電子情報通信学会
パターン認識・メディア理解研究会

第二期PRMUグランドチャレンジ中間報告

9th Aug. 2017@MIRU2017広島

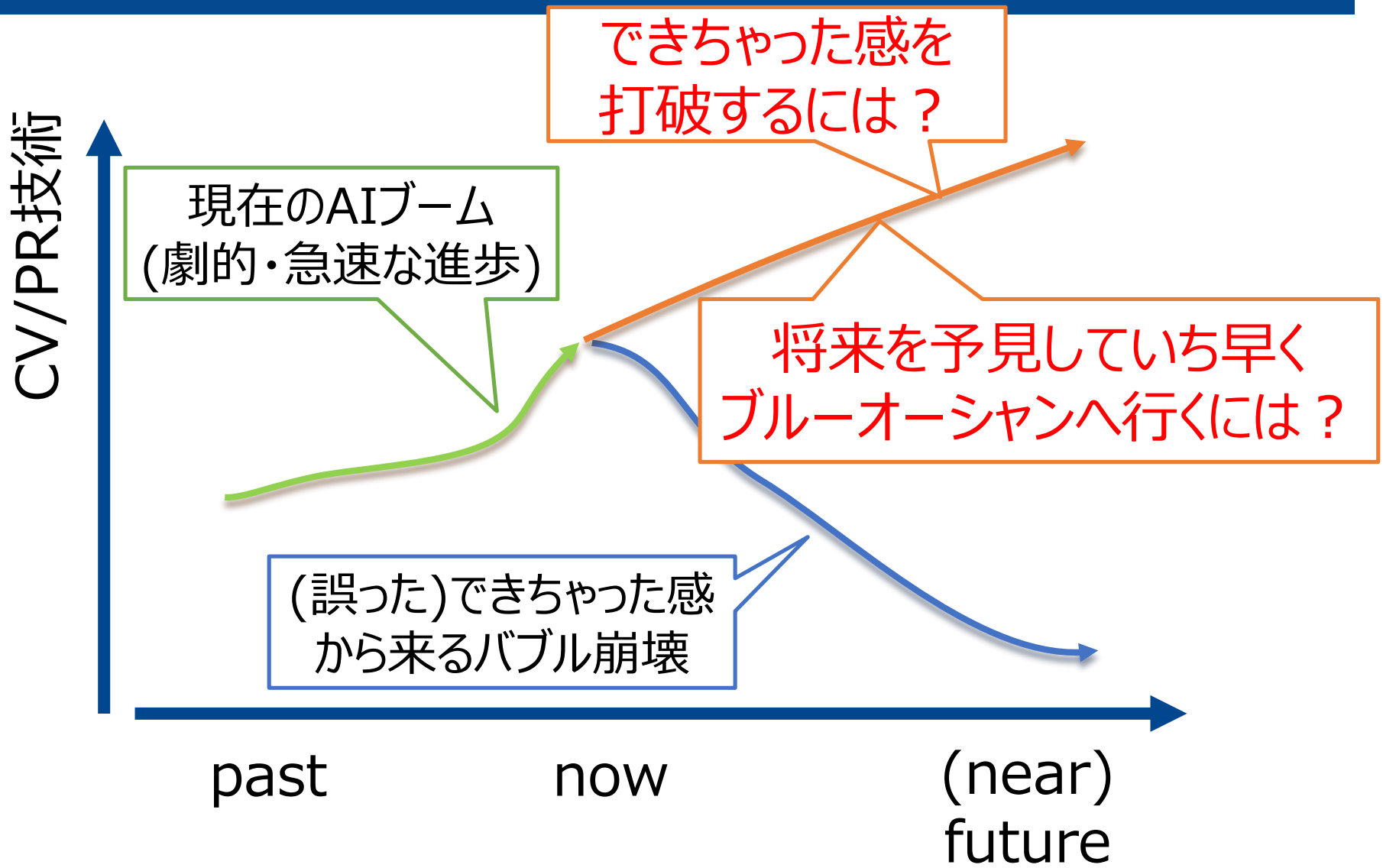
第二期PRMU-GCメンバー
安倍 満@デンソーITラボ，木村 昭悟@NTT-CS研，中澤 篤志@京大，
舩冨 卓哉@NAIST，松下 康之@阪大, 山崎 俊彦@東大 (50音順)
＋ 前田英作・内田誠一(PRMU前委員長・前副委員長)

なぜ今，グランドチャレンジか？
企画趣旨説明

皆さん，忙しいですよね！ ひっぱりだこですよね！

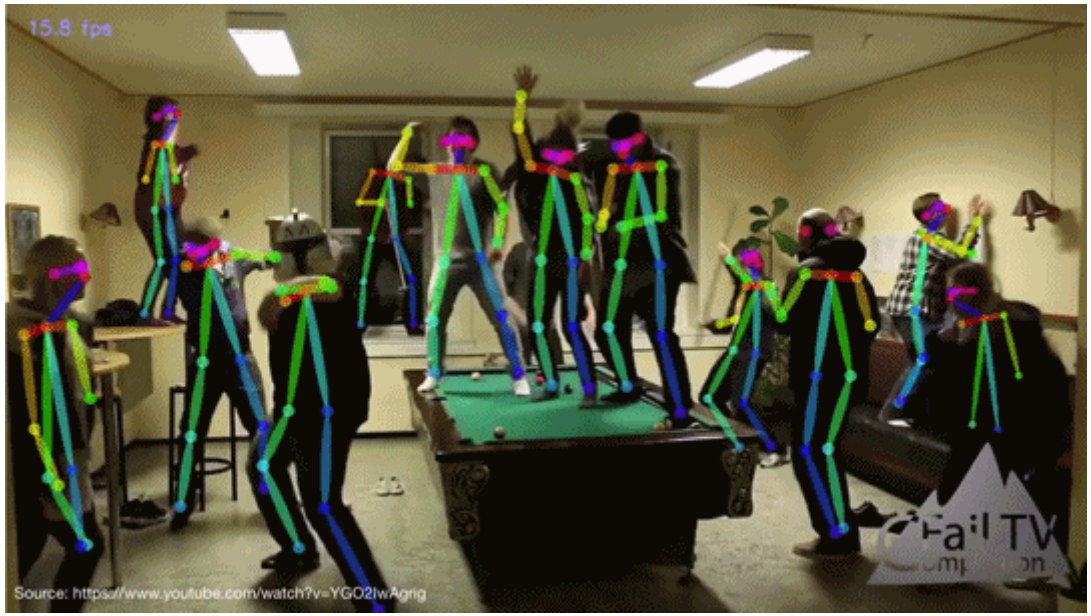
分野外研究者・企業 →

でも，ちょっと近い未来を考えてみましょう (特にこれから産学に羽ばたく若い皆さん！)



というわけで，第二期PRMU-GCの趣旨は…

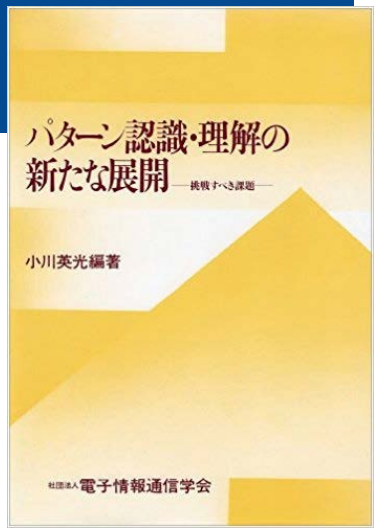
AIブームの今だからこそ，
CV/PRの研究者全体で
「これから」を議論したい



CMU-Perceptual-Computing-Lab/openpose

これまでの経緯

過去にもあったPRMU-GC



- 小川編著「パターン認識・理解の新たな展開」
  - 1994

15年
弱

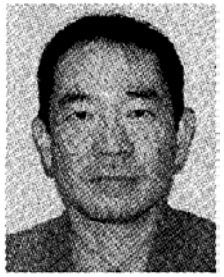
- 第一期PRMU-GC
  - 2007-2009

アンサンブル学習, SVM, MCMC, Particle Filter, SIFT, BoF, 一般物体認識, グラフカット, 近似最近傍探索

モチベーション：
国内でほどほどの研究で満足か？
思わせぶりで役に立たない研究でいいのか？

第一期PRMU-GCメンバ（50音順，敬称略）

- 内田 誠一@九大
- 佐藤 真一@NII
- 佐藤 洋一@東大
- 日浦 慎作@阪大(当時)
- 福井 和広@筑波大



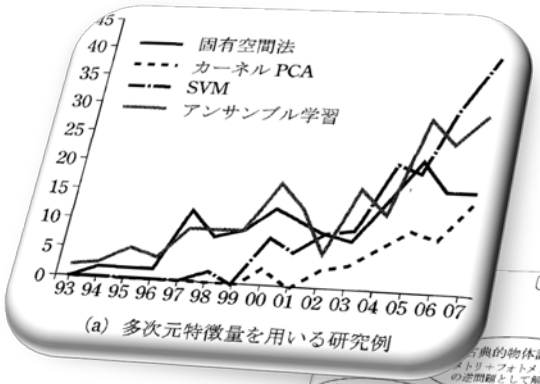
当時アラフォー！

馬場口 登
PRMU委員長(当時)

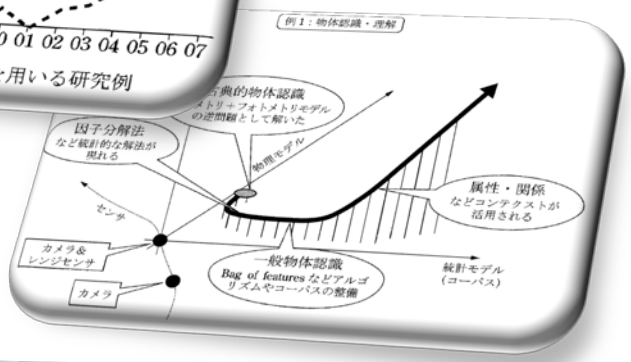
鷲見 和彦
PRMU副委員長(当時)

第一期PRMU-GCの成果

- 過去15年間の進展を整理



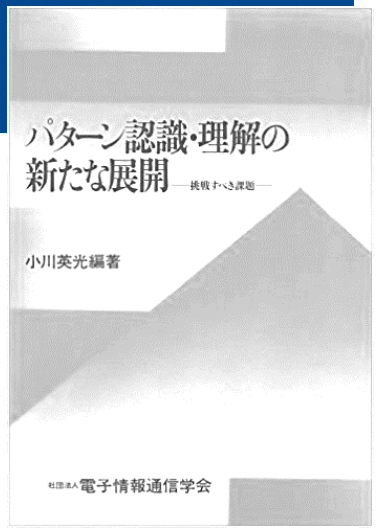
- 技術的展開を３軸で表現



- 今後取り組むべき10課題を提言
  - VQAを予見
  - 実際いずれも取り組まれた

人間の行動や動作の認識理解と評価.
任意の画像に対する自動関連付と構造化.
膨大な視覚情報から意味のある情報の抽出.
不足する視覚情報を与える人の視覚増強.
状況観察による直後の危険予知と通知.
撮影された画像に対する医療上の診断.
人物観察や人物状態からの環境認識.
情景内の文字認識.
地球規模センサ網による価値ある情報の認識.
任意の画像に関する意味記述や，質問応答.

そして第二期PRMU-GCへ



- 小川編著「パターン認識・理解の新たな展開」
  - 1994

15年
弱

- 第一期PRMU-GC
  - 2007-2009

10年
弱

- 第二期PRMU-GC

深層学習，大規模データ
凸最適化，質感情報処理

モチベーション：
ブームの今だからこそ考える
これからのCV/PR

そして第二期メンバ（50音順, 敬称略）

- 安倍 満@デンソーITラボ
- 木村 昭悟@NTT-CS研
- 中澤 篤志@京大
- 舩冨 卓哉@NAIST
- 松下 康之@阪大
- 山崎 俊彦@東大

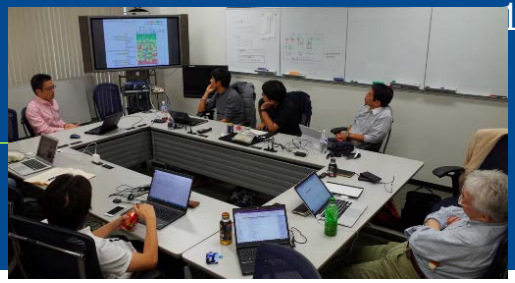
アラフォー！

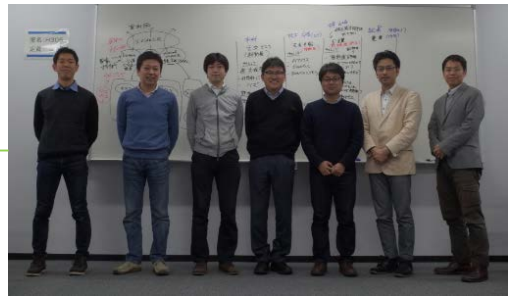
前田 英作
@PRMU前委員長

内田 誠一
@PRMU前副委員長

準備状況

- 2016.10
  - GCメンバーキックオフミーティング@東京
- 2016.12
  - PRMU@鳥取にてキックオフ企画
- 2017.2
  - GCメンバー第二回ミーティング@北海道
- 2017.3
  - GCメンバー第三回ミーティング@名古屋
- 2017.6
  - GCメンバー第一回合宿@熱海





+他分野研究者訪問(後述)

というわけで，
第二期PRMU-GCメンバによる
議論の中間報告を
お楽しみください

↓

全体の方向性 by 安倍さん

↓

メンバそれぞれの思い

ご意見募集！
よりよいものにするためにご協力お願いします！

goo.gl/TgTAjb

提案・感想・文句，
なんでもOK

匿名OK!

