

ビッグデータ処理技術の進化と、 エッジへバーコンピューティング

西川 徹

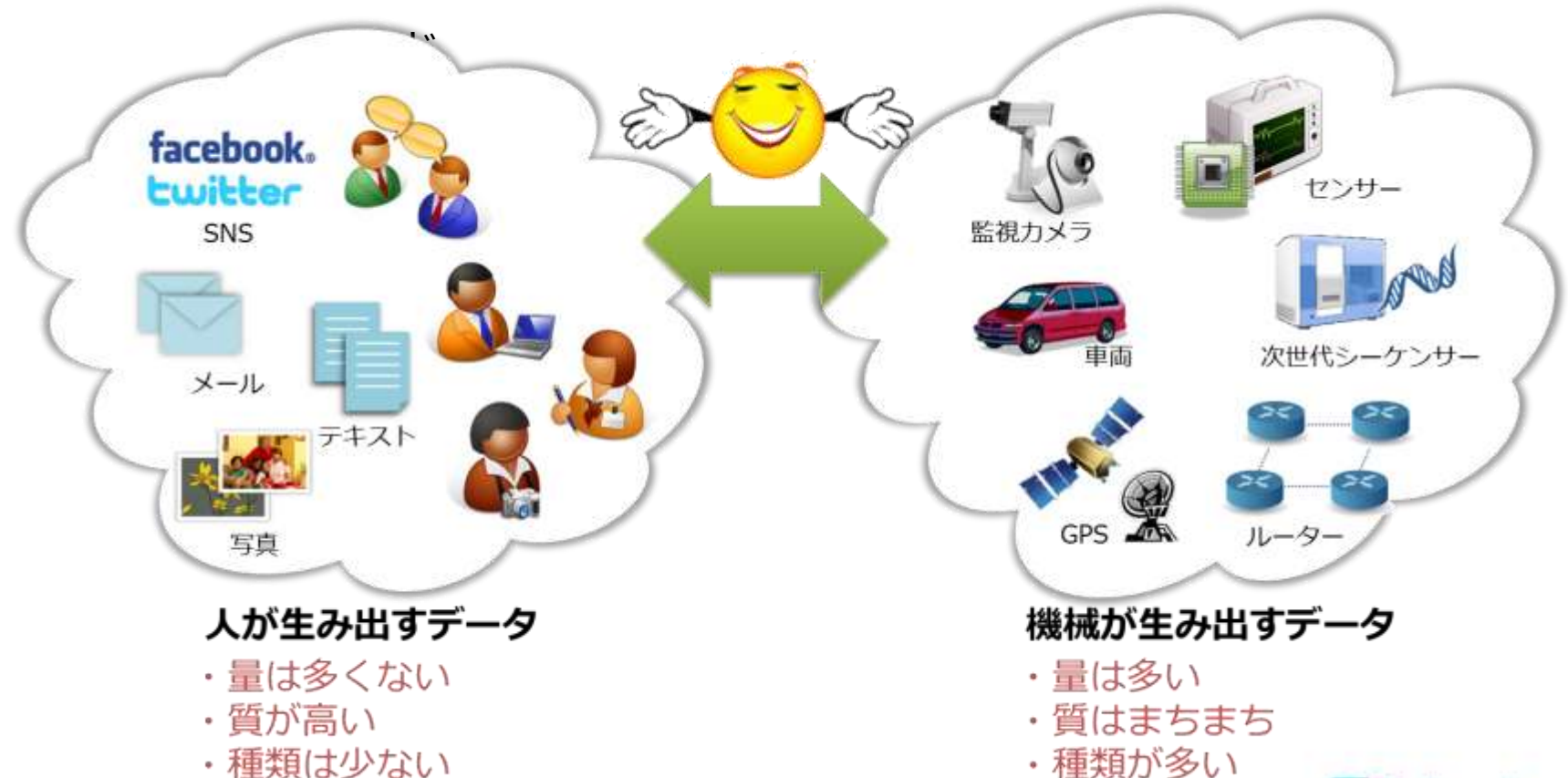
(株) Preferred Infrastructure

代表取締役

<nishikawa@preferred.jp>

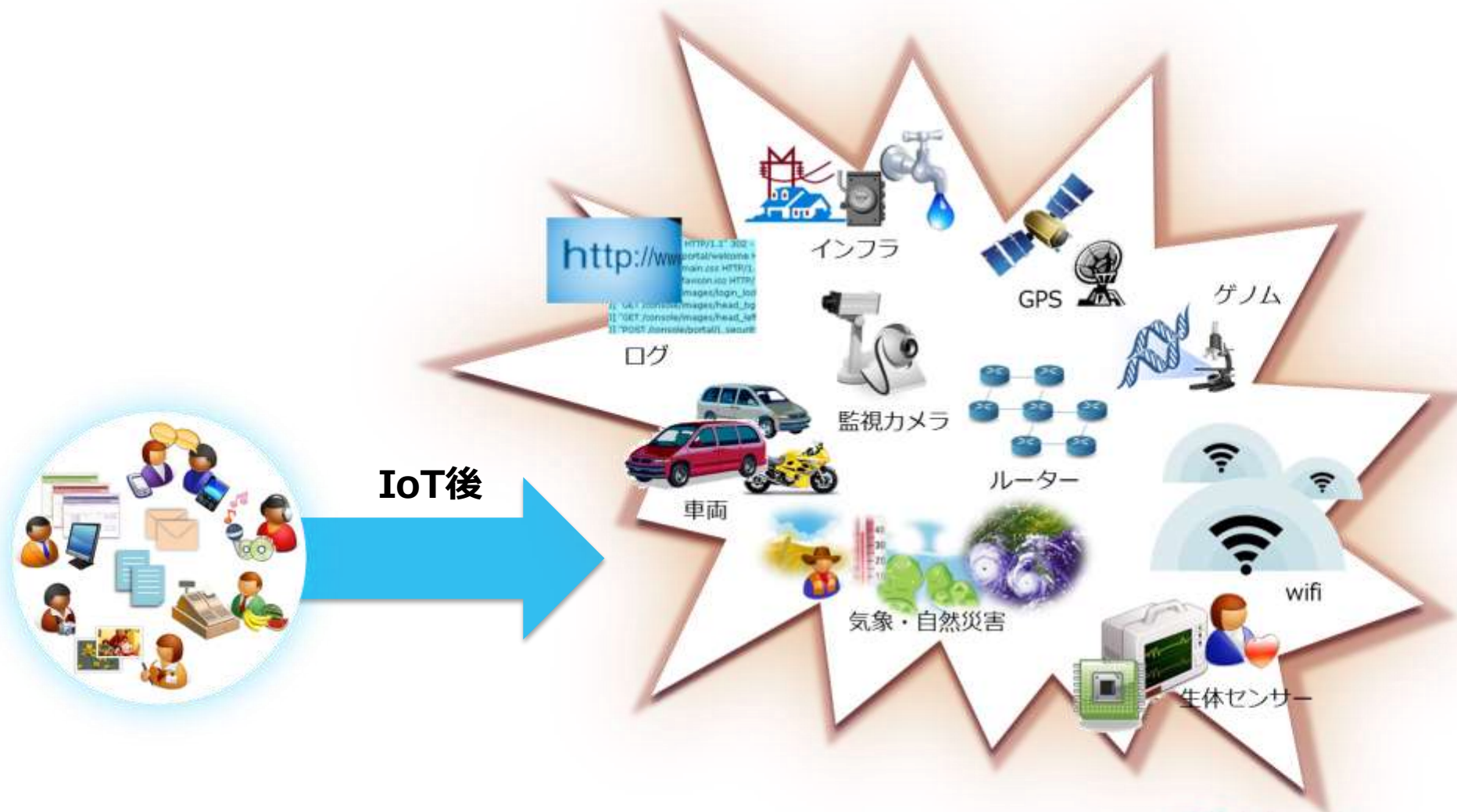
IoT時代の到来で、 IoTデータがビッグデータの主役となる

- 膨大なデータがエッジ側で生成されるようになる
