

車種判別システム構築を通じて見えてきた AIをマーケティング分野で活用する際の 技術的チャレンジ

佐藤 剛宣

AIソリューションアーキテクト

クラウドファン株式会社

2017/5/16

講演内容

- 統計とビッグデータについて
- リアルビッグデータ収集の共通課題
- マーケティングで必要な『違い』の検出
一般物体認識と詳細画像識別について
- マーケティングで必要な精度とは？

統計とビッグデータについて①

ビッグデータに興味がある人がよく期待しているもの



予測値 = $f(\text{要素1}, \text{要素2}, \text{要素3}, \dots, \text{要素N})$

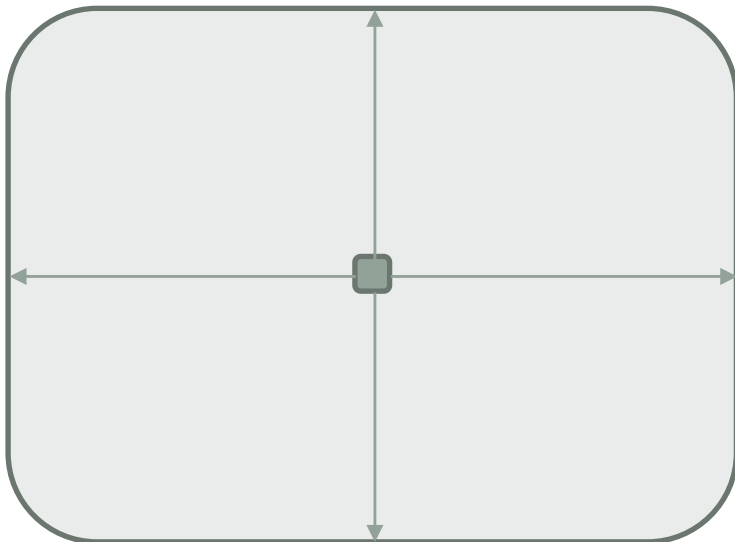


確かに、統計学の目的の一つは、“**the reduction of data**”
Statistical Methods Introduction By Ronald A Fisher, 1925

統計とビッグデータについて②

The Unreasonable Effectiveness of Mathematics in the Natural Sciences
By Eugene Wigner, 1960

“it is important to point out that the mathematical formulation of the physicist's often crude experience leads in an uncanny number of cases to an amazingly accurate description of a large class of phenomena”



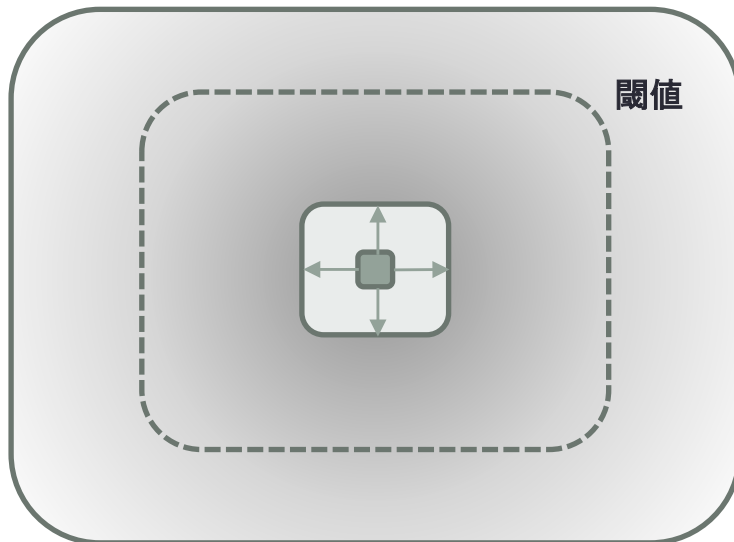
限られた事象から導き出された数式が、
それよりも遥かに多くの事象に応用可能
例) 万有引力の法則

統計とビッグデータについて③

The Unreasonable Effectiveness of Data

By Alon Halevy, Peter Norvig, and Fernando Pereira, Google, 2009

“... we're doomed to complex theories that will never have the elegance of physics equations. But if that's so, we should stop acting as if our goal is to author extremely elegant theories, and instead embrace complexity and make use of the best ally we have: the unreasonable effectiveness of data.”

