

「非モデル生物の遺伝子論文」 選抜の性能検証



Takayuki Suzuki (鈴木貴之), Ryota Yamada (山田涼太) Science Aid株式会社, mail to: takayuki.suzuki@science-aid.com

1. 序論 (Introduction)

【背景】

非モデル生物は、生物多様性の理解や産業応用 (新規酵素、耐性など) の重要な研究対象  
ゲノム解析技術の発展により、非モデル生物のゲノム研究が増加<sup>1,2</sup>

論文データベースから非モデル生物に関する目的の研究や情報を効率的に収集する必要性増加

【課題】

初期スクリーニングには自動化ツールが用いられるが、下記のような問題により、最終的な論文選抜には人手による精査が必要であり、時間と専門知識を要する。

- 遺伝子-生物種の整合性検証が不十分
- 機械的な文脈理解の限界 (研究目的の判断、遺伝子言及の意味合いなど)

【目的】

本研究は、AIエージェントにより「非モデル生物の遺伝子論文」を選抜し、既存手法(PubTator)による選抜との性能比較により、その有用性を検証する。

4. 結果 (Results)

【比較実験結果】

従来手法 (PubTator) vs

AI agent (Biomni) vs AI agent + (改良版Biomni)

指標	従来法 (PubTator)	AI agent (Biomni)	AI agent + (Biomni + 3 new tools)
最終選抜論文数	55	75	44
Precision	60.00%	80.00%	90.91%
処理時間	NA	5.28h	6.09h
費用	NA	\$6.01	\$6.76
使用したモデル	GNorm2	GPT-4.1-mini	GPT-4.1-mini
(補足) 遺伝子	38.97%	59.72%	71.11%
NER/ELのPrecision			
(補足) 生物種	65.55%	63.98%	83.33%
NER/ELのPrecision			

- AI agent + が最も高い適合率を達成。
- 遺伝子・生物種の抽出およびID取得の性能も向上。
- 再現率、正解率については、未検証。

5. 考察 (Discussion)

【偽陽性 (FP) の分析】

手法2のFP (15件) → 手法3のFP (4件)へ削減 (73.3%減、Precision 90.9%)

手法3のFP原因：モデル生物関連(1件), 判断難しい(3件).

手法2のFP概要：遺伝子言及なし(6件), モデル生物関連(5件), 判断難しい(4件)

手法2のFP原因：生物種-遺伝子の整合性チェック不足

手法1のFP原因：ヒト遺伝子誤抽出(10件)、遺伝子言及の誤抽出(10件)

【偽陰性 (FN) の分析】

26件のFN候補を確認 (手法2で検出 → 手法3で見落とし)

35%が系統解析研究で、バーコード遺伝子 (COI、rbcL等) を使用

原因として、系統解析論文を「機能研究でない」と過剰除外、

非モデル生物のDB未登録による検証困難、プロンプト指示の曖昧さ

【LIMITATION】

著者単独キュレーション (セカンドオピニオン未実施)

Recall/Accuracy未評価。3日間サンプル (n=531) による代表性の限界

「非モデル生物遺伝子研究」の定義の曖昧さ (遺伝子の意味合いも含む酵素言及等)

2. 定義とデータセット

【非モデル生物の定義】

ゲノム編集研究のデータベースであるGenome Editing Meta-database<sup>3</sup>に登録されている上位20の生物種を  
研究量が多い生物種と仮定し、それ以外の生物種。

【遺伝子論文の定義】

- 以下6種類のいずれかに該当：  
遺伝子同定解析、オルソログ/ホモログ比較研究、発現解析、系統/進化的解析、機能検証/応用研究、手法/診断技術の開発  
※ 臨床応用研究と基礎研究の境界は曖昧であり、本研究では著者の判断により個別に分類した。
- 非モデル生物のみを対象 (タイトル及び要旨に、具体的なモデル生物の遺伝子名の言及がない論文)

【データセット】

- PubMed (2024/12/01-03) 取得論文: 4,960件  
- 検索クエリ: all[filter] AND pubmed pmc open access[filter] AND journal article[pt]
- PubTatorによる事前フィルタリング後: 531件  
- タイトルと要旨で最も頻出している生物種が非モデル生物

【AI agentに追加したカスタムツール】

- NCBI geneへの問い合わせ機能
  - Entity Linking機能の向上
- NCBI taxonomyへの問い合わせ機能
  - Entity Linking機能の向上
- 結果を構造化データへ整理する機能
  - 安定したデータ生成機能