第二期
PRMUグランドチャレンジ

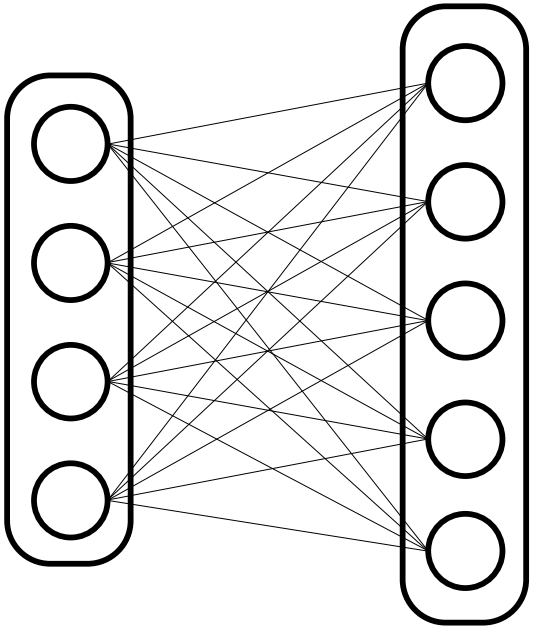
2017/08/09 @ MIRU2017

大きなパラダイムシフト
～パターン認識の問題は解かれたのか？～

ラベル付き大量データ

画像取得先：imagenet (http://image-net.org/)

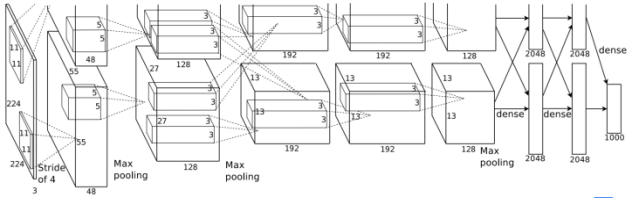
＋



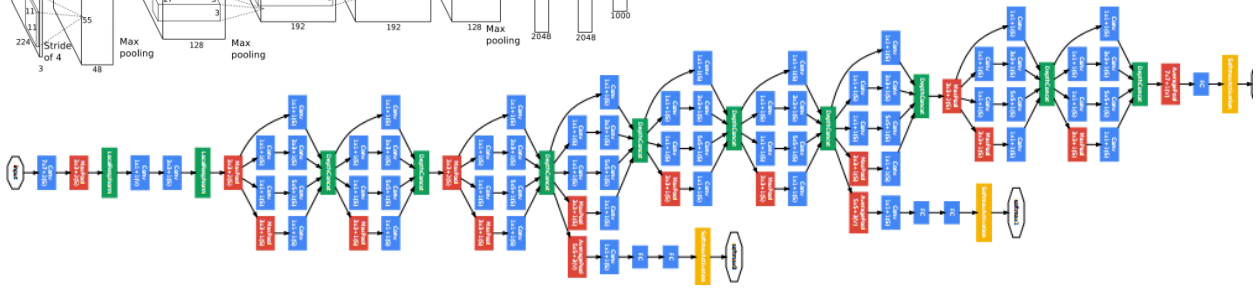
Deep Neural Network

ラベル付き大量データ＋DNN

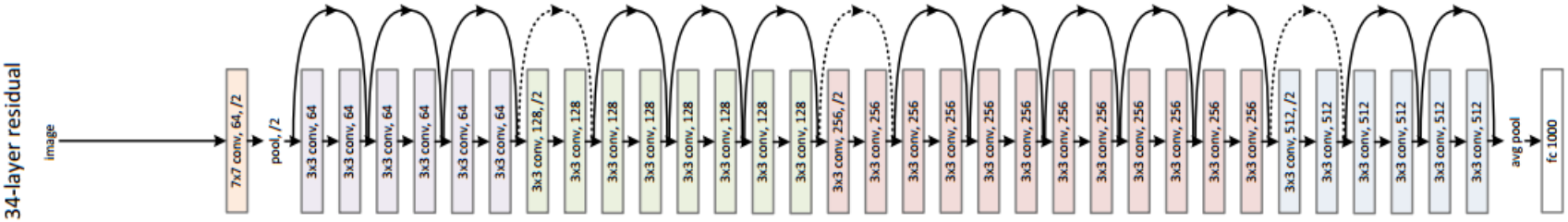
- Classification



ImageNet Classification with Deep Convolutional Neural Networks, NIPS2012



Going Deeper With Convolutions CVPR2015

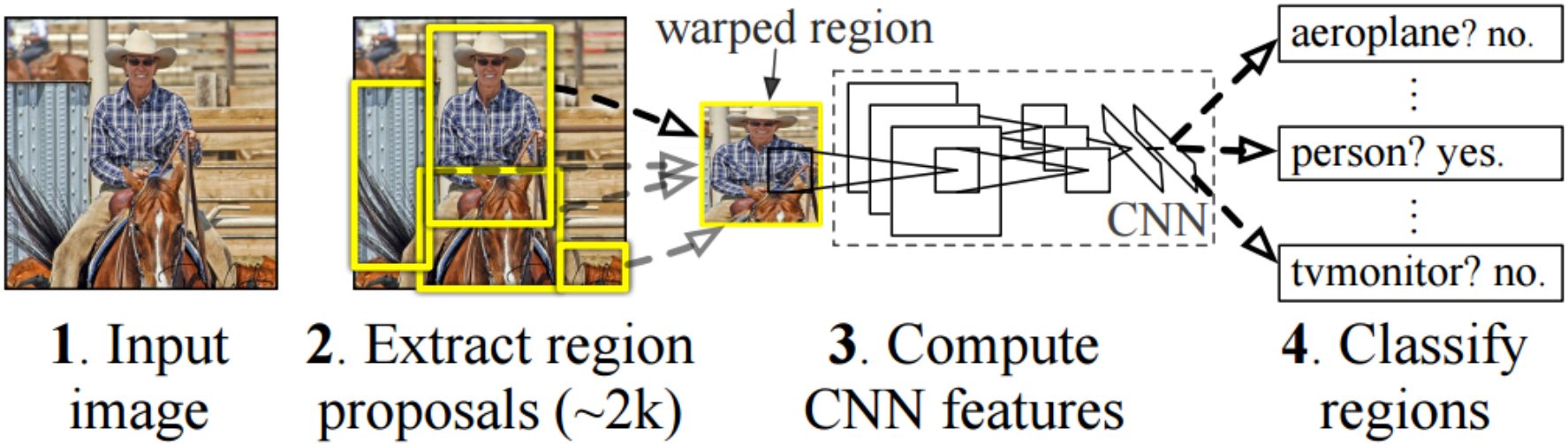


Deep Residual Learning for Image Recognition, CVPR2016

Grand Challengeアンケート→ https://goo.gl/TgTAjb

ラベル付き大量データ＋DNN

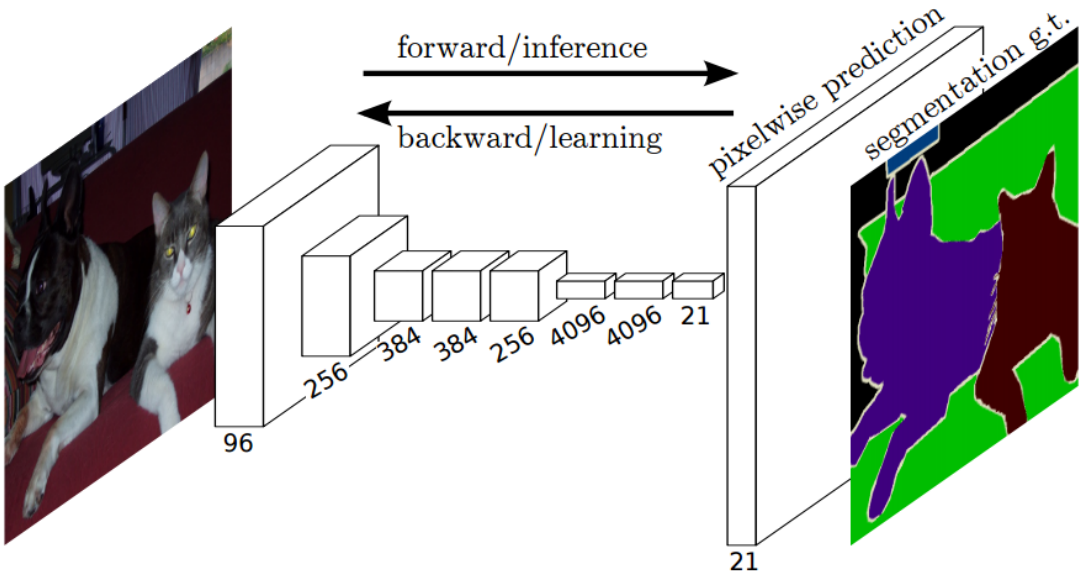
- Object detection



Region-based Convolutional Networks for Accurate Object Detection and Semantic Segmentation, CVPR2014

ラベル付き大量データ＋DNN

- Semantic segmentation



Fully Convolutional Networks for Semantic Segmentation, CVPR2015

ラベル付き大量データ＋DNN

- Human pose recognition

Convolutional Pose Machines, CVPR2016

パターン認識の定石

- ラベル付き大量データ＋DNN
  - Classification
  - Object detection
  - Semantic segmentation
  - Human pose recognition

異なるタスクが同一の定石に当てはまる

パターン認識の定石

- ラベル付き大量データ＋DNN
  - Classification
  - Object detection
  - Semantic segmentation
  - Human pose recognition

定石にさえ当てはまれば、あらゆる
パターン認識の問題は解けてしまうのか？

データを収集することの難しさ

- **データ量の問題**
- **データの偏りの問題（正例>>負例）**
  - 外観検査
  - 医用画像処理
- **収集の危険性／コンプライアンスの問題**
  - 自動運転における安全な学習とは？

画像だけで理解することの難しさ

- A man hugging someone close under a cloudy sky.
- Barack Obama is in this picture.
- The other is probably his wife, Michelle.

画像だけで理解することの難しさ

- 選挙候補者が映っている
- 選挙期間中に投稿された
- 任期が4年である
- 故にこの画像は**選挙運動の一環**

Grand Challengeアンケート→ https://goo.gl/TgTAjb

実問題に適用することの難しさ

- 産業と学術の乖離
  - “Research for Research” vs “too much domain specific”

Finding new research field

Exhibit E2

The MGI Industry Digitization Index

2015 or latest available data

Relatively low digitization — Relatively high digitization

● Digital leaders within relatively undigitized sectors

Sector	Overall digitization[1]	Assets: Digital spending	Assets: Digital asset stock	Usage: Transactions	Usage: Interactions	Usage: Business processes	Usage: Market making	Labor: Digital spending on workers	Labor: Digital capital deepening	Labor: Digitization of work	GDP share %	Employment share %	Productivity growth, 2005–14[2] %
ICT											5	3	4.6
Media											2	1	3.6
Professional services											9	6	0.3
Finance and insurance											8	4	1.6
Wholesale trade											5	4	0.2
Advanced manufacturing											3	2	2.6
Oil and gas											2	0.1	2.9
Utilities											2	0.4	1.3
Chemicals and pharmaceuticals											2	1	1.8
Basic goods manufacturing											5	5	1.2
Mining											1	0.4	0.5
Real estate	●										5	1	2.3
Transportation and warehousing	●										3	3	1.4
Education	●										2	2	-0.5
Retail trade	●										5	11	-1.1
Entertainment and recreation											1	1	0.9
Personal and local services											6	11	0.5
Government	●										16	15	0.2
Health care											10	13	-0.1
Hospitality	●										4	8	-0.9
Construction											3	5	-1.4
Agriculture and hunting											1	1	-0.9

AdTech → Media

FinTech → Finance and insurance

RealEstateTech → Real estate

EdTech → Education

HealthTech → Health care

おもてなし → Hospitality

AgriTech → Agriculture and hunting

www.mckinsey.com

実問題に適用することの難しさ

- 農業系研究者を訪問

  農研機構岡田先生・杉浦先生／東大二宮先生・野下先生／三重大学亀岡先生
  九州大学南石先生／かずさDNA研究所七夕先生

  - 曰く「農業＝ビジネス」
    利益向上が見込める技術でないと現場の共感が得られない
    パターン認識の価値が理解されない→データが集まらない／整理されない→応用が広がらない・・・

- 介護系研究者を訪問

  静岡大学竹林先生・石川先生

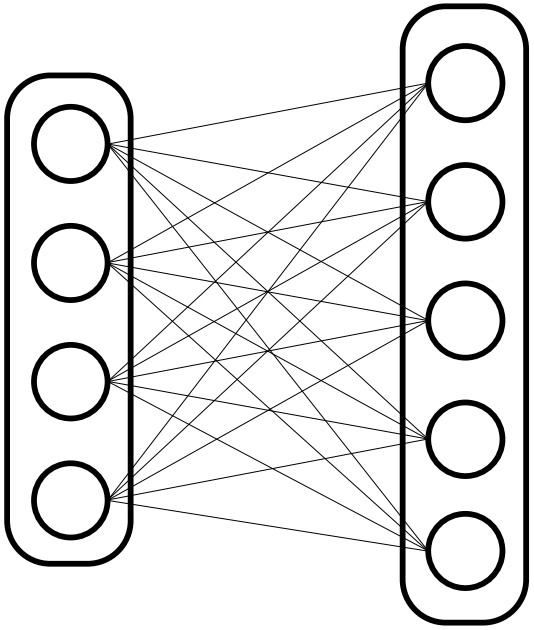
  - 「コミュニケーション」をどうやって定量化し、パターン認識の問題に落としこむか？
    介護エキスパート技術をどのように伝達すればよいか？

パターン認識の定石と現実のギャップ

ラベル付き大量データ

画像取得先：imagenet (http://image-net.org/)

