

データリテラシー

自由席です

データリテラシー

第1回 社会におけるデータ・AI利活用

メディア情報コース
平居 悠（ひらい ゆたか）

自己紹介

隣の人に情報リテラシーの総合課題で作ったwebページを見せながら自己紹介しよう。

データサイエンス・AI教育プログラム

データサイエンスやAIに関する基礎的な知識と技術、及びその知識や技術を他の科目や学修で応用する能力

「データサイエンス・AI教育プログラム」の流れ (※8単位以上修得)



MDASH Literacy
数理・データサイエンス・AI
教育プログラム認定制度
リテラシーレベル



MDASH Advanced Literacy
数理・データサイエンス・AI
教育プログラム認定制度
応用基礎レベル

1 年次秋学期

データリテラシー



2 年次春学期

基礎プログラミングⅠ



2 年次秋学期

基礎プログラミングⅡ



選択必修科目：「AIと社会」「セキュリティ論」「日経講座：デジタル社会論」

選択科目：「数学a」「数学b」「統計学a」「統計学b」「データサイエンス入門b」「数値情報処理b」「画像情報処理」「データ分析手法b」「経営工学a」「経営工学b」「データベース論」「データベース演習」「データ構造とアルゴリズム」「応用プログラミング」「機械学習入門a」「機械学習入門b」

授業概要

2年次以降の講義や演習で必要とされるデータを扱うための基礎的な力を身につける。

本講義の目的

数理・データサイエンス・AIの基礎的素養を身につけ、日常生活、卒業後の仕事の場で使いこなすことができるようになる。

到達目標

- データサイエンス・AIが社会でどのように活用されているのかを理解する。
- データを適切に読み解き理解する力をつける。
- データを適切な可視化手法で他者に説明できるようになる。
- 表計算ソフトを使ってデータを集計・加工できるようになる。

講義について

- 大量のデータの中に存在する法則性を扱う**統計学**の知識
- 実際のデータ分析に必要な**表計算ソフト**やプログラムを扱うスキル

統計的方法

統計学に従ってデータを処理して分析する方法

- **記述統計**：対象となる集団の全個体を観測して得られたデータを整理・要約し、データの規則性から法則を調べる。
- **推測統計**：対象となる集団の一部を取り出して観測し、得られたデータから確率論を使って元の集団の持つ特性を推測する。

授業計画

第1回	社会におけるデータ・AI利活用
第2回	表計算ソフトの基本操作
第3回	セルの参照と集計
第4回	度数分布表とヒストグラム
第5回	代表値と散らばりの尺度
第6回	データのグラフによる表現とデータ比較
第7回	データの分布と標本抽出
第8回	推定
第9回	分割表とクロス集計
第10回	2変量データと相関
第11回	時系列データ
第12回	仮説検定
第13回	回帰分析

授業計画

第1回	社会におけるデータ・AI利活用
第2回	表計算ソフトの基本操作
第3回	セルの参照と集計
第4回	度数分布表とヒストグラム
第5回	代表値と散らばりの尺度
第6回	データのグラフによる表現とデータ比較
第7回	データの分布と標本抽出
第8回	推定
第9回	分割表とクロス集計
第10回	2変量データと相関
第11回	時系列データ
第12回	仮説検定
第13回	回帰分析

表計算ソフト

授業計画

第1回	社会におけるデータ・AI利活用
第2回	表計算ソフトの基本操作
第3回	セルの参照と集計
第4回	度数分布表とヒストグラム
第5回	代表値と散らばりの尺度
第6回	データのグラフによる表現とデータ比較
第7回	データの分布と標本抽出
第8回	推定
第9回	分割表とクロス集計
第10回	2変量データと相関
第11回	時系列データ
第12回	仮説検定
第13回	回帰分析

記述統計

授業計画

第1回	社会におけるデータ・AI利活用
第2回	表計算ソフトの基本操作
第3回	セルの参照と集計
第4回	度数分布表とヒストグラム
第5回	代表値と散らばりの尺度
第6回	データのグラフによる表現とデータ比較
第7回	データの分布と標本抽出
第8回	推定
第9回	分割表とクロス集計
第10回	2変量データと相関
第11回	時系列データ
第12回	仮説検定
第13回	回帰分析

推測統計

成績評価

平常課題（4割）

総合課題（4割）

タイピング（2割）

無断欠席4回以上で不可

成績評価

各課題の評価割合で重み付けした合計点で成績評価

評価	合計点
秀 (S)	90~100
優 (A)	80~89
良 (B)	70~79
可 (C)	60~69
不可 (D)	~59

平常課題（4割）

毎回の提出課題内容で評価

減点方式

各回**8点**満点、1ミスで**6点**、2ミス以上で**5点**、
期限遅れは**4点以下**

総合課題（4割、加点方式、40点満点）

『知の編集力ハンドブック』 p. 103 「基礎演習レポートルーブリック」に基づき評価する。

基礎演習 レポート ルーブリック → 基礎演習終了時点での目標「4」

	定義	5	4	3	2	1
文献の 選択	信頼性が低い、または信頼性の判断が難しい情報源からの引用を避ける。孫引きをしない。自分の主張を支える上で必要かつ十分な文献を引用する。	適切な文献を活用し、自分の意見を支える客観的なデータや事実を適切に示している。	文献の信頼性や孫引きには問題がないが、引用する文献に不足があり、主張に論理の飛躍がある。	信頼性が低い情報源の利用や孫引きが見られる。	信頼性の低い情報源の利用や孫引きが複数ある。	文献からの引用がない。
引用	引用部と自分の書いた文章の区別に曖昧さがなく、引用マークを付けている。	区別に曖昧さがなく、レポートの構成上、必要以上の引用もない。	引用マークをつけ区別しており、区別が曖昧な箇所がない。	引用マークをつけ区別しているが、区別が曖昧な箇所もある。	区別が全体的に不明確で剽窃が疑われかねない。	文献からの引用がない。
文体	「ですます調」と「である調」の混在、話し言葉、体言止め、誤字脱字、ひらがなの多用、文末表現の重複がない。	適切。	問題がほとんどない（1～2か所）。	問題が見られる（3～4か所）。	問題が多い（5～7か所）。	問題が多すぎる（それ以上）。
構成の 文	一文の長さ、主語と述語の関係、句読点の打ち方、段落の使い方に問題がない。	適切。	問題がほとんどない（1～2か所）。	問題が見られる（3～4か所）。	問題が多い（5～7か所）。	問題が多すぎる（それ以上）。
自分の 意見	自分の意見が述べられており、引用部と比較した場合に量的、質的に主となっている。	自分の意見が述べられており、量的かつ質的に主となっている。	自分の意見が述べられており、量的または質的に主となっている。	自分の意見が述べられているが、量的、質的に従になっている。	自分の意見というより単なる感想になっている。	他の資料の丸写しかそれに近い状態で自分の意見がない。
文献 リスト	文献リストを作成しており、文献の記載漏れや、記載項目の抜け落ちがない。	適切。	文献の記載漏れや記載項目の抜けがほとんどない（1～2か所）。	文献の記載漏れや記載項目の抜けが見られる（3～4か所）。	文献の記載漏れや記載項目の抜けが多い（5か所以上）。	文献からの引用なし。または引用しているが、文献リストなし。

タイピング（2割）

trr「日本国憲法」2Hランキングスコアで評価

目安評価	スコア
秀（S）	250~
優（A）	200~249
良（B）	150~199
可（C）	100~149
不可（D）	~99

タイピング練習スケジュール

第2回	ホームポジション
第3回	ローマ字
第4回	英語初級
第5回	日本国憲法（trr試験）
第6回	ホームポジション
第7回	ローマ字
第8回	英語初級
第9回	日本国憲法（trr試験）
第10回	ホームポジション
第11回	ローマ字
第12回	英語初級
第13回	日本国憲法（trr試験）

授業ページ

学生専用サーバ→平居→データリテラシー

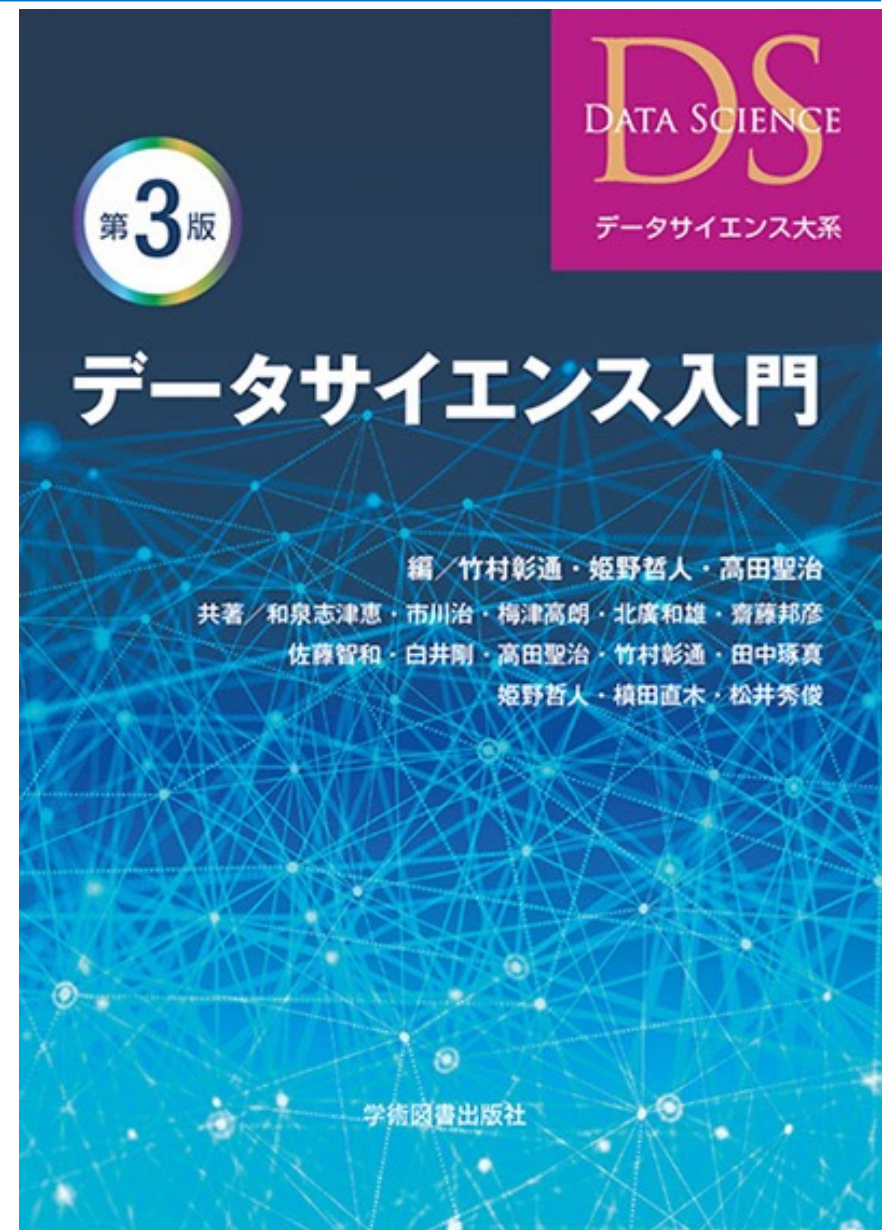
学内：<http://roy.e.koeki-u.ac.jp/>

学外：<https://www.koeki-prj.org/roy/>

データリテラシー（山本先生のページ）

https://www.kitp.org/class/doku.php?id=data_literacy:top

参考書



質問等

- s4「データリテ (2)」 <https://s4.koeki-prj.org/s4-world-2026au.cgi?grp+68>
- メール (yutaka.hirai@koeki-u.ac.jp)
- オフィスアワー（H2研究室木曜2限）

授業計画

第1回	社会におけるデータ・AI利活用
第2回	表計算ソフトの基本操作
第3回	セルの参照と集計
第4回	度数分布表とヒストグラム
第5回	代表値と散らばりの尺度
第6回	データのグラフによる表現とデータ比較
第7回	データの分布
第8回	標本抽出と推定
第9回	分割表とクロス集計
第10回	2変量データと相関
第11回	時系列データ
第12回	仮説検定
第13回	回帰分析

今回の問い

- 現代社会ではどのような変化が起きているのだろうか？
- データ・AIはどこで活用されているのか？
- データ・AI利活用の最新動向はどのようなになっているのか？

今回の目標

データサイエンス・AIが社会でどのように活用されているのかを理解する。

今回学ぶこと

1. 社会で起きている変化
2. 社会で活用されているデータ
3. データ・AIの活用領域
4. データ・AI利活用のための技術
5. データ・AI利活用の現場
6. データ・AI利活用の最新動向

今回の課題

各項目の確認テストに解答する。
解答方法は後述。

今回学ぶこと

1. 社会で起きている変化
2. 社会で活用されているデータ
3. データ・AIの活用領域
4. データ・AI利活用のための技術
5. データ・AI利活用の現場
6. データ・AI利活用の最新動向

データとは

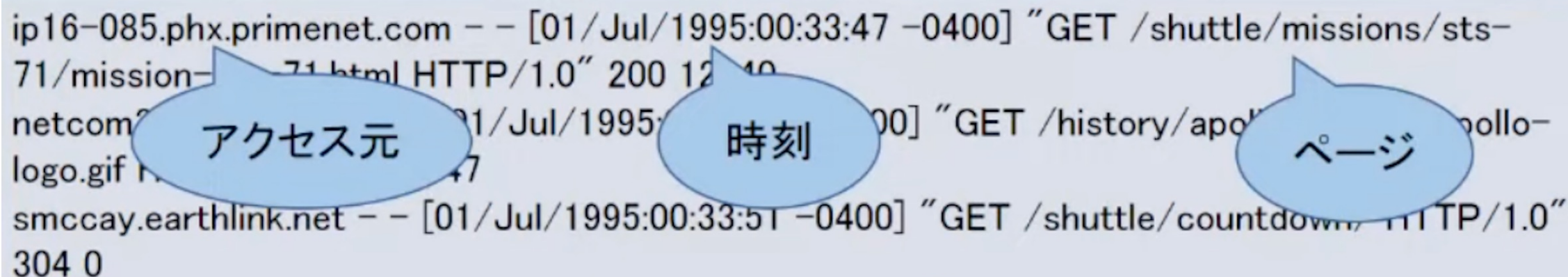
色々な事柄について、主に数値の形で記録したもの

- 身体：身長、体重、体温
- 経済：物価、国内総生産
- 自然：気温、雨量
- アンケート調査：性別、年齢、好みなど

データとは

最近はテキスト、画像など、数値以外のデータが増えている

Webサーバのログファイル



```
ip16-085.phx.primenet.com -- [01/Jul/1995:00:33:47 -0400] "GET /shuttle/missions/sts-71/mission-71.html HTTP/1.0" 200 12340
netcom2 [01/Jul/1995:00:33:47 -0400] "GET /history/apollo-13/logo.gif HTTP/1.0" 200 47
smccay.earthlink.net -- [01/Jul/1995:00:33:51 -0400] "GET /shuttle/countdown HTTP/1.0" 304 0
```