 - 例：映像はカメラ1台で年間100TB、タービンセンサ、ポイントク

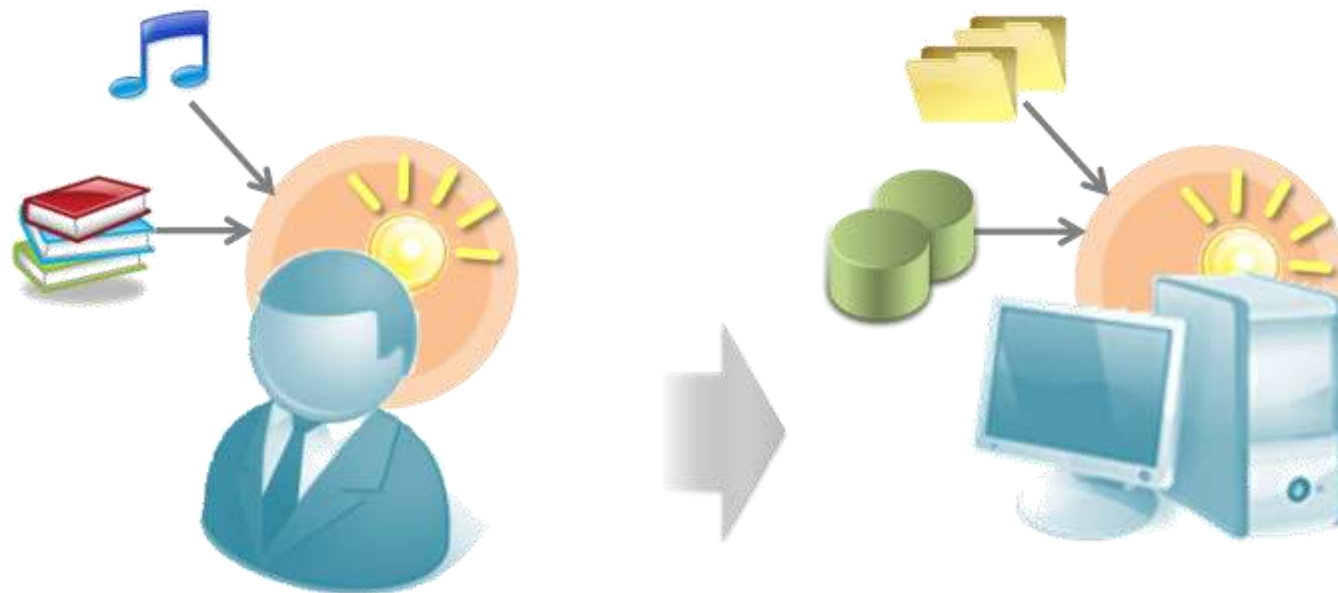


IoT・ビッグデータへの機械学習の適用

データの「種類」の爆発



機械学習技術を活用することにより、
データに対する深い分析の実現を目指す



「学習 = 人間が知識や知能を獲得する過程」を
コンピュータ上で実現しようとする

