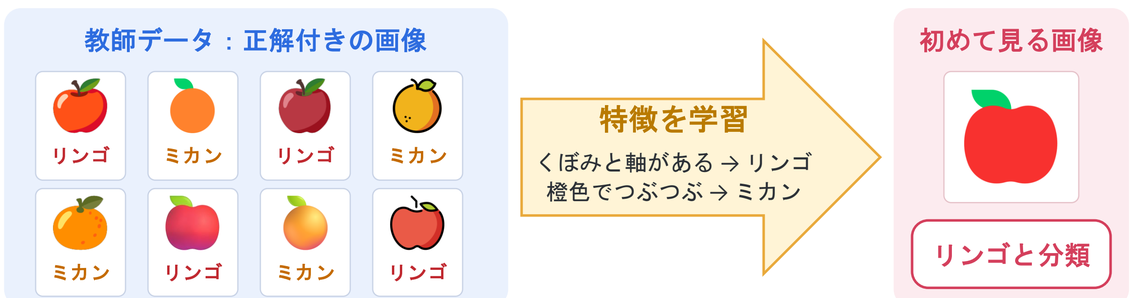
ヒント注入とロールバックによる少数ショットテキスト含意の学習安定化

渡辺颯太　　電気通信大学 情報理工学域 I類 CSプログラム　兼岩研究室

1 機械学習とは

機械学習とは，人工知能（AI）に正解付きの例を大量に見せて学習させる方法である．この例を教師データと呼ぶ．教師データによって，人工知能は例に共通する特徴をつかみ，初めて見るデータも分類できるようになる．



教師データが多いほど，人工知能は正確に分類できる

2 人工知能に読解問題を解かせる

本研究では，物語文を読み，質問の答えを4つの選択肢から選ぶ読解問題を扱う．データにはMCTest500 [1] を用いる．

前提文

ある日，子犬のポピーは外へ遊びに出かけた．ポピーは庭を全速力で走り回った．走ったあと，ポピーはボールで遊び始めた．それに飽きると，ポピーはメアリーとキャッチボールをした．メアリーが一緒に遊びたがってくれて，ポピーはとてもうれしかった．ポピーはメアリーに飛びついて，顔をなめ始めた．ポピーが愛情いっぱいに接してくれて，メアリーもうれしかった．そのあと，メアリーとポピーは家に入り，ごはんを食べて昼寝をした．

質問：メアリーの犬の名前は何か．

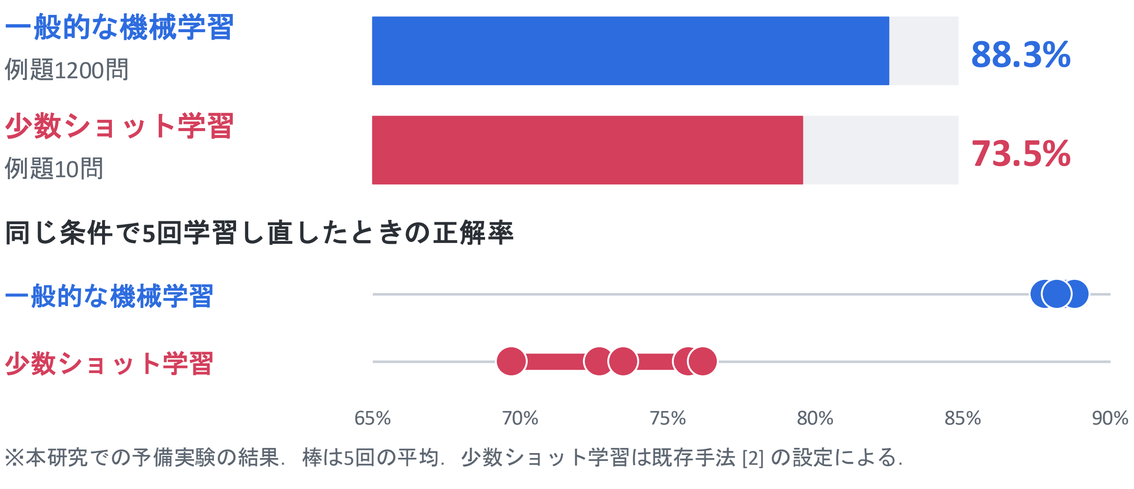
パピー	ハッピー	ジャンプ	ポピー　正解

※実際の問題は英文である．

人工知能は，選択肢を当てはめた文が前提文から導けるかを「はい／いいえ」で判定し，最も「はい」らしいものを答えとする．これをテキスト含意と呼ぶ．

3 少数ショット学習とその難しさ

一般的な機械学習では，大量の例を教師データとして用いる．**ごく少ない例題で学習すること**を少数ショット学習と呼び，本研究では10問の例題だけで600問のテストを解く．



