

農業分野の栽培暦 VQA データセットの構築 —LLM のマルチモーダル質問自動生成能力の評価—

宮脇 一輝¹ 會田 勇斗² 高橋 洗丞² 中川 堯¹ 木村 泰知¹ 門脇 一真³
小林 暁雄⁴ 大友 将宏⁴ 石原 潤一⁴ 馬場 研太⁴ 桂樹 哲雄⁴
¹ 小樽商科大学² ストックマーク株式会社
³ 株式会社日本総合研究所⁴ 農研機構 農業情報研究センター
kimura@res.otaru-uc.ac.jp

概要

本研究では、大規模言語モデル (LLM) を用いて、農業分野の文書「栽培暦」を対象とした Visual Question Answering (VQA) データセットを構築する。JA、農業試験場、農研機構などが発行し文章や図表を含む文書である栽培暦を LLM に入力し、農業知識に関する質問・回答ペアを生成した。また生成結果を手で確認し、LLM のマルチモーダル質問自動生成能力について考察するとともに、不適切な質問や回答を修正した。最終的に、229 文書から 687 件の質問・回答ペアとその矩形情報を含む栽培暦 QA データセットを構築し、農業分野における質問応答研究および知識活用の基盤となるデータ資源が整備された。

1 はじめに

近年の日本では、農業従事者数の大幅な減少と急速な高齢化が進行しており、農業現場における深刻な労働力不足が顕在化している。農林水産省の統計によれば、2024 年の基幹的農業従事者数は、20 年前の 224 万人から 111.4 万人へとほぼ半減しており、そのうち 65 歳以上が約 7 割を占めている¹⁾。さらに、新規就農者数も 43,460 人と、統計開始以降初めて 4.5 万人を下回り、減少傾向が継続している²⁾。

このような状況の下で、営農指導を担う普及指導員は、農業者に対して技術情報の提供や経営支援を行う重要な専門職としての役割を果たしてきた。しかし、担い手の減少に加え、農業経営や技術に対する支援ニーズの多様化・高度化が進む中で、従来の

人的支援のみに依存した対応には限界が生じつつある。そのため、農業現場における AI 技術の活用が、喫緊の課題として求められている。

こうした課題に対応するため、内閣府は「研究開発と Society 5.0 との橋渡しプログラム (BRIDGE)」を通じて、AI 技術を活用した研究開発および社会実装の推進を進めている³⁾。その一環として、農業分野に特化した大規模言語モデル (LLM) の開発や、それを支えるデータセットの整備が進められており [1, 2, 3, 4, 5]、農研機構等による栽培暦 PDF の体系的な収集・公開は、農業知識の集約という点で極めて重要なデータ基盤を提供している。

一方で、これらの栽培暦は PDF 形式で提供されることが多く、作業時期や栽培方法、防除体系などの情報が、文章だけでなく表や図を含む視覚的構造として記載されているため、必要な情報を即座に検索・活用することは必ずしも容易ではない [6]。とりわけ、農業者や普及指導員が実際の営農場面において必要とするのは、「いつ、どの作業を行うべきか」、「特定の条件下でどのような対応が推奨されているか」といった、表や図の内容を正確に読み取ることが前提とした具体的な問いに対する回答である。このような実利用を想定すると、単に文書を蓄積するだけでは、LLM を用いた農業知識の理解や推論に関する研究開発を十分に推進することは難しい。農業分野における質問応答、情報検索、知識推論といった研究課題を体系的に検討するためには、モデルの性能を比較・評価可能な形式で整理された研究用データセットが求められる。

そこで、本研究では、大規模言語モデル (LLM) を活用し、農業分野の栽培暦を対象とした Visual Question Answering (VQA) データセットを構築す

1) https://www.maff.go.jp/j/wpaper/w_maff/r6/r6_h/trend/part1/other/P127_r6_d2_2-3-02.xls

