

大規模マルチモーダルモデルによる画像・PDF形式文書データの構造化

BonBon株式会社 Medical AI R&D Engineer 林 樹永

目次

- 主題
- 前回の取組
- 課題と目的
- 方法
- 結果
- 考察
- まとめ

主題

電子カルテ内の非構造化データの構造化に取り組んでいる

- 電子カルテが医療文書の構造化を可能にした
- 一方で、電子カルテは非構造化データも連携される
 - PDFや画像の形式
 - 規定がない文書
- これらは自由記述形式でいわゆる非構造化データである
- 主に検査結果等経過記録で、構造化によって医療情報の活用が発展する

前回の取り組み

PDF/画像形式の医療文書にOCRを行い、文字列とその座標から構造解析を実施した

画像



文字領域検出



文字認識

```
{
  "description": "WAITING",
  "boundingPoly": {
    "vertices": [
      {
        "x": 344,
        "y": 821
      },
      {
        "x": 2025,
        "y": 879
      },
      {
        "x": 2016,
        "y": 1127
      },
      {
        "x": 335,
        "y": 1069
      }
    ]
  }
},
```

前回の結果: OCR結果からテキストマイニング

DF[結果]S[結果S]

検査目的:

血栓の有無

臨床診断:

心房細動

OCR結果と万病辞書による傷病名と病名及び「～を認め」や「陰性 /陽性」といった事実性の判定を実行

傷病名: 血栓

事実性:

- 左心耳を含め左房内に明らかな血栓様構造物は認めず
 - 左房内層毛毛地工左心耳を含め左房内に明らかな血栓様構造物は認めず
 - 心拍
- 数:70KokohanMincm/secPSPIREINVETER12V13.BEARHYTHIC
AINTE80bpm大動脈弁は三尖、開口良好で有意な逆流は認めず他の弁膜に器質的異常や生理的範囲を越える逆流は認めず
- 部~胸部下行大動脈に可動性のプラークは認めず

超音波所見

※検査中測誤
左心耳血流速度=26.8cm/s。左房内にもややエコーを認めない。
左心耳を含め左房内に明らかな血栓様構造物は認めず。
大動脈弁は三尖、開口良好で有意な逆流は認めず。他の弁膜に器質的異常や生理的範囲を越える逆流は認めず。
Valsalva manoeuvreにて心房内にshunt血流は認めず。
胸部~胸部下行大動脈に可動性のプラークは認めず。

超音波診断

左心耳血流速度=26.8cm/s。左房内にもややエコーを認めない。
左心耳を含め左房内に明らかな血栓様構造物は認めず。

前回の結果:OCR結果の構造解析

OCR結果から事実性の表現を抽出し、事実性の語彙の周辺座標から傷病名を取得した

塗 抹 鏡 検	1	グラム陽性球菌	3+	貧食あり	
	2	白血球	1+		
	3				
	4				
	5				
	6				
	7				
	8				
培 養 ・ 同 定	1	Aerococcus urinae	3+	感受性検査は後日報告します。	
	2	嫌気性菌	陰性		
	3				
	4				
	5				
	6				
	7				
	8				
	9				
	10				



傷病名	事実性
1グラム陽性球菌 貧食あり	3+
塗 2白血球	1+
1Aerococcusurinae 感受性検査は後日報告します。	3+
培2嫌気性菌	陰性

課題と目的

OCR結果からテキストマイニングと構造解析により、傷病名抽出や事実性判定が実現できたが、ルールベースの構造解析は汎用性を欠く

- 大量の辞書を用意する
- 様々な文書構造に合わせた座標解析を用意する
- 文書画像以外の形式に対応できない

画像・PDF形式の自由記述文書データの汎用的構造化をおこなう

方法

- データ受領

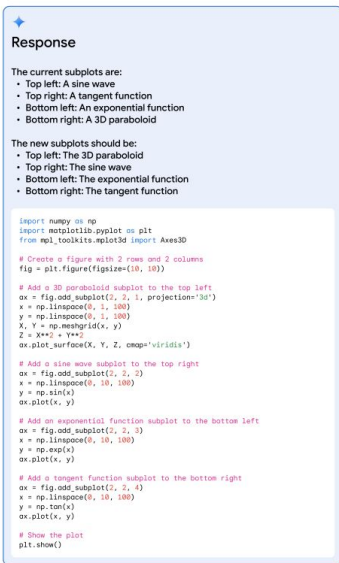
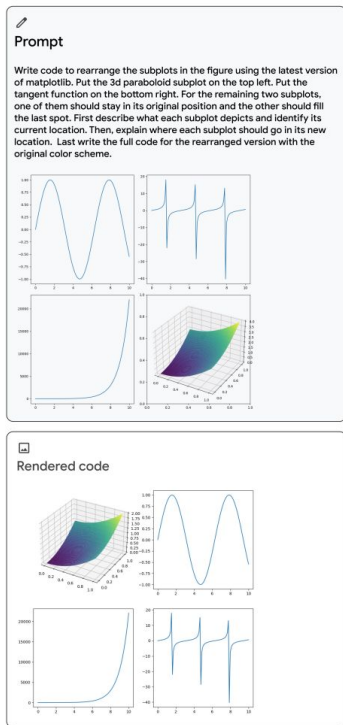
- 日本医療ネットワーク協会 (現 LDI)、宮崎市郡医師会病院 よりデータを受領
- 主要な個人情報マスキング済

