

NTCIR-19 DAGRI: 農業分野のマルチモーダル情報抽出・質問応答タスクの提案

森岡 幹¹ 木村 泰知² 坂地 泰紀³ 門脇 一真⁴ 會田 勇斗⁵ 高橋 洸丞⁵
小林 暁雄⁶ 大友 将宏⁶ 石原 潤一⁶ 馬場 研太⁶ 桂樹 哲雄⁶

¹AIREV 株式会社 ²小樽商科大学 ³北海道大学 ⁴株式会社日本総合研究所

⁵ストックマーク株式会社 ⁶農研機構 農業情報研究センター

tmorioka@airev.co.jp kimura@res.otaru-uc.ac.jp sakaji@ist.hokudai.ac.jp
kadowaki.kazuma@jri.co.jp {hayato.aida, kosuke.takahashi}@stockmark.co.jp
{akio.kobayashi, ohtomo.masahiro841}@naro.go.jp
{ishihara.junichi964, baba.kenta285, katsuragi.tetsuo}@naro.go.jp

概要

農業分野の AI 技術活用の妨げとなっている要因の 1 つである機械が解釈しにくい複雑な表組や図表を多用する文書への解析技術の確立を目指し、農業経営指標および作物生産にあたる年間農作業に関するデータを対象に情報抽出・構造化及び質問応答を行うタスクを提案する。提案するタスクは、NTCIR-19 のパイロットタスクである Data Analytics for aGRicultural Information (DAGRI)¹⁾にて、コンペティションとして開催する。DAGRI は 3 つのサブタスクから構成されており、各タスクの概要と内在している技術的な課題点について述べる。

1 はじめに

大規模言語モデル (LLM) の台頭で、情報抽出や質問応答等のタスクは大きな進歩を遂げ、様々なベンチマークや高い知識水準が求められる試験において人間に匹敵する性能を達成している [1, 2]。近年では、画像をはじめとした他のモダリティと組み合わせるマルチモーダル LLM の研究も盛んに行われており、LLM を適用できるシーンは拡大している。一方で実応用、特に専門分野の深い知識や推論が必要となるシーンでは、LLM を適用するだけでは期待する性能を達成しないことがある [3, 4]。

労働力の高齢化が進み知識継承が社会課題となっている農業分野でも AI 技術活用が期待が高まる一方で、一般知識を学習した LLM では農業分野の専

門的な知識が欠如して正確性の観点で課題が残り、分野特有の知識の転移や適応、専門知識に基づいたモデルの構築が求められている [3, 5, 6]。現在、公的研究機関や JA 等が作成・公開する技術文書は農業普及の貴重な資源となっている。しかしこれらの文書の多くは機械可読な形式ではなく、農業分野における AI 活用の大きな障壁となっている。特に注目すべきは、作物別に整理され PDF 形式で公開されている「標準農業技術文書」である。これらの文書は人間には易しく読めるが、機械が解釈しにくい複雑な表組や図表を多く含んでおり、AI 活用における有用性を制限している。この問題に対処するため、本研究では農業文書に含まれる図表などの複雑なデータを解析し、機械可読な形式へ変換することを目指す。構造化されたデータベースを構築し、このデータベースの上に農業知識の移転を促進する質問応答等のシステムが構築できれば、農業分野の専門知識のデジタル化を推進し、地域コミュニティに根差した知識継承モデルの構築に貢献する。最終的には地域農業における知識共有の新たな枠組を提案し、次世代農家の育成と農業分野の持続可能性向上に向けた基盤を築くことを目指す。

これらの実現に向け本研究では、農業文書のうち農業経営指標に関する標準農業技術文書および作物の生産に関する年間の農作業一覧 (栽培暦) の文書を対象に、所定の情報を探索・抽出して構造化するサブタスクと、機械可読性の低い複雑な図表を含む文書に関して質問応答を行うサブタスクを提案する。提案タスクは、NTCIR-19 のパイロットタスク

1) <https://sites.google.com/view/dagri/home-ja>

DAGRI としてコンペティションを実施する。本稿では、タスクの概要とスケジュールを説明する。

2 関連研究

標準農業技術文書には、特定の作物を栽培するためのスケジュールや使用する技術・資材、および生産に関する収支等の経営情報が記されている。農業技術文書の管理支援を行うシステム FAPS-DB が農研機構により運用されていたが、独自に技術文書を作成・公示する都道府県が少なくないこと、また作成される技術文書中のデータ項目が統一されていないことが課題として挙げられていた [7]。このような背景を受け、本研究ではデータ項目が統一化されておらず複雑にかつ多様に表現されている標準農業技術文書を対象に規格化し、質問応答というタスクを介して農業技術文書への認識・理解の程度を測ることを試みるものである。