アクセス元

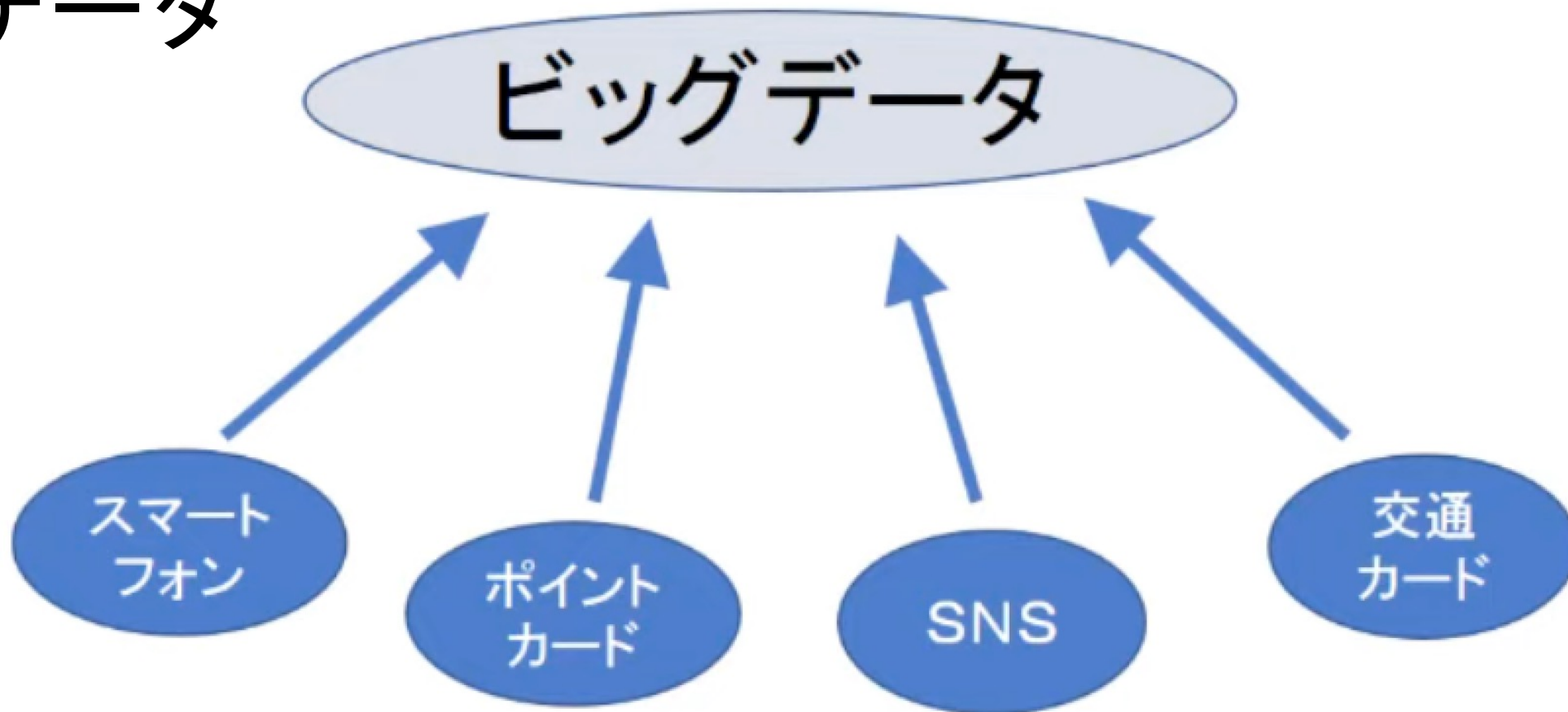
時刻

ページ

<https://ita.ee.lbl.gov/html/contrib/NASA-HTTP.html>

ビッグデータ

ネットワーク上に蓄積された大量かつ多様なデータ



ビッグデータ時代の象徴： スマートフォンの普及

iPhone: 2007年Apple社が
アメリカで発売



現在のスマホは30年前のスーパーコンピュータ

ASCI Red
(1996年)

1秒で1兆回の
計算

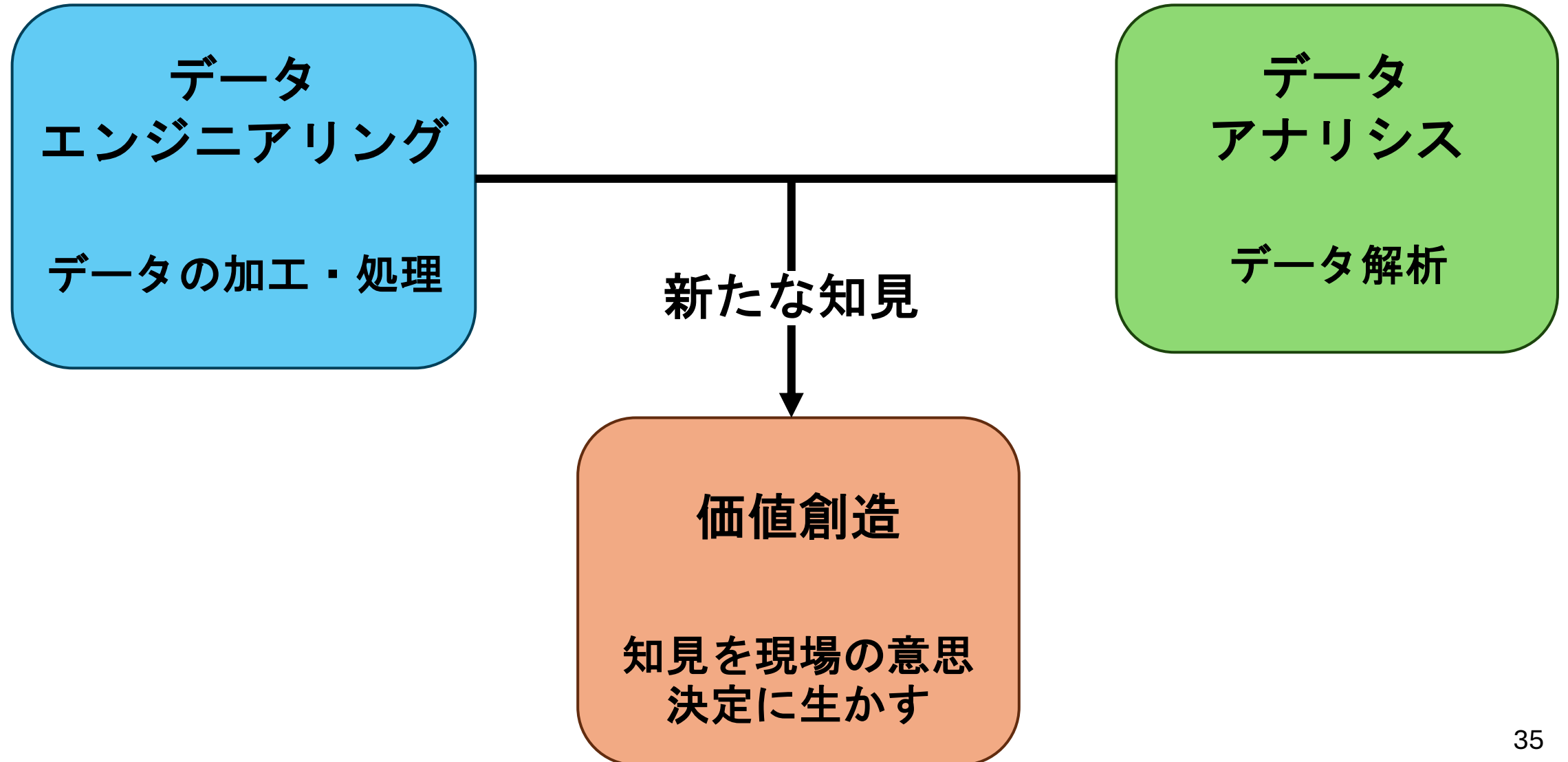


ビッグデータ＝新たな資源

21世紀の石油

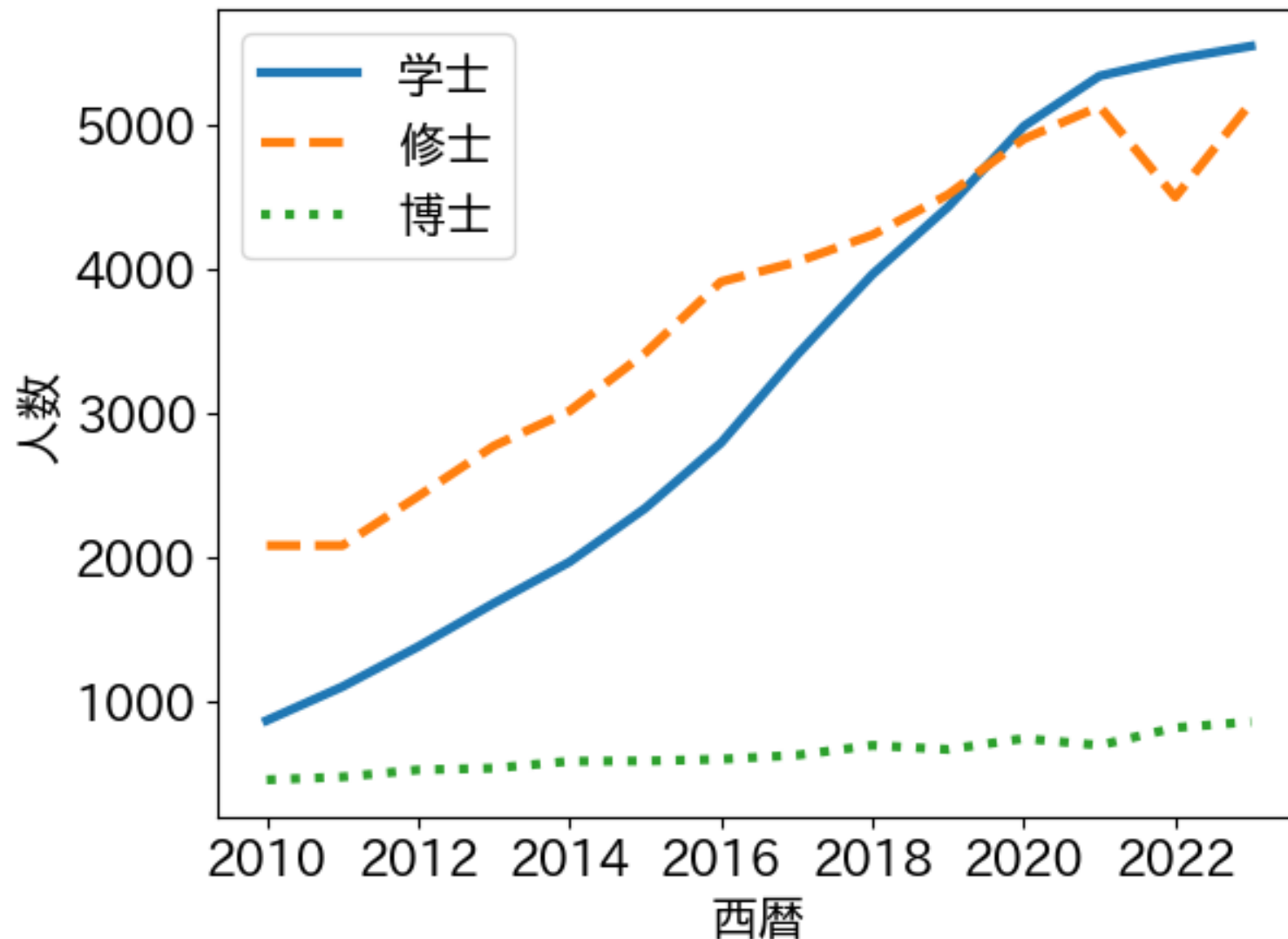
- この新たな資源を生かしたものが競争的優位に立つ
- データとそれを生かす技術の双方が必要
- 日本ではデータを分析する人材が不足

データサイエンスの3要素



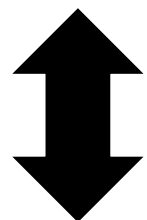
アメリカのデータサイエンティスト育成

アメリカにおける統計学及び生物統計学の学位授与数

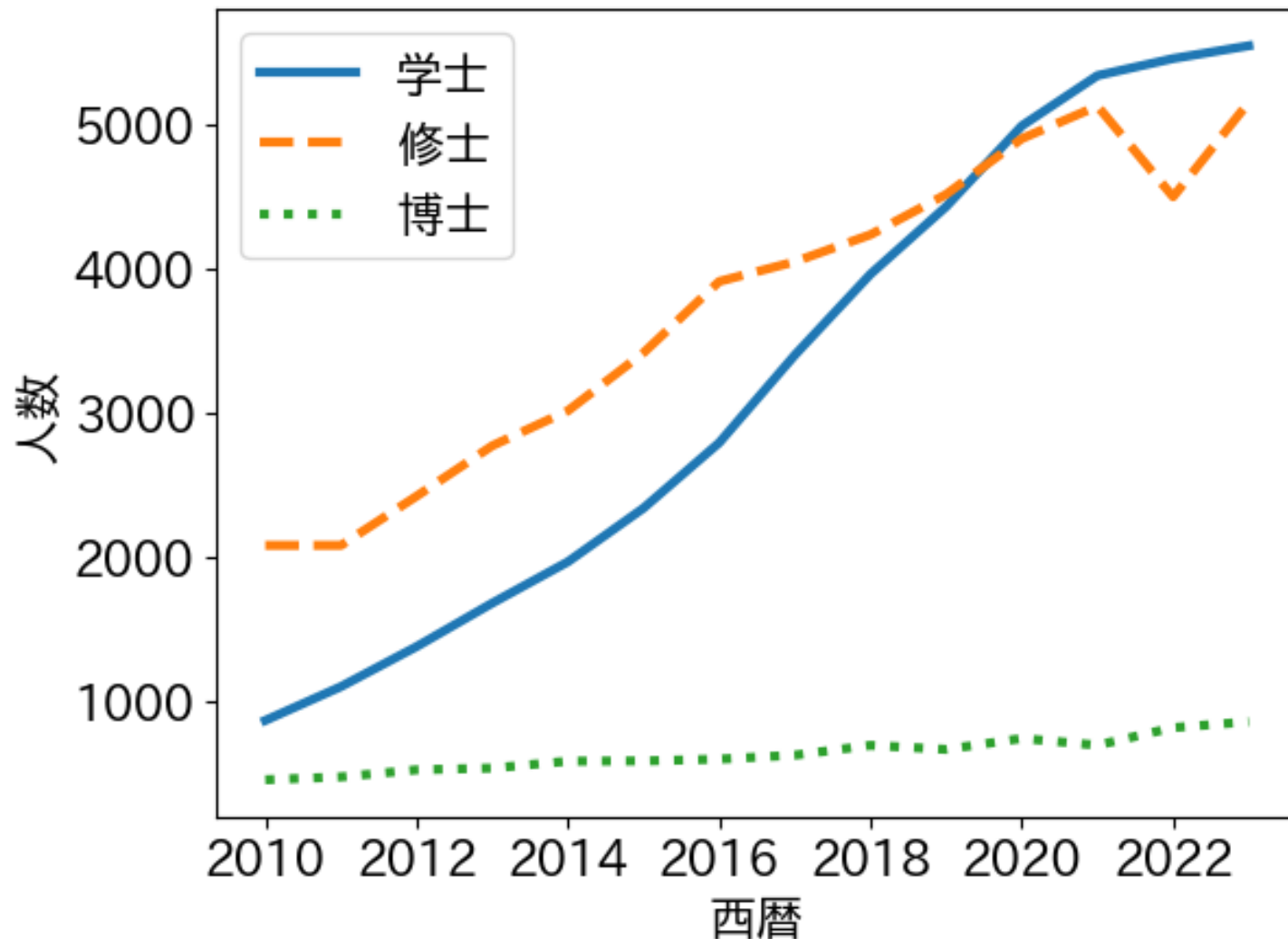


データサイエンティスト育成

アメリカにおける
統計学学位授与数
は上昇傾向にある



日本ではデータサイエンティスト育成は課題となっている



確認テスト 1

ターミナルで

quiz data-1

と入力。全問正解するまで解答を続けよう！

今日の問い

- 現代社会ではどのような変化が起きているのだろうか？
 - ビッグデータが生み出され、データサイエンティストの重要性が急速に上がっている。
- データ・AIはどこで活用されているのか？
- データ・AI利活用の最新動向はどのようなになっているのか？

今回学ぶこと

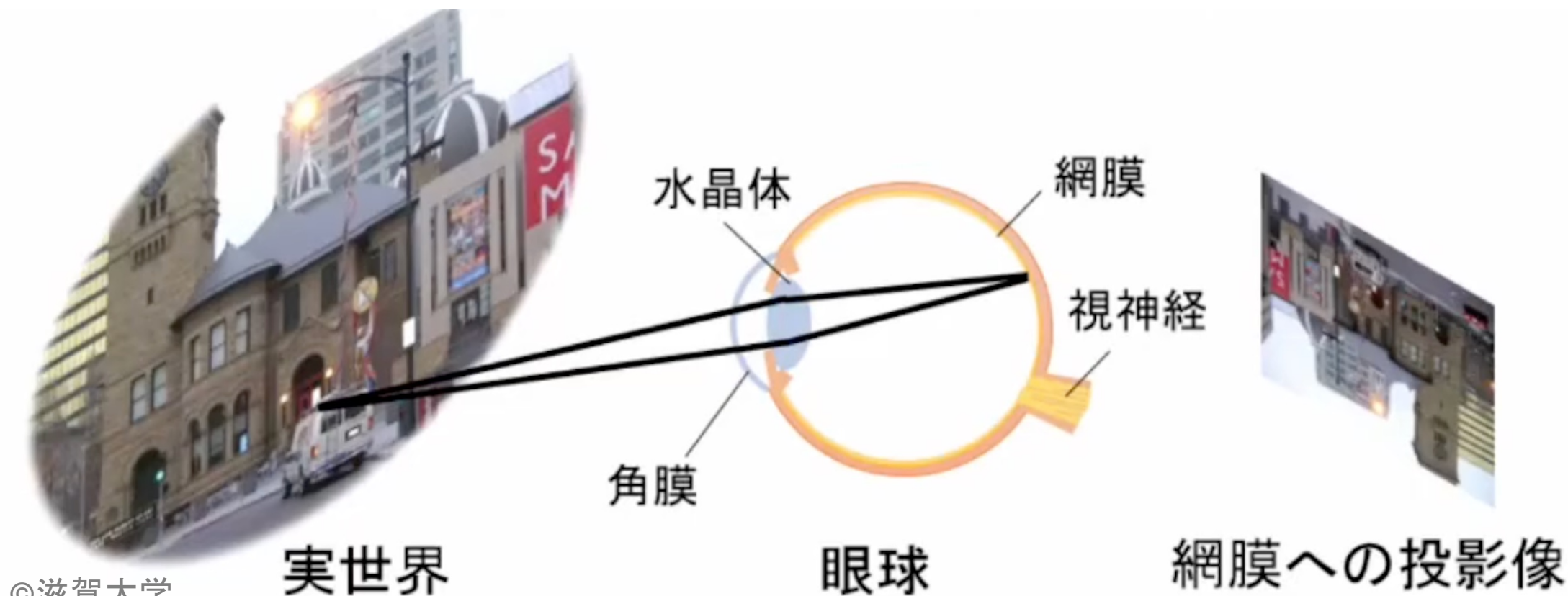
1. 社会で起きている変化
- 2. 社会で活用されているデータ**
3. データ・AIの活用領域
4. データ・AI利活用のための技術
5. データ・AI利活用の現場
6. データ・AI利活用の最新動向

