

人工知能を用いたニュース原稿の自動生成に関する実証実験を実施 ～ディープラーニングによる「AI記者」実現の可能性を検証～

ニュースリリース/NTTデータ

2017年1月27日
株式会社NTTデータ

株式会社NTTデータ（以下：NTTデータ）は、人工知能（以下：AI）を用いて、アナウンサーが読み上げる気象ニュース原稿を気象電文から自動生成する実証実験を2016年9月から4カ月間にわたって実施しました。

近年は、生物の脳構造を参考にして考案された最先端のAIであるディープラーニング技術を画像や音声の自動生成に適用する事例が注目を集めています。しかし、画像や音声と比較して、ビジネスにおいて意味が分かる水準の文書を自動生成することは難しいとされていました。本実証実験では、難易度が高い文書自動生成に挑戦し、自動生成された原稿の品質が実用に耐えうるかの検証を行いました。

本実験では、気象庁が過去に公開した気象電文と、過去にアナウンサーが読んだ気象ニュース原稿をセットにして学習する仕組みを構築し、過去4年分の気象電文から気象ニュース原稿を生成する規則を学習しました。この方法で生成された気象ニュース原稿を評価した結果、日本語の文法は人が読んでも違和感の無いレベルで、意味の正しさにおいては多少の修正が必要なものの、概ね気象電文と同じ内容の文書を作成できることを確認しました。

これまでニュース原稿を自動生成するには、あらかじめ用意されたテンプレート文に、単語や数値を埋め込む方法が主流でしたが、この方法では大量のテンプレート文や単語の埋め込み方を人間が一つ一つ設計する必要があるため、さまざまなパターンに対して網羅的に対応するには限界がありました。一方で、ディープラーニングによる文書の自動生成は多くの人手を必要としないため、設計・開発コストを低減できる可能性があります。

今後、NTTデータでは、AI記者の気象分野における商用化を目指すとともに、企業の決算発表やスポーツ記事等の大量のデータを伴う分野においても新たな実証実験を行い、AI記者の他分野展開を目指していきます。

背景

近年、生物の脳神経回路を参考にしたアルゴリズムであるディープラーニング技術の登場と、コンピュータ演算速度の飛躍的な向上を背景に、さまざまな業種業界においてAIをビジネスに適用する動きが活発化してきています。また、日本においては労働人口が減少し続ける中で、業務の効率化や自動化による生産性の維持・向上が社会的に求められています。

このような社会環境の変化の中で、米国のメディア業界では、2014年よりニュース記事の自動生成技術を商用利用する動きが拡大しています。これらは、あらかじめ用意されたテンプレート文に単語や数値を埋め込むことで記事を自動生成する方式が採用される場合が多いため、予め大量のテンプレート文や単語の埋め込み方を人間が一つ一つ設計する必要があり、複数分野に適用するのは手間がかかる状況でした。

そこで、NTTデータは、人間が文書を執筆する際の規則自体を自動的にデータから獲得できるディープラーニング技術に着目し、メディア業界の事業者と連携して、まずは比較的パターンが複雑でない気象ニュース原稿の自動生成を実現するため、この技術の実証実験に着手しました。

