ACL 2018
参加報告 (2)

2018年9月7日

ヤフー株式会社　清水 徹

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

自己紹介

・清水 徹

- 世の中的な技術調査、レポーティング
- ヤフー知恵袋・ヤフー検索の要素技術開発
- 自然言語処理系の研究開発
  ・今回のACLに short paper採択
  （小林・清水伸幸との共著）

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

ACL 2018 所感

- LSTM RNNが完全にメインストリームに
  - とはいえ、タスクの最高性能を争うような話では より新しいアーキテクチャに及ばず
  - deepでないもの（グラフィカルモデルなど）は探さないと無い
- Transformerが存在感を増す
- 「常識」、commonsense knowledgeの取り込みへ
- マルチモーダルな広がり

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Connecting Language and Vision to Actions

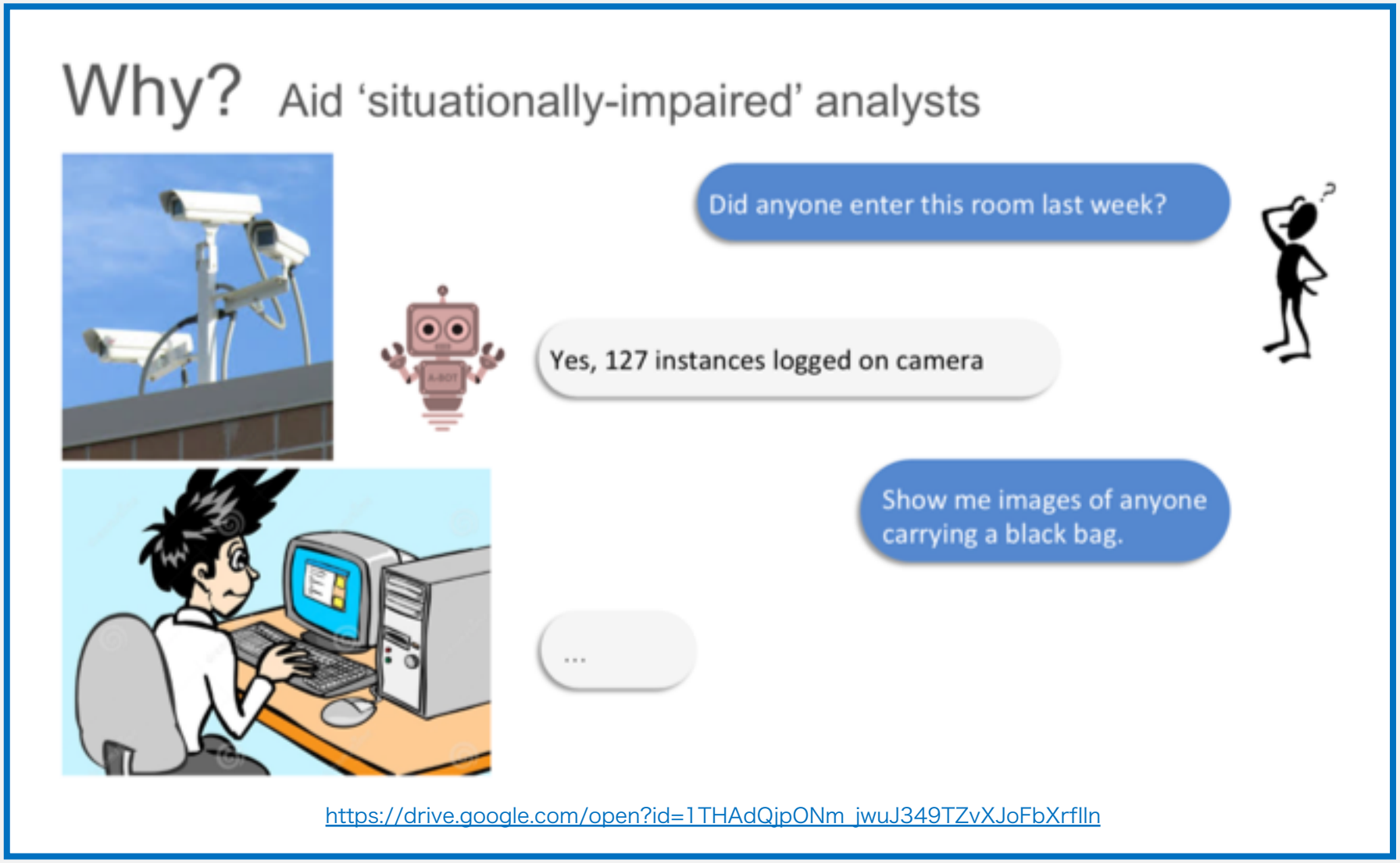
Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

- 今回のチュートリアルセッションの一つ
- Peter Anderson（Georgia Tech）を中心に
- テーマ
  - vision・languageを扱う技術
  - そのような入力のもとで動くagent
- データセットの紹介、タスクへのいざない
- サイト
  - https://lvatutorial.github.io/

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions



※ 本稿では このような青枠内は、全て他資料の引用

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

Why? Personal assistants

And what's next?

Add the chopped zucchini to the pan...

Hey, can you order more zucchini?

Done.

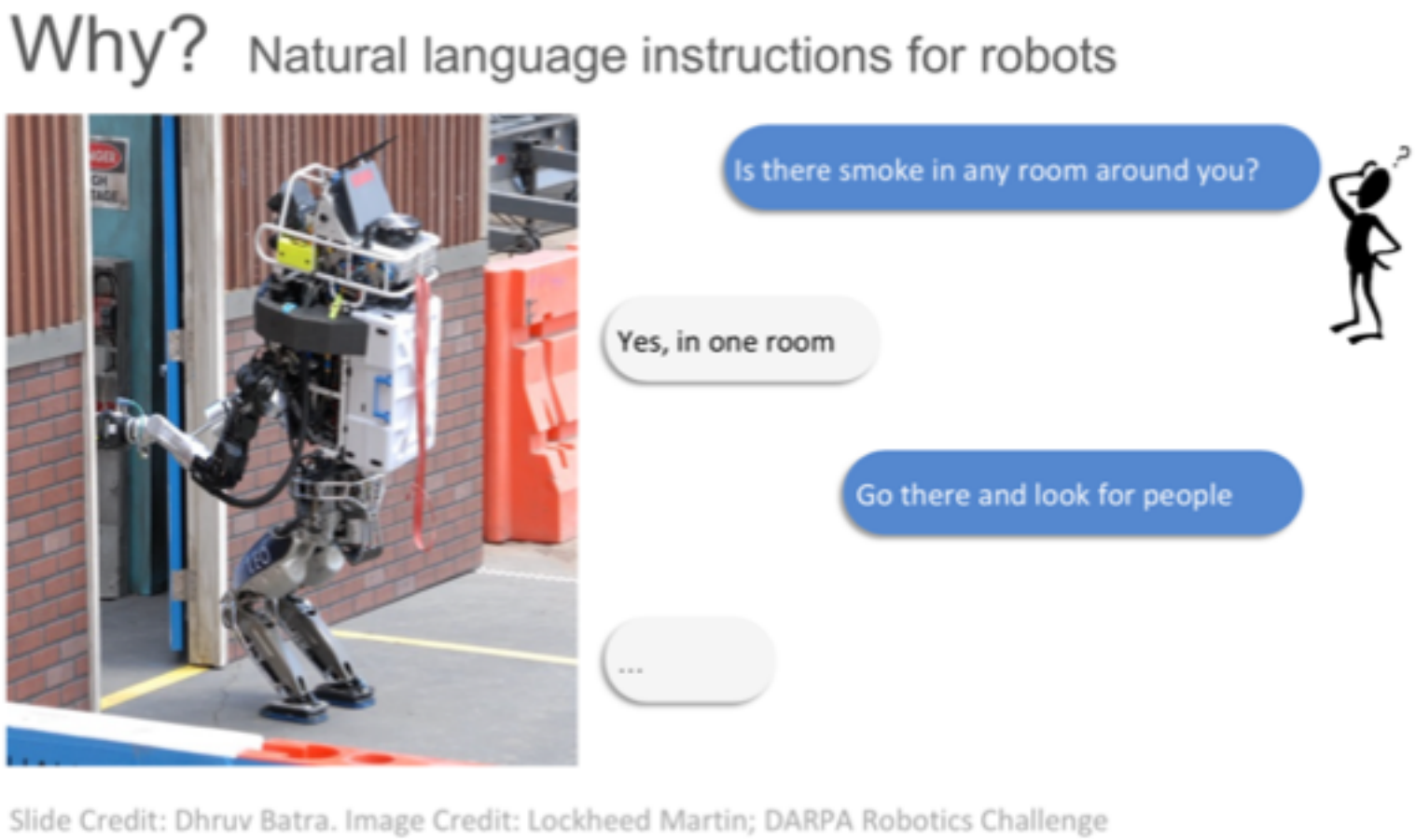
Image Credit:Bosch; Mykie countertop robot