農業分野に特化した言語モデル [5, 6] では、ウェブ上のデータから学習した LLM が農業分野の知識をどの程度有しているか、農業文書を対象に質問応答データセットを構築して検証し、農業分野への適応が必要であることを、適応のためのデータの整備が必要であることを示した。農業分野に特化した言語モデルとしては、対象言語は英語であるが AgriBERT [8]、AgroGPT [9] や AgriGPT [10] が挙げられる。特に [10] は分野特化の LLM のみならず、Retrieval-Augmented Generation (RAG) やエージェント等の LLM の周辺技術を含めたエコシステムとして提案されている。また農作物の成長段階の認識や病害診断のための、農作物画像とテキストを同時に扱えるマルチモーダル LLM に関する研究も行われている [11]。

複雑な図表の認識・情報抽出 今回対象にする標準農業技術文書にはテキストだけでなく、セル結合が多用されている表組や機械が解釈しにくい複雑な図表が含まれている。農業文書中の表組の抽出・認識および CSV 等のデータ形式への再構成に向けて、[12] では、長崎県の標準農業技術文書を対象にベンチマークデータセット TOITA を設計し、[13] では TOITA に対して表構造認識の様々な手法の性能を GriTS [14] に従って評価した。

文書中の表の検出や認識、表データに対する質問応答は農業分野に限ったものではなく、他の分野でも広く研究されている。表データの検出・認識の先駆けとして、EU 及び US における行政文書中の表

データを対象とした ICDAR13 Table Competition [15] が開催された。その後、TableBank [16] や PubTables-1M [17] 等、データセットの大規模化や表データへのアノテーションの緻密化が進み、あわせて、Detectron2 [18] や Transformer [19] 等への適用及び学習済みモデルの公開が行われている。

マルチモーダル質問応答 画像への理解を測るタスクの 1 つに Visual Question Answering (VQA) [20] が挙げられる。VQA の対象となる画像の多くは自然画像である一方で、公的文書や技術文書に出現する画像はグラフなどの図表が多く、標準農業技術文書や栽培暦の画像もグラフ、図表がほとんどである。図表画像を対象にしたベンチマークの 1 つに Figure QA [21] があるが、特定の分野に依らないように質問の種類を限定している。本研究における質問応答システムでは、図表の正しい読み取り・理解だけでなく農業分野の知識への深い理解を求め、実応用を見据えたタスクとして設計する。農業分野以外でも業界・分野の深い理解が求められるベンチマークの開発が積極的に行われている [22, 4]。

3 タスク

NTCIR-19 DAGRI で扱うデータは、農業経営指標に関する文書群、及び年間農作業に関する文書群（栽培暦）である。農業経営指標に関するデータは各都道府県が独自に作成しているが、自治体ごとに独自の項目が存在する、特定の内容が欠損する等、記載のフォーマットは統一されていない。また栽培暦は、テキスト、図表のみならず画像を活用しながら、作物の栽培に関する時期や注意点などを図説している。NTCIR-19 DAGRI では、農業分野の専門知識を理解しデジタル化を目指し統一的なフォーマットに変換・構造化するタスクと、農業分野の知識移転支援の枠組みの 1 つとして複雑な図表を含む文書への質問応答を行うタスクを提案する。タスクの全体像は図 1 のようになっており、3 つのサブタスク、すなわち、構造化に関するサブタスク、統一フォーマット上での質問応答タスク、栽培暦上での質問応答タスクから構成される。各サブタスクの概要、入出力、評価に関する情報について述べる。

3.1 Subtask 1: TableIE

標準農業技術文書には、栽培に必要な労働力や土地、機械設備、資材などの資本、栽培のスケ

■ 標準農業技術文書

作物名: 大豆 (小粒 ユキシズ)				10a 当たり		
作業名	作業時期	使用資材		作業技術		作業人員 家族 雇用
		資材名	使用量	作業機械名	時間 (hr)	
堆肥散布	4/中	豚ふん	0.5 t	マニュアルスプレッダー フロントローダー	0.06	1
心土破碎	4/中~5/上			サブソイラー	0.29	1
耕起	5/上			ローリーハロー	0.16	1
碎土・整地	4/下~5/中			ローリーハロー	0.50	1
は種・鎮圧	5/中	ユキシズ	5.1 kg	ビープランター 鎮圧ローラー	0.30	1
追肥	6/上~6/下	発酵鶏ふん	71.0 kg	ブロードキャスター	0.16	1
中耕・除草	5/中~7/下			草刈り(加チヘーター)	1.15	1

