

PFI 全体セミナー
2013/5/30 @ PFI

一般向けの Deep Learning

岡野原 大輔

株式会社Preferred Infrastructure
hillbig@preferred.jp

DeepLearning

- 専門家向けの情報はいろいろな場所で手に入る
- DeepLearning.net
- Google+ DeepLearning Group
- 人工知能学会 連載解説「Deep Learning（深層学習）」
 - 各分野の利用方法、歴史、理論、実装について
 - 私も第3回実装編を書いています
- Deep Learningの話は一部の機械学習屋向けの話なのか？
→否
- 今日は一般向けにDeep Learningを説明してみます

ブログで書いたらとても注目された

Preferred Research

Brought to you by Preferred Infrastructure, Inc.

ニューラルネットの逆襲

11月
1

By 岡野原 大輔 | Category: 解説 | No Comments

 Tweet 944  B! 559  いいね! 353

岡野原です。Deep Learningが各分野のコンペティションで優勝し話題になっています。Deep Learningは7、8段と深いニューラルネットを使う学習手法です。すでに、画像認識、音声認識、最も最近では化合物の活性予測で優勝したり、既存データ・セットでの最高精度を達成しています。以下に幾つか例をあげます。

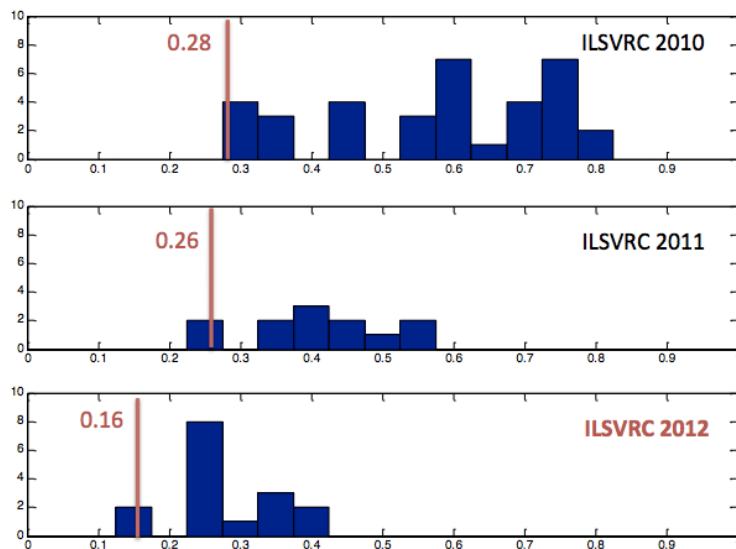
画像認識

- LSVRC 2012 [[html](#)] 優勝チームスライド [[pdf](#)], まとめスライド [[pdf](#)]

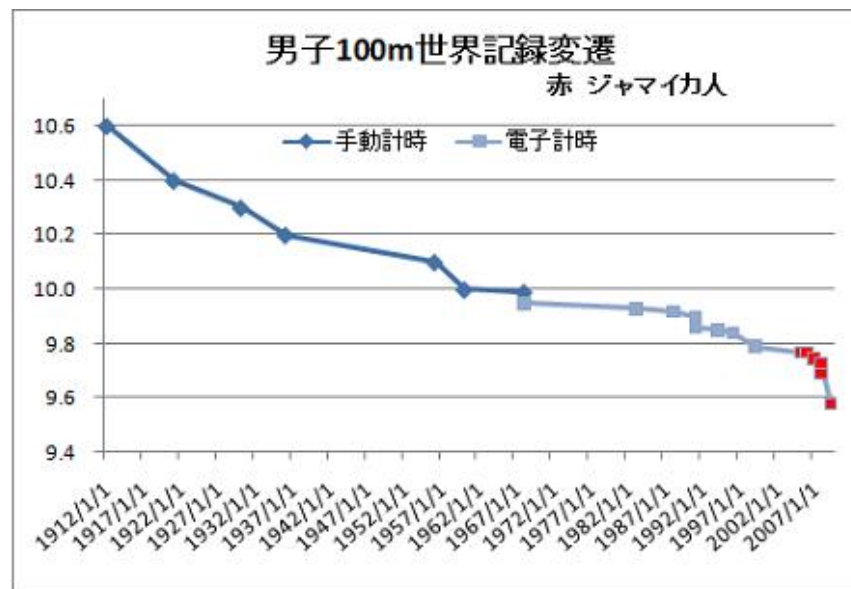
あれから7ヶ月

2012年に衝撃が走る

- 画像認識・音声認識・化合物の活性予測でDNNが最高精度を達成
 - 2位以下をぶっちぎる
 - 100mの世界記録におけるボルトのような立場
 - 非専門家が達成したことにも衝撃
 - 「化合物の活性予測は初めてだけど、試したら出ちゃった」