https://drive.google.com/open?id=1THAdQjpONm_jwuJ349TZvXJoFbXrfIln

※ 本稿では このような青枠内は、全て他資料の引用

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

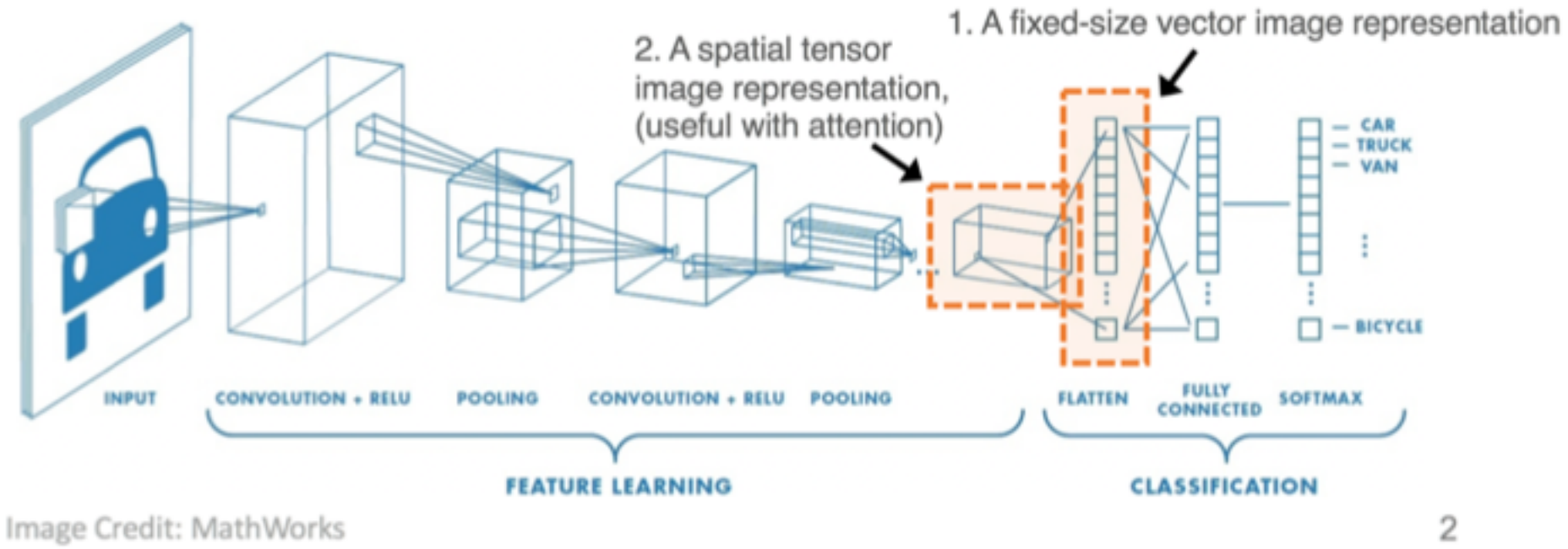


https://drive.google.com/open?id=1THAdQjpONm_jwuJ349TZvXJoFbXrfIln

※ 本稿では このような青枠内は、全て他資料の引用

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

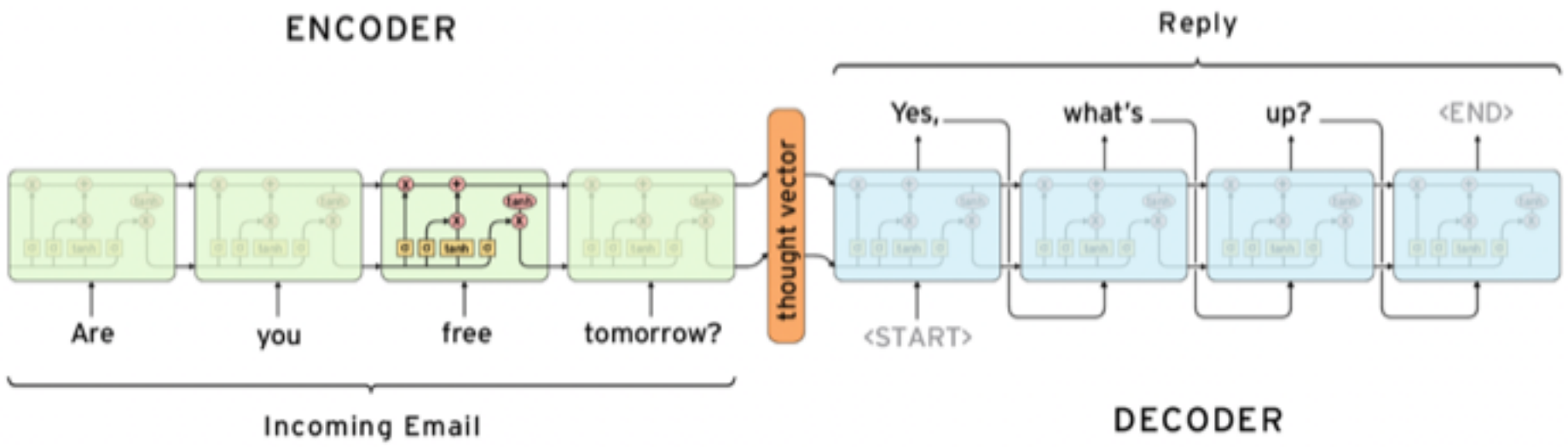


https://drive.google.com/file/d/1ZMw5GHAYrQfNXy2v06rU3K6GxHixXXSN

visionはCNNで

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions



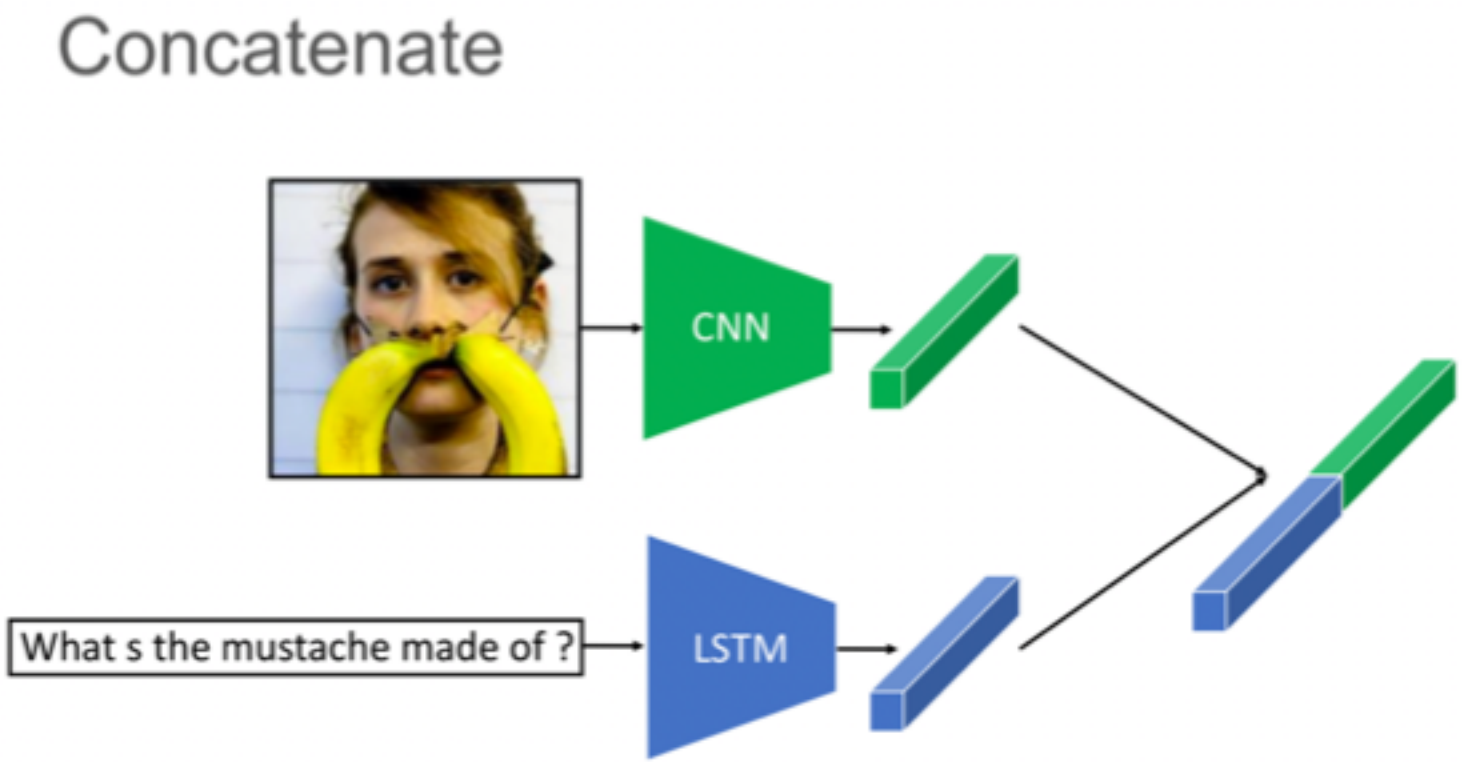
https://drive.google.com/file/d/1ZMw5GHAYrQfNXy2v06rU3K6GxHixXXSN