データサイエンスと画像処理技術

デジタル画像の構成

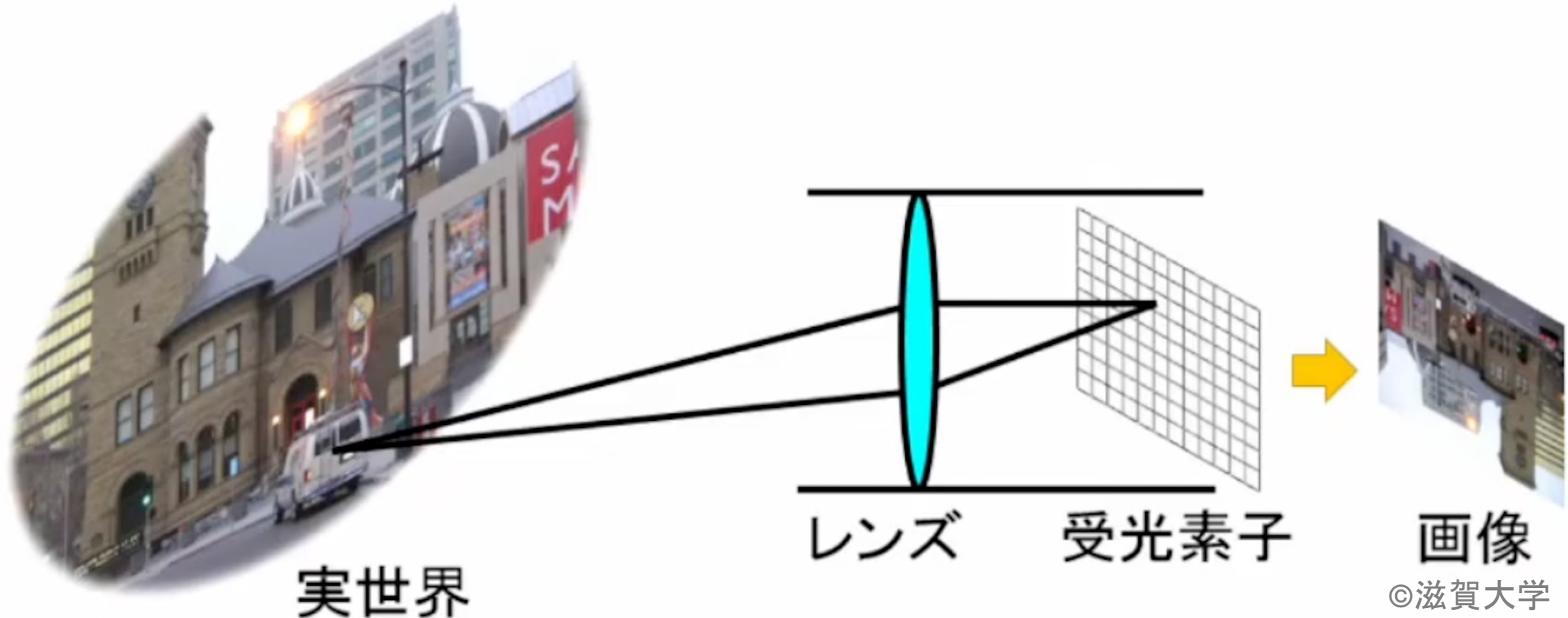
人間の視覚

- 網膜が赤・緑・青の光の量を測定
- 視神経を通して脳に光の量を伝達



デジタルカメラによる撮影

- 網膜の代わりに光の量を測るセンサ（受光素子）を利用
- 受光素子はレンズ裏側に規則的に並べられ、光の量を測定



画素

- 画像は画素と呼ばれるタイルの集まり
- 解像度：一枚の画像を作る画素の数
- 解像度が高いほど細かな形を見分けられる

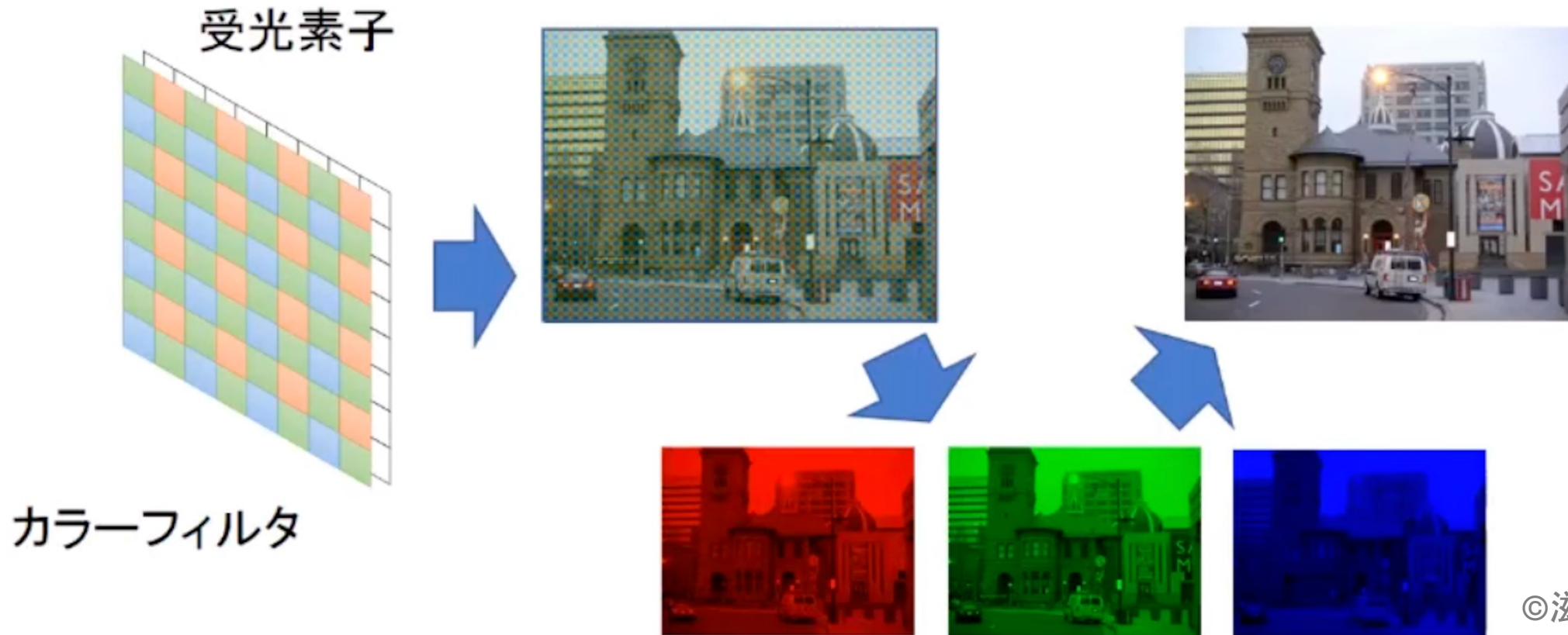
画素が多い = 解像度が高い

画素が少ない = 解像度が低い



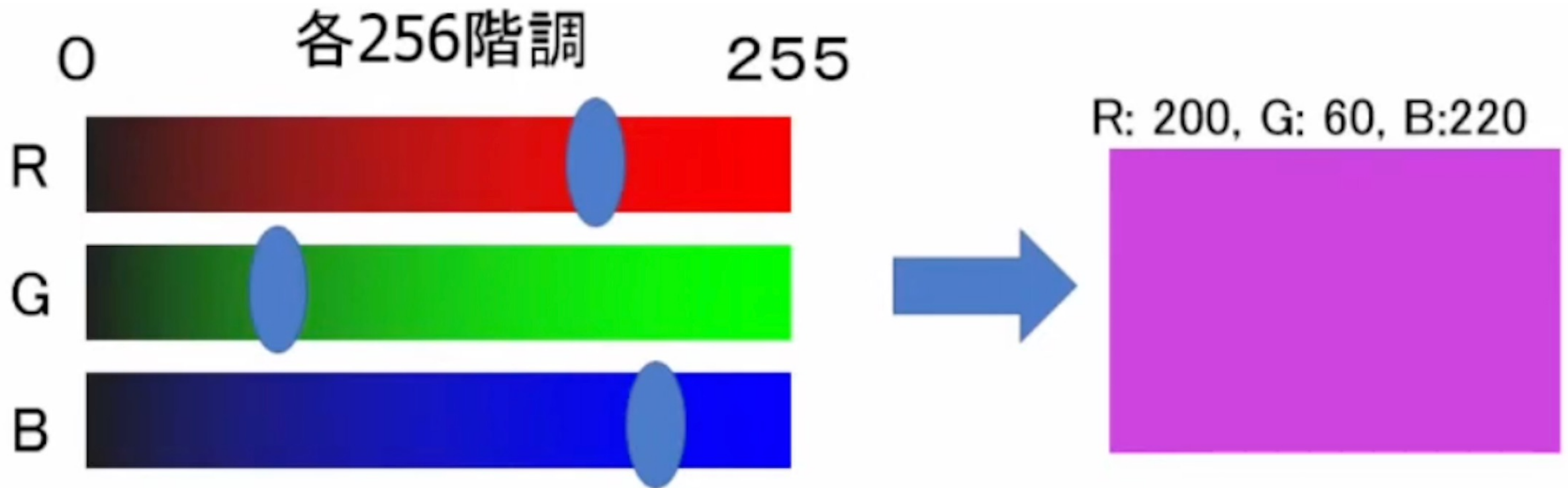
カラー画像の撮影

- 受光素子の前に色のついたフィルターを設置
- 各受光素子が赤・緑・青の光の強さを測る
- 各色の光の強さを画素ごとに合成し、カラー画像を作る



画素の色の表現

- 人間の目の中の細胞も赤・緑・青に反応する細胞に分かれる
- この光の三原色を混ぜ合わせることで全ての色を表現できる



画像表現—データ配列

デジタル化された画像は画素ごとに明るさが格納された数値の配列として保存されている

1画素分の情報

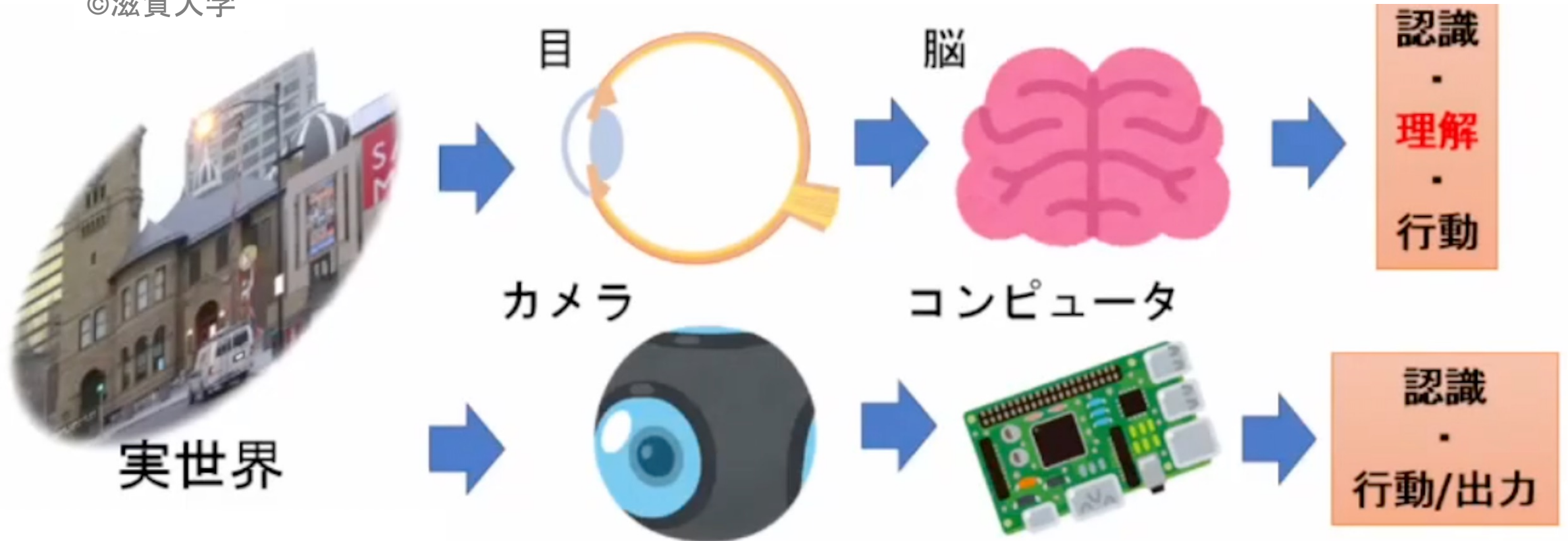


25	25	30	29	85	5		5
29	25	25	29	88	5		5
25	25	25	30	29	3		5
3	25	30	25	25	25		85
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
5	3	5	5	29	25		88

画像解析による人間の視覚・認識機能の模倣

コンピュータビジョン：画像を解析することでコンピュータに様々な機能を実現させる技術

©滋賀大学



画像解析でできること

カメラで撮影された画像から色々な情報を取り出す

- 人間の「位置、動き、視線、人数、年齢、表情、感情」
- 物体の「位置、形、動き、数、カテゴリ、質感」
- シーンの「場所、形状、照明条件、天候」など

画像解析でできること

取り出された情報から様々な機能を実現する

- 指紋認識、顔認識、人数計測
- 測量・計測、景観シミュレーション
- 自動運転、安全運転支援
- 医用画像診断、生活支援、見守り、介護

確認テスト 2

ターミナルで

quiz data-2

と入力。全問正解するまで解答を続けよう！

今日学ぶこと

1. 社会で起きている変化
2. 社会で活用されているデータ
- 3. データ・AIの活用領域**
4. データ・AI利活用のための技術
5. データ・AI利活用の現場
6. データ・AI利活用の最新動向

