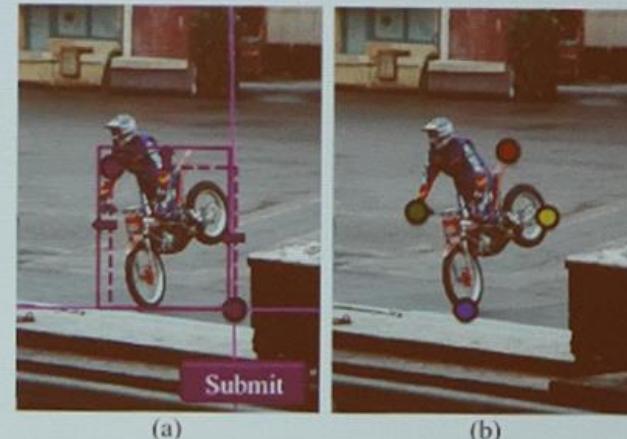


Figure 1. Annotating an instance of motorbike: (a) The conventional way of drawing a bounding box. (b) Our proposed extreme clicking scheme.

Papadopoulos, Dim P., et al. "Extreme clicking for efficient object annotation." Proceedings of the IEEE International Conference on Computer Vision. 2017.

いちからアノテーションしてもらうじゃなくて、スマールセットで推論したデータをアノーターの人に直してもらうようにしている。（あたりまえだけどそっちのほうが楽だよね）

Active Learning

- できる限り精度に影響を与えるデータを選ぶ
- 選択されたデータを人間に問い合わせ、得た正解を訓練データとして追加
- => このサイクルを定期的に行うことで効率化を図る

1.2 Active Learning Examples

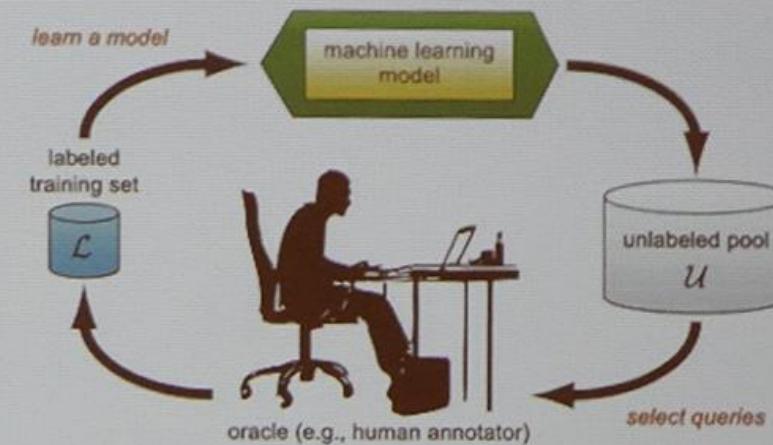


Figure 1: The pool-based active learning cycle.

[Settles, 2009]

インドのナンバープレート認識

インドにおける一般道の ナンバープレート認識について

株式会社エクサヴィザーズ

遠藤太一郎

CCSE conference on computer science
for enterprise



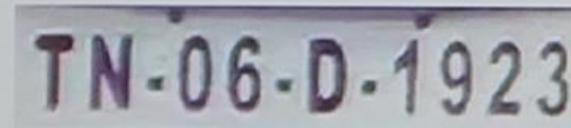
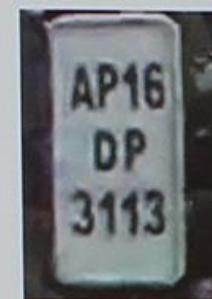
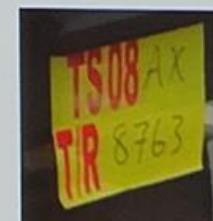
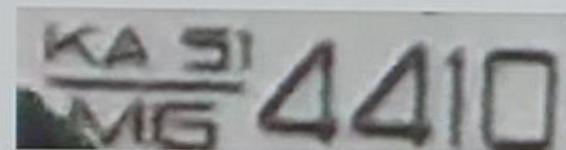
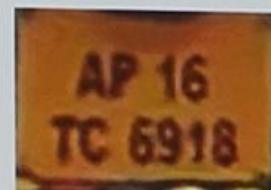
インドのナンバープレート 自由すぎワロタ！WWWWW

PowerPoint ファイル 編集 表示 挿入 フォーマット 配置 ツール スライドショー ウィンドウ ヘルプ 75% A 7月13日(土) 11:07 Taichiro Endo

Mobility

インドのナンバープレートは規格化されていない

日本で使える画像処理的なアルゴリズムを活用することはできない



ナンバープレートの認識、違反者の追跡などを通じて、渋滞緩和をなくしたい。
(国からも支援されているらしい)

PowerPoint ファイル 編集 表示 挿入 フォーマット 配置 ツール スライドショー ウィンドウ ヘルプ 74% A 7月13日(土) 11:09 Taichiro Endo

Mobility

アンドラ・プラデシュ州政府に一括としたソリューションを提供

データ収集
(動画像解析)



解析、予測

交通量／速度



最適化

新しいインフラ設備の際の
都市計画や提案
・新しい"Class 3"の道路を
設けたほうが良い 等

違反者の可視化



違反者への自動的な通知

Y. Ravi Bankar Reddy,
Dy Commissioner of Police
Traffic, Vijayawada City

09666 - 3465600
Cell: 933993114
trafficcp.vj@gmail.com

Letter of Appreciation

TO WHOMSOEVER IT MAY CONCERN

On behalf of the Vijayawada City Police, let me express our sincere gratitude to [KooWards Inc.](#) and [Efftronics](#) for the technical solution provided to the Vijayawada Traffic Police. The ANPR solution has been deployed on 4 cameras in Bandar Road where state-of-the-art deep learning technology has been applied to CCTV footage to analyse the number plates of the vehicle passed in the intersection. Utilizing the data provided from [KooWards Inc.](#), together with [Efftronics](#) has developed a scalable dashboard which promptly analyzes the traffic volume and traffic flow at each intersection where the cameras has been deployed.

The project's efforts will help the Vijayawada State Police in better planning of Traffic to reduce the congestion. Based on the achievements of the project in providing technical assistance to the traffic authorities, we would like to continue cooperation with [KooWards Inc.](#) and [Efftronics](#).

Deputy Commissioner of Police,
Traffic, Vijayawada City.



science

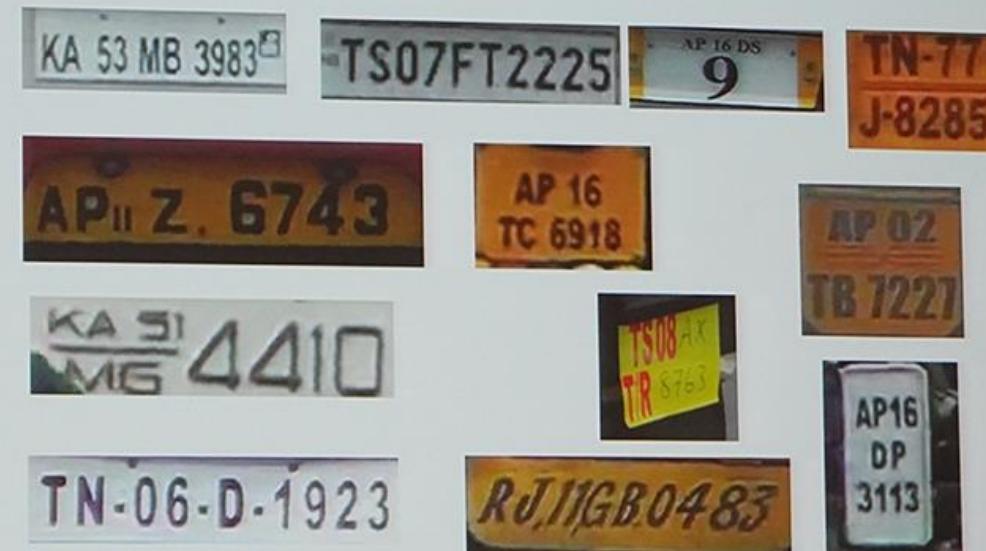
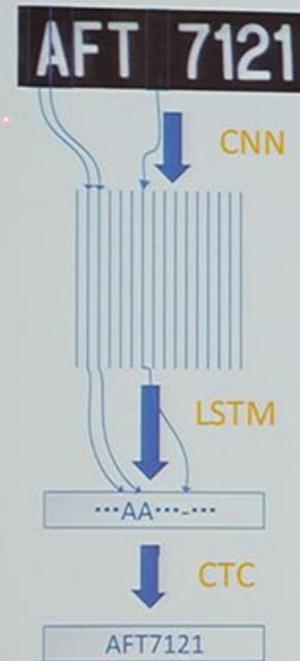
JapanTaxi

LayerX

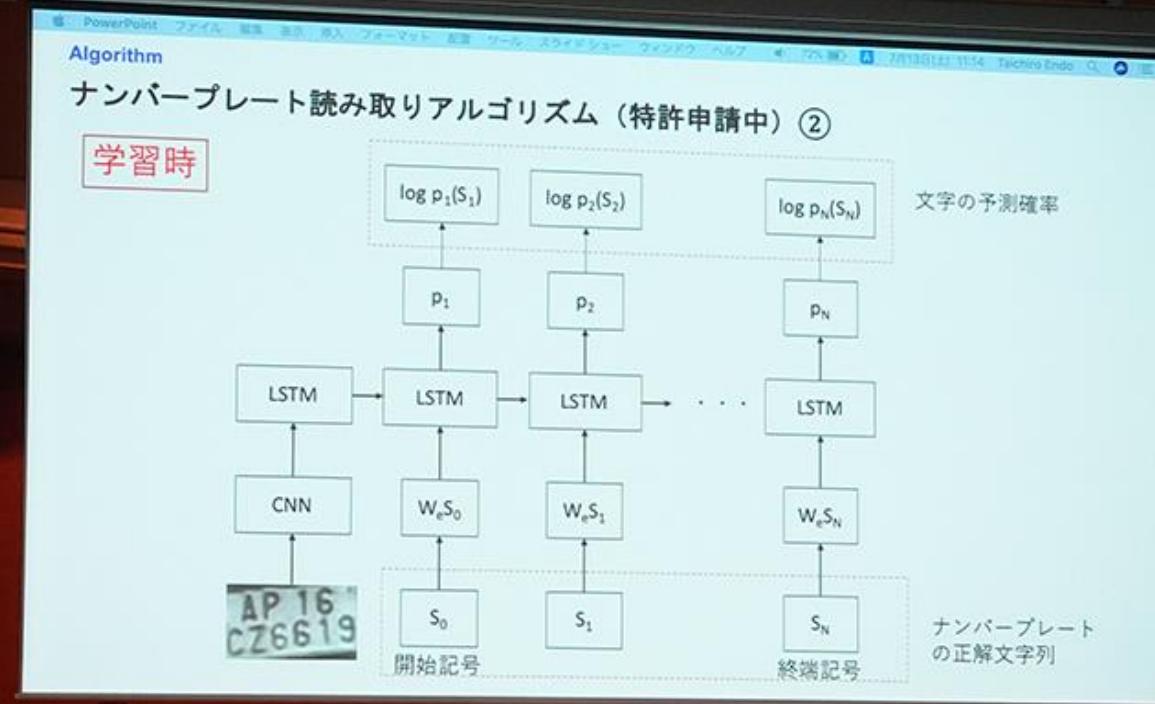
CTC は音声認識で用いられる手法。

Algorithm

Deep Learningを使ったものでも、従来の手法では2段のケースに対応できない



LSTMに入れているけどナンバープレートって数値の羅列だから意味ないじゃないという質問者が居たがオラもそう思った。



講演者は、「ナンバープレートにも意味がある（州を表すなど特定の記号）があるのでよい」といっていたが、州番号とか自明だからそこはルールいいのでは？

Neural Architecture Search とはネットワークのニューラルネットワークの構造自体やパラメーターをまるごと最適化する手法。（そんな大変な！）

C2Cマーケットプレイスにおける Neural Architecture Searchの活用事例

株式会社メルカリ
AI Engineering Team
Software Engineer Intern
熊澤 大樹

CCSE
conference on computer science
for enterprise



アーキテクチャパラメータと損失関数を同時に最適化するよ。
(計算量すごそう。実際に大変)

| Architecture Search問題とは | |
|--|--|
| <u>一般的な学習問題</u> : | 重み w で parameterizeされるモデル : $M = \{f_w(\cdot) : R^n \rightarrow R^m\}$ |
| 入力を x 、出力を y 、ロスを L とすると解きたい最適化問題は | $\min_w L(f_w(x), y)$ |
| <u>Architecture Search問題</u> : | アーキテクチャパラメタ α を含むモデル $M = \{f_{\alpha, w}(\cdot) : R^n \rightarrow R^m\}$ に対する重み w とアーキテクチャパラメタ α の両方を同時に最適化したい |



目標 タスクは違反品の抽出

違反品は法令のによって日々変化するので迅速な対応が必要（チケット転売など）

| Use Case 1: 違反検知モデリング

問題設定：利用規約や法令に違反する出品を早急に検知・削除したい

- 理想: お客様の目に規約違反出品が触れないようにする
- カスタマーサポートが人力で全ての出品をチェックするのは非現実的
- 機械学習エンジニアのタスクは、特定の違反に該当する出品を高いPrecisionで見つけること（意訳：もし誤って正常出品を削除してしまうと、出品者のお客様体験は悪い）

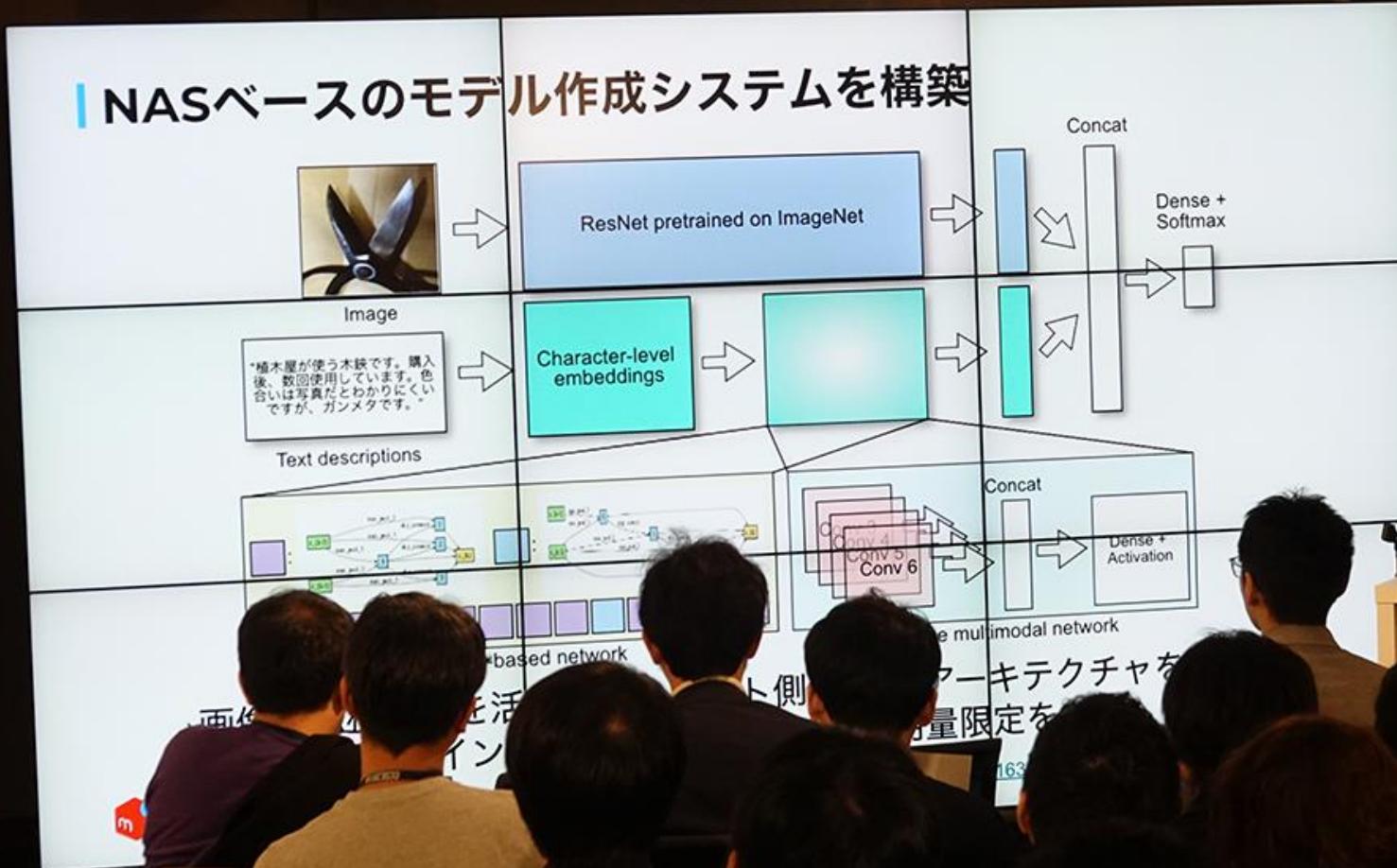
課題

規約違反の検出が可能なモデルを開発したい
規約違反の検出を機械学習で実現したい
規約違反の検出を機械学習で実現したい

CCSE
conference on computer science
for enterprise



ResNet+ 1 character embedding のようよくみえない。



CCSE
conference on computer science
for enterprise



よくなつたらしい。

| 実験結果とその効果

| | | Category | A | B | C | D | E | F |
|-----------|-----------------|----------|--------|--------|--------|--------|--------|---|
| Precision | Proposed method | 0.957 | 0.975 | 0.988 | 0.988 | 0.992 | 0.992 | |
| | Baseline model | 0.930 | 0.961 | 0.980 | 0.965 | 0.989 | 0.978 | |
| | Δ | +0.027 | +0.014 | +0.008 | +0.023 | +0.003 | +0.014 | |
| Recall | Proposed method | 0.539 | 0.851 | 0.938 | 0.901 | 0.965 | 0.937 | |
| | Baseline model | 0.388 | 0.818 | 0.891 | 0.708 | 0.954 | 0.879 | |
| | Δ | +0.151 | +0.033 | +0.047 | +0.193 | +0.011 | +0.058 | |