Error (5 predictions/image)



あらゆる分野でDeep Learningの話がでてくる






- 学会での嵐
 - 主要学会のチュートリアルで紹介
 - ICML（機械学習）, ACL（自然言語処理）, CVPR（画像処理）
 - 学会も誕生！ICLR (元々snowbird workshop)
- ビジネスでの嵐
 - Microsoftが音声認識をDeepLearningベースに置き換える
 - おそらく、Googleなど他の企業も置き換える
 - 中心人物のHinton+ポスドクが立ち上げた会社をGoogleが即買収
 - Googleのエンジニアの中心人物のJeff Deanも携わる
- 新聞などの記事でも取り上げられる

一般物体認識の例 (LSVRC2012優勝チーム)

一般物体のカテゴリ認識

間違えた場合も、納得の
できる間違い方

(右下のtape playerを
携帯電話と間違えている)

																							
lens cap	abacus	slug	hen																				
<table border="1"> <tr><td>reflex camera</td></tr> <tr><td>Polaroid camera</td></tr> <tr><td>pencil sharpener</td></tr> <tr><td>switch</td></tr> <tr><td>combination lock</td></tr> </table>	reflex camera	Polaroid camera	pencil sharpener	switch	combination lock	<table border="1"> <tr><td>abacus</td></tr> <tr><td>typewriter keyboard</td></tr> <tr><td>space bar</td></tr> <tr><td>computer keyboard</td></tr> <tr><td>accordion</td></tr> </table>	abacus	typewriter keyboard	space bar	computer keyboard	accordion	<table border="1"> <tr><td>slug</td></tr> <tr><td>zucchini</td></tr> <tr><td>ground beetle</td></tr> <tr><td>common newt</td></tr> <tr><td>water snake</td></tr> </table>	slug	zucchini	ground beetle	common newt	water snake	<table border="1"> <tr><td>hen</td></tr> <tr><td>cock</td></tr> <tr><td>cocker spaniel</td></tr> <tr><td>partridge</td></tr> <tr><td>English setter</td></tr> </table>	hen	cock	cocker spaniel	partridge	English setter
reflex camera																							
Polaroid camera																							
pencil sharpener																							
switch																							
combination lock																							
abacus																							
typewriter keyboard																							
space bar																							
computer keyboard																							
accordion																							
slug																							
zucchini																							
ground beetle																							
common newt																							
water snake																							
hen																							
cock																							
cocker spaniel																							
partridge																							
English setter																							
																							
tiger	chambered nautilus	tape player	planetarium																				
<table border="1"> <tr><td>tiger</td></tr> <tr><td>tiger cat</td></tr> <tr><td>tabby</td></tr> <tr><td>boxer</td></tr> <tr><td>Saint Bernard</td></tr> </table>	tiger	tiger cat	tabby	boxer	Saint Bernard	<table border="1"> <tr><td>lampshade</td></tr> <tr><td>throne</td></tr> <tr><td>goblet</td></tr> <tr><td>table lamp</td></tr> <tr><td>hamper</td></tr> </table>	lampshade	throne	goblet	table lamp	hamper	<table border="1"> <tr><td>cellular telephone</td></tr> <tr><td>slot</td></tr> <tr><td>reflex camera</td></tr> <tr><td>dial telephone</td></tr> <tr><td>iPod</td></tr> </table>	cellular telephone	slot	reflex camera	dial telephone	iPod	<table border="1"> <tr><td>planetarium</td></tr> <tr><td>dome</td></tr> <tr><td>mosque</td></tr> <tr><td>radio telescope</td></tr> <tr><td>steel arch bridge</td></tr> </table>	planetarium	dome	mosque	radio telescope	steel arch bridge
tiger																							
tiger cat																							
tabby																							
boxer																							
Saint Bernard																							
lampshade																							
throne																							
goblet																							
table lamp																							
hamper																							
cellular telephone																							
slot																							
reflex camera																							
dial telephone																							
iPod																							
planetarium																							
dome																							
mosque																							
radio telescope																							
steel arch bridge																							

ニューラルネット(NN)の歴史 (1/2)