languageはRNNで

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

両者の分散表現をどのように考え合わせるか

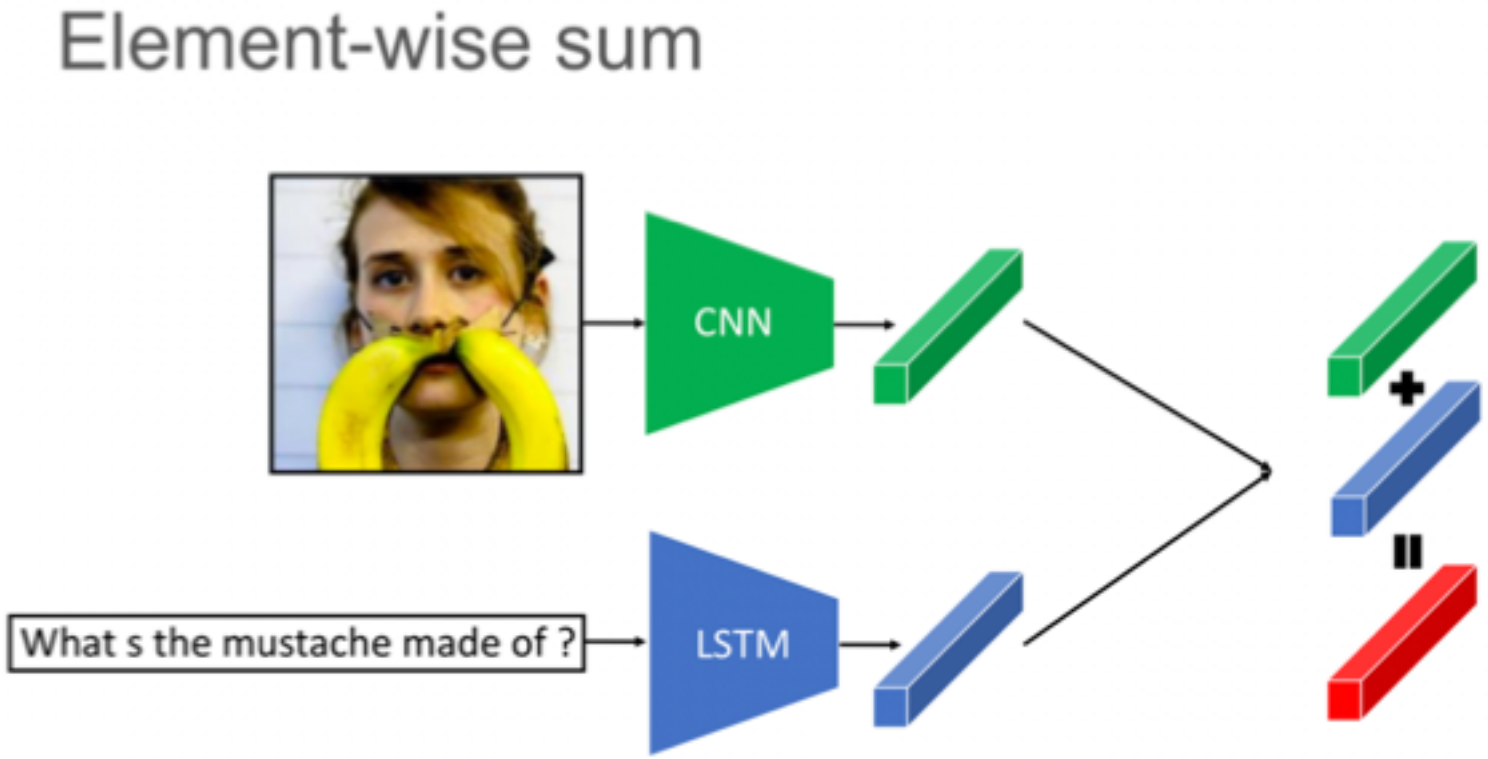


https://drive.google.com/open?id=1uvQ-qtMwc5CV5kkyGHj2ZB26uSMgbz84

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

両者の分散表現をどのように考え合わせるか

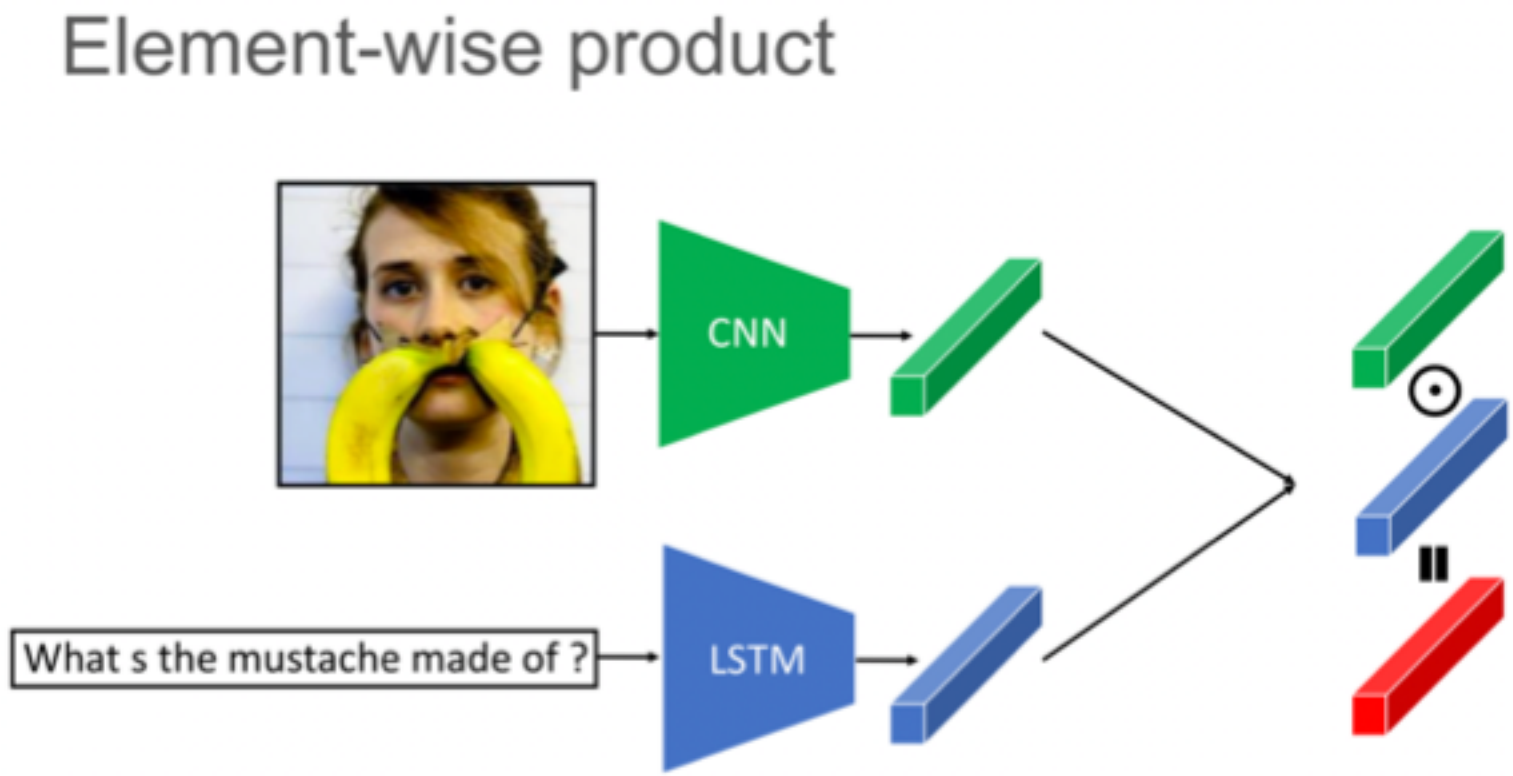


https://drive.google.com/open?id=1uvQ-qtMwc5CV5kkyGHj2ZB26uSMgbz84

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

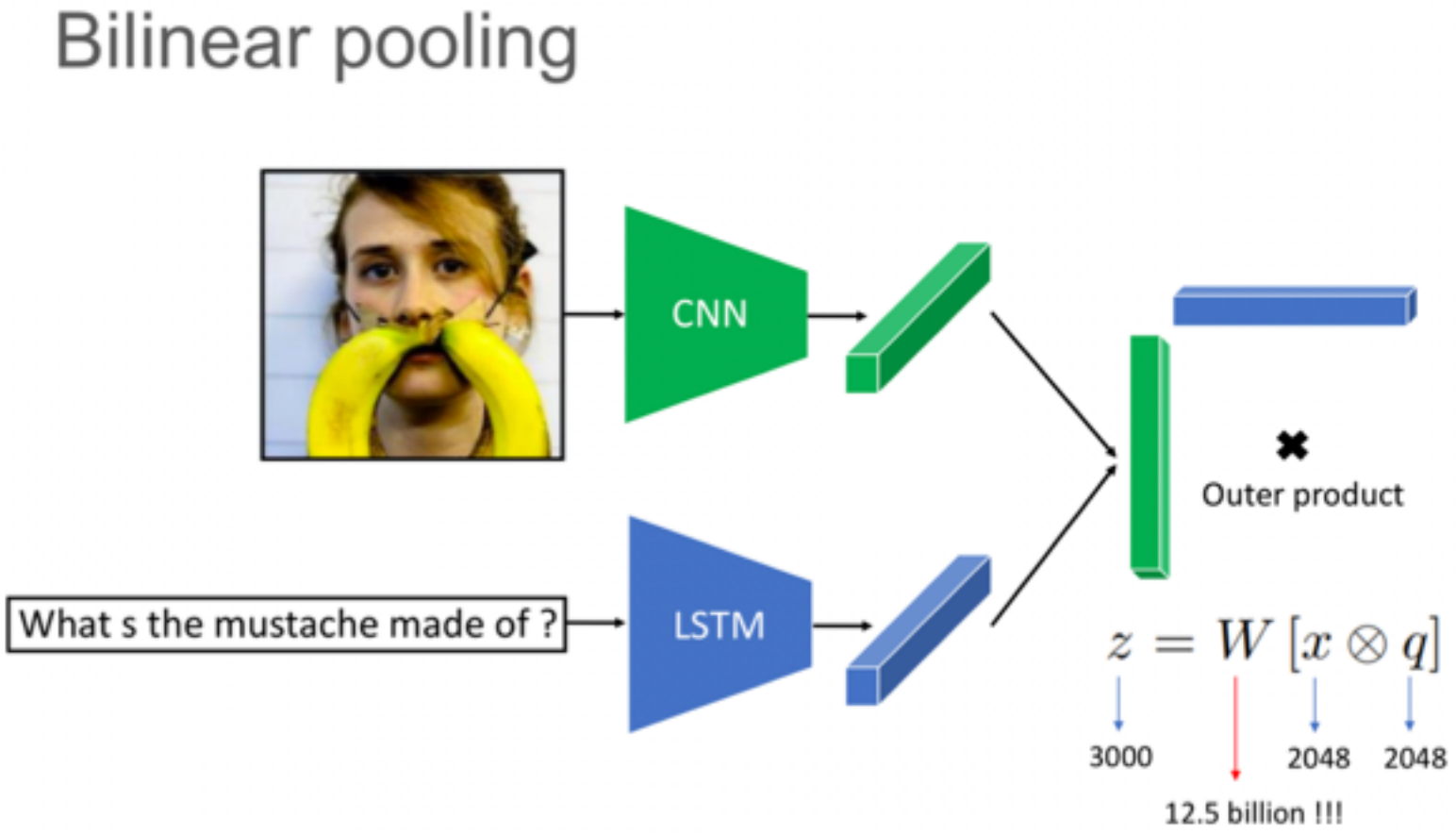
両者の分散表現をどのように考え合わせるか



https://drive.google.com/open?id=1uvQ-qtMwc5CV5kkyGHj2ZB26uSMgbz84

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

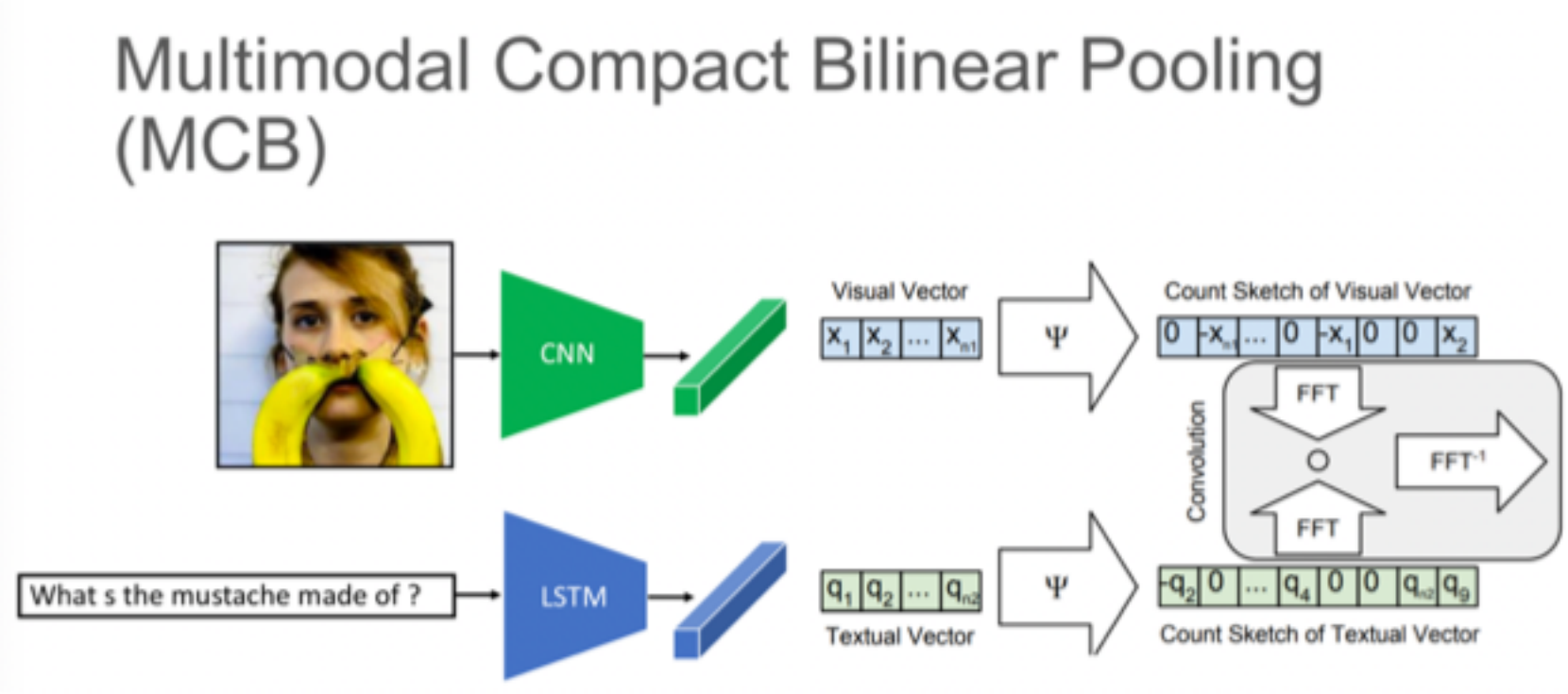
Language and Vision to Actions



https://drive.google.com/open?id=1uvQ-qtMwc5CV5kkyGHj2ZB26uSMgbz84

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions



https://drive.google.com/open?id=1uvQ-qtMwc5CV5kkyGHj2ZB26uSMgbz84

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

- visionとlanguageの分散表現を混ぜる手法： 他、MLB、MFBなど 発展形あり
  - VQA分野を中心に検討
- どれがベストか 結論は出ておらず、注視したい
- 個人的な印象として： 質問応答の 質問と応答の表現を混ぜる手法と相似形（？）
- → そして、このような技術をベースに「自然言語でコミュニケートできる、実世界で移動し課題解決できるエージェントを作る」のが本丸

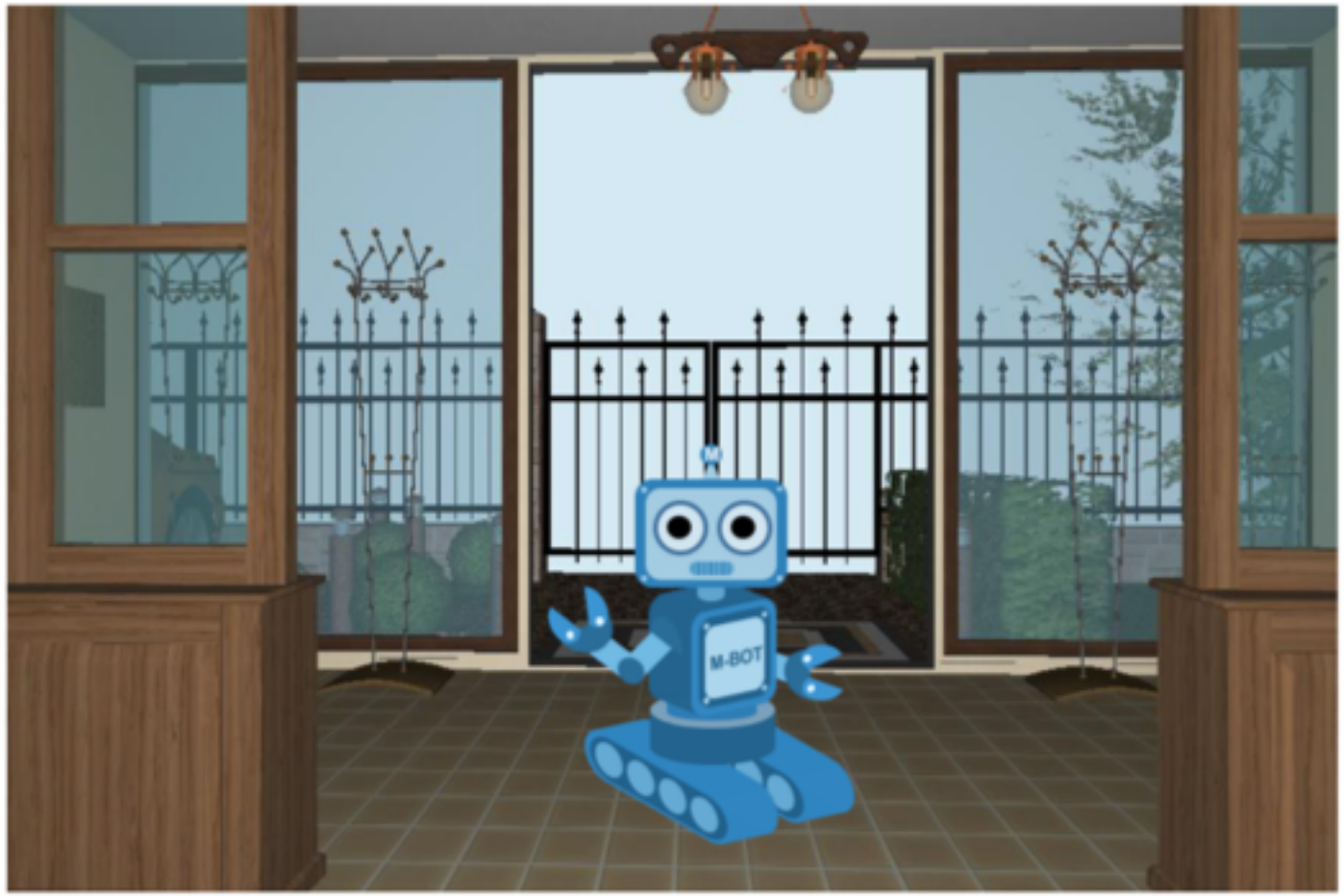
Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