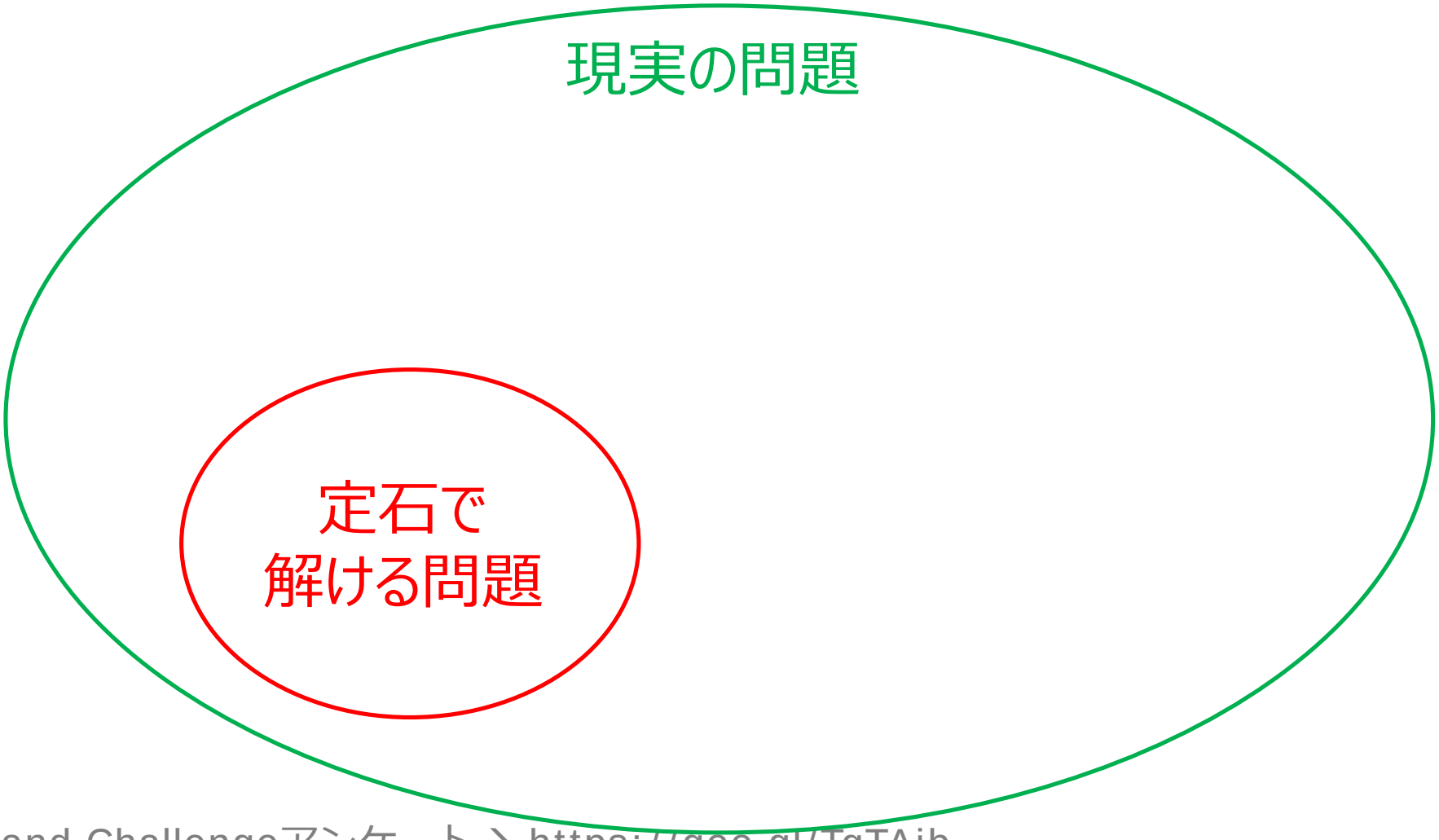
+



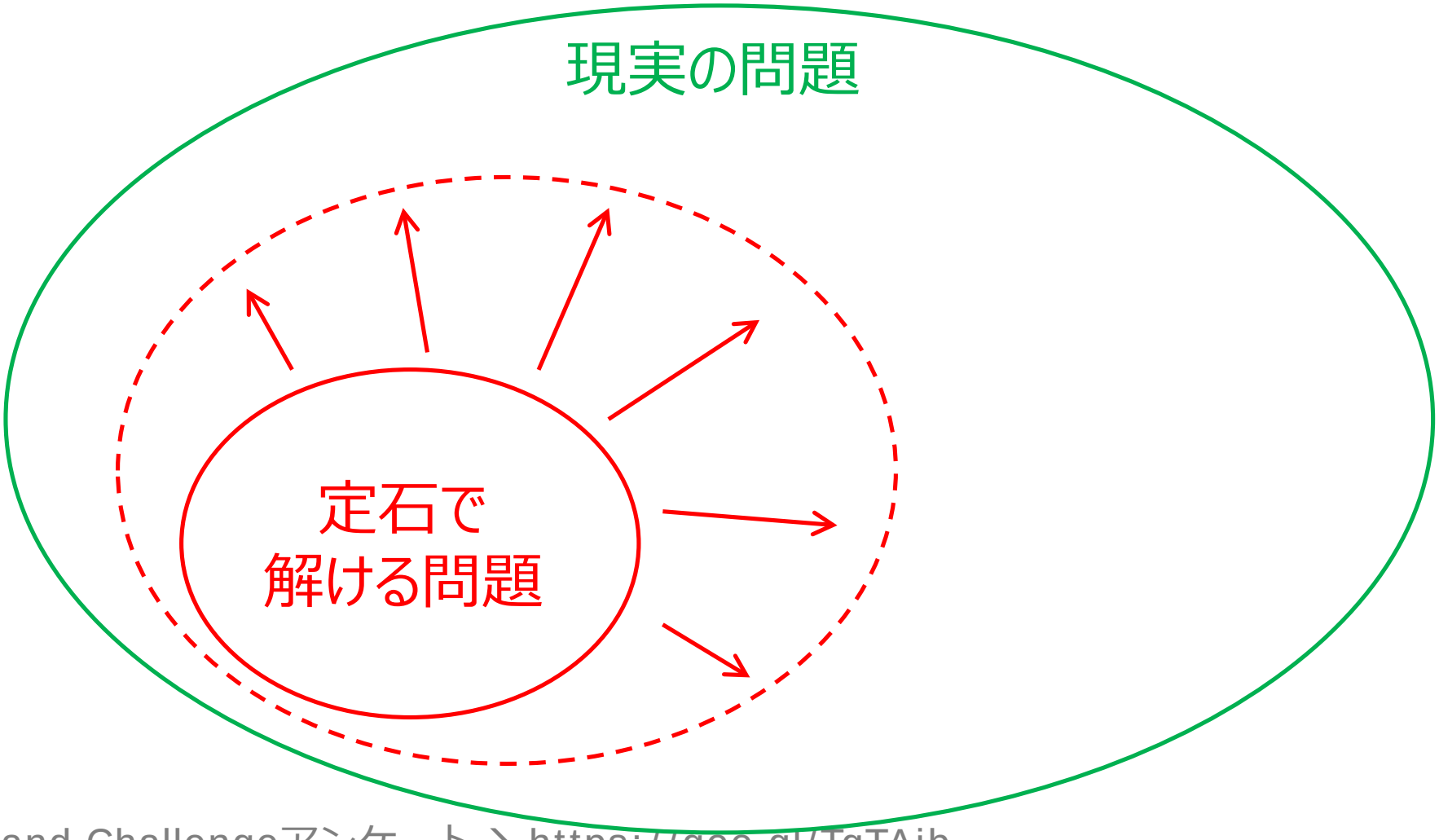
Deep Neural Network

Grand Challengeアンケート→ https://goo.gl/TgTAjb

パターン認識の定石と現実のギャップ



パターン認識の定石と現実のギャップ



パターン認識の次なるステージに向けて

- 解けない問題を再認識
- 解けない問題を解くための切り口を議論

→ Open Idea

計算困難な問題へのパターン認識の応用

大阪大学　松下康之

@MIRU 2017, PRMU-GC企画

https://goo.gl/TgTAjb

教師あり学習を抽象化すると

$$x \rightarrow y$$

例）画像 → オブジェクト名

音声 → テキスト

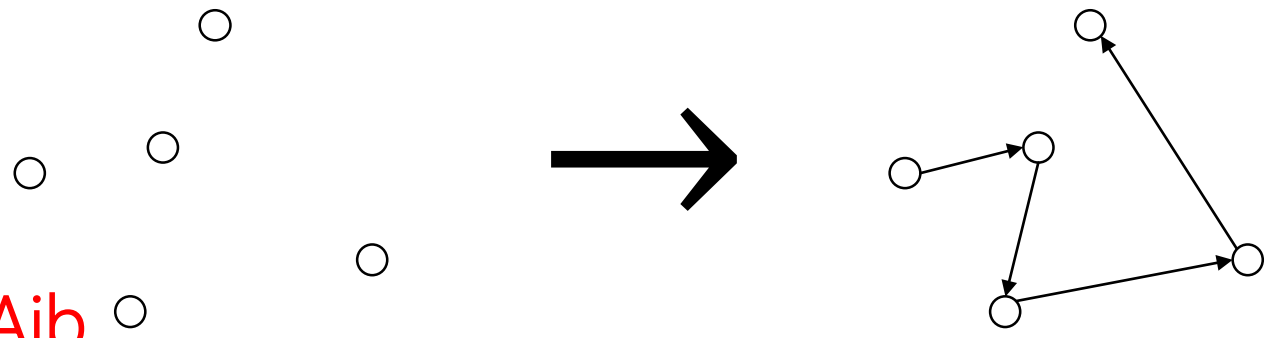
https://goo.gl/TgTAjb

NP困難, EXP問題

計算量の爆発

$$x \rightarrow y$$

例）循環セールスマン問題



https://goo.gl/TgTAjb

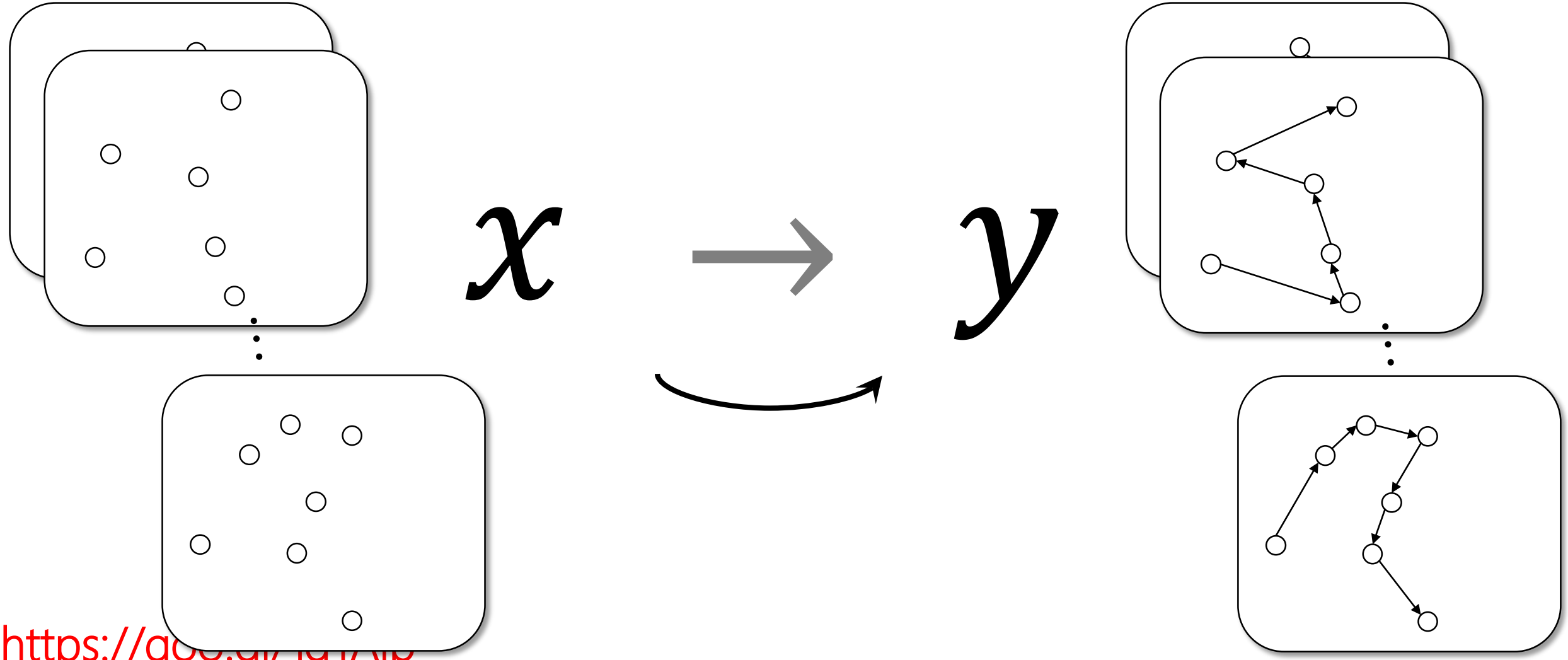
パターン認識 ⇒ Computing

$$x \rightarrow y$$

パターン認識・機械学習で