- 1940年代頃から何度もブームが
 - Perceptron, BackPropagation, ...
 - しかし90年代頃からの長い冬の時代
- 2006年からDeep Neural Netとしての復活
 - Hinton, BengioらによるPreTraining とAutoEncoderの登場
 - 深い階層を持った場合でも勾配が拡散せず学習できる
 - Ngらによる視覚野の働きに似た画像特徴抽出
 - 人の視覚認識の仕組みを部分的に再現できた
 - しかしまだ一部の研究者のみが注目している状況

ニューラルネット(NN)の歴史 (2/2)

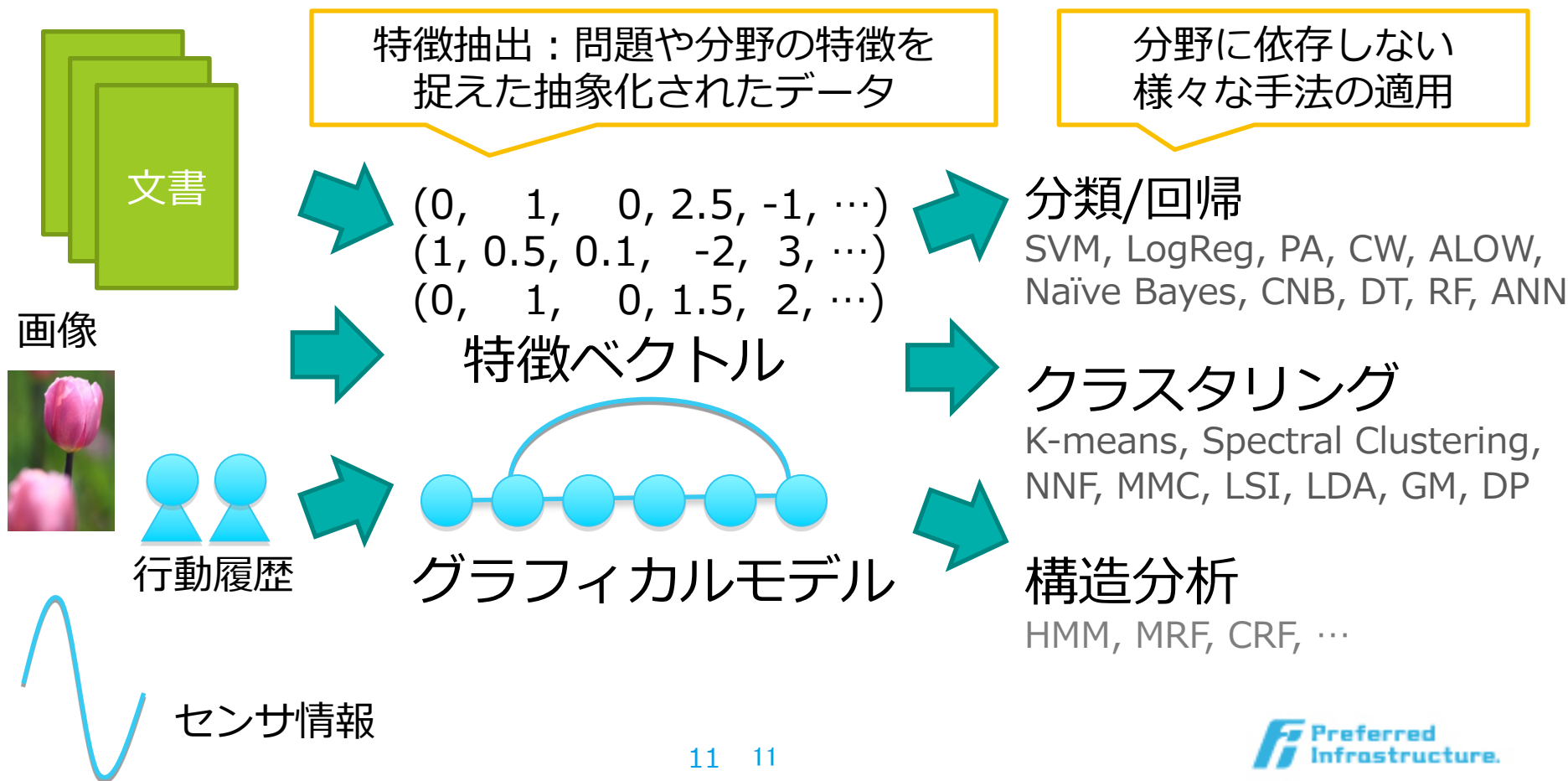
- 2010年以降の発展
 - 多くの分野でのベンチマークテストによる**圧勝**
 - 一般物体画像認識、音声認識、薬物活性予測
 - これまでのstate-of-the-artを大きく凌駕すると共に、非専門家が達成したことに衝撃
 - 大規模NNからの教師無学習（後述）
 - Googleらのは1000万枚の画像を利用してパラメータ数が数十億からなる大規模NNを2000台（16000コア）で1週間で学習
 - 教師無し学習で知識のようなものが得られる