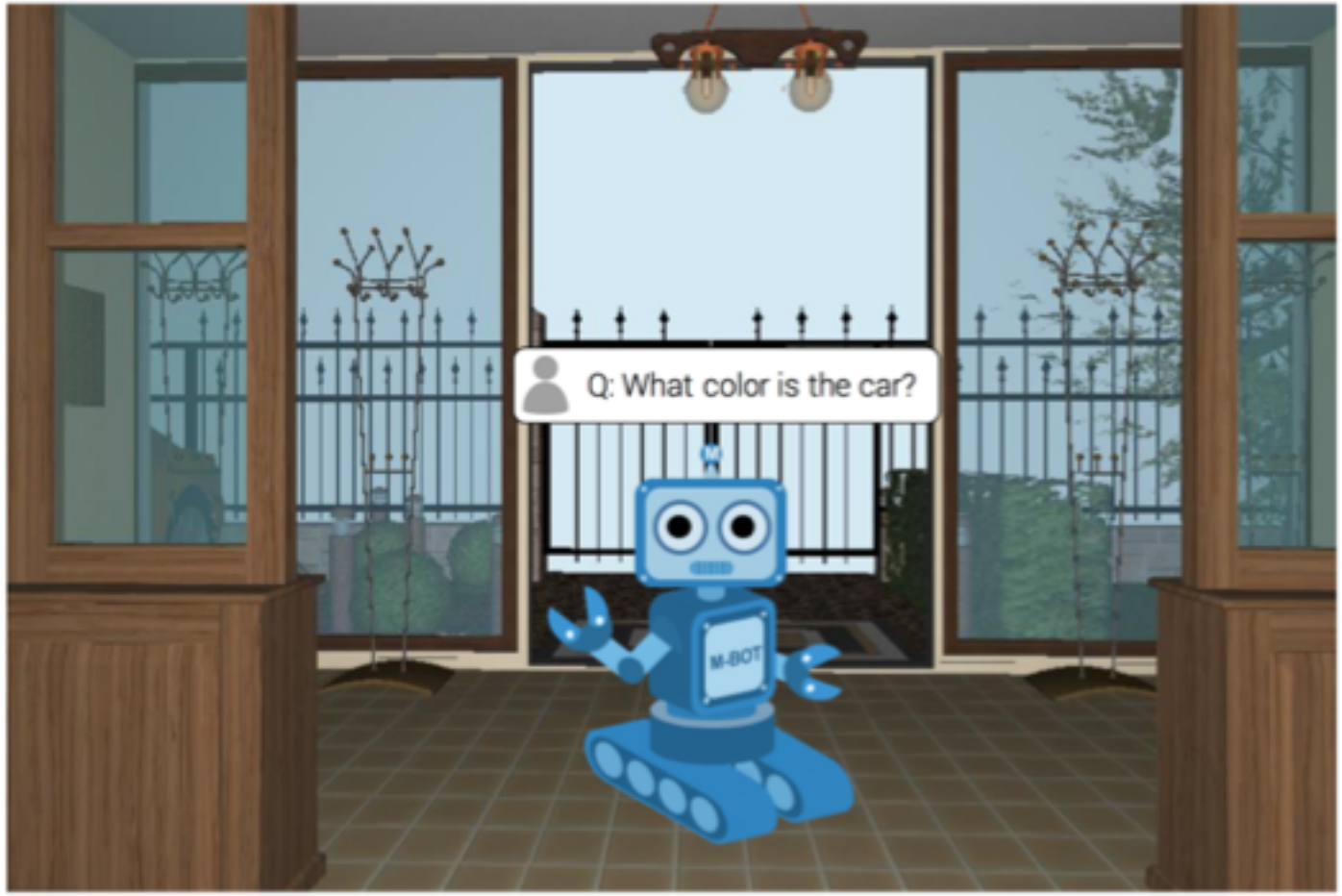
Language and Vision to Actions



https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions



「自然言語で、願わくば音声でQを伝えられる」のがポイント

https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

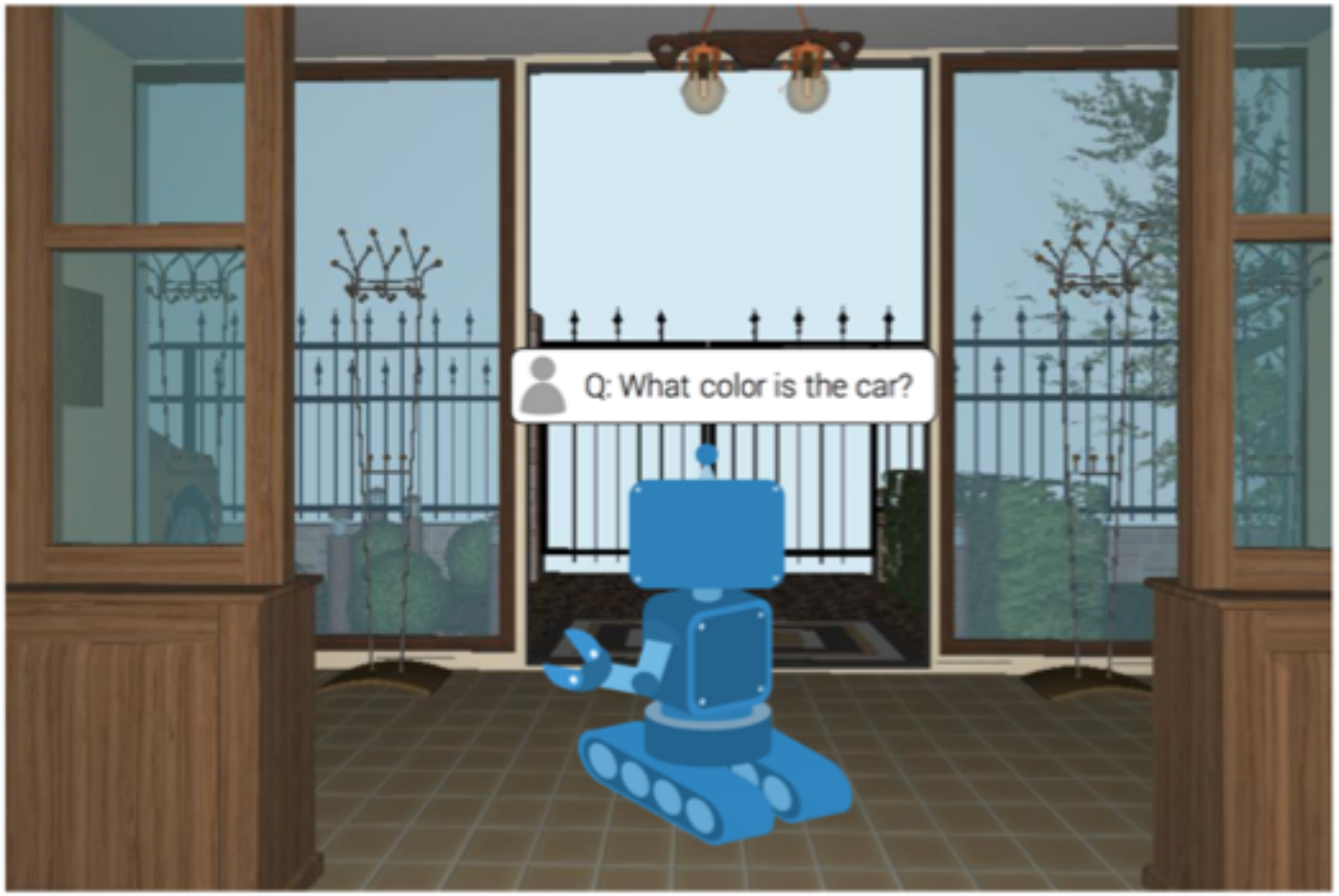
Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