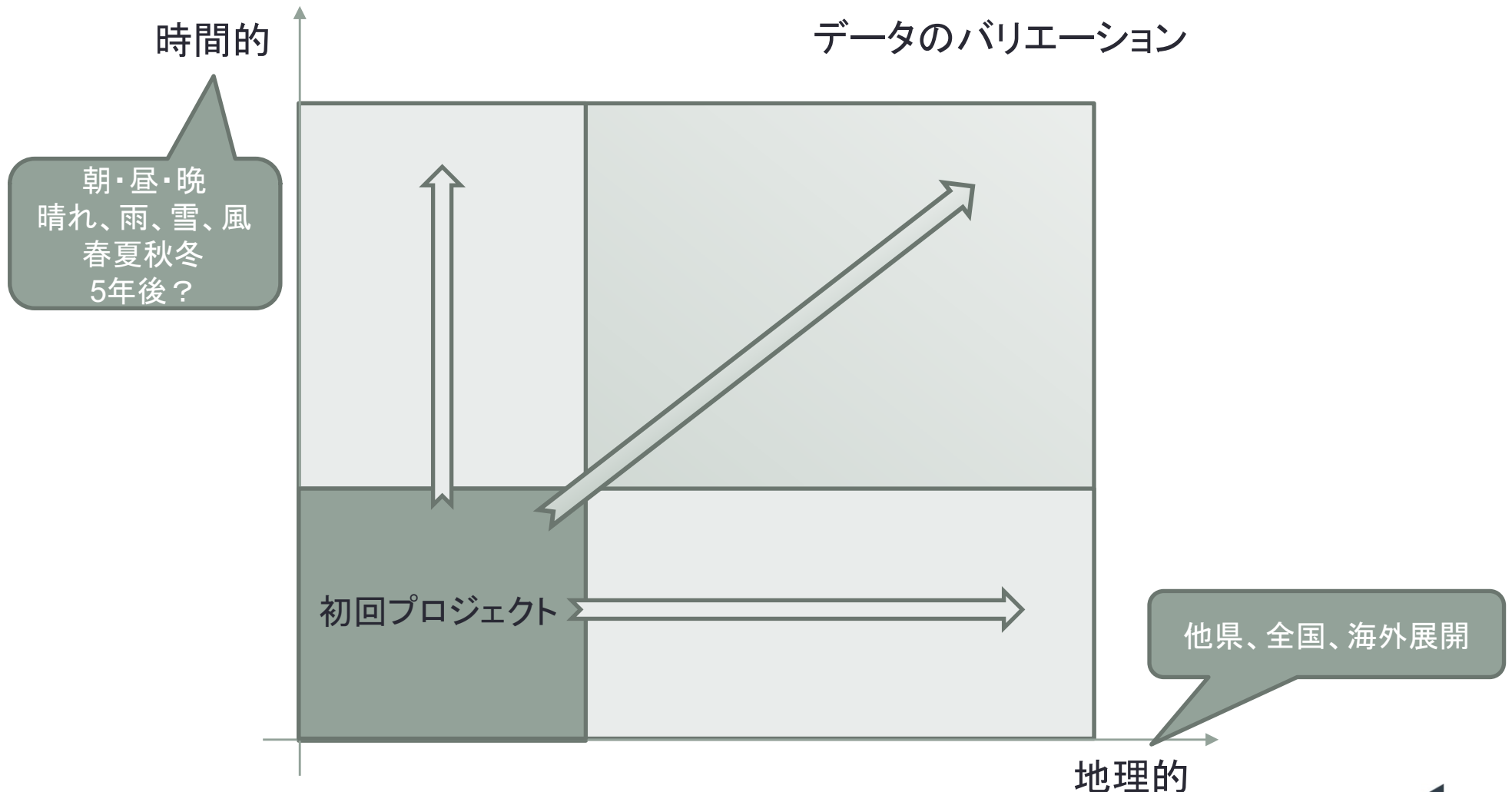


数式で表現できないのであれば、利用できるもの、つまり**大量のデータ**、を使ってモデル化すればよい

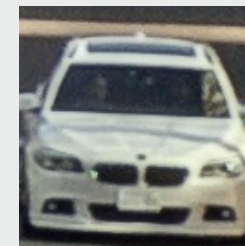
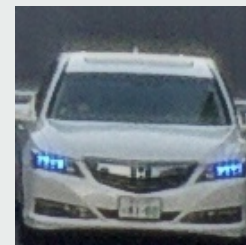
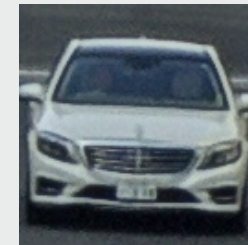
数式とは異なり、**データに含まれないバリエーションを表現することはできない**ため、十分大量なデータが必要

十分な精度を出すのに必要なデータ量は、無限ではなく、ある**閾値**があると考えられている

リアルビッグデータ収集の共通課題



マーケティングで必要な『違い』の検出 一般物体認識と詳細画像識別について①



自動車

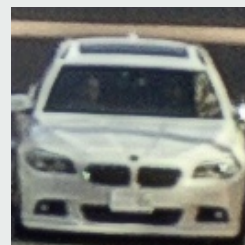
マーケティングで必要な『違い』の検出 一般物体認識と詳細画像識別について②