計算の困難さをバイパスできるだろうか？

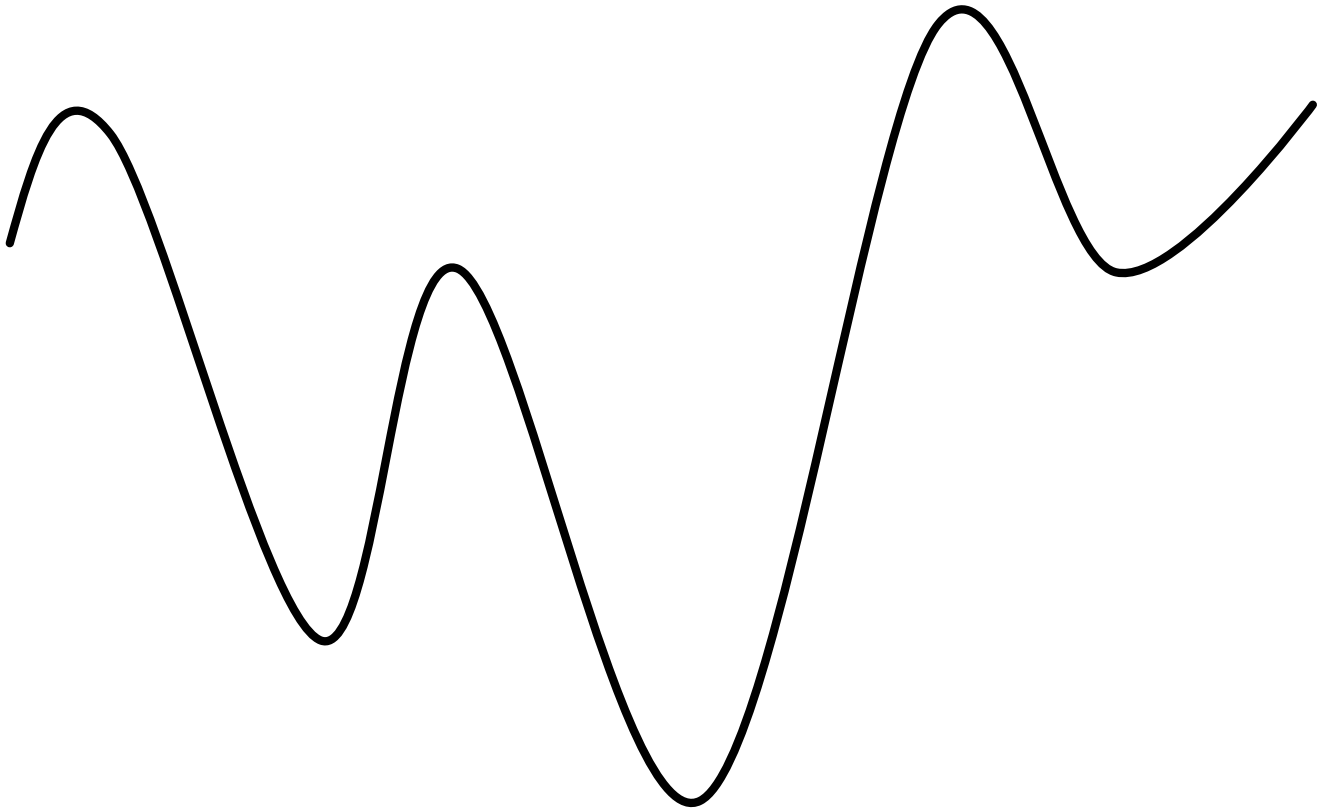
https://goo.gl/TgTAjb

大量の学習データとパターン認識は計算困難さを克服するか？



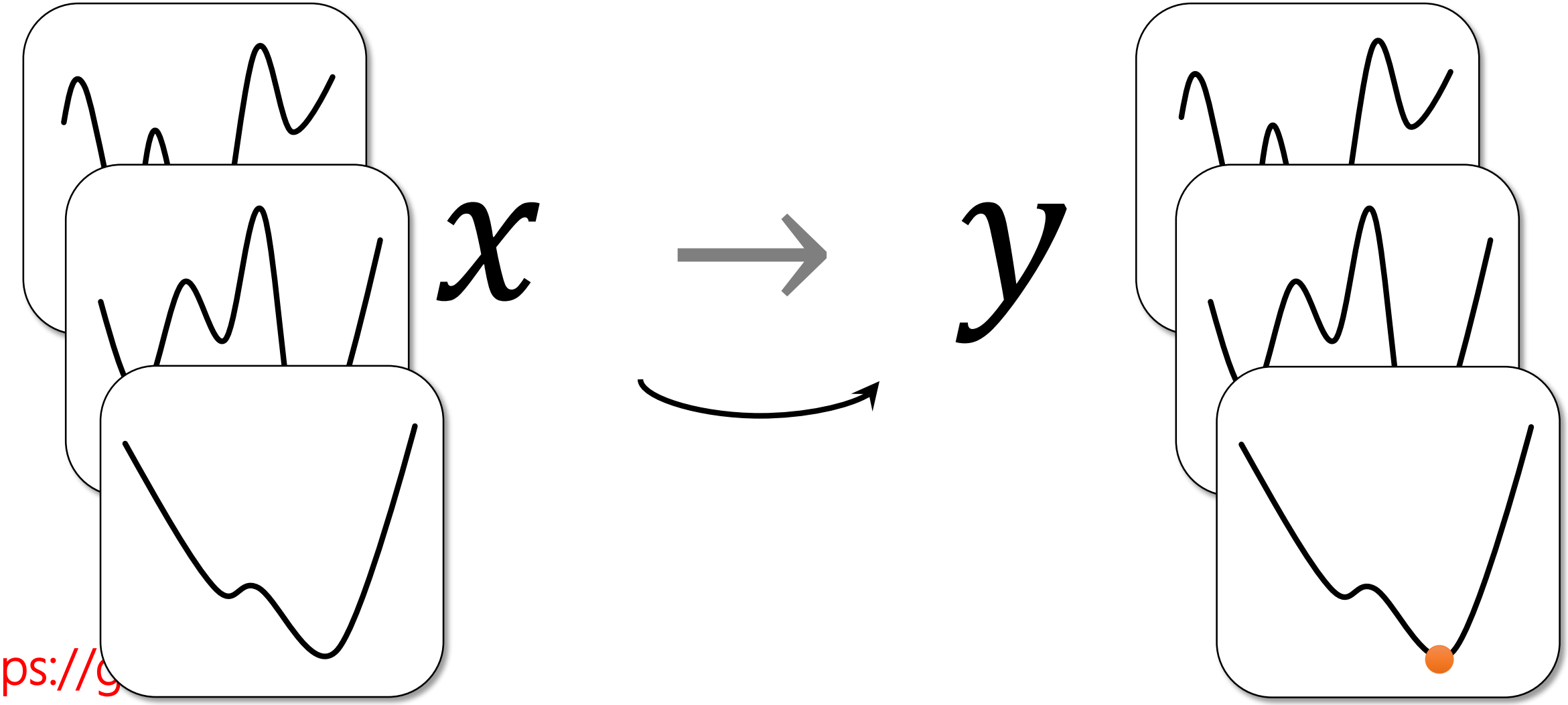
https://goo.gl/ 1g1Kjb

例）非凸関数の大域最適化



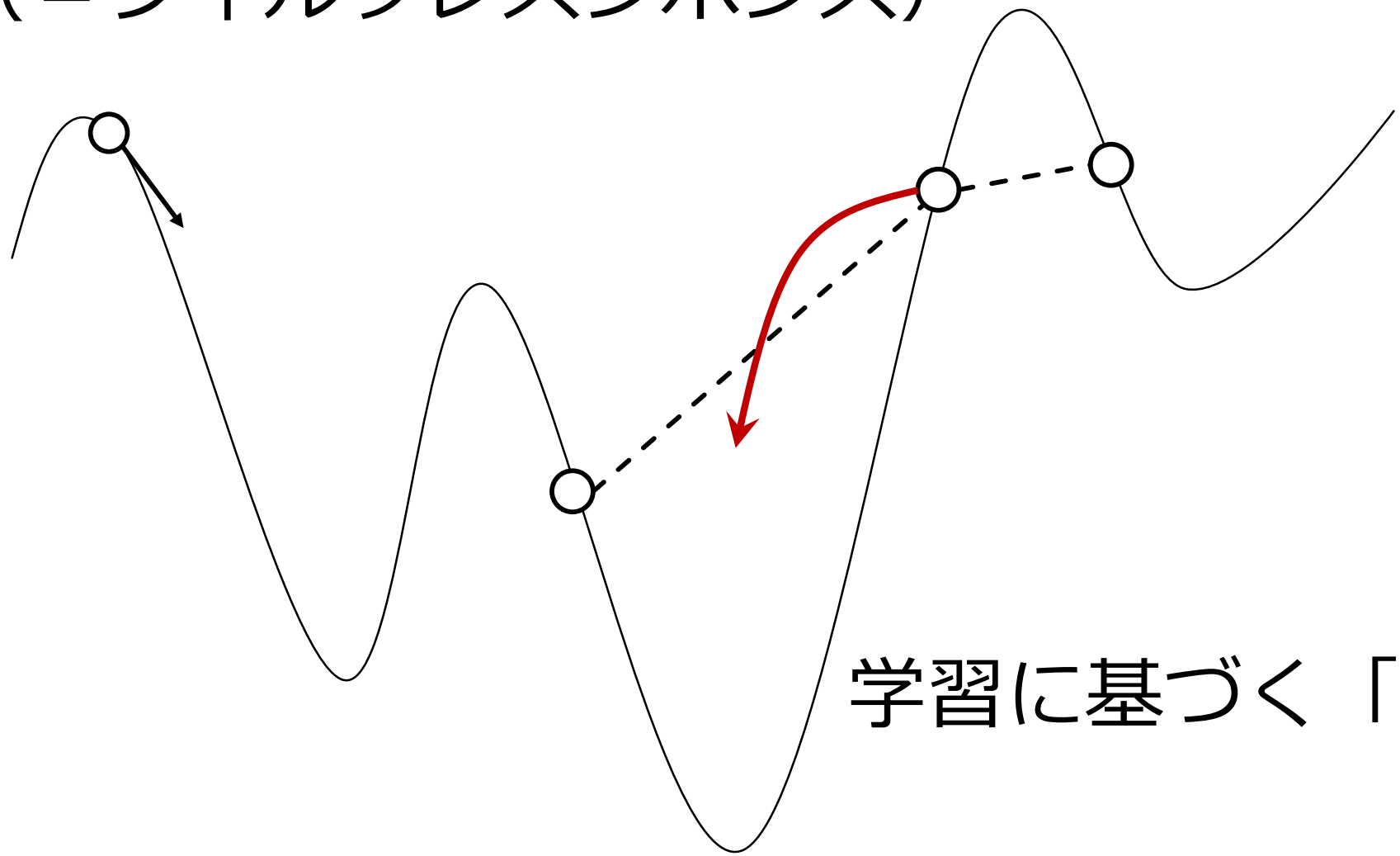
https://goo.gl/TgTAjb

例）非凸関数の大域最適化



https://g

勾配（＝フィルタレスンポンス）



学習に基づく「勾配」は？

https://goo.gl/TgTAjb

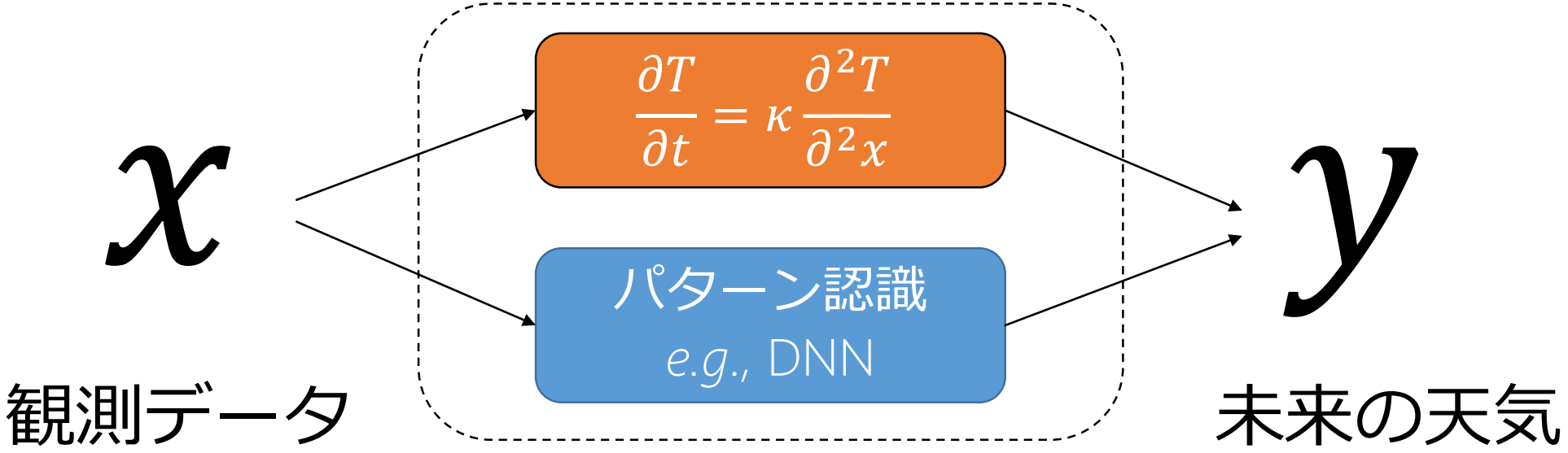
パターン認識で未来予測は？ - 天気予報

$$x \rightarrow y$$

観測データ　　未来の天気

https://goo.gl/TgTAjb

パターン認識で未来予測は？ - 天気予報

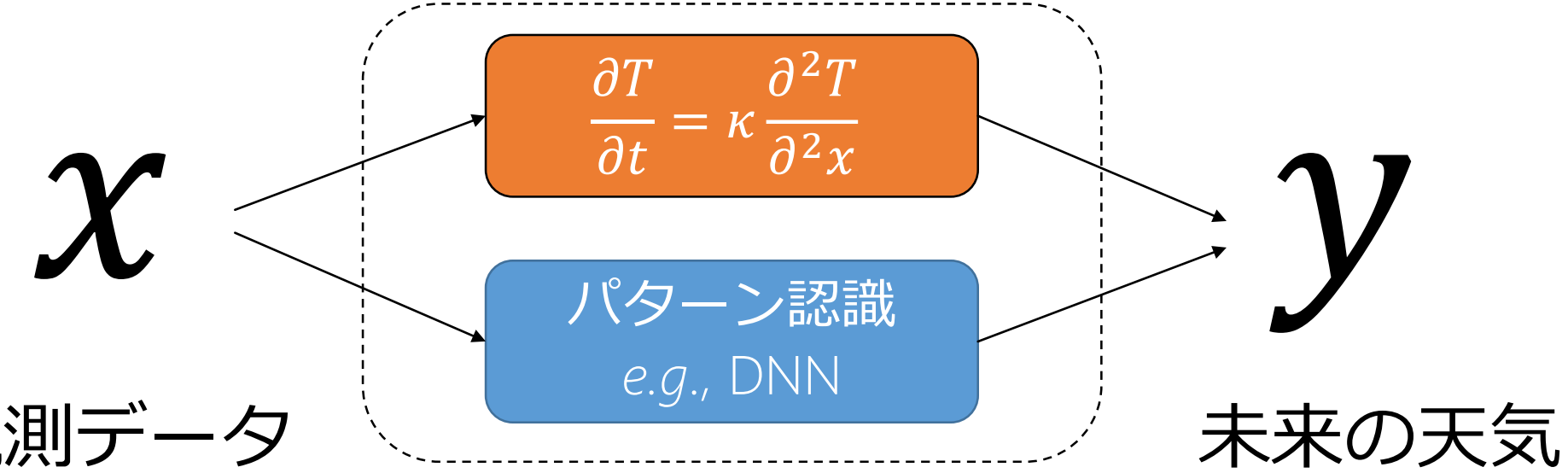


https://goo.gl/TgTAjb

パターン認識で未来予測は？ - 天気予報

演繹的アプローチ(Deduction)

公理から合理的推論によって結論を導く



帰納的アプローチ (Induction)

