

# 子どもの言語獲得・学習のモデル化とシミュレーションを目指して

竹内 彰一

人工知能・ソフトウェア技術研究センター  
千葉工業大学

# 人工知能・ソフトウェア技術研究センター

STAIR Lab. (Software Technology and Artificial Intelligence Research Laboratory)

2015年4月1日創立

所長：米澤明憲 東大名誉教授

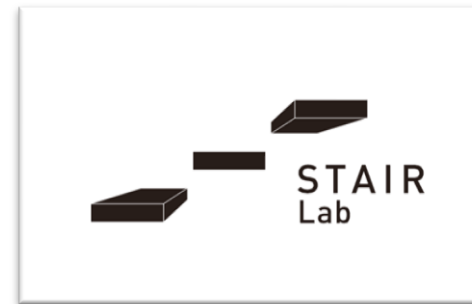
研究専門機関

- 学部・大学院とは独立の研究機関
- 人も重複なし
- 学生なし

研究テーマ

- 人工知能
- ソフトウェア技術

研究員 募集中



<http://stair.center>



<https://www.facebook.com/stairlab>

# 自己紹介： 竹内彰一

STAIR Lab. 主席研究員 工学博士

## 経歴

- 1979年東京大学大学院工学系計数工学修士卒
  - 神経回路網の研究
- 1979年三菱電機中央研究所
  - 1982年から86年まで第五世代コンピュータプロジェクト(ICOT)に出向
- 1991年ソニーコンピュータサイエンス研究所シニアリサーチャ
- 1996年ソニー株式会社アーキテクチャ研究所所長
- 2005年ソネット(株)シニアリサーチフェロー
- 2015年千葉工大 ステアラボ 主席研究員

人工知能、第五世代、ニューラルネットワーク、コンピュータヒューマンインタラクション、ビッグデータ解析、Webマイニング、ネットワークサービスの研究開発等、いろいろ

## 現在の研究テーマ

- 幼児の言語獲得、会話・経験を重ねるごとに賢くなる人工知能、ニューラルネットワーク全般、画像・シーン認識

子どもの言語獲得・学習のモデル化とシミュレーションを目指して

本日の話のあらすじ

- 人工知能、機械学習、DNN
- DNN
- NN言語モデル
- 子どもの言語獲得とDNN

# 人工知能

## 心の二つのモデル

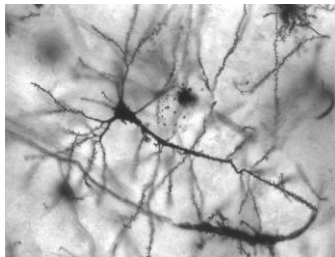
トップダウン / 記号論理アプローチ

$man(\text{“socrates”})$ .

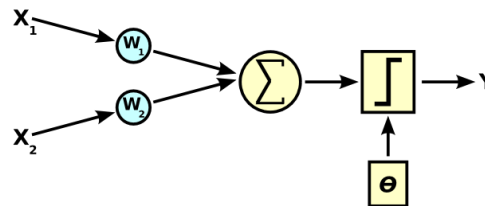
$\forall x$  s. t.  $man(x)$  **implies**  $mortal(x)$ .

?  $mortal(\text{“socrates”})$

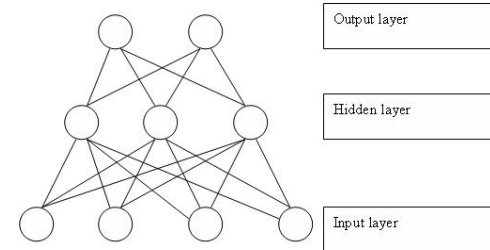
ボトムアップ / 脳モデルアプローチ



By MethoxyRoxy - Own work, CC BY-SA 2.5



By User:Yearofthedragon, CC BY-SA 3.0



By Anna Schroeder - Own work, CC BY-SA 3.0

# それぞれのアプローチ

## 二つの心のモデル

|            | 脳モデルアプローチ   | 記号論理アプローチ                    |
|------------|---|------------------------------|
| 目標         | 学習  | 問題解決                         |
| 目標達成のための方法 | 学習 = 多数のパラメータを持つ系の最適化<br>例からの自動的・機械的な学習<br>心の状態 = sub NNの状態 | 記号処理言語や論理で<br>問題解決方法をプログラミング |
| 概念とは       | NNの中のニューロンの活性化のパターン<br>分散表現 (ベクトル表現)                        | 記号                           |
| 思考とは       | NNの中で各ニューロンの活動が連鎖することで起きているだろう                              | 記号を操作することで実現できる              |
| ツール        | DNNシミュレーションフレームワーク  | 記号処理言語<br>定理証明系              |
| 得意分野       | 主に認識、制御<br>最近は言語もゲームも                                       | 推論<br>問題解決<br>言語             |

最近、得意分野が重なりつつある：AlphaGo、機械翻訳、など

# 人工知能の歴史

## トップダウン / 記号論理アプローチ

- ずっと主流

## ボトムアップ / 脳モデルアプローチ

- 1958年にPerceptron (1層NN) 発表 → 第一次ニューラルネットワークブーム
  - ❖ 1969年にミンスキーらによってPerceptronの限界が示されブーム終了
  - ❖ 単層では大したことができない → 多層への期待
- 1980年代前半、コネクショニストらがHopfield modelやbackpropagationを発表 → 第二次ニューラルネットワークブーム
  - ❖ 多層NNの研究が開花
  - ❖ 1980年代後半に終了。(2層、3層を超えるNNを訓練するのが難しかった)
- 1980年代中頃から機械学習研究が活発化
  - ❖ 「学習」にフォーカスした数理的アプローチ
  - ❖ DNNの数学的基盤形成に大きな貢献
- 2012年 ImagenetでHintonらのDeep CNNが圧勝したことをキッカケに → 第三次ニューラルネットワークブーム
  - ❖ Dataset → {1000 x 1000}, DNN → 22層

# 第三次ニューラルネットワークブームの中身

## 「教師あり学習 Supervised Learning」が格段の進歩

- 最適化問題として定式化し数値計算で答えを見つける方法が定着
  - {input, output}の組をたくさん与えてモデルを訓練する
  - 失敗を損失関数として定義して、これを最小化するようにモデルのパラメータを調整する
- 最適解は確率的最急降下法で探す
  - 訓練用のデータを繰り返し使いながらパラメータを調整する
  - **大量の訓練用データ → より良い局所解**
- 理論的にはほとんどの場合、局所最適解でつかまるが、**多層NNの場合、局所最適解がそれほど悪くもない**ことが最近わかってきた
- 何より計算機が第二次ブームの時とは大違い
  - ムーアの法則の恩恵
    - 30年間で $2^{20}$ 倍=**100万倍** (116年→1時間)
    - VAX11/780=**1DMIPS** → Intel Core i7 5960X=**238,310 DMIPS** (27年→1時間)
  - かつてない**多層のNN**、かつてない**大規模データセット**が可能に
  - タスクによっては人間に勝つことも



# Deep Neural Networks



# Deep Neural Networks

代表的な二種類のアーキテクチャ

## Convolution Neural Network

- 空間認識に優れ、画像認識への応用が多い
- 入力データから、1次特徴、2次特徴と順次特徴抽出を階層的に積み上げ、それらの特徴を使って上位の層で多層パーセプトロンの分類問題を解く