3. AIエージェント (Biomni)

【Biomniによるプラン生成の例】

```
I'll proceed stepwise.

Updated plan:

1. [✓] Extract species and gene entities from input text and PubTator hints.
2. [ ] Confirm taxonomy ID and class for species "tobacco" (common name).
3. [ ] Validate gene "P16" in tobacco using query_ncbi_gene.
4. [ ] Classify research types.
5. [ ] Construct species_gene_list.
6. [ ] Save annotation with save_confirmed_paper_annotation.
```

【AIエージェントとは】

与えられた目標に対して自律的に行動・判断するAIシステム。

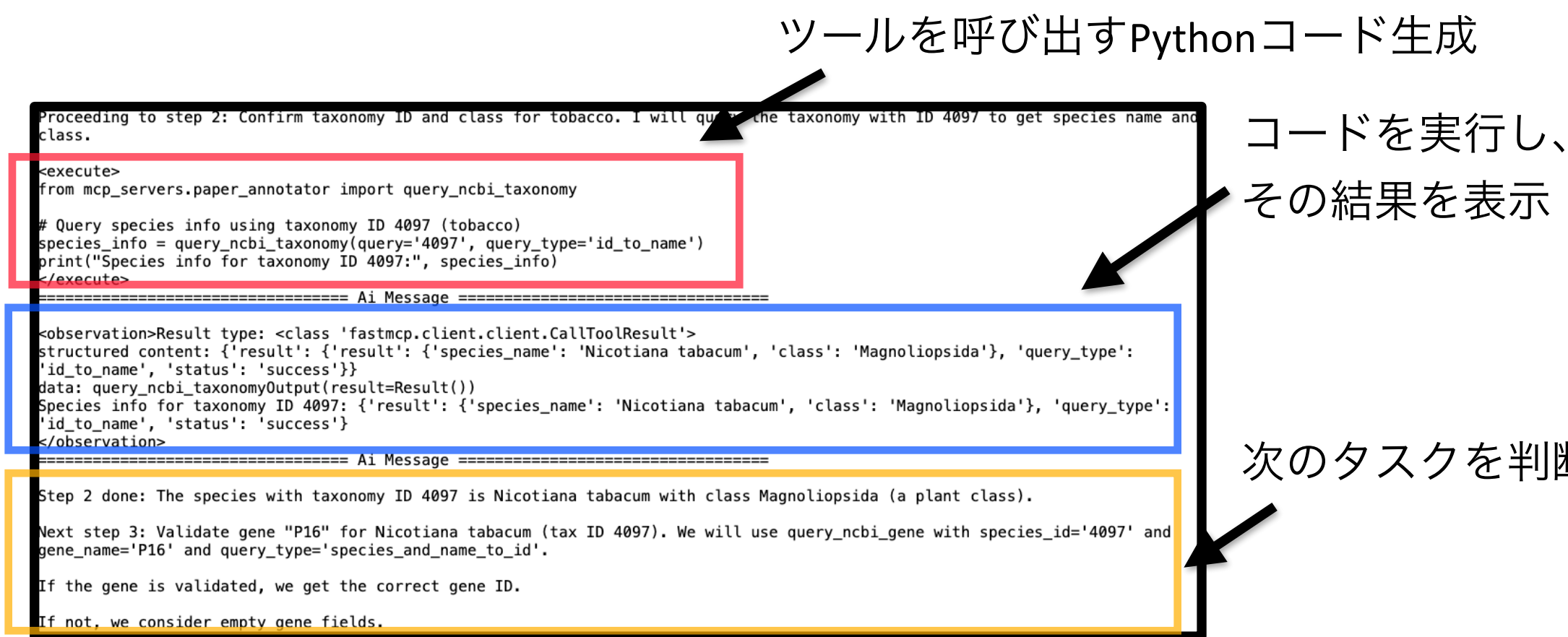
- OpenAI: Agents are systems that independently accomplish tasks on your behalf.
- Google: AI agent can be defined as an application that attempts to achieve a goal by observing the world and acting upon it using the tools that it has at its disposal.