経験的データを収集して結論を導く

https://goo.gl/TgTAjb

まとめ

- パターン認識・機械学習は計算困難さを克服するか？
  - NP困難問題、EXP問題への応用
- 未来予測問題への応用
  - 長期天気予報は可能か？
- 機械学習は帰納的
  - 演繹的アプローチとの融合の形とは？

$$x \rightarrow y$$

https://goo.gl/TgTAjb

柔らかなパターン認識を目指して

デンソーアイティーラボラトリ
安倍 満

2017/08/09

https://goo.gl/TgTAjb

再掲：導入部のスライドより抜粋

データを収集することの難しさ

- **データ量の問題**
- **データの偏りの問題（正例>>負例）**
  - 外観検査
  - 医用画像処理
- **収集の危険性／コンプライアンスの問題**
  - 自動運転における安全な学習とは？

https://goo.gl/TgTAjb

『固いパターン認識』の限界

https://goo.gl/TgTAjb

『固いパターン認識』の限界

画像

パターン
認識器

舵角

https://goo.gl/TgTAjb

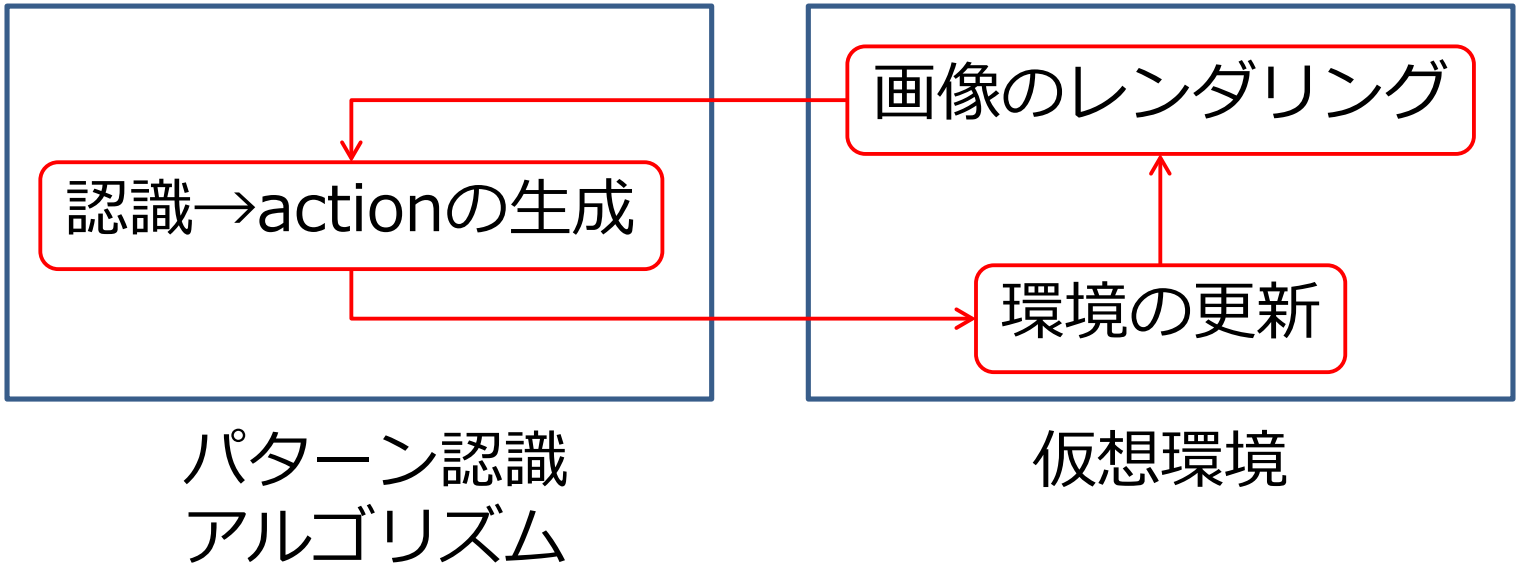
『固いパターン認識』の限界

画像　　舵角

- **失敗から学ぶ**ことが難しい
  （学習データを集めるために危険を冒すわけにはいかない）

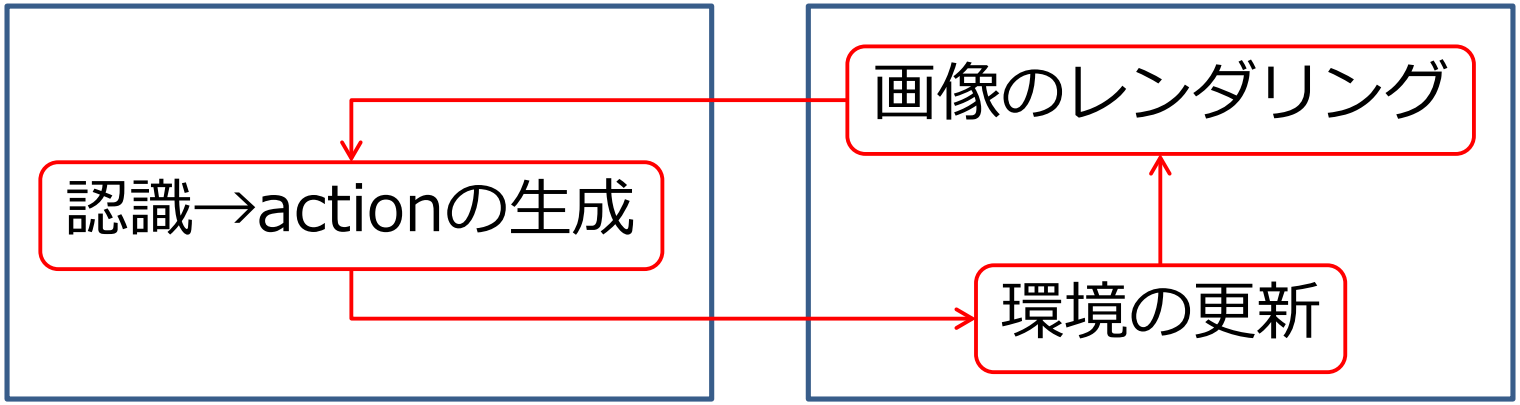
https://goo.gl/TgTAjb

『柔らかなパターン認識』を目指して



https://goo.gl/TgTAjb

『柔らかなパターン認識』を目指して



パターン認識
アルゴリズム

仮想環境

CG + Machine Learning

https://goo.gl/TgTAjb

CG + Machine Learning → Next CV

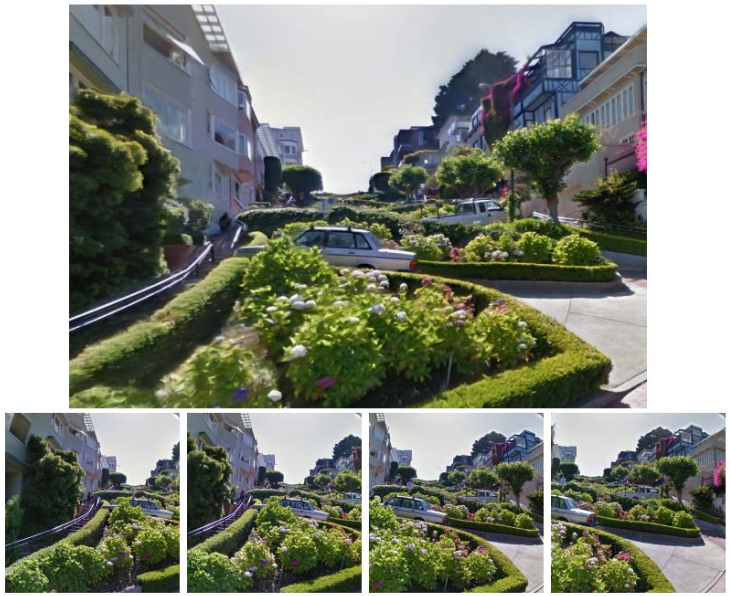
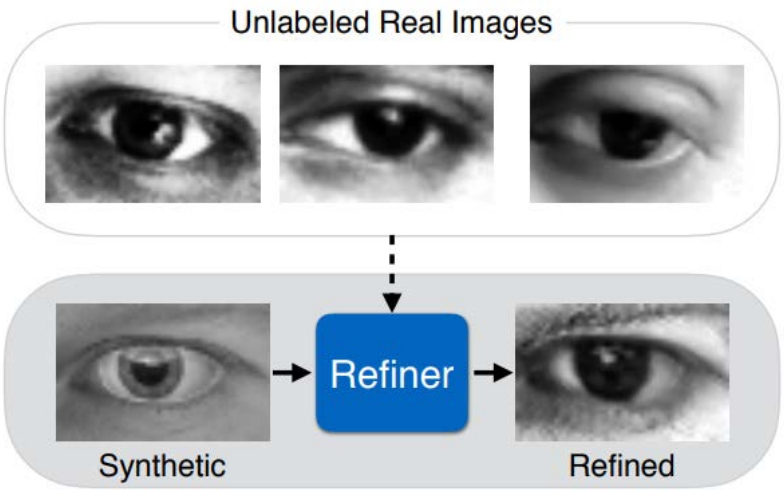
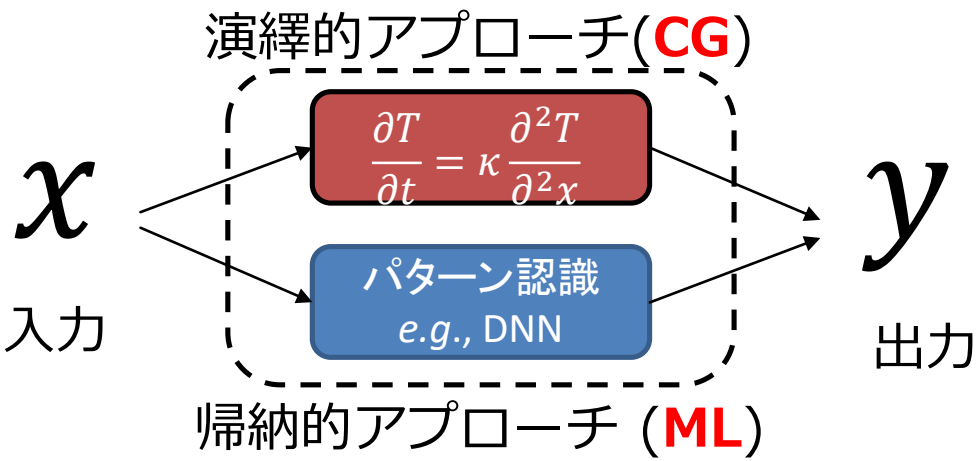


Figure 1: The top image was synthesized from several input panoramas. A portion of four of the inputs is shown on the bottom row.

DeepStereo: Learning to Predict New Views from the World's Imagery (Google, CVPR2016)



SimGAN, CVPR2017



https://goo.gl/TgTAjb

Challenges

- 質＝リアリティ
  - ✓ 画質の再現
  - ✓ 物理法則の再現
  - ✓ エージェントの動作ふるまいの再現
- 量＝スケーラビリティ
  - ✓ 実世界画像から仮想世界を生成（image-based rendering）
- 仮想環境→実環境への転移
- 能動的な学習の戦略

https://goo.gl/TgTAjb

【 前編: https://goo.gl/9hVKKc 】

Innovative R&D by NTT

メディア理解の多重性とその解消に向けて（続編）

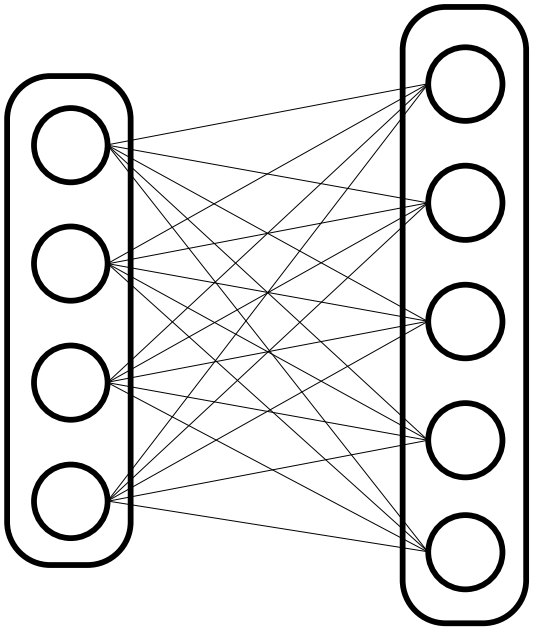
木村　昭悟　（きむら　あきさと）
NTTコミュニケーション科学基礎研究所
E-mail: kimura.akisato@lab.ntt.co.jp

Copyright©2014 NTT corp. All Rights Reserved.

パターン認識の「定石」

ラベル付き大量データ

＋



Deep Neural Network

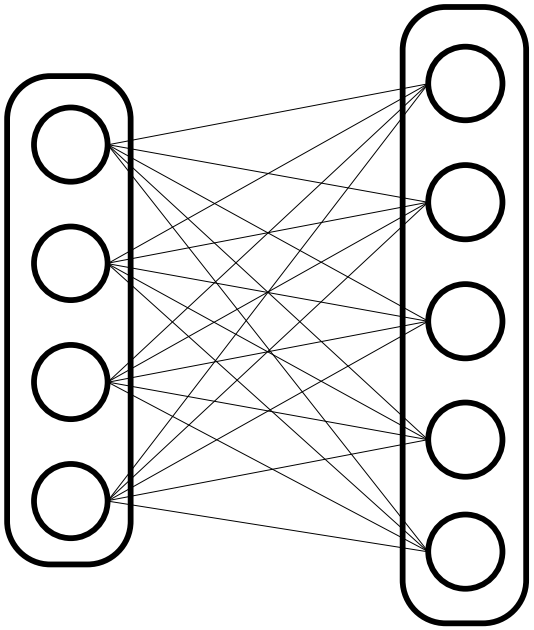
【ご意見は https://goo.gl/TgTAjb へ】

Copyright©2014 NTT corp. All Rights Reserved.

パターン認識の「定石」を拡張する

ラベル~~付き~~大量データ
なし

＋



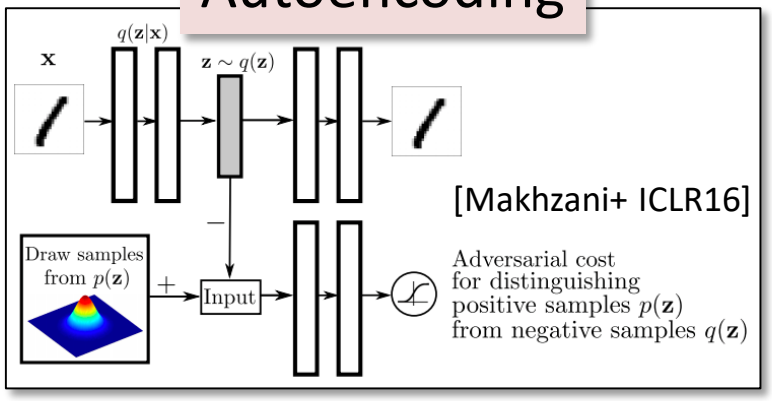
Deep Neural Network

【ご意見は https://goo.gl/TgTAjb へ】

Copyright©2014 NTT corp. All Rights Reserved.

