紹介する論文

- The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision
- 画像xNLPのマルチモーダル
- 少量のデータでも高い精度を出せる
- Wikipedia: Knowledge-Powered Conversational Agents
  - QAシステムに関する研究
  - 絵文字を使ってるので気になった
  - 大規模なデータセットの公開
The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision

Jiayuan Mao
Chuang Gan
Pushmeet Kohli
Joshua B. Tenenbaum
Jiajun Wu

MIT CSAIL and IIIS, Tsinghua University
MIT-IBM Watson AI Lab
Deepmind
MIT BCS, CBMM, CSAIL
MIT CSAIL
ざっくり言うと

- Visual QAタスクにおいてend-to-endで学習を行う手法 (Neuro-Symbolic Concept Learner)を提案
  - オブジェクトに対するラベルなどのデータなしで、視覚的な概念(形や色)の認識、言語の意味解析を行う
  - 少量の学習データでも精度が高い
  - 回答を出力するまでのプロセスが説明可能

Figure 1: An example image-question pair from the VQS dataset and the corresponding execution trace of the proposed NS-CL.
Q: What is the shape of the red object?

Answer: Cylinder
Groundtruth: Box

学習時のインプットは、画像と、画像に対する質問文(Q)、正解(Groundtruth)のみ
Neuro-Symbolic Concept Learnerの概要

Q: 何色のオブジェクトですか？
A: 緑色

Q: 紅色のオブジェクトの形状は何ですか？
A: 球体
Neuro-Symbolic Concept Learnerの概要

① 画像に対してオブジェクト検出が行われ、オブジェクトを表現するベクトル Visual Representationを得る

Q: 何の形をしている赤い物体か？

A: 球体

Semantic parsing

Program
Query(Shape, Filter(Red))

Neuro-Symbolic Reasoning
Q: What is the shape of the red object?

A: Sphere

2. Qに対して、コンセプト（形や色など）を表すベクトル Concept Embeddingを得る
3. Qに回答するためのプログラム列を得る
Neuro-Symbolic Concept Learnerの概要

Q: What is the shape of the red object?  

A: Sphere

Semantic parsing

Object detection(Mask R-CNN)  
Feature extraction(ResNet-34)

Program
Query(Shape, Filter(Red))

④ Neuro-Symbolic ReasoningモジュールがVisual Representation, Concept Embeddings, Programsを受け取り、回答を出力する
Qに対して正しいProgramを出力できた場合、Programは以下の2ステップとなる

(1) 赤色のオブジェクトでフィルターする
(2) 赤色のオブジェクトの形を求める
Q: What is the shape of the red object?

Neuro-Symbolic Reasoning モジュール

(1) 赤色のオブジェクトでフィルターする

Visual RepresentationからColor Representationを得る。

Concept Embeddingsの赤を表すベクトルとのcos類似度を計算し、
最も類似度の高いオブジェクトに絞り込む

Q: What is the shape of the red object?
Q: What is the shape of the red object?

Visual RepresentationからShape Representationを得る。

Obj2に対する形を表すベクトルと、Concept Embeddings中の形を表すベクトルとを比較し最もcos類似度の高い形を出力する

1. Object detection (Mask R-CNN)
2. Feature extraction (ResNet-34)
3. Semantic parsing
4. Program
   
   Query(Shape, Filter(Red))

形のベクトルに変換するNN

Shape Representation

Sphere

Cube

Concept Embeddings:
- cube
- sphere

Q: What is the shape of the red object?
Neuro-Symbolic Reasoningの学習

Visual Representation, Concept Embeddingsに対してはBack Propagation, Programに対しては強化学習で学習をすすめる

Q: What is the shape of the red object?

A: Cube

Groundtruth: Sphere

Semantic parsing

Program

Query(Shape, Filter(Red))

Object detection(Mask R-CNN)
Feature extraction(ResNet-34)

Backprop

Reinforce
一連の学習はCurriculum Learningで実行される。

最初は簡単なデータ
（オブジェクト数などが少ないもの）
で学習をし、適切に回答を出力できるように
だったら次の難易度に移行し学習が行われる。

A. Curriculum concept learning

- **Initialized** with DSL and executor.

- **Lesson1**: Object-based questions.
  
  Q: What is the shape of the red object?
  A: Cube.

- **Lesson2**: Relational questions.
  
  Q: How many cubes are behind the sphere?
  A: 3

- **Lesson3**: More complex questions.
  