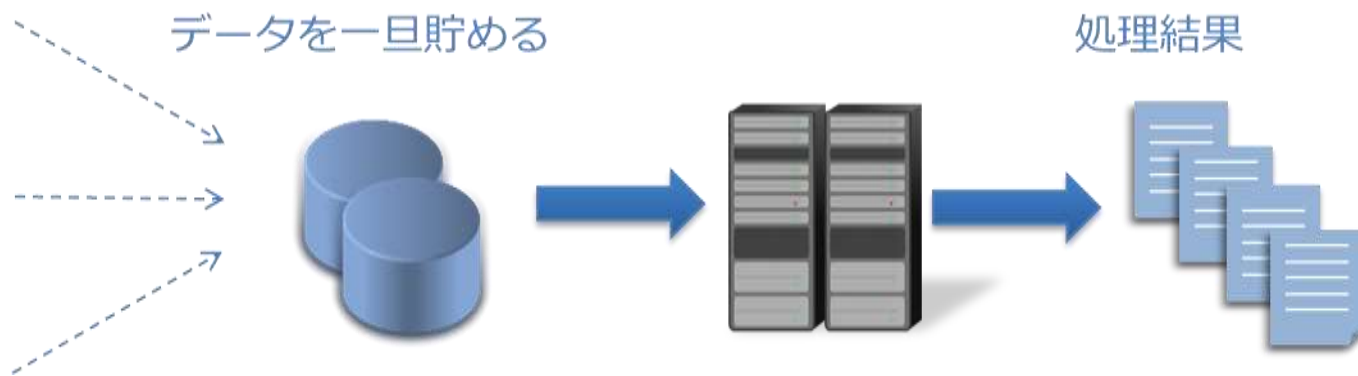
機械学習技術

- データがあるところ、どこでも使える
- 様々な分野の問題に利用可能

適用分野

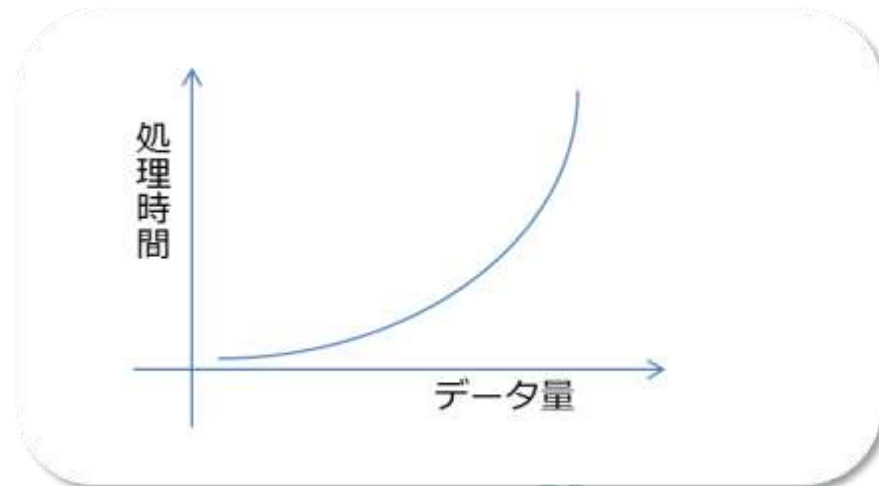
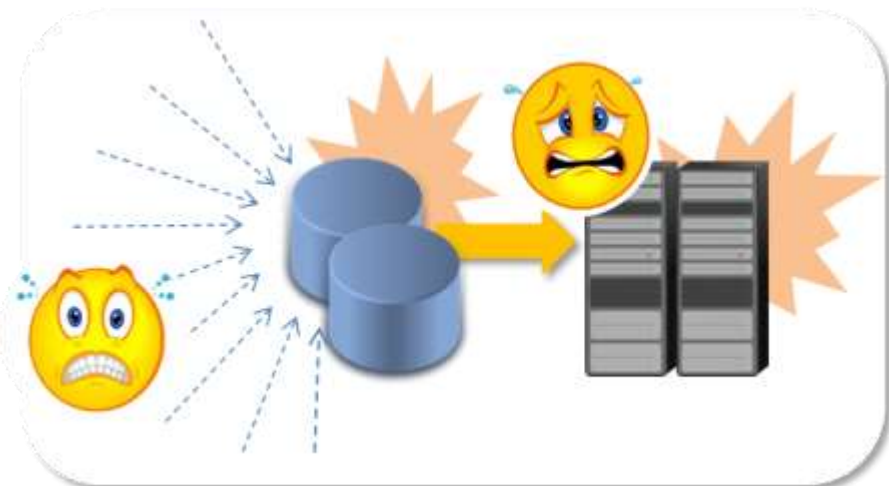
レコメンデーション クラスタリング	分類、識別	市場予測	評判分析
情報抽出	文字認識	ロボット	画像解析
遺伝子分析	検索ランキング	金融	医療診断