*正（違反）クラスの閾値を0.9に設定したときのPrecisionとRecallスコア

- 各カテゴリについて旧モデルからの改善を確認
- 内製のデータETLプラットフォームに組み込むことで、データ取得プロセスをクロスデータソースまでの全行程を自動化
- 新規データソース登録後、プロイ完了までにかかる時間は月単位→週単位

CCSE

conference on computer science
for enterprise

CyberAgent.
GREE

LIFULL

POSY
WIZARD5



エッジデバイスの最適化を実施したい。

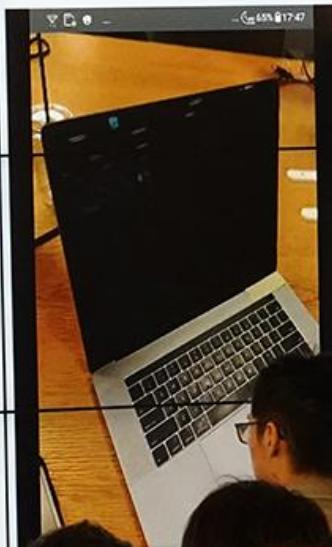
特にエッジでは内蔵されている推論器など様々なので各種デバイスにあった最適化を実施したい

Use Case 2: Neural Networkのデバイス最適化

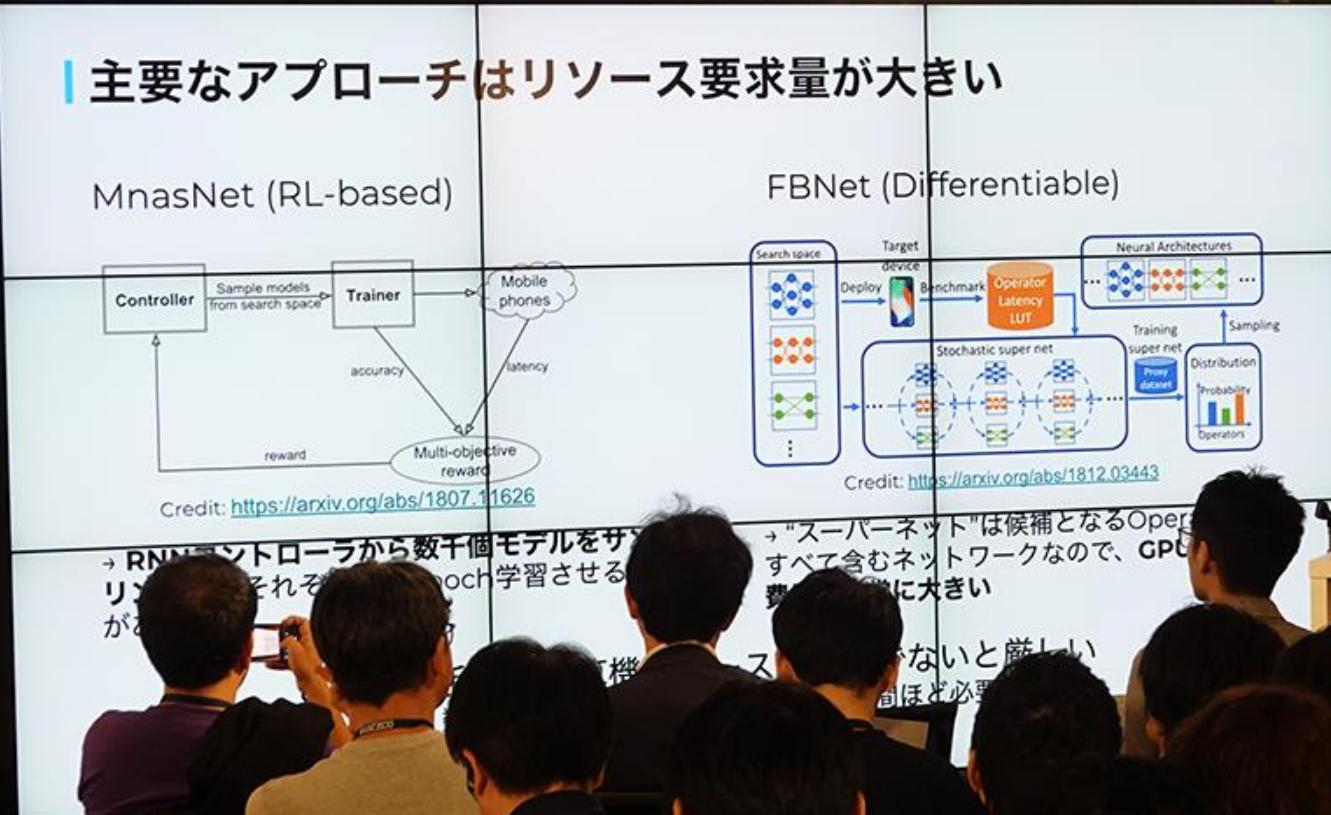
問題設定：EdgeでのNeural Network推論を行いたいタスクが存在する

- リアルタイムに変化するUX実現のためには厳しいLatencyの制約が存在
- デバイス・チップセットの世代や組み合わせは多種多様であり、“One-size-fits-all”なモデルは存在しない (See: <https://arxiv.org/abs/1801.06637>)

→ Multi-Object Detectionを用いて、エンジニアリングに対する変化を最小限に抑えつつ、各デバイスごとに最適化を行っている



MnasNet/FBNetというものが最適化法で有名らしいが計算コストがえげつない1000GPUもってこい。



Single-path NAS というのなら比較的コストがひくく実施可能なのでこれでやった。
(最適化パラメータをCNN内部で表現できる形式のようちょっとよく調べないとわからん)

選択した手法と実験結果

Single-Path NAS

(1) Prior NAS work:
Multi-path search space

(2) Proposed NAS method:
Single-path search space

- NAS problem: expensive path-level selection
- Supernet: each op as separate path/layer
- #parameters/layer: all weights across all paths

- NAS problem: efficient kernel-level selection
- Supernet: all ops in single "superkernel" layer
- #parameters/layer: #weights of largest op only

Reference: <https://arxiv.org/abs/1904.02877>

| Device | SoC Generation (Snapdragon) | Model | ImageNet Top-1 Accuracy* | Latency (ms)* |
|--------|--------------------------------|-------------|-----------------------------|---------------|
| A | 845 | SPNAS | 74.48 | 77.90 |
| A | 845 | MobileNetV2 | 71.80 | 76.36 |
| B | 808 | SPNAS | 73.07 | 113.92 |
| B | 808 | MobileNetV2 | 71.80 | 162.82 |
| C | 670 | SPNAS | 73.15 | 92.14 |
| C | 670 | MobileNetV2 | 71.80 | 111.85 |
| D | 801 | SPNAS | 71.93 | 84.65 |
| D | 801 | MobileNetV2 | 71.80 | 111.85 |

→ 候補の全kernelをルタ内にconvolutional層で大きく削減することに成功したTPU v2

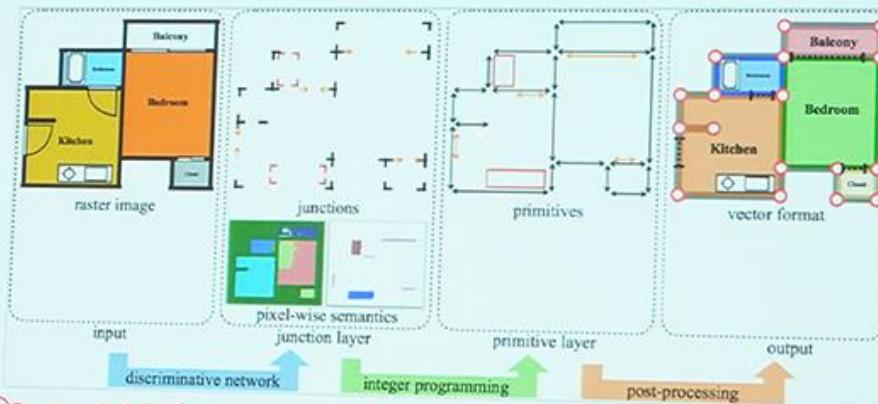


不動産系への応用 間取り図から部屋のセグメントを推論する



うーん、微妙かも。とりあず、不動産の情報からなにかすることはたくさんありそう
(谷脇さんに相談) (*'ω`*)

Key idea



- ①ResNet-152による
セグメンテーション・
ジャンクションポイント
検出
- ②整数計画法による
制約の導入

CCSE conference on computer science
for enterprise



CCSE

conference on computer science
for enterprise

経営者が考える企業の研究とは

東京大学 伊藤謝恩ホール



Sansan 株式会社
取締役

常樂 諭



株式会社サイバーエージェント
取締役

内藤 貴仁



株式会社エクサウェイズ
代表取締役社長

石山 洸

午後からのメインのトークセッション（みんな集合していた）
研究チームを持っている会社の経営者層の人のトークセッション



トークセッションは各社 10分ぐらいで会社や研究開発の説明を実施
最初は CyberAgent
(Cyber Agentさんは相当いけてるだろうという感じなの要期待)



CCSE conference for enter

CyberAgent. Ropponen Institute of Technology

sansan LIFULL EXAWIZAROS LINE

C.A.ってCygamesの親会社だったんだーとたまに思う。例。

事業領域



インターネット広告



メディア



ゲーム

C.A. は AILab というのがあって、75名在籍してるらしい。すごい人数だな。
あと職場環境とか羨まします (*'ω`*)

R&D組織

R&Dメンバー計：75名
- 博士取得者：16名
- 社会人博士：4名



CCS

CyberAgent.
sansan

AI Lab



ence

Taxi

ayerX

C

CA Cyb

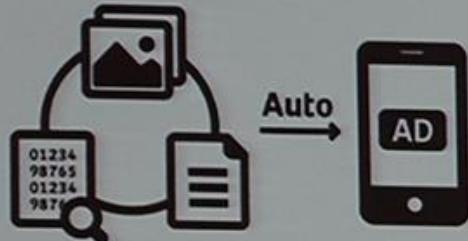
sansar



研究分野

広告を作る

Creative Research

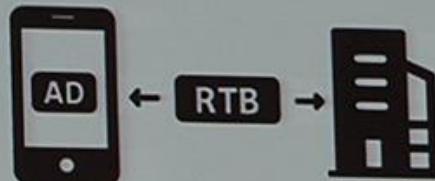


自然言語処理
コンピュータビジョン
グラフィックス
機械学習

コンピュータグラフィックス
スキャニング
モデリング

広告を分析・評価する

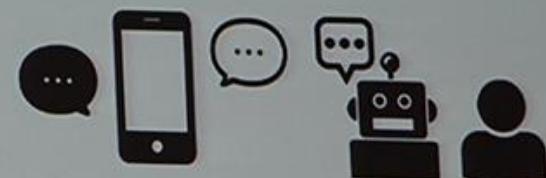
Advertising Economics



因果推論
計量経済学
機械学習
最適化

広告を開拓する

接客対話エージェント



自然言語処理
HAI
HCI
仮想エージェント
社会心理学
機械学習



基本的に全般的に研究しているが、やっぱりアドテク方面に力を入れているっぽい。
広告の自動生成などはやっぱり興味があるようで。リリースできる段階まできたとおっしゃっていた。
既存のアドのアノテーターは数百人規模でいるらしい。（沖縄などでやっているっぽい）
そこを自動化したいらしい。（まあそうなるわな）

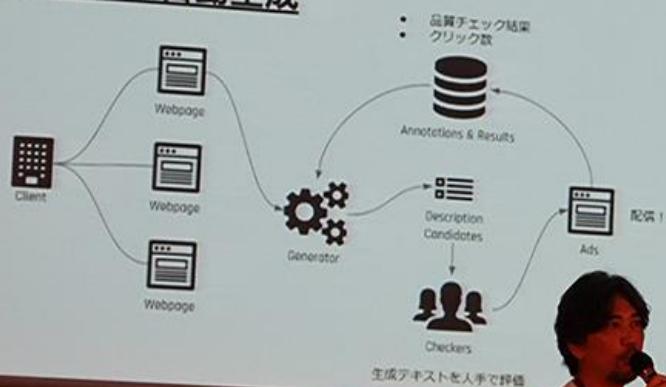
広告テキスト自動生成、バナーの構成自動化（アドビ先生がやっているやつな）

広告の自動生成に関する研究例

- バナーの人物自動差替え
- ポーズの考慮 等



広告テキストの自動生成



フォトグラメトリー技術



3Dモデルの研究



science
PanTaxi
LayerX
BEJA

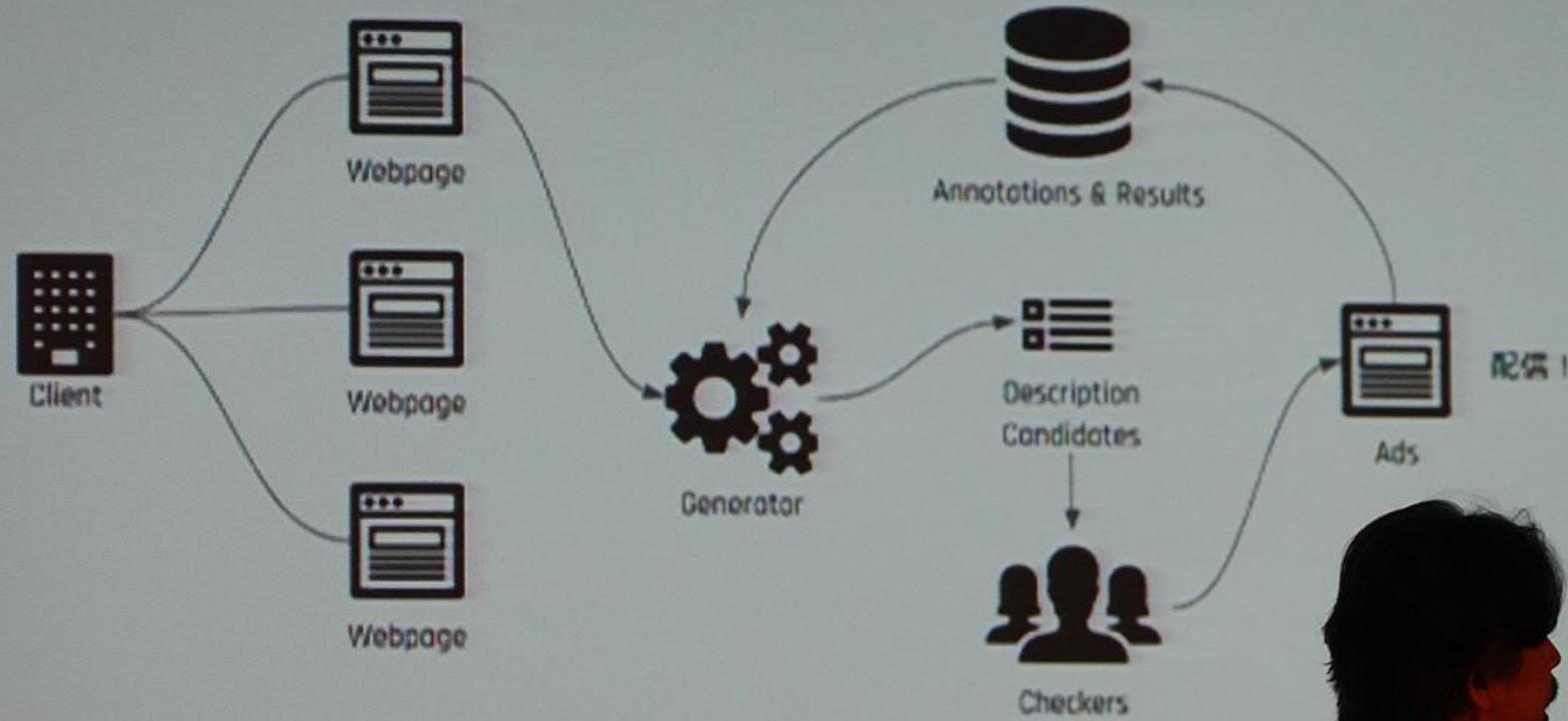
A man in a white t-shirt is speaking into a microphone while standing in front of the presentation screen.