- 構造解析

- Google Cloud PlatformのVertex AI サービスで管理されている大規模マルチモーダルモデルを用いて
入力に画像と指示文を与え構造解析を行った

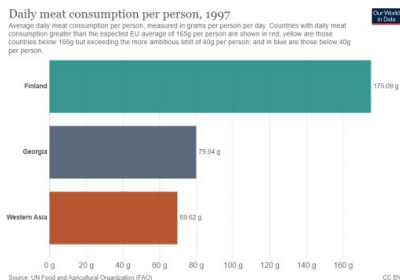
補足: 大規模マルチモーダルモデルによる画像推論

画像、自然言語処理等の様々な形式のデータセットを用いた機械学習タスクで訓練したモデル
訓練の結果、画像データから内容の抽出やグラフの説明をする等の推論能力を持つ



GPT-4 visual input example, Chart Reasoning:

User What is the sum of average daily meat consumption for Georgia and Western Asia? Provide a step-by-step reasoning before providing your answer.



GPT-4 To find the sum of average daily meat consumption for Georgia and Western Asia, we will follow these steps:

1. Identify the average daily meat consumption for Georgia.
2. Identify the average daily meat consumption for Western Asia.
3. Add the values from steps 1 and 2.

Step 1: From the provided chart, we see that the average daily meat consumption for Georgia is 79.84 grams per person per day.
Step 2: From the chart, we see that the average daily meat consumption for Western Asia is 69.62 grams per person per day.
Step 3: Add the values from steps 1 and 2.
79.84 g (Georgia) + 69.62 g (Western Asia) = 149.46 g

So, the sum of average daily meat consumption for Georgia and Western Asia is 149.46 grams per person per day.

Table 14. Example prompt demonstrating GPT-4's visual input capability. The prompt consists of a question which requires reading information from a chart and performing computation.

GPT-4 visual input example, Pixel to Paper Summaries:

User Below is part of the InstructGPT paper. Could you read and summarize it to me?



GPT-4 The InstructGPT paper focuses on training large language models to follow instructions with human feedback. The authors note that making language models larger doesn't inherently make them better at following a user's intent. Large models can generate outputs that are untruthful, toxic, or simply unhelpful.

To address this issue, the authors fine-tune language models on a wide range of tasks using human feedback. They start with a set of labeler-written prompts and responses, then collect a dataset of labeler demonstrations of the desired model behavior. They fine-tune GPT-3 using supervised learning and then use reinforcement learning from human feedback to further fine-tune the model. The resulting model, called InstructGPT, shows improvements in truthfulness and reductions in toxic output generation while having minimal performance regressions on public NLP datasets.

The authors conclude that fine-tuning with human feedback is a promising direction for aligning language models with human intent.

User Could you explain the process described in Figure 2?

GPT-4 Figure 2 in the InstructGPT paper illustrates the three steps of their method:

1. Supervised fine-tuning (SFT): They collect demonstration data, which consists of labeler-written prompts and desired output behavior. This data is used to fine-tune GPT-3 with supervised learning.
2. Reward model (RM) training: They collect comparison data, where a labeler ranks the outputs from several model samples (from best to worst). This data is used to train the reward model.
3. Reinforcement learning using Proximal Policy Optimization (PPO): They optimize a policy against the reward model using reinforcement learning. For each new prompt sampled from the dataset, the policy generates an output. The reward model calculates a reward for the output, and the reward is used to update the policy using the PPO algorithm.

These three steps are iteratively performed to train the InstructGPT model, which aims to generate outputs that align better with human preferences and follow instructions more effectively.

Table 17. Example prompt demonstrating GPT-4's visual input capability.

補足: 大規模マルチモーダルモデルで試みる機械学習技術

Document Question Answerという技術である
指示文と画像を入力に、画像に書かれている情報を抽出する

Inputs

Question

What is the idea behind the
consumer relations efficiency
team?

Document Question Answering Model

Output

Answer

Balance cost efficiency with
quality customer service

July 31, 2003

Consumer Relations Efficiency Team

Team:	Nancy Bosland	Shonda McMurry	Veronica Walker
	Gary Hicks	Nancy Montgomery	Yvette Willard
	Todd Holbrook	Brice O'Brien	
	Cindi Hunter	Doreen Wallop	

Objective: Balance cost efficiency with quality customer service

Our Process:

- Examined Consumer Relations cost structure:
 - West Interactive - Automation
 - Holliston - Live Rep
 - Y.A. - Live Rep
 - Chowitters
- AT&T
- Brand Chargebacks
- Veritas

Reasons for Consumer Contacts:	%
Request Catalog/Order Form	55%
Order Status	19%
Mailing List Request (add, change, update add, target change)	15%
Promotional Questions (lists, rebills, orders, questions, terms, Web issues, etc.)	2%
Product Issues	4%
Non-acceptance DM, conversion, fulfillment	1%
FLP Issues - card request, statement issues	1%
Other	1%