機械学習を大規模化する上での2つの難しい課題

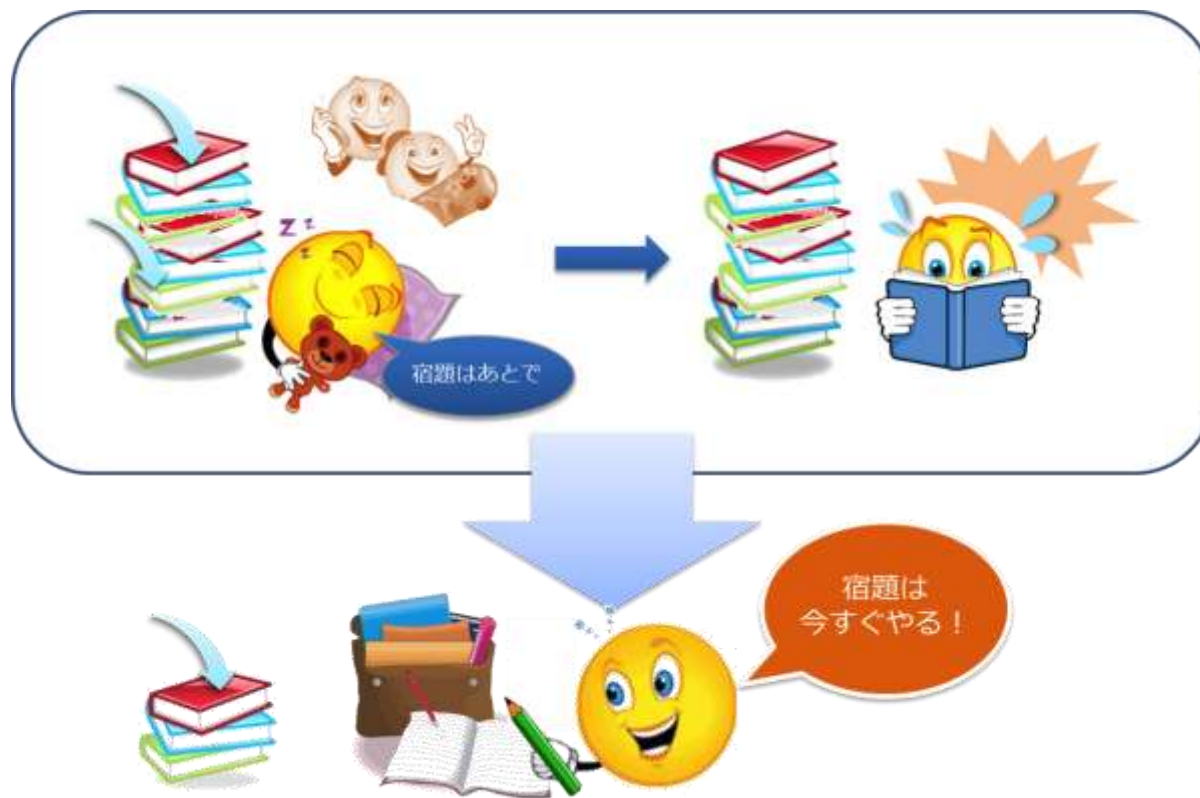


入出力がボトルネックとなる

計算量がボトルネックとなる

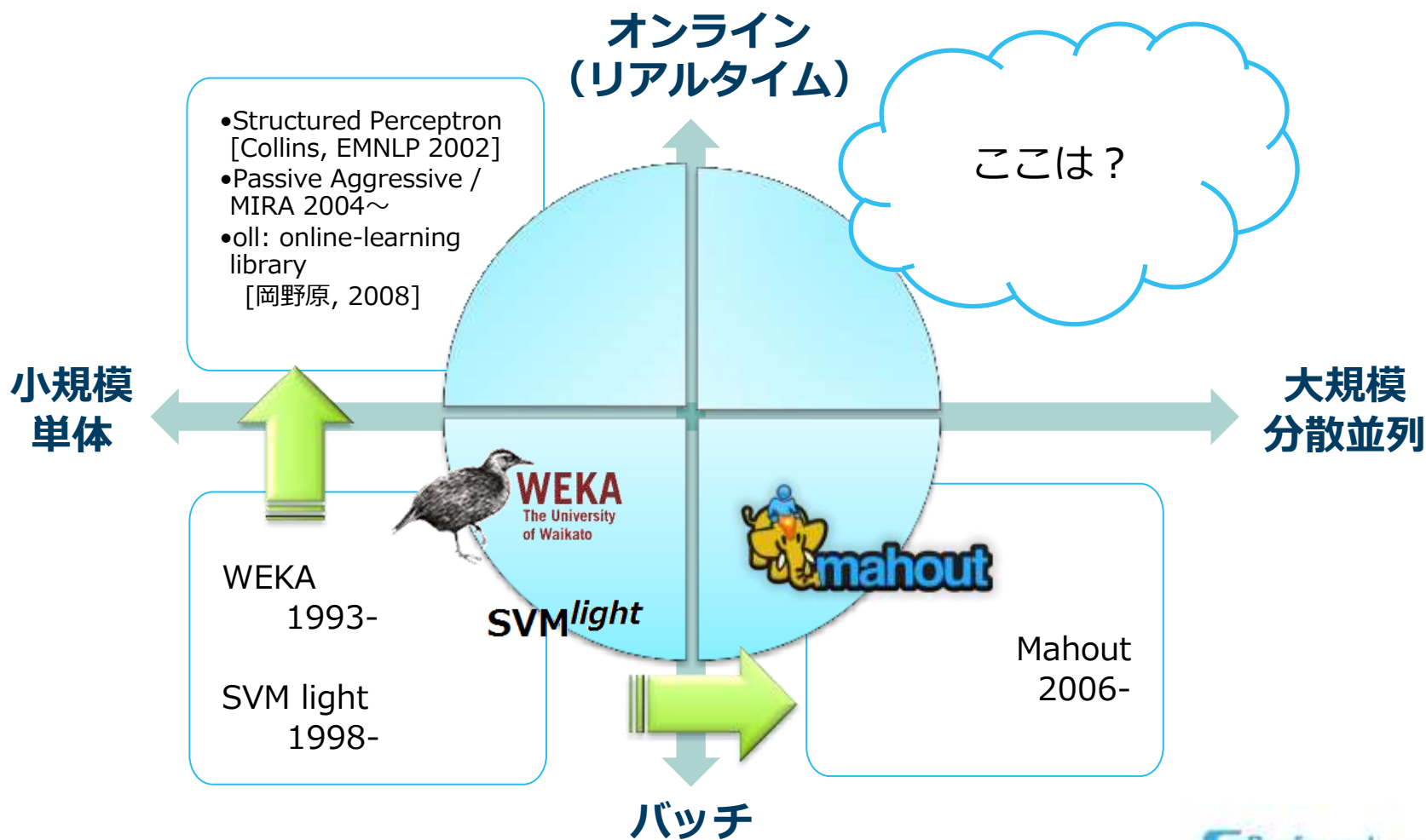


分析処理のオンライン化・ストリーム化

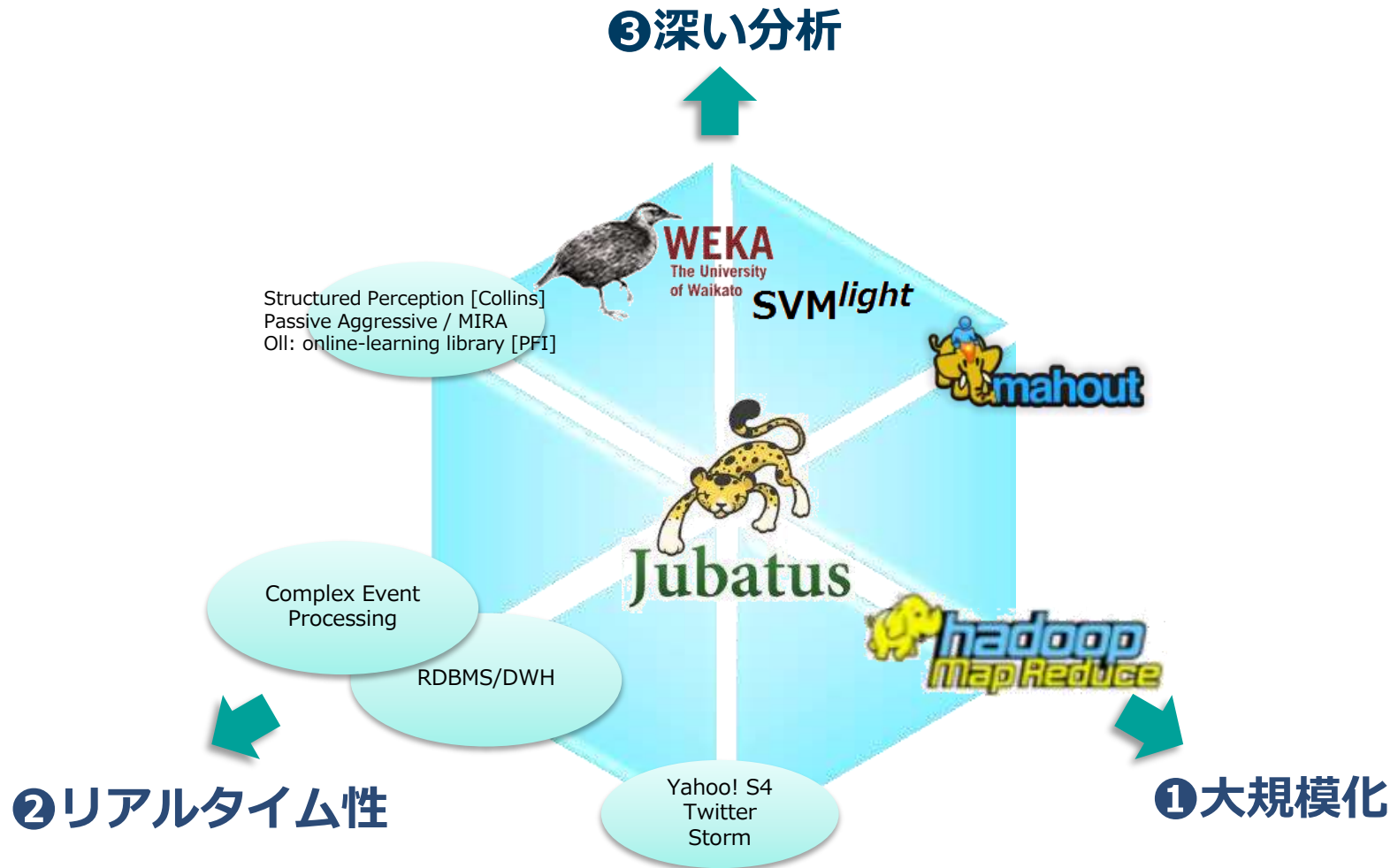