DSC

A man in a black t-shirt with 'DSC' on it is seated in the audience.

広告テキストの自動生成

- 品質チェック結果
- クリック数



生成テキストを人手で評価



研究開発メンバはほんとに研究だけやっているメンバーもいるらしい
その人たちはどちらかというとトップカンファレンスに採択されるがの目標らしい。

研究成果

研究成果は積極的に外部発表へ。
各学術カンファレンスへ積極的に投
稿

▼投稿先一部

| | | |
|------------|-------|-------|
| CVPR | ICML | ADKDD |
| AAAI | HAI | MLSE |
| IJCAI-ECAI | MIRU | HRI |
| CHI | IROs | HCII |
| SIGGRAPH | ICDAR | JSAI |



Sansan さん

この会社、私のイメージでは完全に

<https://www.youtube.com/watch?v=ZkLqWeky5ml>

Company Profile

| Sansanの紹介



これまたいい環境。うっほ。

Company Profile

| 働く環境



Space ONE



Sansan Innovation Lab

uter science

EE JapanTaxi

y LayerX

CA ABEJA

CCSE

CyberAgent.

sansan

cy

De

トークセッション内容

内容は司会者の質問に登壇者が経営者が答えるという形式であり。テーマは「経営者が考える企業の研究」というテーマ。

Qes :

なぜ企業で研究を実施しているのか？

Ans :

三名それぞれ、危機感がまずひとつ、GAFAを代表するような大手の資本の会社に対応すべきためにこのままの経営では危機感がある。技術的にも人材的にもそれで研究を実施することで次の技術やトレンドの先取りするというのと、研究を実施できるようなよい人材を確保するのが目的。また企業のブランディングの一貫でもやっている（これも広い意味でリクルーティング）

Qes :

研究者の評価や研究の意味についてはどのように判断しているのか？

Ans :

Sansanさんはあくまで実装して既存のサービスにのせられるかどうかをまずみている。

Qes:

研究チームの社内外の軋轢や摩擦についてはどのように解消してるか？

Ans:

研究職や一般のエンジニアとの軋轢については、C.A. ではロケーションを別しているので直接両者が合わないようになっている！（すばら）社外との調整（共同研究）などは最初に意識合わせ（目線合わせが）重要。

ドライブレコーダーの映像から、ガソリンスタンドの料金を読み取る研究

ドライブレコーダーの Scene Text Recognitionにおける Multi-task Learning

次世代モビリティ事業部 モビリティ研究開発グループ

高橋 文彦

JapanTaxi

CCSE conference on computer science
for enterprise



ドライブレコーダーの映像からガソリンスタンドの料金を読み取りたいらしい
※ガソリンスタンドの料金を情報として提供したいらしい。
Japan taxiさん

ガソリンスタンド料金の自動認識



JapanTaxi

10

CCSE conference on computer science
for enterprise

CyberAgent. mercari GREE JapanTaxi

sansan TIS Gunosy LayerX

cybozu LINE ABEJA CA ABEJA

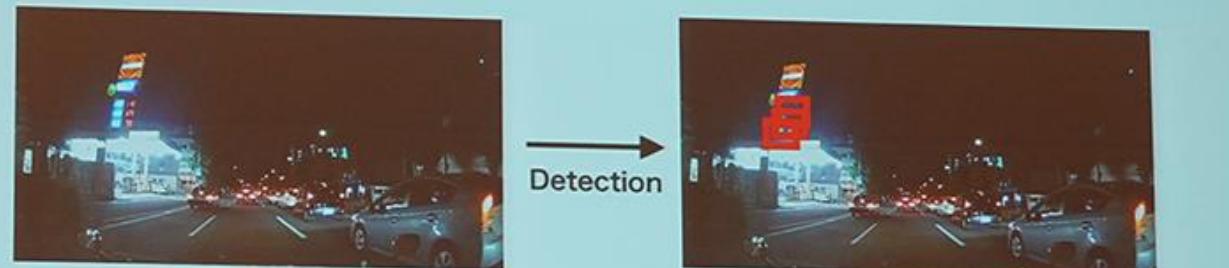
DeNA masonic RICOH

mercari



まず標識を Detection てから、数値を読む
問題としては Scene Text Recognition という問題に帰着される。

Scene Text Recognition



118 129 98

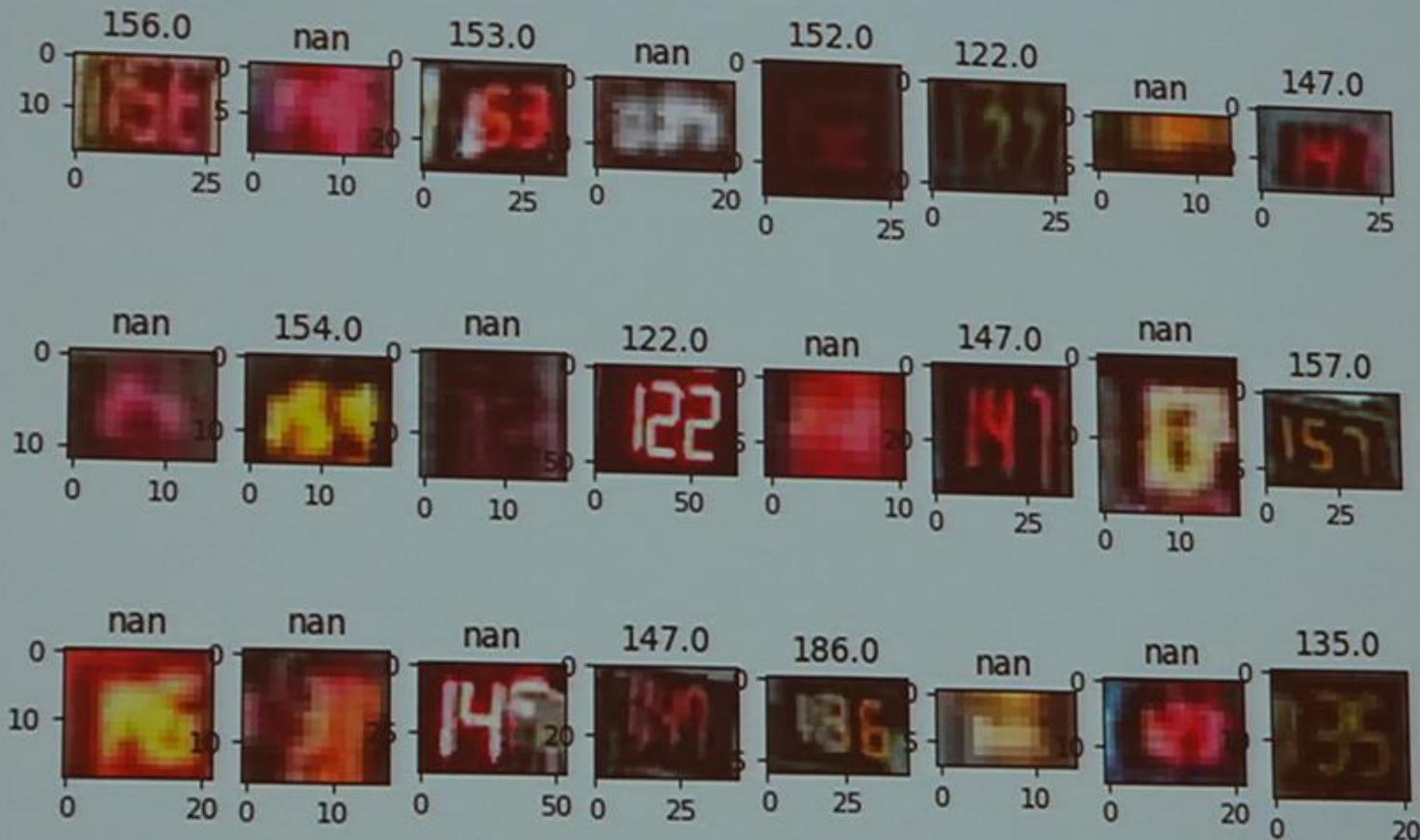
Recognition ← ↓ →

118 129 98

CCSE conference on computer science
for enterprise



ガソリンスタンド料金認識は難しい - 解像度の低さ -



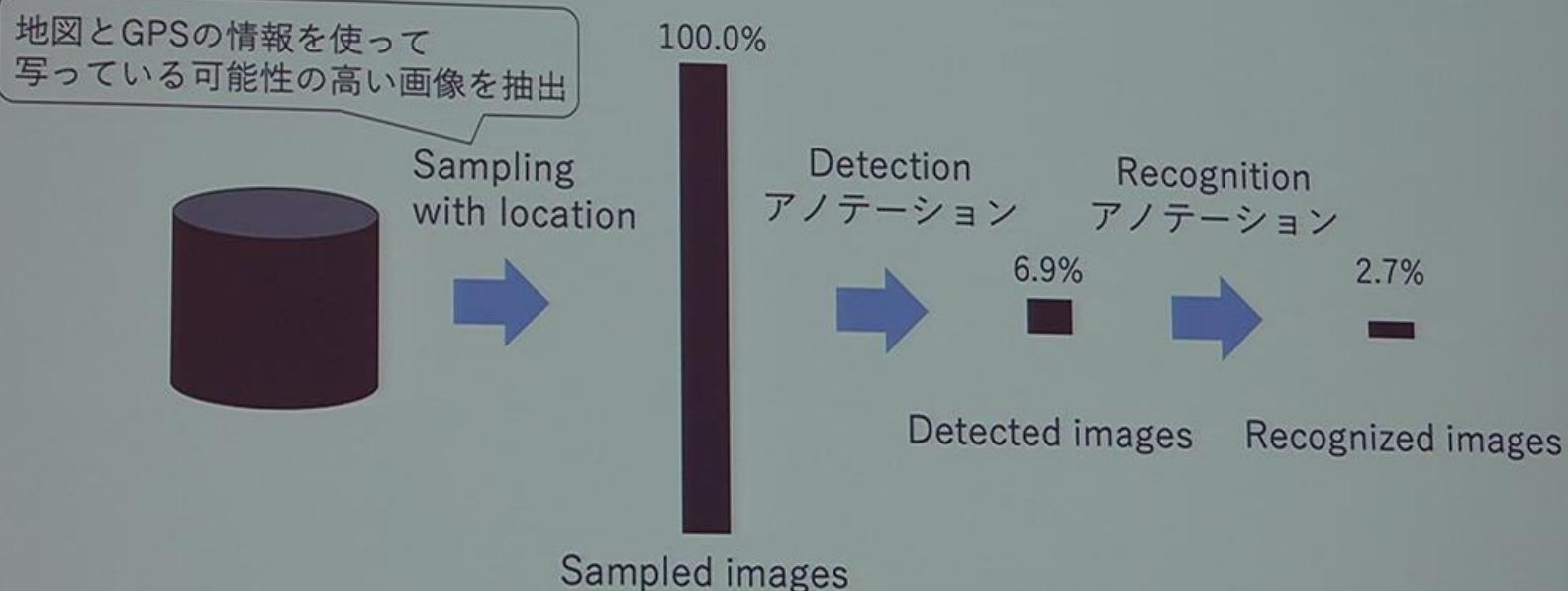
そうそう、数値なので文脈がない
インドのやつ文脈どうのこうのいっていたけど、、番号とかLSTM無理やろ！

ガソリンスタンド料金認識は難しい – 前後文脈がない –



単語は前後の文字から推測できるが、数字は推測できない

課題：学習データを収集するのが大変



地図を利用しても**3%程度**しか学習データが集まらない

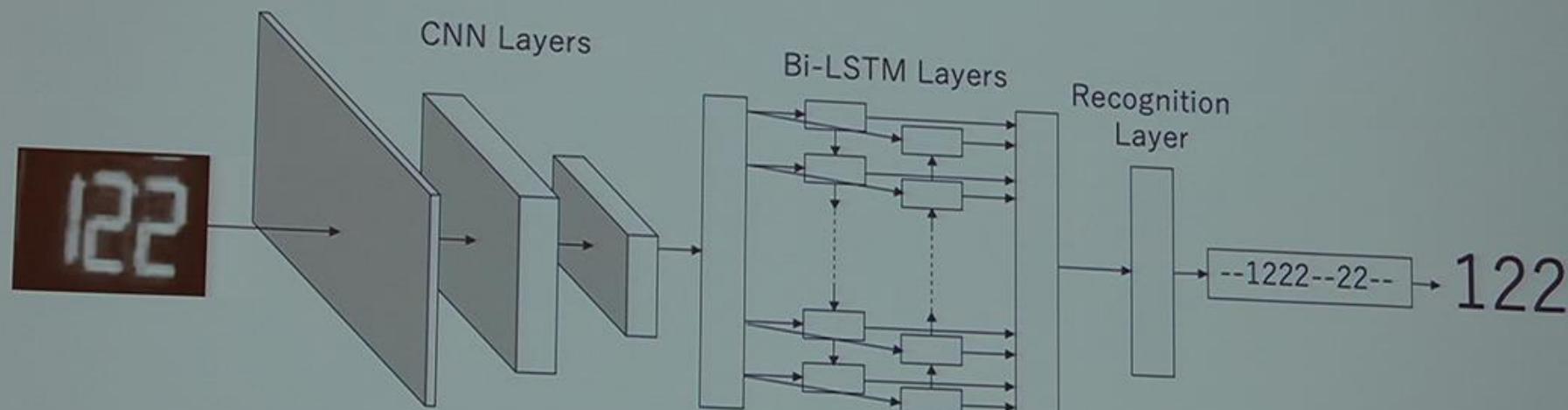
JapanTaxi

14

ガソリンスタンドの地図情報を利用して、その地域をフィルタしてから探しているらしいが、数が少ないらしい。（情報元はタクシーのドライブレコーダー）

一般的なSean Text Recognitionのモデル

CNN-RNN Model (baseline)