機械学習の先進的な事例

企業分析

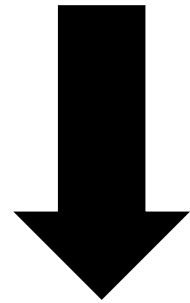
マーケティング

どのような価値を提供すればターゲット市場のニーズを満たせるかを探り、その価値を生み出し、顧客にどどけ、そこから利益を上げること

フィリップ・コトラー『コトラーのマーケティング講座』

マーケティング

定量的にお金の流れを評価



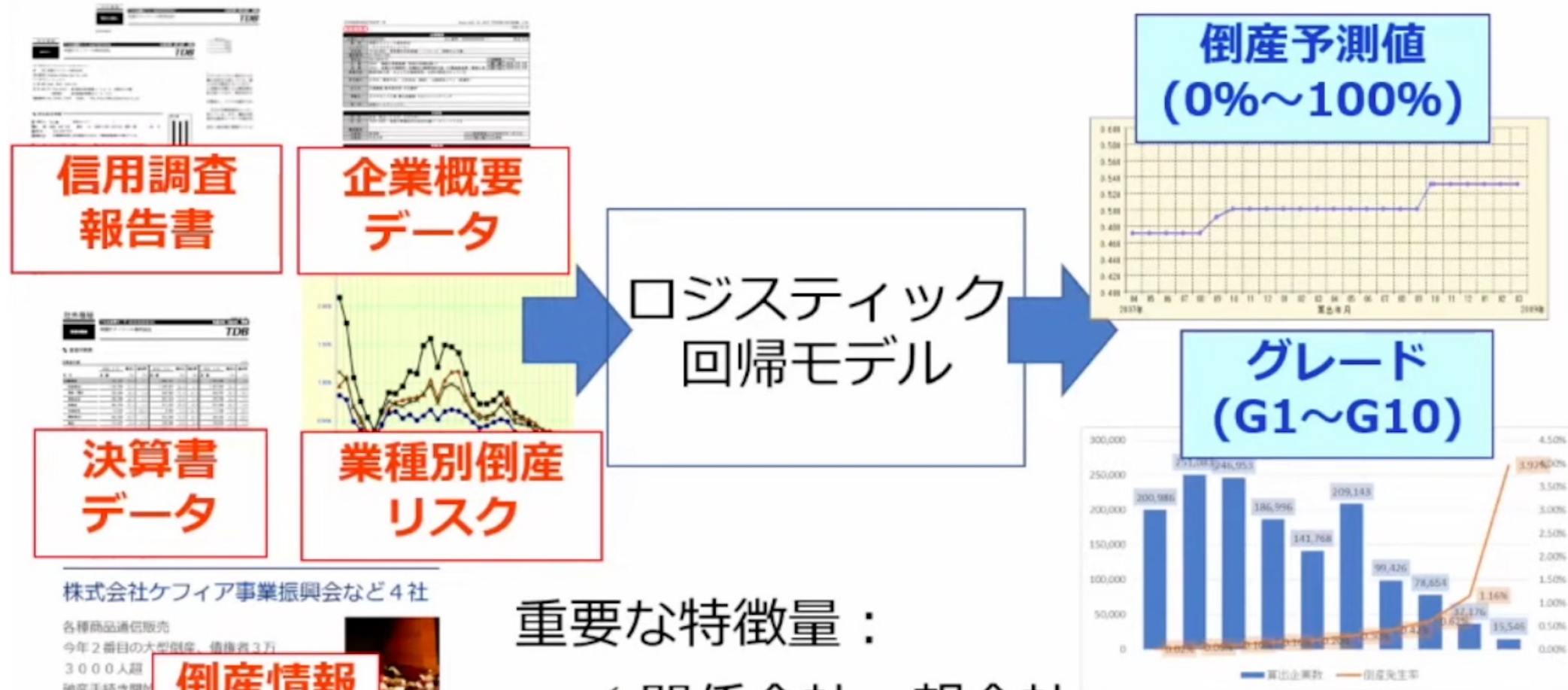
- 損失を最小限に抑えられる
- 利益を最大限大きくできる

機械学習と信用リスク

企業は、損失を最小限にすることを目的として、取引先の信用リスクを評価

- 販売先企業が倒産してしまうと、代金回収ができない（貸し倒れ）
- 一般的には、信用調査会社から、新規取引先企業の貸し倒れの危険度（信用リスク）の評価を得て、取引継続の判断を行う。

事例 倒産予測モデル



重要な特徴量：

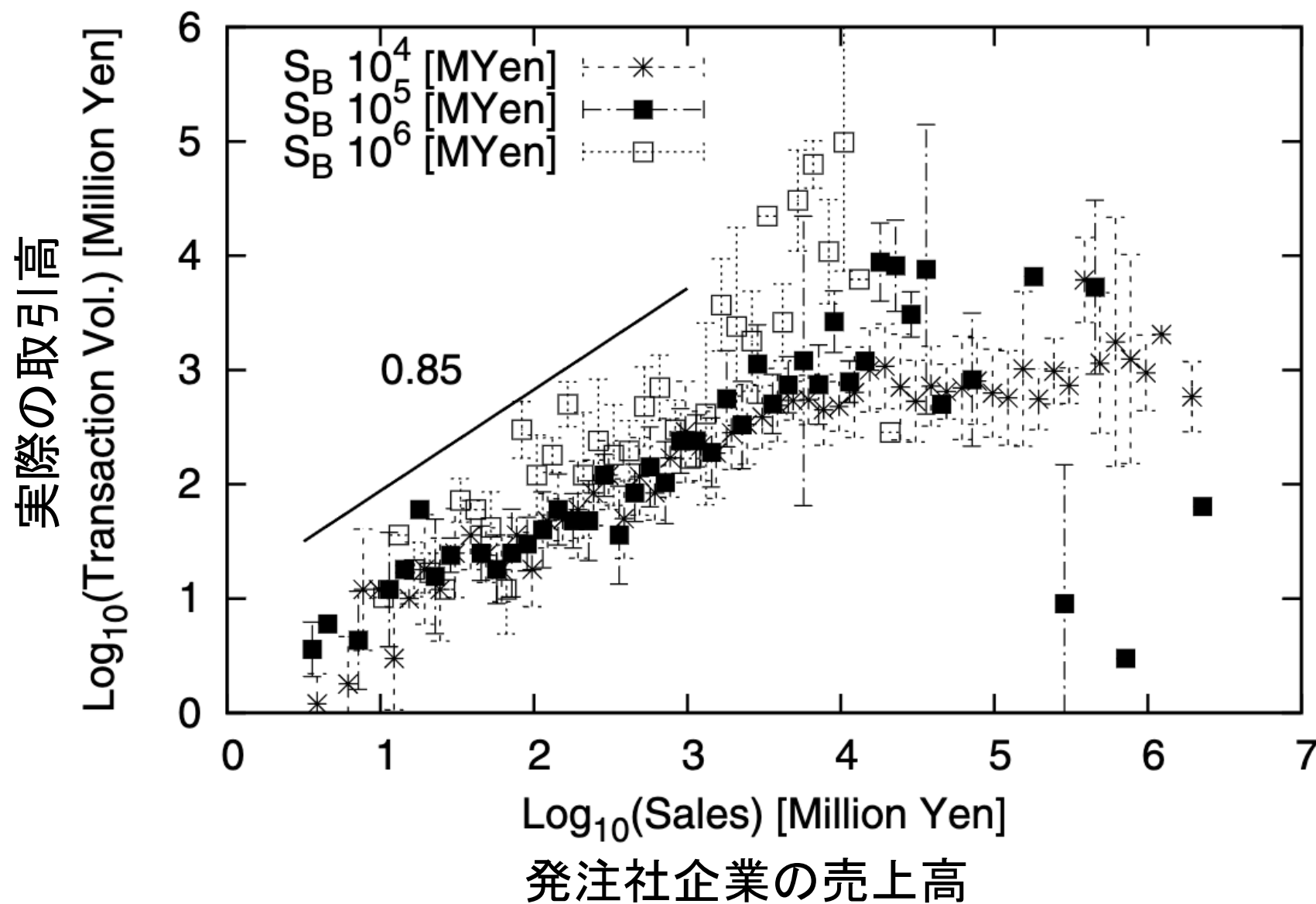
- ✓ 関係会社・親会社
- ✓ 自己資本
- ✓ 業種別倒産リスク
- ✓ 代表者特徴

機械学習と地域活性化

自治体は、地域の利益最大化のために、地域の中で中核となる企業の支援を行う

- 地域の中で取引が集中しており、地域外とも取引を行っている企業（コネクターハブ企業）を、企業間取引データから選定
- 地域での集中度合いを定量的に評価するため、企業間での取引高を推定

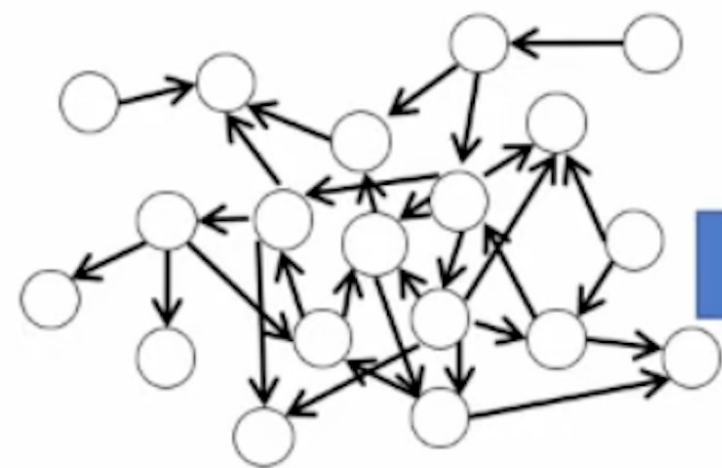
事例 売上高と取引高の関係



売上高と取引高は比例関係にある

事例 取引高推計モデル

企業間取引
ネットワーク

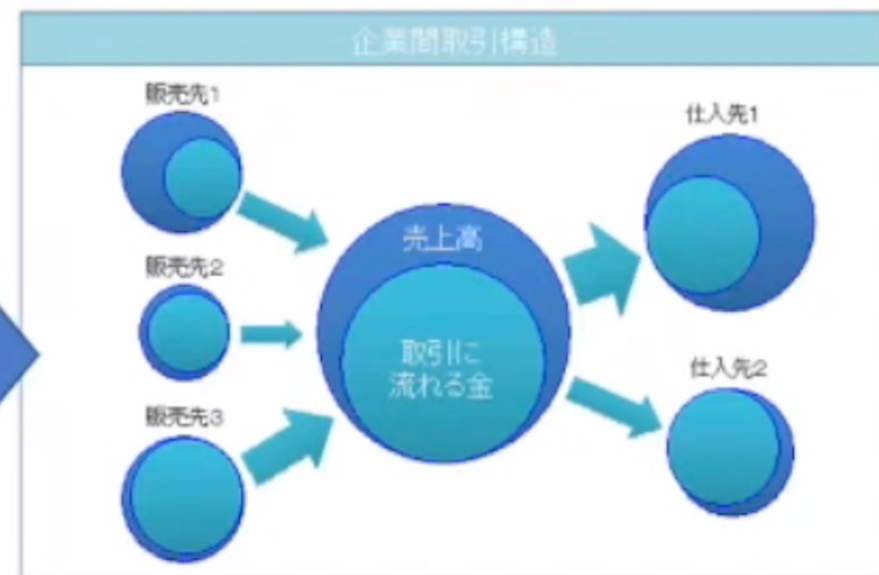


重力モデル
・パラメータ
最適化

○ : 企業
→ : お金の流れ

特徴量 :

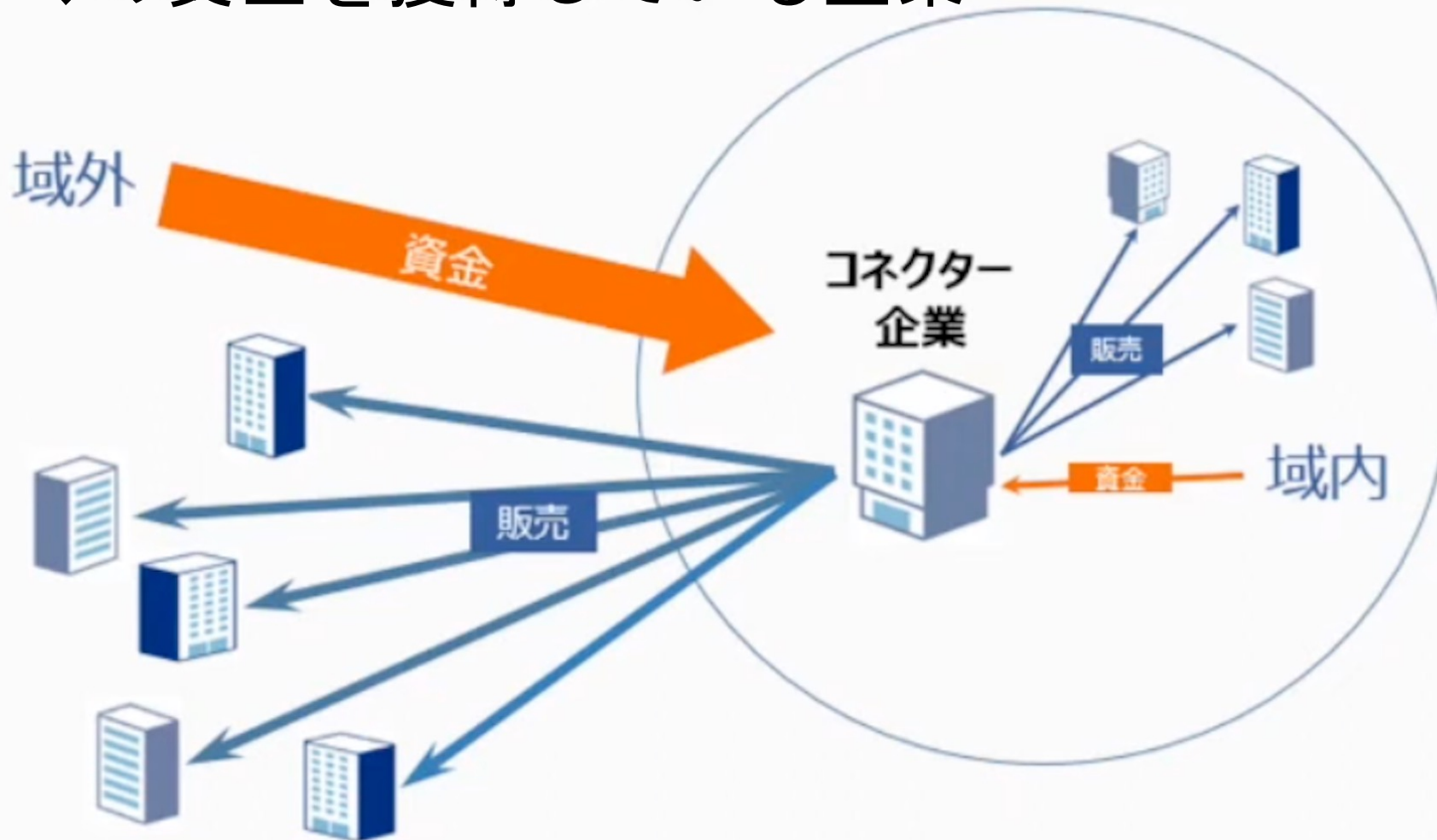
- ✓ 企業規模
- ✓ 取引本数



➡ : 企業間取引高

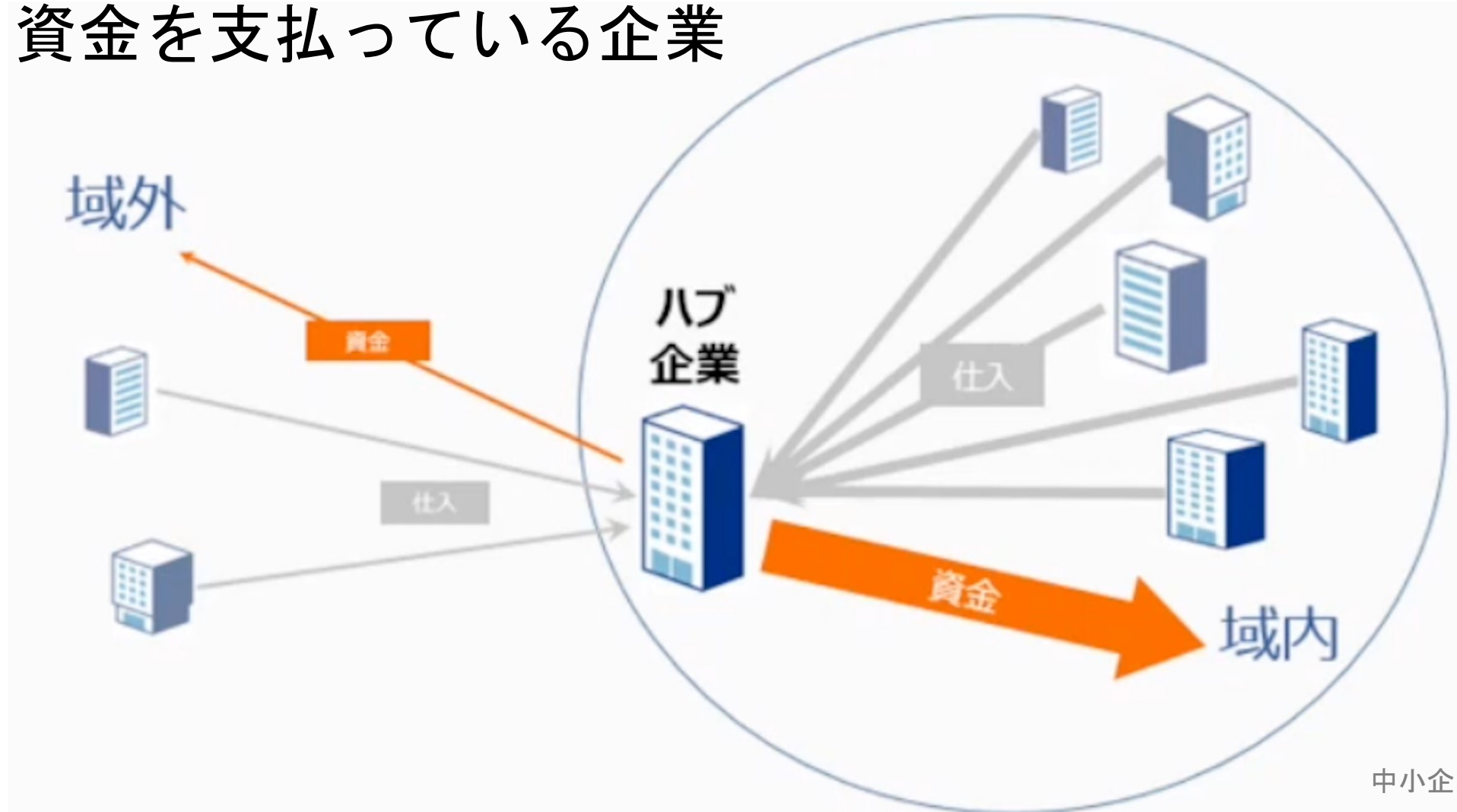
事例 地域の中核となる企業の選定指標

コネクター企業：販売によって域外から域内以上により多くの資金を獲得している企業



事例 地域の中核となる企業の選定指標

ハブ企業：仕入れによって域内に域外以上により多くの資金を支払っている企業



確認テスト 3

ターミナルで

quiz data-3

と入力。全問正解するまで解答を続けよう！

今回の問い

- 現代社会ではどのような変化が起きているのだろうか？
 - ビッグデータが生み出され、データサイエンティストの重要性が急速に上がっている。
- データ・AIはどこで活用されているのか？
 - 画像解析、企業分析などで活用されている。
- データ・AI利活用の最新動向はどのようなになっているのか？