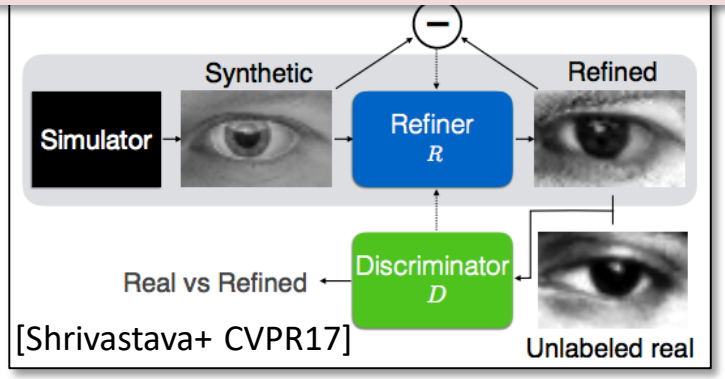
正解データ不足問題は解けつつある

Autoencoding



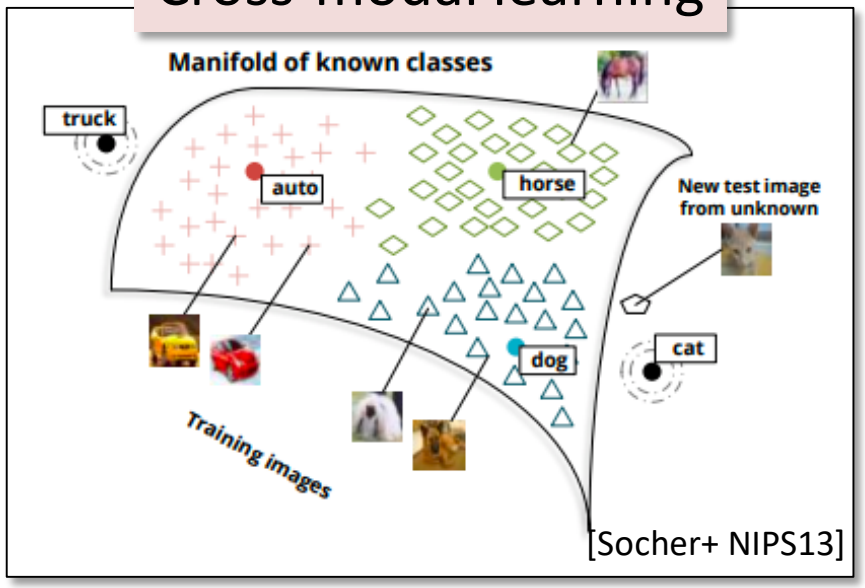
[Makhzani+ ICLR16]

Synthesizing training samples



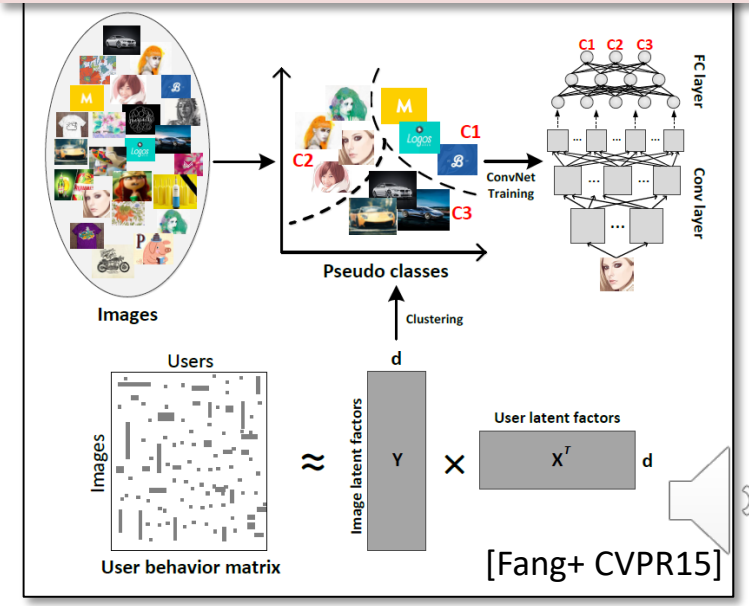
[Shrivastava+ CVPR17]

Cross-modal learning



[Socher+ NIPS13]

Weakly supervised learning



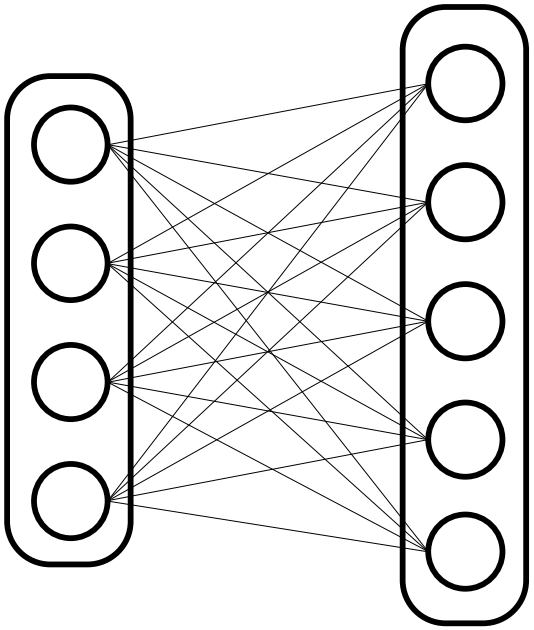
[Fang+ CVPR15]

「定石」では解けない問題

ラベル~~付き~~大量データ

不定

＋



Deep Neural Network

【ご意見は https://goo.gl/TgTAjb へ】

Copyright©2014 NTT corp. All Rights Reserved.

「定石」では解けない問題

- パーソナライズ認識：
  正解ラベルが状況によって異なる認識
- 「究極の選択」マシン：
  正解ラベルがどこにもない認識

ラベル~~付き~~大量データ
不定

Deep Neural Network

【ご意見は https://goo.gl/TgTAjb へ】

Copyright©2014 NTT corp. All Rights Reserved.

パーソナライズ認識

- 唯一無二の正解を求める画一的な認識から，個人や状況によって正解が異なる認識へ．

[Ordonez+ ICCV13]

そんな難しいことを言われても困る

ハナゴンドウ　grampus griseus

そんなことは見ればわかるから

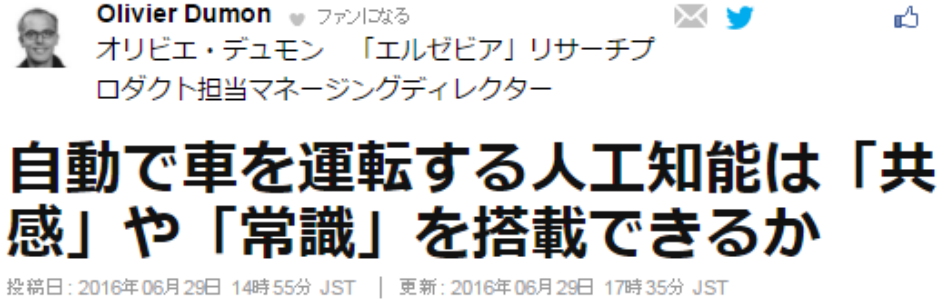
dolphin　イルカ

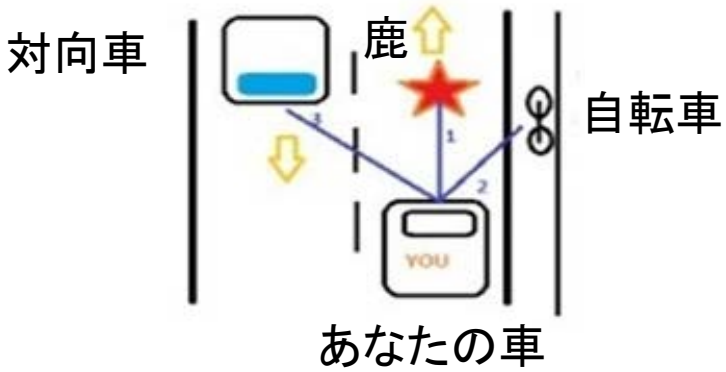
【ご意見は https://goo.gl/TgTAjb へ】

Copyright©2014 NTT corp. All Rights Reserved.

究極の選択マシン

- 唯一無二の評価尺度を信仰する認識から，複数の評価尺度の選択的利用を行う認識へ．





- 時速80kmで移動中の自動運転車
- 突然大型の鹿が目の前に飛び出した．
- 右側には自転車，左の反対車線には車．

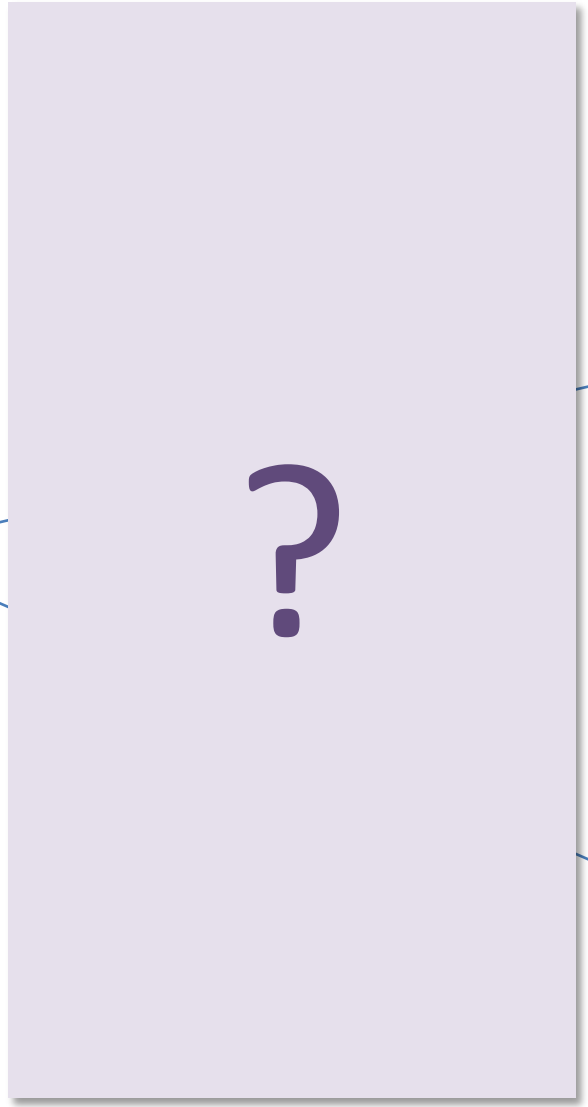
1. 全力ブレーキ → 鹿は確実に死ぬ，自分も重傷もしくは死亡，車は大破．
2. 右ハンドル → サイクリストは重傷もしくは死亡，自分は無事．
3. 左ハンドル → 自分と対向車の運転手の死亡確率が50%．

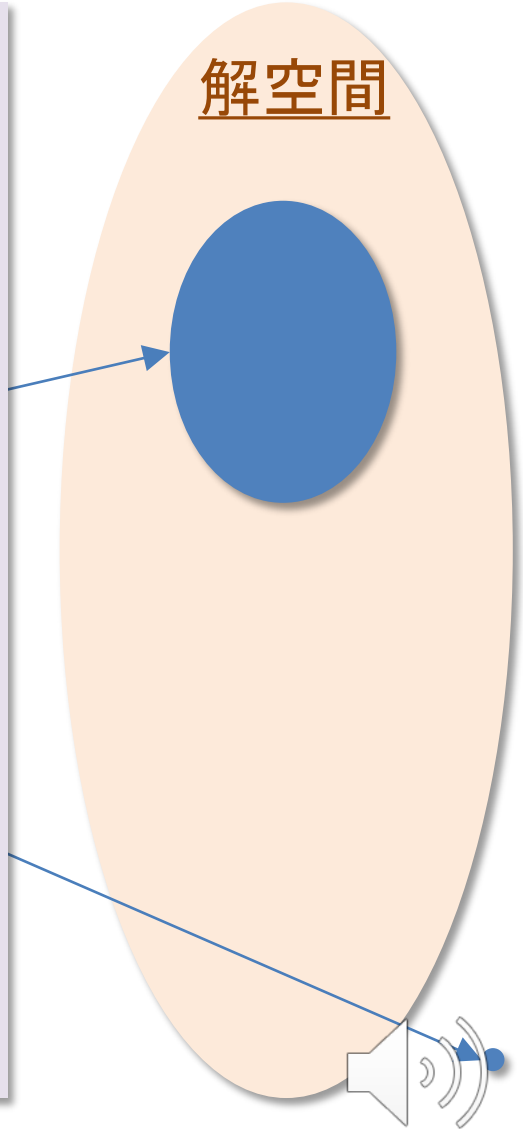
http://www.huffingtonpost.jp/olivier-dumon/empathy_technology_b_10727[illegible]52.html

【ご意見は https://goo.gl/TgTAjb へ】

Copyright©2014 NTT corp. All Rights Reserved.

正解不定問題の構図





【ご意見は https://goo.gl/TgTAjb へ】

Copyright©2014 NTT corp. All Rights Reserved.

正解不定問題の構図

Innovative R&D by NTT

認識対象

外側の世界

Commonsense
環境・文脈

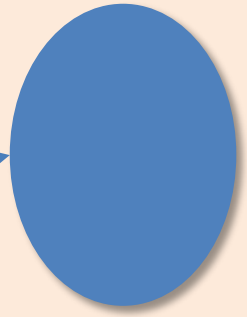
送り手

Personality
周辺コンテンツ

受け手

保有知識・嗜好
評価尺度

解空間



【ご意見は https://goo.gl/TgTAjb へ】

Copyright©2014 NTT corp. All Rights Reserved.

正解不定問題への道筋

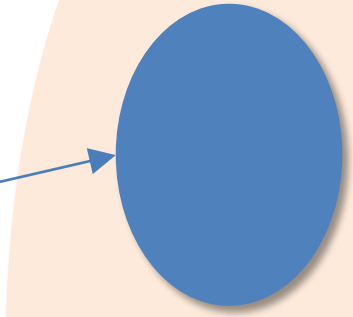
認識対象

【知識獲得】
解空間を広くする
【強化学習】
解空間を知る

【編集】
解の任意性を削減
【対話】
解を絞り込む

【最適化】
解を見つける
【推論】
解空間を狭める

解空間



【ご意見は https://goo.gl/TgTAjb へ】

Copyright©2014 NTT corp. All Rights Reserved.