## Recurrent Neural Network

- 時間認識に優れ、言語モデルへの応用が多い
- 時系列データを隠れ層に圧縮して蓄積
- 問題(教師あり学習)に応じて蓄積内容が調整される

関数

プログラム

CNN

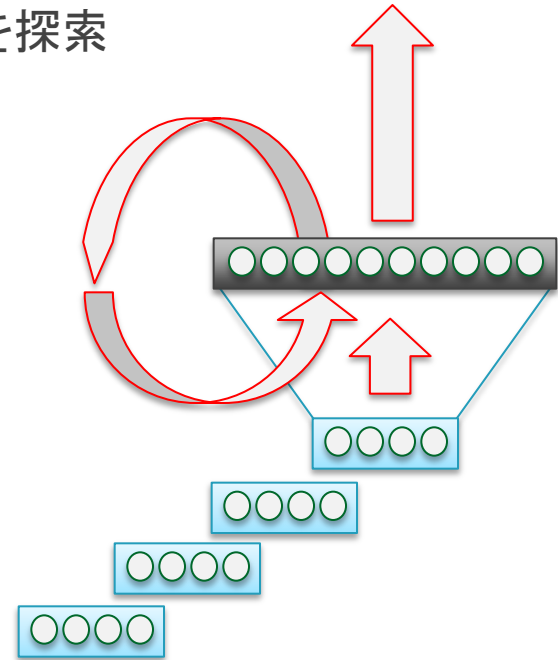
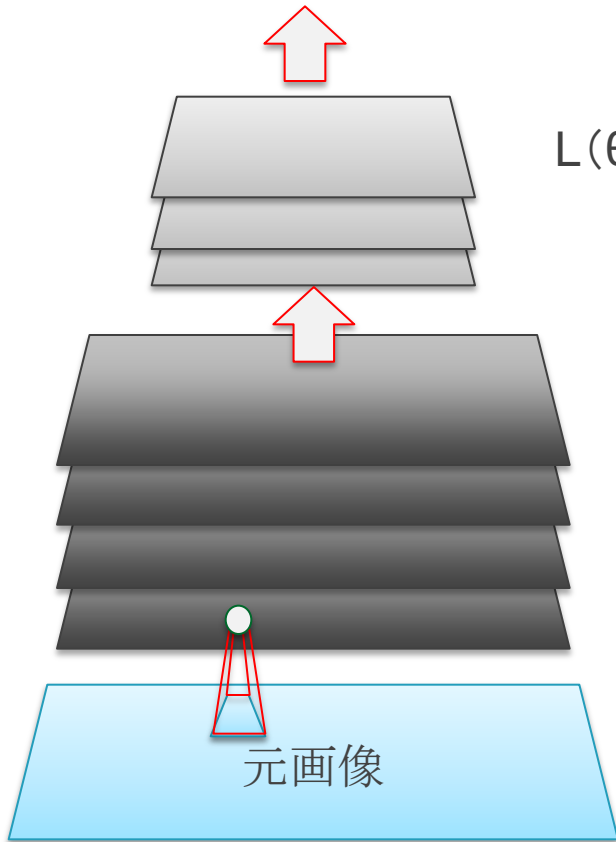
Supervised Learning

RNN

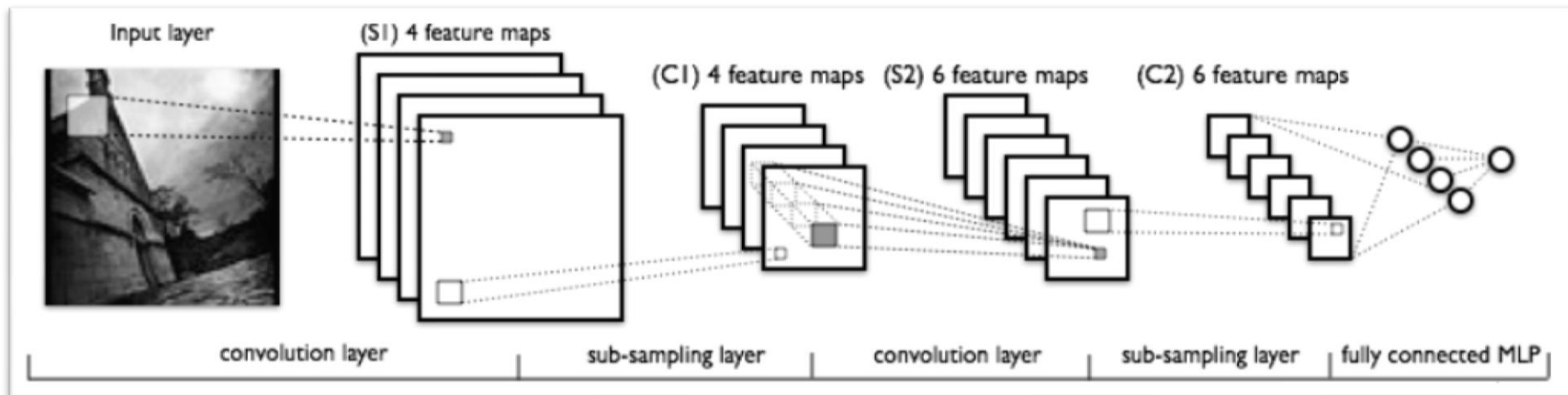
誤差関数  $L(\theta)$

$L(\theta)$ を最小にする $\theta$ を探索

Back propagation



# CNN



様々な局所的な特徴の抽出

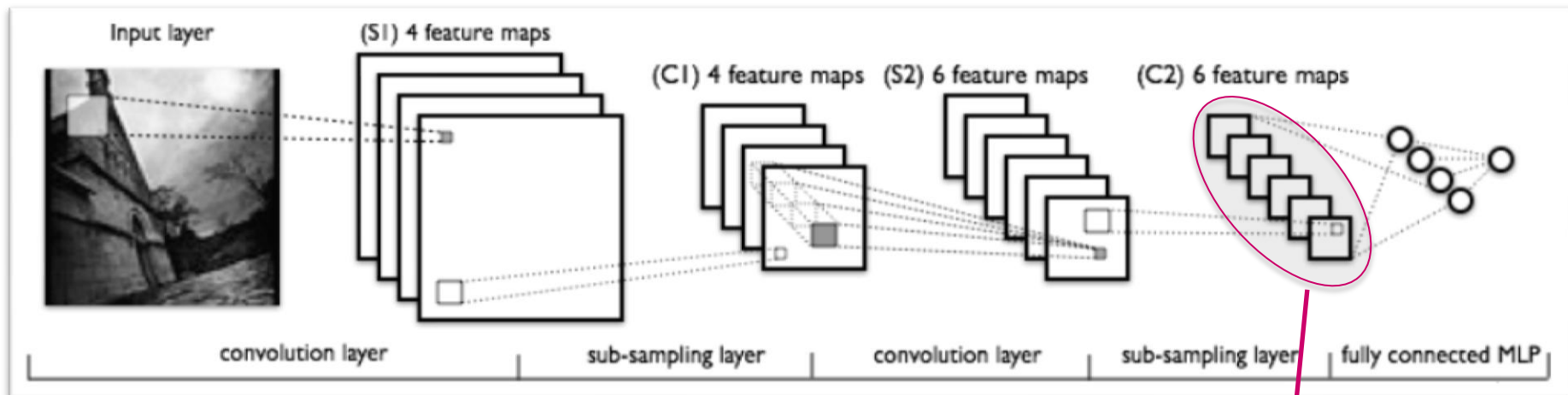
分類問題を解く

分類問題で高い精度を出すためには  
どんな特徴を抽出したらよいか

従来の画像認識では専門家が  
問題に適した特徴を手作りで  
開発していた  
--- feature engineering ---

CNNは  
学習により  
抽出すべき特徴を  
決定する

# CNNは学習により抽出すべき特徴を決定する



様々な局所的な特徴の抽出

分類問題を解く

分類問題を解くために  
最適化された特徴集合

〔学習された画像の意味〕

# ImageNet Large Scale Visual Recognition competition

かつてない大規模なベンチマークを使ったテクノロジーコンペティション

## 画像を1000個のカテゴリーに分類するタスク

- カテゴリー数がそれまでのコンペより一桁多い

## 訓練データ

- カテゴリーごとに1000+枚の画像が与えられる
- これもこれまでのコンペにない数
- しかもFlickr等から集めた普通の画像
- 総画像枚数は128万枚

## 勝者

- 2012年にHintonらのSupersessionがNo.1になって以来、Oxford U., Google, MSのDeep CNNモデルがNo.1を維持
- 2015 ChampのTOP5 accuracyは 96.4%