その自律性により、タスク実行の際にどの情報を使うか、どの順番で進めるかを状況に応じて柔軟に判断するシステムであることが特徴。

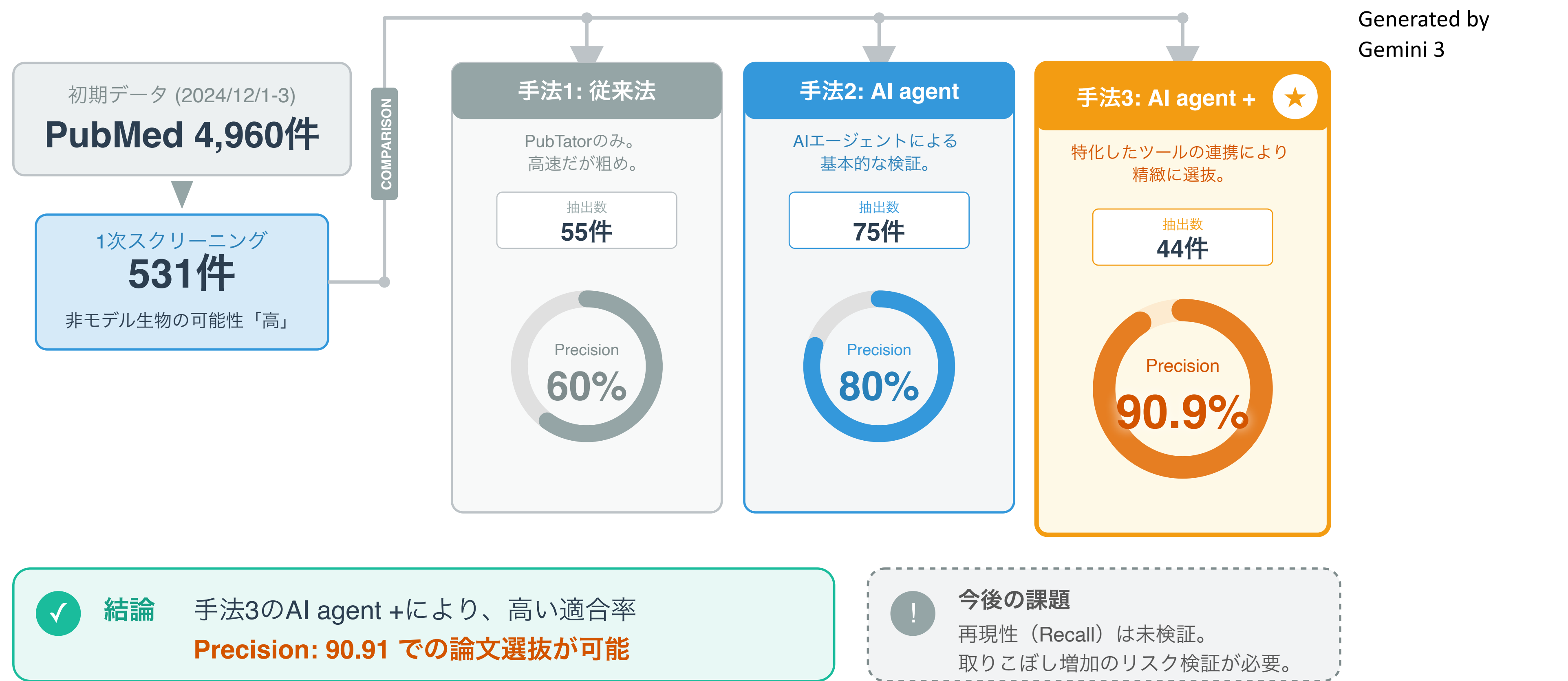
Googleによる定義にあるように、AIエージェントは手持ちツールの利用により外世界の観察を通じてタスク達成を試みる。

【Biomniの機能】

- 300以上のツールが実装
- CodeActによりAIエージェントが実行した処理が明確
- 汎用性・拡張性が高い



「非モデル生物の遺伝子論文選抜」のための手法比較



※ 開発した改良版Biomniと評価に使ったデータは、12月中に公開予定です

6. 結論と展望

【結論】

遺伝子・生物種DBへのアクセス機能などの新規ツール連携をしたAI agent + (手法3)は、Precision 90.91%を達成し、既存手法およびデフォルトAI agentを大幅に上回った。  
自律的なツール実行と結果考察により、複雑な文脈判断が可能になった。

【今後の展望】

- 評価指標の完全化：
  - Recall/F1-scoreの計測による評価。全531論文の完全キュレーションによるAccuracy検証
- データセットの拡充：
  - より長期間のランダムサンプリングや分野・時期のバイアス検証
- ベンチマーク化：
  - 別の専門家レビューによる品質保証、データセット公開とコミュニティ評価、AI agent文献キュレーション能力の標準評価タスクとして確立