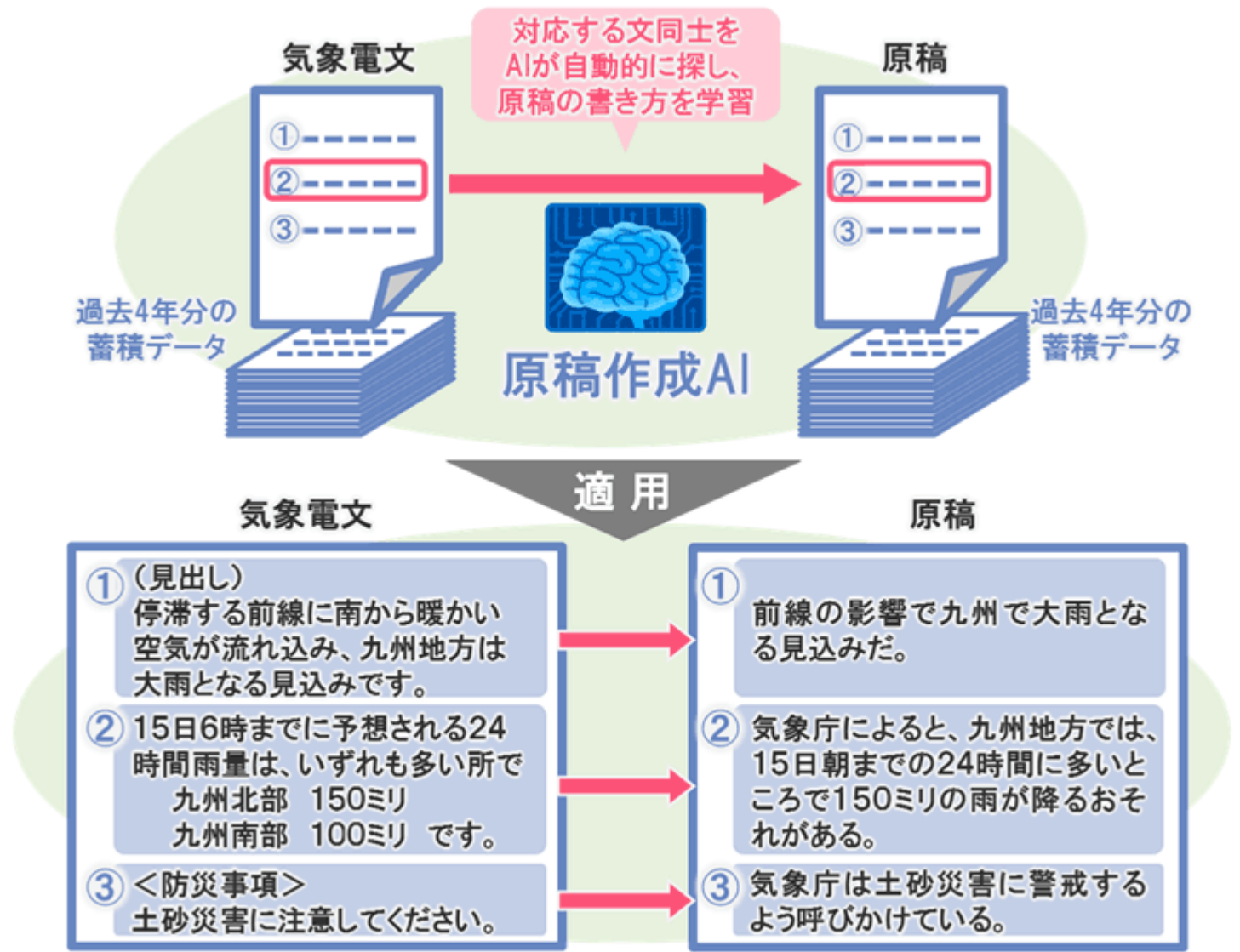
取り組み概要

本実証実験では、まずは気象庁が過去に公開した気象電文とアナウンサーが読んだニュース原稿をそれぞれ4年分用意し、原稿作成の規則性をディープラーニングで学習することで、ニュース原稿を生成するAIを構築しました。このAIに、新たな気象庁の気象電文を読み込ませると、AIが学習した結果を基にして、新たな気象ニュース原稿を自動生成するようになります。

さらに、NTTグループのAI「corevo™」の高精度の日本語解析技術を組み合わせることで、より自然な日本語の生成を実現しています。

次に、自動生成された気象ニュース原稿の「日本語文法の正しさ」と「意味の正しさ」を評価しました。その結果、「日本語文法の正しさ」は、4点満点中3.86点（NTTデータ独自の採点基準）で、人が読んでも違和感が無いレベルに達し、「意味の正しさ」は、4点満点中3.07点（NTTデータ独自の採点基準）で、自動生成された気象ニュース原稿をわずかに修正することで、元の気象電文と矛盾しないレベルに達していることを確認しました。

この結果を踏まえ、今後は、「意味の正しさ」を向上させるほか、気象分野における商用化を目指すとともに、他分野展開のための新たな実証実験を進めます。



図：AIによるニュース原稿自動生成のイメージ

システムの特長とメリット

システム開発コストの低減

ディープラーニング技術を用いることで、開発者が個別の処理方法を設計・開発することなく、大量の教師データ^注から規則性を学習し、原稿を自動生成できるようになります。また、教師データの種類を変更すれば、そのデータに対応して文書を自動生成することが可能です。これらのことから、基となる教師データがあれば、気象以外にも、企業の決算発表やスポーツ記事を書くAI記者を新たに誕生させることも可能となり、開発者が個別の処理方法を設計・開発するコストを低減することが可能です。

原稿作成業務の効率化

AI記者が気象電文からニュース原稿を自動生成するため、人間の記者は、一から原稿を作成する必要が無くなり、原稿作成業務の効率化が可能になります。このことから、人手で行うべき業務により集中することができるようになります。

ニュース配信の速報性向上

AI記者は入力データを基に、記事を高速かつ大量に自動生成することが可能です。これにより、速報性が求められるスポーツニュースや災害情報のリアルタイム配信、地方のニュースの積極的な配信が可能になります。

今後について

今後、NTTデータでは、気象分野におけるニュース原稿自動生成AIの商用化を目指すとともに、企業の決算発表やスポーツ等、大量のデータを伴う分野においても新たな実証実験を行い、AI記者の他分野展開を目指していきます。

注釈

注 「教師データ」とは、入力に対して適切な出力を生成できるようにAIを訓練するためのデータ。

- 「corevo」は日本電信電話株式会社の商標です。
- その他、文中の商品名、会社名、団体名は、各社の商標または登録商標です。

本件に関するお問い合わせ先

報道関係のお問い合わせ先

株式会社NTTデータ
広報部
風間
TEL：03-5546-8051

製品・サービスに関するお問い合わせ先

株式会社NTTデータ
ITサービス・ペイメント事業本部
放送・情報サービス事業部
山内、福井
TEL：050-5546-9061

本実証の技術に関するお問い合わせ先

株式会社NTTデータ

技術革新統括本部

技術開発本部

エボリューションITセンタ

小間、駒井

TEL：050-5546-9741