省スペース・効率的なアルゴリズムを使う

大規模データに機械学習を適用する、2つのアプローチ

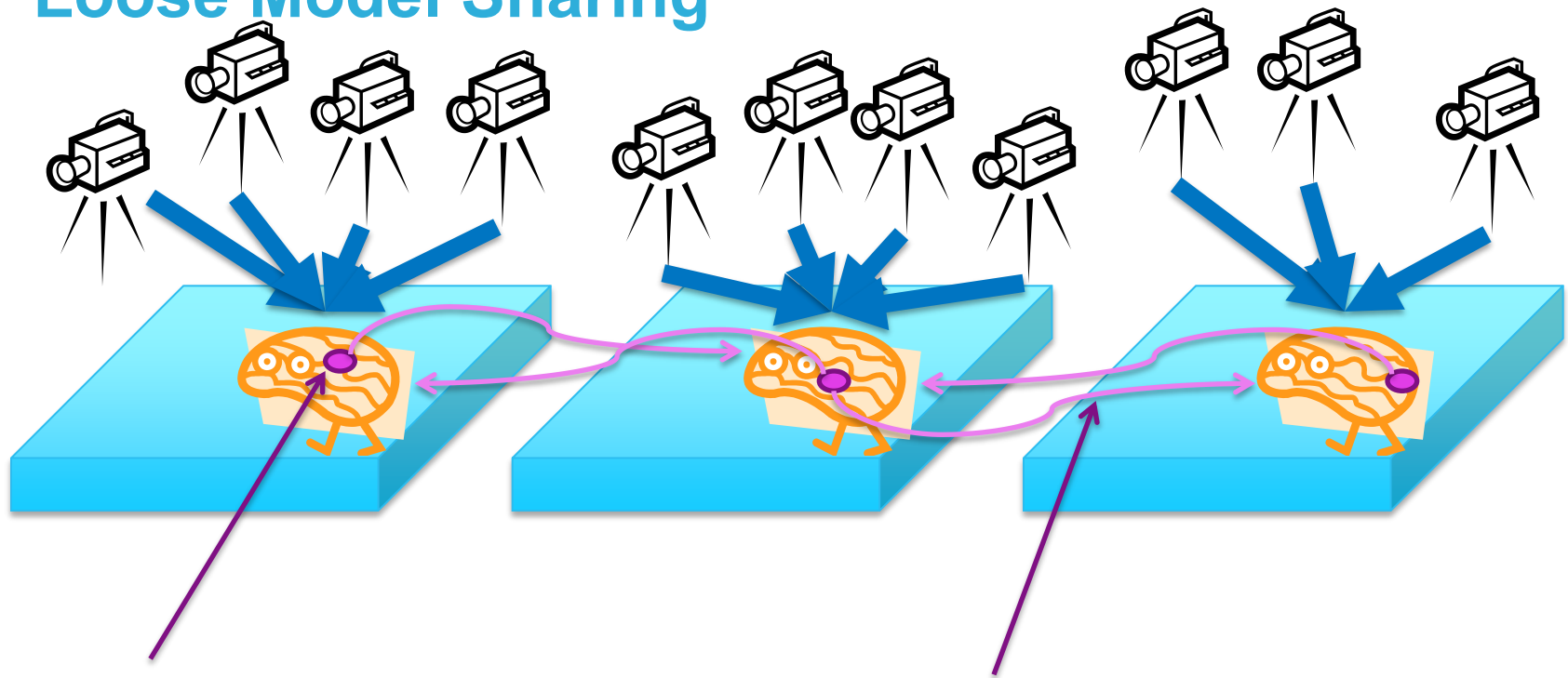


Jubatus



Jubatus supports Scalable Parallel Machine Learning

“Loose Model Sharing”

