

薬剤鑑別アプリケーションの技術的検討

2021年7月17日

(株) アリス 取締役 伊東祐孝、上田 裕
NPO法人広域連携医療福祉システム機構 理事長 周藤安造
NPO法人広域連携医療福祉システム機構 理事 沖 眞
開志専門職大学情報学部 講師 西川浩平

1. 薬剤鑑別アプリケーションの要件

- 薬剤のヒート画像から、薬剤名を表示
- 薬剤の錠剤画像から、薬剤名を表示
- ヒート及び錠剤画像は、スマホで手動で撮影
- スマホで撮影後、薬剤名を表示
- 撮影薬剤名と、先行薬の薬剤名を表示
- 検証時の費用を抑える

ヒート画像



錠剤画像



2. 薬剤のヒート画像から薬剤名表示

- (1) ヒート画像から文字列抽出 (OCR)
- (2) ヒート上の文字列から薬剤名を検索 (薬剤データ分析)
- (3) 先行薬の検索 (薬剤データ分析)
- (4) アプリケーション構成

3.1.1 薬剤のヒート画像から文字列抽出 —OCR製品(1)—

オープンソースとクラウドサービスから、TesseractとGoogle Cloud Visionを採用

	Tesseract	Google Cloud Vision API	MS Azure Computer Vision API	Amazon Rekognition
形態	オープンソース (2005年)	クラウドサービス		
言語	検証環境では、日本語含む	日本語含む56ヶ国語	日本語含む26ヶ国語	英数字+記号のみ
価格	無料	\$1.5/1,000枚 (月1,000枚まで無料)	\$1.5/1,000枚 (月5,000枚まで無料)	\$1.0/1,000枚 (最初12ヶ月間 月5,000枚まで無料)

3.1.1 薬剤のヒート画像から文字列抽出 —OCR製品(2)—

(1) Tesseract

Tesseractは、米Hewlett-Packard(HP)が1985~1995年にかけて開発していたOCRエンジン。ネバダ州立大学ラスベガス校(UNLV)が1995年に開催したOCRソフトウェアの精度を評価するコンテストで、トップ3の成績を記録したという。しかしHPがOCR事業からの撤退を決めたことから、「Tesseractはほこりをかぶっていた」(Google)。

その後「HPの関係者がTesseractをお蔵入りさせるよりも公開した方がよいと考え」(Google)、UNLVの情報科学研究所と共同で2005年にオープンソース化した。Googleはバグ修正などで協力し、2006年1月27日にSourceForge.netに登録した。
【日経XTECニュースより】

(2) Google Cloud Vision

Google Cloud Platform(グーグル クラウド プラットフォーム、GCP)とは、Googleが提供しているクラウドコンピューティングサービスである。Google 検索やYouTubeなどのエンドユーザー向けのサービスでも、同じインフラストラクチャーが利用されている。さまざまな管理ツールに加えて、一連のモジュール化されたクラウドサービス(90以上)が提供されており、コンピューティング、データストレージ、データ分析、機械学習などのサービスがある。

人工知能

Cloud AutoML

Cloud TPU

Cloud machine Learning Engine

Cloud Job Discovery

Dialogflow Enterprise

Cloud Natural Language

Cloud Speech-to-Text

Cloud Text-to-Speech

Cloud Translation API

Cloud Vision API

(今回使用する、OCR API)

Cloud Video Intelligence

【ウィキペディアより】

3.2.1 OCR検証 —テスト環境—

(1) 薬剤のヒート画像



drug01.jpg



drug02.jpg



drug03.jpg



drug04.jpg

(2) OCR環境

1) Tesseract

Tesseract Open Source OCR Engine v5.0.0-alpha.20191030 with Leptonica

2) Google Cloud Vision

クラウドサーバからPHP言語を使用して、Webサービスとして利用

3.2.2 OCR検証 —Tesseract—

画像ファイル名	drug01.jpg	drug02.jpg	drug03.jpg	drug04.jpg
OCR結果	<p>剛M野央 寺の英遊リー—ロエA1ヒビ皿 へで \\NN SNKN 55 \\NANA NT 5 \\人へ、人 NNへNAS ...人3 3拓 ctysts5eepXt6) AS し さ 、迷@来WW yel9821802 る NMM 必の青田 、IINNNNN はさどにNNNSNNNN 旭 2 じY 3 通。</p>		<p>1 葉に3 ceelr626271 6 川軍川川HL 「Of」3u09 雁外十息 たく子エとへキキエン</p>	<p>ア— レンるるc —アアマやて ヤヤターシヤンシシタ SSJSS505451 い でででそでで ででですす でででででででで でででマ— で でででででそでそででで でででででででででで ででマ の =老 3 ジラタタタタスクイク ウツナタタタタケッテン アアアアアアタタタタタスク スアアア アアアアアアアアアアアア そそをををををををををを やてででヤタルででヤッタンタ ヤヤシヤンでシイマヤタマテテ ター— ててててててマママタタイタ スベ ツヤヤタツイン 緒フイタ でででででででででで ででですすでででで トでででででででで マ SM NN NANA N</p>