n04270147\_1257.JPEG



n04270147\_1274.JPEG



n04270147\_1338.JPEG



n04270147\_1347.JPEG



n04270147\_1355.JPEG



n04270147\_1358.JPEG



n04270147\_1382.JPEG



n04270147\_1459.JPEG



n04270147\_1474.JPEG



n04270147\_1479.JPEG



n04270147\_1528.JPEG



n04270147\_1530.JPEG



n04270147\_1540.JPEG



n04270147\_1567.JPEG



n04270147\_1591.JPEG



n04270147\_1592.JPEG



n04270147\_1606.JPEG



n04270147\_1615.JPEG



n04270147\_1633.JPEG



n04270147\_1644.JPEG



n04270147\_1648.JPEG



n04270147\_1722.JPEG



n04270147\_1734.JPEG



n04270147\_1740.JPEG



n04270147\_1758.JPEG



n04270147\_1775.JPEG



n04270147\_1793.JPEG



n04270147\_1796.JPEG



n04270147\_1807.JPEG



n04270147\_1810.JPEG



n04270147\_1834.JPEG



n04270147\_1842.JPEG



n04270147\_1843.JPEG



n04270147\_1854.JPEG



n04270147\_1869.JPEG

# Imagenet 1000 categories

|         |     |        |    |         |   |
|---------|-----|--------|----|---------|---|
| dog     | 118 | 爬虫類    | 35 | 海洋生物    | 8 |
| 装置      | 77  | 楽器     | 28 | 両生類     | 8 |
| 道具      | 61  | 昆虫     | 27 | その他の生き物 | 7 |
| 鳥類      | 59  | 家具     | 24 | 蜘蛛      | 7 |
| vehicle | 55  | 人工物    | 24 | 武器      | 6 |
| 哺乳類     | 52  | 食べ物    | 22 | cat     | 5 |
| 建築物     | 51  | 船      | 18 | 靴       | 5 |
| 入れ物・容器  | 49  | 魚類     | 16 | 甲冑      | 5 |
| 衣類      | 45  | 自然物    | 15 | 織物      | 5 |
| 雑貨      | 44  | ボール    | 11 | 軟体動物    | 5 |
| 動物      | 43  | 台所用品   | 9  | カニ      | 4 |
| plant   | 38  | コンピュータ | 8  | 甲殻類     | 3 |
|         |     |        |    | 人間      | 3 |

犬だけで118種類、鳥だけで59種類  
きめ細かい認識 ( Fine Grained Recognition ) にも強い



# 簡単な応用：きめ細かな花の認識

ImageNetに植物は38種類含まれるがそのうち花は2つしかない

400種の花の認識へ適用する実験をやってみた

- ModelはImageNet用のモデル GoogleNet with Batch Normalization を使う
  - <https://github.com/soumith/imagenet-multiGPU.torch>
- 花画像はImageNetの元のDB (image-net.org) からダウンロード
  - 目標：一種類あたり1000枚以上の画像を集める
  - 残念ながらそういう種はなかなかない。700で妥協
  - ノイズ (誤分類) が多いのでそれを掃除する作業が必要

結果 → [flowers.stair.center](http://flowers.stair.center)

- 花の種類は407種類 (データソースの関係で欧米の花に偏っている)
- 精度は 99% (テスト用データの値)

