映像×LLMによる報告書作成業務の自動化

劉 健全　山崎 智史　宮野 博義

要 旨

大規模言語モデル（LLM）の技術発展に伴い、交通や金融、物流、製造、建設、小売り、医療などさまざまな分野でLLMの活用が高く期待されています。これらの分野では、テキストだけでなく、音声や静止画、動画など複数の異種データを取り扱っているため、テキスト以外の複数データにも対応可能なLLM活用に関する技術開発と産業応用が求められています。世の中では、静止画に対応したLLMの開発が急速に進み、豊富な情報が含まれる動画への対応も重要視されつつあります。本稿では、NECにおいて長時間の動画処理を可能にしたLLM活用に関する最新の技術開発と、それを適用した報告書作成業務の自動化などの産業応用、及び今後の取り組みについて紹介します。

KeyWords　LLM／画像認識／映像要約／短縮動画／説明文章／報告書

1. はじめに

2022年11月に米国OpenAI社が発表した「ChatGPT」というAIを契機として、生成AI（アイディアを生み出すAIの分野）が急激に盛り上がりを見せており、現在は人工知能（AI）の「第4次ブーム」[1]に突入したとも言われる状況になっています。ChatGPTは、ユーザーの質問や指示をもとにテキストなどを創る生成AIの一種です[2]。このAIが成り立っているのは、大規模言語モデル（Large Language Model、以下、LLM）という技術のおかげです。

LLMとは、大量の文章データを使って学習した、非常に高度なAIであり、翻訳や文章要約など言語にかかわる処理を可能にする汎用AIを支える巨大な言語モデルです。特に、LLMは数百万から数十億以上のパラメータを保持しており、人間のような高度な知能をうまく表現できます。LLMは2018年頃にBERT[3]とGPT[4]といったモデルの発表より登場し、翻訳やテキスト分類、文章要約、文脈理解などさまざまなタスクで優れた性能を発揮しています。ChatGPTの登場によって、更なるLLMの技術開発が進んでおり、交通や金融、物流、製造、建設、小売、医療などさまざまな業界で活用が期待されています。これらの業界では、テキスト文章だけでなく、音声や画像、動画も扱っているため、さまざまなデータを理解できるモデルの開発とLLM活用が求められています。

特に、デジタル化が進んでいる現代社会では、たくさんの動画データが日々記録されています。これらの記録動画の分析から現場の状況を理解し、説明文章の生成や事故報告書の作成などの業務効率化が求められていますが、記録動画は十分に活用できていないのが現状です。そこで、動画から欲しいシーンを探し出し、説明できるようなLLM活用法が求められています。

しかし、画像向けのLLM開発は進んでいても、豊富な情報が含まれる動画全体を理解するAIの構築は、まだ難しい状況です。そこで、NECはこれまでに培ってきた自社の強みである映像認識のAI技術を生かし、動画解析向けLLM活用法の開発に先陣を切って挑戦しています。

本稿では、NECにおいて長時間の動画処理を可能にしたLLM活用に関する最新の技術開発（第2章）と、それを適用した報告書作成業務の自動化などの産業応用（第3章）、及び今後の展望（第4章）について紹介します。

2. 映像×LLMに関する技術開発

第2章では、NECが開発した映像×LLMに関する最新

の技術[5) 6)]を解説します。この技術（記述的映像要約技術と呼びます）は、映像認識AIとLLMを組み合わせ、長時間の動画から利用者の要望に沿ったシーンのみを抽出し、更にそのシーンを説明する要約文まで自動生成する技術です。本技術は、「ユーザー視点の物語がある映像要約」[7)]というコンセプトから研究開発をはじめ、長い映像を丸ごと見なくても、短時間で知りたい情報だけを効率的に得られるようにする技術です。

本技術の開発では、「ユーザー視点（User-centric）」と「物語がある（Narrative）」という点に着目してシステムの実装を進めました。

「ユーザー視点」の実現に関しては、ユーザーが質問や指示文を明示的に与え、LLMの活用によりユーザーの意図をうまく解釈したうえで、映像内でユーザーが求めたいシーンだけを抽出します。例えば、ドライブレコーダーの動画を考えてみます。保険調査員の視点からすれば、事故原因の究明や被害状況を把握するためのシーンが重要となります。一方、交通警察の視点では、取り締まり対象となる違反行為や安全運転に関するシーンを特に押さえておきたいでしょう。映像からどんな意味を見出すのか、どんな価値を引き出したいのかに対して、画一的な映像分析には限界が生じるため、ユーザー視点を加えることが不可欠です。