作物名: 春まき小麦 (はるきり)				10a 当たり		
作業名	作業時期	使用資材		作業技術		作業人員 家族 雇用
		資材名	使用量	作業機械名	時間 (hr)	
堆肥散布	前年 9/上	海鮮コーラップ 発酵鶏ふん	650 kg 67.5 kg	マニュアルスプレッダー フロントローダー ブロードキャスター	0.22 0.11	1 1
堆肥散布	前年 3/下	豚ふん	1.27 t	マニュアルスプレッダー フロントローダー	0.36	1
堆肥散布	4/中	豚ふん	0.73 t	マニュアルスプレッダー フロントローダー	0.24	1
心土破碎	4/中			サブソイラー	0.05	1
耕起	4/上 4/中			チゼル チゼル	0.08 0.09	1 1

■ 栽培ごよみ

月 旬	3月			4月			5月			6月			7月					
	中	下	上	中	下	上	中	下	上	中	下	上	中	下	上			
ステージ	出芽・緑化・硬化						生育期			有効分げつ決定期 幼穂形成期(7/13~14頃)								
水管理	除草剤散布後は4日以上止め水												中干し			低温時深水管理		
作業内容	種子準備 種子消毒 床土準備			催芽・播種 加温出芽・育苗 転作準備 耕起			基肥施用 代掻き・田植え 育苗補助 除草剤散布			取置き苗除去 穂切り 雑草刈			灌水管理 雑草刈 追肥					

■ Subtask 1: Table IE

品目: 大豆 / テーブル名: 作業技術一覧

作業技術名	使用機材名一覧	機材使用時間	...
堆肥散布	・ マニュアルスプレッダー ・ フロントローダー	0.06	...
心土破碎	・ サブソイラー	0.29	...
...

品目: 春まき小麦 / テーブル名: 作業技術一覧

作業技術名	使用機材名一覧	機材使用時間	...
堆肥散布	・ マニュアルスプレッダー ・ フロントローダー	0.22	...
堆肥散布	・ ブロードキャスター	0.11	...
...

検索
/ 情報抽出
/ 構造化

複数テーブルを
参照する
質問応答

■ Subtask 2: Multi-Table Multi-Hop QA

大豆と春まき小麦で共通する作業機械名を列挙せよ

マニュアルスプレッダー, フロントローダ...

■ Subtask 3: FiT-QA

除草剤の散布は何月ごろから行いますか

5月中旬

マルチモーダル
質問応答

図1 DAGRI タスクの全体像. 3つのサブタスクについて対象となるデータ及びタスクの概要・出力例を示している.

ジュールを作物ごとに策定した経営指標と、経営指標に基づき栽培する作物とその規模、栽培にあたり生じる収支目標などを定めた経営類型が記載されている。Subtask 1では、各都道府県の農業経営指標に関するファイル群から、経営類型および経営指標に関する情報を抽出・解析し、統一フォーマットへと変換・構造化を行う。Subtask 1の入力、出力は以下の通りである。

入力	農業技術文書:	各自治体が公開する PDF/Excel 文書
	タスク指示書:	本研究で都道府県ごとに定めたガイドライン
出力	統一フォーマット:	経営類型, 経営指標 (作物・品種別)

項目は [23, 24] で定めた統一フォーマットに準じ、経営類型では「前提表」, 「栽培規模」, 「経営収支」, 「資本装備と減価償却費」, 作物ごとの経営指標では「経営指標」, 「作業技術」, 「作業時間」を JSON 等の

構造化ファイルとして出力する。入力のファイル群から経営類型、経営指標の各項目を漏れなく正確に探索・抽出できるかを評価する。評価指標は適合率・再現率・F1 スコアを予定している。

統一フォーマットに関する仕様及び詳細の説明は [23] を、Subtask 1の詳細は [24] を参照されたい。

3.2 Subtask 2: Multi-Table Multi-Hop QA

Subtask 2は普及指導員の業務の一つである、農業者が育てている作物やその栽培規模、経営費や労働時間といった農業を営む状況の把握及び営農計画の策定支援に着目し、普及指導員を支援するための「Multi-Table Multi-Hop QAに基づく営農シミュレーションタスク」である。

農業経営に関する知識や指針は、「農業経営類型」と呼ばれる文書において、セル結合や注記なども含む複雑な表で提供されることが一般的である。本タスクでは、収支・作業体系・労働時間など異なる観