ImageNetのモデルは汎用的にいろいろな画像分類タスクに使える

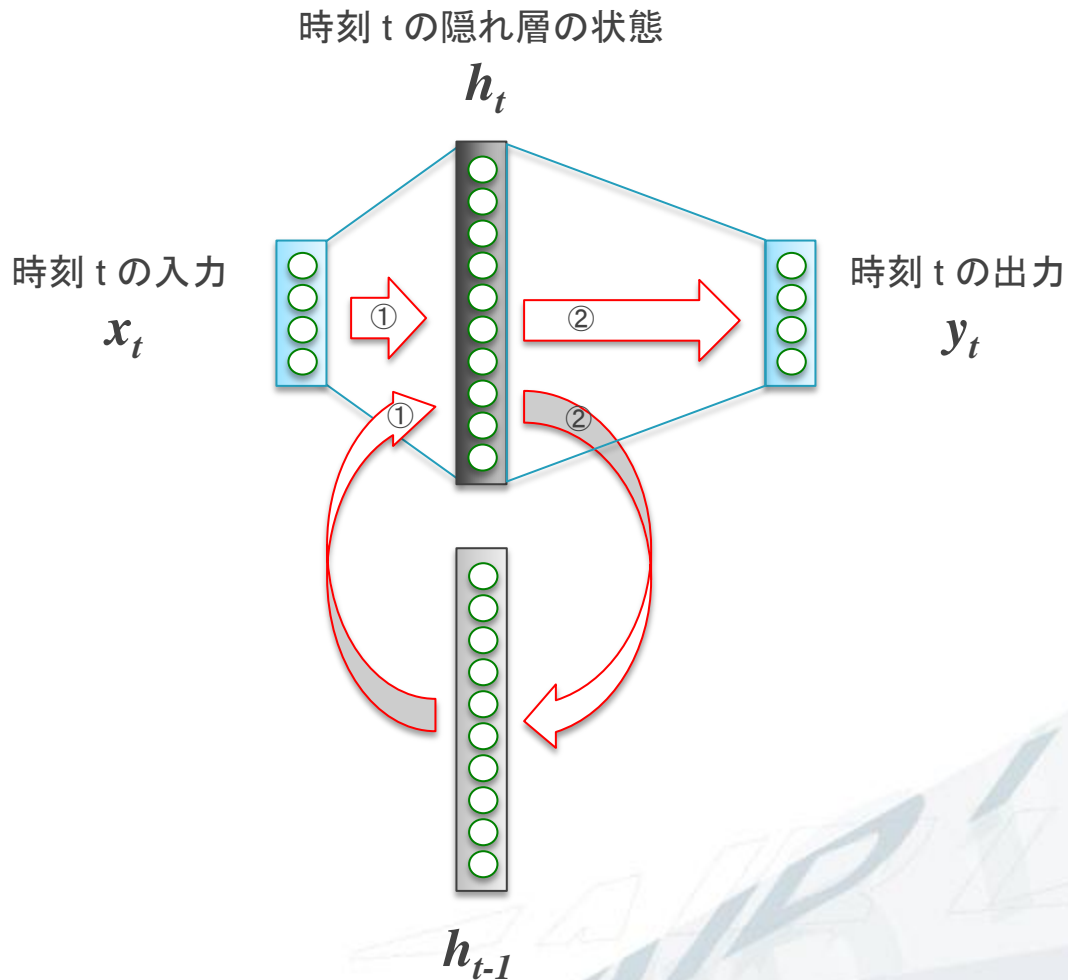
# RNN

可変長の入出力

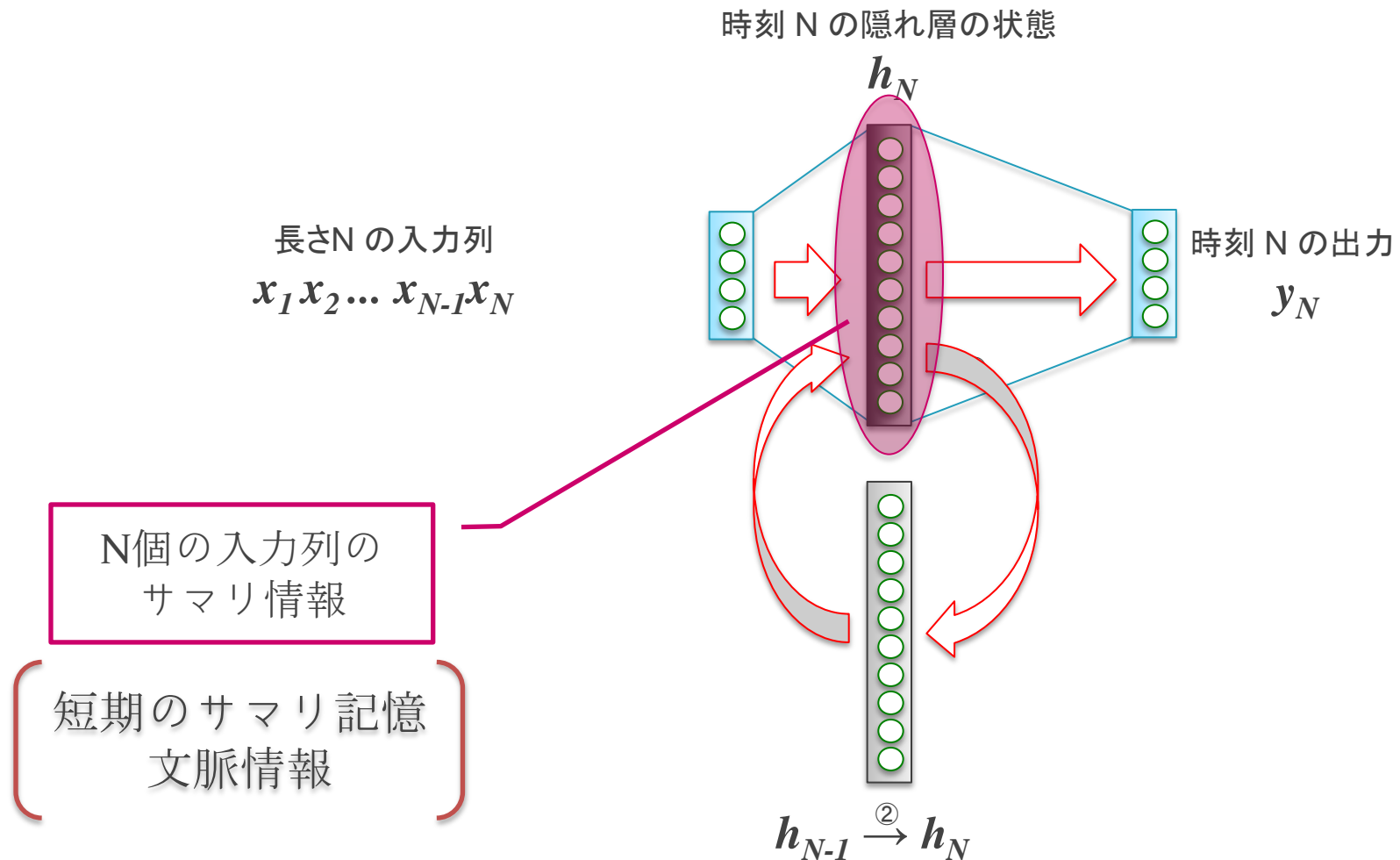
内部状態を保持

Turing complete

- 任意のProcedureを実現できる ( 適当に結線すれば )



# RNN は任意長の時系列のサマ리를 固定長のベクトルに記憶する



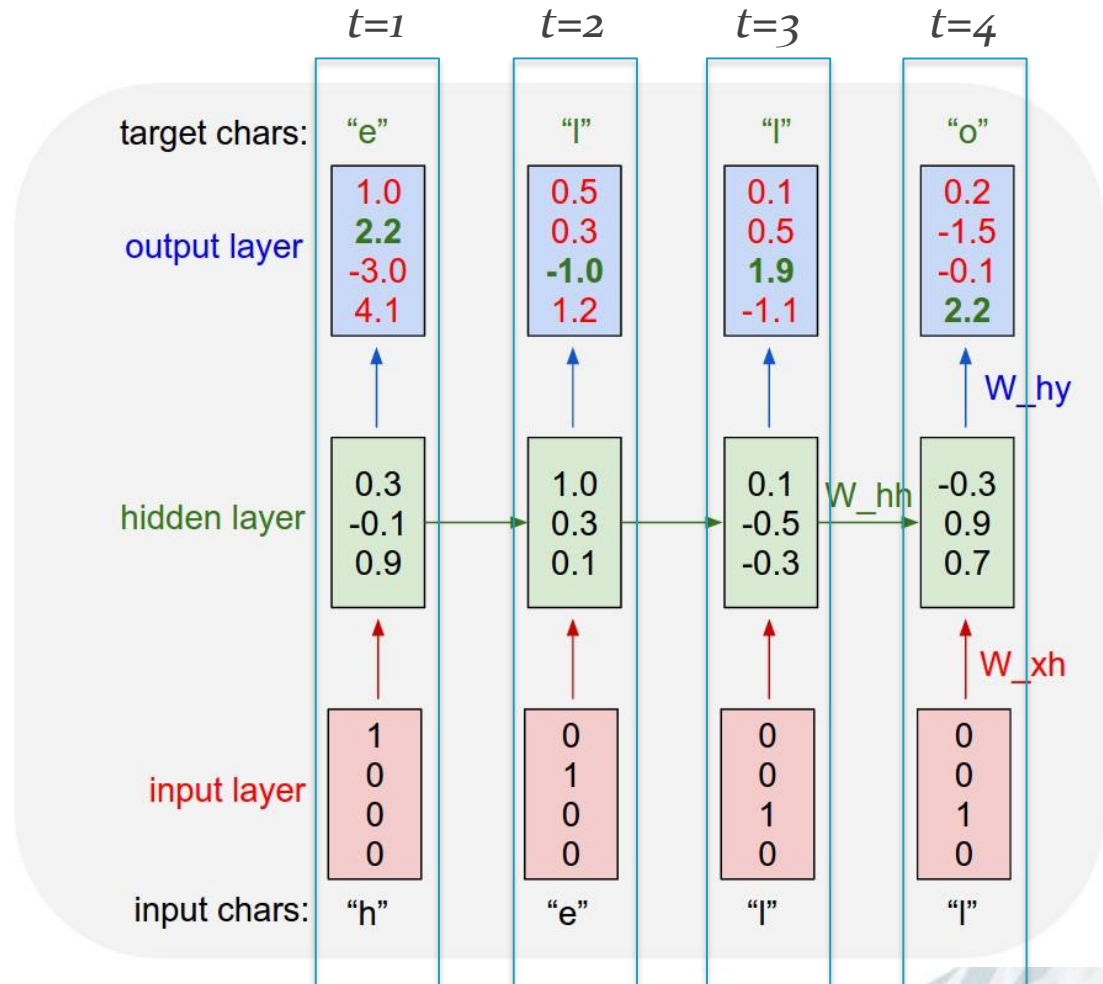


# Neural Networks 言語モデル

記号と意味

# RNNによる文字からの文生成

Andrej Karpathy



# トルストイ「戦争と平和」(英文)から文字のつながりを学習 文脈を考慮した文字と文字のつながりを学習する

■ [100] tyntd-iafhatawiaoirdemot lytdws e ,tfti, astai f ogoh eoase rranbyne 'nhthnee e plia  
tklrgd t o idoe ns,smtt h ne etie h,hregtrs nigtike,aoaenns lng

■ [300] "Tmont thithey" fomesscerliund Keushey. Thom here sheulke, anmerenith ol sivh l  
lalterthend Bleipile shuwy fil on aseterlome coaniogennc Phe lism thond hon at. MeiDimorotion  
in ther thize."