また、「物語がある」要約文の生成に関しては、複雑な物事を人が理解しやすくするために、単なる言語的な表現と情報の羅列だけではなく、物事にかかわる因果関係を明確にすることが必要です。これを実現するには、単なるLLM任せのような活用ではなく、たくさんの認識エンジンを映像に適用して認識した人や物、行動、環境、事象などの実在情報に基づいて、映像内で何が起きているか、ストーリーテリングのような形で、分かりやすく示せるように要約文を生成します。

次に本技術を実現した基本フレームワークと、技術の特長、及び初期の検証結果についてそれぞれ紹介します。

2.1 基本フレームワーク

今回NECが開発した技術における処理の流れを**図1**に示します。技術実現の基本フレームは、映像認識AIとデータ検索システム、LLM、及び短縮動画と説明文章の生成モジュールから構成されます。

まず、前処理として複数・長時間の動画を多数の映像認識AIエンジンに入力します。そこで、映像認識AIエンジンは、映像内の人物、車、建物などのさまざまな物体や環境と、それらの変化を個別に認識します。そして、これらの認識結果を時系列的につなぎ合わせ、動画シーンをコンパクトに表現する独自のグラフ構造を作成し、データ検索システム内の映像データベースに蓄積します。

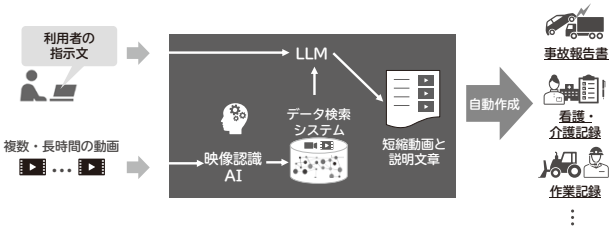


図1　技術実現の基本フレームワークにおける処理の流れ

次に、オンラインの問い合わせ処理を行います。利用者が前処理済みの長時間動画に対して、求めたい情報や問い合わせの条件を自然言語のテキスト（利用者の指示文）でシステムに入力します。システム内部において、LLMが利用者の指示文を意味的に理解し、検索問い合わせ用の複数条件に分解してユーザーの意図や要望を正しくとらえます。

それから、ユーザーの意図や要望を表すテキストとデータ検索システム内のグラフ構造との高速かつ高精度な照合を行い、ユーザーの意図と一致した複数の部分グラフのみを抽出します。そして、部分グラフと紐付いているシーンのみを元の動画から切り出して短縮動画を生成します。同時に、切り出されたシーンを構成する、前処理で映像認識AIエンジンより認識済みの結果（人や物、車、建物など）に基づいて、LLMを用いて短縮動画の物語を説明する要約文章を生成します。

最後に、各業界で定められた報告書や記録書のフォーマットに合わせて、当該部分の画像や動画、文章を適切に埋め込むことで、事故報告書や看護・介護記録書、施工・作業記録書を自動的に作成します。

2.2 技術特長

次に、今回開発した技術が持つ3つの特長を紹介します。

特長1：効率的にシーンを見つけ出し、報告書の作成を迅速化

本技術は、映像認識AIとLLMを組み合わせて、動画の各シーンを理解することが可能です。具体的には、100以上の映像認識AIを活用し、シーンを構成する人物、車、建

物、動物、樹木などの自然物、天気などのさまざまな物体や環境と、それらの変化を個別に認識します。その認識結果だけをLLMで分析することで、動画全体を分析する場合と比較して、利用者が求めるシーンを効率的に見つけ出すことができ、目視による動画の繰り返し確認が不要となります。

特長２：動画の正確な解釈で、専門家と同品質の報告書を生成

本技術は、生成文章の品質を向上させるために、対象分野のサンプル映像を使ってLLMを事前にファインチューニングします。例えば、ドライブレコーダーの動画に適用する場合、事前に道路交通関係の動画を分析します。これによりLLMが専門知識を備え、動画内で起きた出来事を正しく理解できるようになります。その結果、生成AIの正確性の課題であるハルシネーションに対応しながら、信頼性の高い報告書を作成できるようになります。