2) <http://www.maff.go.jp/j/tokei/kouhyou/sinki/index.html>

3) https://www8.cao.go.jp/cstp/bridge/bridge_2512.pdf

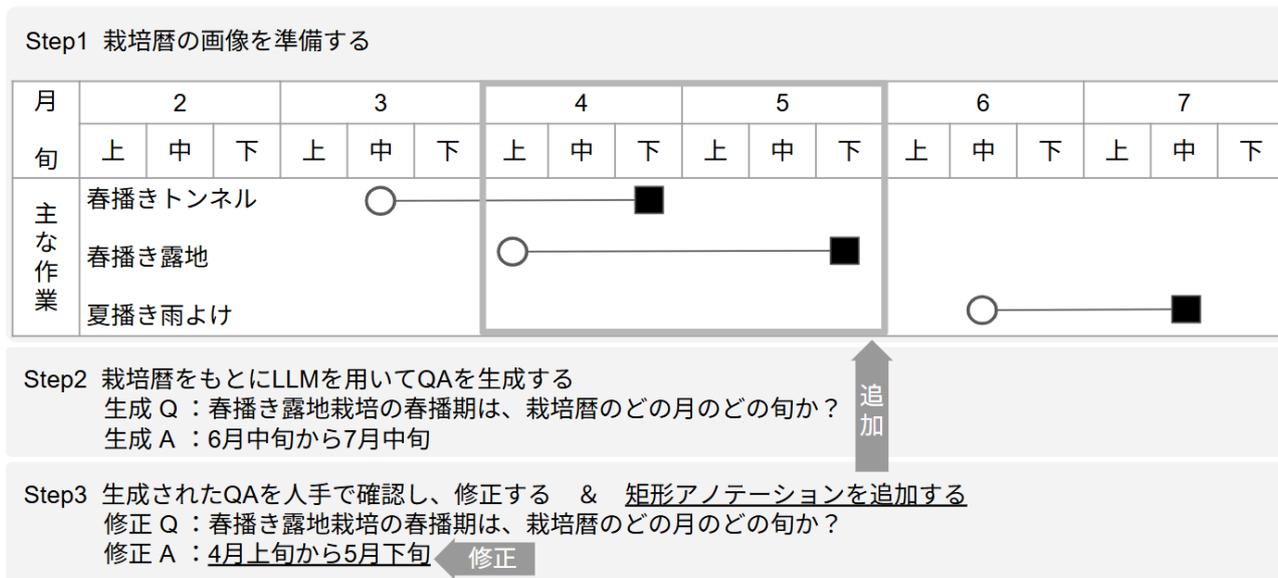


図1 LLMのマルチモーダル質問自動生成による栽培暦VQAデータセット構築の流れ

る。具体的には、栽培暦に記載された作業時期、品種選定、施肥、防除、病害虫対策などの農業知識を対象とした質問・回答ペアを、LLMであるQwen/Qwen3-VL-30B-A3B-Instructにより生成するとともに、その生成結果に対して人手による確認および修正をし、さらに矩形をアノテーションすることで、信頼性の高いQAデータセットを作成する。図1に、本研究における栽培暦QAデータセット構築の全体像を示す。

構築の結果、使用したLLMが生成した質問・回答ペアのうち約7割は正しく農業知識を問うものであった。また、最終的に、229件の栽培暦から687件の正しい質問・回答ペアを含む栽培暦QAデータセットが構築された。このデータセットは、今後NTCIR-19のタスクにおいて活用する予定である。

2 関連研究

質問応答 (Question Answering, QA) 分野では、SQuAD [7, 8] をはじめとした大規模かつ高品質なデータセットの整備が研究の進展を促してきた。近年では、複数文書や複数文を横断した推論能力を評価するマルチホップQAが提案され、HotpotQA [9] などのデータセットが提供されている。

テキスト情報に加えて画像や図表を対象とするVisual Question Answering (VQA) においては、VQA v2 [10, 11] などのデータセットが提供されている。他のマルチモーダルQAとして、業務手順や分岐条件を視覚的に表現したフローチャート画像に着目し

