

プレゼンテーションを自動で行う 3D エージェントの作成についての予備調査

Preliminary Research on the Creation of a 3D Agent for Automated Presentations

海野 琉聖

Ryusei Unno

広島市立大学 言語音声メディア工学研究グループ

Language and Audio Media Engineering Research Group, Hiroshima City University

概要 この世界にはたくさんの文書ファイルが存在する。ユーザーが1つ1つ読み解くことは大変であり、文書を基にプレゼンテーションを作成することは時間のかかることである。また、解説を見る時にはキャラクターが存在することで楽しく学習することができ、分かりやすく飽きにくい体験が可能になると考えられる。

正しい解説のためには、文書から情報を適切に抽出する必要がある。キャラクターについてはテキストによる解説が出来た上での追加要素なので、今回は情報を取り出す方法と、解析する手段について調査・検討を行う。

1 はじめに

現代社会において、良いプレゼンテーションは効率的な情報伝達と意思疎通の役割を持つ。しかしプレゼンテーションの準備には時間と労力がかかるため、自動化することは必要がある。

近年、3D 技術と人工知能 (AI) 分野の大きな進歩により、人間の仕事を機械が代替する可能性が広がっている。特に、文章生成に関しては世界中の企業が取り組む様子から、多様な分野での活動が期待される。

そこで課題になるのは、文書ファイルからの正確な情報抽出の困難さである。PDF については、視覚的なレイアウト情報が記述されており、機械がテキストを正しく読み取ることが難しい。さらに、グラフや図表については人に伝える上では欠かせない要素である。しかし、現在は機械が理解するには難しい分野である。今回は、この情報抽出を中心に調査を進める。

本研究の目標は、文書からの適切な情報抽出と、その情報を利用した自動解説を行う 3D エージェントの作成を行うことである。現時点ではキャラクターの表示および、自動音声生成までを対象としている。

2 背景と技術的課題

2-1. プレゼンテーション自動化の背景

プレゼンテーション作成にはコストがかかる。また、プレゼンテーションが個人の技術に左右されるという問題がある。

2-2. 技術的課題

まず、機械が文書の構造解析を行うことが困難である。PDF は人が読みやすいように構成されているため、機械が読み取ることを考慮していないためである。テキストが複数の行にまたがっている場合、それらが1つの文であることも認識しなければならない。また、文書にはテキストだけでなく、画像、グラフ、表などが混在しており、構造を関連付ける方法も必要である(図 1)。

次にグラフの画像から座標を取得することについて。例えば折れ線グラフでは頂点の座標を取得したいが、頂点を検出する方法が判明していない。また、グラフの領域を検出できるようにしないと、画像のサイズによって異なる数値の頂点の座標を取得しても実用的な座標に変換できないと考えられる。他の棒グラフや円グラフについては別のアプローチが必要。

現在、チャート解析の先行研究として Chart-to-text[1]がある。(図 2)にチャートの例を示す。この研究ではチャート画像と関連する情報から要約文を生成する AI モデルであり、人の評価では答えとなる要約文に対し、60%程度正しく生成出来るという結果が出ている。

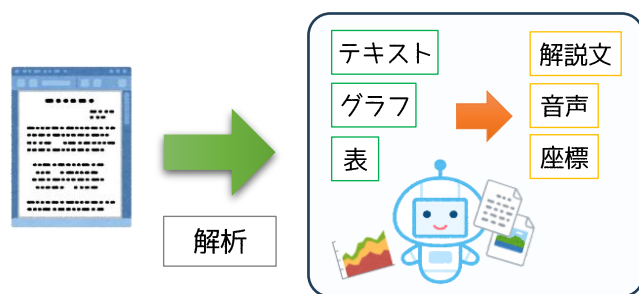


図 1 技術背景

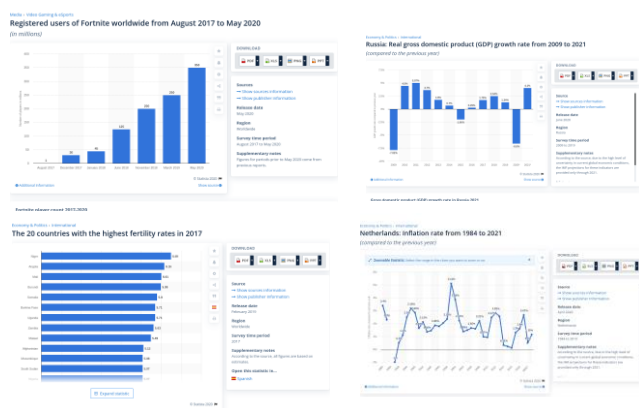


図 2 Chart-to-text の使用画像例

3 注目すること

グラフについてキャラクターがアクションを起こせるようにしたい。例えば、グラフの上を移動させることである。しかし、グラフには様々な種類があり、場合によっては 2 つの折れ線が同時に描画されているなど、数多くの状況に