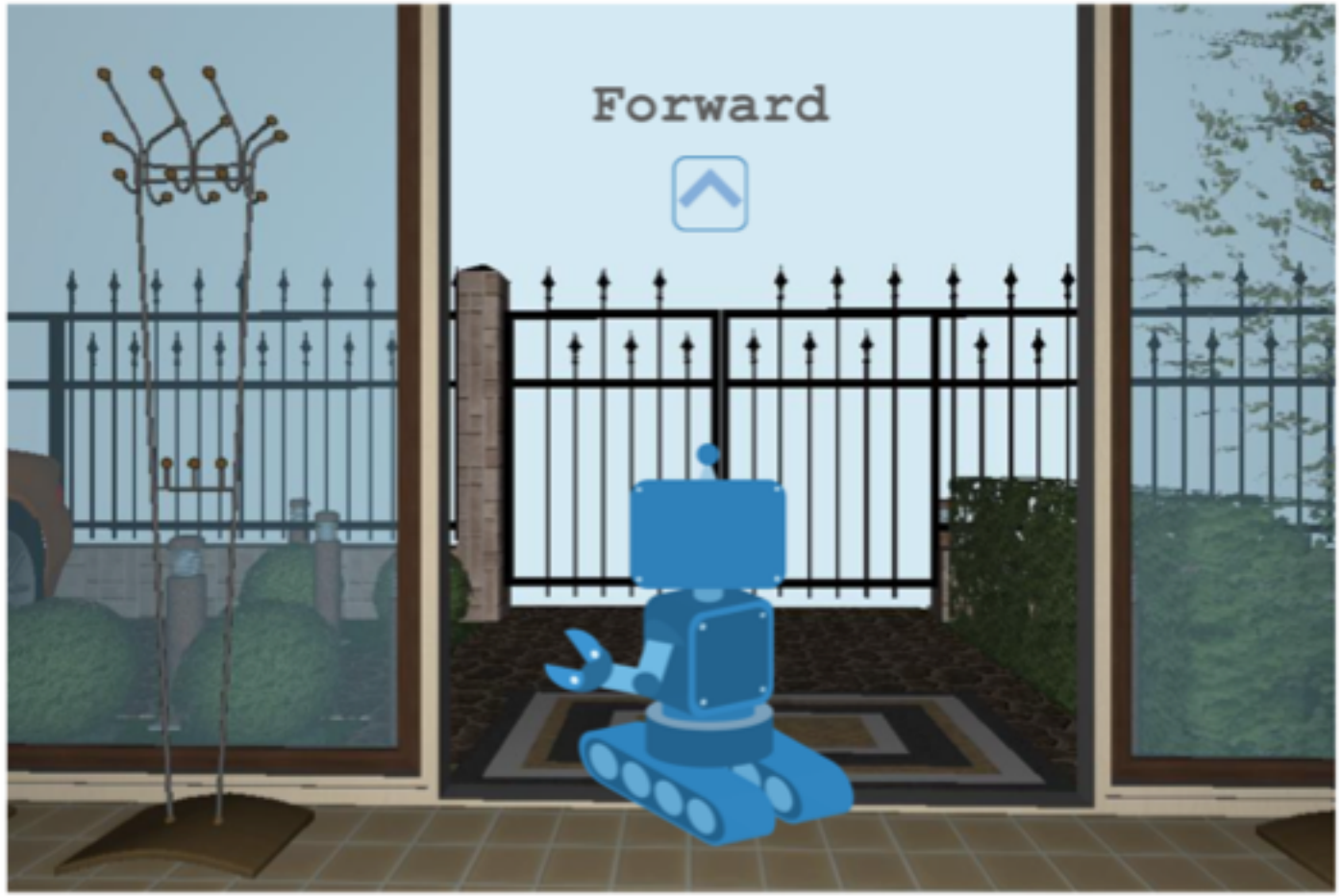


問題解決の
ために ある程度
自律的に行動

https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

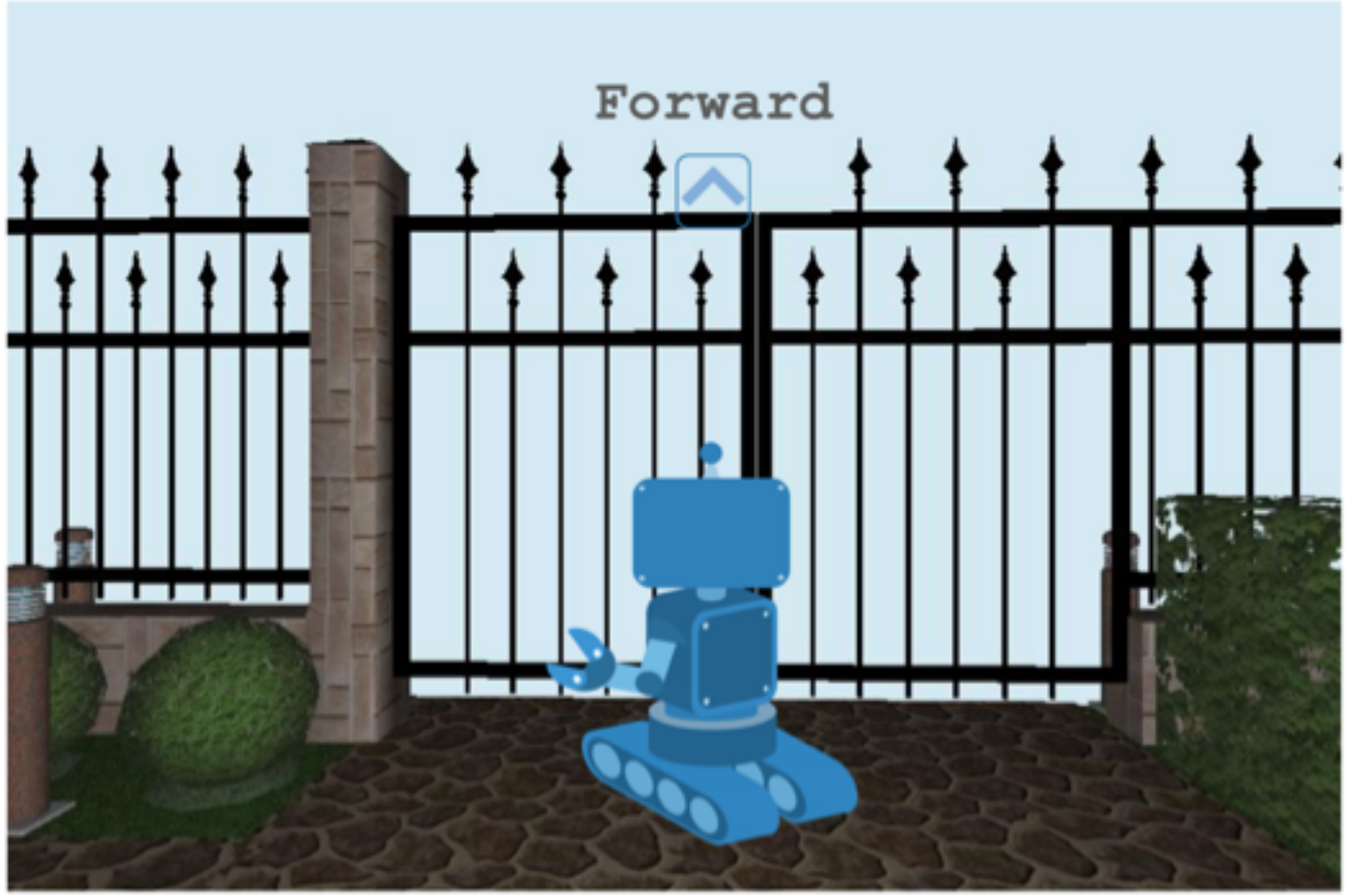


navigation

https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

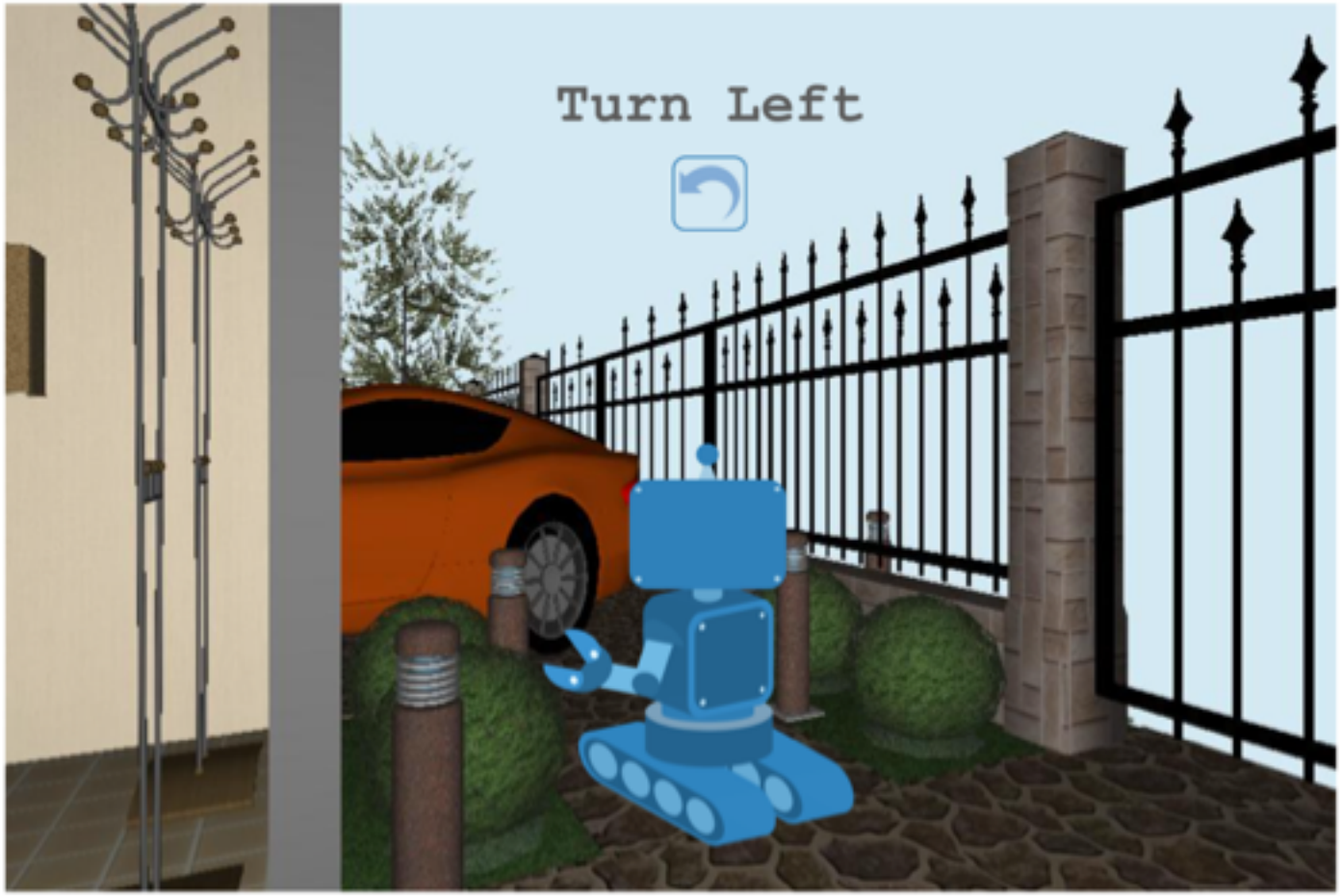
Language and Vision to Actions



https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

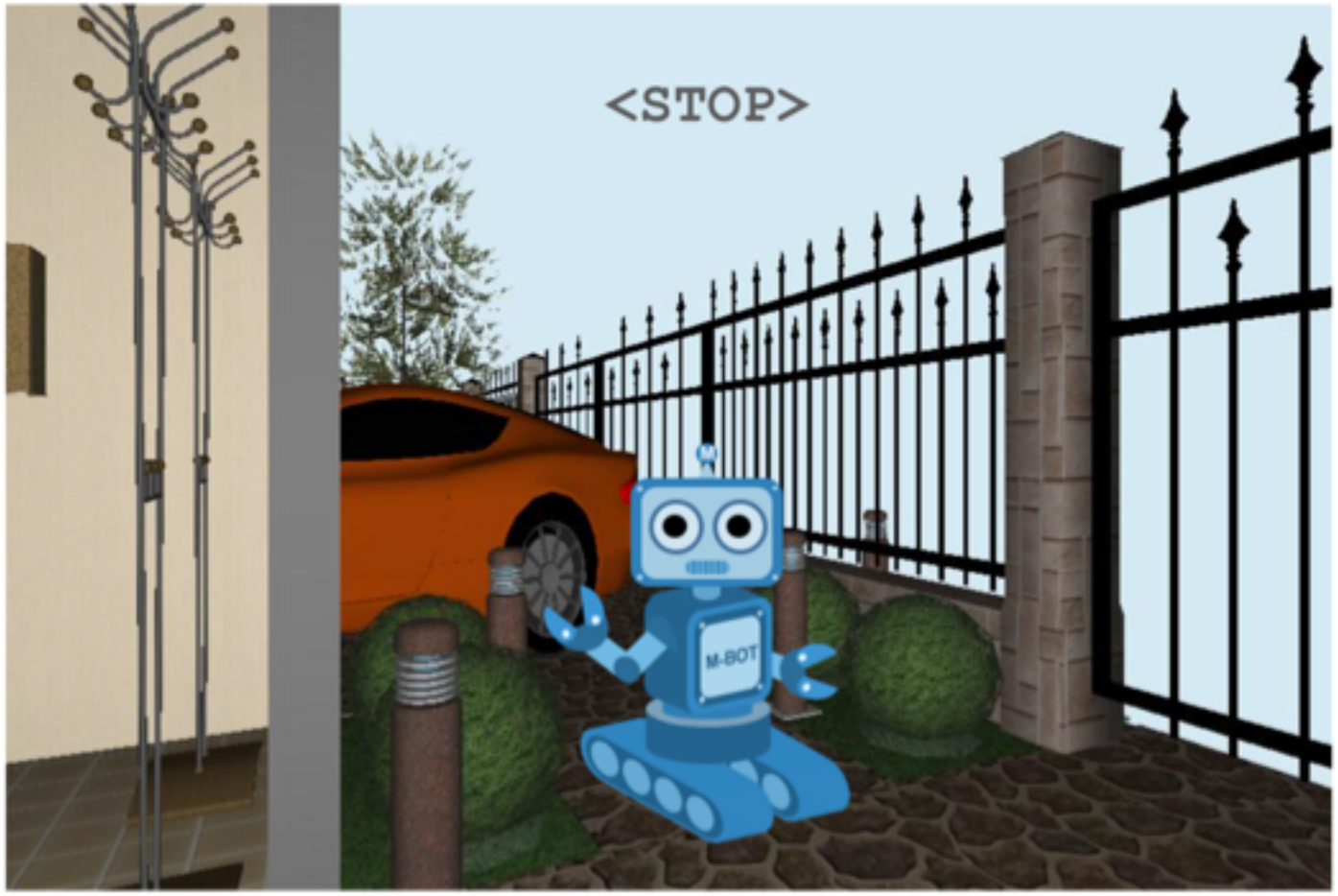
Language and Vision to Actions



https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

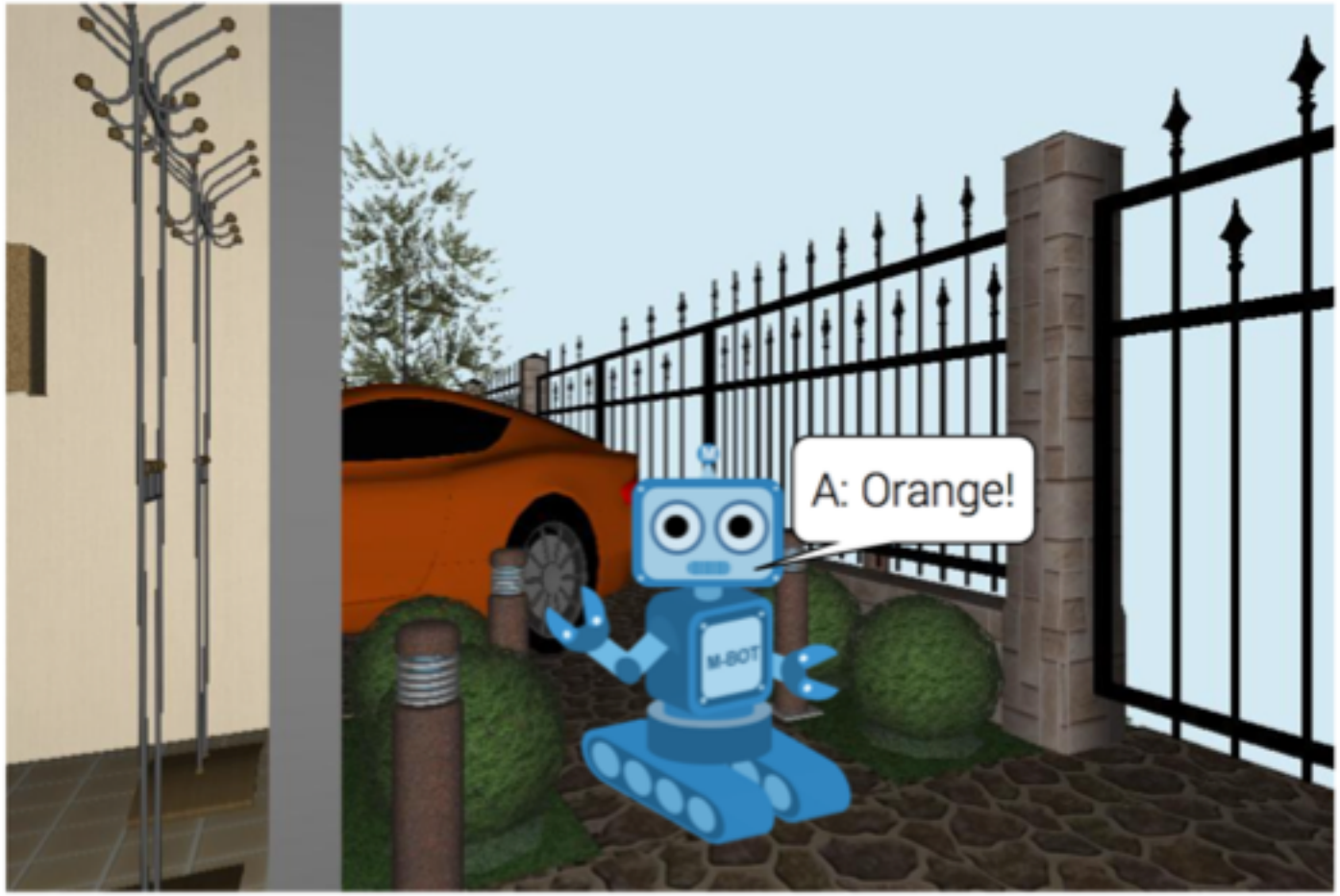
Language and Vision to Actions



https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions



回答も
自然言語で、
音声で

https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

ns

EQA Dataset: Questions in Environments

github.com/facebookresearch/**House3D**

Yi Wu, Yuxin Wu, Georgia Gkioxari, Yuandong Tian

Interactive environment built on the SUNCG dataset (Song et al., 2017)

https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

Language and Vision to Actions

AI2-THOR (Kolve et al., 2017)

https://drive.google.com/open?id=1G9TMQT6Rx-ekTO0w30iYd8yS4noDw3Nw

video credit: AI2-THOR

Q：
電子レンジの中に
コップ入ってる？

Copyright © 2016 Yahoo Japan Corporation. All Rights Reserved.