質問 春まき小麦と大豆の栽培において、5月の作業を4月に行った場合、4月に必要な人員数は何倍になるか答えよ。但し、一人あたりひと月に180時間稼働するとする。

抽出・計算
必要人員数:
 シナリオA: ((4月春小麦)+(4月大豆)) / 180
 シナリオB: ((4月+5月春小麦)+(4月+5月大豆)) / 180
比較:
 シナリオB / シナリオA

回答 3.8倍

作業時間表 (春まき小麦) 作業時間表 (大豆)

図2 Multi-Table Multi-Hop QA の例

点の複数の表を参照しつつ、農業収益費、経営に占める償却費率など経営に関する12の指標値を回答するとともに、その回答根拠となる表や項目・計算式を明確にすることを目的とする。

Subtask 2の入力,出力は以下の通りである(図2)。

入力	ファイル:	作業時間表(春まき小麦).json, 作業時間表(大豆).json
	クエリ:	春まき小麦と大豆の栽培において、5月の作業を4月に行った場合、4月に必要な人員数は何倍になるか答えよ。但し、一人あたりひと月に180時間稼働するとする。
出力	回答:	3.8倍

評価は、文字列一致の正解率により行う。Subtask 2に関する詳細は[25, 26]を参照されたい。

3.3 Subtask 3: FiT-QA

Subtask 3は、テキスト・図表・画像から構成される「栽培暦」のPDF文書を画像として入力し、提示された質問に対して適切な回答を生成するマルチモーダル質問応答タスクである。栽培暦とは、特定の作物について、播種、定植、施肥、防除、収穫といった主要な栽培工程を、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。栽培暦には作業日程などの時系列情報を含む図表が多く含まれるが、これらの情報を既存のマルチモーダルLLMが適切に解釈することは難しく、正確な出力を得ることが困難であることが指摘されている。そこで、Subtask 3では、栽培暦に含まれる時系列的な情報の理解に着目したマルチモーダル質問応答タスクを設計した。

Subtask 3の入力,出力は以下の通りである(図3)。

入力	質問:	1. 幼穂形成期は何月の何旬ですか？
	画像:	2. 48-145-003_page_1.png
出力	回答:	7月中旬

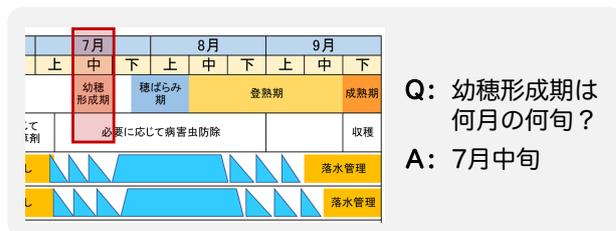


図3 FiT-QA の例

評価は、BLEU、および、LLM-as-a-Judgeにより行う予定である。また、NTCIR-19のタスク終了後には人手による評価も検討している。Subtask 3に関する詳細は[27, 28]を参照されたい。

4 スケジュール

本稿で提案した3つのサブタスクは、国立情報学研究所が開催するNTCIR-19の1タスクとして実施する。NTCIR-19 DAGRI タスクは以下のスケジュールで開催する予定である:

- 2026年1月 サンプルデータ公開
- 2026年2月 第1回参加者向け説明会
- 2026年3月 プレ・コンペティションデータ公開
- 2026年4月 第2回参加者向け説明会
- 2026年3-6月 プレ・コンペティション (Dry Run)
- 2026年7月 本コンペティション (Formal Run)
- 2026年8月 評価結果公開, タスク概要論文配布
- 2026年9月 NTCIR-19 タスク参加者論文提出
- 2026年11月 NTCIR-19 最終稿提出
- 2026年12月 NTCIR-19 カンファレンス

5 おわりに

本稿では、農業分野におけるAI活用の障壁となる課題について紹介し、NTCIR-19 DAGRIにおける3つのサブタスクについて述べた。各タスクの詳細について個別の報告[23, 24, 25, 26, 27, 28]も実施されるので、あわせて参照いただきたい。農業分野に特化はしているが、分野によらず普遍的な課題も含むタスクでもあり、興味を引くタスクがあった場合は是非とも参加いただきたい。

現在はプレコンペティションに向けてデータセットをはじめとしたリソースの準備を進めており、予備テストを実施した上で、2026年7月に本テストを実施する予定である。

謝辞

本研究は、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, および, JST, RISTEX, JPMJRS25L2 の支援を受けて行った。