3.2.3 OCR検証 —Google Cloud Vision(1)—

画像ファイル名	drug01.jpg	drug02.jpg	drug03.jpg	drug04.jpg
OCR結果	<p>A RORUV ST7T1 2.5 ng DSEP. 高コレステロール血症の薬 (01)04987081784134 ロスバ スタチン 2.5mg IDSEP DSEP 高コレステロール血症の薬 (01)04287081784134 スタチン? 2.5mg ロスバ スタチン 2.5mg IDSEP DSEP 高コレステロール血症の薬 (91)04987081784134 スタチン 2.5mg IDSEP DSEP ロスバ スタチン 2.5mg IDSEP 高コレステロール血症の薬 (01)04987081784134 スタチン 2.5mg IDSEP DSEP 88 高コレステロール血症の薬 to 359 ¥1762) ¥735) ¥2,497 ¥1,762 4750 ます 69</p>	<p>Betanis 25 ベタニス 遊活動 錠剤 25mg ベタニス 25mg 25mg おしだす。 ベタニス ベタニス 25mg astellas アステラス製薬 ベタニス 25mg ベタニス 遊活動 錠剤 25mg (01)04987233745006 と当 ¥2,497 社.762 (097夫 ¥2 ¥10000 お針 預収書とさせて 2000 帯群 的銭伝票 ===== り引換 当日限りです 釣銭番 90750</p>	<p>Fexofenadine Hydrochloride 60 フェキソフェナジン 60mg[G] (01)04987792981285 60 フェキソフェナジンOUジェキソフェナ ジン EERE EERE 60mg[G] 塩酸塩錠 60mg[G] 60 フェキソフェナジン フェキソフェナジン 塩酸塩錠 60mg[G] 60mg[G] 60 フェキソフェナジン フェキソフェナジン 塩酸塩錠 60mg[G] おしだす。 (01)04987792931285 と当 真当 ¥2,497 ¥1,762 ¥2,497 10000 5 894 TO 1/30 53 8 4 590</p>	<p>INE2BG3 ムラ カシヨウヨウサン 加味速達散 ツムラ カシヨウヨウサン 加味速達散 2.5g 2.5g 24 24 (01)04987138802491 (01)04987138802491 ¥2,497 M1,799 10000 催收票 3768 POS番号01 2019/11/30 的銭権 おっり月付 2019/11/30(土)10:21 釣りの清算は当日限りです 9/11/30 2.80845907503 的銭番号</p>

3.2.3 OCR検証 —Google Cloud Vision(2)—

drug01.jpg画像を、右回転で10度づつ測定

右回転の角度	0度	10度	20度	30度
OCR結果	A RORUV ST7T1 2.5 ng DSEP. 高つしステロール西症の錠 (01)04987081784134 ロスバ スタチン 2.5mg IDSEP DSEP 高コレステロール血症の薬 (01)04287081784134 スタチン ? 2.5mg ロスバ スタチン 2.5mg IDSEP [DSEP] 高コレステロール血症の薬 (91)04987081784134 スタチン 2.5mg [DSEP] ロスバ スタチン 2.5mg IDSEP [DSEP] 高コレステロール血症の薬 (01)04987081784134 スタチン 25mg 10SEPI ロスバ スタチン 2.5mg IDSEP	ROUVV ST7T1N 2.5 ng DSEP 高つしステロール西症の錠 (01)04987081784134 ロスバ スタチン 2.5mg IDSEP DSEP 高コレステロール血症の薬 (01)04987081784134 スタチン おし 2.5mg ロスバ スタチン 2.5mg IDSEP [DSEP] 高コレステロール血症の薬 (91)04987081784134 スタチン 2.5mg ロスバ スタチン 2.5mg IDSEP [DSEP] 高コレステロール血症の薬 (01)04987081784134 スタチン 2.5mg 10SEPI ロスバ PP スタチン 2.5mg IDSEP	ROUVV ST7T1N 2.5 ng DSE9 高つしステロール西症の錠 (01)04987081784134 ロスバ スタチン 2.5mg IDSEP DSEP 高コレステロール血症の薬 (01)04987081784134 スタチン おし 2.5mg [DSEP] ロスバ スタチン 2.5mg IDSEP 高コレステロール血症の薬 (91)04987081784134 スタチン 2.5mg [DSEP] ロスバ スタチン 2.5mg IDSEP [DSEP] 高コレステロール血症の薬 (01)04987081784134 スタチン 2.5mg LOSEPI ロスバ スタチン 2.5mg IDSEP PP	IROGOV STETIO 2.5 ng DSE9 高つしステロール西症の錠 (01)04987081784134 ロスバ スタチン 2.5mg TDSEP DSEP 高コレステロール血症の薬 (01)04987081784134 スタチン 2.5mg [DSEP] ロスバ スタチン 2.5mg IDSEPI 高コレステロール血症の薬 (91)04987081784134 スタチン 2.5mg [DSEP] ロスバ スタチン 2.5mg IDSEPI [DSEP] 高コレステロール血症の薬 (01)04987081784134 スタチン 2.5mg OSEPI ロスバ スタチン マ金属 2.5m IDSEP PP