La

Leave the bedroom, and enter the kitchen. Walk forward, and take a left at the couch. Stop in front of the window.

https://drive.google.com/open?id=15QmLgcndu4bYMCxVpEA8EcvEEDILVmmq

「指示に従わせる」系
navigationタスク

Environments

	Underlying Dataset	Photo-Realism	Cognitive Realism	Physics / Actionability	Customizable	Large-scale
House3D (Wu et al., 2017)	SUNCG	✗	✓	✗	✓	✓
HoME (Brodeur et al., 2017)	SUNCG	✗	✓	✓	✓	✓
MINOS-SUNCG (Savva et al., 2017)	SUNCG	✗	✓	✗	✓	✓
AI2-THOR (Kolve et al., 2017)	Unity assets	✓	✓	✓	✓	✗
Matterport3DSim (Anderson et al., 2018)	Matterport3D	✓	✓	✗	✗	✗
MINOS-Matterport3D (Savva et al., 2017)	Matterport3D	✓	✓	✗	✗	✗

https://drive.google.com/open?id=1NuX64KqJPBrzcfk83MXQcAnwhTo68DRt

Machine Reading for Question Answering

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Machine Reading for Question Answering

- ワークショップの一つ
- 「質問応答のためのmachine reading技術」
- Gao (MSR)、Blunsom (DeepMind)、Socher (Salesforce)などの招待講演
- サイト
  - https://mrqa2018.github.io/

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

SQuAD

- Stanford Univ., Rajpurkarらによる質問応答データセット・タスク
- ver 1.1で 10万問、ver 2.0で「答えがない」設問を含む 5万問を追加
  - Know What You Don't Know: Unanswerable Questions for SQuAD (Rajpurkar et al.)
- ImageNetの質問応答版（あるいは自然言語処理版）な様相、industryからのタスク参加も活発
- サイト： https://rajpurkar.github.io/SQuAD-explorer/

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

SQuAD

- SQuAD: 100,000+ Questions for Machine Comprehension of Text (Rajpurkar et al.), 2016
- Wikipediaから、良質と思われる536記事をピックアップ、その中の各パラグラフについてcrowd workerが質問・回答ペアを考える形で作成

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

SQuAD

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity. The main forms of precipitation include drizzle, rain, sleet, snow, graupel and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals within a cloud. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

文書中に正解が埋まっており、そのスパンを指定する形で回答

https://arxiv.org/pdf/1606.05250.pdf

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

SQuAD

SQuAD 1.1

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar et al. '16)	82.304	91.221
1 Jul 11, 2018	QANet (ensemble) *Google Brain & CMU*	**84.454**	**90.490**
2 Jul 08, 2018	r-net (ensemble) *Microsoft Research Asia*	84.003	90.147
3 Jun 20, 2018	MARS (ensemble) *YUANFUDAO research NLP*	83.982	89.796
4 Mar 19, 2018	QANet (ensemble) *Google Brain & CMU*	83.877	89.737
5 Jun 21, 2018	MARS (single model) *YUANFUDAO research NLP*	83.122	89.224

https://rajpurkar.github.io/SQuAD-explorer/

YAHOO! JAPAN

SQuAD

SQuAD 2.0

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Aug 15, 2018	Reinforced Mnemonic Reader + Answer Verifier (single model) *NUDT* https://arxiv.org/abs/1808.05759	**71.699**	**74.238**
2 Aug 21, 2018	FusionNet++ (ensemble) *Microsoft Business Applications Group AI Research* https://arxiv.org/abs/1711.07341	70.345	72.555
3 Aug 21, 2018	SAN (single model) *Microsoft Business Applications Research Group* https://arxiv.org/abs/1712.03556	68.653	71.441
4 Jul 13, 2018	VS^3-NET (single model) *Kangwon National University in South Korea*	68.438	71.282

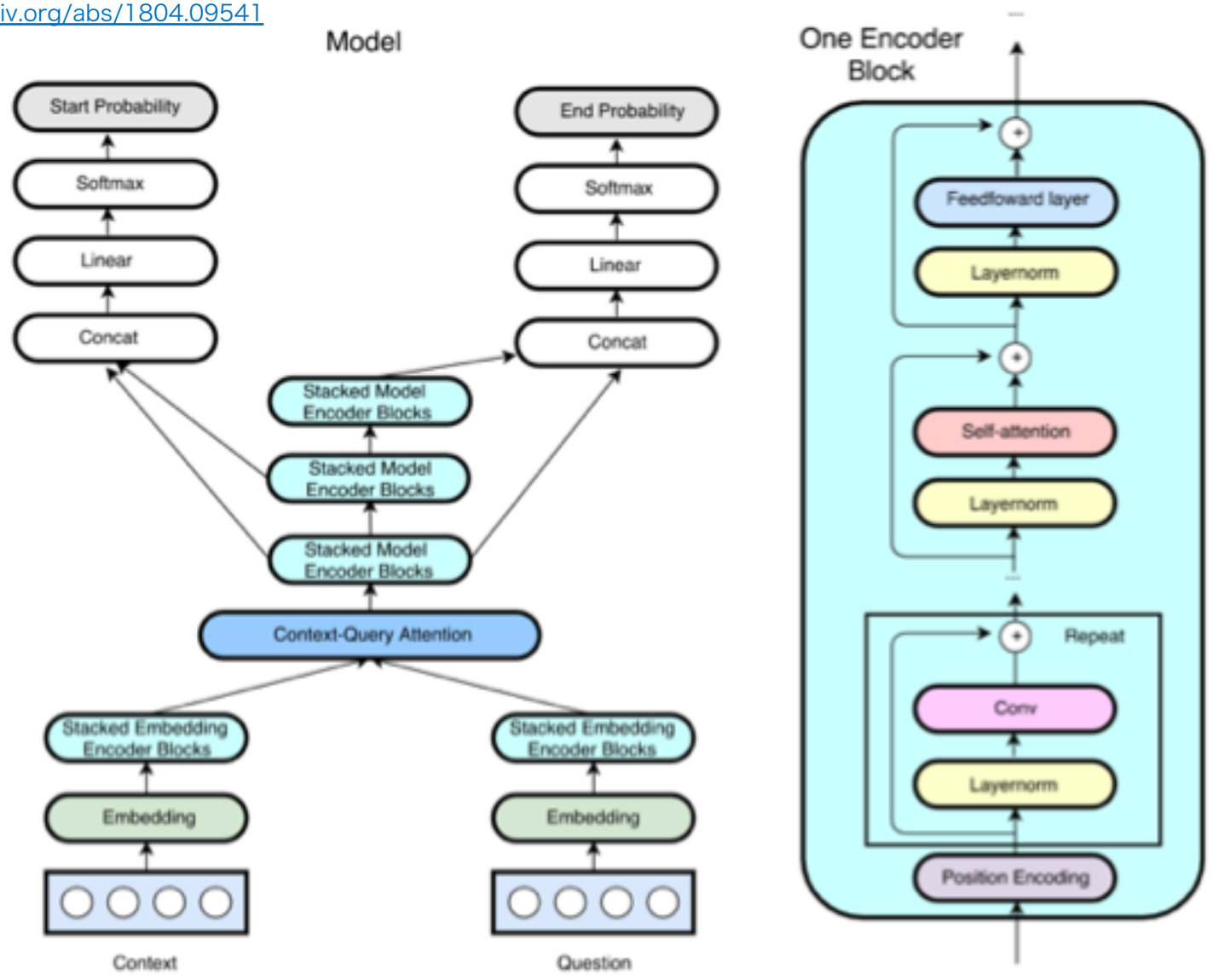
https://rajpurkar.github.io/SQuAD-explorer/

SQuAD — 既存の手法・モデル

QANet

トータル 130層

https://arxiv.org/abs/1804.09541

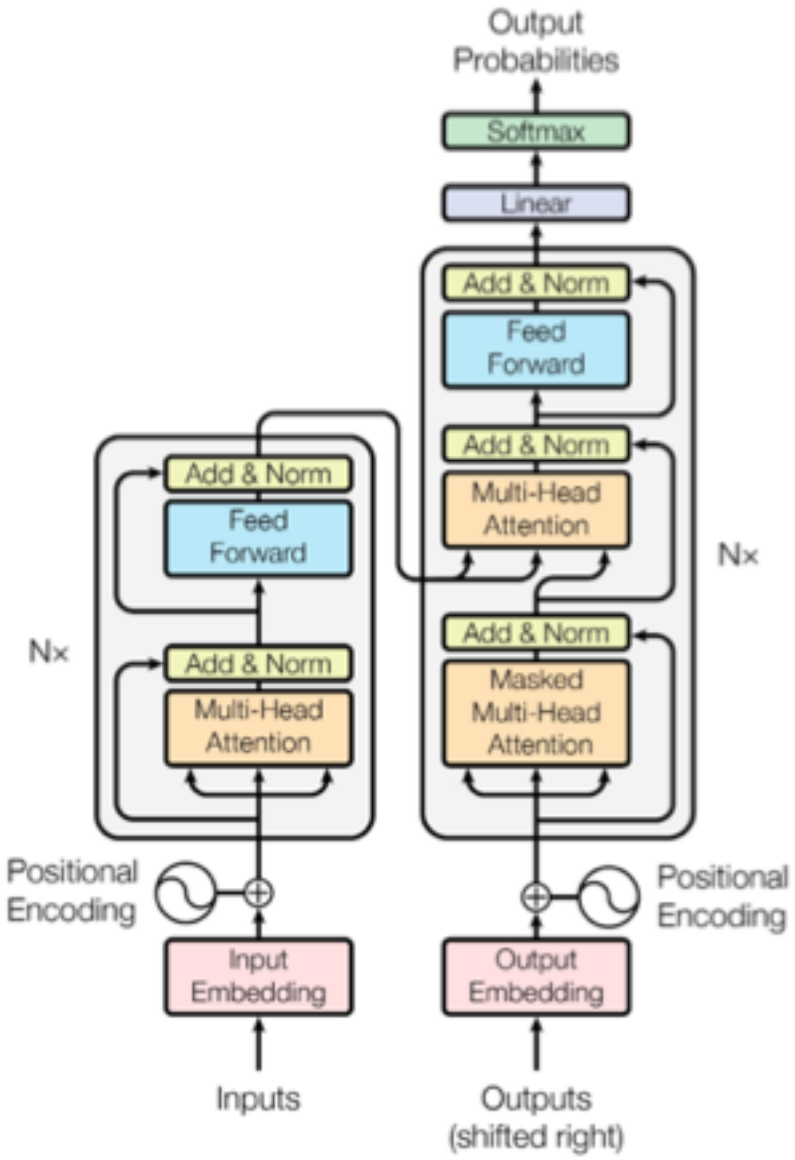


Transformer

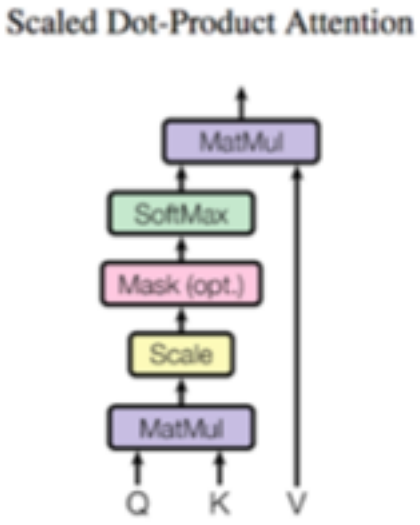
- Attention Is All You Need (Vaswani et al.), 2017
- 機械翻訳タスクに Transformerを導入、CNNもRNNも使わずに当時のSOTAを実現
- self-attention
- URL： https://arxiv.org/pdf/1706.03762.pdf

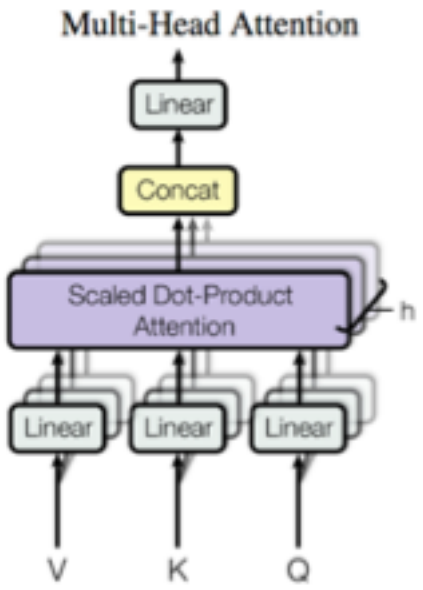
Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Transformer