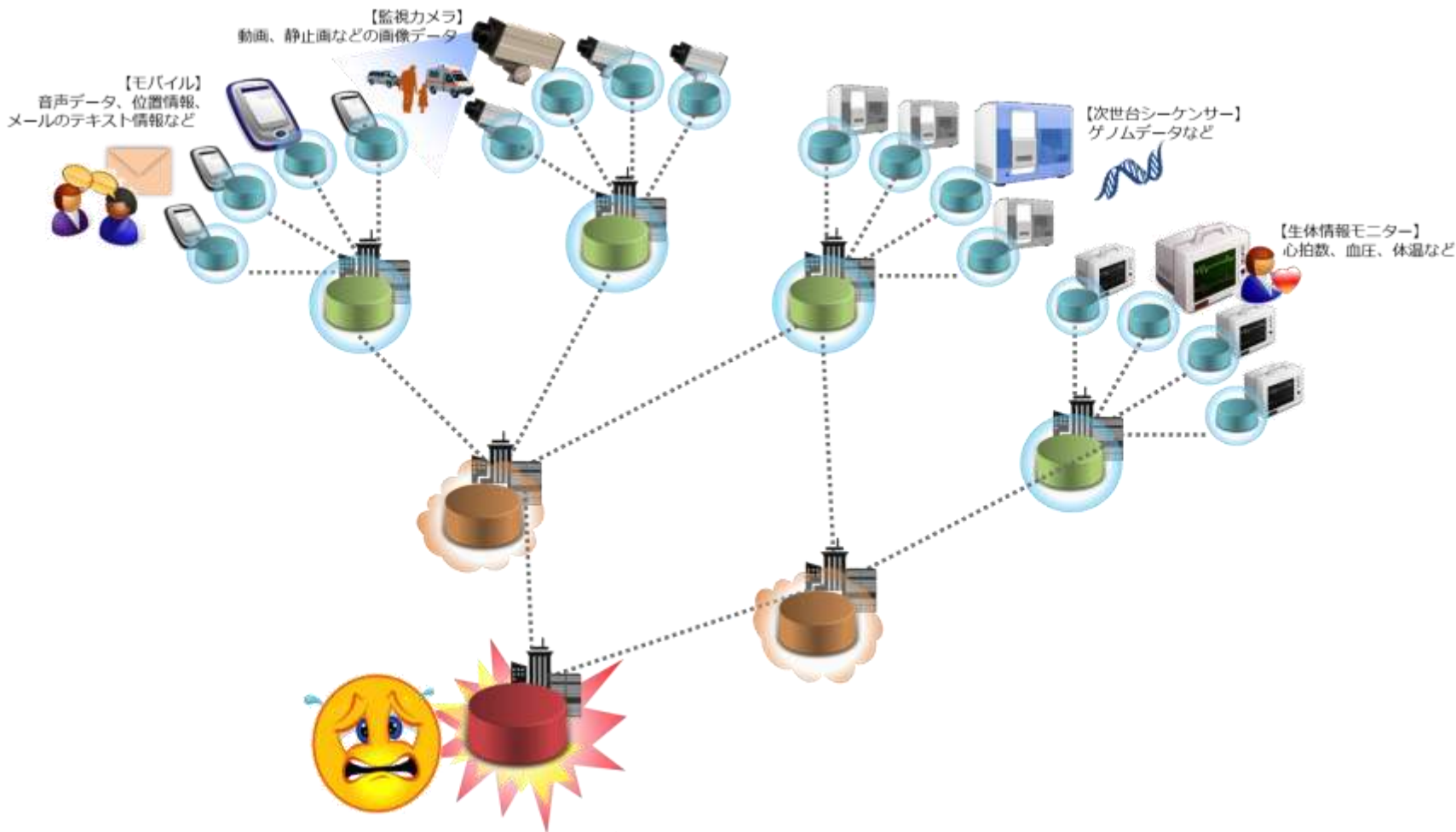


Each node updates its model incrementally by learning from raw data

Nodes only exchange differences between models in each node

データを集めるのではなく、
データが分散したまま分析・活用する

データを集約するアプローチは、 近い将来うまくいかなくなる



丸山宏, 岡野原大輔 Edge-Heavy Data: CPS・ビッグデータ・クラウド・スマホがもたらす次世代アーキテクチャ
GICTF総会 特別講演 2012, <http://www.gictf.jp/doc/20120709GICTF.pdf>

分野によっては、法律的な問題もある

- データを集めようとする、法律や政治の問題が避けられない
- それを解決しようとかんばっている内に、諸外国に技術的に遅れをとってしまう

エッジヘビーコンピューティング

1000 Petabytes/Year > 200 Petabytes

In Edge Devices

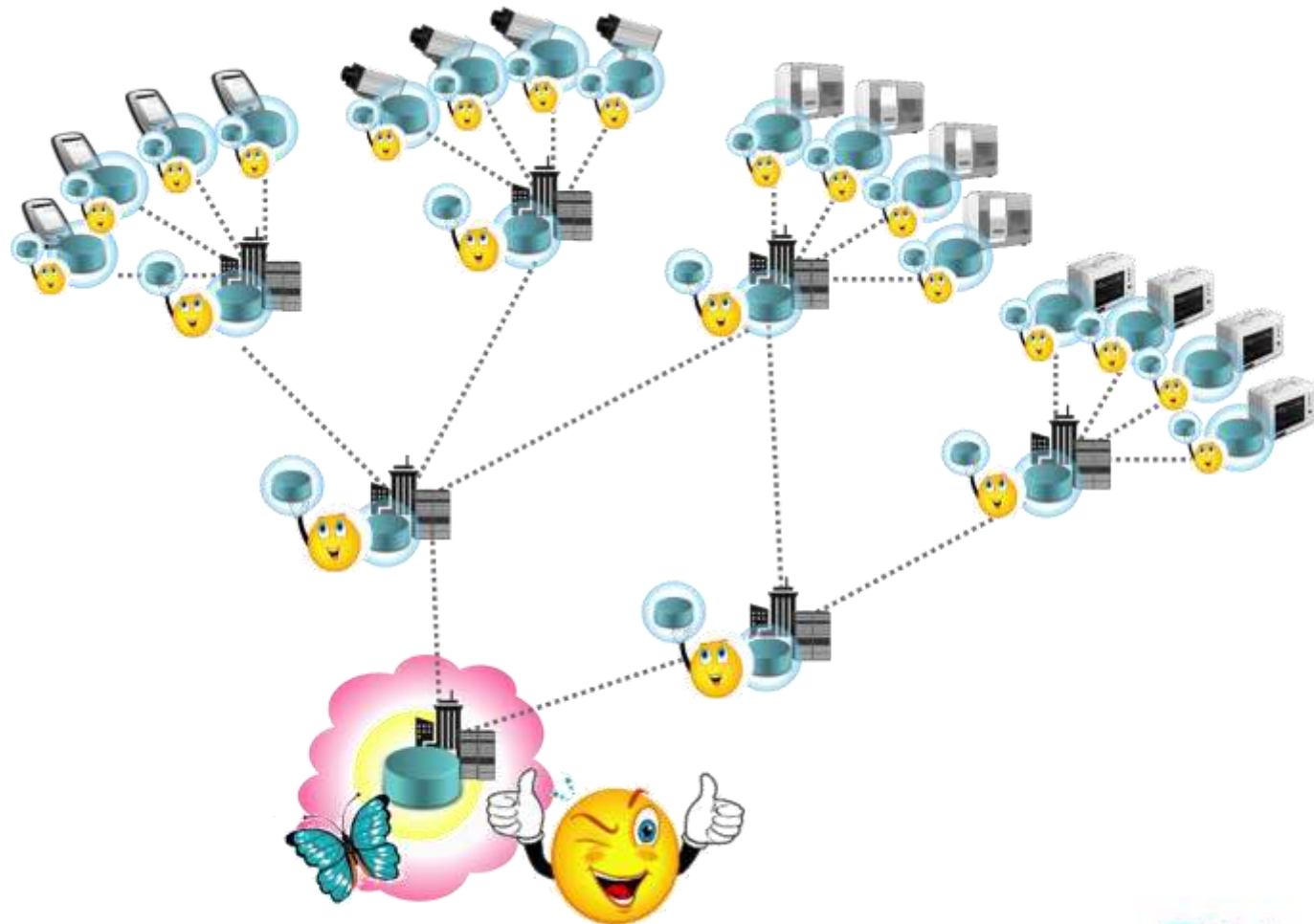
(Surveillance Cameras and Smartphones in Japan)

