インシュアテックイノベーション　連載第3弾

AI-OCR活用から生成AI(ChatGPT)との共生まで⑤

アイリックコーポレーション（IRRC）フェロー
保険・ヘルスケアDX担当　畔柳 主税

AI―OCRの「読取精度100％」への期待と現実

保険会社からAI―OCRの活用の相談・RFPへの参加打診が続いている。そんな中で、最近は、AI―OCRに過大な期待をし過ぎて、プロジェクトが延期になるケースも増えている。その一方で、AI―OCRの限界を理解して、現実的な解決策を作り、現場で上手に活用する会社もある。保険会社向けのAI―OCRのRFPへの参加・導入PJ（プロジェクト）に15以上参加して、保険会社の実態を熟知している弊社の諸吉に、保険会社におけるAI―OCR導入の成功のポイント等について、その経験をもとにレポートしてもらう。

インフォディオ社と共同で精度改善ミーティング（右が諸吉）

AI―OCR技術の導入は、保険会社のデジタルトランスフォーメーション（DX）の進展に大きく貢献する可能性を持つ。その効果を最大限に引き出すためには、技術の正確な理解と適切な活用方法が重要となる。当社が提供するAI―OCRは、保険証券や健康診断結果表、診療明細書などの非定型帳票に対応している。これらの帳票には、多様なレイアウトや記載項目が存在し、その複雑性をAIが解決することにより、業務効率化・生産性向上が期待される。

保険会社の業務部門がユーザーの場合、お客さまへの利便性（CX）の向上と査定業務の効率化を期待することになる。AI―OCR導入の評価においては、人による手作業で処理、目検確認の代替となり得るか、という点が重要である。

執筆時点の技術においては、AI―OCRによる精度は100％の時もあれば、80％の時もある。例えば、画像の鮮明さ、罫線の有無などが精度に影響し、低い解像度の画像や、特殊なレイアウトや項目を含む場合は、精度が低下する。そのため、ユーザーとなる保険会社からは、人によるチェックをゼロにするのではなく、複数人体制のデータ化における工数を減らす手段としての相談をいただいている。

OCR結果の「信頼度」を使った効率化のご要望に応える場合もあるが、「信頼度」自体の信頼性はどのように担保するかが問題となる。つまり、信頼度が100％であれば、人のチェックを省略できる、という考え方もあるが、その100％は常に完全ではないのがAI活用の現実である。

「正解データ」でAIの学習方向性決める

精度向上に関わる正解データ作成の取り組み

AI―OCRの精度改善の取り組みについてその一部を紹介したい。そもそも、AI―OCRはどこにAIを活用しているのか。お客さまからは、「OCRを使っているうちにAIが自動的に学習してくれるのか？」というご質問をいただくことが多い。このご質問の答えとして、「理屈としてはそのようになるが、AIが自動的に学習するための準備が必要であり、そのためには多くの『正解データ』の作成が必要になる」ということをお伝えしている。

「正解データ」とAIによるデータの抽出結果を突合させ、精度の低い項目、傾向を分析し、AIの学習の方向性を決めるのである。

AIに対する学習の方向性を決めたうえで、複数の新しい帳票を学習に利用することにより、AIの学習が強化されていく。

精度改善イメージ図

「正解データ」を作成する工程をいかに時間短縮できるかが、OCR結果の分析スピードを上げることに関わってくる。健康診断書を例にその難しさをお伝えしたい。

健康診断の結果表は全国の各医療機関が利用している健診管理システムによってレイアウトが異なっている。検査項目数は50項目程度から多いものでは200項目を超える場合もある。定性的な表現の表記ゆれも存在する。同じ医療機関でも医師が診察時に手打ち入力をしている場合は表現が異なることがある。

ユーザーである保険会社は、査定において必要になる項目については正解として保持している。しかし、個人情報の取り扱い上、当社のようなOCRベンダーに提供するためのハードルは高く、全ての項目を網羅しているわけではない。そのため、当社では契約において匿名加工情報として認められたデータのみを利用し、できるだけ多くの項目の「正解データ」の作成を行っている。当然、各医療機関とデータ連携ができるわけではないため、「正解データ」の作成はPDFなどの結果表の画像データからの作成となる。専門的なデータを取り扱う人的リソースが必要となる。