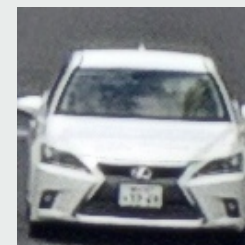
AUDI



MERCEDEZ



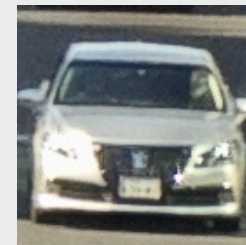
BMW



LEXUS



HONDA



TOYOTA

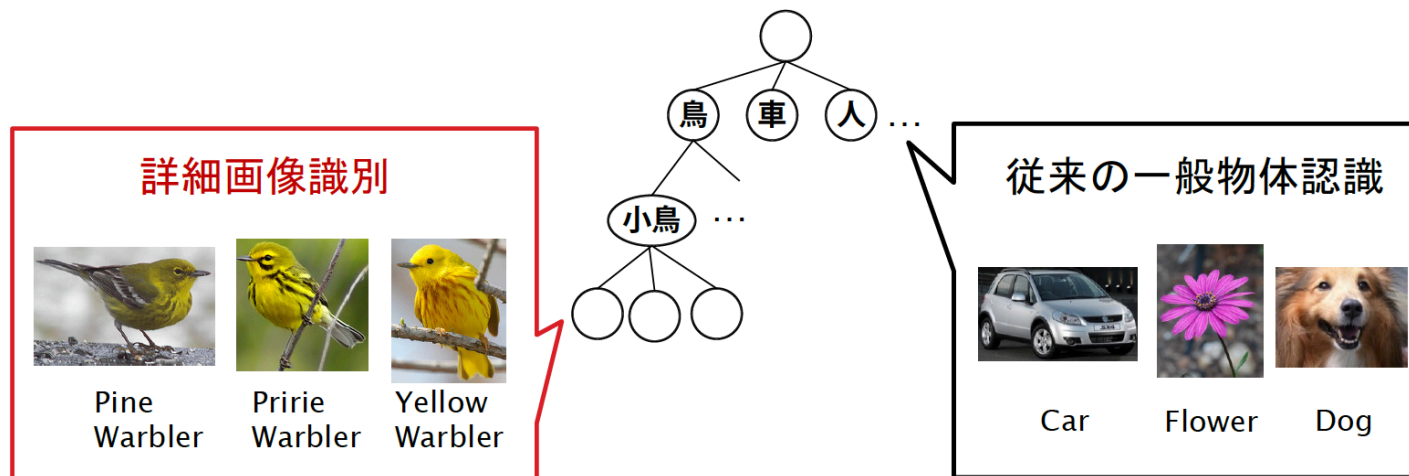
自動車

マーケティングで必要な『違い』の検出 一般物体認識と詳細画像識別について③

詳細画像識別

Fine-grained Visual Categorization (FGVC)
Subordinate-level Categorization

- ▶ **ある特定の対象領域における高粒度の多クラス画像識別**
 - 意味的・視覚的に非常に類似したクラスを扱う
 - ・ クラス間の違いは極めて細かく、局所的である場合が多い
 - 一般的な人間を遥かに超えるレベルの詳細な識別能力を目指す