なぜ深層学習がこれほど成功したか

- Hinton曰く：「Bengioらが90年代に培った手法」
+ 「大規模データ」 + 「DropOut」
 - 特に大きなブレークスルーがあったわけではない
- 学習手法の改善
 - PreTraning, AutoEncoder, Dropout, Maxout、学習率調整
- 実装技術の改善
 - GPGPUや大規模クラスタの利用
- ニューラルネットはこれまで注目されていなかっただけ
+ これまでの学習手法の煮詰まり感
 - 既存手法（線形分類器、ベイズ、カーネル法）に関してやれることはほぼやった。特徴抽出は比較的手付かず

機械学習が急速に広まった理由 特徴表現と分析処理の分離

- 異なる分野で同じ機械学習手法が適用可能



特徴抽出

- 入力データから特徴を抽出し特徴ベクトルで表す
 - テキスト、画像、音声、数値
 - 各領域の専門家による職人芸であった

特徴ベクトル

世の中ではビッグデータというキーワードが注目されていますが、一口にビッグデータといっても立場や観点によって定義は様々です。



ビッグデータ	2
世の中	1
キーワード	1
定義	1

周囲が黄色	1
中心が茶	1
花びらの割合	0.7
高さ	150

分析結果

IT関連

ひまわり

他の要素は0とする

DeepLearning

深層学習による特徴抽出（＝表現学習）

- 機械学習は次の2つのステップからなると説明した
 - STEP1 入力データからの特徴抽出
 - STEP2 特徴に対する学習・推論
- 特徴抽出は今でもドメイン知識や人手による試行錯誤が必要
 - 自然言語処理、画像処理、行動履歴 ... 毎に異なる技
 - Feature Engineeringとも呼ばれる一種のアート
 - どれを使うか、どう組み合わせるのか、値はどうするのか
- 特徴抽出も自動化できないか？
 - 特徴抽出は機械学習の実用上なボトルネック
 - 人手は本当に最適か、さらに改善できないか？

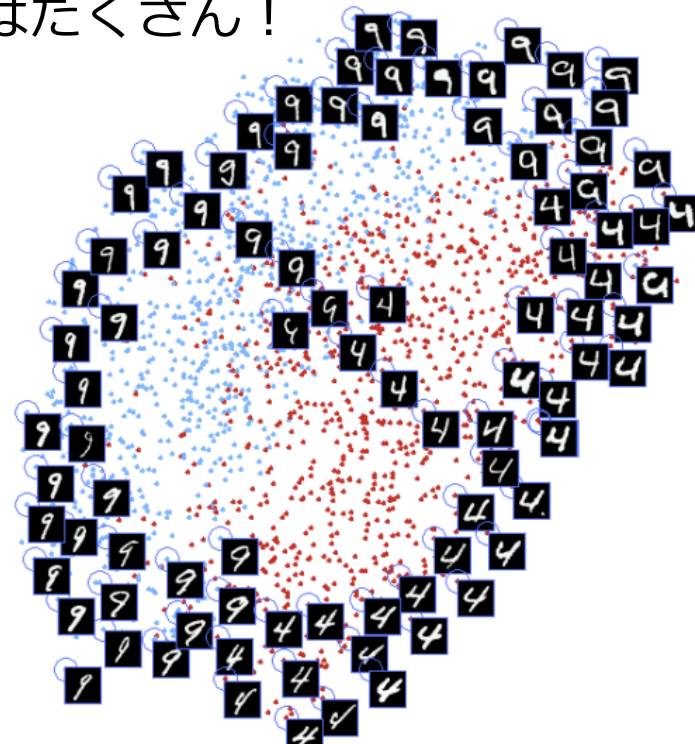
⇒ **表現学習**

表現学習

「実世界の観測情報から本質的な情報を抽出し表現する」方法を学習

- 世の中のデータは一見、複雑で量が多く処理が困難なように見える
- 本質的な情報を抽出し表現できればメリットはたくさん！
 - その後の問題が簡単な問題となる
 - 可視化しやすい
 - 一般化しやすい
 - 原因もわかりやすい
 - データサイズも少ない