- CTC(Connectionist Temporal Classification)で文字列を予測
 - 音声認識でも使われる系列予測の手法

JapanTaxi

[An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition,
Baoguang Shi et al., IEEE Trans. Pattern Anal. Mach. Intell. 2017]

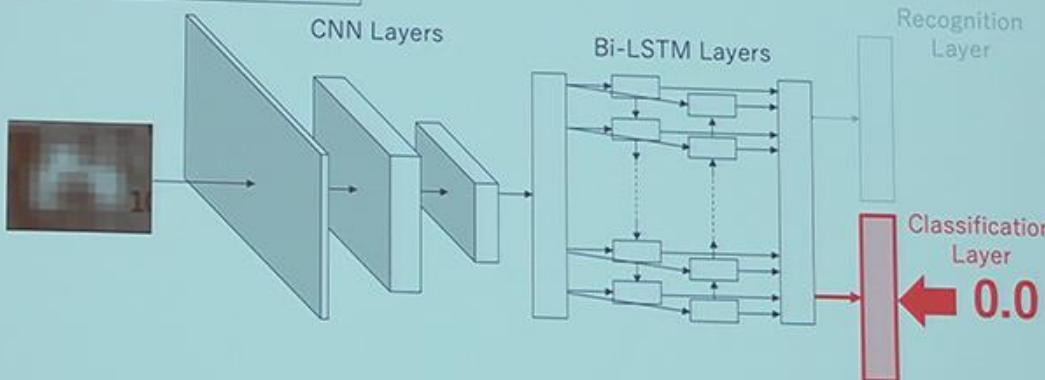
18

この手の問題はCTCを利用するのがよいらしい。

通常のCTCに対して、数値が写っているかうつっていないかのラベルを同時に学習する（マルチタスク学習）することで精度を向上させている。
うつっていないときは赤字のところは0だけど、CTCの出力層はごく微妙な数値を誤差逆伝搬しているらしい。
 $\text{Loss} = (1-\alpha) \text{CTC-loss} + \alpha \text{classification_loss}$ のように定義するのが一般的だと思うが。。

Multi-task Learning of Scene Text Recognition

数字が写っていない画像の学習



- ClassificationのlossからのみBack Propagation
- ClassificationのLoss functionはMean Squared Error

JapanTaxi

22

CCSE conference on computer science
for enterprise



そもそも学習データがすくないなら、ドライブレコーダーを搭載した車1台借りてご近所のガソリンスタンドを徘徊したらいいじゃね？（もっとデータとれるだろ）それぐらいお金あるだろ？ おたく本業ですよね？学生じゃないだから。。。トホホ

背景：ドライブレコーダーの動画からガソリンスタンドの料金を自動認識

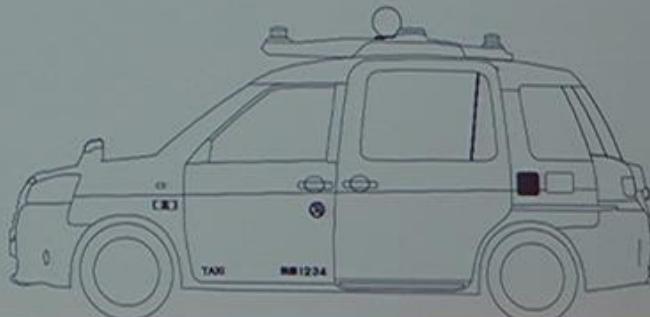
課題：学習データを取集するのが大変

結論：

- 数字じゃない形（データが大量）も学習することで少ないデータで精度が向上
 - 学習データが少ない時にもMulti-task Learningは有効
- 実用上、数字が書かれているかどうかの判定は必要なので、1つのモデルに詰め込めたのは一石二鳥だった

JapanTaxi

WE'RE
HIRING!



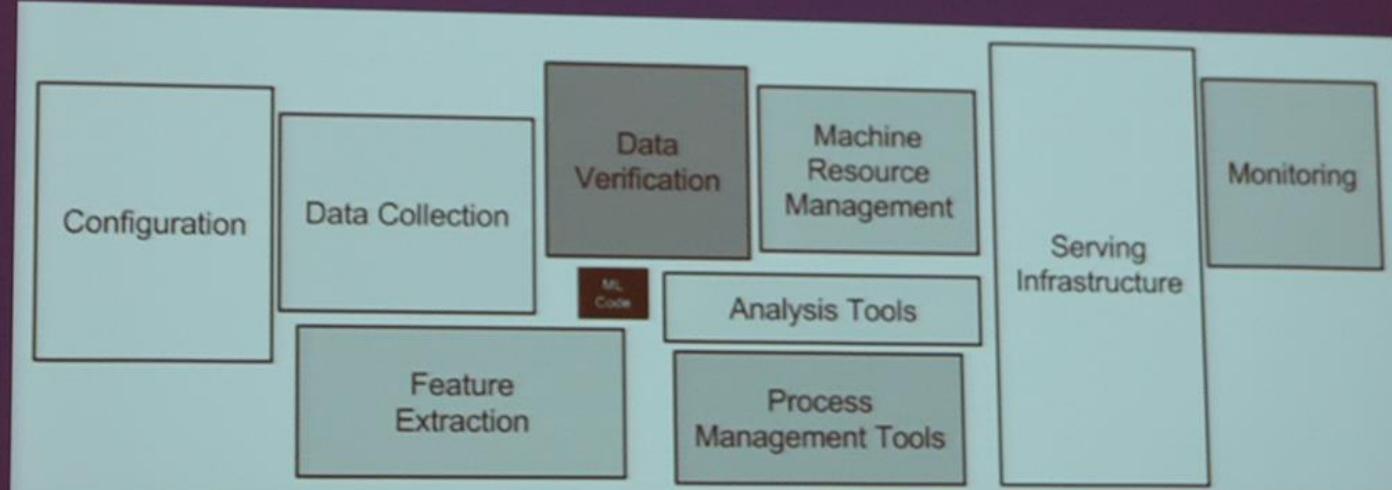
ABEJA の ML Ops のお話



Cameras forbidden

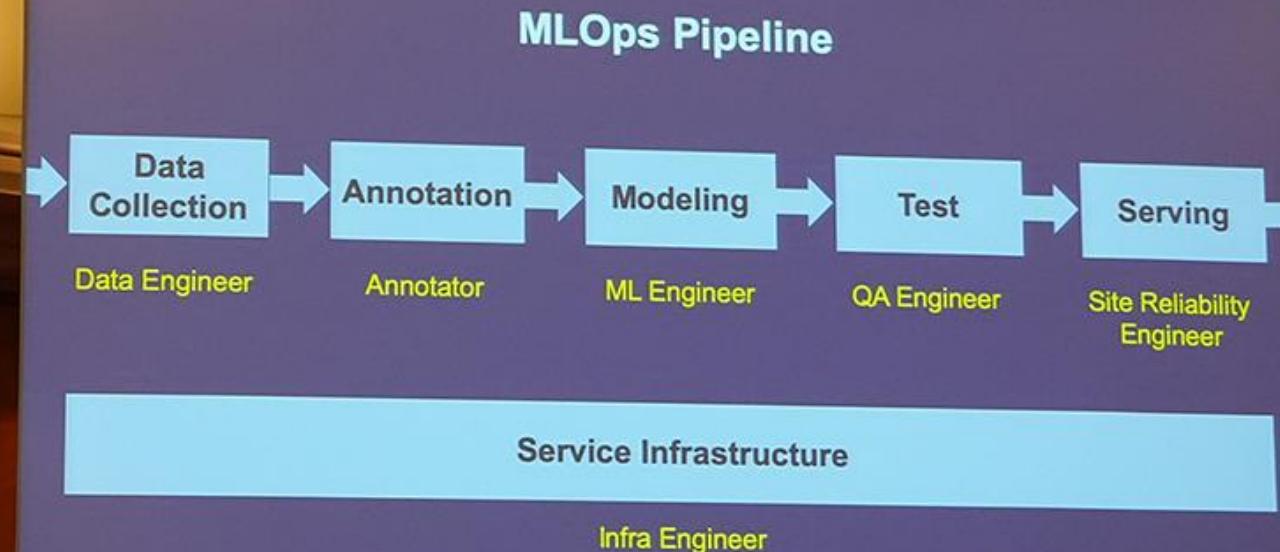
What is MLOps?

MLOps = Machine Learning Workflow + DevOps



Sculley et al. 2015

ML Ops にはいろんな人材が必要



CCSE conference on computer science
for enterprise

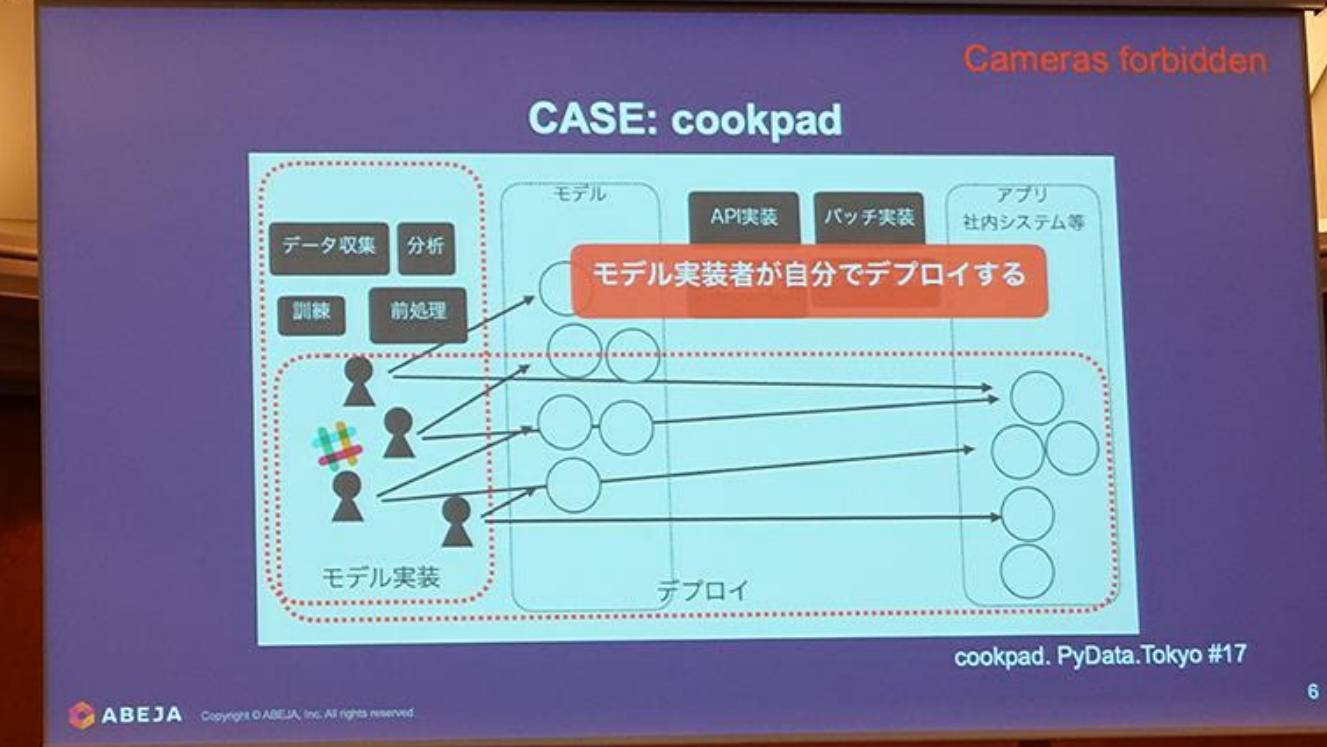


CCS



だが実際はモデル実装者無双！（そななんだなーみんな）

CCSE conference on computer science
for enterprise

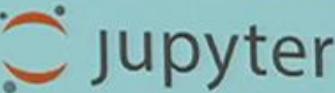


イケてるワークフロー（私はこれでkubernetesの勉強はじめました）

Recent Trend



Airflow



jupyter



argo



SELDON



Kubeflow



Istio



kubernetes



Copyright © ABEJA, Inc. All rights reserved

7

What do we need for MLOps?

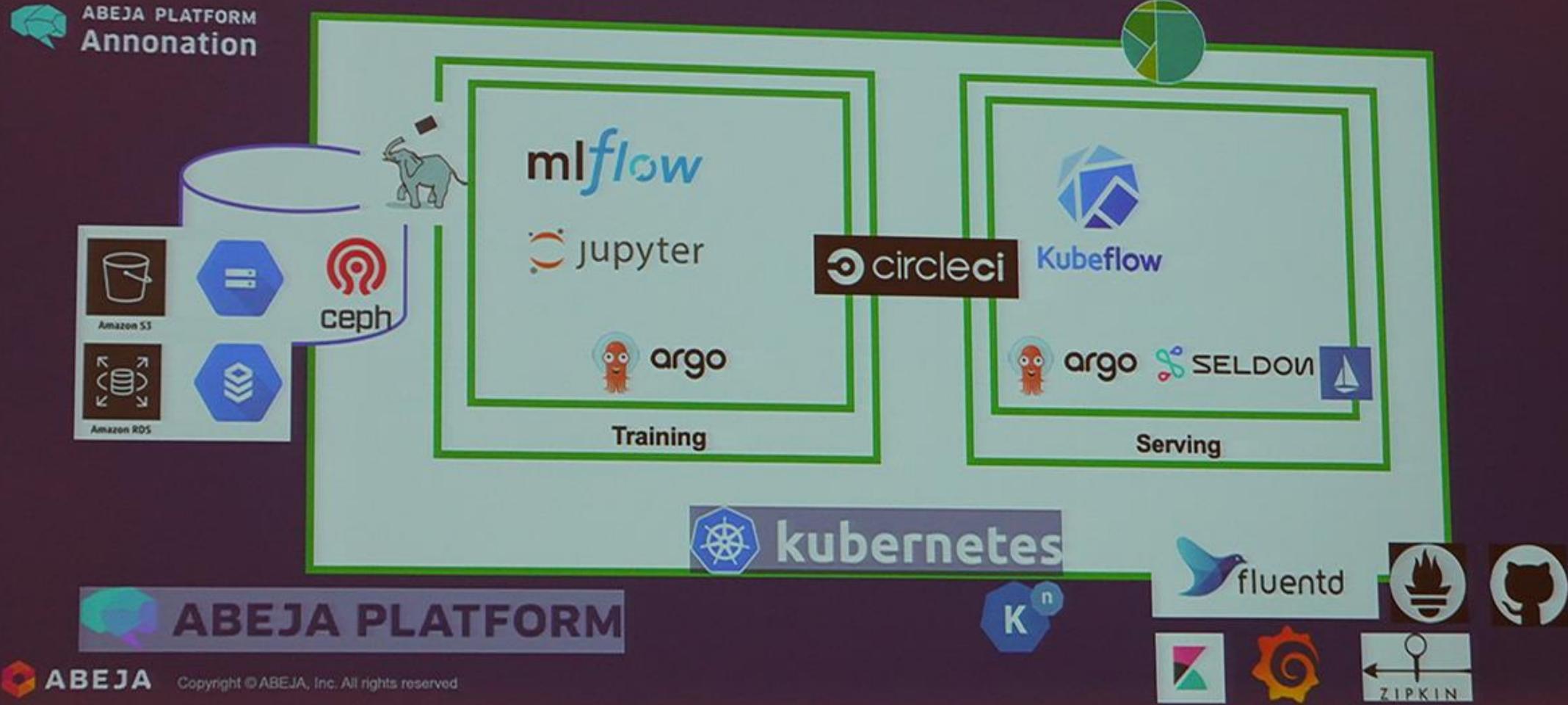
Machine Learning Workflow

- Reproducibility = 機械学習結果を再現したい
- Feedback = 機械学習モデルを継続的に改善したい

DevOps

- Controllability = 出力 / 結果 を制御したい
- Resource Management = 計算資源を有効活用したい
- Stability = 安定的に運用したい
- Automation = 自動化して運用コストを下げたい
- Measurement = ROI / KPI を計測したい
- Observability = 監視したい