In Huge Computing Cloud

(300,000 nodes, each node has 2TB HDD, redundancy is 3)

データを「貯めない」、「一カ所に集めない」
この前提のもとで、
深い分析を実現するコンピューティングを
実現する。

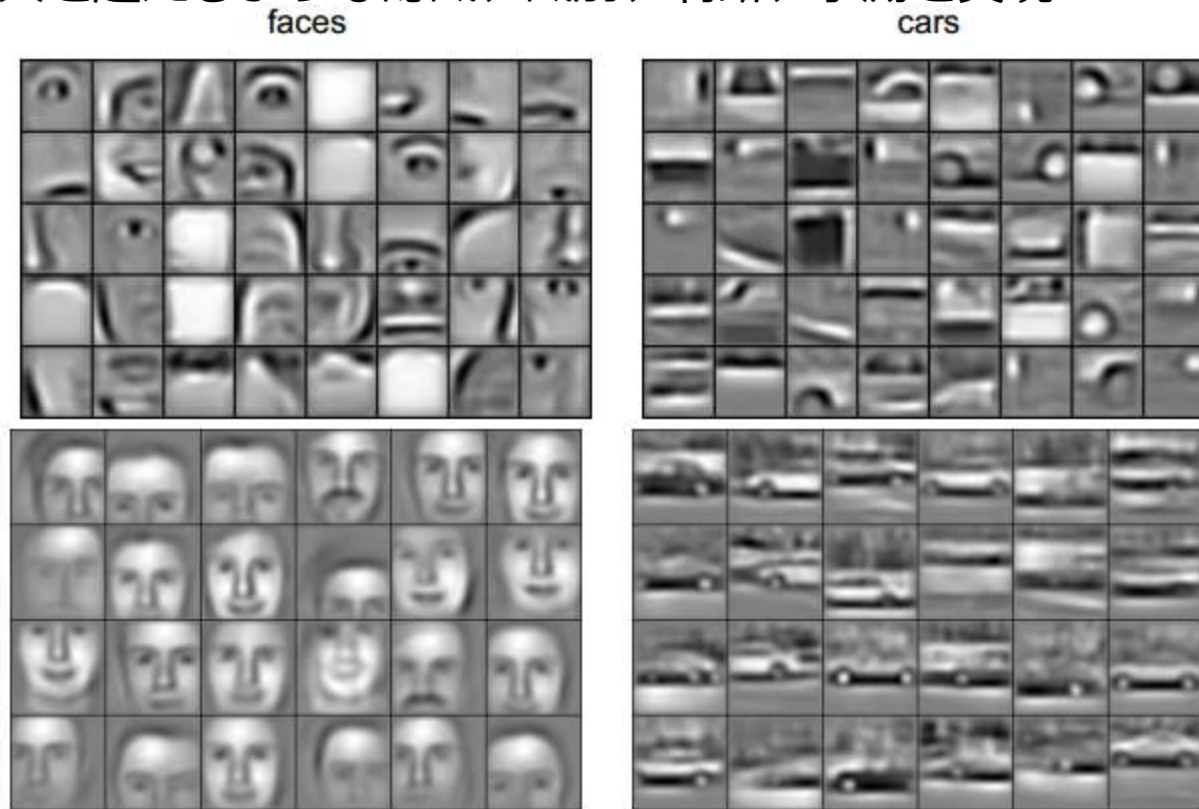
エッジ側が、もっとIntelligentになった、 階層型のアーキテクチャ



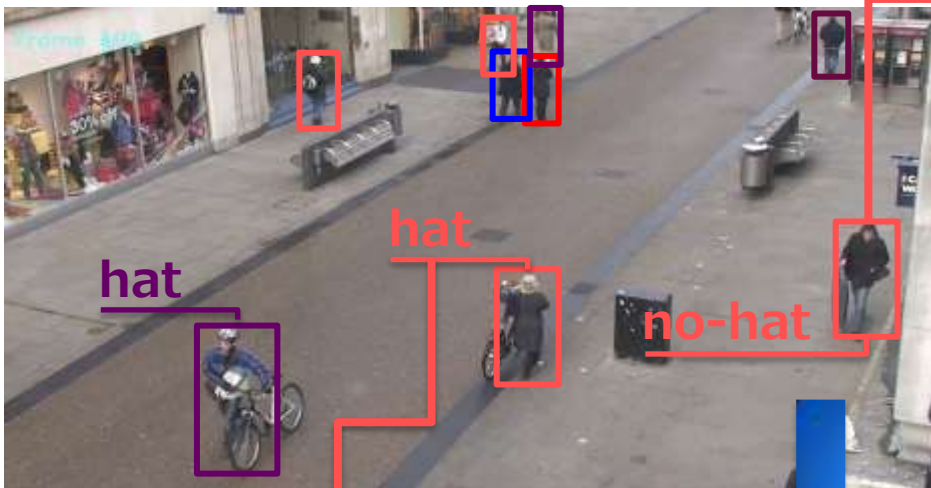
Deep Learningと エッジヘビーコンピューティング

深層学習 (Deep Learning)

- これまでの機械学習は分野毎に専門家による特徴抽出がされていた
- 特徴抽出すらも学習により獲得し、生のデータから直接認識
- 人を超えるような認識、識別、判断、予測を実現



Functions and Improvements of Visual Recognition



1. Attribute Classification

Date	Error of Each Tag		
	Raising Hands	Gender	Hat
Feb. 2014	19%	-	-
May. 2014	2%	<i>New!</i> 10%	<i>New!</i> 3%