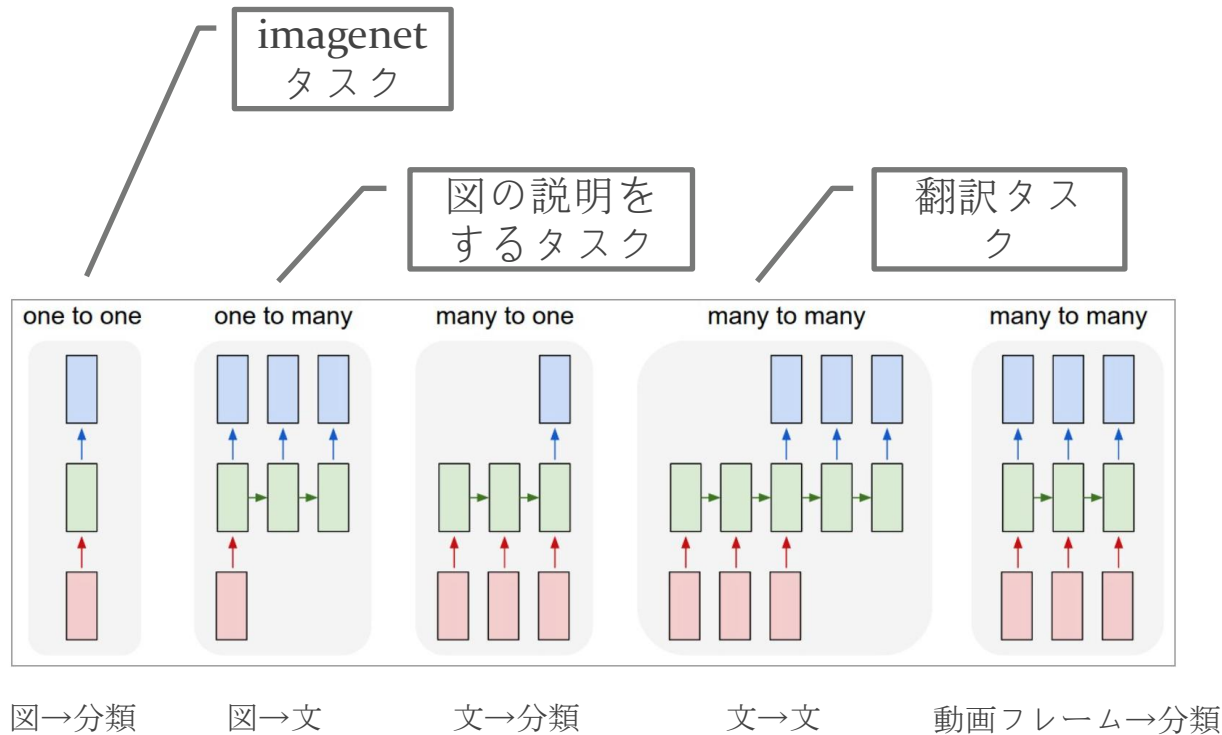
■ [700] Aftair fall unsuch that the hall for Prince Velzonski's that me of her hearly, and behs to  
so arwage fiving were to it beloge, pavu say falling misfort how, and Gogition is so overelical and  
offer.

■ [2000] "Why do what that day," replied Natasha, and wishing to himself the fact the princess,  
Princess Mary was easier, fed in had oftended him. Pierre aking his soul came to the packs and  
drove up his father-in-law women

スペース、句読点の使い方、単語綴りなどが徐々に学習されている

# Deep RNNのいろいろな言語タスクへの展開

## RNN Encoder – Decoder : Sequence to Sequence Mapping



The Unreasonable Effectiveness of Recurrent Neural Networks  
Andrej Karpathy

# RNN Encoder - Decoderによる翻訳

## Sequence to Sequence Learning with Neural Networks

I. Sutskever, O. Vinyals, Q. V. Le @google

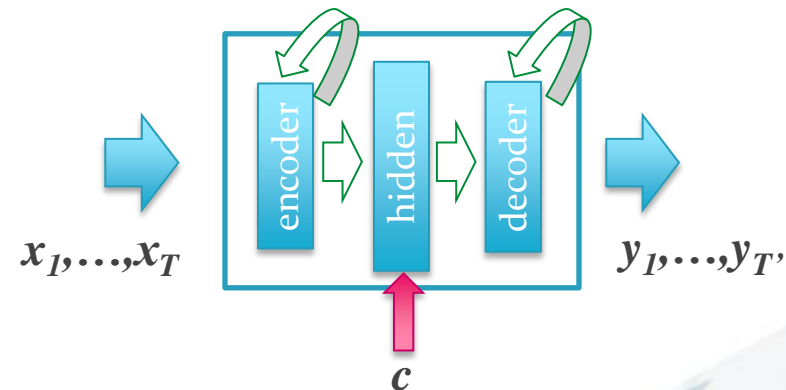


### Model

- 求めるもの =  $p(y_1, \dots, y_T | x_1, \dots, x_T)$ 
  - $x$  が原文、 $y$  が翻訳文、 $T \neq T'$
- 入力文をEncoder RNNに逆順にfeedし、最後の隠れ層の状態を  $c$  とする
- 入力文を全部 feedした後の意味ベクトルから出力文を Decoder RNN で順次生成
- $P(y_1, \dots, y_{T'} | x_1, \dots, x_T) = \prod p(y_t | c, y_1, \dots, y_{t-1})$

### Architecture

- Encoder, Decoderともに4層のRNN
- 各層に1000ユニット
- 単語ベクトル：1000次元
- 入力文のfeedは逆順



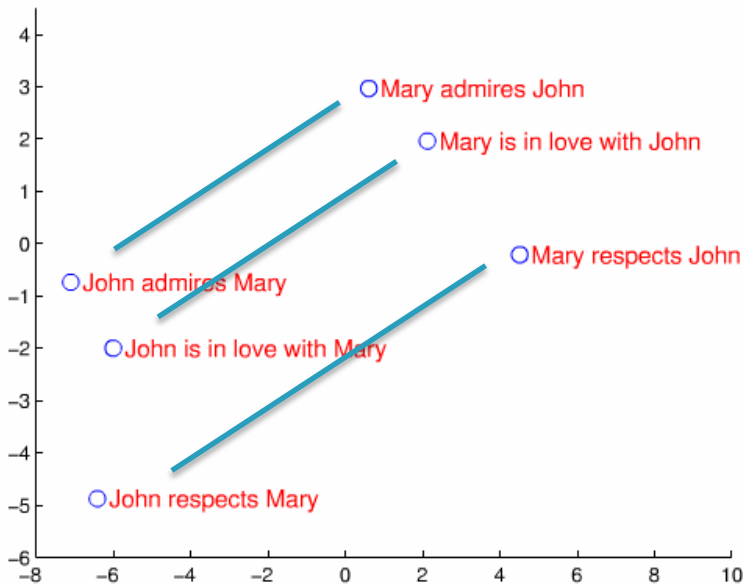
### Dataset

- WMT'14 英仏翻訳データセットのサブセット
  - 12M個の文の対 (ワード数 348M仏語、304M英語)
  - 語彙数：英語 160K、仏語 80K (頻出後を選択、それ以外はUNK)



# Distributed Representation

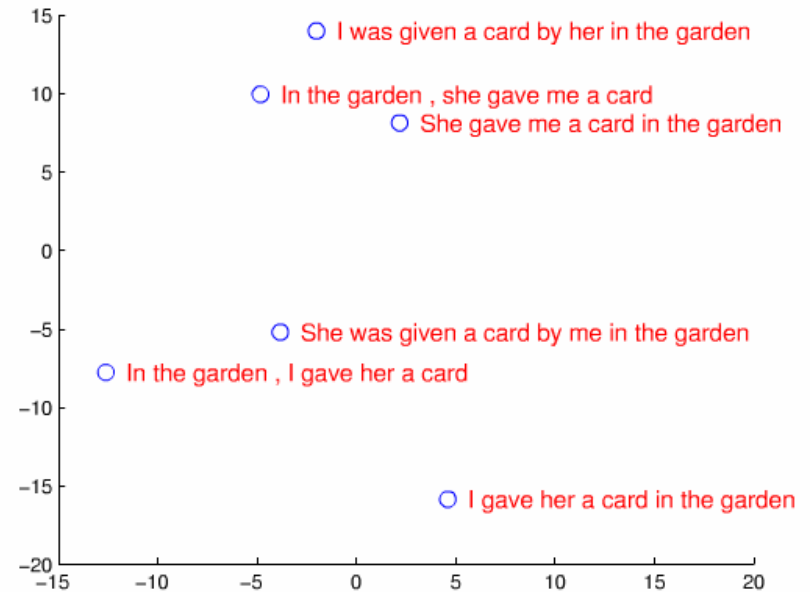
入力文が隠れ層でどのようなベクトルに変換されたか  
PCA 主要2成分をマップ