手書き文字データ(元々256次元)
2次元で表現した場合

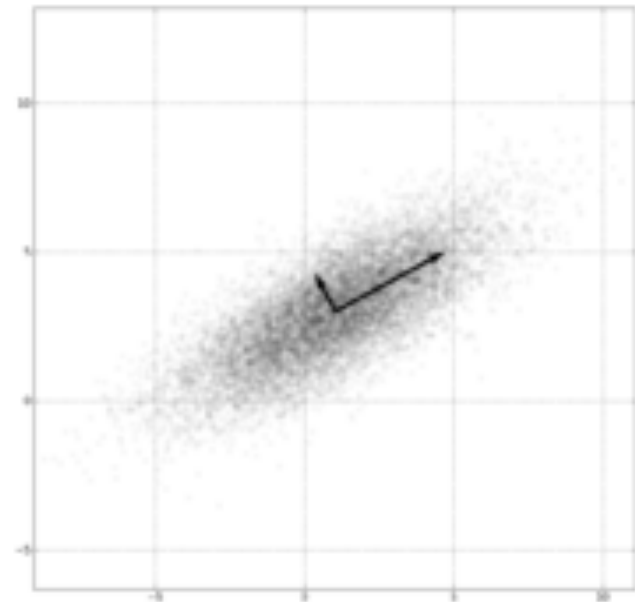


情報表現の代表選手

主成分分析 (PCA: Principal Component Analysis)

- 高次元データの集合を少数の主成分のベクトルで表す
- 各行がデータに対応する行列 X に対する特異値分解が $X = U\Sigma V^T$ の時、 U が主成分ベクトル、 $V\Sigma^T$ が主成分値
- 元のデータを二乗誤差最小での近似に対応