<https://huggingface.co/tasks/document-question-answering>

03/17/98 TUE 06:16 FAX 7328631723 AFG:RJR Q:003

SPECIAL POS / PDI REQUEST FORM

Requesting RJR Manager A.P. GROLL Date 03/19/98
Region Number 1200 Voice Mail Number 21565
Store / Chain Name LOVE STORES No. of Stores 20
Requesting (Check One) ☒ Produced POS / PDI ☐ Digital Art Mechanical Only
Is this an existing item? ☐ No ☒ Yes Or a new item? ☒ No ☐ Yes Due Date Required 03/15/98
Description of Request (Give as much detail as possible)
CAMEL SHOPPING BASKETS

Drawing of Request (Attach separate drawing if necessary and sample if available):

Exact Size: _____" (H) _____" (W)
Size excluding dead areas: _____" (H) _____" (W)
Identify Dead Areas (Hidden by frames, etc.) _____" (Top) _____" (Bottom) _____" (Side/End)
Quantity Requested 50 SKU Pack: 1
Ship To Location (If this request is to be warehoused by RJR, please write RJR in the name area)
Name RJR - NEW YORK METRO DCU
Address MANHATTAN CENTER - 400 MANHATTAN CENTER PARKWAY
City EDISON State NJ Zip Code 08837
Attention A.P. GROLL
Complete the below information only if art is being requested for local production:
Store / Chain Contact Name _____ Phone _____
Printer / Supplier Contact Name _____ Phone _____
After approval by Region Sales Manager, e-mail or fax form to your Area Manager of Operations.
Allow a minimum of 5 to 8 weeks for special requests.
RSM Approval [Signature] Date 3-17-98
AMO Approval _____ Date _____
TO BE COMPLETED BY WINSTON-SALEM
DATE REQUEST RECEIVED _____ REGIONAL DATE _____ HAS NO. ASSIGNED _____ BU/PAK _____
WORKSHEET _____ PU NUMBER ASSIGNED _____
SUPPLIER ASSIGNED _____ DUE DATE IN WAREHOUSE _____
PROJECT ESTIMATED COST _____ PROJECT ACTUAL COST _____ DL CODE ASSIGNED _____

Source: <https://www.industrydocuments.ucsf.edu/docs/rmj000>

Store / Chain Name LOVE STORES

Requesting (Check One) ☒ Produced POS / PDI

Is this an existing item? ☐ No ☒ Yes Or a new item?

Description of Request (Give as much detail as possible)

03/17/98 TUE 06:16 FAX 7328631723 A7

SPECIAL POS / P

Requesting RJR Manager A.P. GROLL

Region Number 1200

Q: Is it an existing item ?

Question types: *form* and *yes/no*

A: yes

Q: What is the date given at the top left?

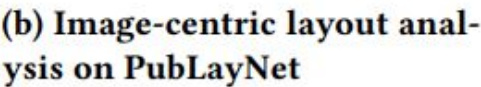
Question types: *layout*

A: 03/17/98

Layout Language Modelという技術にも取り組む
文書画像から項目名とその要素を抽出する

Layout Language Modelという技術にも取り組む
文書画像から項目名とその要素を抽出する

(a) Text-centric form understanding on FUNSD



補足: Google Cloud PlatformのVertex AI

Google Cloud PlatformのVertex AIは機械学習に特化した様々なサービスが使える
様々な学習済みモデルをAPI形式で利用可能である

Model Garden

[EXPLORE GENERATIVE AI](#)

[VIEW MY MODELS](#)

[DEPLOY FROM HUGGING FACE](#)

Modalities

Language48

Vision86

Tabular6

Document5

Speech1

Video4

Multimodal6

Tasks

Generation56

Classification51

Detection33

Extraction17

Recognition17

Translation12

Embedding3

Segmentation7

Retrieval1

Open vocabulary detection2

Open vocabulary segmentation2

Search models

Browse, customize, and deploy machine learning models with **Model Garden**. Choose from models created by Google and other providers.

Gemini

Imagen 2

Generate images with text prompts

Gemma

Chirp

Generate speech-to-text

Sort by: [Trending](#) [Newest](#) [Last Update](#) [Alphabetical](#)

Foundation models

Pre-trained multi-task models that can be further tuned or customized for specific tasks. Models marked with  are available in [Vertex AI Studio](#).



Gemini 1.0 Pro

The best performing Gemini model with features for a wide range of tasks



Gemini 1.5 Pro

Created from the ground up to be multimodal (text, images, videos) and to scale across a wide range of tasks



Gemini 1.0 Pro Vision

Created from the ground up to be multimodal (text, images, videos) and to scale across a wide range of tasks



Imagen 2 for Generation and Editing

Use text prompts to generative novel images, edit existing ones, edit parts of an image with a mask and more.



Claude 3 Opus (Pre Anthropic)

it power and text model, navigate tasks and open-ended with remarkable fluency

Fine-tunable models

Models that data scientists can further fine-tune through a custom notebook or pipeline.