  Q: Does the red object left of the green cube have the same shape as the purple matte thing?
  A: No

- **Deploy**: complex scenes, complex questions
  
  Q: Does the matte thing behind the big sphere have the same color as the cylinder left of the small matte cube?
  A: No.
実験

- CLEVRデータセットに対して精度評価
  - 立方体や球などが複数配置された画像と、質問、回答がペアになったデータセット

- 訓練データ: 70K, 開発データ: 15K, 評価データ: 15K
  - 訓練データの全てを使用した場合、一部(7K)を使用した場合でそれぞれ精度比較
実験

100%のデータで99.6%の精度、
10%のデータ、7Kのイメージと70Kの質問ペアでも98.9%の精度を達成。
考察

Programにすることで、回答を出力するまでのプロセスを明示することができ、説明可能になる

| Concept     | Program    | Result       |
|-------------|------------|--------------|
| Gray Cylinder | Filter     |              |
| Behind      | Relate     |              |
| Cyan Cylinder | Filter     |              |

Example A.

Q: Do the cyan cylinder that is behind the gray cylinder and the gray cylinder have the same material?

| Concept     | Program | Result |
|-------------|---------|--------|
| Gray Cylinder | Filter  | Yes (0.92) |

Example B.

Q: There is a small blue object that is to the right of the small red matte object; what shape is it?

| Concept     | Program | Result       |
|-------------|---------|--------------|
| Small Red Matte Object | Filter  |              |
| Right       | Relate  |              |
| Small Blue Object | Filter  |              |
| Shape       | Query   | Cube (0.85)  |
まとめ

- 視覚的な概念（形や色）、単語や文の意味解析をそれぞれ個々に学習する手法 Neuro-Symbolic Concept Learnerを提案

  - 必要なのは画像とQとAのみ

- 少ない学習データ量でも高い精度を実現

- 回答を出力するまでのプロセスを説明できることを示した
of Wikipedia: Knowledge-Powered Conversational Agents

Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, Jason Weston
(Facebook AI Research)
ざっくり言うと

- Wizard of Wikipediaデータセットの作成、公開
  - 知識を問うような質問のデータセット、評価指標がない
  - Wikipediaから知識を検索し、会話を行う作業を実施し作成
- 質問応答システムの手法 Transformer Memory Network を提案
  - オープンドメインな対話において、上手く知識を扱えるようにしたいが、現状の対話システム(e.g. seq2seq, transformerなど)では、知識や記憶保持に限界がある
The Wizard of Wikipedia dataset

- 二人の人間がチャットで雑談をする際のデータを収集する
- 知識を与える人といろいろ聞く人の対話データ
- 約21万発話のデータが収集できた

| Topic: Armadillo |
|------------------|
| Wizard: I love animals and think armadillos are awesome with their leathery shell. |
| Apprentice: I don’t think I’ve ever seen an armadillo in real life! |
| Wizard: I’ve seen them at the zoo. Armadillo means little armored one in Spanish. |
| Apprentice: Are they native to a Spanish-speaking part of the world? |

Knowledge:

- Armadillos are New World placental mammals in the order Cingulata ...
- The word “armadillo” means “little armoured one” in Spanish.

- The nine-banded armadillo ("Dasypus novemcinctus"), or the nine-banded, long-nosed armadillo, is a medium-sized mammal found in North, Central, and South America.

Wizard: Yes, they are most commonly found in North, Central, and South America
発話生成の手法で得られた発話例

Model: Two-Stage Generative Transformer MemNet

| Topic: Circus (seen) | Model: have you ever been to a circus? it is a company of performers who put on diverse entertainment shows. |
|----------------------|---------------------------------------------------------------------------------------------------------|
| Human:               | I went to the circus once when I was a child actually.                                                    |
| Model:               | that’s great! entertainment is a form of activity that holds the attention and interest of an audience. |
| Human:               | That’s true. Have you ever been to a circus?                                                              |
| Model:               | I have not, but I know they were built with various types of seating ring, and sometimes a stage.         |
| Human:               | I went to one held in a stadium, so we had box seats.                                                      |
| Model:               | that’s awesome! i’m sure you’ll have to go to the circus, with appropriate music.                         |
| Human:               | Yes, circuses usually have their own kind of music.                                                        |
| Model:               | circuses can be done by circuses.                                                                         |

| Topic: Granny Smith (seen) | Model: I really love Granny Smith apples, they’re my favorite type of apple. |
|-----------------------------|--------------------------------------------------------------------------------|
| Human:                      | I love granny smith apples. they have hard, light green skin and a crisp flesh. |
| Model:                      | Yes, I really enjoy them. I also like Honeycrisp apples but they’re so expensive! |
| Human:                      | they’ve been grown for thousands of years in asia and europe, and were brought to north america by european colonists (*) |
| Model:                      | Oh really? They’ve been around way longer than I thought! |
| Human:                      | they’re also consumed raw, it’s one of the most popular cooking apples. |
| Model:                      | i always eat them raw, I hate cooked apples for some reason, even in pie. |
| Model:                      | i’m not sure if it’s cooked, but I do know that they can be cooked raw. |
データセットのリンク

https://parl.ai/projects/wizard_of_wikipedia/