What should we do for MLOps? (My Answer)



AIによるアニメ中割生成



CCSE conference on computer science
for enterprise

CyberAgent. Rakuten Institute of Technology mercari

sansan LIFULL

cybozu EXAWIZARDS LINE AB

DeNA LEAPMIND



PRESS席

S席

(発表者の人えげつないエンジニア。Chainer のyoloV2とかの実装者)

自己紹介

- 名前 : 李 天琦
- Twitter : @Leetenki
- Facebook : Leetenki

■ 経歴

2016年4月～ : DeNA入社
2016年6月～ : Japanリージョンゲーム事業本部 (サーバ開発)
2016年10月～ : AIシステム部異動 (コンピュータビジョン研究)



アニメ中割

アニメ現場の問題



限られた制作時間

クオリティの担保

限られた予算規模

アニメータの人材不足

アニメ中割

アニメーターの描く中割



- ・ 中割りの枚数は30分アニメ1話あたり**3500～4000枚**
- ・ 1枚描くのには**数時間**かかる大変な作業

*出典：SSS・STL・WAO 2017「ずんだホライズン」

参考：アニメーターWeb レポート (<http://animatorweb.jp/>)

どうでもいい話だが、30分アニメでこれまでに最高に使った作画枚数は12,000枚 2002年制作「ラーゼフォン」第15話「子どもたちの夜」（演出・脚本 磯光雄）だと言われている。一般的に作画枚数が金額がかかるので東映などはリミット枚数がある。

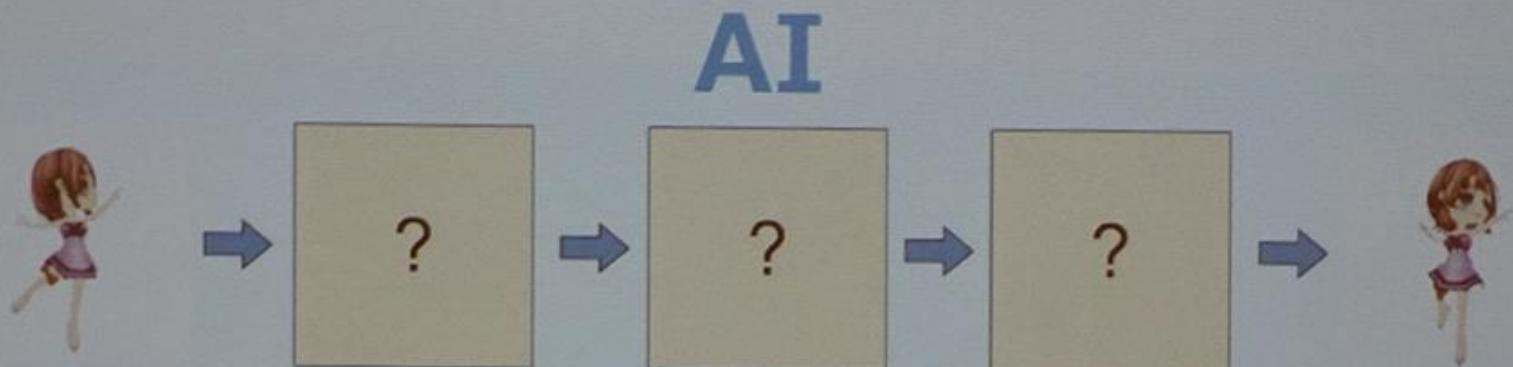


AIでアニメで中割ができるらしいよね！
という話。実際にアニメのツールでは、モーフィングの一種としてこの機能が実装
されているものもあるが、あくまでモーフィング。

本研究で議論されているように構造変化がはげしいものは適用できない

アニメ中割

アニメ中割の自動化

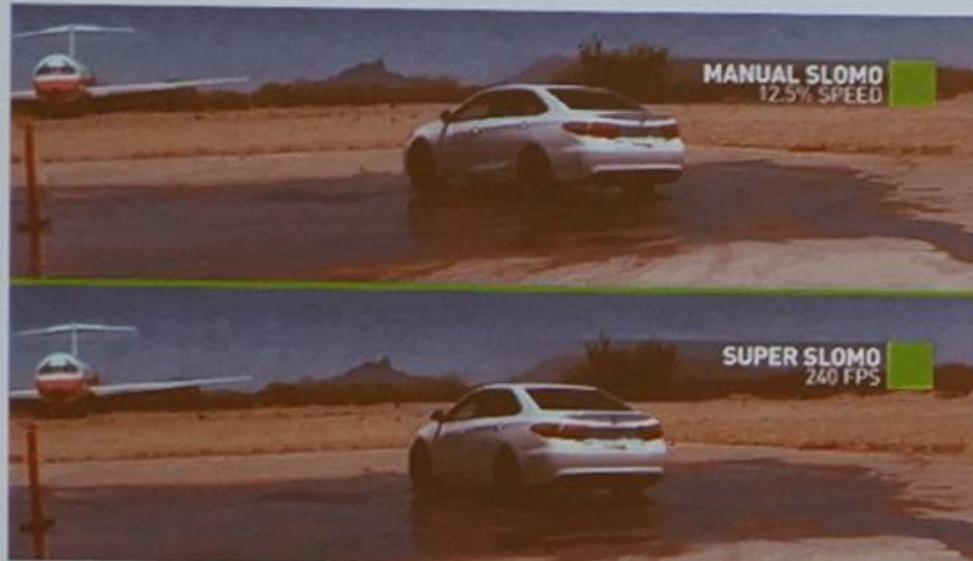


アニメではないがフレームを補完する技術はすでにある
(nvidiaがやっている) 非常に高フレームレート映像を作成
Frame Interpolation / ハイパスローモーション映像のような映像。ヌルヌル。

現状世界最高性能の Frame Interpolation 技術

■ Super SloMo (Jiang+, CVPR'18)

- 30/60FPSの入力動画から中間フレームを予測し、8倍の240/480FPSの動画を生成



Research at NVIDIA: Transforming Standard Video Into Slow Motion with AI
<https://youtu.be/MjViy6kyiqs>

Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation.

Huaizu Jiang, Deqing Sun, Varun Jampani, Ming-Hsuan Yang, Erik Learned-Miller, Jan Kautz. In CVPR 2018.

Frame Interpolation

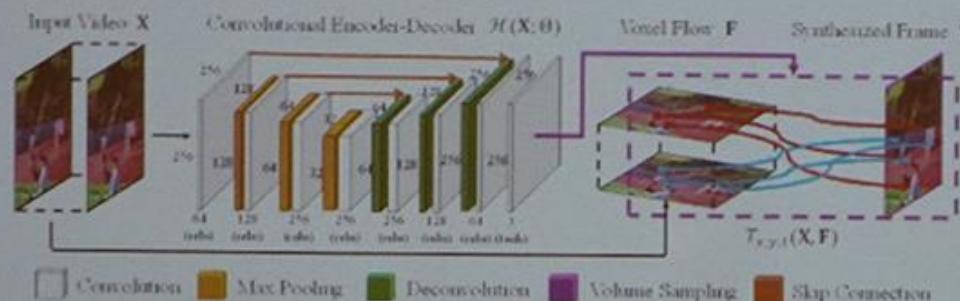
現状世界最高性能の Frame Interpolation技術 Neural Network による Optical Flow算出・Frame補完

■ Super SloMo (Jiang+, CVPR'18)

- Optical Flow を計算、中間フレームを合成して、さらにリファイン
- 30/60FPSの動画の中間フレームを生成し、240/480FPSの動画を生成

■ Deep Voxel Flow (Liu+, ICCV'17)

- Optical Flow を計算し、中間フレームを合成
- Super SloMo と同等性能



Video Frame Synthesis using Deep Voxel Flow. Ziwei Liu, Raymond A. Yeh, Xiaou Tang, Yiming Liu, Aseem Agarwala. In ICCV 2017.
Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation. Huaizu Jiang, Deqing Sun, Varun Jampani, Ming-Hsuan Yang, Erik Learned-Miller, Jan Kautz. In CVPR 2018.

Table 4: Results on the UCF101 dataset.

| | PSNR | SSIM | IE |
|------------------------|--------------|--------------|-------------|
| Phase-Based [18] | 32.35 | 0.924 | 8.84 |
| FlowNet2 [1, 9] | 32.30 | 0.930 | 8.40 |
| Deep Voxel Flow | 32.46 | 0.930 | 8.27 |
| SepConv [20] | 33.02 | 0.935 | 8.03 |
| Super SloMo(Adobe) | 32.84 | 0.935 | 8.04 |
| Super SloMo | 33.14 | 0.938 | 7.80 |

Frame Interpolation

Optical Flow

- 2枚の画像間における**pixelの移動量を表したベクトルマップ**
- 激しく動くpixelほど強く反応する



アニメの中割にこれらの技術が使えない理由は
低フレームワークで物体の構造変化が激しいため
(高フレームワークの映像では物体の構造変化があまりない)

既存手法の課題

アニメの中割では既存の Frame Interpolation 適用が困難

- 既存手法 : **高FPS入力** (30~60 → 240~480)
- アニメ中割 : **低FPS入力** (3~7 → 12~30)

※ 低FPSで**構造変化が大きい**画像間の補完生成が必要



既存のオプティカルフローに+して
構造変化を学習できるようにボーンや目やシルエットのいちなどの構造情報をと
らえるような情報を付与してマルチタスク学習を実施

提案手法

Optical Flowに加え、構造情報も利用したマルチタスク学習モデル

■ 入力画像 :



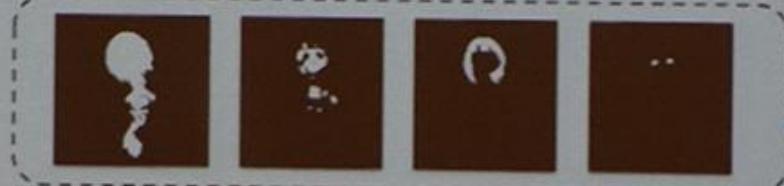
■ Optical Flow :



■ Pose Keypoint :

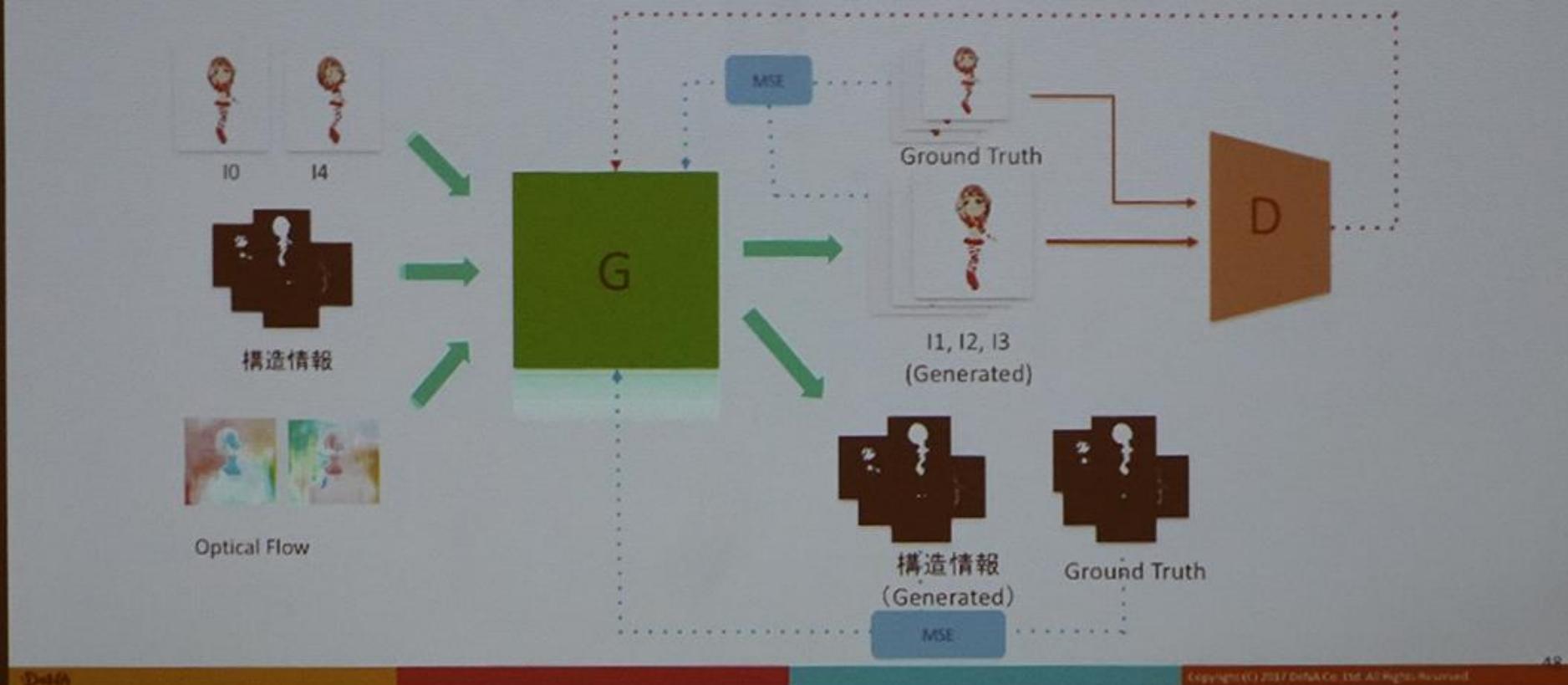


■ 各部位のマスク :