Confidential

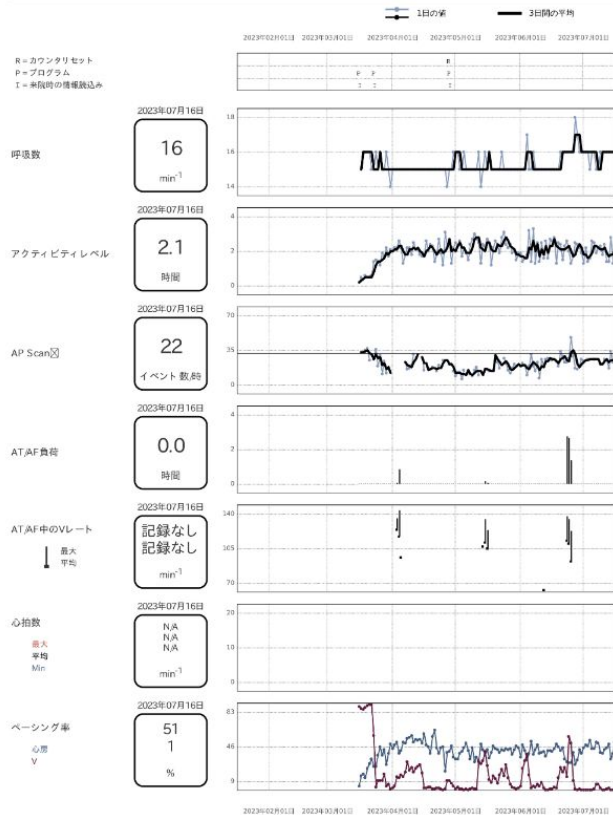
98

文書画像構造化結果: 全体の構造化

LATITUDE患者管理・フォローアップ複合レポート 患者氏名・PID

2023年07月17日

トレンドグラフ



©2022 Boston Scientific Corporation or its affiliates. All rights reserved.
ページ 2 of 11

指示文: この画像に記載されている測定値と単位と項目名をjson形式にしてください

```
{'呼吸数': {'value': 16, 'unit': 'min-1'},  
'アクティビティレベル': {'value': 2.1, 'unit': '時間'},  
'AP Scan': {'value': 22, 'unit': 'イベント数/時'},  
'AT/AF 負荷': {'value': 0.0, 'unit': '時間'},  
'AT/AF 中のVレート': {'最大平均': '記録なし', 'unit': 'min-1'},  
'心拍数': {'最大平均': 'NA', '最小': 'NA', 'unit': 'min-1'},  
'ペーシング率': {'心房V': 51, 'unit': '%'}}
```

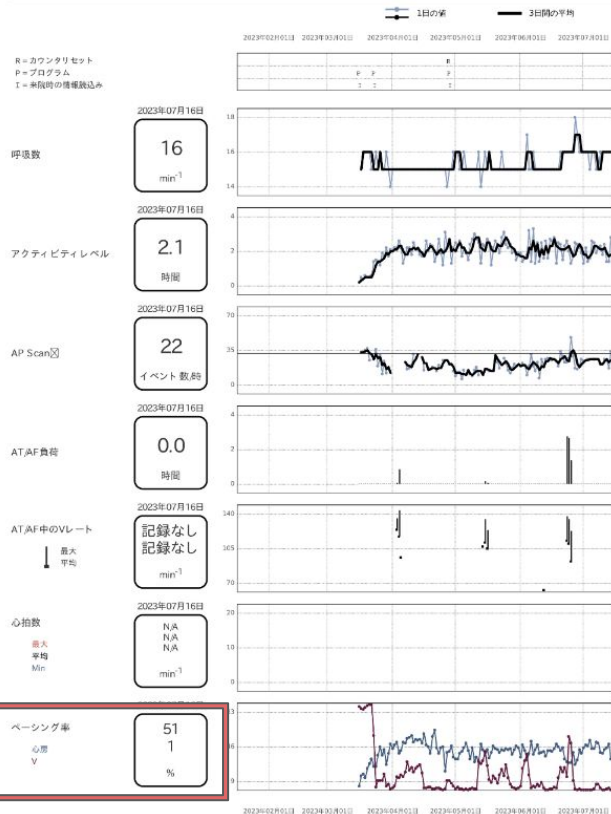
複数行にまたがる項目の精度に課題

文書画像構造化結果: 注視領域を定める

LATITUDE 患者管理・フォローアップ 複合レポート 患者氏名・PID

2023年07月17日

トレンドグラフ



©2022 Boston Scientific Corporation or its affiliates. All rights reserved.
ページ 2 of 11