意味が同じ文どうしは距離が近い

述語間の関係は、主語目的語にかかわらず保存される

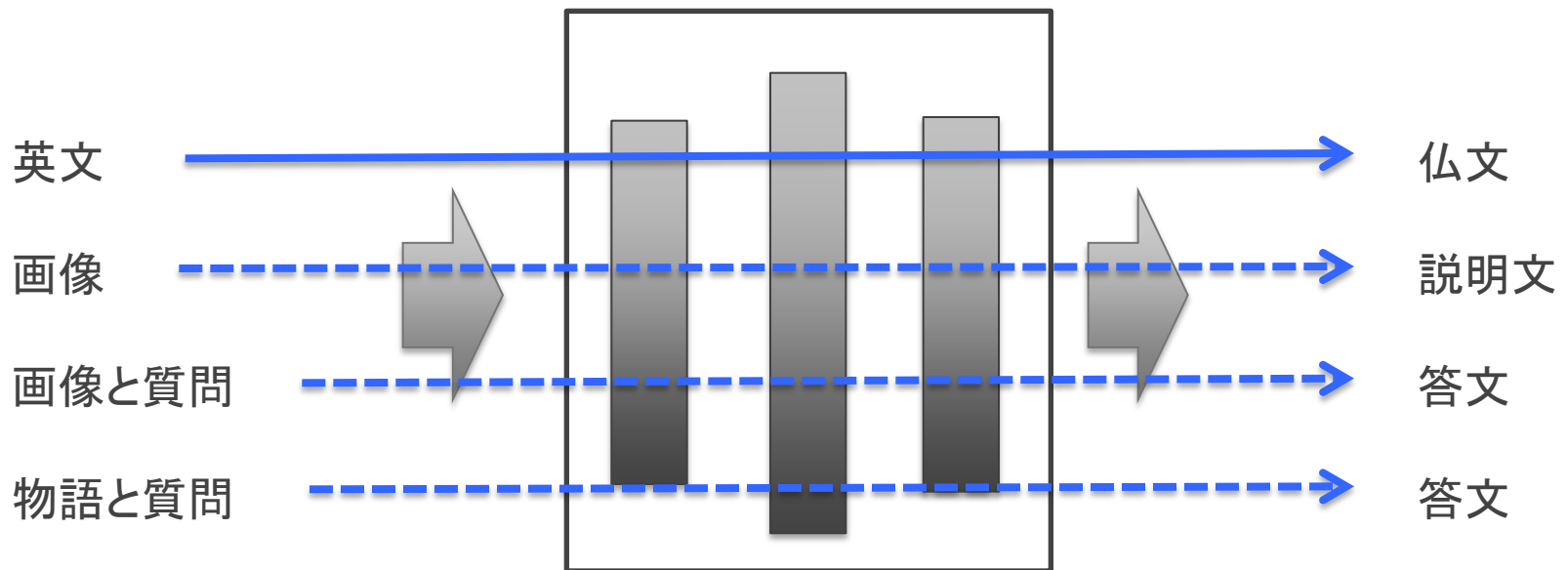
$$"M a J" - "J a M" + "J ilw M" = "M ilw J"$$



受動態・能動態によっても意味が変わらないことを学習

# 意味の分散表現 ( Distributed Representation )

記号論アプローチともっとも異なるポイント



↑  
意味  
(ベクトル表現)

どのような表現になるかは  
学習によって決まる

# 見たものを言葉で説明する

Show and Tell: A Neural Image Caption Generator

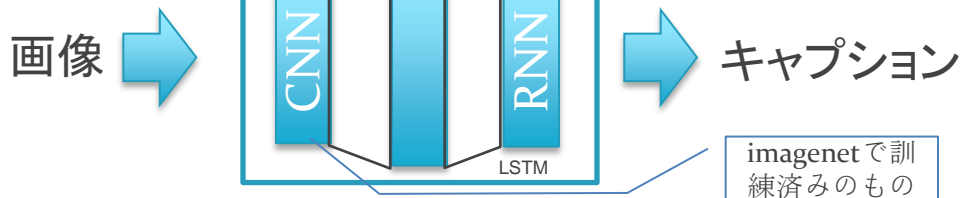
O. Vinyals, A. Toshev, S. Bengio, D. Erhan (Google)

画像が与えられるとそのテキスト記述を自動生成する

Model

- $p(\text{正解の単語列} \mid \text{画像})$ を最大化するよう訓練

Architecture



Performance

| Approach               | MS COCO   | PASCAL (xfer) | Flickr 30k | Flickr 8k | SBU       |
|------------------------|-----------|---------------|------------|-----------|-----------|
| Im2Text [23]           |           |               |            |           | 11        |
| TreeTalk [17]          |           |               |            |           | 19        |
| BabyTalk [15]          |           | 25            |            |           |           |
| Tri5Sem [11]           |           |               |            | 48        |           |
| m-RNN [20]             |           |               | 55         | 58        |           |
| MNLM [13] <sup>4</sup> |           |               | 56         | 51        |           |
| SOTA                   |           | 25            | 56         | 58        | 19        |
| NIC                    | <b>67</b> | <b>59</b>     | <b>66</b>  | <b>63</b> | <b>28</b> |
| Human                  | 69        | 69            | 68         | 70        |           |

state of the art



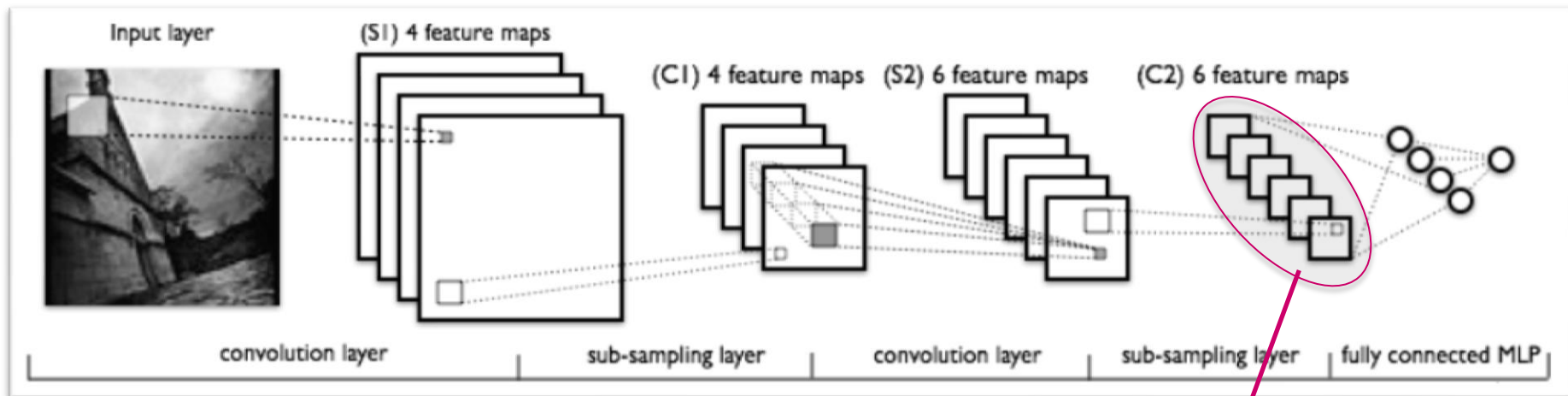
A group of people shopping at an outdoor market.

There are many vegetables at the fruit stand.

| Dataset name        | size  |        |      |
|---------------------|-------|--------|------|
|                     | train | valid. | test |
| Pascal VOC 2008 [6] | -     | -      | 1000 |
| Flickr8k [25]       | 6000  | 1000   | 1000 |
| Flickr30k [32]      | 28000 | 1000   | 1000 |
| MSCOCO [19]         | 82783 | 40504  | -    |
| SBU [23]            | 1M    | -      | -    |

画像の特徴から文を生成するルールを人間が作るより、画像と文の組み合わせからDLする方が良い結果を出せた

# CNNは学習により抽出すべき特徴を決定する



様々な局所的な特徴の抽出

分類問題を解く

ここの分散表現を使って  
RNNでテキストを生成する

〔画像の意味〕



**A person riding a motorcycle on a dirt road.**



**Two dogs play in the grass.**



**A skateboarder does a trick on a ramp.**



**A dog is jumping to catch a frisbee.**



**A group of young people playing a game of frisbee.**



**Two hockey players are fighting over the puck.**



**A little girl in a pink hat is blowing bubbles.**



**A refrigerator filled with lots of food and drinks.**



**A herd of elephants walking across a dry grass field.**



**A close up of a cat laying on a couch.**



**A red motorcycle parked on the side of the road.**



**A yellow school bus parked in a parking lot.**



**Describes without errors**

**Describes with minor errors**

**Somewhat related to the image**

**Unrelated to the image**

|  |
|--|
| A man throwing a frisbee in a park.<br><b>A man holding a frisbee in his hand.</b><br><b>A man standing in the grass with a frisbee.</b>                   |
| A close up of a sandwich on a plate.<br>A close up of a plate of food with french fries.<br>A white plate topped with a cut in half sandwich.              |
| A display case filled with lots of donuts.<br><b>A display case filled with lots of cakes.</b><br><b>A bakery display case filled with lots of donuts.</b> |

Table 3. N-best examples from the MSCOCO test set. Bold lines indicate a novel sentence not present in the training set.