マルチタスク学習

提案手法: 全体図

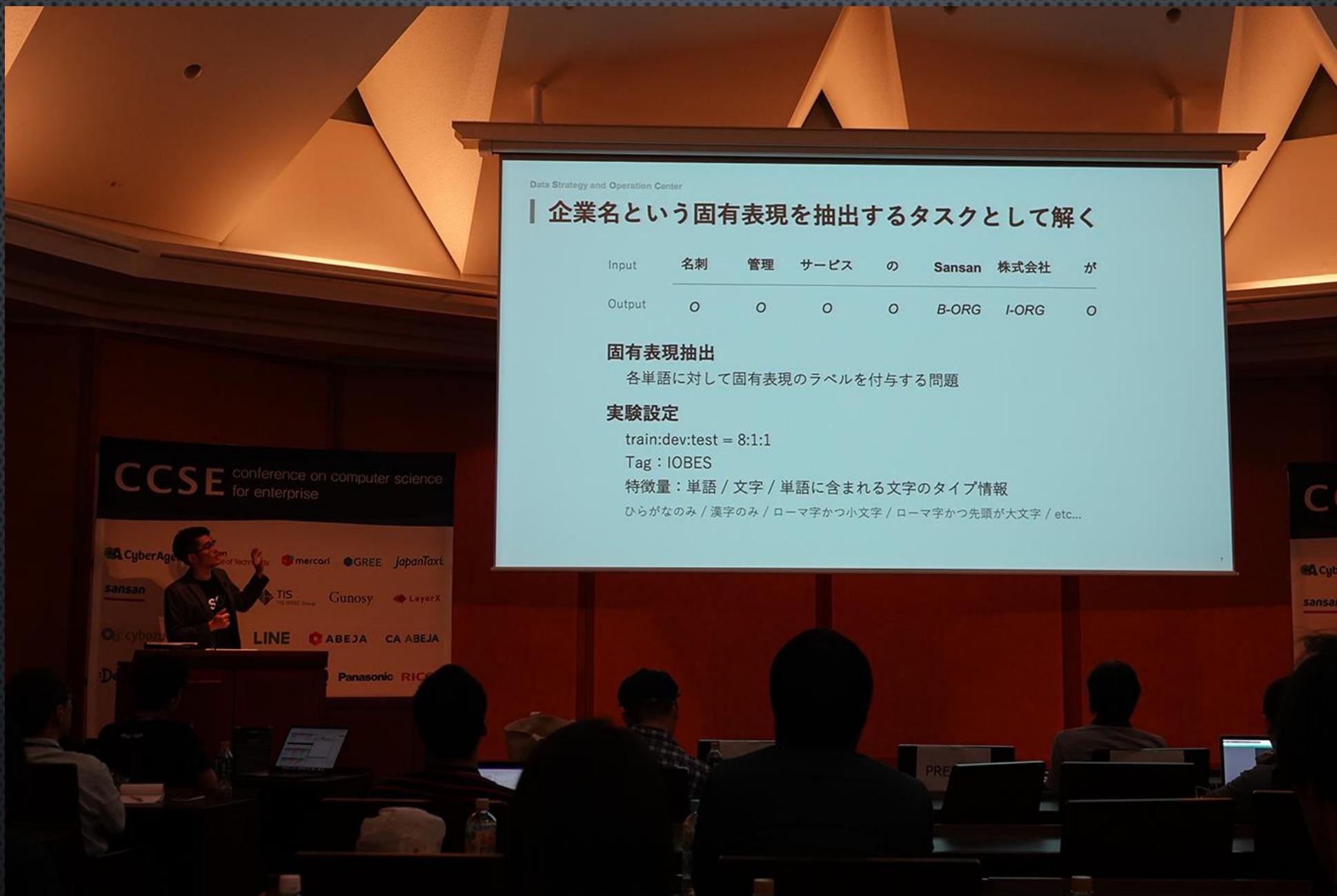


結果は youtubeに動画をuploadしました（限定公開）
<https://youtu.be/z-X8ELizznI>

ただし、作画のダイナミズムはこんなロジックじゃないんだけどね！
参考資料として、電腦コイルの作画を添付しておく。



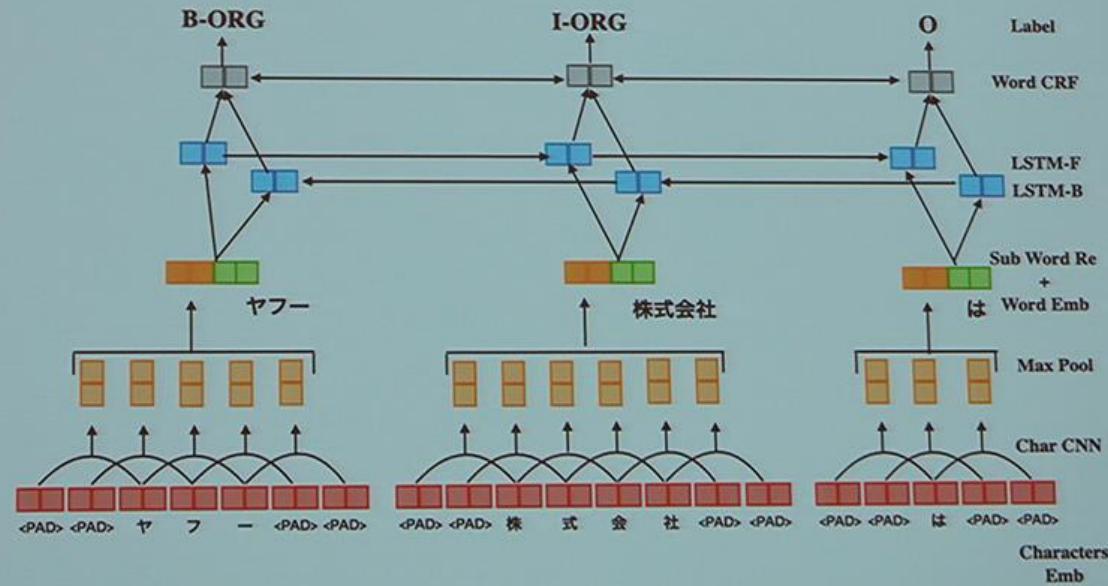
ニュース情報からの企業名（固有表現抽出）



biLSTM-CNNs-CRFでやる
※CNN->biLSTMを通じて特徴情報を抽出して最後にCRFで解く
(結構古い手法)

Data Strategy and Operation Center

| Bidirectional LSTM-CNNs-CRF [Ma & Hovy 2016]



science

JapanTaxi

LayerX

CA ABEJA

Data Strategy and Operation Center

結果

| Model | Dictionary | Character Features | Test Easy | | | Test Hard | | | Average F1 |
|-------------|------------|--------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | | Precision | Recall | F1 | Precision | Recall | F1 | |
| Exact Match | - | - | 0.346 | 0.779 | 0.480 | 0.172 | 0.262 | 0.208 | 0.344 |
| BiLSTM-CRF | IPADic | No | 0.870 | 0.827 | 0.848 | 0.830 | 0.793 | 0.811 | 0.830 |
| BiLSTM-CRF | Unidic | No | 0.865 | 0.811 | 0.837 | 0.845 | 0.769 | 0.804 | 0.821 |
| BiLSTM-CRF | IPADic | Yes | 0.839 | 0.862 | 0.850 | 0.855 | 0.786 | 0.819 | 0.835 |

ベースラインの辞書マッチと比較して高い精度

単語に含まれる文字情報を追加したモデルが最も精度が良くなつた

Computer science

GREE JapanTaxi

osy LayerX

C

Cy

sans

| まとめ

1枚の名刺の先にあるニュースを正確にユーザに届けたい

- ・ニュース記事からの企業名抽出というタスク

深層学習を用いた固有表現抽出

- ・Bidirectional-LSTM-CNNs-CRFのモデルを作成
- ・単語や文字情報、単語の特徴を考慮した特徴量を用いることで企業名を表現

今後の発展

- ・固有表現抽出自体の精度向上
多様な企業名、多様なニュース記事、商品名やサービス名との区別
- ・企業名の語義曖昧性を解消するためのEntity Linkingとの連携

CCSE conference on computer science
for enterprise



CCSE

メルカリの写真検索を支えるバックエンド

株式会社メルカリ 中河 宏文

mercari

CCSE conference on computer science
for enterprise

CyberAgent GREE JapanTaxi

sansan

TIS Gunosy LayerX

NEC

ABEJA CA ABEJA

Panasonic

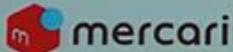
ESS席

| What is 写真検索

- ・ 写真検索とは、いわゆる画像検索機能
- ・ アプリから写真を元に商品を検索する
- ・ 商品名を知らなくても画像から商品を検索できる



動画リンク: <https://youtu.be/kTni8EvOCgI>



4

JapanTaxi

LayerX

CCS

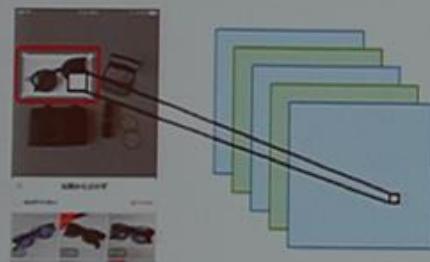
CyberAgent

sansan

写真検索は DNNで特徴ベクトルを作つて最近傍検索(ANN)を実施している
(まあ普通に考えたらそうすうわな)

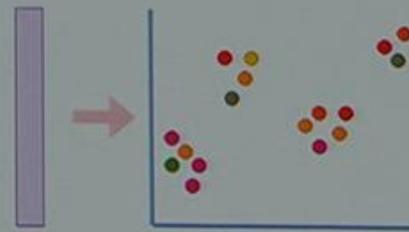
| 基本的な写真検索の仕組み

1



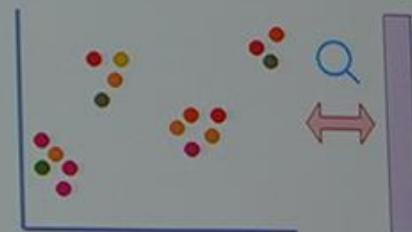
Deep Neural Networks
(DNN)を使用して商品画像
から特徴ベクトルを取得

2



取得した特徴ベクトルを
Approximate Nearest
Neighbor Index(ANN Index)
に追加して画像indexを構築

3



検索時には同じく商品画像から
DNNを介して特徴ベクトルを取
得し、ANN Indexから検索

| What is Kubernetes

- Kubernetes（以下k8s）とは
オープンソースのコンテナ・
オーケストレーションシステ
ム
- k8sにはCustom Resource
Definitionと呼ばれる独自の
リソースを定義できる機能が
あり、開発者はその機能を介
してk8sの機能を拡張できる



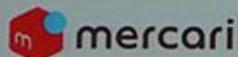
kubernetes

ネットワークの学習などはこちらのPFで実施されているようです。

| ML Platform Lykeion

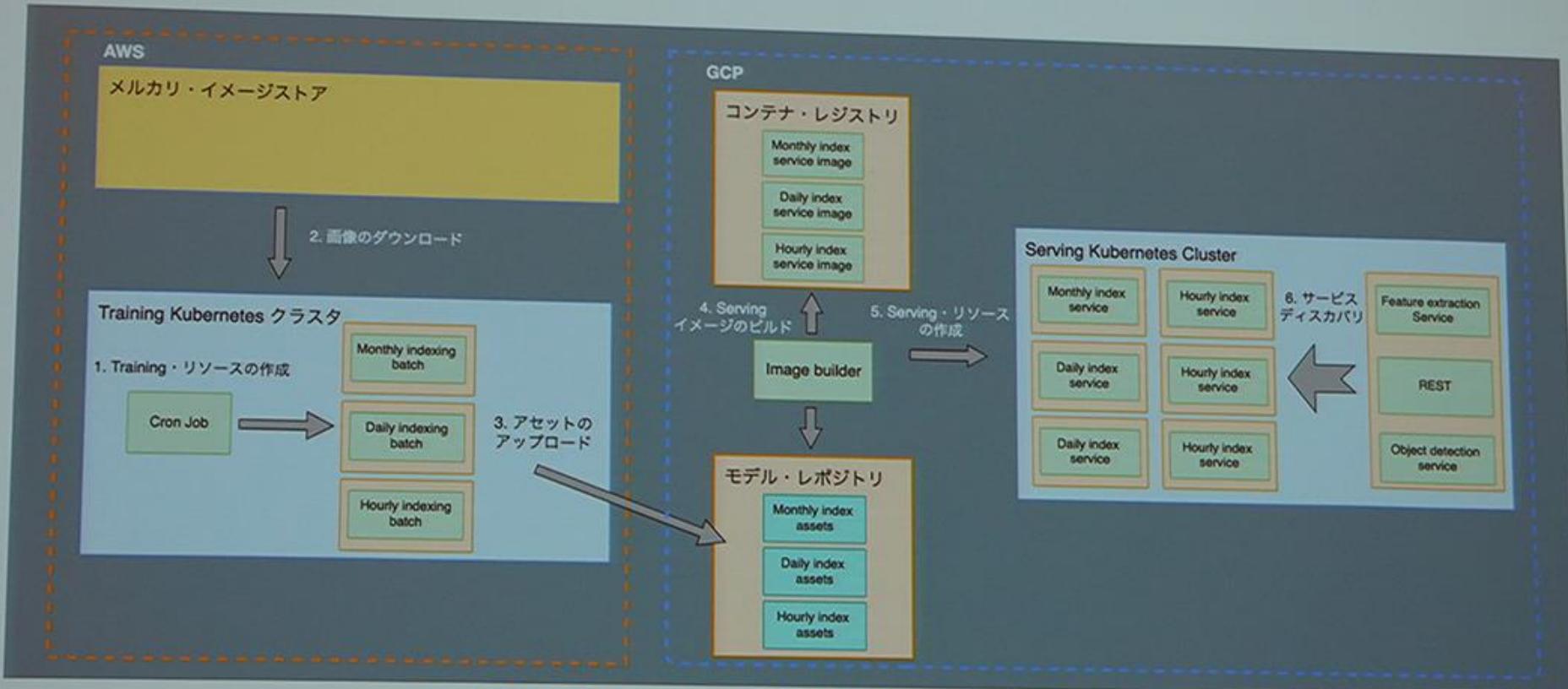
写真検索はLykeionと呼ばれる内製のML Platform上に構築されており、下記の機能はPlatform側の機能を使用している

- Training/Serving CRD & カスタムコントローラ
- コンテナベース・パイプライン
- Training/Serving コンテナイメージ・ビルダー
- モデル・レポジトリ



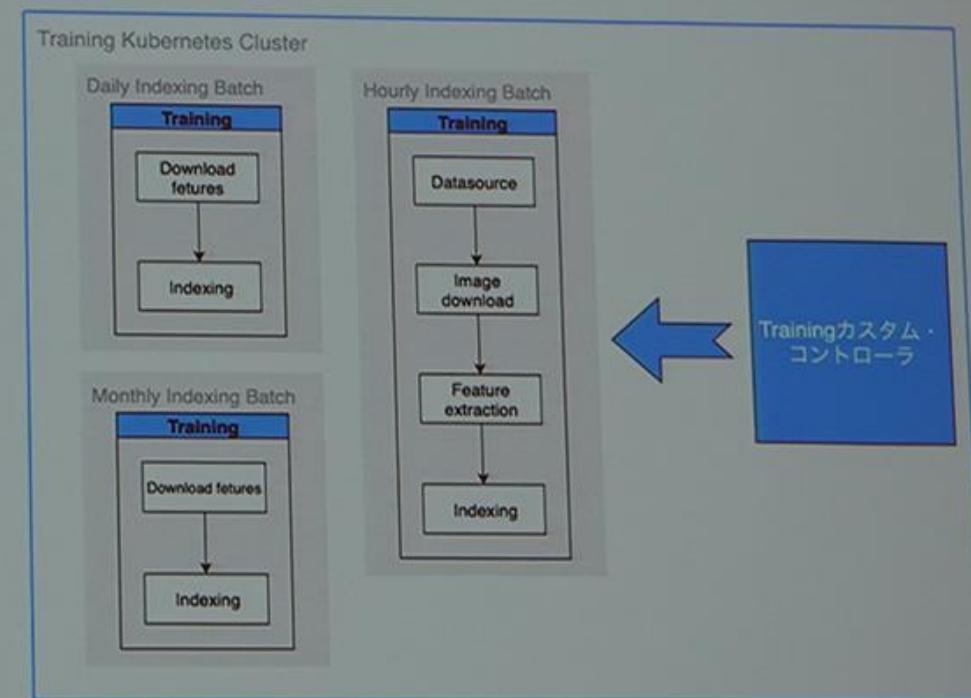
7

Architecture概要図



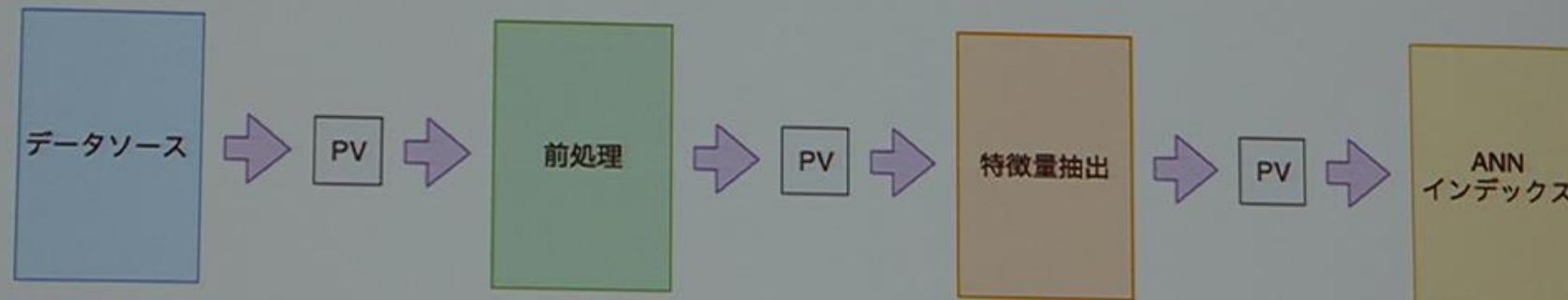
| Training・リソースの作成

- **Training custom resource**をCronJobが作成
- カスタム・コントローラがCRD・リソースで設定されたコンテナベース・パイプラインを実行
- 実行するバッチ単位としてはHourly, Daily, Monthlyが存在