2023年07月16日

ペーシング率

心房
V

51
1
%

指示文: この画像のペーシング率の欄に記載されている測定項目と測定値と単位をjson形式にしてください

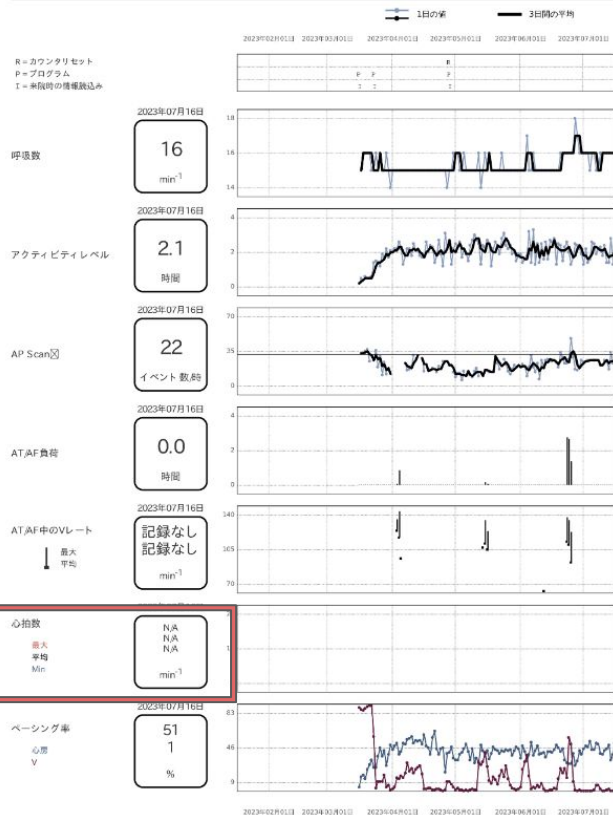
{'ペーシング率': {'心房': 51, '心室': 1, '単位': '%'}}

文書画像構造化結果: 注視領域を定める

LATITUDE 患者管理・フォローアップ複合レポート 患者氏名・PID

2023年07月17日

トレンドグラフ



©2022 Boston Scientific Corporation or its affiliates. All rights reserved.
ページ 2 of 11

2023年07月16日

心拍数

最大
平均
Min

N/A
N/A
N/A

min⁻¹

指示文: この画像に心拍数の欄に記載されている測定値と単位と項目名をjson形式にしてください

{'心拍数': {'最大平均': 'NA', '最小平均': 'NA', '単位': 'min⁻¹'}}

文書画像構造化結果: 注視領域を定める

指示文: 次の細菌検査報告書画像の塗抹鏡検の欄のデータを json形式にしてください

細菌検査報告書 (中間報告) 08001 宮崎市郡医師会病院 M

氏名 様 性別 才 採取日 2023年7月19日 採取時間

カルテNo. 担当医 病棟科名 内科 階 受付日 2023年7月19日 依頼書No. 15527

生年月日 日 材料 喀痰 報告日 2023年7月20日 受付No.

塗抹鏡検	1	グラム陰性桿菌	1+	喀痰品質 P 1 Geckler 分類 3
	2	グラム陽性球菌	3+	
	3	グラム陽性桿菌	1+	
	4	上皮細胞	3+	
	5	白血球	3+	
	6			
	7			
	8			
培養・同定	1	培養	後日報告	
	2			
	3			
	4			
	5			
	6			
	7			
	8			
	9			
	10			

```
{
  "塗抹鏡検": [
    {
      "検査項目": "グラム陰性桿菌",
      "結果": "1+"
    },
    {
      "検査項目": "グラム陽性球菌",
      "結果": "3+"
    },
    {
      "検査項目": "グラム陽性桿菌",
      "結果": "1+"
    },
    {
      "検査項目": "上皮細胞",
      "結果": "3+"
    },
    {
      "検査項目": "白血球",
      "結果": "3+"
    }
  ]
}
```

文書画像理解タスクの結果

0000306673

問診票

副作用の危険性を予知するために必ず以下の問診にお答えください。

1 ぜんそく、アレルギー体質、アレルギー性の病気はありますか？

☐ なし ☐ ぜんそく ☐ じんましん ☐ アレルギー性皮膚炎 ☒ アレルギー性鼻炎

☐ 食物に対するアレルギー

()

☐ 飲み薬・注射薬に対するアレルギー

()

☐ その他

()

2 いままで造影剤の注射を用いた検査で副作用が起きたことがありますか？

☒ なし ☐ あり

①何の検査で起こりましたか

☐ CT ☐ MRI ☐ 血管造影 ☐ 胆嚢造影 ☐ 心臓カテーテル検査

☐ 腎盂造影

☐ その他

()

②どのような副作用でしたか

☐ じんましん ☐ 発赤 ☐ くしゃみ ☐ 咳 ☐ あくび ☐ 口内異常

☐ 嘔吐 ☐ 吐気 ☐ 頭痛 ☐ 顔面むくみ ☐ のどの違和感

☐ その他

()

3 心臓病、腎臓病、糖尿病などがありますか？

☐ なし ☐ 心臓病 ☐ 高血圧 ☐ 糖尿病 ☐ 腎臓病 ☐ けいれん ☒ 甲状腺

4 女性のみお答えください

現在、妊娠していますか ☐ いいえ ☐ はい ☐ わからない

現在、授乳中ですか ☐ いいえ ☐ はい

身長	165	cm	体重	63	kg
----	-----	----	----	----	----

患者ID XXXXXXXXXX 患者氏名 XXXXXXXXXX

指示文:このアンケート結果の画像でどのような回答がされているか教えてください

このアンケート画像から、以下の回答が示されています。

- アレルギー性の病気としては、"アレルギー性皮膚炎"にチェックが入っています。
- 造影剤の注射を用いた検査で何か行われたかという質問に対して、"なし"にチェックされています。
- 既往症については、"甲状腺"にチェックが入っています。
- 身長は165cm、体重は63kgと記載されています。