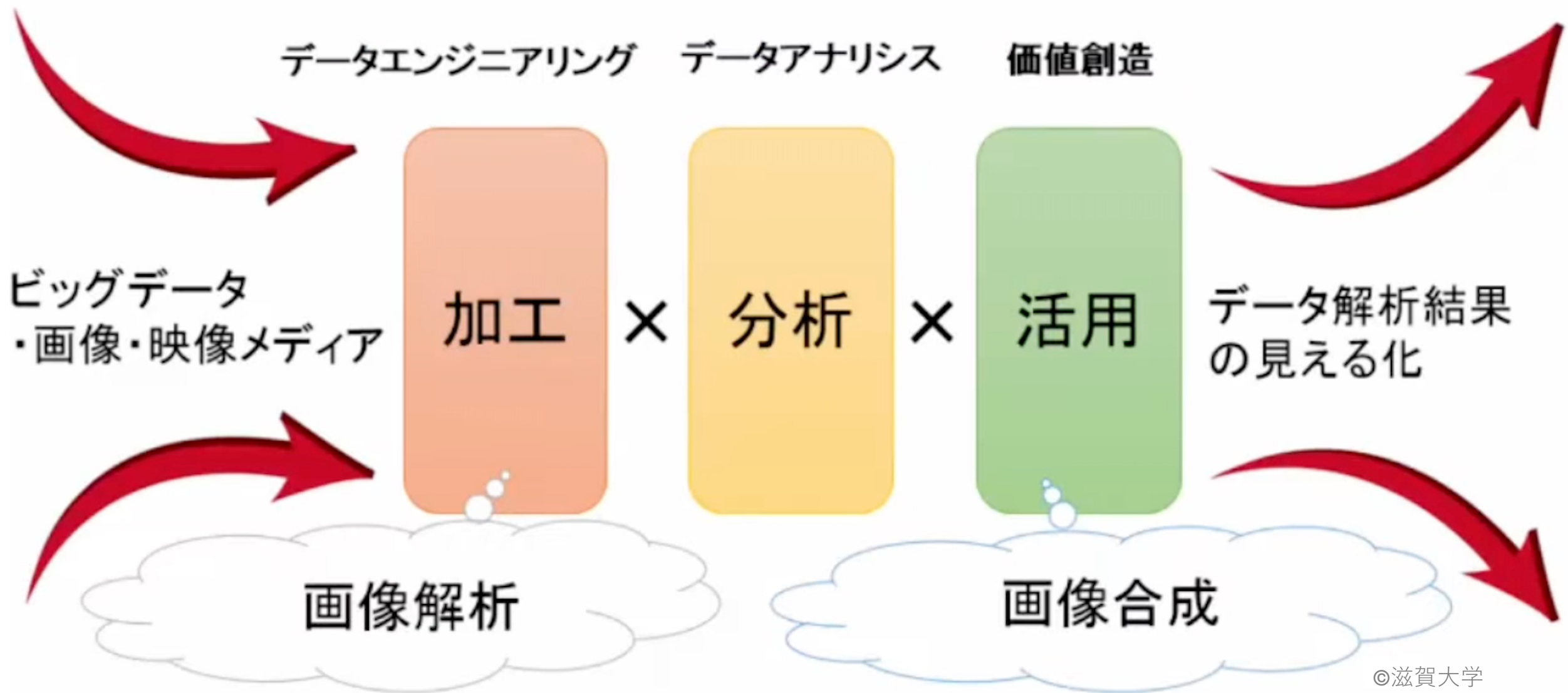
今回学ぶこと

1. 社会で起きている変化
2. 社会で活用されているデータ
3. データ・AIの活用領域
- 4. データ・AI利活用のための技術**
5. データ・AI利活用の現場
6. データ・AI利活用の最新動向

データサイエンスと画像処理技術

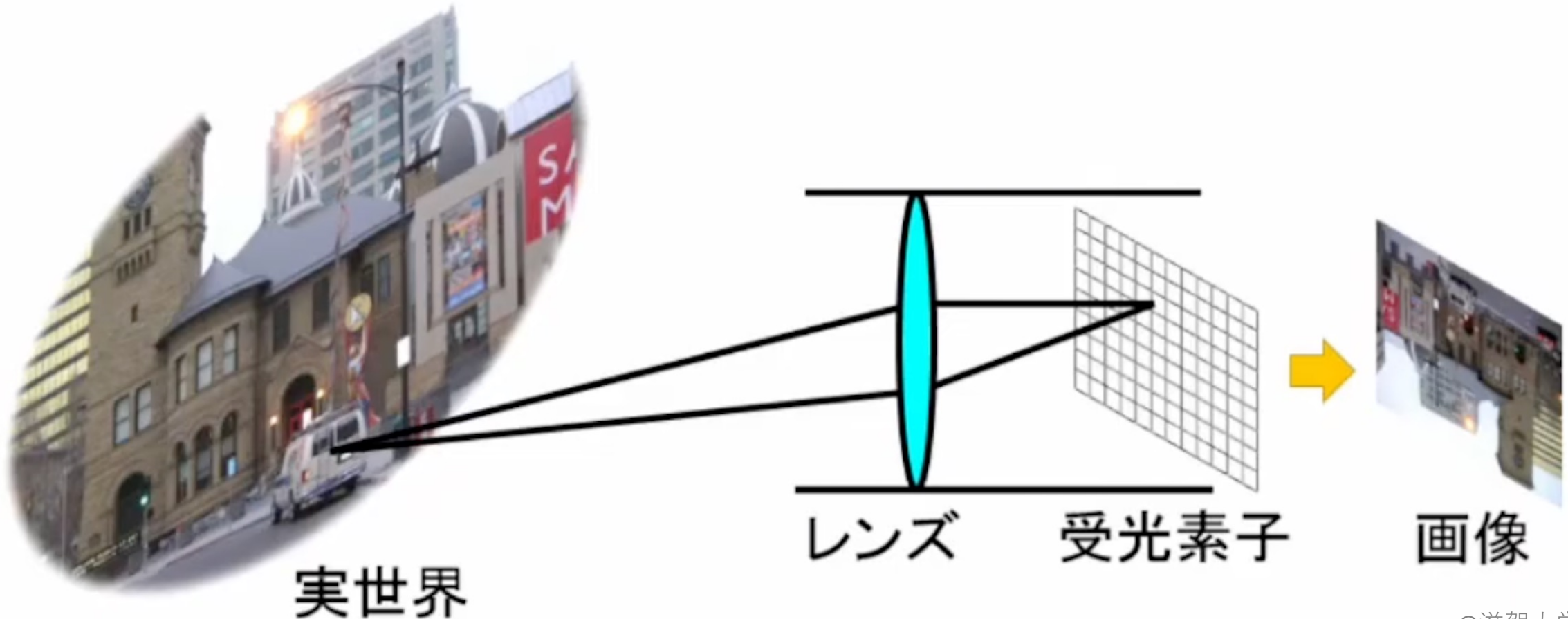
画像処理の応用

データサイエンスと画像処理



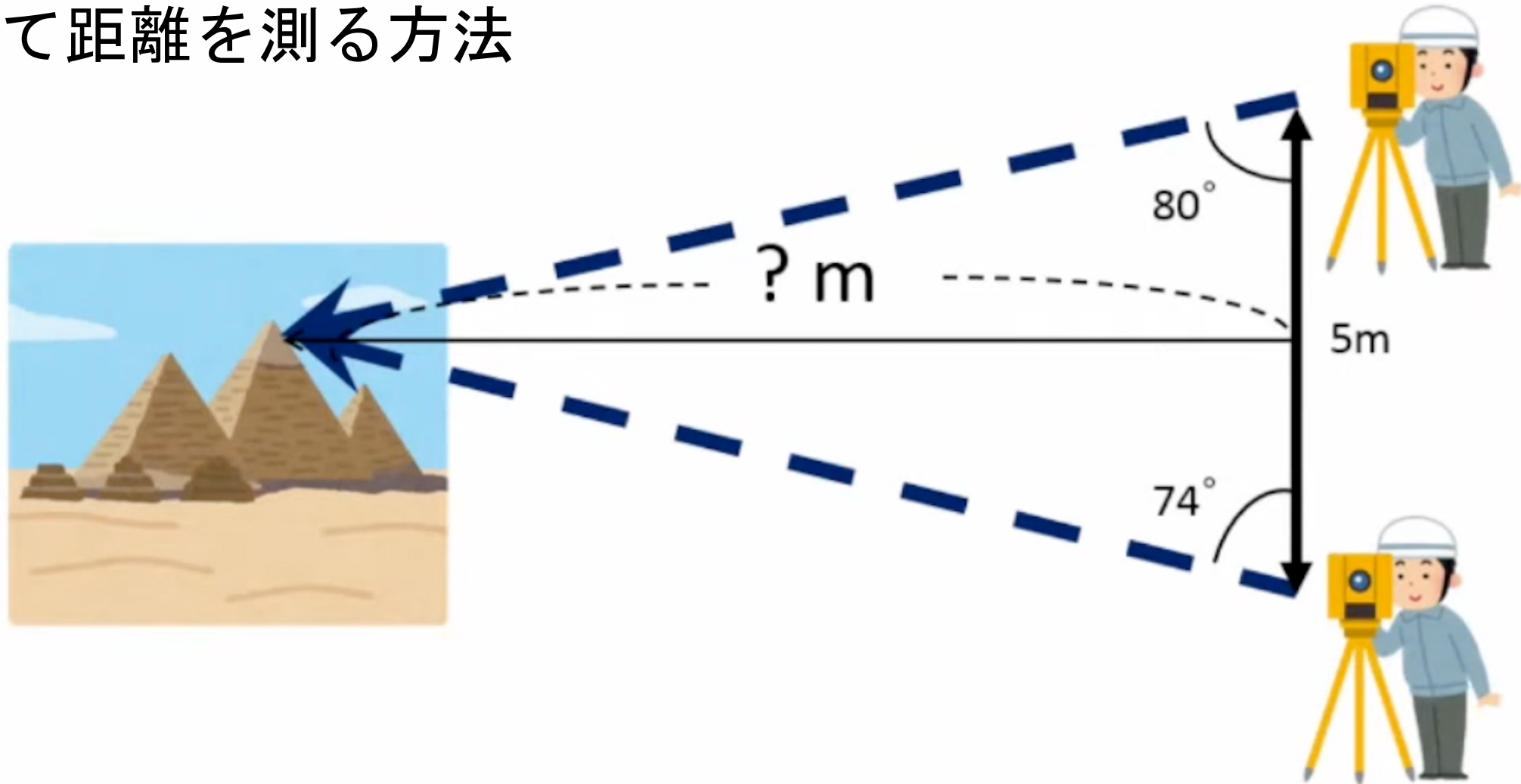
画像解析：写真撮影で失われるもの

- 写真は3次元世界を2次元世界に変換したもの
- 距離の情報は失われる



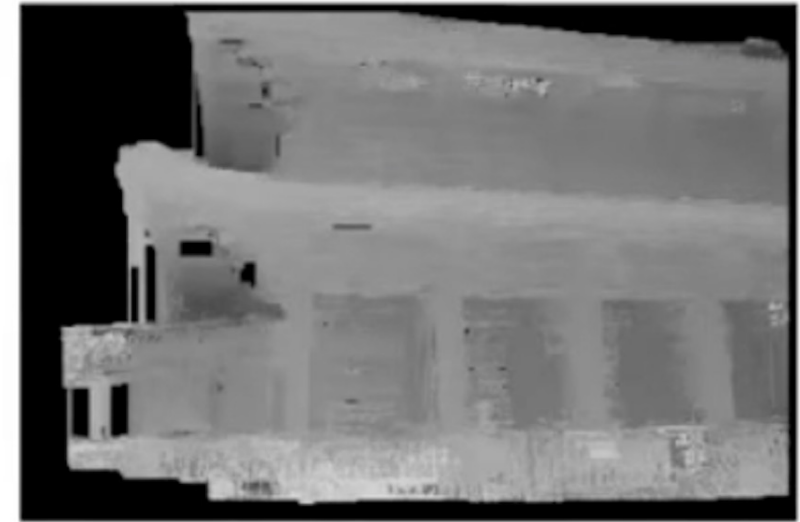
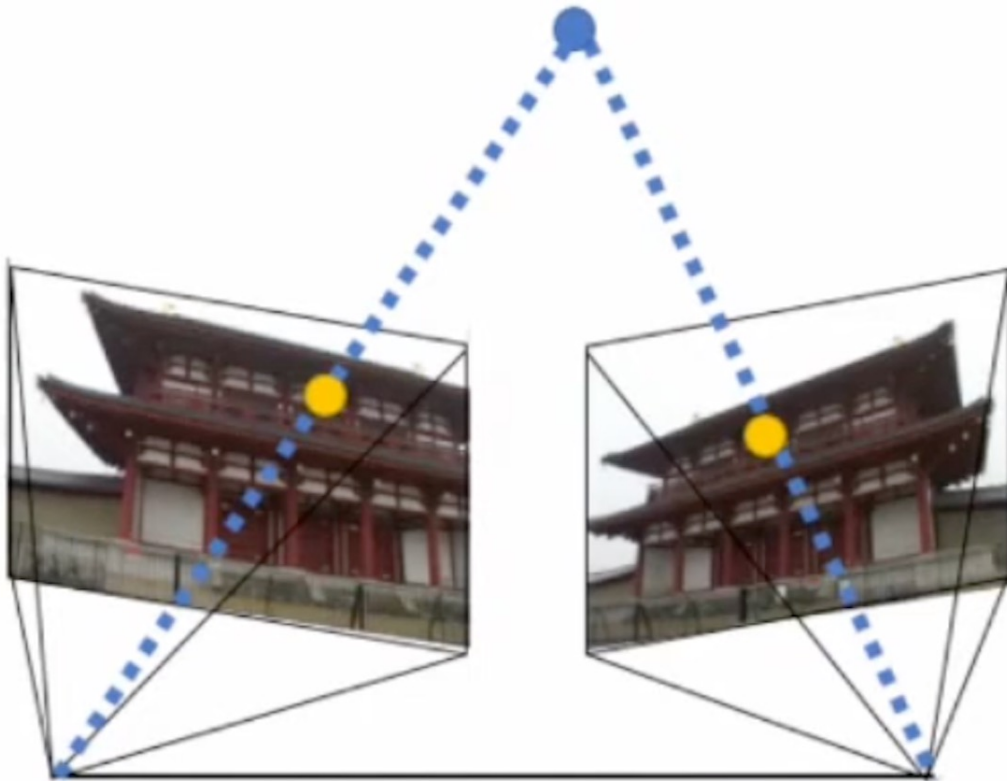
画像解析：三角測量