たJSynFlow [12] データセットなどがあげられる。

農業分野においては、AI農業社会実装プロジェクト⁴⁾の中で各都道府県が公表する農業のオープンデータが取りまとめられてきた [2, 3, 4, 5, 6]。農業分野に特化したQAとして、板倉ら [13] は農業経営類型を用いたデータセットを半自動で構築した。

3 栽培暦とは

栽培暦とは、特定の作物について、播種、定植、施肥、防除、収穫などの主要な栽培工程を、地域の気候条件や作期に基づいて時系列的に整理した農業技術資料である (栽培暦のサンプルは付録A.1に示す)。栽培暦は、作物の安定生産や品質確保を目的として、JA、農業試験場、農研機構、地方自治体などの公的・準公的機関を中心に作成されるもので、地域特性を反映した実践的な栽培指針として農業現場で広く利用されてきた。そこには、月別または生育段階別の作業内容・時期に加え、品種選定の目安、施肥量、防除方法、病害虫への対応、気象条件に関する留意点など、経験知と科学的知見が体系的に集約されている。資料の形式としては、年間カレンダー、作物や水管理に関する図、複数の表、病害虫や作物・農業機械等の写真、テキストや箇条書き、注記などを組み合わせた構造を持つものも多く、視覚的な理解が容易となるように設計されている。一方、レイアウト依存性が高いため、機械可読性は必ずしも高いとはいえない。

4) https://www8.cao.go.jp/cstp/bridge/keikaku/r5-20_bridge_r7.pdf

栽培暦は、農業分野における時系列性、条件分岐、地域依存性といった複雑な情報構造を内包する知識集約型文書であり、その自動処理の発展は、自然言語処理の研究者のみならず、農業分野の研究者にとっても有用と考えられる。

3.1 栽培暦の公開状況と収集状況

本研究では、わが国の農業関連機関が公開する栽培暦資料を体系的に収集した。収集対象となった発行元は、JA、農研機構、地方自治体、農業関連団体など計 111 機関である。

収集した資料はいずれも PDF 形式で、その総数は 1,094 ファイルである。また、収集した PDF のページ数分布を分析した結果、1 ページ構成の資料が 800 件と全体の大半を占めていた。一方で、10 ページ以上の比較的長い資料も一定数存在し、最大では 265 ページに及ぶ栽培暦が 1 件確認された。

4 データセット構築

4.1 マルチモーダル質問応答の自動生成

本研究では、質問応答 (QA) データを自動的に生成するための前処理として、文書画像からの情報抽出および記述生成を行った。栽培暦は表や図、注記を含む複雑なレイアウトを有する文書であるため、テキスト情報のみならず視覚的構造を考慮した処理が必要となる。まず、収集したすべての栽培暦 PDF ファイルを PNG 形式の画像に変換し、画像ベースでの解析が可能な処理環境を整備した。この変換により、ページ全体のレイアウトや図表配置を保持したまま入力として扱うことが可能となり、後続の視覚言語モデルによる処理に適した形式とした。さらに、変換後の各画像に対して、視覚言語モデル (Vision-Language Model: VLM) を用いて内容記述 (description) を生成した。図 2 に画像から質問応答に必要な情報を抽出する例を示す。具体的には、Qwen/Qwen3-VL-30B-A3B-Instruct⁵⁾ (以下、Qwen3-VL) を用い、画像中に含まれるテキスト情報 (作業時期、栽培工程、注意事項など) を自然言語で記述させた⁶⁾。

5) <https://huggingface.co/Qwen/Qwen3-VL-30B-A3B-Instruct>

6) Qwen3-VL に加えて、DeepSeek-OCR や PaddleOCR-VL も用いて画像中のテキスト情報抽出を試みたが、DeepSeek-OCR や PaddleOCR-VL は日本語の文字化けの発生頻度が高く、栽培暦において 1 枚の画像から出力される平均的な文字数が Qwen3-VL と比較すると 0.8 倍程度少なくなるため、