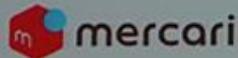
| コンテナベース・パイプライン

- ・各工程を個別のコンテナ・イメージで実行
- ・ライブラリの依存関係など環境ナイーブなMLパイプラインの問題を解決
- ・パイプラインDAGはYAMLで記述 各工程の入出力はPersistent Volume (以下PV) を介す



| 画像のダウンロード

- ・ S3上に存在するメルカリ・イメージストアから商品画像をダウンロード
- ・ パイプライン上もっとも時間がかかる工程（画像数が膨大なため）
 - ・ そのためPVに一定期間キャッシュする事によって再インデックスが必要な時には素早くパイプラインを回せるようにしている

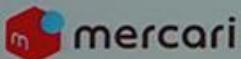
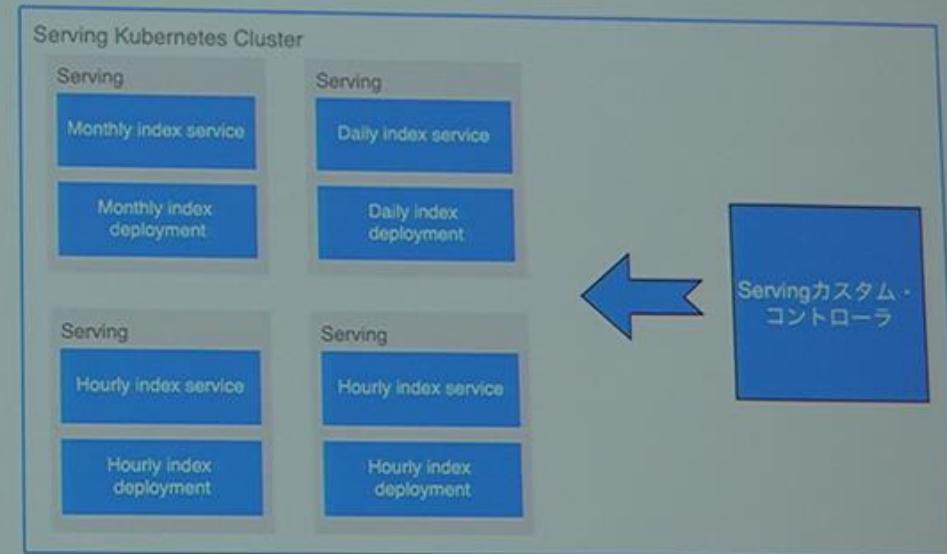


| アセットのアップロード

- ・パイプラインの成果物である特徴ベクトルとANN Indexをモデル・レポジトリに保存
- ・全ての成果物はバージョン管理された状態で保存される
- ・モデル・レポジトリはGCS上に構築

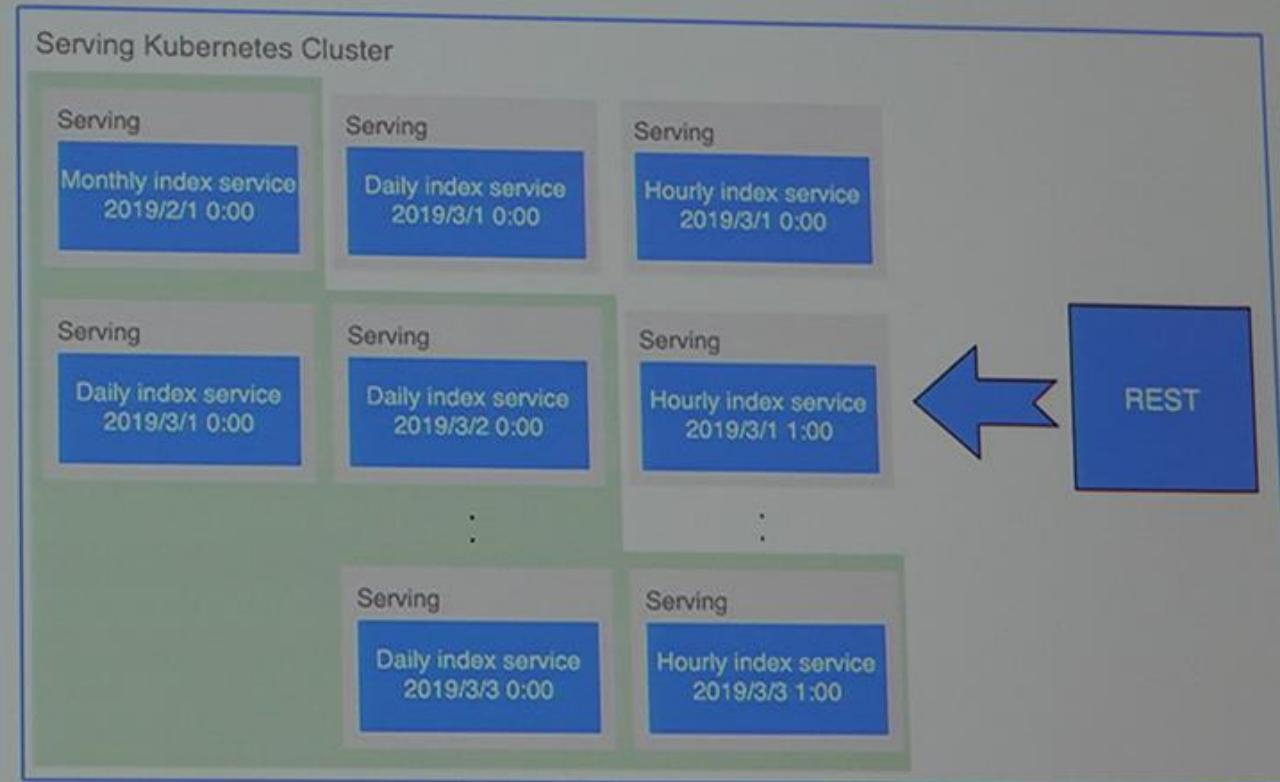
I Serving・リソースの作成

- Image Builderはコンテナ・イメージをビルドしたあと、Servingカスタム・リソースを作成
- Servingカスタム・コントローラはCRD・リソースの設定を元に必要なDeployment、Service等を作成
- 本システムでは構築されたANN Indexを個別のIndexサービスとしてデプロイ



サービス・ディスカバリ

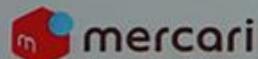
- クラスタ上に存在するIndexサービスをk8sを介して自動的に取得する
- なるべく大きな粒度のIndexを使用するよう、異なる期間・粒度のIndexサービス(Hourly, Daily, Monthly)を自動的に組み合わせる
- REST <-> Indexサービス間のプロトコルはgRPCを使用



とりあえず、まとめるとこうだ。（あまりよくわかってない）

| 写真検索のバックエンド・インフラ

1. コンテナ・ベースの再現性の高いシステム
2. k8sのCRD/カスタム・コントローラやサービス・ディスカバリ等の機能を活用
3. **Batch Execution as Custom Resource**等、ML Platformで実現されている機能を使用し、ロバストなシステムを構築
4. クラウド・インフラをk8sで抽象化する事によって、各クラウド・ベンダの良いとこ取りをしている



プレスリリース

Gunosyの研究論文がデータマイニングの国際会議「KDD2019」にて採択

Gunosy

Gunosy の研究論文がデータマイニングの 国際会議「KDD2019」にて採択

～ 広告クリエイティブ作成支援のためのマルチタスク学習と
Conditional Attention による CV 予測 ～

株式会社Gunosy（本社：東京都港区、代表取締役CEO：竹谷祐哉、以下、Gunosy）は、Gunosy内の「Gunosy Tech Lab（読み：グノシー テック ラボ、以下、同ラボ）」にて、「広告クリエイティブ作成支援のためのマルチタスク学習と Conditional AttentionによるCV予測」（原題：Conversion Prediction Using Multi-task Conditional Attention Networks to Support the Creation of Effective Ad Creatives）の研究結果を発表した論文が、データマイニングに関する国際会議ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2019（KDD 2019）の Applied Data Science Trackに採択されたことをお知らせいたします。

KDDは、データマイニングに関する国際会議の1つであり、データマイニング領域では世界一の注目度、難易度を誇る会議とされています。第25回目となる2019年の会議はアラスカ州アンカレッジにて8月4日から8日まで開催される予定です。

<https://gunosy.co.jp/news/172>

(C) Gunosy Inc. All Rights Reserved.

PAGE | 2

問題の設定がよくわからていなかつたのだが、実際にimpを出さないでも
より効果的なクリエイティブの創出ができるらしい？

Background

Gunosy

Ad creatives in display advertisements

Play an important role in delivering product information to customers.

Ad Creative:

An object for ad including
short texts and image.

Conversion (CV):

User actions, such as
purchasing items
or downloaded apps.

Ads with many CVs are more effective.

→ Creating high-performing ad creative
is expected in advertisers.



Background | The purpose of this study

Gunosy

To support the creation of ad creatives with many CVs, and we propose a **new framework** for these support, including accurate prediction of ad creative text conversions **in advance**.

For the creation of high-performing ad creatives, rather than predicting zero conversions, we would like to **accurately predict high-conversion creatives as such**

Ad creatives are mainly text and image.

We **focus on the text**, and predicting its CVs.

→ Because it is difficult to replace ad images, but easy to replace text.

Creative
#1 (**Bad**)

Easy to **click** but
Hard to **conversion**

Creative
#2 (**Good**)

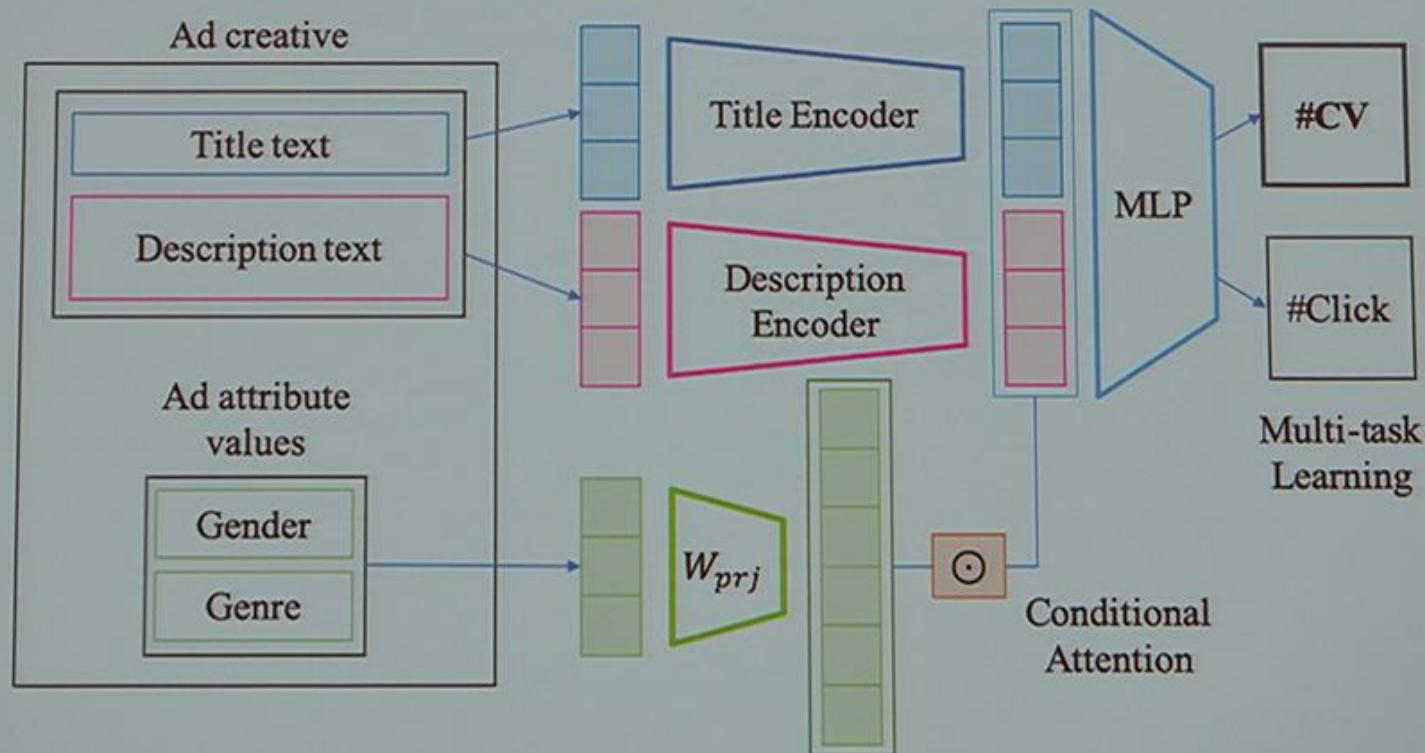
Easy to **click** and
Easy to conversion



Methodology | Overview

Gunosy

The proposed framework consists of **three key ideas**:
Multi-task learning, Conditional attention, Attention highlighting.



属性によってどこが注視されているのかわかる。

Methodology | Attention highlighting

Gunosy

Attention highlighting visualizes important words based on conditional attention.

If the words contributing to conversions are clarified, advertisers will be able to easily create high-performing ad creatives.

For All

1000万人が選ぶ！みんなが遊んでいるゲーム10選
スマホに入れておきたい無料ゲームを限定

For Women

1000万人が選ぶ！みんなが遊んでいるゲーム
スマホに入れておきたい無料ゲームを限定

For All

-10kgのダイエットに成功！痩せる理由はこれ
女子に人気の方法で効果を実感

For Men

1000万人が選ぶ！みんなが遊んでいるゲーム
スマホ

For Women

-10kgのダイエットに成功！痩せる理由はこれ
人気の方法で効果を実感

For All

有名芸能人監修。簡単にできる料理レシピ本
一人暮らしの男性にもおすすめ！

For Women

人気の方法で効果を実感
ダイエットに成功！痩せる理由はこれ

For Women

有名芸能人監修。簡単にできる料理レシピ本
一人暮らしの男性にもおすすめ！

For Men

有名芸能人監修。簡単にできる料理レシピ本
一人暮らしの男性にもおすすめ！

(C) Gunosy Inc. All Rights Reserved.

PAGE | 14

Experiment | Dataset

Gunosy

Use real-world creative data from Gunosy Ads ('17/8 - '18/8)

- Removed creatives with a low number of impressions.
- Convert words to vectors using pre-trained word2vec.
- The number of clicks and conversions is log-normalized.

After these preprocessing, **14,000** creatives were included.

Features included in the ad creative dataset.

| | Features | Feature Description | Details |
|------------|--------------------|---------------------------|-----------------------------|
| Texts | title | Title texts | Avg. 15.44 ± 3.16 chars |
| | description | Description texts | Avg. 32.69 ± 5.43 chars |
| Attributes | genre | Genre of the creatives | 20 types |
| | gender | Gender of delivery target | 3 types (M / W / Others) |

他の結果と比較しているが、SVMでも結構精度でるんやな。
いつも思うやけど、SVMでこの精度でてらええやんとおもったりする。

Experiment

Gunosy

For accurately predicting high-conversion creatives as such,
we adopted **NDCG**, which is evaluation metrics for ranking.

| Model | NDCG [%] | | | |
|-------|------------------------------|--------------|--------------|--------------|
| | All | | #CV top 1 % | |
| | single | multi-task | single | multi-task |
| SVM | | 96.72 | | 83.73 |
| MLP | 96.68 | 97.18 | 82.97 | 84.12 |
| GRU | Vanilla | 96.54 | 97.00 | 76.39 |
| | Attention | 96.76 | 97.11 | 83.00 |
| | Conditional Attention | 96.77 | 97.20 | 87.11 |
| | | | | 87.14 |

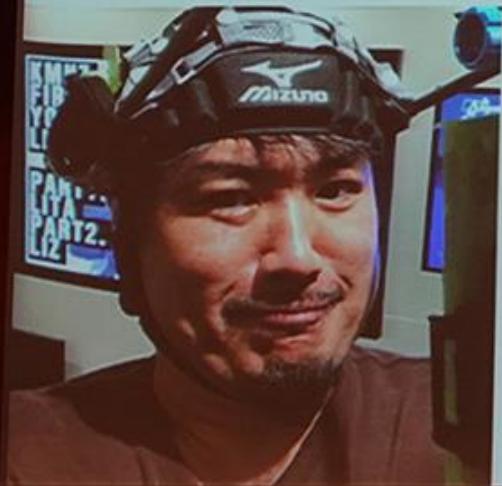
Compared with baselines, the model with multi-task learning and conditional attention seems to implicitly learn features that contribute to conversion prediction.