参考文献

- [1] OpenAI. GPT-4 technical report, 2024.
- [2] Jungo Kasai, et al. Evaluating GPT-4 and ChatGPT on Japanese medical licensing examinations, 2023.
- [3] 馬場研太ほか. 農業経営指標策定に向けた大規模言語モデル活用の可能性—農業文書の表形式への自動構造化実験—. 農業情報研究, Vol. 34, No. 2, pp. 51–67, 2025.
- [4] Issa Sugiura, et al. EDINET-Bench: Evaluating LLMs on Complex Financial Tasks using Japanese Financial Statements, 2025.
- [5] 板倉亮真ほか. 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.
- [6] 板倉亮真ほか. 地方農業に特化した QA データセットの構築と検証. 人工知能学会全国大会論文集, 第 JSAI2025 巻, p. 1Win477, 2025.
- [7] 佐藤正衛. 農業技術体系データの作成・利活用の現状と課題. 農作業研究, Vol. 56, No. 3, pp. 197–203, 2021.
- [8] Saed Rezayi, et al. AgriBERT: Knowledge-infused agricultural language models for matching food and nutrition. In **Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22**, pp. 5150–5156. International Joint Conferences on Artificial Intelligence Organization, 7 2022. AI for Good.
- [9] Muhammad Awais, et al. AgroGPT: Efficient agricultural vision-language model with expert tuning. In **Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision**, 2025.
- [10] Bo Yang, et al. AgriGPT: a large language model ecosystem for agriculture, 2025.
- [11] Piaofang Yu, et al. A framework for agricultural intelligent analysis based on a visual language large model. **Applied Sciences**, Vol. 14, No. 18, 2024.
- [12] 阿部瑞稀ほか. 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 発表論文集, 2025.
- [13] 中村彩乃ほか. 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 発表論文集, 2025.
- [14] Brandon Smock, et al. GriTS: Grid table similarity metric for table structure recognition. In **International Conference on Document Analysis and Recognition**, pp. 535–549. Springer, 2023.
- [15] Max Göbel, et al. ICDAR 2013 table competition. In **2013 12th International Conference on Document Analysis and Recognition**, pp. 1449–1453, 2013.
- [16] Minghao Li, et al. TableBank: Table benchmark for image-based table detection and recognition. In **Proceedings of the Twelfth Language Resources and Evaluation Conference**, pp. 1918–1925, Marseille, France, May 2020. European Language Resources Association.
- [17] Brandon Smock, et al. PubTables-1M: Towards comprehensive table extraction from unstructured documents. In **CVPR 2022**, August 2021.
- [18] Yuxin Wu, et al. Detectron2. <https://github.com/facebookresearch/detectron2>, 2019.
- [19] Brandon Smock, et al. Table Transformer. <https://github.com/microsoft/table-transformer>, 06 2021.
- [20] Stanislaw Antol, et al. VQA: Visual question answering. In **2015 IEEE International Conference on Computer Vision (ICCV)**, pp. 2425–2433, 2015.
- [21] Samira Ebrahimi Kahou, et al. FigureQA: An annotated figure dataset for visual reasoning. **CoRR**, Vol. abs/1710.07300, , 2017.
- [22] Qianqian Xie, et al. FinBen: a holistic financial benchmark for large language models. In **Proceedings of the 38th International Conference on Neural Information Processing Systems, NIPS '24**, Red Hook, NY, USA, 2024. Curran Associates Inc.
- [23] 小林暁雄ほか. DAGRI Data format: 農業経営データ解析タスクのための統一フォーマット構築. 言語処理学会 第 32 回年次大会 発表論文集, 2026.
- [24] 森岡幹ほか. DAGRI Subtask 1: Table IE 複雑な図表を含む農業文書から情報抽出・構造化するタスクの提案. 言語処理学会 第 32 回年次大会 発表論文集, 2026.
- [25] 板倉亮真ほか. 農業経営データに対する Multi-table QA ベンチマークの構築. 言語処理学会 第 32 回年次大会 発表論文集, 2026.
- [26] 中嶋楓花ほか. DAGRI Subtask 2: Multi-Table Multi-Hop QA に基づく営農シミュレーションのためのタスク提案. 言語処理学会 第 32 回年次大会 発表論文集, 2026.
- [27] 宮脇一輝ほか. 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—. 言語処理学会 第 32 回年次大会 発表論文集, 2026.
- [28] 中川董ほか. DAGRI Subtask 3: 栽培暦を対象とした Figures and Tables Question Answering (FiT-QA) の提案. 言語処理学会 第 32 回年次大会 発表論文集, 2026.