- `redsvd`を使えばコマンド1発で巨大な行列もPCAできるよ！

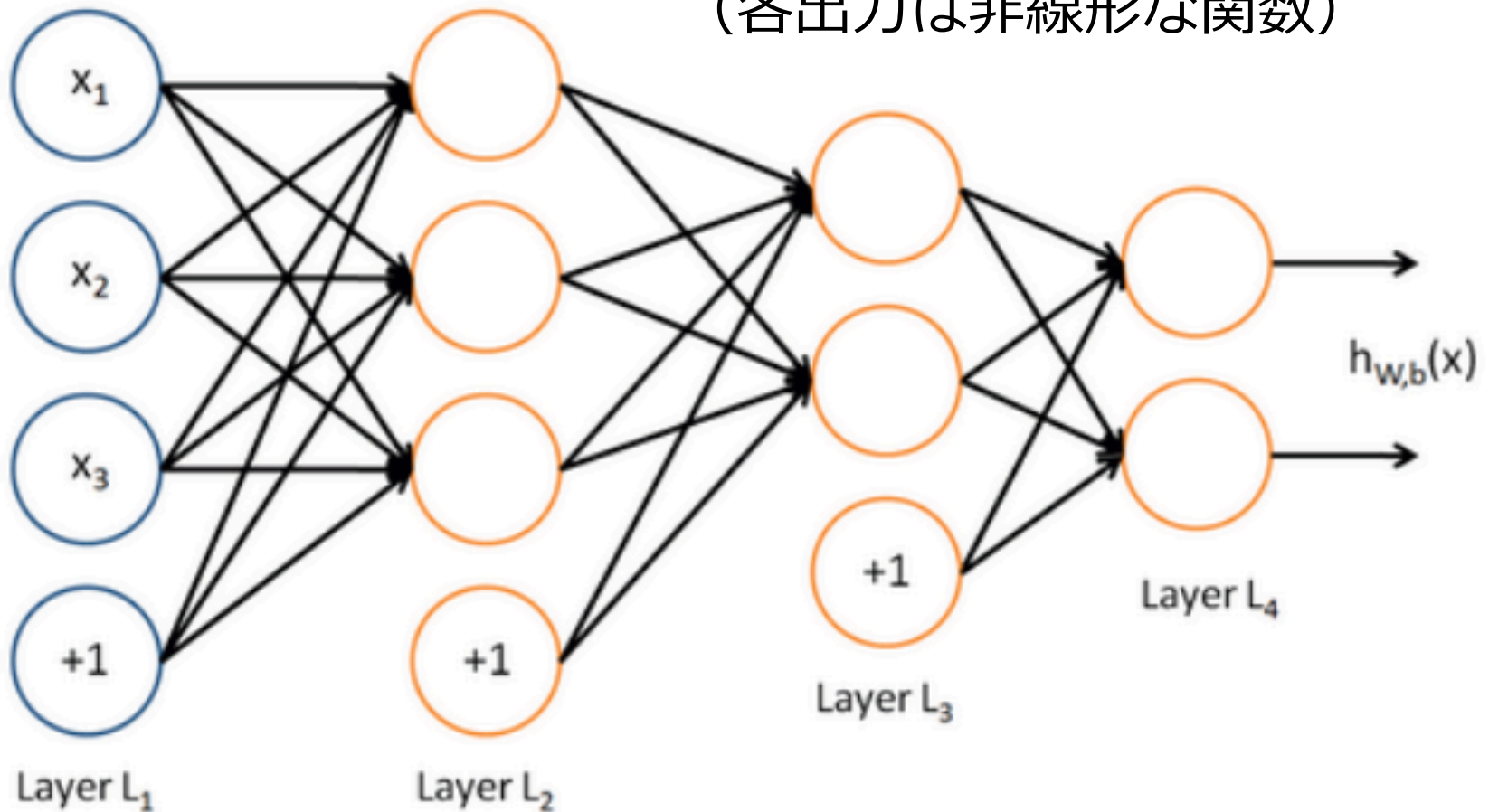


情報表現

- データ圧縮
 - データをうまく表現することにより、情報を失わず同じ情報をより短い符号長で表す
- データ変換
 - フーリエ変換・ウェーブレット変換
 - Burrows Wheeler変換
- データ要約
- クラスタリング

深層学習の場合

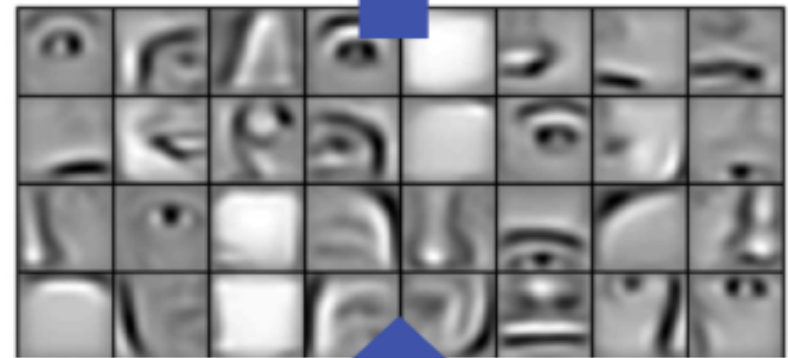
階層的に情報表現をより
「総合的」かつ「抽象的」
な情報に変換していく
(各出力は非線形な関数)



深層学習の場合 顔の学習の例

- 深いレイヤーはパーツを組み合わせた全体の学習

深いレイヤー
全体的な特徴
より抽象的なもの



- 浅いレイヤーは単純なパーツを学習
浅いレイヤー
小さい部品
より具体的なもの

表現学習をとことんやったら？

DistBelief [J. Dean+, NIPS 13]