3.3 OCR検証 —結果—

今回のOCRとしては Google Cloud Vision を使用して、薬剤鑑別アプリケーションの構築を行う

- 1) 水平軸が取れていない画像でも、文字として判読している
- 2) ヒートで使われているフォント種類は不明で、多くのフォントが使われている可能性がある
- 3) 無料では無いが、比較的安価に使用することができる
(今回の検証でのGoogle Cloud Vision費用は0円)

機能	1,000 ユニットあたりの料金		
	最初の 1,000 ユニット/月	1,001~5,000,000 ユニット/月	5,000,001 以上の ユニット/月
テキスト検出	無料	\$1.50	\$0.60

4.1 薬剤データ分析 ー薬剤データ(1)ー

薬剤のヒートに表示されている文字列から、薬剤を決定する「薬剤データ」が手元に無い為、下記のデータで代用する

(1) 厚生労働省がHPで公開している薬価データ

「薬価基準収載品目リスト及び後発医薬品に関する情報について(令和元年9月30日まで)」として、EXCEL形式でダウンロード可能
今回の検証では、内容薬を使用する。

- ・先発医薬品と後発医薬品のデータ有り
- ・先発医薬品と後発医薬品には、直接親子関係の表記は無いが、成分等の照合で推測が可能
- ・薬剤のヒートに表示される情報は無いが、成分・量・品名から推察する

	データ数
内用薬	10,941
注射薬	4,032
外用薬	2,466
歯科用薬剤	28

4.1 薬剤データ分析 ー薬剤データ(2)ー

成分: フェキソフェナジン塩酸塩の28品中、サンプルで7品表示

区分	薬価基準収載医薬品コード	成分名	規格	品名	メーカー名	診療報酬において加算等の算定対象となる後発医薬品	先発医薬品	同一剤形・規格の後発医薬品がある先発医薬品	薬価	経過措置による使用期限	備考
内用薬	4490023F1016	フェキソフェナジン塩酸塩	60mg錠	フェキソフェナジン塩酸塩60mg錠		後発品			12.20		
内用薬	4490023F1024	フェキソフェナジン塩酸塩	60mg錠	アレグラ錠60mg	サノフィ		先発品	○	57.40		
内用薬	4490023F1237	フェキソフェナジン塩酸塩	60mg錠	フェキソフェナジン塩酸塩錠60mg 「明治」	Meiji Seika ファルマ	後発品			23.20		
内用薬	4490023F1245	フェキソフェナジン塩酸塩	60mg錠	フェキソフェナジン塩酸塩錠60mg 「モチダ」	ニプロファーマ	後発品			23.20		
内用薬	4490023F1253	フェキソフェナジン塩酸塩	60mg錠	フェキソフェナジン塩酸塩錠60mg 「CEO」	セオリアファーマ	後発品			23.20		
内用薬	4490023F1251	フェキソフェナジン塩酸塩	60mg錠	フェキソフェナジン塩酸塩錠60mg 「JG」	日本ジェネリック	後発品			23.20		
内用薬	4490023F1270	フェキソフェナジン塩酸塩	60mg錠	フェキソフェナジン塩酸塩錠60mg 「サワイ」	沢井製薬	後発品			23.20		