タカとハヤブサはどこが違う？～新たな認識領域「詳細画像識別」の展開と応用～
東京大学大学院情報理工学系研究科 創造情報学専攻 中山研究室 中山英樹より

マーケティングで必要な精度①

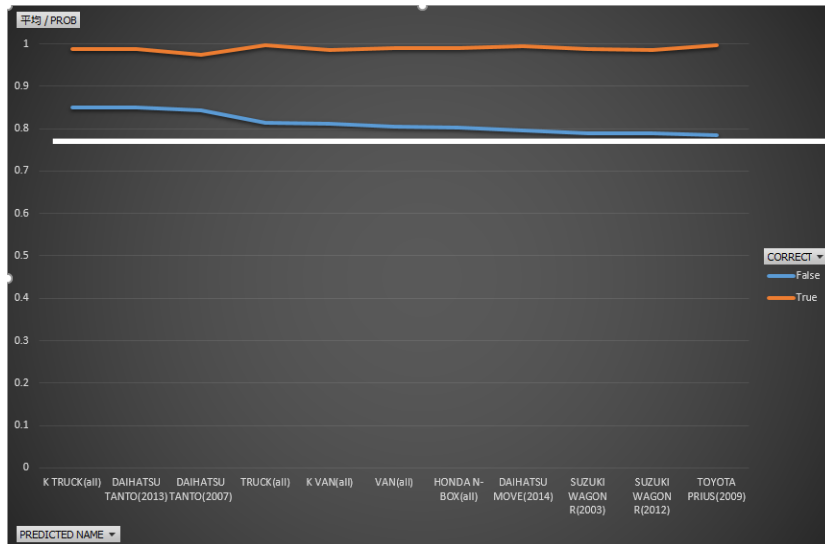
➤ 交通量調査系

- ✓ 全体の数字のカウントが目的
- ✓ 見過ごしよりはミスカウント
- ✓ False Negative (見過ごし) を抑制したい

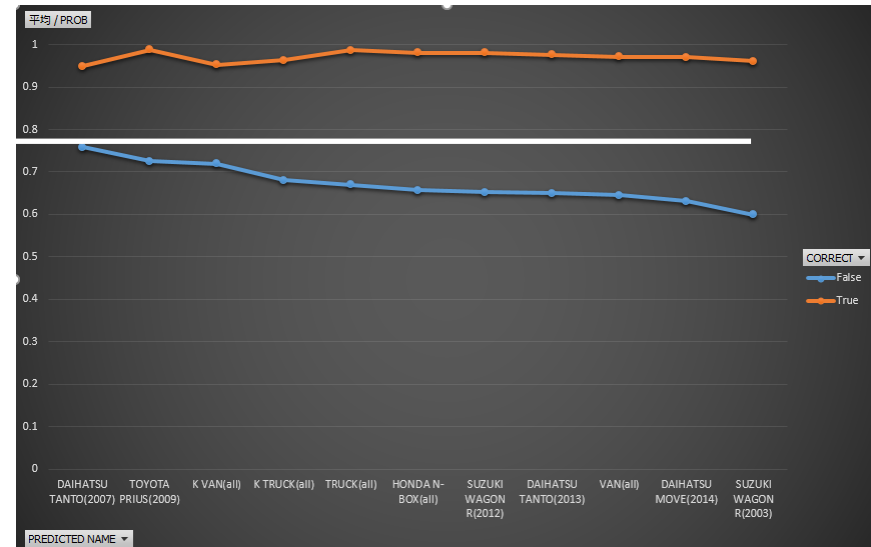
➤ ターゲット広告系

- ✓ ターゲティングが目的
- ✓ 間違ったターゲットよりは見過ごし
- ✓ False Positive (間違ったターゲット) を抑制したい

マーケティングに必要な精度②



False Positiveの抑制なしモデル
(11車種、青線: False Positive時の確率)



False Positiveの抑制ありモデル
(11車種、青線: False Positive時の確率)

ありがとうございました
tsato@cloudian.com