最高スコアの文は80%の確率でtraining setにあった文

トップ15位までの生成文を調べると、50%の確率で新規の文が見つかる。しかもその文は十分な品質を持っていた

訓練例にない新しい文を生成できた！

Word embeddingはキャプションに出現する単語の共起関係に影響を受ける

word embeddingが近い単語はCNNの視覚的特徴でも似たオブジェクトに対応付けられる

### Word Embedding

| Word     | Neighbors                         |
|----------|-----------------------------------|
| car      | van, cab, suv, vehicule, jeep     |
| boy      | toddler, gentleman, daughter, son |
| street   | road, streets, highway, freeway   |
| horse    | pony, donkey, pig, goat, mule     |
| computer | computers, pc, crt, chip, compute |

Table 6. Nearest neighbors of a few example words

# 画像の内容を日本語で語る

## Andrej KarpathyのNeuralTalk2を使って実験

- <https://github.com/karpathy/neuraltalk2>

## 大規模なデータセットを用意

- 画像：約16万枚（MSCOCO）
- 日本語キャプション：画像1枚につき5つ（クラウドソースで作成）

## キャプションが60万集まったところの中間成果

- [こちら](#)



# 画像についての質問に答える

Are you talking to a Machine?

Dataset and Methods for Multilingual Image Question Answering

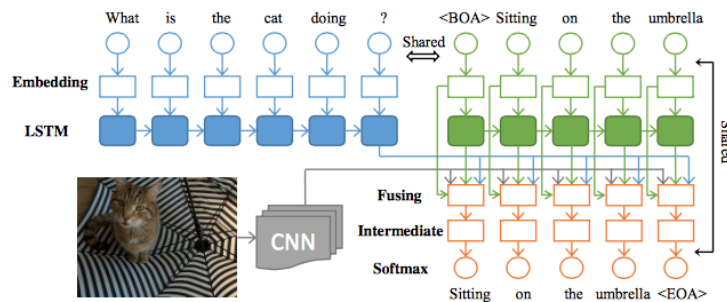
Haoyuan Gao, Junhua Mao, Jie Zhou, Zhiheng Huang, Lei Wang, Wei Xu (Almost Baidu Research)

画像とそれに写っている内容についての質問文を与えると、答え(文)を返す

Model

- $p(\text{正答の単語列} \mid \text{画像、質問文})$ を最大化するよう訓練

Architecture



LSTM<sub>1</sub>

質問文を処理

LSTM<sub>2</sub>

回答文を処理  $p(w_n \mid w_1, \dots, w_{n-1})$

CNN

Imagenetで事前に訓練して、固定回答生成

$p(w_n \mid w_1, \dots, w_{n-1}, \text{Image}, Q)$

Performance

- Dataset FM-IQA(Freestyle Multilingual Image Question Answering)を作った
  - **120,000 images, 250,000 自由中国語QA文、その英語翻訳文**
- チューリングテストで評価
  - 64.7%のケースで、人間の審判はmQAのアンサーと人間のそれを区別できなかった
  - 回答の正確さは 1.454 / 2.0 (人間の回答の正確さは 1.918 / 2.0)



|          |   |   |  |   |   |
|----------|---|---|--|---|---|
| Image    |  |  |  |  |                        |
| Question | 公共汽车是什么颜色的?<br>What is the color of the bus?                                      | 黄色的是什么?<br>What is there in yellow?   | 草地上除了人以外还有什么动物?<br>What is there on the grass, except the person?                  | 猫咪在哪里?<br>Where is the kitty?   | 观察一下说出食物里任意一种蔬菜的名字?<br>Please look carefully and tell me what is the name of the vegetables in the plate? |
| Answer   | 公共汽车是红色的。<br>The bus is red.  | 香蕉。<br>Bananas.   | 羊。<br>Sheep.   | 在椅子上。<br>On the chair.  | 西兰花。<br>Broccoli.   |

Figure 1: Sample answers to the visual question generated by our model on the newly proposed Freestyle Multilingual Image Question Answering (FM-IQA) dataset.

|            |  |  |   |  |  |
|------------|--|--|---|--|--|
| Image      |  |  |  |  |  |
| Question   | 帅哥在干什么?<br>What is the handsome boy doing?   | 这是什么?<br>What is there in the image?   | 盘子里有什么水果?<br>Which fruit is there in the plate?                                     | 这是什么车?<br>What is the type of the vehicle?   | 公交车停在那干吗?<br>Why does the bus park there?  |
| GT Answer  | 在抓飞盘。<br>Trying to catch the frisbee.  | 草原上的马群。<br>Horses on the grassland.  | 苹果和橙子。<br>Apples and oranges.   | 公交汽车。<br>Bus.  | 准备维修。<br>Preparing for repair.   |
| mQA Answer | 冲浪。<br>Surfing.  | 这是牛。<br>They are buffalos.   | 香蕉和橙子。<br>Bananas and oranges.  | 火车。<br>Train.  | <OOV>。<br><OOV> (I do not know.)   |

Figure 6: Failure cases of our mQA model on the FM-IQA dataset.

## A Neural Conversational Model

*O. Vinyals, Q.V Le @google*

### Model

- SutskeverらのSeq2Seq Model

### Dataset

- コールセンターの会話：30M 単語、ユーザとオペレーターの会話
- 映画のセリフ：62M 文 ( 923M 単語 )、前の文から次の文を予測させるように学習

### Train

- 次単語予測学習

# 会話例

## Conversation 2: Simple Q&A (contexts and multiple choice)

*Human: my name is david . what is my name ?*

短期記憶

*Machine: david .*

更新可能な記憶

*Human: my name is john . what is my name ?*

*Machine: john .*

ある程度の持続

*Human: are you a leader or a follower ?*

*Machine: i 'm a leader .*

*Human: are you a follower or a leader ?*

*Machine: i 'm a leader*

# 会話例

## Conversation 3: General knowledge Q&A

*Human: who is skywalker ?*

*Machine: he is a hero .*

*Human: who is bill clinton ?*

*Machine: he 's a billionaire .*

*Human: is sky blue or black ?*

*Machine: blue .*

*Human: does a cat have a tail ?*

*Machine: yes .*

*Human: does a cat have a wing ?*

*Machine: no*

*Human: can a cat fly ?*

*Machine: no .*

*Human: how many legs does a cat have ?*

*Machine: four , i think .*

*Human: how many legs does a spider have ?*

*Machine: three , i think .*

*Human: how many legs does a centipede have ?*

*Machine: eight .*

*Human: what is the color of the sky ?*

*Machine: blue .*

*Human: what is the color of water ?*

*Machine: water .*

*Human: what is the color of blood ?*

*Machine: it is the same as a black eye .*

*Human: what is the usual color of a leaf ?*

*Machine: it is a green one .*

*Human: what is the color of a yellow car ?*

*Machine: yellow .*

*Human: how much is two plus two ?*

*Machine: four .*

*Human: how much is ten minus two ?*

*Machine: seventy - two .*

三人称単数の使い分け

is\_a関係の記憶

has\_a関係の記憶

色を表す単語と色を記述する構文の記憶

# 会話例

## Conversation 6: Opinions

*Human: what do you think about tesla ?*

*Machine: he 's a good conductor .*

*Human: what do you think about bill gates ?*

*Machine: he 's a good man .*

*Human: What do you think about messi ?*

*Machine: he 's a great player .*

*Human: what do you think about cleopatra ?*

*Machine: oh , she 's very regal .*

*Human: what do you think about england during the reign of elizabeth ?*

*Machine: it was a great place .*

主語に応じて、 he / she / it を使い分けている

# Neural Conversation

## 明示的なあるいは記号的な知識表現なしの会話

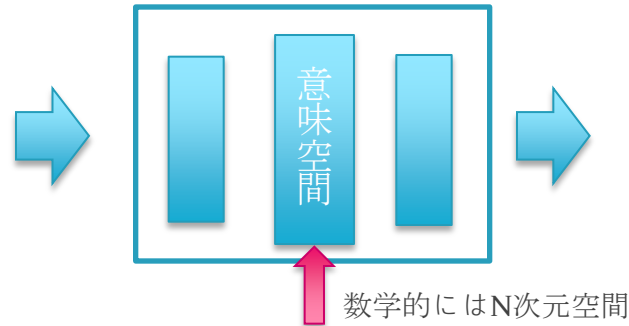
### 結果

- 短期記憶で答えることのできる質問を回答できる
- 単純なファクトの記憶
- 常識レベルの推論
- コンテキストの理解
- 自然な言葉を紡ぎだしている
  - 次単語予測学習が大きく貢献