※本研究での予備実験の結果．棒は5回の平均．少数ショット学習は既存手法 [2] の設定による．

例題が少ないと正解率が下がり，学習のたびに結果が大きくぶれる

4 既存の少数ショット学習：UFO-ENTAIL

UFO-ENTAIL [2] は，**10問の例題だけ**で学習する手法である．例題から「はい」と「いいえ」の平均像を作り，テスト問題がどちらに近いかで判定する．

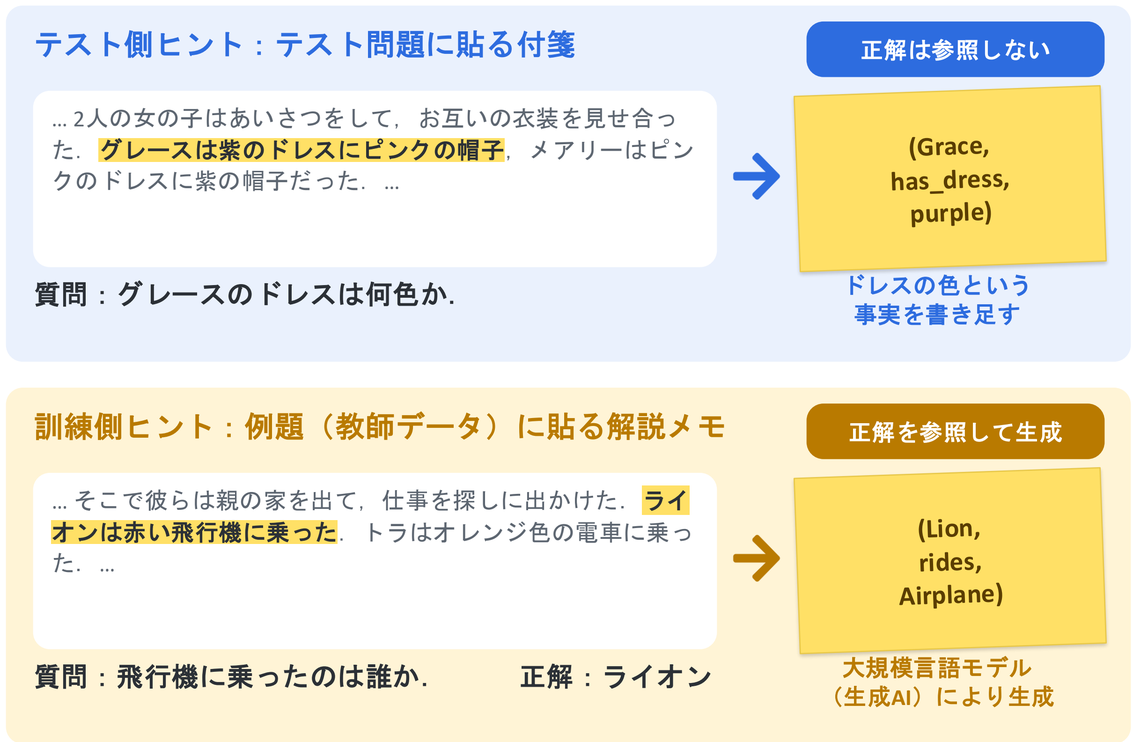


論文での正解率：10問の例題で74.2%

手がかりが10問分しかないため，判断を誤りやすく学習も不安定になる

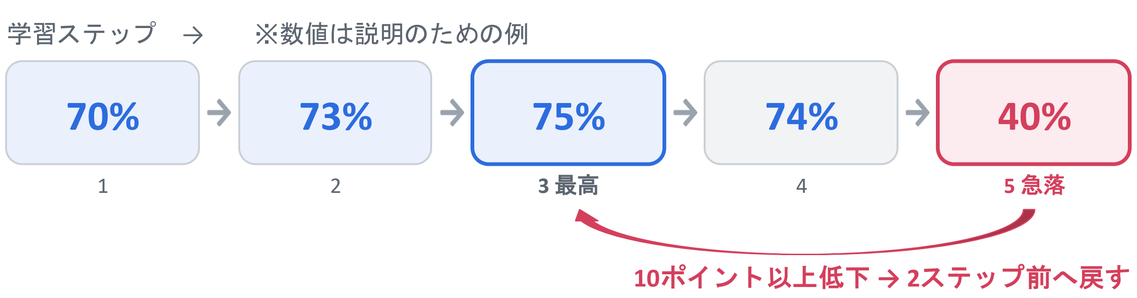
5 提案手法①：ヒント注入

前提文から質問に関係する事実を取り出し，ヒントとして入力に書き足す．付箋を貼るイメージであり，貼る場所は次の2つである．



6 提案手法②：ロールバック

例題が少ないと，学習の途中で正解率が急に下がることがある．そこで，ゲームのセーブとロードのように，正解率が急に下がった場合は少し前の状態に戻す．

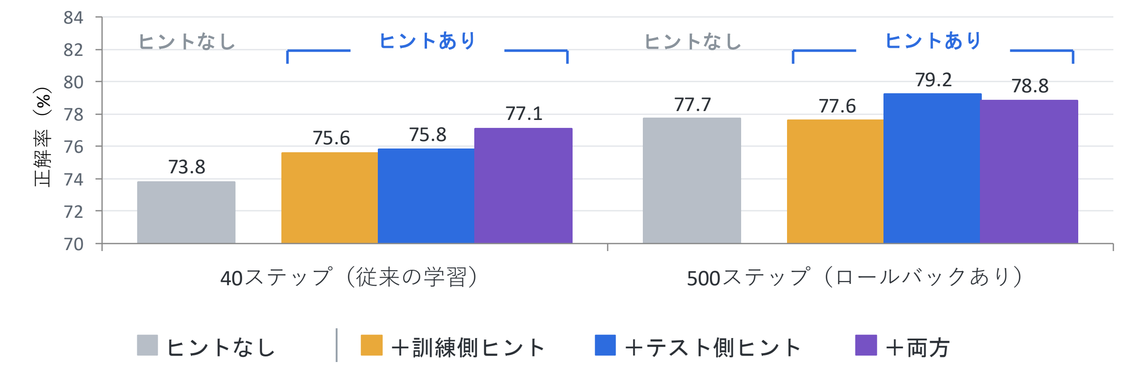


急落しても立て直せるため，500ステップまで学習を続けられる

7 実験結果

テスト側ヒントにより **＋11問**（77.3% → 79.2%，5回の平均）

ロールバックにより **＋24問**（73.0% → 76.9%，10回の平均）



40ステップ：どちらのヒントも有効であり，両方を用いた場合が最高（77.1%）

500ステップ：テスト側ヒントのみの場合が最高（79.2%）

1200問で学習した場合の88.3%との差を，例題10問のまま11.0ポイントから9.1ポイントに縮めた．※グラフは3回の平均．テストは600問．縦軸は70%から．

8 まとめと今後の課題

- テスト側ヒントは，正解を参照せずに正解数を増やした
- 訓練側ヒントは，短い学習でのみ効果が見られた
- ロールバックにより，学習途中の正解率の低下を抑えた

今後の課題：ヒントが正解と一致しなくても効く要因の解明

[1] M. Richardson, C. J. C. Burges, E. Renshaw. MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text. EMNLP 2013.
[2] W. Yin, N. F. Rajani, D. Radev, R. Socher, C. Xiong. Universal Natural Language Processing with Limited Annotations: Try Few-shot Textual Entailment as a Start. EMNLP 2020.
果物の絵：OpenMoji (CC BY-SA 4.0)，Microsoft Fluent Emoji (MIT License)，Noto Emoji (Apache License 2.0)