2つの違う位置から物を見たときにできる三角形を使って距離を測る方法



画像解析：2枚の画像を使った三角測量

- ステレオ法：画像を使って三角形を作ると距離を測る
- 画素ごとに測った距離を画像にすると距離がわかる画像（奥行画像）ができる



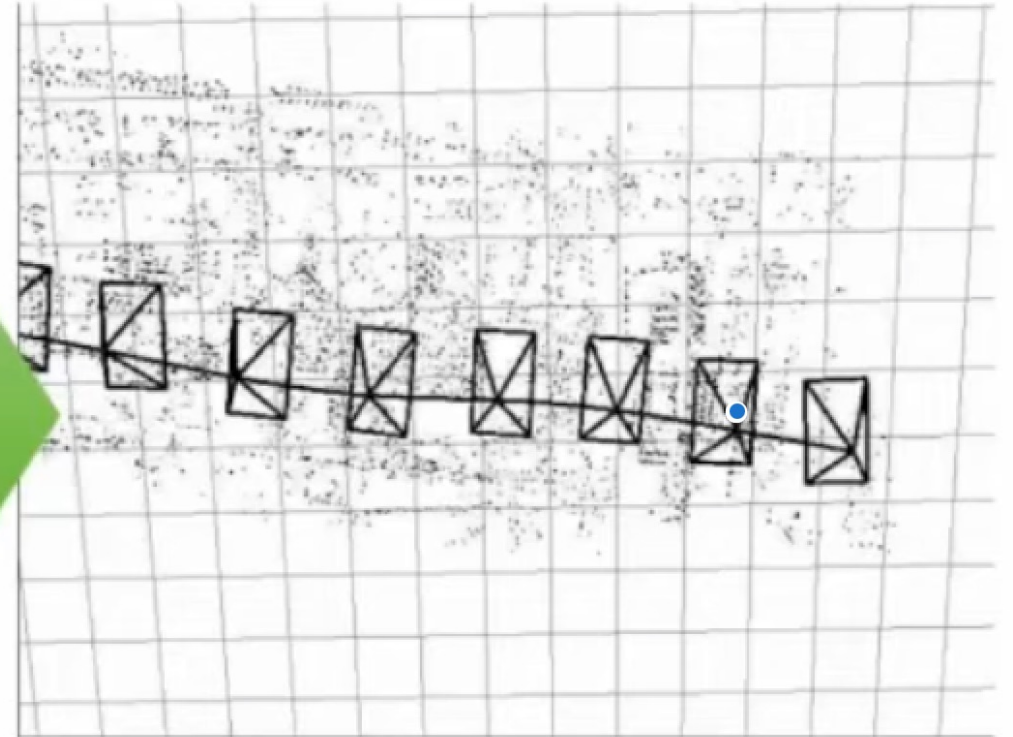
奥行画像

画像解析：動き情報から三次元復元

動画を解析し、カメラの動きや撮影シーンの三次元的な情報を取り出せる

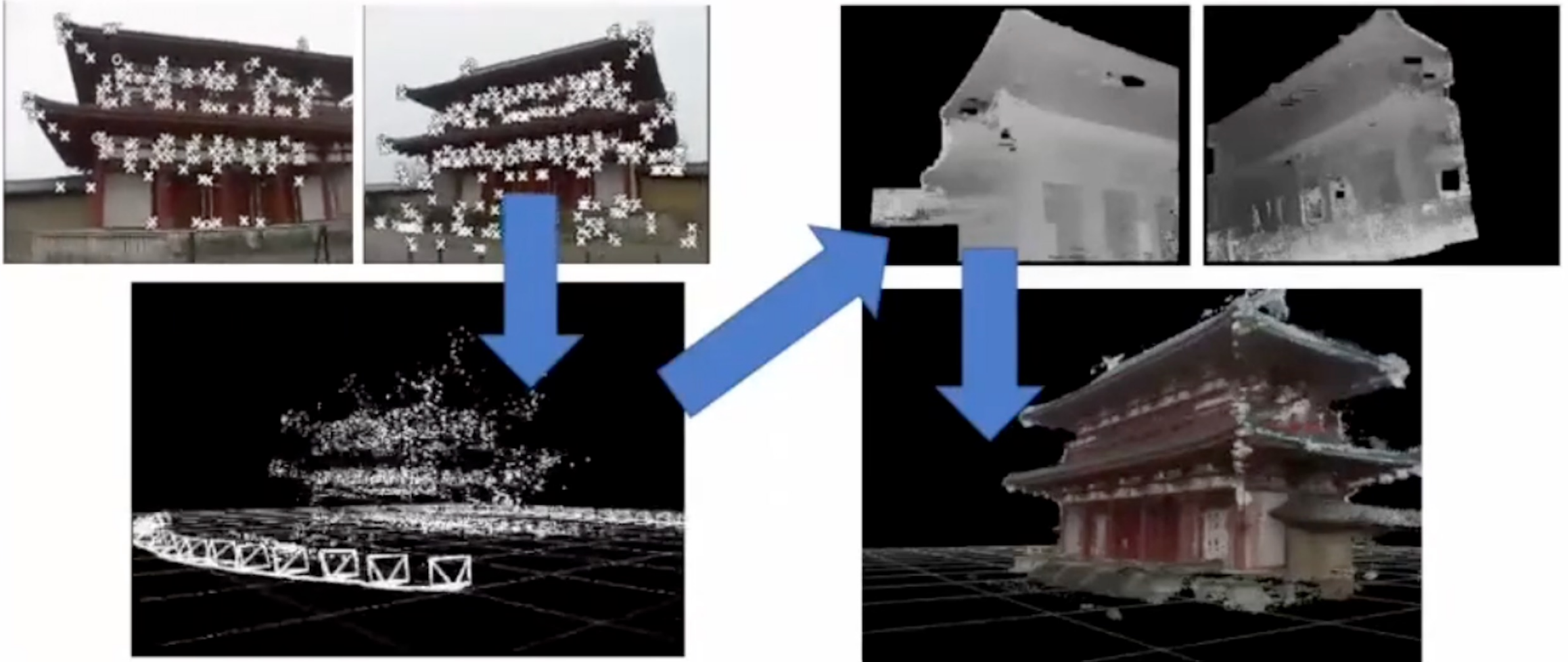


入力動画像



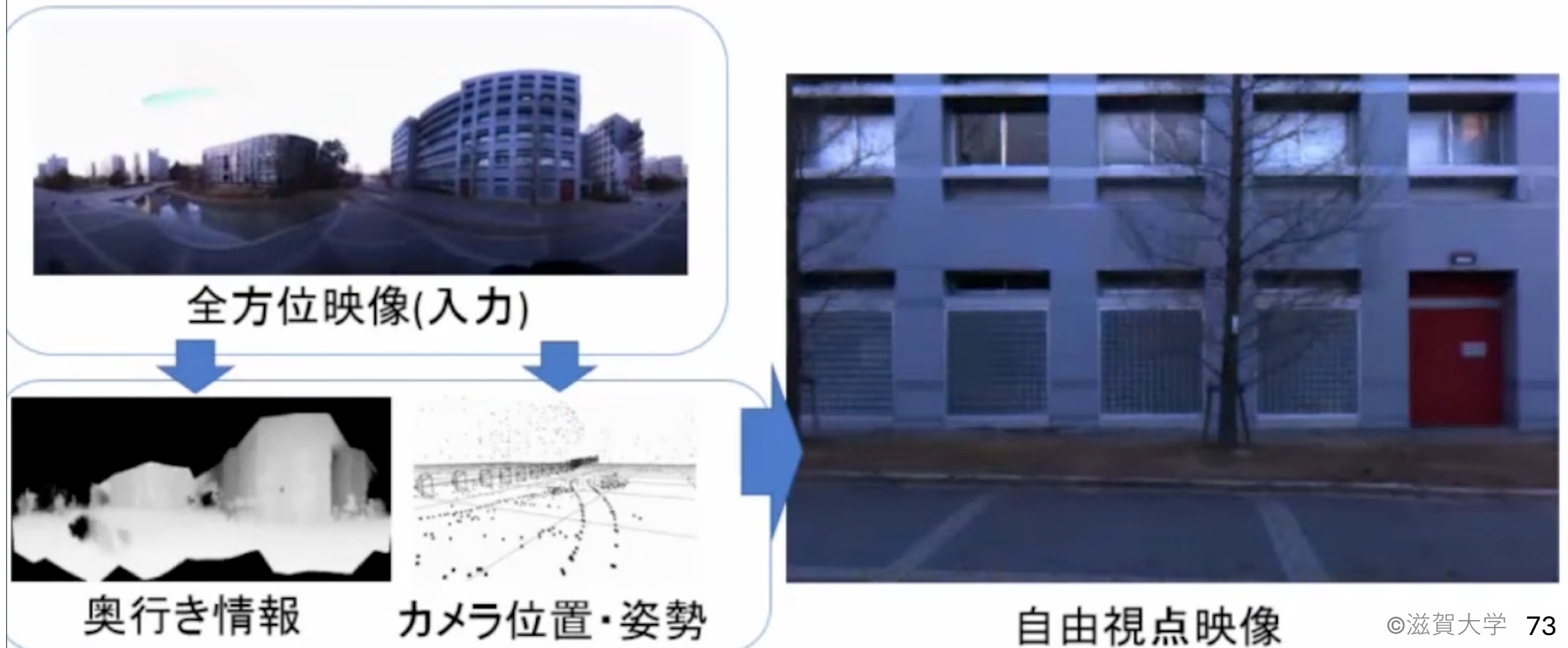
復元された三次元情報

画像解析：動画解析による三次元モデル作成



画像合成：自由視点映像生成

画像処理によって推定された三次元情報を活用すると、元々撮影していない自由な視点からの映像を合成できる



画像合成：自由視点画像生成による遠隔ロボットの操縦



ロボット



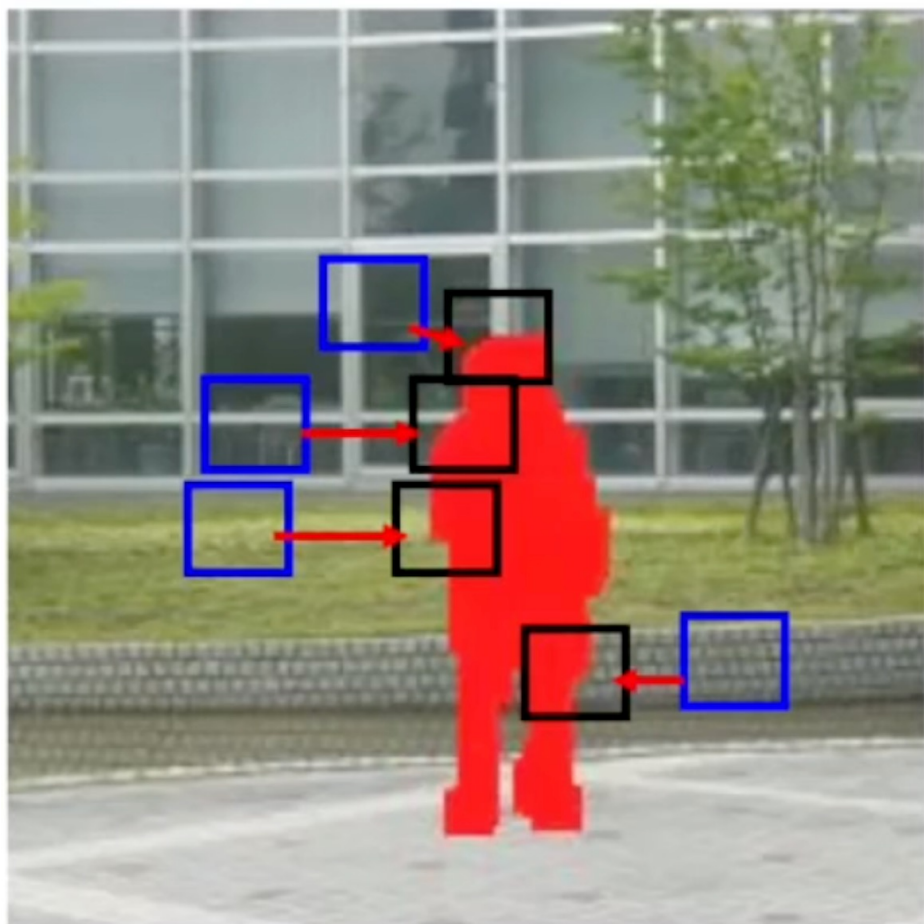
全方位カメラ

距離センサ



画像合成：画像修復

写真に写ってしまったり見えない場所の情報を復元できる



画像合成：移動物の自動消去

自由視点の技術と画像修復の技術を組み合わせて映像から動いている人だけを自動的に消すことができる

入力画像



処理後の画像



画像合成：リアルタイムでの物体消去

シーンの三次元情報を推定すれば、目の前の物体をその場で消してしまうこともできる

©滋賀大学



Input

Output

確認テスト 4

ターミナルで

quiz data-4

と入力。全問正解するまで解答を続けよう！

今回学ぶこと

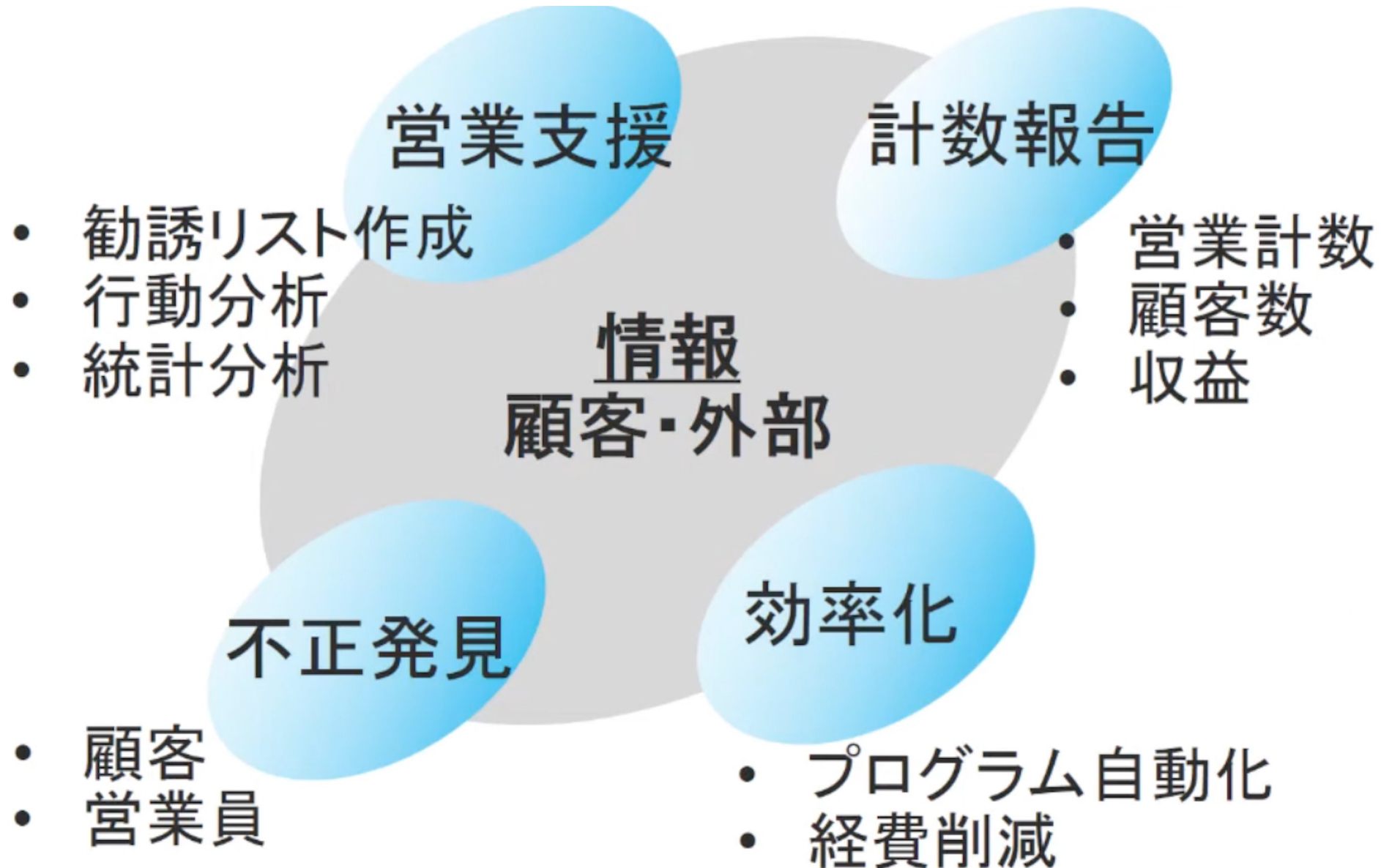
1. 社会で起きている変化
2. 社会で活用されているデータ
3. データ・AIの活用領域
4. データ・AI利活用のための技術
- 5. データ・AI利活用の現場**
6. データ・AI利活用の最新動向

