

マルチモーダルAI技術による業務効率化

◆NECの「映像認識×生成AI」技術がCEATEC 25周年特別賞を受賞

2024年10月、日本最大級のITとエレクトロニクスの展示会「CEATEC 2024」が開催された。出展された技術の中から、本年のCEATEC開催テーマである「Innovation for All」に相応しく、特にAI関連テクノロジーとその活用において優れていると評価される案件として、NECの「映像認識×生成AIによる報告書作成業務の変革」がCEATEC 25周年特別賞を受賞した。本技術は、長時間の映像データに映った人物や車、建物などの変化を映像認識AIによって認識・理解し、NEC開発の生成AIである「cotomi（コトミ）」の大規模言語モデル(LLM)を用いて、ユーザーの目的に応じた短縮動画と説明文章を自動生成する世界初の技術である。本技術をドライブレコーダー（ドラレコ）の動画分析に活用することで、事故発生時の状況や発生に至った経緯などを説明する文章と短縮動画を生成するだけでなく、事故調査報告書をフォーマットに合わせて自動作成する（図1）。保険会社が利用した場合、従来と比較して報告書の作成時間半減を見込む。

従来の映像認識AI技術は、物体認識や人物検知など映像上に枠などをつけて可視化するにとどまり、その映像にどのような価値があるか示唆する技術ではなかった。対して、NECの「映像認識×生成AI」技術では、映像認識AIで映像群を文字化し、それら文字群を生成AIによりユーザーにとって意味のある情報に変換し文章化することで、映像の意味を理解できるようになった。

NECの技術の適用分野はさまざま考えられる。たとえば、交通・運送業界での事故情報データベースの自動作成、放送・配信業界でのコンテンツの自動作成、医療・介護業界での看護記録の自動作成支援や製造・建築業界での作業や施工の記録自動作成支援がある。ただし、業種や映像のジャンルに応じて映像認識AIの作り込みやLLMのファインチューニングは必要になる。

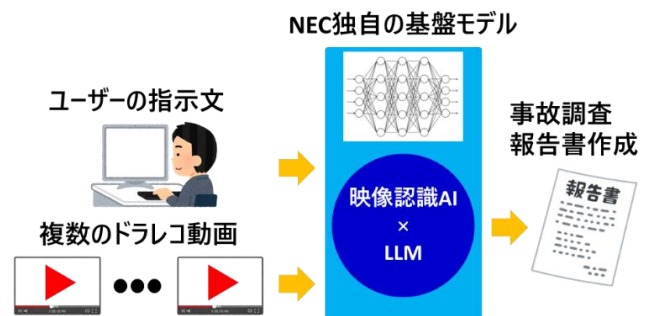


図1 ドライブレコーダー映像から事故調査報告書を自動作成（出所：NECプレスリリースをもとにARC作成）

◆生成AIの登場で画像認識の技術開発が進む

24年7月、パナソニック コネクトは、画像認識の世界最高峰の学会、CVPR2024のコンペで第2位の評価を獲得したと発表した。本コンペでは、3分間の家事の動画の中で人物が何をしようとしているのかをAIが推測して回答する。そのためには、まず、時間の経過と共に変化する動画の中から重要な行動やその目的を認識・判断し、次に、質問文と回答文を解釈する必要がある。従来の画像認識AI技術は、その場その時の事象（空間的情報）を把握することはできても過去からの経緯（時間的情報）を理解することは難しかった。この課題を解決するため、パナソニック コネクトは、画像と言語情報を扱う、複数の生成AI（物体検出の専門家、家事の専門家、専門家の意見を取りまとめる監督役など）にディベートさせることで回答精度を高めたシステムを開発した。

24年10月、リコーは、経済産業省が推進する国内の生成AIの基盤モデルの開発力強化を目的としたプロジェクト「GENIAC」に参加する事業者として採択されたと発表した。企業内文書は、効率的な利用や新たな価値・イノベーションを生み出すための活用が期待されている。しかし、企業内文書はテキストだけではなく、図や表組、画像等も含むため、画像認識を必要とする文書の読解精度が課題となっている。そこで、「GENIAC」の生成AIの基盤モデルの開発に対する計算資源の提供支援を受けて、リコーは、さまざまな企業内文書群を読み取るマルチモーダルLLMの本格的な開発を開始する。なお、マルチモーダルLLMとは、テキスト以外の図・表・画像などの複数の情報（モダリティ）も統合的に扱うことのできる大規模言語モデル（基盤モデル）である。

◆マルチモーダルAI技術で、データの意味や価値の理解が深まる

今後は、映像データだけでなく、音声データやセンサーデータなどを連結（マルチモーダル化）して、LLMと融合したマルチモーダルAI技術の開発が進む。マルチモーダルAI技術によって、さまざまな種類のデータの意味や価値の理解が深まる。たとえば、ドラレコと速度計の記録から自動車事故が起こった原因を詳細に分析したり、作業員視点の映像と手順書から製造業の熟練技能者の「暗黙知」を「形式知」としたりなどである。マルチモーダルAI技術は、自動化やDX化が遅れていた産業領域で業務効率化に不可欠な技術となりそうだ。 【永田紘基】