キンギョソウ
栽培上の留意点
1) 上手な作型、品種の導入で3番花迄収穫可。
2) 高温期は品種の導入に十分注意する。

月	8	9	10	11	12	1
年内出荷	∴	⊙			■	
1月出荷			∴	⊙		■

(∴: 播種 ⊙: 定植 ■: 採花)

↓
VLMを用いた
内容記述(description)の生成

キンギョソウ
栽培上の留意点
1) 上手な作型、品種の導入で3番花迄収穫可。
2) 高温期は品種の導入に十分注意する。
(以下、作型カレンダー図。「年内出荷、1月出荷」の2段。月別の目盛りは8~1。凡例は行末に記載)
(∴: 播種 ⊙: 定植 採花)

年内出荷
播種「∴」...およそ8月ごろに配置
⊙定植.....9月ごろ
採花 (黒い区間~点) ...11月に配置 (年内に収穫)

1月出荷
播種「∴」...8~9月
⊙定植.....10月
採花.....1月にマーク

図 2 画像から質問応答に必要な情報を抽出する例

次に、内容記述で使用したモデルと同一の Qwen3-VL で、各ページから 3 ペアずつ QA 生成を行った。QA 生成の際に Qwen3-VL には画像と生成した内容記述と QA 生成の指示文を与え、QA の回答が簡潔になるように指示文で調整した (付録 A.2 参照)。回答を短く調整したのは、将来的に本研究で作成したデータをもとに栽培暦に関する QA ベンチマークを設計し llm-as-a-judge を評価に使用した場合に、冗長性バイアスを排除するためである。

4.2 アノテーションの方法

4.1 節で自動生成した結果に対し、著者 3 名がアノテーションツールである CVAT (Computer Vision Annotation Tool) を用い、下記の作成方針 3 点に基づいて内容の確認および修正、ならびに矩形アノテーションの追加を行った。

1. 回答は簡潔な表現にする。
修正前「秋冬どり 30~50 日前」
修正後「30~50 日前」
2. 質問 (Q) において指定されている回答形式に基づいて、回答 (A) の表現を対応させる。
(例: Q 「どんな状態？」 → A 「○○な状態」)
3. 画像から読み取られる情報や表現で QA ペアを

Qwen3-VL を使用して作成したデータを最終版とした。

構成する。

この確認では、各 QA ペアが参照する情報が記載された画像内の該当箇所に対して、対象の栽培暦と照合しながら、質問文および回答文の事実関係の妥当性、表現の適切性、ならびに QA としての明確性を確認した。

さらに、QA と視覚的根拠との対応関係を明示するために矩形アノテーションを追加した。これは、各質問回答がどの範囲の情報に基づいて生成されたものであるかを明確にするためのものである。

4.3 アノテーション件数

前節までに述べた手順を実施した結果、JA、農業試験場、農研機構、地方自治体などの公的・準公的機関が発行する栽培暦 229 件を対象として、合計 687 件の質問・回答ペアからなる栽培暦 QA データセットを構築した。

5 LLM の質問自動生成結果と考察

表 1 生成された質問回答に対する人手による修正の有無

質問の修正	回答の修正	件数	割合 (%)
なし	なし	476	69.3
あり	なし	49	7.1
なし	あり	88	12.8
あり	あり	74	10.8
合計		687	100.0

質問回答の自動生成において、人手による修正の有無別に分類した件数を表 1 に示す。「正しい（修正なし）」に分類された事例は、図表情報から自然言語への変換が適切に行われている場合に多く見られ、その結果として、修正を必要としない質問回答が生成される傾向が確認された。一方、誤り分析の結果、図中の数値自体は正しく読み取られているものの、その数値を問う質問文の表現が不適切であるために修正が必要となるケースが確認された。

自動生成の誤り例として、質問のみを修正した例を図 3 に示す。生成された畝（うね）づくりに関する質問は「畝の幅は何 cm ですか?」というものであり、回答の「30cm」に対応する値ではないため、「条間は何 cm ですか?」と修正した。このように、図に関する文章の記載がない画像から情報を抽出する場合は、図中の数値が持つ意味の誤理解が発生しやすいと考えられる。