金融

銀行が所有するデータ

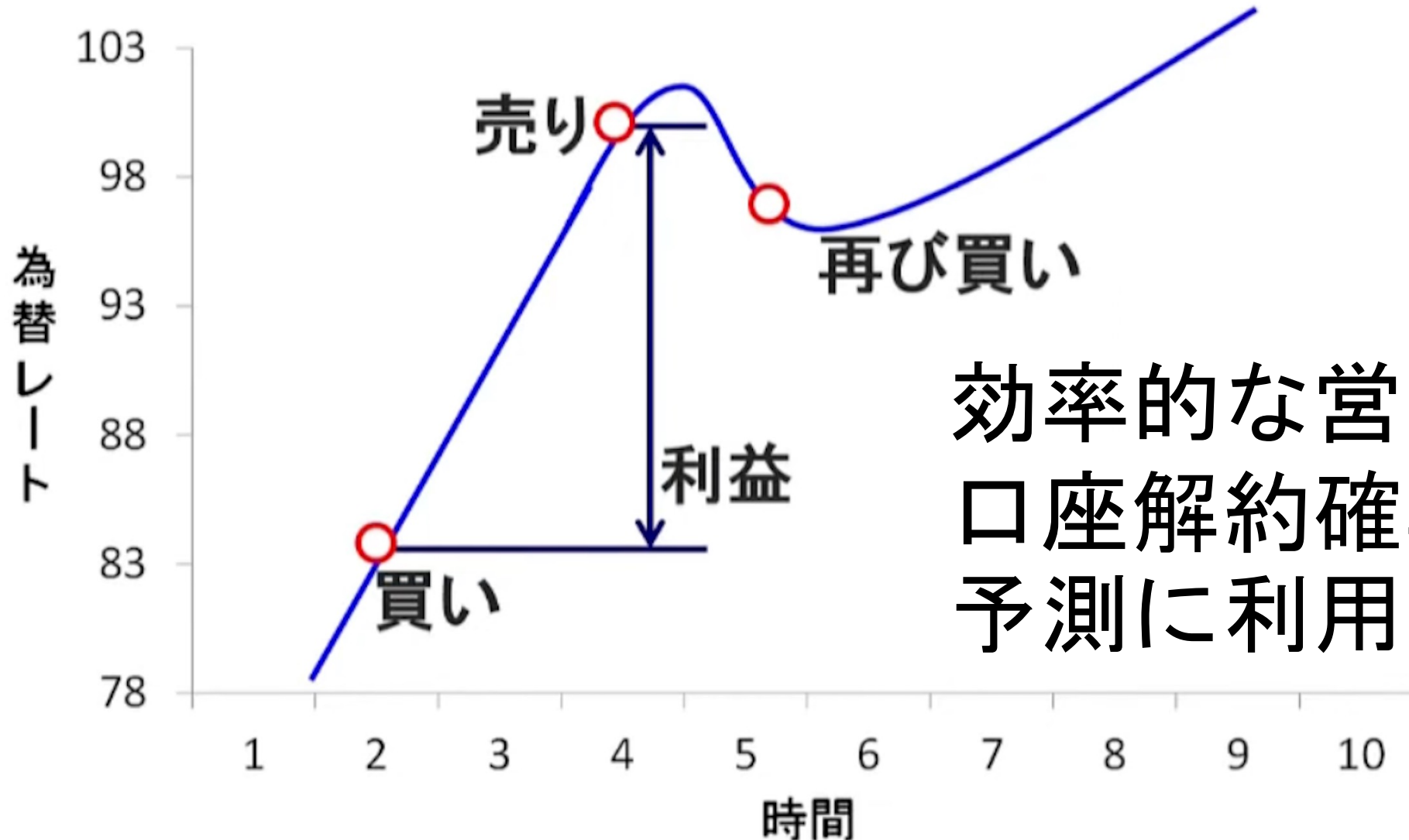
- 名前、誕生日、住所、電話番号
- 預金残高
- 取引履歴
- 預金、投資信託、ローン等
- 収益の情報
- 顧客との会話内容記録

データを使ってできること



分析事例（顧客別損益分析）

外貨預金の場合



効率的な営業活動や
口座解約確率などの
予測に利用

分析事例（解約率モデルと防止施策の実施）

- 直近の入出金状況
- サービスの利用頻度
- 年齢、口座開設年数など

に回帰分析を行い、口座解約予測を行う

確認テスト 5

ターミナルで

quiz data-5

と入力。全問正解するまで解答を続けよう！

今回学ぶこと

1. 社会で起きている変化
2. 社会で活用されているデータ
3. データ・AIの活用領域
4. データ・AI利活用のための技術
5. データ・AI利活用の現場
6. データ・AI利活用の最新動向

機械学習の先進的な事例

マーケティング

BtoC企業が保有しているビッグデータ

BtoC (Business to Consumer): 企業が一般消費者を対象に行う
ビジネス形態

消費者の属性と行動（購買履歴）が分かるビッグデータ

- 小売店（デパート、コンビニ、量販店など）
 - 売り上げデータ
- 鉄道
 - ICカードによる乗降データ、購入履歴
- Eコマース
 - Web上での購入、クリック数、サイトへの訪問履歴

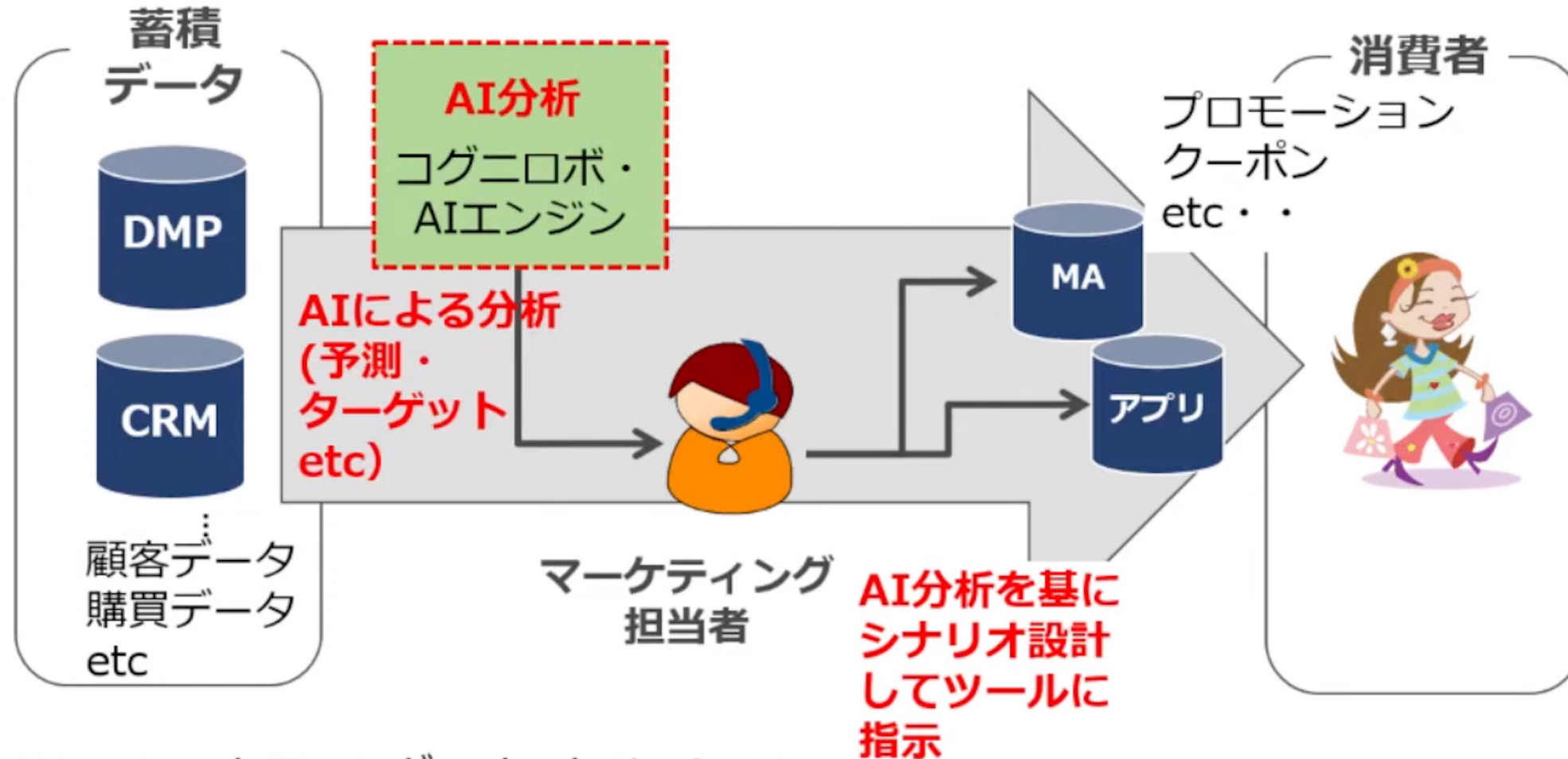
企業の課題

過去テクノロジーでの限界

性別や年齢など情報以外は把握できないなど、統計的手法で得られる顧客データのシンプルな特徴に基づいたシナリオ設計やプロモーションでは限界

AIを活用した複雑な特徴量を掛け合わせた新しい視点での分析が必要

マーケティング実務によるAI活用事例



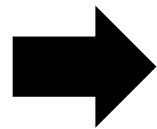
MA : マーケティング・オートメーション
DMP: データ・マネジメント・プラットフォーム
CRM: カスタマー・リレーションシップ・マネジメント

予測モデルの実例

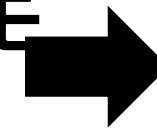
入力

出力

購入商品の
組み合わせ



年間来店日数予測モ
デル（AIモデル）



年間来店日
数の予測

予測モデルの実例

誰に新たに何を購入してもらえれば、年間来店日数が増加するかを予測

顧客Aの購買実績

ワイン	SDカード	ゲーム	化粧品	
1	0	1	0	



顧客Aの仮想実績

ワイン	SDカード	ゲーム	化粧品	
1	0	1	1	

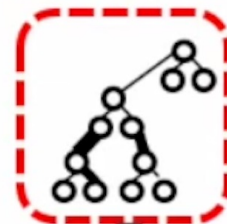
顧客Aが化粧品を
購入したと仮定

来店予測モデル



年間来店日数の予測 = 3日

来店予測モデル



年間来店日数の予測 = 7日

<

予測モデルの実例

©滋賀大学

購入商品					年間来店日数
顧客A	商品A				●日
顧客B	商品A	商品B			▼日 #日
顧客C	商品A		商品B		×日 &日
顧客D	商品A			商品D	△日 ▼日
顧客X	商品A	商品B			◆日 @日
顧客Y	商品A	商品B	商品C		○日 ◆日 *日