当社は、保険代理店ビジネスによって、長年にわたり生命保険業界をリードしてきた仕組みをもつ。多くのお客さまが加入している保険証券をお預かりしデータ化することで、お客さまが分かりやすく理解できるためにビジュアル化するものだ。保険証券のデータ化においては多くの時間を必要とするが、当社はお客さまに喜ばれるためのデータ化に注力してきたことから、複雑な帳票から必要なデータを探して入力する専門チームをもっている。現在においては、保険証券についてもOCRによって効率的にテキスト抽出したうえで「正解データ」を作成しているが、保険証券の「正解データ」作成業務で培った仕組みや体制が、健康診断書などの他の複雑な帳票における対応につながっているのである。

精度向上における全体最適

健康診断書など医療機関が発行する帳票は、医療機関ごとにレイアウトが異なることを述べたが、その違いは検査項目の順番や実施有無の違いもあれば、「前回」と「今回」の検査結果の順番の違い、西暦と和暦、定性表現の違い（「異常なし」「問題なし」「特記事項無し」）などさまざまである。

AI―OCRによる最終的な結果抽出にも得手不得手があるため、帳票によって精度のばらつきが出てしまう。精度改善を求めるユーザーは、精度が低い特定の帳票が目立ち、その帳票の改善に注力してしまう傾向にある。特殊な傾向を持つ帳票の精度改善にAIの学習を振り分けた場合、他の帳票の精度を低下させてしまうこともあり、全体最適が保てないことがある。そのため、改善を行うためには十分な調査と分析が必要になり時間を要することもあるが、ユーザーに対してはその分析や対応方法について丁寧に説明することを心掛けている。

当社はAI―OCRの開発元であり、当社子会社のインフォディオ社と共同で営業・SE・エンジニアよる精度改善のミーティングを定期的に行っている。OCRエンジンのバージョンアップの影響も議論し、特定の帳票の部分的な精度改善とAI―OCR製品全体としての継続的な精度向上を図っている。

ユーザーで学習させるAI―EXの開発

ここまで、精度改善の仕組みや体制について当社の取り組みを紹介したが、冒頭にあるようなお客さまが想像するような、AIによる自動学習できる仕組みについても、提供をはじめており、今後も対象帳票の拡大を行っていく予定である。その仕組みは、インフォディオ社が開発した「AI―EX」であり、すでに一般に業務利用する請求書などの帳票においては実績があるが、保険会社が利用するような特殊な帳票においても利用を順次拡大していきたいと考えている。

「AI―EX」はユーザーのフィードバックを画面上で直接OCRの学習に活用することができるが、従来からの事前に学習させたOCRと同時利用することも可能である。同時に二つの結果を得て、信頼度が高い方の結果を利用することで、全体精度が大きく向上することが期待される。

マイナポータルによる医療データ活用との共存

健康診断書や診療明細書などは将来的には、どの医療機関で受診した結果においても、マイナポータル上で参照することができるようになり、本人の承認を得れば保険会社の査定に利用できる仕組みが整いつつある。

しかし、全ての保険の契約者にマイナポータルでの連携ができるようになるまでは、これまでどおり健康診断結果の写しの提出を求める運用が併用されると考えられ、スマホを使った写真撮影などでAI―OCRでの利用のニーズはしばらくあり続けるであろう。日々の製品改善の取り組みをしながら、新技術の活用と時代のニーズに適合するサービス開発に一層取り組んでいきたい。

【畔柳主税（あぜやなぎ・ちから）氏のプロフィル】
静岡県富士市生まれ・東工大卒。石油会社のIT部門から2008年より保険業界向けのITソリューション・DXの企画・営業に携わる。持ち味は企業コラボ。

（毎月第4水曜日掲載）