また、質問と回答の両方を修正した例を図 4 に示

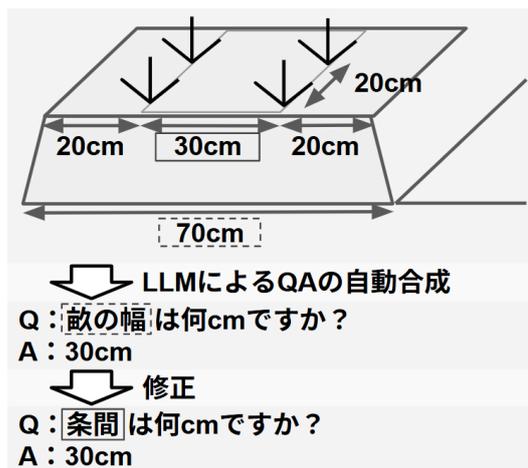


図 3 自動生成の誤りの例（質問の修正あり）

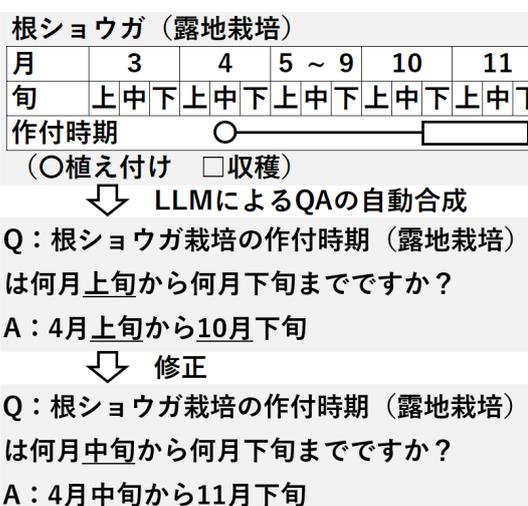


図 4 自動生成の誤りの例（質問・回答の修正あり）

す。生成された根ショウガ栽培に関する質問は「何月上旬から何月下旬まで」と期間を問うものであるが、正解は「4月中旬から11月下旬」であるため、質問も「中旬」を含むよう修正し、また生成された回答は期間を示す数値が誤っていたため、その修正を行った。

6 おわりに

本研究では、大規模言語モデル（LLM）と人手確認を組み合わせることで、公的・準公的機関が発行する栽培暦を対象とした信頼性の高い質問回答（QA）データセットを構築した。構築したデータセットは、作業時期や施肥、防除、病害虫対策など、農業実務に即した知識を網羅している。今後、NTCIR-19 DAGRI Subtask 3 の質問回答データセットとして活用する予定である。

謝辞

本研究は、JST, RISTEX, JPMJRS25L2 の支援、および、内閣府研究開発と Society5.0 との橋渡しプログラム (BRIDGE) 「AI 農業社会実装プロジェクト」 JP23836805 の補助を受けて行った。