### 課題

- 目的関数が会話の目的とは異なる
- 世界の知識（長期記憶）がない
- 一貫性がない
  - 訓練用のデータセットが多数の人間の発言・文章を含むため

# 概念はどこにできたのか



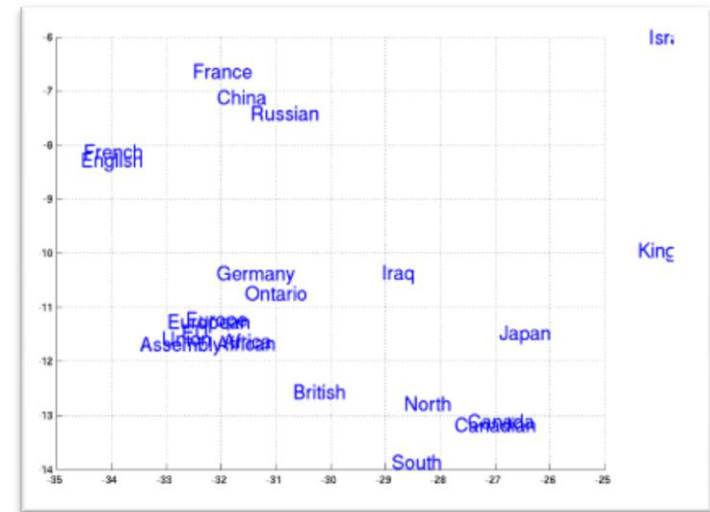
意味空間(隠れ層)内に「概念」に相当するものができた

どんな風に存在しているか

- 個々の「概念」は意味空間の中の局所領域に対応する
- この意味空間には多数の概念が疎に散らばっている
  - 多数の銀河系が疎に散らばっている宇宙のように
- おそらく句も文も映像的な記憶も同じ

ここで言う概念は

- 哲学的な「概念」という概念を指しているわけではない
- 統計的に大きな割合を占める刺激
  - ある単語、ある句、ある画像、ある顔、ある音、、、、学習で獲得された記憶の塊



## 意味空間

状態空間が2次元だとすると、franceやiraqはその空間の中のある局所領域に相当

# NNNLM vs 子どもの言語獲得





# 子どもの言語獲得・学習のモデル化とシミュレーションを目指して

## 研究概要

### ゴール

- 言語を喋り理解する人工知能を実現する
- 固定した語彙と構文を使いこなせるというのではなく、語彙も構文も経験とともに拡大していくことが可能な言語能力
- 言葉を通して知識を吸収する能力

### アプローチ

- 子供の言語獲得に学び、モデルを作ろう

### 理由

- 獲得・発達過程を観察することで人間の言語処理の原理がわかる
- 「発達」のメカニズムを取り込みたい

### 手段

- Deep Neural Networks

# なぜ DNN か

## 従来の会話する人工知能研究との違い

子どももDNNもベースは脳、ともに白紙の状態から言語を獲得する

ゼロからブートストラップするモデルという点で共通項は多い

- 生まれた時に脳の中に存在しない
- 例から学習しなくてはならない
- 脳の学習機構を活用している

従来AIでは「記号、その意味、記号間の関係、文構造と意味の対応」などを事前にアドホックに設計する

- 自然言語処理アルゴリズム、辞書、知識表現、KB、等々

# 子どもの言語獲得

今井むつみ氏やスティーブン・ピンカー氏の著書を参考に

新生児は体内で母語の音韻の学習をしている

- 生まれたばかりの新生児がしゃべったということは未だかつてない

18ヶ月ころに急激に進歩

- 語彙爆発(1日に憶える語彙数が急激に増える)
- この時期、神経結合が大量に作られる(脳生理学的事実)

モデルがリッチになると能力が増える

母語を学習できない子どもはいない(健康であれば)

- 第2外国語の学習に失敗する子どもや大人はたくさんいるにもかかわらず

従来AIの言語モデルはこっちに似てる

チンパンジーに人間の子どものと同じ環境を与えても言語(手話も)を獲得できない

- 脳の構造が違うのか
- チンパンジーの脳とDeep NNの何が違うのか

チンパンジーの脳 < RNN E-D ?

「言葉」らしい言葉

- 単語や順番のない単語列は言葉と言えない(チンパンジー、ピジン語)
- 二語文、三語文にして初めて言葉と言える(クレオール語)
- 述語をマスターすると、表現力が格段に増す

RNNは「言葉らしい」言葉を生成?

子どもが獲得する順序に脳内の言語モデルがうかがえる

- 名詞(形のあるもの→形のないもの) → 形容詞 → 動詞

RNNもこの順序で獲得している?

言語は自己拡張能力があり、言語を獲得するとはそれも含めての話

まったく未解決

# 言葉の発達：初期

固有物への命名、自分の感覚中心（動詞も形容詞も）

## ■ モノの名前

- アヒル、「パパは写真」「グレープフルーツはお月さま」
- 汎化ができていない

## ■ 動詞

- 「くっく」「ピョーン」
- 動作という概念はあっても、関連する名詞や擬音を使う

## ■ 形容詞

- 「おいしい」「まずい」「いたい」
- 自分の感覚に関連した形容詞は言うが、「大きい」などの客観的な形容詞はまだ

# 言葉の発達：二歳前（語彙爆発）

汎化や類似性を実験的に多用し限界を調整しているように見える

## ■ モノの名前

- 「しらすぎ = 白いカラス」「ガスタンク = ボール」「コンデンスミルク = イチゴのしょうゆ」
- 汎化の発達、類似性の活用（→ 比喻）

## ■ 動詞

- 「蹴る = 足で投げる」「唇を噛む = 歯で唇を踏む」
- 汎化、類似性の活用

## ■ 形容詞

- 「血压がやすい」
- 汎化

# NNNLM vs 子どもの言語獲得

## NNNLM

- 教師付き学習 + 誤差関数の最急降下法による最適化は強力
- 誤差逆伝搬は極めて効率的
- 大規模コーパスが必須
  - 大量のパラメータをoverfitting回避して調節するため

## 人間の学習はそこまで効率的ではない

- 目的関数や誤差があいまい
  - 従って強力な最適化が進行しているわけでもないので overfitting も起きない
  - 大規模コーパスはかえって過負荷となって負の効果
- しかし安定した最終性能
- 初期は脳に十分な処理能力がない
  - 細胞数、結合数において初期は不十分なのではないか
  - 時間とともに段階的に増えるのかもしれない

## 評価関数はコミュニケーションの成否

- 教師信号？相手からのフィードバック？

# 子ども並みの言語モデルに向けて

## (超) 楽観的な展望

### 対象とする世界

- 家とその周辺
- タイムスパン：1週間以内のイベント
- 語彙数：1000程度（2歳児相当）

### 能力

- モノやコトの認識ができる（1000個のカテゴリー）
- モノやコトに関する質問応答ができる
- 10個の文以上（10インタクション以上）のコンテキストの保持とそれに基づく会話
- 自発的発話（発見（変化）、フォーカス（興味））

### モデル

- RNN+
- 入力 は 文 と 画像、出力 は 文
- 文、画像のデータセットは100万例以上
- supervised and unsupervised 学習
- 記憶機構の拡充：容量と参照
- motivation：強化学習（学習と報酬）

### + スケールアウト

- システムの規模を拡大して機能を損なうことなく語彙数を拡大させることが可能に