- Improvement given by Deep Learning
- Will improve further with more learning data

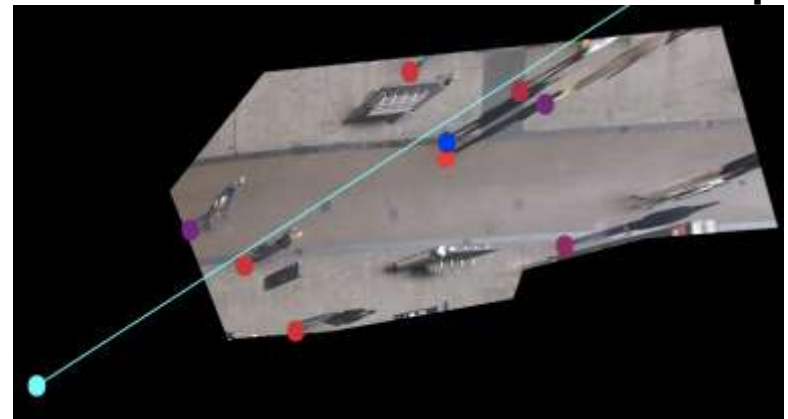
2. Person Detection

Date	Error @ 0.1 fppi*
Feb. 2014	40%
May. 2014	19%

*fppi means false positives per image.

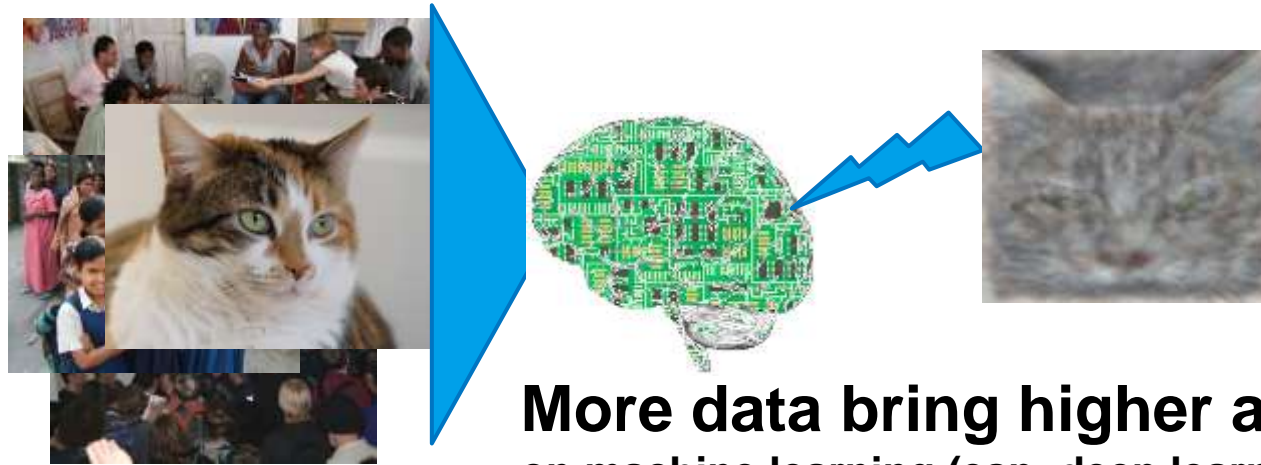
- Will improve further with more learning data

3. Position Estimation on the Map



- Higher accuracy compared to other position estimation methods based on e.g. Wi-Fi strength, geomagnetism
- Can be improved by using multiple cameras (to be added)

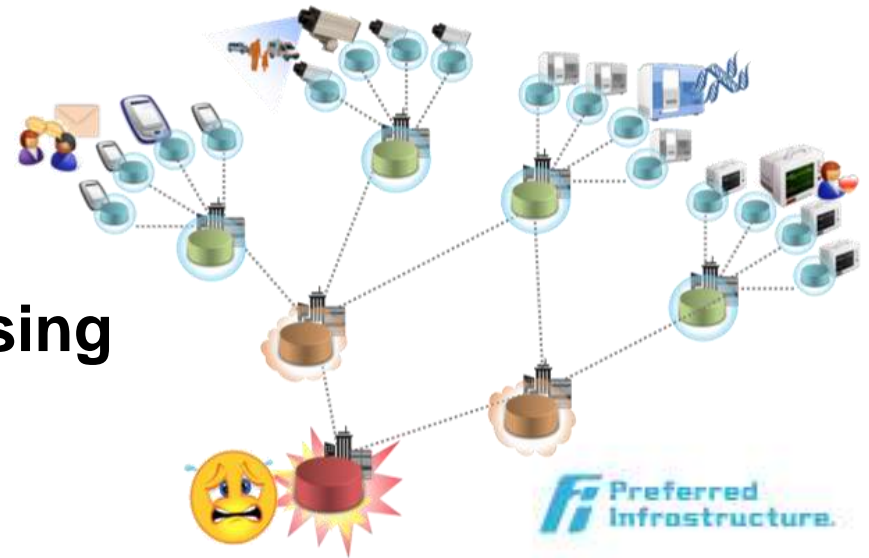
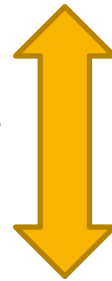
Evolution of IoT/loE occurs together with Evolution of Machine Learning



More data bring higher accuracy
on machine learning (esp. deep learning)

Distributed & Cooperative Intelligence can bridge

In IoT Era,
edge-heavy data is increasing enormously.





**Preferred
Networks**

IoT時代に向けた新しいコンピュータを創造する あらゆるモノに知能をもたせ、分散知能を実現する



Preferred Networks, Inc. (PFN)

- 設立：2014年3月
- 場所：本郷（東京都文京区）、San Jose (CA, USA)
- 取締役：西川徹、岡野原大輔、長谷川順一
- ミッション：

IoT時代に向けた新しいコンピュータを創造する
あらゆるモノに知能をもたせ、分散知能を実現する

- ▶ IoT事業にフォーカスするため、
株式会社Preferred Infrastructure（2006年3月創業）よりスピンオフした

オンラインとオフラインを「リアルタイム」でつなぐ



モノとオンラインが
リアルタイムに「協調」する



Copyright © 2006-2011

[Preferred Infrastructure All Right Reserved.](#)