CECSE 2019

conference, on computer science
for enterprise

VTuber向け特殊表情コントローラの 可能性と機械学習的アプローチ

グリー株式会社 GREE VR Studio Lab
Director 白井暁彦



Akihiko SHIRAI, Ph.D
GREE VR Studio Lab
Twitter@o_ob @VRStudioLab



GREE VR Studio

記者会見やライブデモにも活用



VTuber社員
「いそら真実」 [@_Isoramanami/](https://www.youtube.com/@_Isoramanami/)



→ 決算説明会（朝日新聞社YouTubeより）



外部イベント用仕様



iPhone新機種のキラーアプリ的扱い

YouTube「[KDDI iPhone XS/XS Max発売イベント](#)」で検索

Vtuberのこういう表情って他の人がコントローラーで制御していることが多いようです。（知らなかった）いとうさんは目が★になるのがすきです。

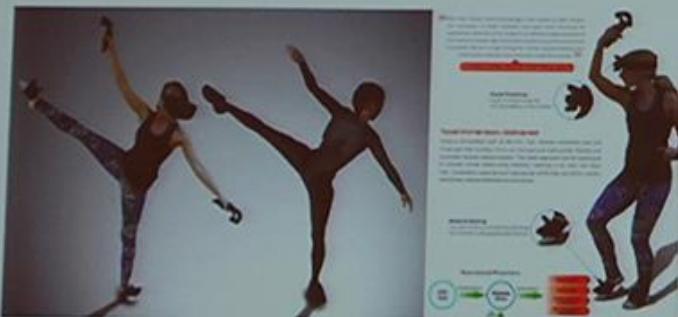
REALITY 公式番組における特殊表情の例



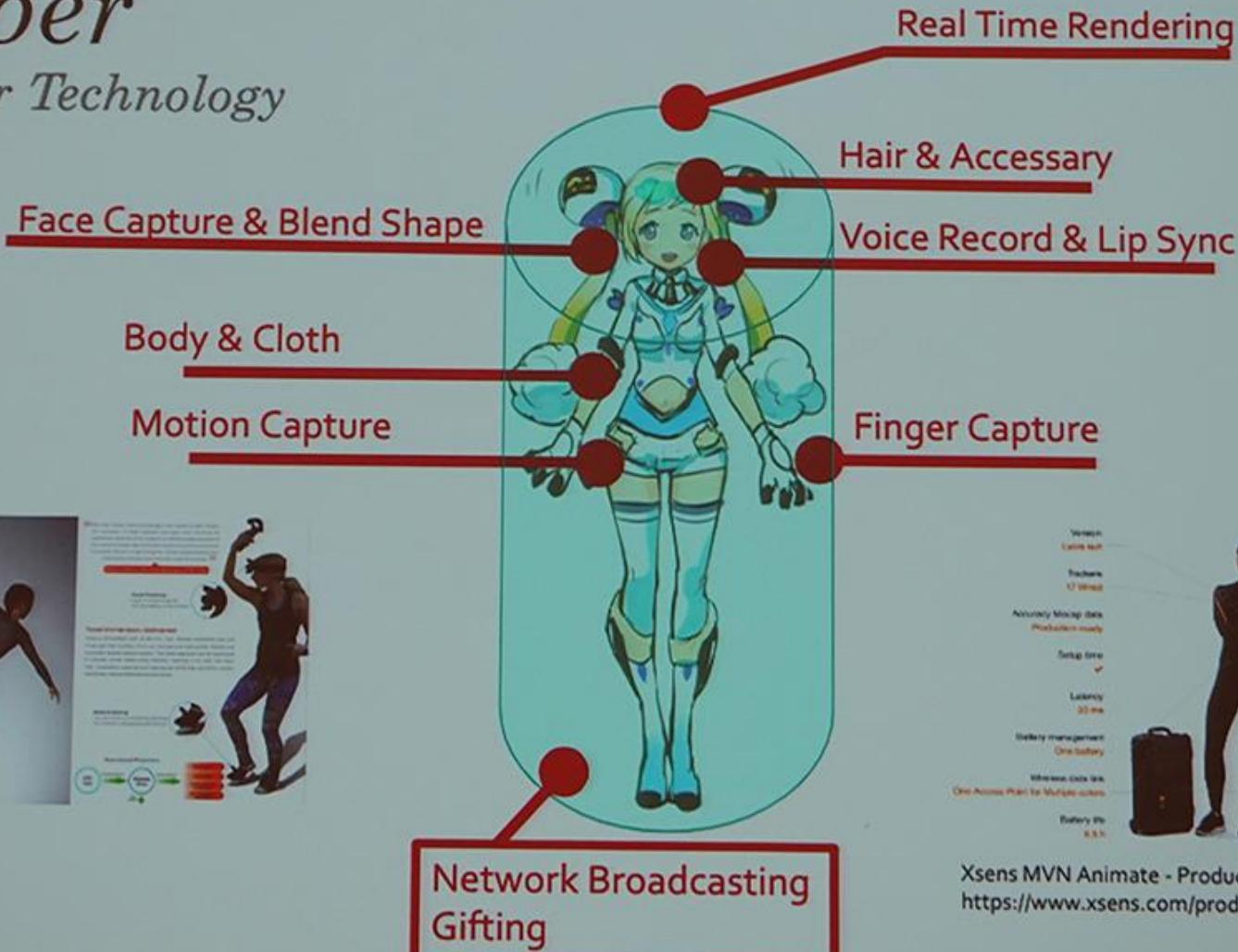
プロ用システムだとこんなおおがかりなシステムのようです。
すごい。

VTuber

3D Avatar Technology



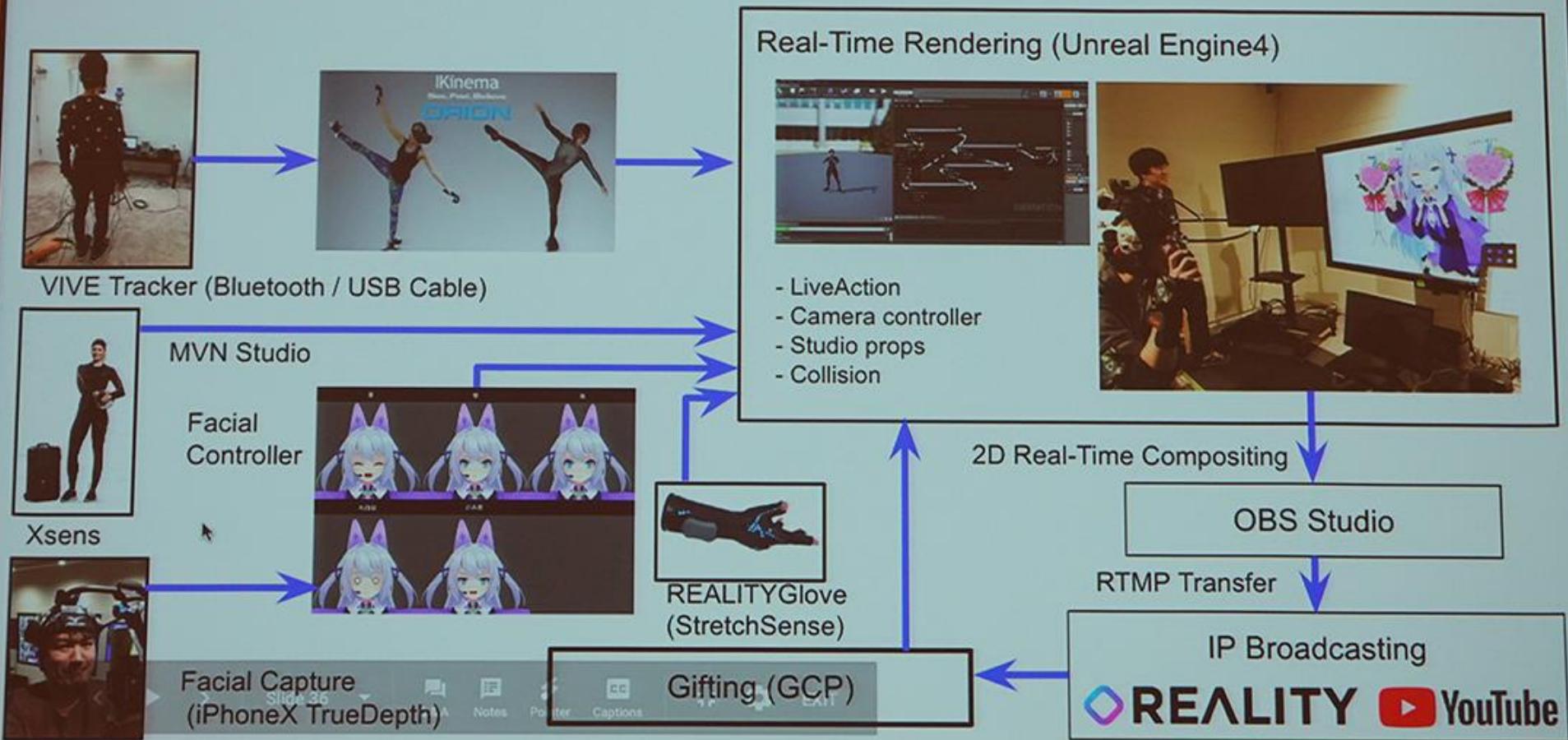
<https://ikinema.com/orion>



Xsens MVN Animate - Products - Xsens 3D motion tracking
<https://www.xsens.com/products/xsens-mvn-animate/>

プロ用システムだとこんなおおがかりなシステムのようです。
すごい。

Pipelines for VTuber Live Entertainment in REALITY Studio



CO

Cyber

sansan

通常顔

(自然な表情・軽く笑み)
(自然なスピードで発声練習)
これから、表情テストをはじめます。次の調子は大丈夫ですか？それぞの表情を、10秒かけて演じてください。



おねだり

(頬を引いて上目遣い)
実は～...
今日はセンパイに、おねがいが
あります～～...
PS4..ほしいなあ～～
(上目遣い・カメラ目線)



星目(目を大きく開く)

買ってくれるんですか？
よっしゃー！スパイダーマンが遊べるぞ～！
センパイってばちょういな～！



><(バツ目)顔

(参考書で叩かれて)
痛一っ！
ううう.....いたたた...
カドが、カドが入りました
よ.....
(10秒間 痛みをこらえる)



照れ顔

ええ～？なになに？
私が？かわいいから？
でへへ...も～、
しょうがないですね～



怒り顔

ちょっと！
何するんですか！！
私にこんなことして、まさか
ただで済むとは思ってないです
よね！？



恐怖/青ざめ

うわああっ！
じょっ、冗談ですう...
ごめんなさい、ごめんな
さい、調子に乗りました
あ～.....！



嬉しい顔

やった——！PS4！！
センパイありがと——！
わーー！ばんざーい！！
わーわ———！



威圧顔(静かに怒)

あの、センパイ、私をバカにし
てるんですか？
こう見えてけっこう真剣に怒っ
てるんですけど！



無気力顔

.....はあ.....つかれた.....



企み・ドヤ顔

どううですか？もっと感謝
してもいいんですよ？
それと、約束は守っても
らいますからね～



白丸目(呆けた顔)

へ？買ってもいいけど条件が
ある？
ふんふん、はあ、へえー.....



キス顔(唇を突き出して)

え、目を閉じて？
くちびるを突き出して？
こうですか？
ん-----。
(合計10秒がんばって！)



泣き顔

オーイオイオイ.....
センパイがいじめるよおー...
おがあざあーーーん.....



通常顔(口閉じ)

(自然な表情・軽く笑み)
(口を閉じて
10秒お待ちください)

これで1テイク終了です
お疲れ様でした！

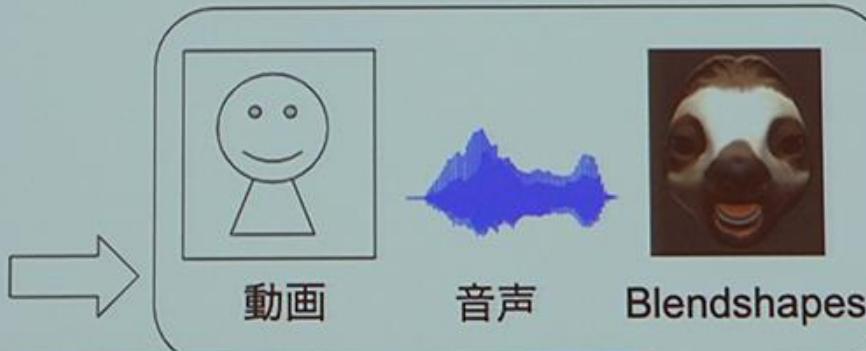


地獄のような
トレーニングを
多様な出演者で実施
(画像は例です)

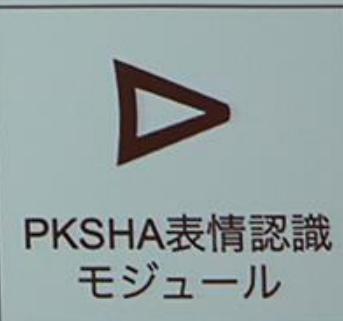
こういう特殊表情変化をiPhoneXなどの表情トラッキング情報から
機械学習で変化させたいらしい。

PKSHA Technology社とのコラボレーション研究

- 一人の演者に、15表情の演技をしてもらう（8ティク/表情）
 - iPhone Xで撮影。動画・音声・Blendshapesを録画する
- 録画データをPKSHA表情認識モジュールで学習・評価
 - 実験用環境で収録した（生配信のデータ収録・アノテーションは困難）



学習・
評価



結果の精度は以下のような感じ。まだまだのよう。がんがれ。

評価結果

- メトリックは F-measure (precision と recall の調和平均)
- frame by frame で 62.83 % の精度を達成
- 分類クラスを削減すると精度が向上する
- 撮影条件が変わると精度が大きく下がる
- multi-frame による予測は精度が出なかった

| Method | Dataset | 15 classes | 8 classes | 5 classes |
|----------------|--------------|----------------|----------------|----------------|
| Frame by Frame | Take 6 (遠距離) | 35.28 % | 56.95 % | 66.52 % |
| Frame by Frame | Take 7 (通常) | 62.83 % | 86.50 % | 90.50 % |
| Frame by Frame | Take 8 (斜め) | 14.34 % | 24.40 % | 66.52 % |
| Multi-frame | Take 6 (遠距離) | 30.47 % | 41.31 % | 60.53 % |
| Multi-frame | Take 7 (通常) | 43.81 % | 58.57 % | 73.81 % |
| Multi-frame | Take 8 (斜め) | 27.21 % | 49.82 % | 72.08 % |

全体的な所感

- ・研究開発のほとんどが DNN を利用した研究および応用でだったりするでのなんか、個人的には重箱のすみをつつくものばかりで、うーんというか、これは**テクいわー** というがあまりなかったので残念。他の学会もこんな感じだったりするとおもなーという感じ。
LeapMind だけ、CNNをFPGAで実装するとかいう頭のおかしい発表をしていが、あれぐらいぶっ飛んでいたりする人がいたら嬉しい。（しかし人気なし、普通のWeb系の人はFPGAって何？ フア？？ という感じだろ）
- ・運用上はどこも、一人無双で一騎当千の活躍を実施してるところが多くまあ～にたよなもんんだなーという感じがした。またそういった意味で ML Opsを専属とするエンジニアの求人は増えそう。
- ・研究開発にかけてる姿勢も会社毎で違っていてインターン人が発表していたりしたのが印象的だった
- ・みんながつかっているシステム構成がわかってよかったです。Kubernetes はまじでみんな使えたほうがよい。Kubernetes は既存のマネージドサービスを利用するのがよいみたい。マネージド・サービスの場合は AWSよりも GCP (google) のほうがよいらしこ。KubernetesはもともとgoogleのプロダクトなのでGCPのほうがこなれている。

おわり