ついても考慮しなければならない。現時点では、Chart-to-text や OpenCV, PDFMiner などが利用出来ると考えている。Chart-to-text はグラフの解説文生成に、OpenCV は直線検出等の画像解析に、PDFMiner は PDF 文書中のテキストの抽出に使用する。

テキストの領域検出には AI モデルもあり、Layout-Parser[2](図 3)が使えるのではないかと考えた。このツールでは、セクションごとに分かれていると考えられる範囲を別々に認識し、中の文字を読み取ることも出来る。これを使用することで、順番に文書の説明を行うことを可能にし、テキストの抽出も行おう事ができると考える。

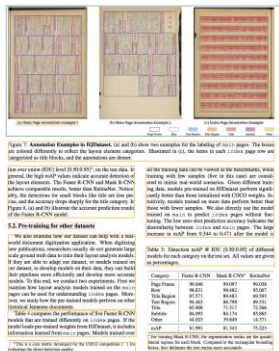


図 3 Layout Parser のサンプル画像

4 今後の方向性

・得られた知見

自然言語処理の状況と、文書の読み取りに関する AI について。最近では BERT や GPT 等の大規模モデルの台頭により、言語生成の精度は非常に向上している。適切な文章を与えることが出来れば、十分な説明の生成も行えるだろう。一方で、文書から自動的に解析することは未だ実験段階であり、グラフ画像についても完成と呼べるツールは存在していない。

プレゼンテーションの必要性和、自動化にあたり考える必要があったことについて。発表資料の作成には、図の用意や聞き手に分かりやすいように設計するなど多くの手間がかかること。自動化によってそのコストが軽減され、キャラクターを取り入れることはエンターテイメント性の向上へと繋がり、ユーザーがより集中できる環境になるという考え。しかし、仮にこのツールが実用的になり誰でも使えるようにする場合は、データの著作権やプライバシーに関して適切に管理しなければならない事も考えなければならない。

・今後の計画

技術を探し、評価する。3D エージェントの設計と動かした方の調査。ユーザーインターフェースの開発、システムの統合とその方法の調査を行う。

特にグラフとキャラクターを連携させることについては、グラフ上をキャラクターに歩かせたり(図 4)、変化が大きいところには反応をさせたりといったアクションの実装を目指す。実現するにあたっては[2-2 技術的課題]で述べたように、グラフを解析する方法を探す必要がある。座標を取得

できれば、値に沿ってキャラクターを動かすことでアニメーションを実現できる。

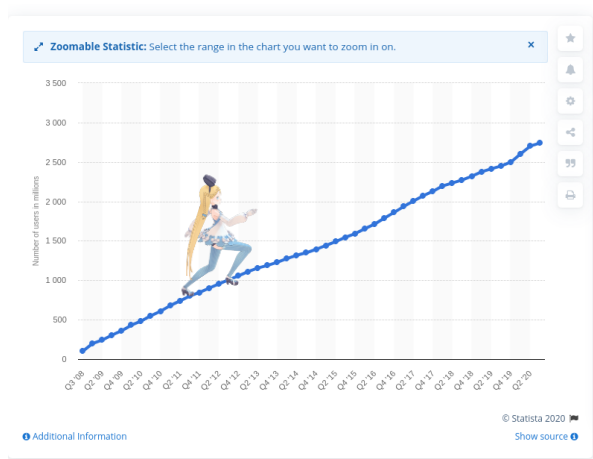


図 4 グラフの上を歩くイメージ

使用モデル「アリシア・ソリッド」[3]

5 まとめ

本報告では、今後行う研究についての調査について述べた。発展途上の分野だが、提供されている物を活用して、形にしたいと考えている。現状、文章生成に限れば GPT-4o に文書を渡す方法もあるが、今回は表やグラフのような視覚データに着目した。

6 参考文献

[1] Shankar Kantharaj, Rixie Tiffany Ko Leong, Xiang Lin, Ahmed Masry, Megh Thakkar, Enamul Hoque, Shafiq Joty, “Chart-to-Text: A Large-Scale Benchmark for Chart Summarization”, ACL 2022 Main Conference, Thu, 14 Apr 2022, <https://arxiv.org/abs/2203.06486v3>

[2] Zejiang Shen, Ruochen Zhang, Melissa Dell, Benjamin Charles Germain Lee, Jacob Carlson, Weining Li, “LayoutParser: A Unified Toolkit for Deep Learning Based Document Image Analysis”, Mon, 21 Jun 2021, <https://arxiv.org/abs/2103.15348>

[3] アリシア・ソリッド, “Alicia Solid”, <https://3d.nicovideo.jp/alicia/>