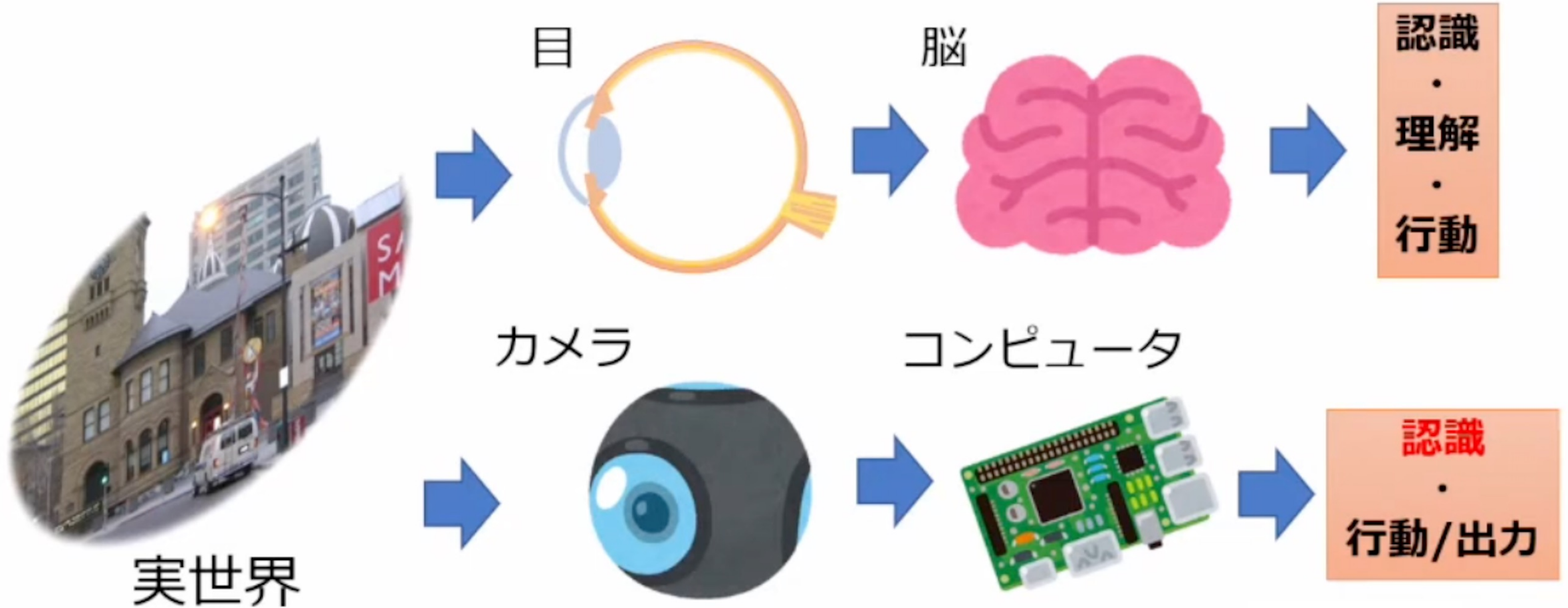
顧客Aは、商品Aを購入した場合、年間の来店日数は●日であるとAIが予測

機械学習の先進的な事例

画像

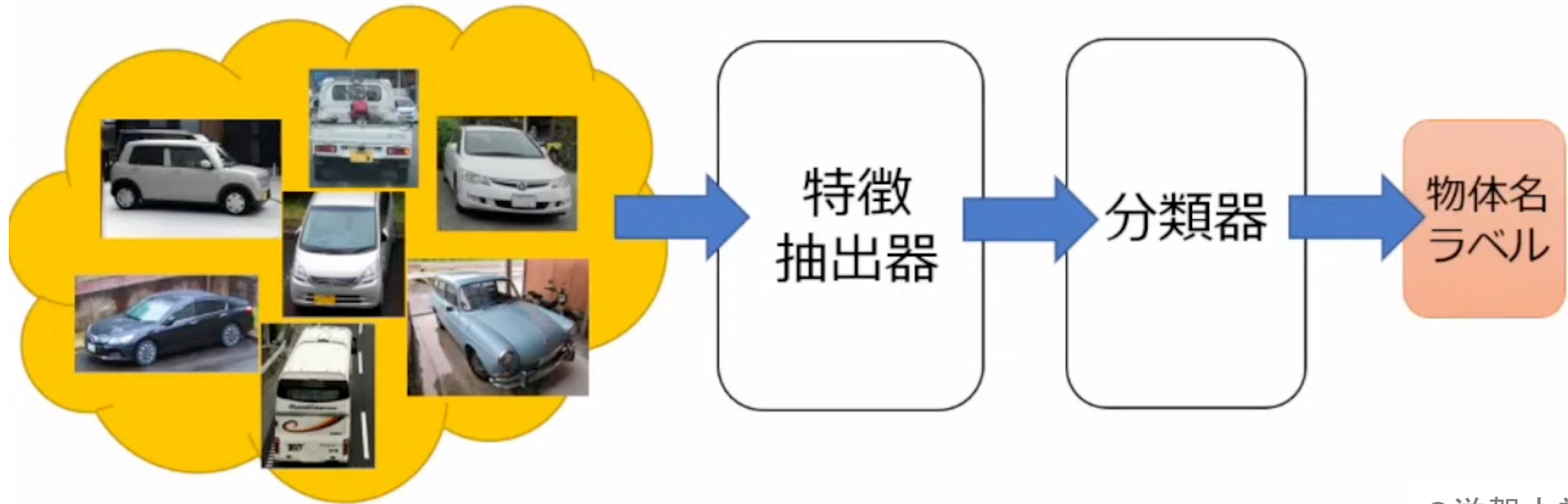
機械学習とコンピュータビジョン

コンピュータに人間のような視覚機能を与える



物体認識の方法（一般的な機械学習）

物体認識：画像中に撮影された物のカテゴリや名前を推定するもの



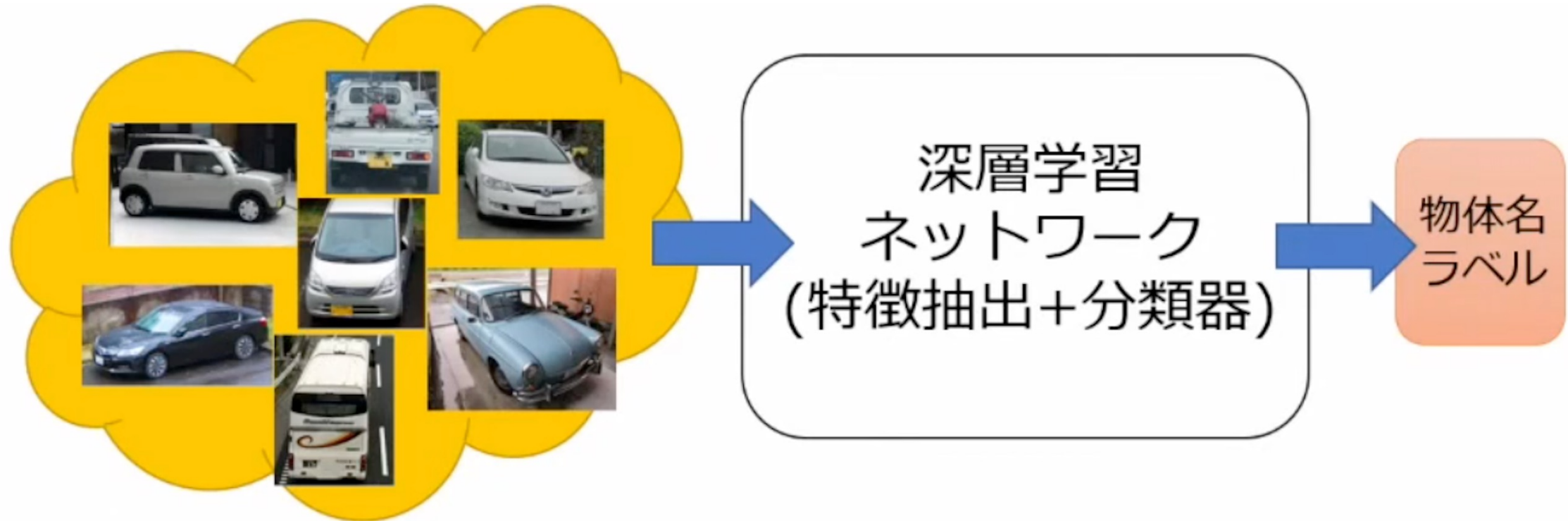
画像群

©滋賀大学

画像から特徴を抽出し、その特徴と物体名を関連づけるように分類器を学習させる

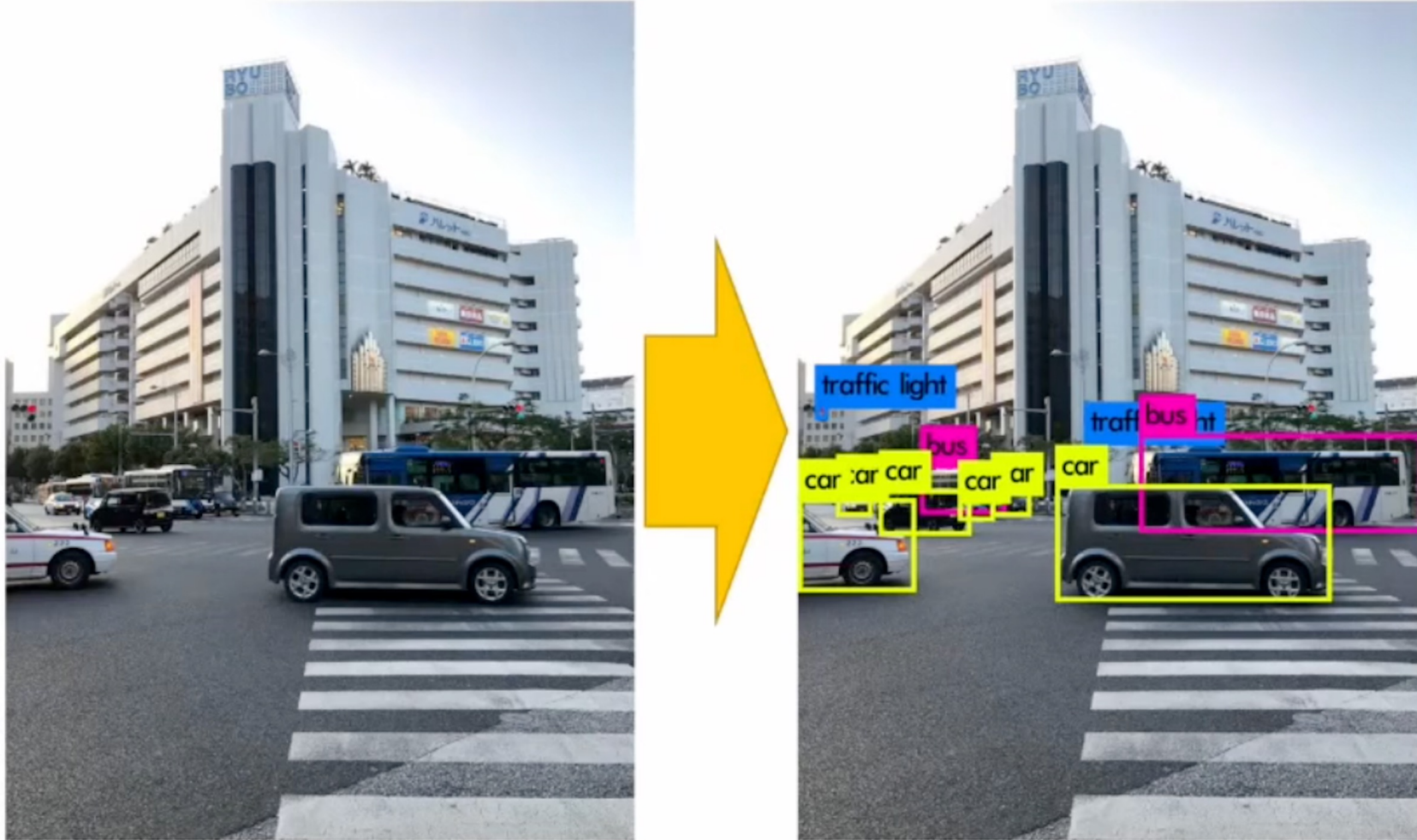
物体認識の方法（深層学習）

学習用の画像データが大量にある場合には、特徴の抽出から識別までを一括で行う深層学習が利用されることが多い

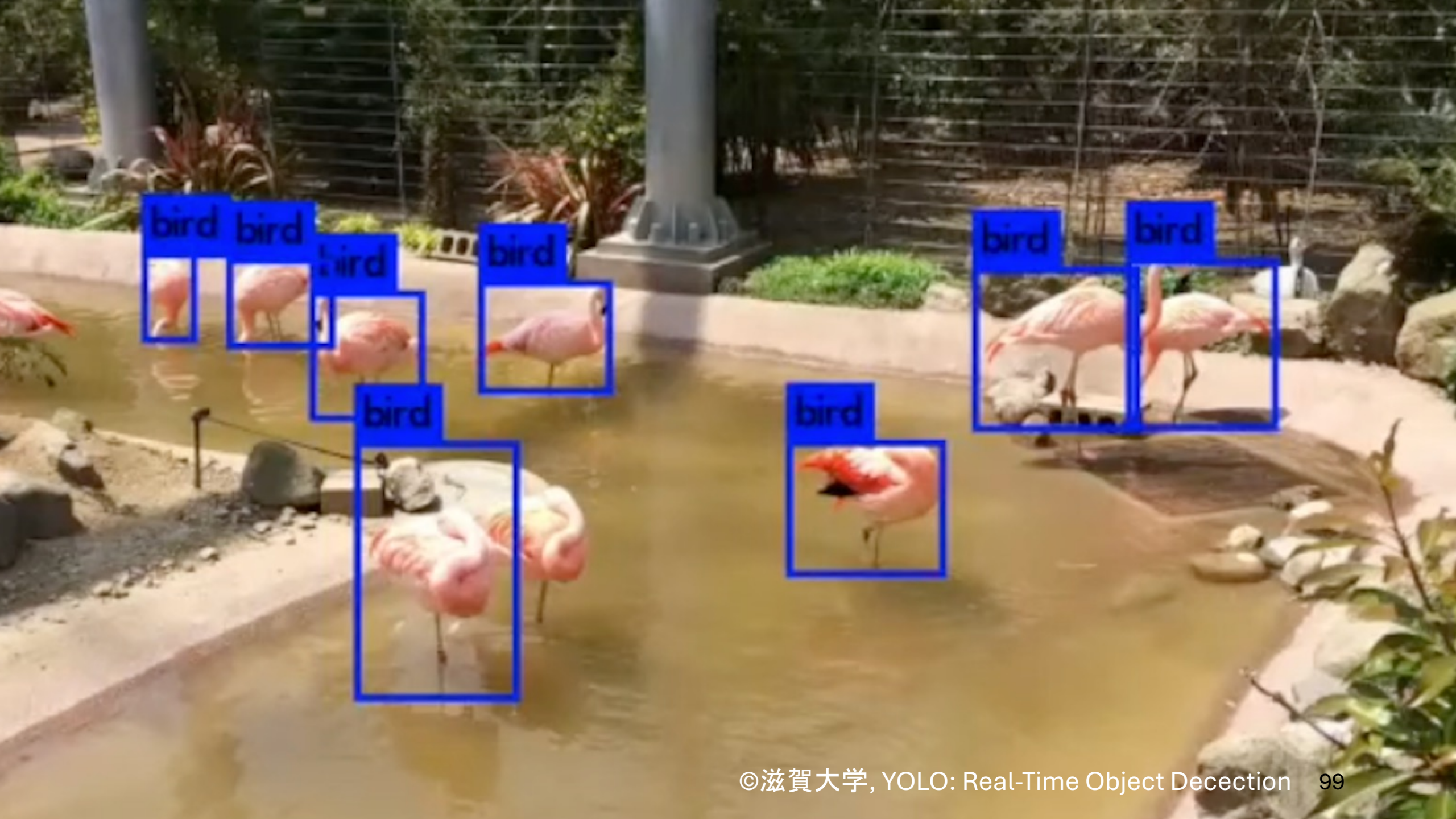


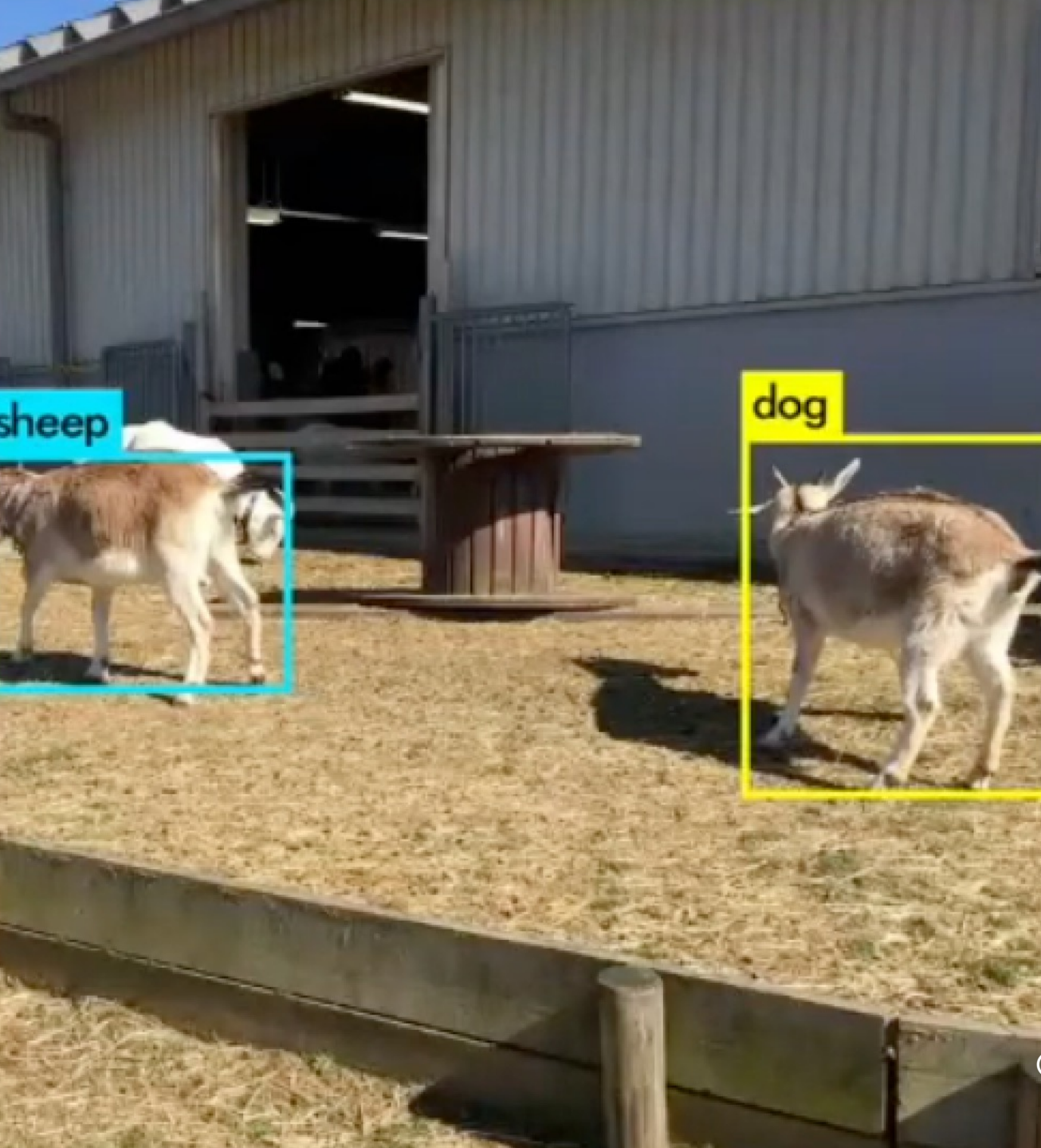
物体名ラベル付きの画像群

物体認識の例



YOLO: Real-Time Object Detection による物体認識結果
* <https://pjreddie.com/darknet/yolo/>





画像からのキャプション生成の例



夕食を囲む
人々

Ushiku et al. (2015)

GAN（敵対的ネットワーク）

画像の生成器と識別器の2つを使って
画像の生成と認識を自動で繰り返して
学習するネットワーク

GAN（敵対的ネットワーク）

生成器は識別器を騙せるよう学習
識別器は正しい判断ができるよう学習



確認テスト 6

ターミナルで

quiz data-6

と入力。全問正解するまで解答を続けよう！

今回の問い

- 現代社会ではどのような変化が起きているのだろうか？
 - ビッグデータが生み出され、データサイエンティストの重要性が急速に上がっている。
- データ・AIはどこで活用されているのか？
 - 画像解析、企業分析などで活用されている。
- データ・AI利活用の最新動向はどのようなになっているのか？
 - マーケティングや物体認識などで実際に使われている

今回の目標

データサイエンス・AIが社会でどのように活用されているのかを理解する。

今回学んだこと

1. 社会で起きている変化
2. 社会で活用されているデータ
3. データ・AIの活用領域
4. データ・AI利活用のための技術
5. データ・AI利活用の現場
6. データ・AI利活用の最新動向

次回

第 1 回	社会におけるデータ・AI利活用
第 2 回	表計算ソフトの基本操作
第 3 回	セルの参照と集計
第 4 回	度数分布表とヒストグラム
第 5 回	代表値と散らばりの尺度
第 6 回	データのグラフによる表現とデータ比較
第 7 回	データの分布と標本抽出
第 8 回	推定
第 9 回	分割表とクロス集計
第 1 0 回	2 変量データと相関
第 1 1 回	時系列データ
第 1 2 回	仮説検定
第 1 3 回	回帰分析