Copyright©2014 NTT corp. All Rights Reserved.

情報量の高いセンシング

奈良先端科学技術大学院大学
舩冨 卓哉
2017/08/09

https://goo.gl/TgTAjb

民生カメラに実装される高度な撮影

theta360.com

	民生
画角/解像度	4π/4K
視点数	13×13
Framerate	1200
波長	RGB/FIR

www.lytro.com

www.nikon-image.com

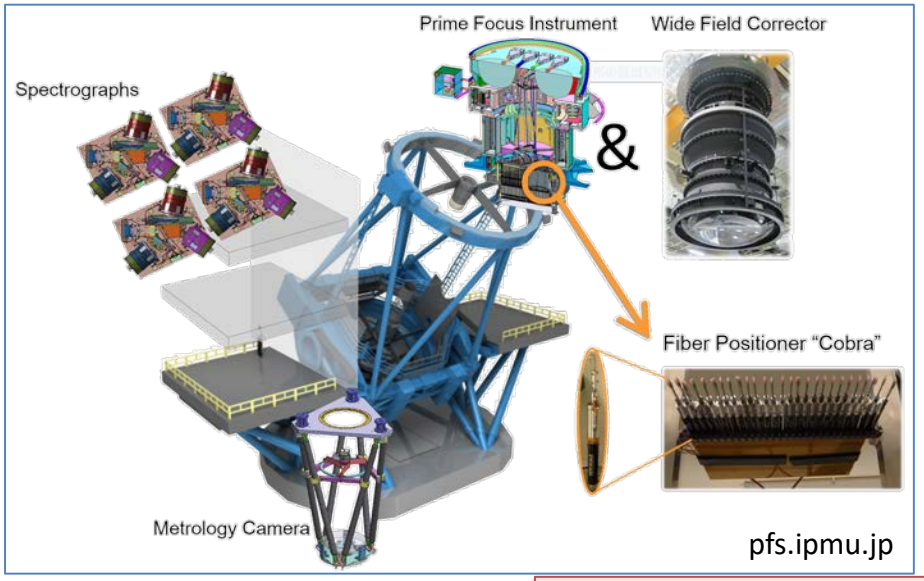
www.catphones.com

研究はやりやすい・適用範囲は広いが，
- これで満足していていいのか？
- これに特化した処理だけでいいのか？

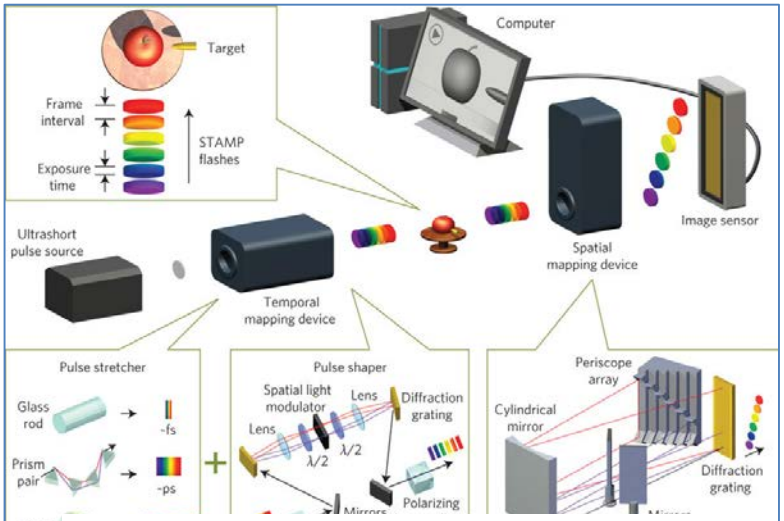
https://goo.gl/TgTAjb

科学技術分野で利用される最先端

	民生	
波長	RGB/FIR	380-1260nm/2Å (Prime Focus Spectrograph)
Framerate	1200	1T fps (STAMP)



pfs.ipmu.jp



Sequentially timed all-optical mapping photography

民生品とはスケールが異なる

https://goo.gl/TgTAjb

究極のカメラ：Plen-optic センシング

Plen-optic function: $P\left(\theta, \phi, \lambda, t, \boldsymbol{P}_{x,y,z}\right)$

	変数	
画角	θ, ϕ	The Panoptic Studio CMU
視点数	$\boldsymbol{P}_{x,y,z}$	
Framerate	t	30fpsより，もっと高速に
波長	λ	RGBより，もっと多波長で

そう単純にはいかない
- 光の量が激減：劣悪なSNR
- 膨大で冗長なデータ

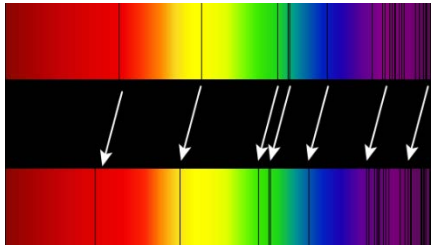
https://goo.gl/TgTAjb

そんな膨大なデータで何が見たい？

撮ったデータのほとんどが冗長

その中のごく少数の異常（＝高情報量）を見たい

– Photonが返ってきた時刻が欲しい
  - ほとんどが真っ黒
– フラウンホーファー線のずれ量を見たい
  - 天体速度によるドップラー効果や重力
– 反射特性：Diffuse＋Specular
  - きらりと光る反射の強さや範囲



commons.wikimedia.org

https://goo.gl/TgTAjb

Challenges：情報量の高いセンシング

データを収集することの難しさ

- **データ量の問題**
- **データの偏りの問題（正例>>負例）**
  - 外観検査
  - 医用画像処理
- **収集**
  - 自動

観測と解析を同時に行い，
情報量が高い事象を効率よく
センシングする機構を考えられないか？
– 負例の効率的収集にも繋がる

https://goo.gl/TgTAjb

予測不可能なシステムとしての人間

- 「人を見る」研究はどこに進むか? -

中澤　篤志

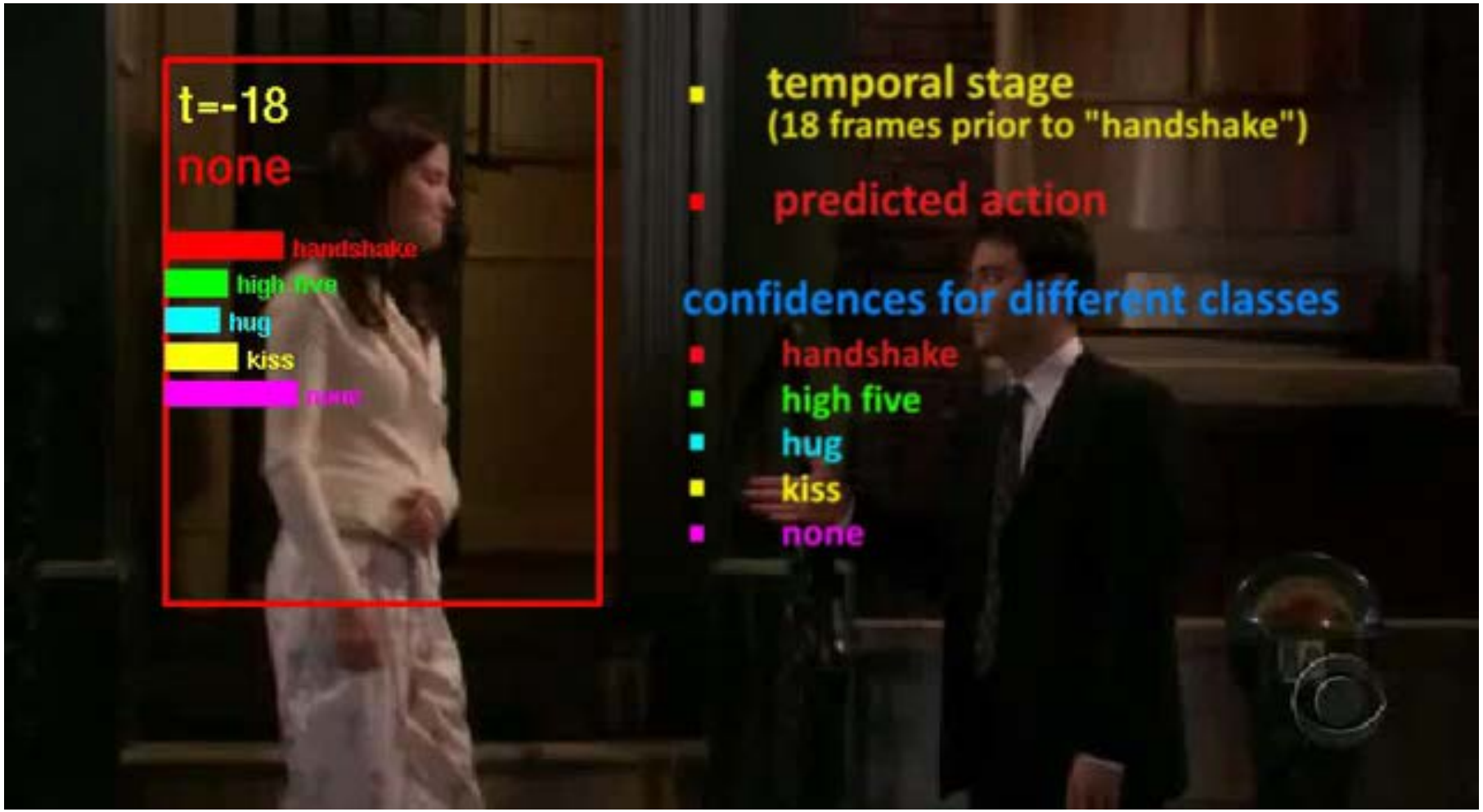
京都大学大学院情報学研究科

- **安倍さんのスライド（GCメンバの現状認識）**
  1. 基本的にEnd-to-End学習で解ける問題はDNNで学習すれば解ける（であろう）
     e.g. 画像からの人の姿勢推定
  2. そのスキームに持ち込めない問題をどう取り扱うか？
     ・例えば、正例を得るのが難しい場合
     ・正例・不例がアンバランスな場合
     ・背景知識が必要な場合

- **人と人とのコミュニケーション（例えば介護）**
  - 何が人の行動を規定するのか？

人の行動予測

- 現在の人の行動予測 → 限られたシナリオ・時間フレーム



A Hierarchical Representation for Future Action Prediction, ECCV14

介護シーン（実問題）

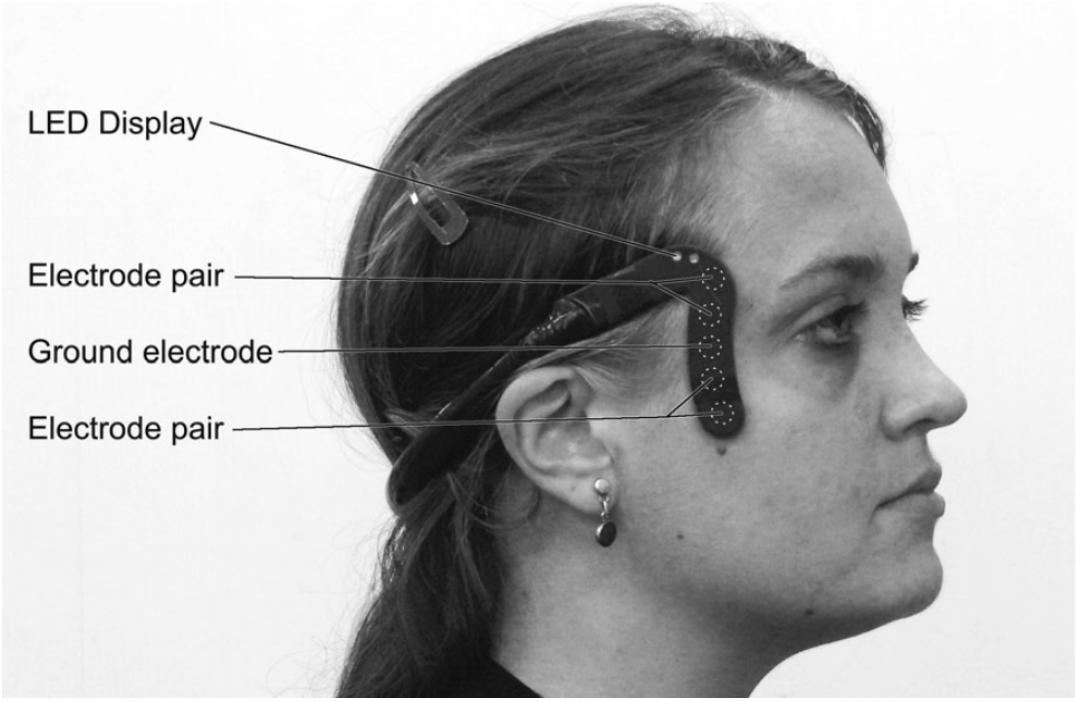
- 同じ目的（介護）をしているが何が違うのか?
  - 人に何をしたらどういった反応をするか？ **→ スキル = 行動予測問題**
- **より長期的な行動予測（ストーリー性のある行動予測）**

介護の中身をオープンに～ハイテク・理論が現場を変える～
NHKクローズアップ現代
(2016年2月3日)

ASD児・認知症の表情

発達障害児の表情の可視化[Suzuki2015]