特長３：大規模なコンピュータを使うことなく、数秒間で報告書を作成

本技術は、1時間以上の動画から、目的のシーンの動画と説明文章を数秒間で作成可能です。そのために、NECが開発したコンパクトで高性能なLLMと、高速なデータ検索システムを活用しました。

前述した基本フレームワークに基づいて、クラウド環境またはオンプレミス環境で実行可能、かつ利用者がWebブラウザからアクセス可能なシステムを開発しました。ドライブレコーダーの動画分析を実演デモとして**図２**に示します。デモ画面では、大型トラックが赤信号を無視して交差点に進入したことにより、黒い一般乗用車と衝突した交通事故が発生した瞬間を表示しています。同時に、本技術が発生した衝突事故について、事故の発生要因を分析してその原因を説明する文章も自動的に生成しました。実験では、同システムを用いて本技術の効果確認を行いました。

2.3 検証結果

本技術の効果を確認するため、ドライブレコーダーの動画から事故調査報告書を作成するユースケースに適用し、検証を行いました。検証の結果、従来は人手で行っていた事故及び事故原因となったシーンの探索や、報告書案の作成を自動化し、報告書作成に掛かる時間を半減できることを確認しました。

3. 映像×LLMに関する産業応用

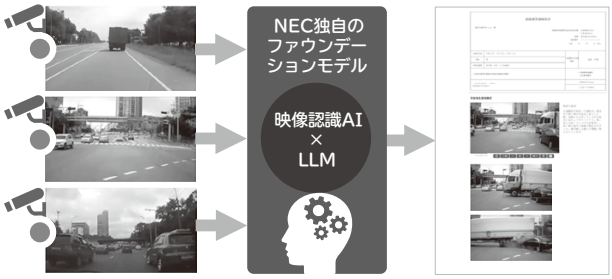
第３章では、本技術（映像×LLM）が適用可能な産業分野とユースケースについて紹介します。

まずは、**図３**に示すように本技術を交通と金融の領域に適用します。本技術をドライブレコーダー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自動で生成可能となります。また、それらをもとに、損害保険金請求や交通安全指導などに向けた事故調査報告書を、フォーマットに合わせて自動作成します。これにより、検証結果にも示したように、従来は手作業で行っていた報告書の作成に掛かる時間を半減できます。

また、交通と金融の領域以外にも、医療や介護、製造、建設、航空、小売など、さまざまな場面で安全管理や業務の効率化を目的とした動画の利活用が進んでいます。しかし、長時間の動画の確認と、作業のヒヤリハットや改善点に関する報告書の作成に、膨大な工数と時間が掛かることが課題となっています。本技術（映像×LLM）の活用により、さまざまな物体や環境から構成され、かつ時間経過とともに変化する複雑なシーンを含む動画から、説明文章

デモ用画像の出典: https://www.youtube.com/watch?v=YBbutvif1W8

図２　本技術の実演デモ画面



デモ用画像の出典: https://www.youtube.com/watch?v=YBbutvif1W8

図３　応用事例：交通事故調査報告書の自動作成

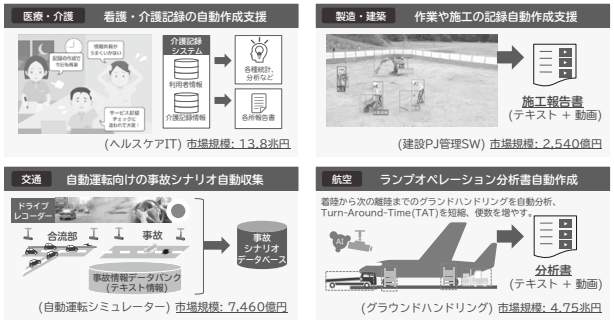


図4　医療や介護、製造・建築などの領域適用イメージ

と報告書の自動作成を支援することが可能です。

これらの産業分野への適用イメージは、**図4**に示します。例えば、工場の製造ラインを撮影しているカメラ映像に適用すれば、完成品検査などの重要なポイントで作業チェックを効率化できます。1日24時間分の映像を確認しなくても、本技術が仕上げた報告書に目を通すだけで確認が可能になります。他にも、看護師や介護士の方などの日報作成や、店舗でのシフト確認、空港でのグラウンドハンドリングなど、映像確認を効率化する技術として幅広く応用できます。また、BtoC分野でも活用可能です。例えば、スポーツの試合映像のなかから特定の選手だけを追いかけたダイジェスト映像の作成も、効率化することができます。