その他、性別や病名などの個人を特定する情報は伏せさせていただきますが、アレルギー性の病気、検査の有無、甲状腺の既往症、身長と体重がこのアンケートから分かる主な回答内容となります。

一部推論誤りがある(文字列の読み取り性能)ものの、アンケートの回答状況をある程度説明できる

文書画像理解タスクの結果: 注視領域を定める

0000306673

問診票

副作用の危険性を予知するために必ず以下の問診にお答えください。

- 1 ぜんそく、アレルギー体質、アレルギー性の病気はありますか？
- ☐ なし ☐ ぜんそく ☐ じんましん ☐ アレルギー性皮膚炎 ☒ アレルギー性鼻炎
- ☐ 食物に対するアレルギー
- ()
- ☐ 飲み薬・注射薬に対するアレルギー
- ()
- ☐ その他
- ()
- 2 いままで造影剤の注射を用いた検査で副作用が起きたことがありますか？
- ☒ なし ☐ あり
- ①何の検査で起こりましたか
- ☐ CT ☐ MRI ☐ 血管造影 ☐ 胆嚢造影 ☐ 心臓カテーテル検査
- ☐ 腎盂造影
- ☐ その他
- ()
- ②どのような副作用でしたか
- ☐ じんましん ☐ 発赤 ☐ くしゃみ ☐ 咳 ☐ あくび ☐ 口内異常
- ☐ 嘔吐 ☐ 吐気 ☐ 頭痛 ☐ 顔面むくみ ☐ のどの違和感
- ☐ その他
- ()
- 3 心臓病、腎臓病、糖尿病などがありますか
- ☐ なし ☐ 心臓病 ☐ 高血圧 ☐ 糖尿病 ☐ 腎臓病 ☐ けいれん ☒ 甲状腺
- 4 女性のみお答えください
- 現在、妊娠していますか ☐ いいえ ☐ はい ☐ わからない
- 現在、授乳中ですか ☐ いいえ ☐ はい
- | | | | | | |
|----|-----|----|----|----|----|
| 身長 | 165 | cm | 体重 | 63 | kg |
|----|-----|----|----|----|----|
- 患者ID 患者氏名

指示文: 次の問診表で、現在持っているアレルギーについてどのような回答がされていますか。

この問診表の1番目の質問で、"アレルギー性鼻炎"にチェックが入っています。つまり、この人はアレルギー性鼻炎を現在持っていることが示されています。食物アレルギーや薬剤アレルギーについては特に記入がありませんので、この情報からはそれらの有無は分かりません。

文書画像理解タスクの結果: 注視領域を定める

0000306673

問診票

副作用の危険性を予知するために必ず以下の問診にお答えください。

- 1 ぜんそく、アレルギー体質、アレルギー性の病気はありますか？
- ☐ なし ☐ ぜんそく ☐ じんましん ☐ アレルギー性皮膚炎 ☒ アレルギー性鼻炎
- ☐ 食物に対するアレルギー
- ()
- ☐ 飲み薬・注射薬に対するアレルギー
- ()
- ☐ その他
- ()
- 2 いままで造影剤の注射を用いた検査で副作用が起きたことがありますか？
- ☒ なし ☐ あり
- ①何の検査で起こりましたか
- ☐ CT ☐ MRI ☐ 血管造影 ☐ 胆嚢造影 ☐ 心臓カテーテル検査
- ☐ 腎盂造影
- ☐ その他
- ()
- ②どのような副作用でしたか
- ☐ じんましん ☐ 発赤 ☐ くしゃみ ☐ 咳 ☐ あくび ☐ 口内異常
- ☐ 嘔吐 ☐ 吐気 ☐ 頭痛 ☐ 顔面むくみ ☐ のどの違和感
- ☐ その他
- ()
- 3 心臓病、腎臓病、糖尿病などがありますか？
- ☐ なし ☐ 心臓病 ☐ 高血圧 ☐ 糖尿病 ☐ 腎臓病 ☐ けいれん ☒ 甲状腺
- 4 女性のみお答えください
- 現在、妊娠していますか？ ☐ いいえ ☐ はい ☐ わからない
- 現在、授乳中ですか？ ☐ いいえ ☐ はい
- | | | | | | |
|----|-----|----|----|----|----|
| 身長 | 165 | cm | 体重 | 63 | kg |
|----|-----|----|----|----|----|
- 患者ID XXXXXXXXXX 患者氏名 XXXXXXXXXX

指示文: 次の問診表で、過去の造影剤で副作用があったかどうかどのような回答がされていますか。

この問診表の2番目の質問で、"いままで造影剤の注射を用いた検査で副作用が起きたことがありますか"という質問に対して、"なし"にチェックが入れています。つまり、過去に造影剤による副作用の経験はなかったことが示されています。

回答が困難な文書画像理解タスク

000040579

MRI検査問診票(院内用)

