NTTデータ先端技術株式会社 様

言語理解AIの開発を TMSで加速させる

次世代技術のひとつとして、言語理解AIの開発と活用を進めているNTTデータ先端技術様。人間が書いた文章をコンピュータに理解させる「言語理解AI」を、ビジネスの現場にいかに効率的に導入するか。そのためのツールとして、**Text Mining Studio**（以下、**TMS**）を活用している。

ソフトウェアソリューション事業本部
デジタルソリューション事業部
AIサービス開発担当
家亦 真弘 様

ソフトウェアソリューション事業本部
デジタルソリューション事業部
AIサービス開発担当
城塚 音也 様

Interview

ビジネス文書を読み、内容を理解するAI

御社で開発されている言語理解AIについて教えてください。

城 塚 NTTデータ先端技術では自然言語ソリューション「INTELLILINK バックオフィス NLP」を提供しています。ソリューションが提供する各種の言語理解APIを組み合わせることでオフィスでの非定形作業、知的作業の自動化や高度化が可能です。
その活用事例として、コールセンターでのよくある質問に自動回答するシステムがあり、8割近くの高い正答率を実現しています。また、ビジネスで交わす契約書や公的サービスの申請書の専門的な記載を理解し、潜在するリスクや内容の整合性のチェックにも活用されています。
契約書やFAQ、議事録など、ビジネスでは数多くのテキストデータが使われていますが、我々のチームは最新の自然言語処理技術を活用して、自然言語で書かれたテキストから情報を抽出したり、分類を付与したり、要約を作成したりといった、人間の業務を自動化するソリューションの開発を行っています。

近年、自然言語の解析技術はかなり進化しているそうですね。

城 塚 自然言語を理解するAIは、世界的に30年以上前から研究が行われていますが、なかなか実用レベルの性能が得られませんでした。近年、人間の脳の神経細胞にヒントを得た数理モデルであるニューラルネットワークを活用し、膨大なテキストデータを学習させることで、文脈に応じた単語の意味の違いを精緻に捉えられるようになりました。その結果、例えば機械翻訳ではTOEIC試験で900点を超える性能を達成しています。一方で、精度の高い言語理解を実現するためには、目的のタスクに合った質の高い学習データの用意が欠かせません。

その学習データ作成にTMSをお使いになったのですね。

家 亦 学習には、大きく分けて、教師なし学習、教師あり学習の2つがあります。言語理解AIに行ってほしいタスクに合わせた追加学習では、教師あり学習が有用です。しかしながら、「教師データ」と呼ばれる、正解があらかじめ付与された学習データを大量に準備する必要があります。それぞれのデータにひとつずつ正解を付けていく作業は時間がかかりますし、また人によって正解がばらついてしまうという品質の問題もあるため、なんとかして作業の効率化や品質の向上を図りたいというのが**TMS**利用の発端です。**TMS**で作業を効率化・標準化できれば、学習データ作成の経験が浅いエンジニアでも、一定レベルの学習データ作成が可能になり、AI開発期間の短縮を期待できます。

［裏面に続く］

PROFILE

NTTデータ先端技術株式会社 様

NTTデータ先端技術は、NTTデータグループの技術面を支える中核会社として1999年に設立。基盤・ソフトウエア・セキュリティの3本柱のソリューション事業を通じて、お客様に価値を提供することを目指している。詳細な情報についてはhttps://www.intellilink.co.jp/を参照。