- 発達障害（ASD)児：表情が表出しにくい
- 表情筋を検出することで可視化 → 他人に「笑い」を表出できる
- **母親はデバイスが無くても分かっている**

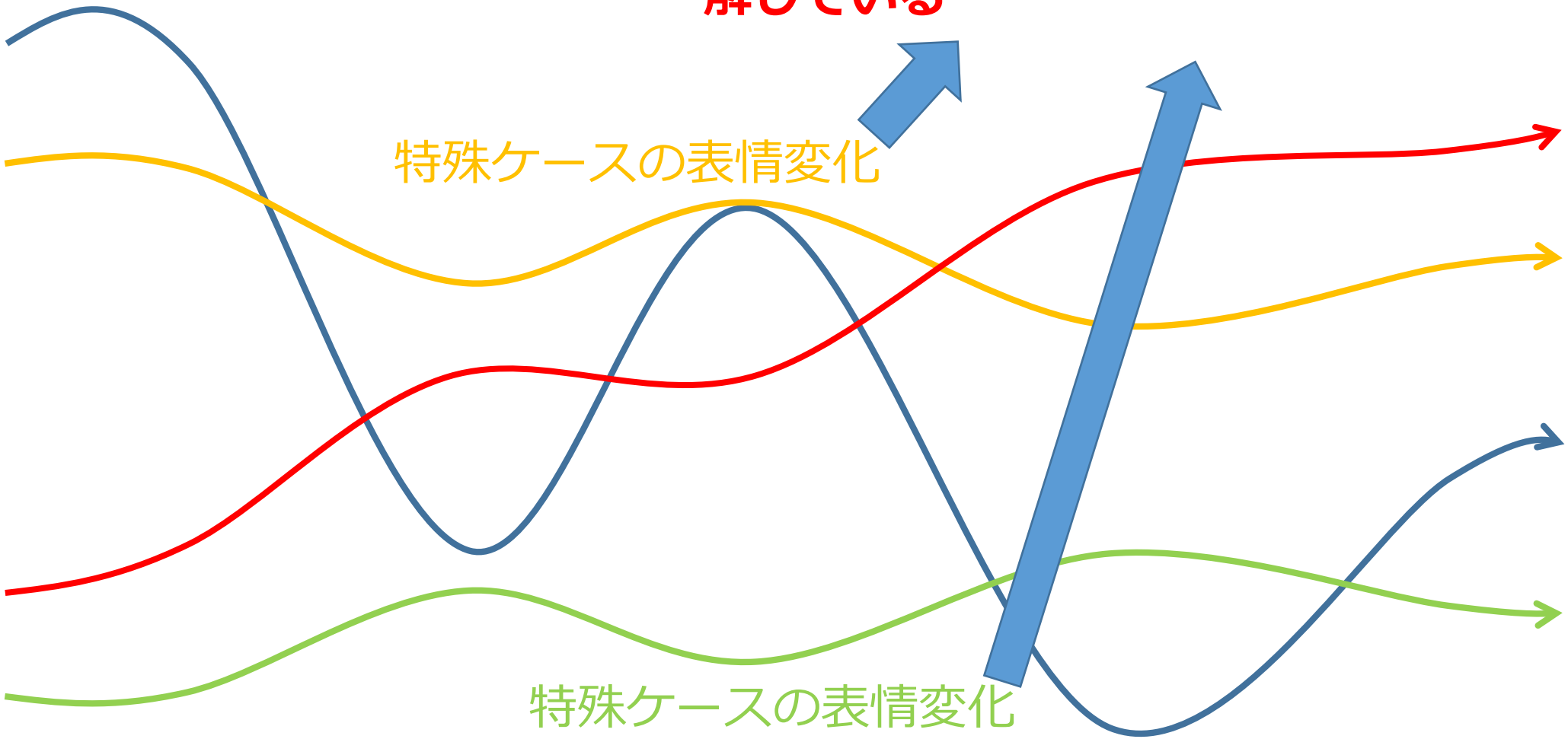


認知症介護ケアの可視化[Nakazawa2016]

- 表情が表出しにくい
- **看護者・家族はデバイスが無くても長期的観察から表情（機嫌・気分）が分かっている**

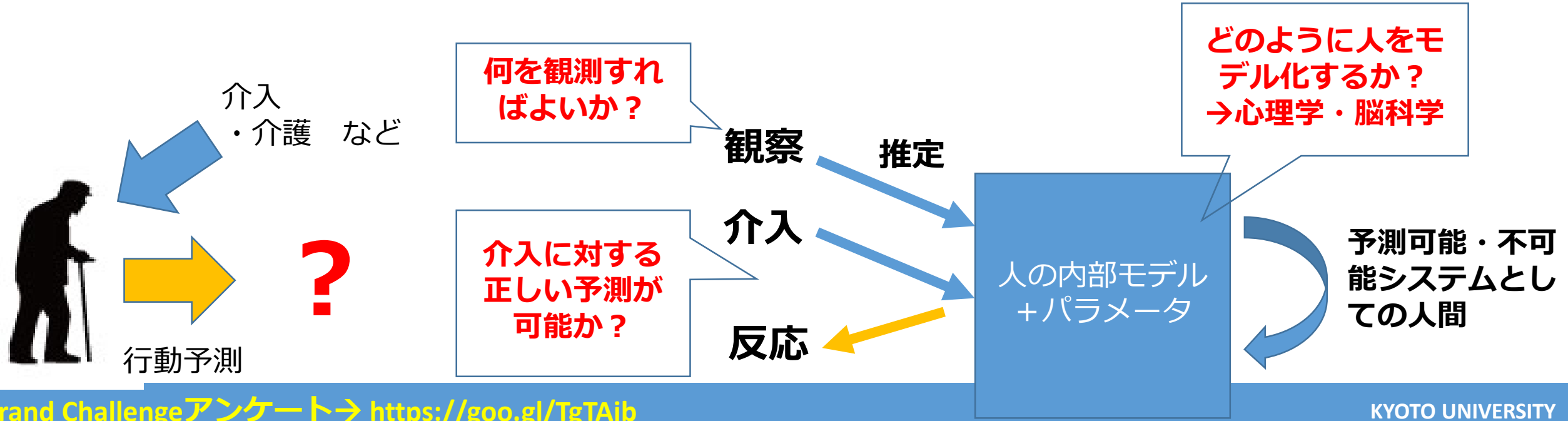
表情の変化認識

・・とはいえ、我々は「この人はこういう人」という形で「表情バイアス」を理解している



脱・統計的機械学習＝個別化の時代へ

- 今までの機械学習(End-to-End学習)
  人の行動を**統計的な傾向**から「多くの人がどう行動するか？」を予測する（最尤推定）
- 今後
  1．個々の人にパーソナライズドされたパターンを認識できる・予測する。
  ⇒この人たちの行動を予測するためには、どのような情報が必要で、どのような考え方（思考モデル）を持っているか？⇒「**予測可能システム**」としての人間
  2．より長期的な行動を予測する（シナリオも予測できる←よりパーソナルな情報も含む）
- **研究課題：人の内部モデルをどう作るか？　そのモデルパラメータを推定するのは何を測ればよいのか？**
  画像だけで足りるのか？　画像以外で何を得られればよいのか？



マルチメディア領域のグランドチャレンジ
– PRMU第二期グランドチャレンジ –

Toshihiko Yamasaki

Associate Professor,
Department of Information and Communication Engineering,
Graduate School of Information Science and Technology,
The University of Tokyo

https://goo.gl/TgTAjb

東京大学
THE UNIVERSITY OF TOKYO

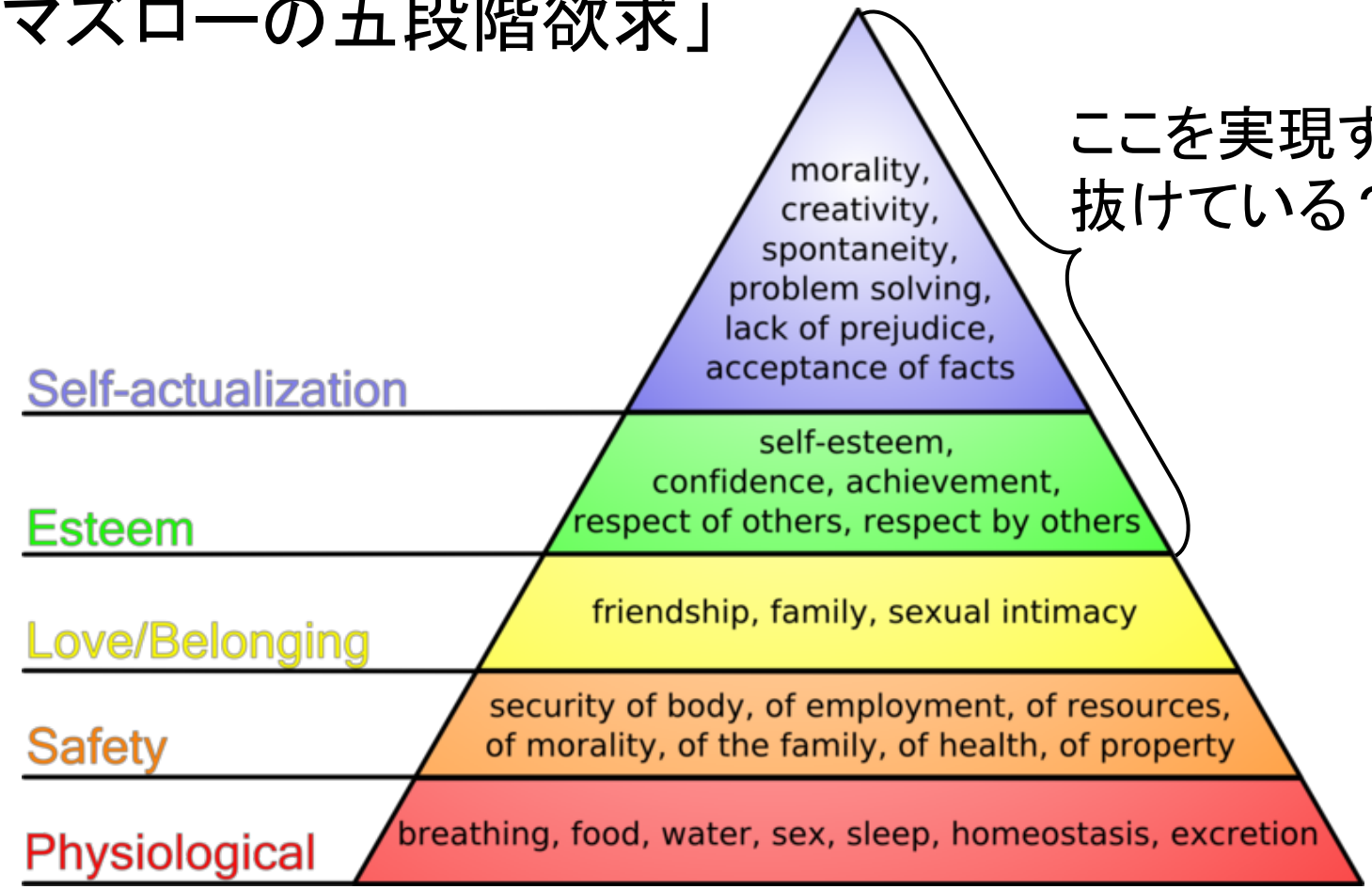
次のグランドチャレンジ

- 「安心・安全」のためのパターン認識から
  「幸福・満足」のためのパターン認識へ
- Whatの認識からHow, Whyの認識へ
- 強い産学連携

https://goo.gl/TgTAjb

「幸福・満足」のためのパターン認識

「マズローの五段階欲求」



ここを実現する技術が抜けている？

https://goo.gl/TgTAjb https://en.wikipedia.org/wiki/Maslow%27s_hierarchy_of_needs

Whatの認識からHow, Whyの認識へ

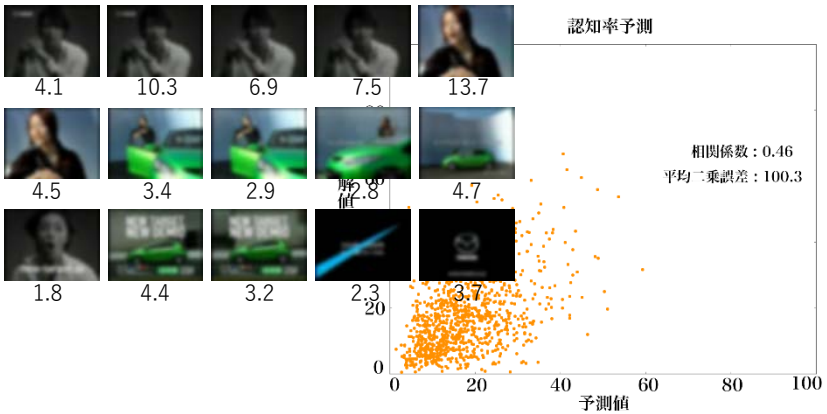
介護シーン（実問題）

- 同じ目的（介護）をしているが何が違うのか?
  - 人に何をしたらどういった反応をするか？→ スキル ＝ 行動予測問題

介護の中身をオープンに ～
ハイテク・理論が現場を変
える～
NHKクローズアップ現代
(2016年2月3日)

@中澤先生

http://roanoke.edu/Images%5Cmcsp%5Cted.jpg



https://goo.gl/TgTAjb

強い産学連携

Exhibit E2

The MGI Industry Digitization Index

2015 or latest available data

Relatively low digitization — Relatively high digitization

● Digital leaders within relatively undigitized sectors

AdTech → Media

FinTech → Finance and insurance

ReTech → Real estate

EdTech → Education

HealthTech → Health care

Omotenashi → Hospitality

AgriTech → Agriculture and hunting

Sector	Overall digitization[1]	Assets: Digital spending	Assets: Digital asset stock	Usage: Transactions	Usage: Interactions	Usage: Business processes	Usage: Market making	Labor: Digital spending on workers	Labor: Digital capital deepening	Labor: Digitization of work	GDP share %	Employment share %	Productivity growth, 2005–14[2] %
ICT											5	3	4.6
Media											2	1	3.6
Professional services											9	6	0.3
Finance and insurance											8	4	1.6
Wholesale trade											5	4	0.2
Advanced manufacturing											3	2	2.6
Oil and gas											2	0.1	2.9
Utilities											2	0.4	1.3
Chemicals and pharmaceuticals											2	1	1.8
Basic goods manufacturing											5	5	1.2
Mining											1	0.4	0.5
Real estate	●										5	1	2.3
Transportation and warehousing	●										3	3	1.4
Education	●										2	2	-0.5
Retail trade	●										5	11	-1.1
Entertainment and recreation											1	1	0.9
Personal and local services											6	11	0.5
Government	●										16	15	0.2
Health care											10	13	-0.1
Hospitality	●										4	8	-0.9
Construction											3	5	-1.4
Agriculture and hunting											1	1	-0.9

www.mckinsey.com

次のグランドチャレンジ

- 「安心・安全」のためのパターン認識から
  「幸福・満足」のためのパターン認識へ
- Whatの認識からHow, Whyの認識へ
- 強い産学連携

https://goo.gl/TgTAjb

ご意見募集！
よりよいものにするためにご協力お願いします！

goo.gl/TgTAjb

提案・感想・文句，
なんでもOK

匿名OK!