4.2 薬剤データ分析 ー薬剤データ作成ー

厚生労働省がHPで公開している薬価データから、以下のデータ加工して薬剤データを作成

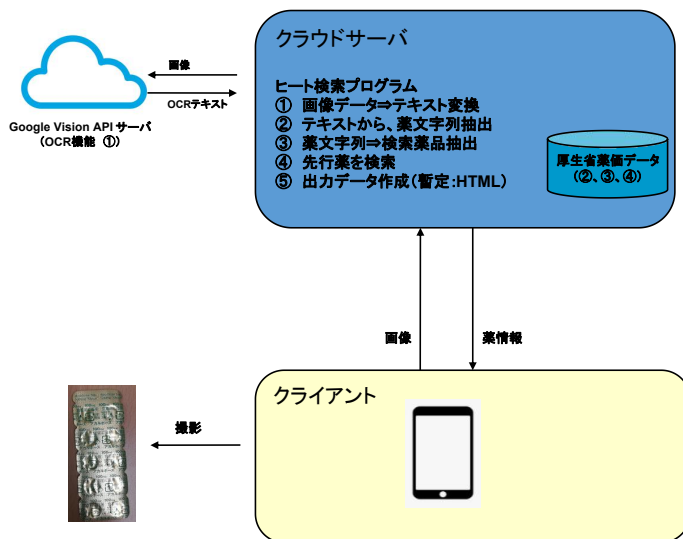
- 1) 先発品と後発品の薬剤を抽出して、親子関係を作成
- 2) 品名から、メーカー名抽出
例) JG⇒日本ジェネリック、サワイ⇒沢井製薬等
- 3) 成分、規格の情報整理

4.3 薬剤データ分析 ーOCR結果の照合一

OCRから抽出された文字列と、薬剤データの照合ロジック

- 1) 最小文字数(2文字)で頻度1は削除
- 2) 不要文字列の削除
例) プラ等
- 3) 薬剤データの照合
- 4) 照合時に、類似文字の考慮
例) 「リ」と「ソ」、「u」と「μ」等
- 5) 照合頻度値を数値化して、複数の候補薬剤名の順位作成
- 6) 照合された薬剤が後発品の場合には、先発品の検索

5.1 薬剤鑑別アプリケーション システム構成



5.2 薬剤鑑別アプリケーション 実行結果

検索文字列：JG(13) フェキシフェナジン(7) 60mg(7) 塩酸塩錠(4)

薬価基準収載品目リスト及び後発医薬品に関する情報

検索薬品

区分	薬価基準収載医薬品コード	成分名	規格	品名	メーカー名	診療報酬において加算等の算定対象となる後発医薬品	先発医薬品	同一剤形・規格の後発医薬品がある先発医薬品	薬価	経過措置による使用期限	備考	先発品薬品コード
内用薬	4490023F1261	フェキシフェナジン塩酸塩	60mg1錠	フェキシフェナジン塩酸塩錠60mg「JG」	日本ジェネリック	後発品			23.20			4490023F1024

先発品薬品

区分	薬価基準収載医薬品コード	成分名	規格	品名	メーカー名	診療報酬において加算等の算定対象となる後発医薬品	先発医薬品	同一剤形・規格の後発医薬品がある先発医薬品	薬価	経過措置による使用期限	備考	先発品薬品コード
内用薬	4490023F1024	フェキシフェナジン塩酸塩	60mg1錠	アレグラ錠60mg	サノフィ		先発品	○	57.40			

5.3 薬剤鑑別アプリケーション — 検証結果 —

任意の39種類の薬剤を使用しての検証

(1) 正解率

19薬剤 / 39薬剤 = 48%

(2) 誤認識例

1) OCRが認識しない 9薬剤 / 39薬剤 = 23%

例)



0.5g07.



アルファカルシド0.5gサワイ9.5ug

2) 薬データにデータが無い 8薬剤 / 39薬剤 = 20%

例) 「タケキャブ 20」、「Mylan」⇒「ファイザー」等

6. まとめ

(1) 薬剤のヒート画像から、薬剤名を表示

- 1) ヒート表示名データベースが必須
- 2) OCR変換前の、画像変換処理が必要
- 3) 100%に近い検索性能が必要

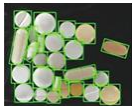
(2) 錠剤画像から、薬剤名を表示

- 1) 錠剤表示名データベースが必須
- 2) 錠剤撮影時のテクニックが必要
⇒ 刻印の錠剤もあり、文字の識別が困難

A. 資料 PROOFIT 1D 富士フィルム株式会社

1. 製品名 一包化監査支援システム「PROOFIT 1D(プルーフイット ワンドース)」
2. 発売日 2019年1月11日
3. 標準ユーザー渡し価格(税別)
機器本体システム一式:10,000,000円
(標準的な設置環境への設置費用含む)

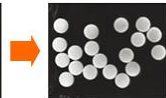
【図2】錠剤やカプセル剤を高精度に抽出し状態を確認



【図3】ならし機能



錠剤が混みあっている状態



ならし機能で錠剤を散らす



一包化監査支援システム
「PROOFIT 1D」