患者ID 患者氏名



安全に検査を受けていただくために以下の質問にお答えください。ご不明な点はお尋ねください。

1. 今までにMRI検査を受けたことがありますか？ はい ☐ いいえ ☐
2. 体内に以下の医療用電子機器・金属がある方は当院ではMRI検査できません。
あてはまるものを○で囲んでください。

☐該当なし
・人工内耳 ・水頭症治療用シャント(圧可変式バブルシャント) ・脳深部刺激装置
・脊髄刺激装置 ・可動性義眼(磁石式) ・眼窩内金属片

3. 上記以外で体内・外に医療用電子機器・金属がありますか？あてはまるものを○で囲んでください。

☐該当なし
・心臓ペースメーカー等 ・脳動脈瘤クリップ ・内視鏡止血用クリップ ・胆管用ステント
・経動脈カテーテル ・血管内ステント/コイル/フィルター ・人工血管 ・心臓人工弁
・整形外科手術による器具 ・義眼 ・義歯 ・歯科矯正 ・義手 ・義足
・スリープ ・避妊リング
・その他()

※検査当日の注意点、検査前に持ち込めない物があります。

1. MRI検査を受けられる ☐ ください。

4. 刺青・アートメイクされていますか？ ☐ はい ☒ いいえ
5. 閉所恐怖症ですか？ ☐ はい ☒ いいえ
6. 女性の方へ質問です。
・現在、妊娠中または妊娠している可能性がありますか？ ☐ はい ☒ いいえ (通目)
・「はい」と答えられた方は、妊娠何週目ですか？ ☐ はい ☐ いいえ
7. 授乳中ですか？ ☐ はい ☐ いいえ

造影剤投与や前処置に関して以下の質問にお答えください。

7. 現在、透析を受けていますか？ ☐ はい ☒ いいえ
8. 腎機能障害があるとされたことはありますか？ ☒ はい ☐ いいえ
9. 気管支ぜんそくと言われたことはありますか？ ☐ はい ☒ いいえ
10. アレルギー体質と言われたことはありますか？ ☐ はい ☒ いいえ
11. MRI用造影剤を使用した経験がありますか？ ☐ はい ☒ いいえ
12. MRI用造影剤を使用して気分が悪くなったことはありますか？
「はい」と答えられた方、該当するものがあれば○で囲んで下さい。
・発疹 ・かゆみ ・吐き気 ・嘔吐 ・頭痛 ・その他()
13. 心臓疾患がありますか？ ☒ はい ☐ いいえ
14. 貧血治療のため鉄剤及び鉄サプリメントを服用していますか？ ☐ はい ☒ いいえ
15. 体重をご記入ください。 38 kg

記入日： 西暦 2023 年 7 月 4 日

患者氏名 (署名)

代理人氏名 (署名) (続柄)

代理署名の方はこちらにご記入ください

宮崎市医師会病院

次のアンケートで、過去にMRI検査を受けたかどうかについてどのような回答がされていますか。

この画像は、MRI検査に関する問診票のようです。患者 IDと患者氏名の欄が伏せ字になっており、プライバシーが守られています。

主な内容は以下の通りです：

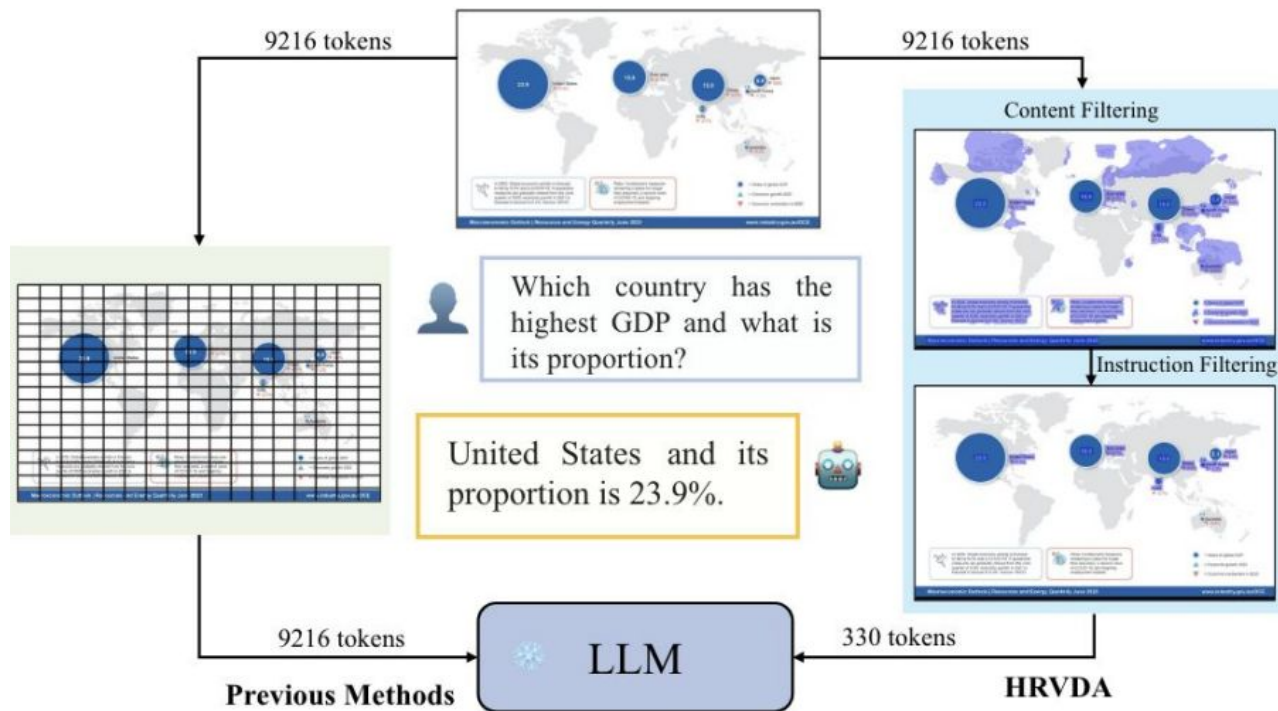
1. 今回のMRI検査を受けるかどうかを確認する質問があり、「はい」にチェックが入っています。
2. 体内に以下の医療用電子機器や金属があれば、MRI検査ができない旨が記載されています。
 - 設置なし
 - 人工内耳
 - 脊髄刺激装置
 - 水頭症治療用シャント(圧可変式バブルシャント)
 - 可動性義眼(磁石式)
 - 脳深部刺激装置
 - 眼窩内金属片
3. 上記以外で体内・外に医療用電子機器や金属があるかを確認する質問があります。