4. むすび

本稿では、技術面と応用面から映像×LLMに関する研究開発の取り組みについて説明しました。全世界的にデジタル化が浸透しつつある現在、映像×LLMの活用は、記録映像の分析から現場の状況を理解し、説明文章の生成や事故報告書の作成などの業務効率化に欠かせない生成AIの産業応用の重要な存在になりつつあります。

今後は、各産業分野での適用に求められるスペック（性能やコスト）に合わせて本技術の改善を継続し、さまざまな場面で確実に導入できる高い技術を実現することを目標として、映像×LLMの活用から広がる効率的な社会の実現を目指し、研究開発に取り組んでいきます。

* ChatGPTは、米国OpenAI社の商標です。
* その他記述された社名、製品名などは、該当する各社の商標または登録商標です。

参考文献

1) 松原 仁ほか：人工知能「第4次ブーム」の胎動 ChatGPTは世界をどう変えるのか，日経ビジネス，2023.6
2) 島津 翔：百家争鳴、勃興する「創るAI」，日経ビジネス，2023.3
3) Jacob Devlin et al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018 https://arxiv.org/abs/1810.04805
4) OpenAI: Improving language understanding with unsupervised learning, 2018.6 https://openai.com/research/language-unsupervised
5) NECプレスリリース：NEC、映像認識AI×LLMにより、動画から説明文章を自動生成する技術を、世界で初めて開発，2023.12 https://jpn.nec.com/press/202312/20231205_01.html
6) NEC研究開発：短縮動画＋説明文で要約 映像認識AI×LLM，2023.12 https://jpn.nec.com/rd/technologies/202314/index.html
7) Yongkang Wong et al.: Compute to Tell the Tale: Goal-Driven Narrative Generation, MM '22: The 30th ACM International Conference on Multimedia, pp.6875-6682, 2022.10 https://doi.org/10.1145/3503161.3549202

執筆者プロフィール

劉 健全
ビジュアルインテリジェンス研究所
ディレクター＆主幹研究員

山崎 智史
ビジュアルインテリジェンス研究所
主任

宮野 博義
ビジュアルインテリジェンス研究所
所長

NEC 技報のご案内

NEC技報の論文をご覧いただきありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報（日本語）　NEC Technical Journal（英語）

Vol.75 No.2　ビジネスの常識を変える生成AI特集
～社会実装に向けた取り組みと、それを支える生成AI技術～

ビジネスの常識を変える生成AI特集によせて
生成AI技術への取り組み ～基盤から応用、ルール作りまで～

◇ 特集論文

急速に広まる生成AIの市場適用
NECにおける生成AIの取り組みについて
電子カルテと医療文書の作成支援による医師業務効率化
映像×LLMによる報告書作成業務の自動化
映像分析と生成AIによるリアル世界の行動理解
サイバー脅威インテリジェンス生成自動化
生成AIの社内活用を進めるNEC Generative AI Service（NGS）
ソフトウェア・システム開発への生成AIの活用
LLMとMIで革新する素材開発プラットフォーム
LLMと画像分析を活用した被災状況の把握

生成AIの可能性を高める基盤技術
日本語性能に優れたNECの大規模言語モデル（LLM）
NECの生成AIを支える国内企業で最大規模のAIスーパーコンピュータ
より安全な大規模言語モデル（LLM）を目指して
データを秘匿したまま連携できる連合学習技術とLLMへの適用可能性
大規模言語モデル（LLM）によるFew-shotクラスタリングの可能性
オープンドメイン常識推論のための知識拡張型プロンプト学習
AI連携とオーケストレーションのための基盤ビジョンLLM
クエリを考慮した新規手法により関連する企業データを減らしLLM APIの使用コストを最適化

AI技術が社会へ浸透するために
AI標準化・ルールメイクに関する動向とNECの取り組み
人権尊重に向けたNECのAIガバナンスの取り組み
RCModelを用いたAIリスクマネジメントのための人材育成事例

◇ NEC Information
2023年度C&C賞表彰式典開催

Vol.75 No.2
（2024年3月）

特集TOP