

背景・取り組み

- 教育現場では、国語の読解や理科の記述問題などを中心に短文記述式問題が広く扱われている
- 実用上の期待を受けながらも、短文記述式答案の自動採点 (SAS) に関する研究は少ない
 - データセットの欠如が最大の要因
- 初の日本語SASデータセットを提供 (Mizumoto+'19)^[1] (2020年7月から学術利用に限り提供開始)
 - 採点項目ごとの項目点や得点の根拠を示す根拠箇所が付与されたSASデータセットは世界初
 - 計29問 (97 採点項目)、各問ごとに500~2000答案を有する世界最大規模のSASデータセット

データ例 (問題、答案)

問題

傍線部(1)「こうした～築いてきた」とあるが、それはどういうことか。70字以内で説明せよ。

採点基準 (一部抜粋、要約)

採点基準は非公開

A 言葉は他社との意思疎通を可能にする媒体であることが書けている…6点

B 言語には個別の経験を抽象化する作用があることが書けている…3点

C 世界との一体感が失われるということが書けている…6点

答案

項目点と根拠箇所がアノテートされている唯一のSASデータセット

言葉は様々な人とのやり取りを行う際の媒体となるが、その一方で

A 根拠箇所 (6点)

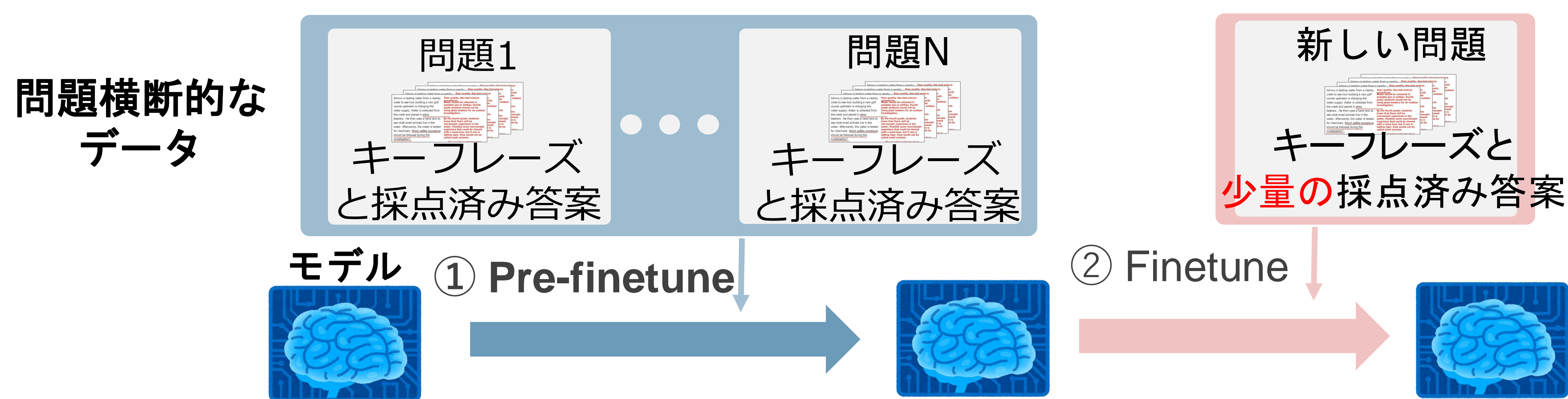
言葉は抽象的な記号でもあるので、世界との直接的な関わりを遮断してしまうこと。

B 根拠箇所 (3点)

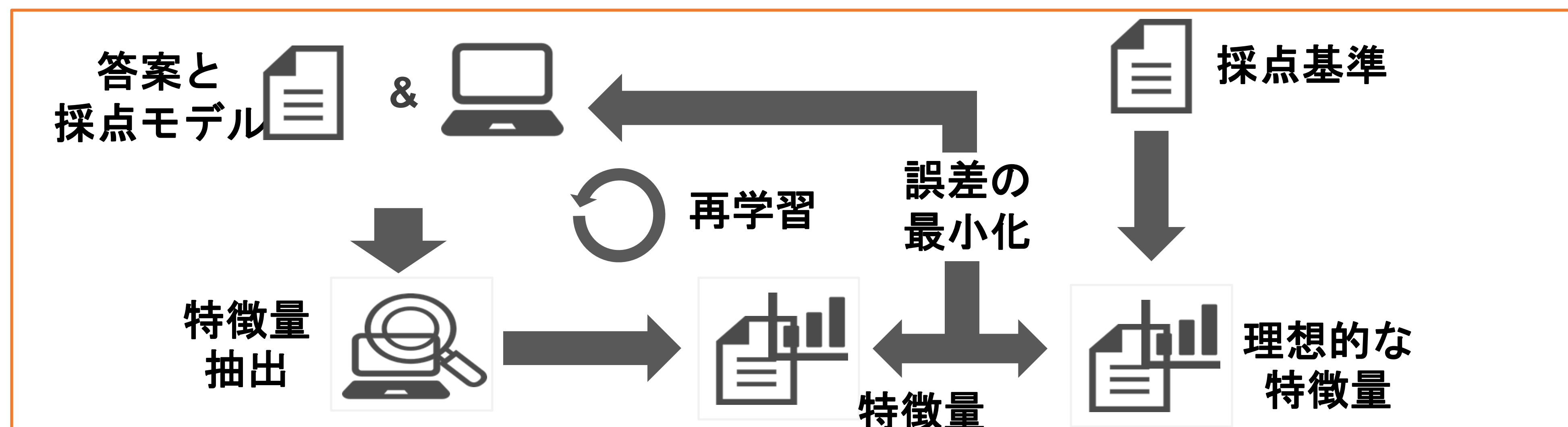
C 根拠箇所 (6点)

研究での使用例

- 問題横断的なSASモデルの学習 (Funayama+'23)^[2]
 - 採点基準 (キーフレーズ) を活用した問題横断的なモデルの学習により、性能を維持したままモデルの学習に要するデータ数を半減 (=低コストでのモデル構築)



- Short-cut learningの抑制 (Asazuma+'23)^[3]
 - モデルが得点予測に用いた答案の特徴量と採点基準を比較することで、superficial cueによるshort-cut learningを抑制する手法を構築



- 他にも、フィードバック生成 (Phan+'23; 岩瀬ら'23; Furuhashi+'24)、根拠箇所の活用 (Takano+'22; Sato+'22; Wang+'21)、学習データの削減・拡張 (加藤ら'20; 羽賀ら'23; 宮田ら'23; 春日ら'23)、不確かさの推定 (高橋ら'23) など

参考文献

[1] Tomoya Mizumoto, Hiroki Ouchi, Yoriko Isobe, Paul Reiser, Ryo Nagata, Satoshi Sekine, Kentaro Inui. Analytic Score Prediction and Justification Identification in Automated Short Answer Scoring, BEA 2019

[2] Hiroaki Funayama, Yuya Asazuma, Yuichiroh Matsubayashi, Tomoya Mizumoto and Kentaro Inui. Reducing the Cost: Cross-Prompt Pre-Finetuning for Short Answer Scoring, AIED2023

[3] Yuya Asazuma, Hiroaki Funayama, Yuichiroh Matsubayashi, Tomoya Mizumoto and Kentaro Inui. Adhere to the Rubric: A Method for Building Trustworthy Short Answer Scoring Models. HELMeTO 2023