ある層の計算を行う際、直下の層全体にattentionをかけ、サマリーする形で各要素を計算（self-attention）

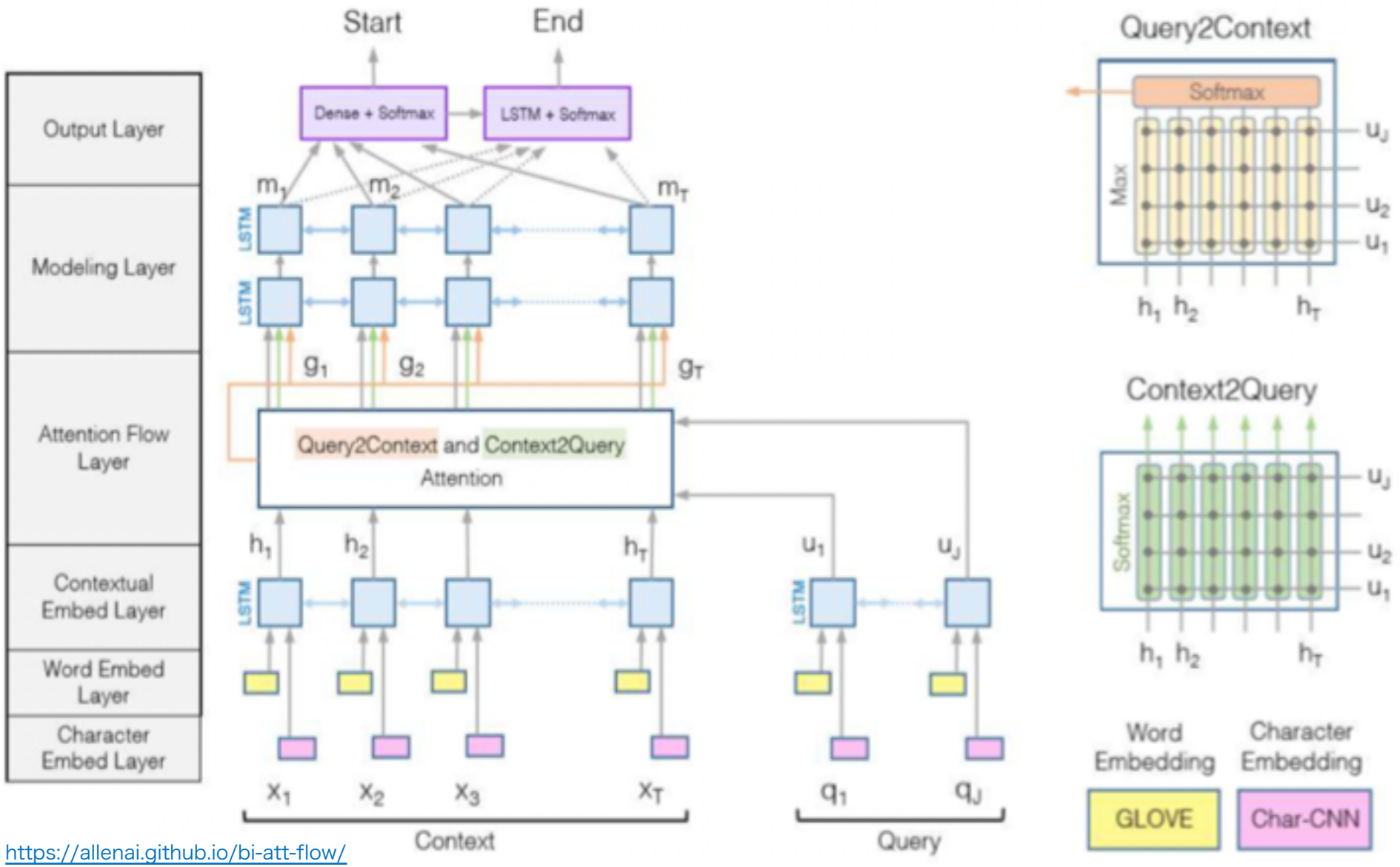




https://arxiv.org/pdf/1706.03762.pdf

SQuAD — 既存の手法・モデル

BiDAF

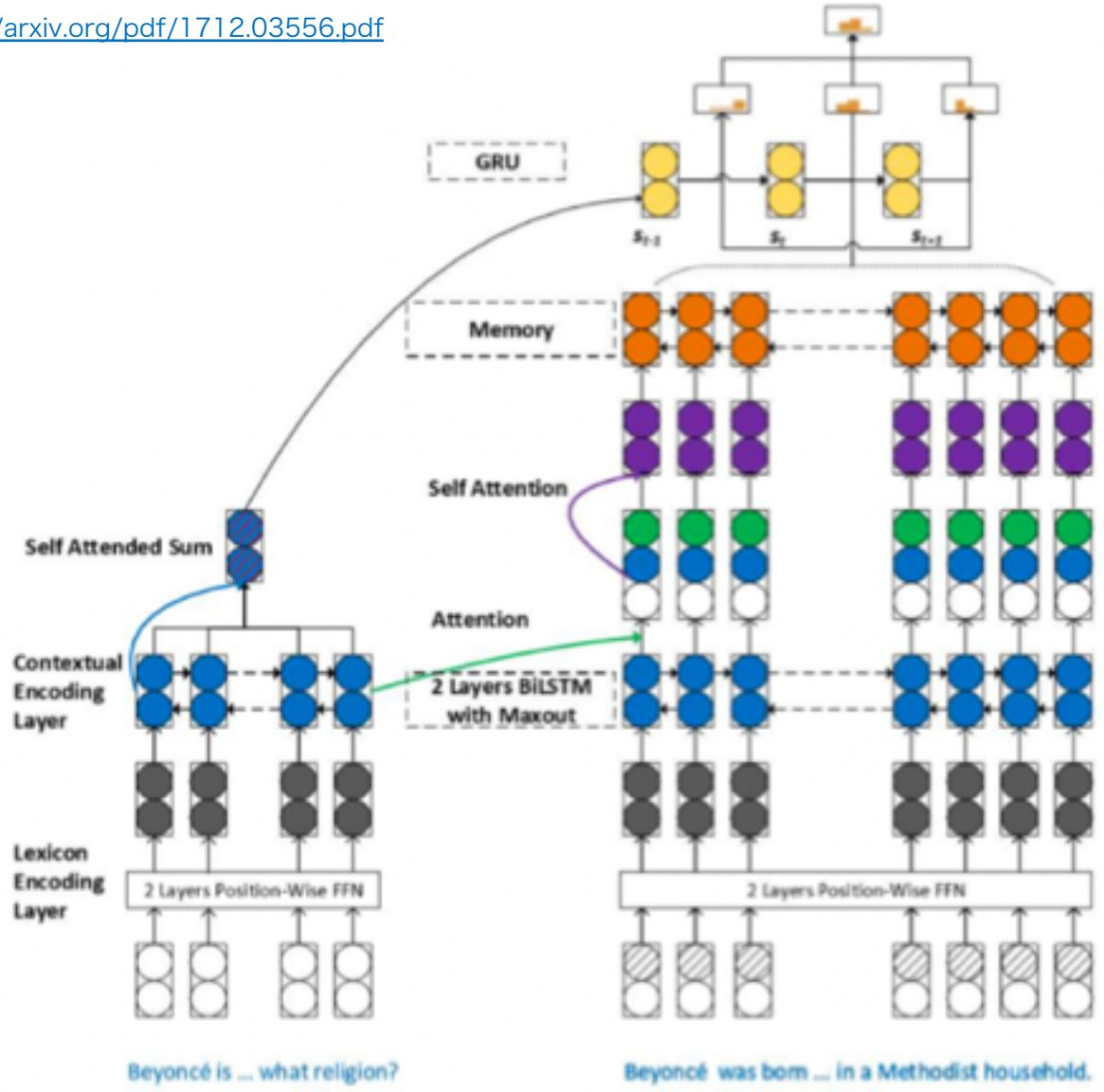


https://allenai.github.io/bi-att-flow/

SQuAD — 既存の手法・モデル

SAN

https://arxiv.org/pdf/1712.03556.pdf



Natural Language Decathlon

- MRQAにおける R. Socherの招待講演
- 10種類の異なる教師ありNLPタスクを、question answeringの枠組みに載せて 1つのモデルで対応
  - 各タスクが 他に対するpretrainingになるようなセッティング
- モデルの総合力を多角的に評価

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Natural Language Decathlon

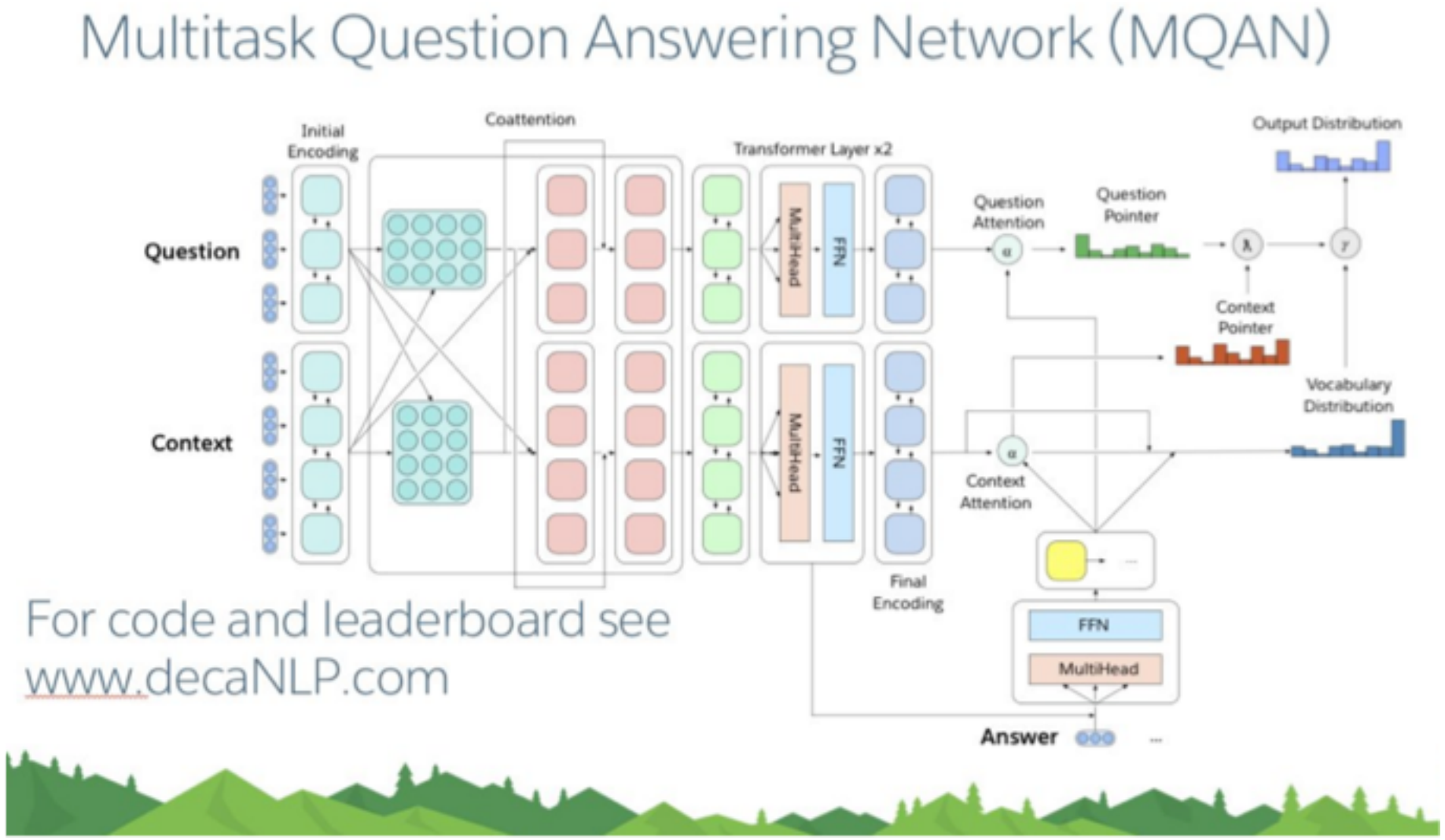
Multitask Learning as Question Answering

- Question Answering
- Machine Translation
- Summarization
- Natural Language Inference
- Sentiment Classification
- Semantic Role Labeling
- Relation Extraction
- Dialogue
- Semantic Parsing
- Commonsense Reasoning

- Meta-Supervised learning: From {x, y} to {x, t, y} (t is the task)
- Use a question, q, as a natural description of the task, t, to allow the model to use linguistic information to connect tasks
- y is the answer to q and x is the context necessary to answer q

https://twitter.com/RichardSocher/status/1021918049136304128/photo/1

Natural Language Decathlon



https://twitter.com/RichardSocher/status/1021918049136304128/photo/1

Winograd Schema Challenge

- Turing test的な、machine intelligenceの評価手法の一つとして、2014年 E. Goostmanにより提案
- 言語面、commonsense knowledge / reasoningにフォーカス

The trophy would not fit in the brown suitcase because it was too **big** (*small*).

What was too **big** (*small*)?