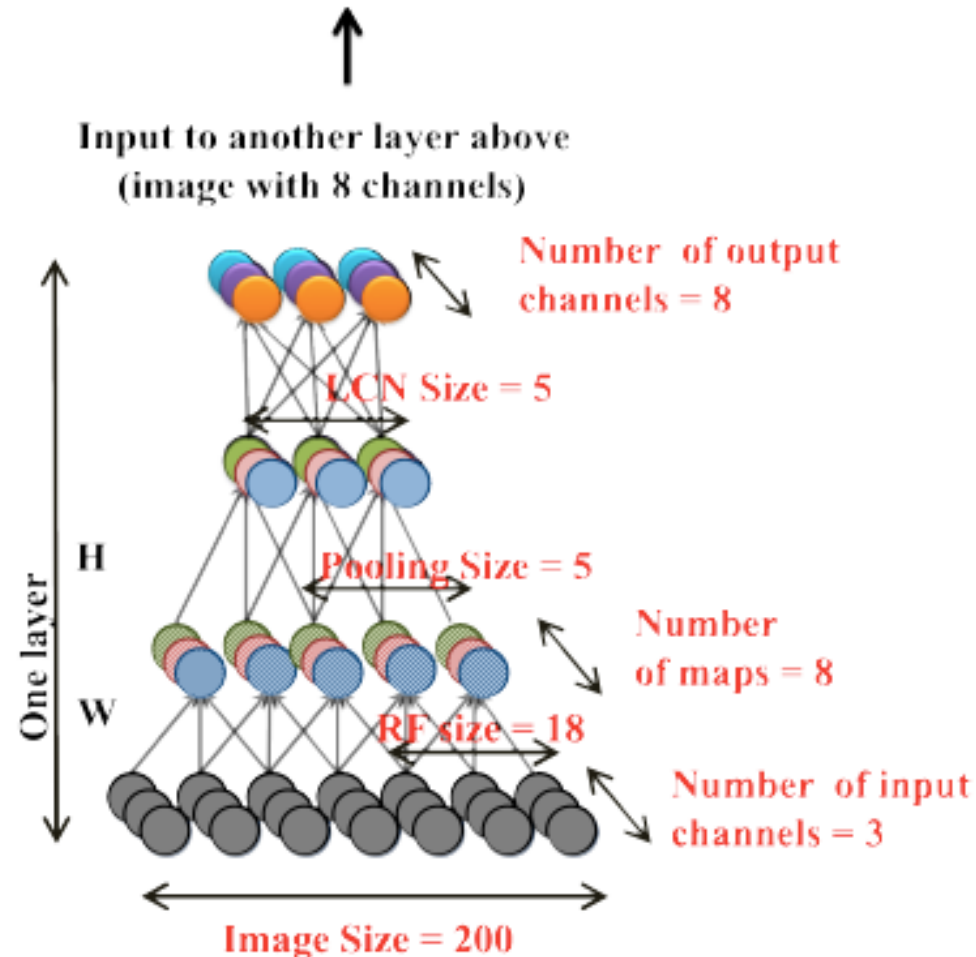
非同期での大規模学習をサポート

- オンラインの並列学習 (SGD)
- バッチ学習の並列学習 (L-BFGS)
- 16000コアでの並列学習が可能

Youtubeから得られた 200 x 200の
画像 1000万枚に対して教師無し学習

(AutoEncoderと似たRICAと呼ばれる
方法でDNNを学習)

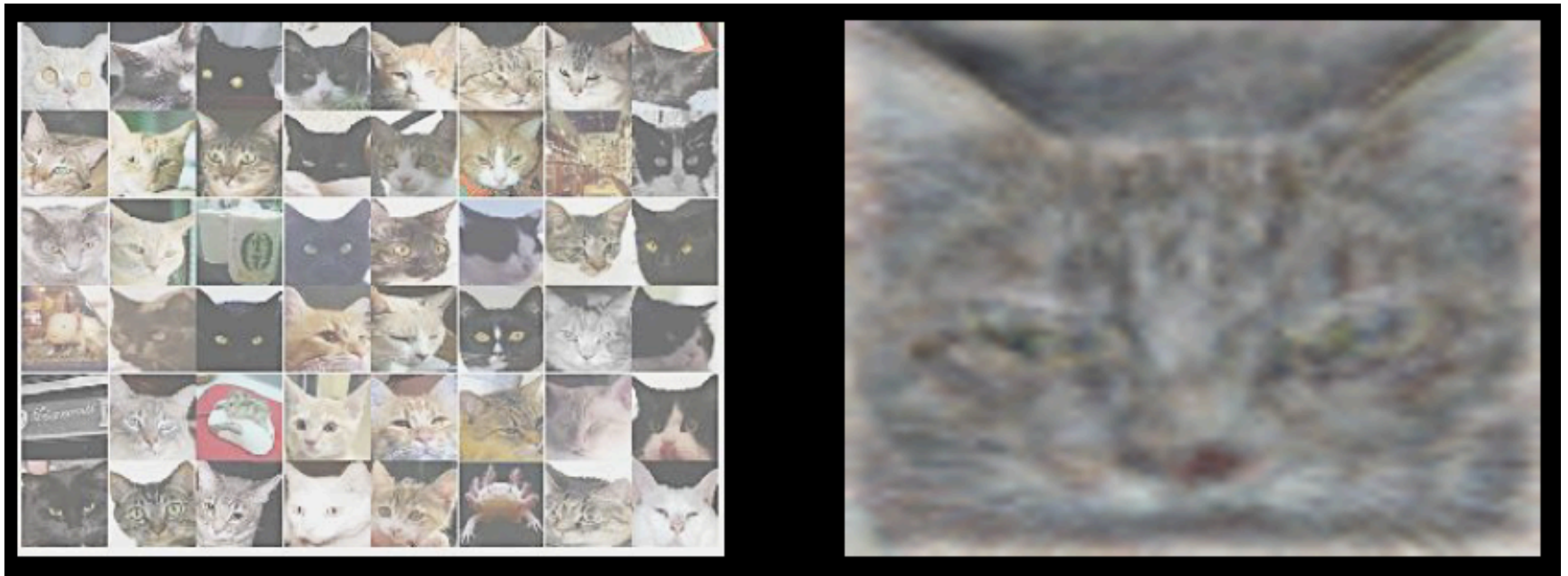
次項以降で学習で得られた
最上部のニューロンが反応する画像
を紹介



右：学習で得られたニューロンが最も反応する画像
左：このニューロンに反応したテスト画像



右：学習で得られたニューロンが最も反応する画像
左：このニューロンに反応したテスト画像



自然言語の場合 [Richard+]