参考文献

- [1] 石原潤一, 小林暁雄, 桂樹哲雄, 大友将宏, 橋本祥, 阪本浩太郎, 杉村安都武, 米丸淳一, 安藤まや, 後藤美智子, 関根聡, 川村隆浩. 国産農業用 LLM のためのインスタクションデータ構築と構築された LLM システムの評価. 言語処理学会第 31 回年次大会, 2025.
- [2] 杉山陽菜乃, 阿部瑞稀, 中村彩乃, 前多陸玖, 坂口遥哉, 佐藤栄作, 木村泰知, 小林暁雄, 大友将宏, 石原潤一, 桂樹哲雄, 川村隆浩. 農林業基準技術に含まれる表を対象とした PDF から CSV へ変換する際の課題分析. 言語処理学会第 31 回年次大会論文集, pp. 3954–3959. 言語処理学会, 2025.
- [3] 阿部瑞稀, 杉山陽菜乃, 中村彩乃, 前多陸玖, 坂口遥哉, 佐藤栄作, 木村泰知. PDF 形式の農業技術文書を用いた表構造認識ベンチマーク TOITA. 言語処理学会第 31 回年次大会, 2025.
- [4] 中村彩乃, 杉山陽菜乃, 阿部瑞稀, 前多陸玖, 坂口遥哉, 佐藤栄作, 木村泰知. 農林業基準技術文書を対象とした PDF 解析ツールの表構造認識の性能評価. 言語処理学会第 31 回年次大会, 2025.
- [5] 大友将宏, 石原潤一, 橋本祥, 桂樹哲雄, 二宮芳継, 小林暁雄, 坂地泰紀, 川村隆浩. データペーパー: 各都道府県が提供する農業関係オープンデータ. 言語処理学会第 31 回年次大会 発表論文集, 2025.
- [6] 杉山陽菜乃, 木村泰知. 農林業関連の文書に含まれる図表を対象とした質問応答に向けた分析. 第 41 回ファジィシステムシンポジウム, 9 2025.
- [7] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. SQuAD: 100,000+ questions for machine comprehension of text. In **Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 2383–2392, 2016.
- [8] Pranav Rajpurkar, Robin Jia, and Percy Liang. Know what you don't know: Unanswerable questions for SQuAD. In **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**, pp. 784–789, 2018.
- [9] Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W. Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. HotpotQA: A dataset for diverse, explainable multi-hop question answering. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 2369–2380, 2018.
- [10] Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C. Lawrence Zitnick, and Devi Parikh. VQA: Visual question answering. In **Proceedings of the IEEE International Conference on Computer Vision (ICCV)**, pp. 2425–2433, 2015.
- [11] Yash Goyal, Tejas Khot, Douglas Summers-Stay, Dhruv Batra, and Devi Parikh. Making the V in VQA matter: Elevating the role of image understanding in visual question answering. In **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, pp. 6904–6913, 2017.
- [12] 佐々木浩. JSynFlow: LLM を活用した日本語フローチャート VQA データセット合成. In **The 39th Annual Conference of the Japanese Society for Artificial Intelligence**, 2025.
- [13] 板倉亮真, 坂地泰紀, 野田五十樹, 小林暁雄, 大友将宏, 石原潤一, 桂樹哲雄. 生成 AI のための農業データセット構築とモデル評価. 言語処理学会第 31 回年次大会 発表論文集, 2025.

A 付録

A.1 栽培暦のサンプル



図 5 栽培暦のサンプル: 農研機構から刊行されている「にじのきらめき」栽培暦 改訂版⁷⁾

A.2 使用したプロンプト

画像内容の記述文を作成するプロンプト

<image>

この画像に含まれるテキストや図表をできるだけ忠実に読み取り、改行や段落も適宜保ちながら markdown 形式で詳細に書き起こしてください。

QA 生成をするプロンプト

<image>

以下は、この画像から取得した OCR テキストです:

{記述文}

この画像と上記テキストの内容に基づいて、日本語で 3 組の Q&A (質問と回答) を合成してください。

制約条件:

1. 質問 (question) は複雑でも構いませんが、必ずこの画像とテキストの内容からだけ答えられるものにしてください。

外部の知識や推測を必要とする質問は禁止です。

2. 回答 (answer) は必ず「短いフレーズ・単語・数値」などの簡潔な表現にしてください。

・文末に「です」「ます」「だ」「である」などを付けた文章形式は禁止です。

・例: "2024 年", "3 ページ", "安全手順", "温度設定", "10 件" など。

・1~10 語程度で、できるだけ短くしてください。

3. 出力は JSON 形式のオブジェクト 1 つだけにしてください。余計な文章や説明、コードブロックは絶対に書かないでください。

以下の形式にしてください:

```
{
  "qa_pairs": [
    {"id": 1, "question": "...", "answer": "..."},
    {"id": 2, "question": "...", "answer": "..."},
    {"id": 3, "question": "...", "answer": "..."}
  ]
}
```

7) https://www.naro.go.jp/publicity_report/publication/pamphlet/tech-pamph/134301.html