Answer 0: the trophy

Answer 1: the suitcase

- URL： http://commonsensereasoning.org/winograd.html

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Pretraining Sentiment Classifiers with Unlabeled Dialog Data

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

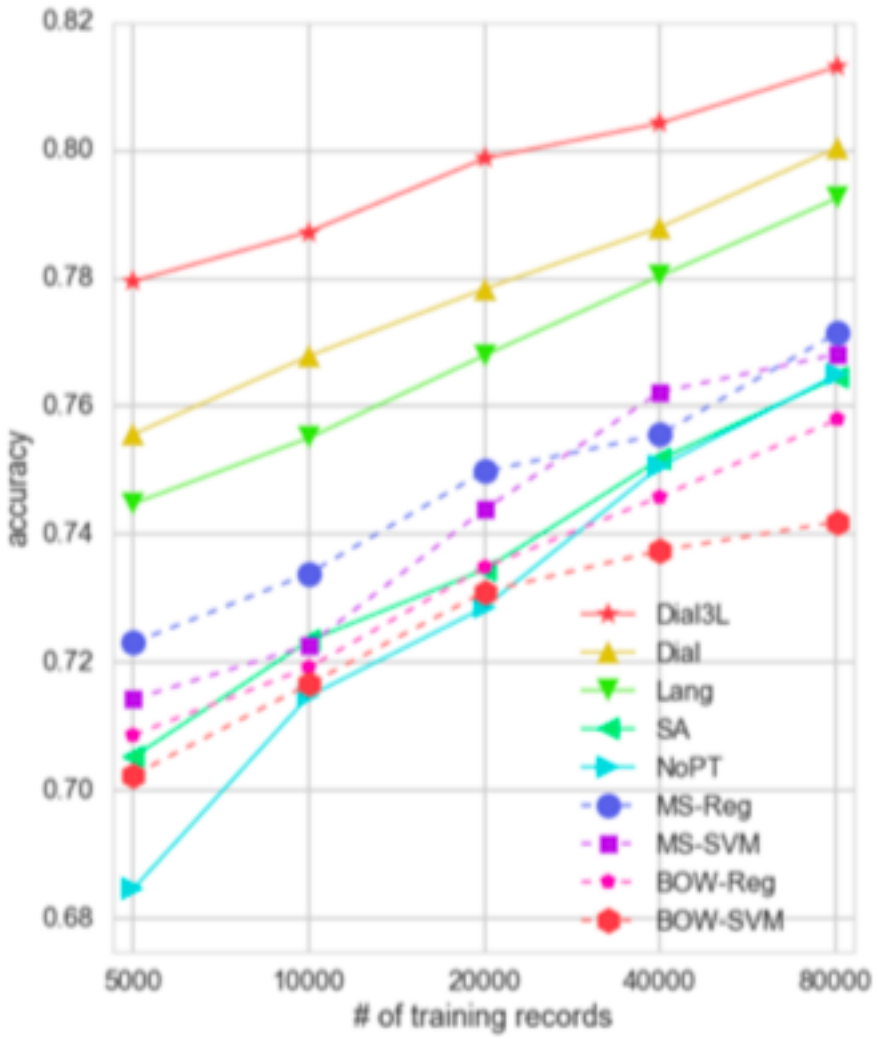
Pretraining Sentiment Classifiers with Unlabeled Dialog Data

- 計算力的には 大規模なモデルの学習が可能になっている一方、ラベル付きデータの作成は大変なまま、量としても小規模にとどまる
- タスク： tweet感情分析
- semi-supervisedな手法を追求
- 対話データを用いた事前学習が特に効果的であることを 先行研究との比較で示す
- URL： https://aclanthology.info/papers/P18-2121/p18-2121

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Pretraining Sentiment Classifiers with Unlabeled Dialog Data

	5k	10k	20k	40k	80k
Dial3L	**77.96**	**78.73**	**79.82**	**80.42**	**81.31**
Dial	75.57	76.79	77.84	78.80	80.04
Lang	74.49	75.51	76.80	78.04	79.26
SA	70.53	72.34	73.45	75.18	76.46
DSE2M	67.71	68.88	70.47	73.08	75.75
DSE2M-Dial	70.72	72.47	74.67	76.01	77.94
DSE2M-Lang	71.28	72.36	74.10	75.55	77.58
DSE6M	67.79	68.47	70.42	72.72	74.86
DSE6M-Dial	69.55	71.34	72.98	75.07	77.02
DSE6M-Lang	70.68	72.33	74.01	75.52	77.30
NoPT	68.47	71.48	72.86	75.07	76.50
MS-Reg	72.31	73.39	74.99	75.57	77.15
MS-SVM	71.43	72.27	74.39	76.23	76.82
BOW-Reg	70.87	71.93	73.49	74.59	75.80
BOW-SVM	70.25	71.67	73.11	73.75	74.20



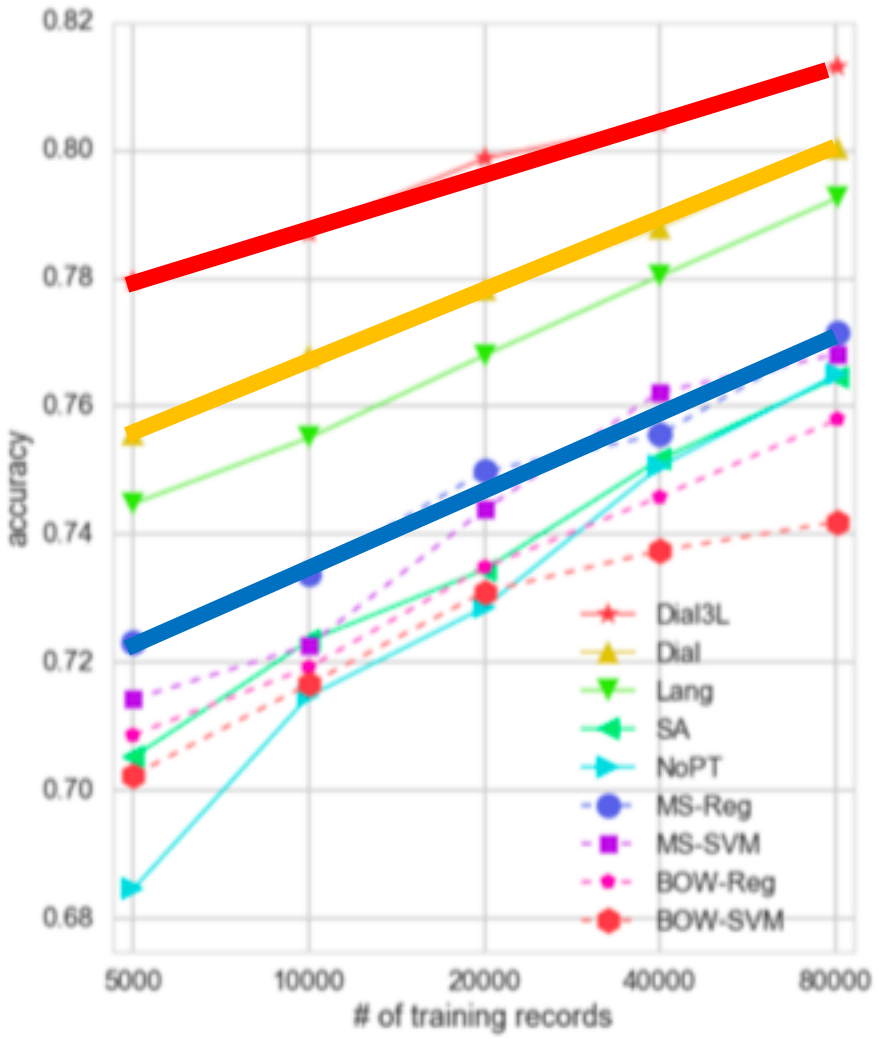
Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Pretraining Sentiment Classifiers with Unlabeled Dialog Data

提案手法が他を上回っている
だけでなく

よりデータ量の少ないところ
でこそ、事前学習を経たLSTM
大規模モデルが優位に

					80k
					1.31
					0.04
					9.26
					6.46
					5.75
					7.94
					7.58
					4.86
					7.02
					7.30
NoPT	68.47	71.48	72.86	75.07	76.50
MS-Reg	72.31	73.39	74.99	75.57	77.15
MS-SVM	71.43	72.27	74.39	76.23	76.82
BOW-Reg	70.87	71.93	73.49	74.59	75.80
BOW-SVM	70.25	71.67	73.11	73.75	74.20



LSTM 3層

LSTM 1層

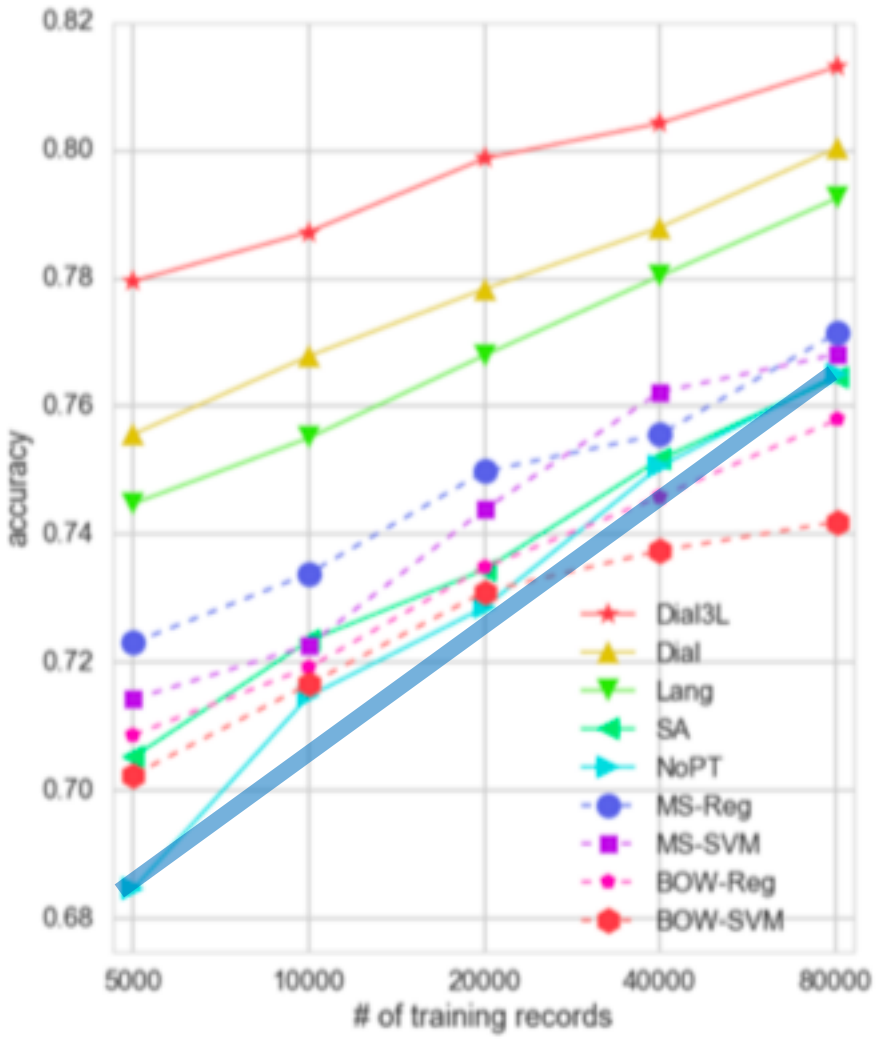
旧来の
手法

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Pretraining Sentiment Classifiers with Unlabeled Dialog Data

事前学習なしだと、比較の土俵に乗らない

	5k	10k	20k	40k	80k
[illegible]	[illegible]	[illegible]	[illegible]	**[illegible].42**	**81.31**
[illegible]	[illegible]	[illegible]	[illegible]	[illegible].80	80.04
[illegible]	[illegible]	[illegible]	[illegible]	[illegible].04	79.26
[illegible]	[illegible]	[illegible]	[illegible]	[illegible].18	76.46
[illegible]	[illegible]	[illegible]	[illegible]	[illegible].08	75.75
DSE2M-Dial	70.72	72.47	74.67	76.01	77.94
DSE2M-Lang	71.28	72.36	74.10	75.55	77.58
DSE6M	67.79	68.47	70.42	72.72	74.86
DSE6M-Dial	69.55	71.34	72.98	75.07	77.02
DSE6M-Lang	70.68	72.33	74.01	75.52	77.30
NoPT	68.47	71.48	72.86	75.07	76.50
MS-Reg	72.31	73.39	74.99	75.57	77.15
MS-SVM	71.43	72.27	74.39	76.23	76.82
BOW-Reg	70.87	71.93	73.49	74.59	75.80
BOW-SVM	70.25	71.67	73.11	73.75	74.20



LSTM
事前学習
なし

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.

Universal Language Model Fine-tuning for Text Classification

- J. Howard, S. Ruder
- AWD-LSTMベースの言語モデリングによる事前学習、教師あり分類タスクへの転移
  - 枠組みとしては Dai and Li 2015にあった事前学習と同様
  - gradual unfreezing、slanted triangular learning rate などのテクニックを導入し、複数タスクで SOTA
- URL： http://aclweb.org/anthology/P18-1031

Copyright © 2018 Yahoo Japan Corporation. All Rights Reserved.