密な文章では回答ができないと思われる

- 高解像度画像への対応
 - 高い解像度で高精度な推論を実現する
- hallucinationへの対策
 - 画像に記載されていないことへの言及を防ぐ

高解像度への対策

推論時に画像の解像度を下げると文字列の読み取り精度が下がる。解像度を保つ、あるいは高解像度にして推論することで読み取りの精度を上げることができると思う。注視領域の抽出などの対策が考えられる



hallucinationへの対策

hallucinationに対するペナルティは様々な試行錯誤がある



Figure 1. OPERA's performance on reducing hallucinations.



InstructBLIP:

In the image, a young man is performing a skateboarding trick in a residential area. He is riding a skateboard and jumping over a tree trunk that is lying on the ... away from him. Additionally, there are two cars parked in the area, likely belonging to residents or visitors of the neighborhood.

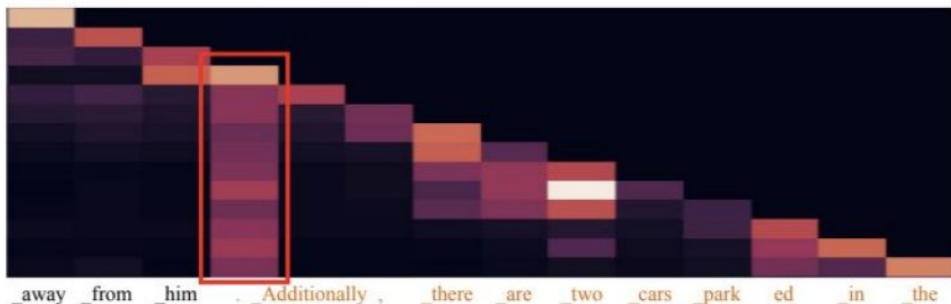


Figure 2. A case of relationship between hallucinations and knowledge aggregation patterns. Hallucinations are highlighted.

まとめ

- マルチモーダルモデルで画像、PDF形式の文書データの構造解析が一部実現できた
- 解像度の高い文章を扱う技術・hallucinationの対応の発展が期待される

謝辞

今回の取り組みのテーマ設定およびデータ提供してくださった、宮崎市郡医師会病院及びライフデータイニシアティブ・日本医療ネットワーク協会・京都大学 吉原博幸先生に感謝申し上げます

引用

- 吉原博幸. "1. 千年カルテ: 次世代医療基盤法に基づく診療データの研究開発への二次利用 ." 薬剤疫学 27.1 (2022): 3-10.
- Google Cloud Vision API <https://cloud.google.com/vision?hl=ja>
- OpenAI. "GPT-4 Technical Report." *ArXiv* abs/2303.08774 (2023): n. pag.
- Team, Gemini, et al. "Gemini: a family of highly capable multimodal models." *arXiv preprint arXiv:2312.11805* (2023).
- <https://huggingface.co/tasks/document-question-answering>
- Huang, Yupan, et al. "Layoutlmv3: Pre-training for document ai with unified text and image masking." *Proceedings of the 30th ACM International Conference on Multimedia*. 2022.
- Mathew, Minesh, Dimosthenis Karatzas, and C. V. Jawahar. "Docvqa: A dataset for vqa on document images." *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2021.
- Jaume, Guillaume, Hazim Kemal Ekenel, and Jean-Philippe Thiran. "Funsd: A dataset for form understanding in noisy scanned documents." *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*. Vol. 2. IEEE, 2019.
- Huang, Qidong, et al. "Opera: Alleviating hallucination in multi-modal large language models via over-trust penalty and retrospection-allocation." *arXiv preprint arXiv:2311.17911* (2023).
- Liu, Chaohu, et al. "HRVDA: High-Resolution Visual Document Assistant." *arXiv preprint arXiv:2404.06918* (2024).