- 別PDFを参照

深層学習による情報表現の獲得

- 層の数が多いNNによって、観測データから本質的なデータを抽出した、内部表現/潜在表現/特徴量を抽出する
- 一度抽出さえできれば簡単な線形分類器でも学習可能
- 一度抽出したらいろんなタスクに利用可能
 - 特定の課題に特化しているわけではないので
- 元々のデータに潜在的な構造があるため
 - 高次元の観測データが低次元の部分空間や多様体上に集中して分布
→データの分布に沿った座標系を取ることで効率的に表現可能
(主成分分析の例)

Disentangling

- 「もつれをほどける」「ほぐれる」という意味
- 元々のデータが高次元中で絡まっている状態を、うまくほぐして表現することで低次元で簡単な情報にする
 - 実際、その後簡単な線形分類器で高精度を達成する
 - 先程のGoogleの例はImage netでの最高精度を達成
- データに対する仮説：観測データを生み出してている情報源は少数のコンセプト・属性からなる
 - なので、そのような少数のコンセプト・属性を見つければ良い

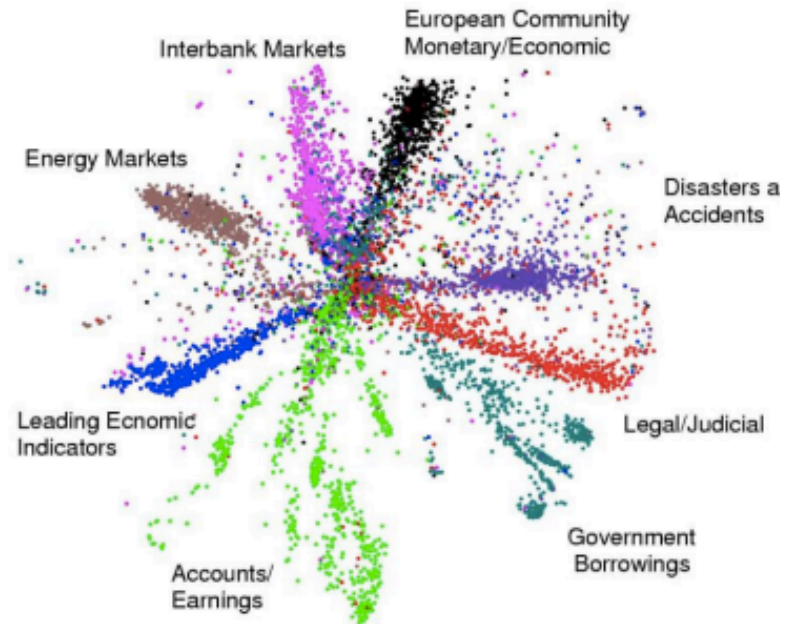
文書の2次元へのマッピング

右側のDNNベースの手法はきれいに分かれる

LSA 2-D Topic Space



Autoencoder 2-D Topic Space



DNNの信じる世界 = Principal of Compositionality

- 複雑な現象の意味は、それを構成する部品の意味の組み合わせである
 - 人間（生物）の多くの認識がこれを満たす
 - テキスト、画像、音声
- DNNはPrincipal of Compositionalityを利用している
 - 情報表現も、単純な部品の組み合わせで複雑な情報も表現できる
 - 部品は上位の複数の部品で共有される
 - これらの関係再帰的に効率よく学習できる

まとめ

深層学習 = 表現学習

- DNNにより、表現学習を行える
 - 人間の認識がうまくいく分野はうまく学習しているように見える
 - 構文解析, Sentiment Analysis
 - 音声認識
 - 画像認識、映像認識
 - 化合物分類（職人が見ると一発で分かる）
- しかし ...
 - 精度が欲しいだけの場合は必ずしもDNNじゃなくても良い
 - 3層程度のfully-connected NN + dropout + maxoutで最高精度
 - <http://rodrigob.github.io/>
- DNNの研究・実用化はまだ道半ば