TMSをどのように活用されたのでしょうか。

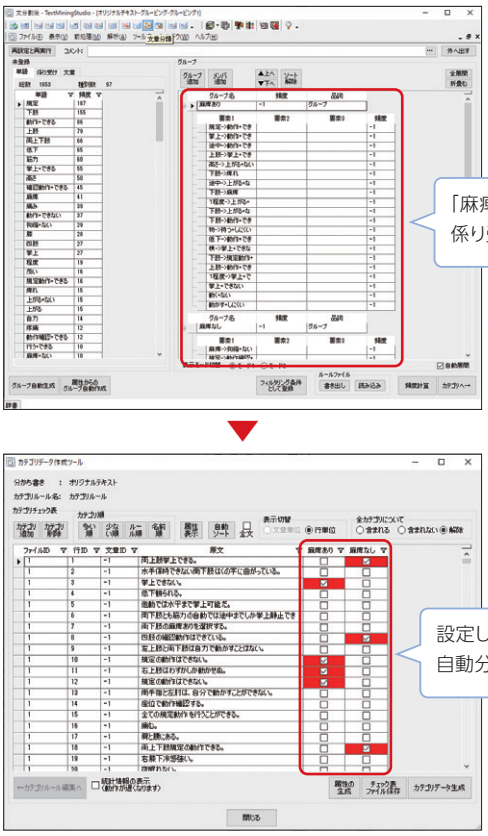
家 亦 介護認定審査の言語理解AIの開発を例にご説明します。介護認定審査では、ケアマネージャーが対象者を調査したレポートを自治体職員が多くの手間をかけてチェック・判定していましたが、AIを使うことでその効率化を実現しました。例えば、調査レポートに「両下肢とも自動では途中までしか挙上静止できない」という記載があった場合、この文章をAIで自動判別し、“麻痺がある”というカテゴリに分類することで、介護認定業務を効率化できます。このような言語理解AIの学習には、上記のような例文に対して“麻痺がある”“麻痺がない”という正解データを付与した「教師データ」を大量に準備しなければなりません。

大量の学習データを準備するのに、人間が文章を読んで正解データを付与するのは非効率的です。また、複数人が教師データを準備すると、各人によって判断基準が異なってしまう恐れもあります。そこで**TMS**を活用することにしました。

TMSは用語の抽出機能が強力で、辞書整備にあまり時間をかけなくても適切な単位で専門用語を抽出できます。また、「何が」「どうした」という係り受け関係を抽出し、カテゴリに自動分類することも可能です。上述の例ですと、「両下肢」と「挙上静止できない」という係り受け関係が文章中にあった場合、“麻痺がある”というカテゴリに分類する、という具合です。

文章では様々なことばが使われているので、各カテゴリに分類する係り受け関係のパターンを多く準備する必要がありますが、**TMS**ではこういった作業をマウスでドラッグ&ドロップするだけで簡単に作成できます。以前は目検による手作業でしたので、新たなことばを発見したときはレポート全体を再度チェックする必要がありました。**TMS**なら再チェックも容易で漏れがありません。

TMSによる学習データ作成例



TMSの高度な解析機能は言語理解AI開発に活用できる

どのようなメリットがありましたか。

城 塚 **TMS**はこの作業で初めて使いましたが、操作方法をチュートリアルで確認することから始めて作業終了まで、1人が1週間弱で終了しています。以前の手作業でしたら、2〜3人で3週間はかかっていたでしょう。チーム内ではExcelでマクロを組んだりもしましたが、それをメンテナンスする手間も大変です。“モチはモチ屋”の言葉通り、プロが開発したツールの方が使いやすいと実感しました。

TMSの魅力は何といっても多くの機能が簡単な操作で実現可能な点にあります。例えば、クラスタリングでk-meansでは思った結果が得られないとき、ソフトクラスタリングで分析し直すといったこともマウスひとつで簡単にできます。実は私もテキストマイニングツールの黎明期にテキストマイニングツールを開発し、商品化したことがあるのですが、その経験からも**TMS**は良くできたツールだと感心しています。

今後の展開についてお聞かせください。

城 塚 **TMS**は、言語理解AI構築の初期段階に膨大なテキストの傾向を把握する際のツールとして有用です。我々はGoogleの言語理解モデルBERTなどを言語理解AIに活用していますが、大量のテキストの前処理を**TMS**で行った上でBERTを使って高度な分析をするといった、組み合わせによる使い方が効率的だと考えています。

言語理解AIを現場適用する際の最も大きな課題のひとつが学習データの準備です。ともすれば学習データ準備に半年以上もかかるときがあり、システム導入の大きなネックとなっていました。半教師あり学習や能動学習といった必要な正解データの量を削減する手法もありますが、やはり相当な数の正解データを人手で作成することは避けて通れません。それが、**TMS**を使えば品質の高い学習データが短期間に作成できるようになります。我々は**TMS**を言語理解AIの普及を加速させるキーツールのひとつとして期待しています。

NTT DATA 株式会社NTTデータ数理システム

営業担当

〒160-0016 東京都新宿区信濃町35番地 信濃町煉瓦館1階　TEL 03-3358-6681（お問い合わせ：平日10:00-16:00）

〈e-mail〉sales@ml.msi.co.jp　〈URL〉https://www.msi.co.jp/solution/analytics/index.html

096-202307