

自然言語処理技術の産業への展開

Social Implementation of Natural Language Processing Technology



豊橋技術科学大学 IT 活用教育センター特任教授

井佐原 均

通商産業省工業技術院電子技術総合研究所、郵政省通信総合研究所、独立行政法人情報通信研究機構を経て、現職。産業日本語研究会世話人会代表。

1 はじめに

新型コロナウイルスの感染拡大に伴い、筆者が世話人会代表を務める産業日本語研究会が令和2年3月6日に開催を予定していた「第11回産業日本語研究会・シンポジウム」は中止となった。このシンポジウムのテーマは「AIが支える産業日本語」であった。コロナ禍の中、人と人の対面が制約される今、人とコンピュータ（AI）の対話、あるいは人と人とのコミュニケーションを支援するコンピュータ（AI）の必要性が高まっている。AI技術が基盤となって実社会で役に立つ、産業として成り立つサービスが求められている。

本稿では、このようなサービスを開発するプロジェクトの例として、筆者が研究リーダーを務める、自然言語処理技術を用いたシステムの社会実装を目指すプロジェクトについて概説する。

2 知の拠点あいち重点研究プロジェクト

「知の拠点あいち重点研究プロジェクト」は愛知県が公益財団法人科学技術交流財団に委託し実施しており、大学等の研究シーズを活用して県内主要産業が有する課題を解決し、新技術の開発・実用化や新産業の創出を促進する産学行政連携の研究開発プロジェクトである。現在は第3期が令和元年度からの3年間で進行中である。プロジェクトの全容は、<http://www.astf-kha.jp/project/>を参照されたい。第3期のプロジェクトには以下の3つのサブプロジェクトがある。

- ① 近未来自動車技術開発プロジェクト
- ② 先進的 AI・IoT・ビッグデータ活用技術開発プロジェクト
- ③ 革新的モノづくり技術開発プロジェクト

先進的 AI・IoT・ビッグデータ活用技術開発プロジェクトは、さらに9つのプロジェクトに分かれる。本稿で説明する筆者らの「分野適応技術による自然言語処理技術のビジネス展開」は、その一つである。

3 分野適応技術による自然言語処理技術のビジネス展開

このプロジェクトでは、多目的に利用可能な知識抽出 AI システムを基盤として、多言語情報発信（機械翻訳）システムおよび人間的な対話を実現するシステムを開発する。これにより共通の基盤を利用した多種の応用システムを効率的に実現できる。

このプロジェクトでは、以下の3つのテーマで研究を行っている。

- ① 知識抽出 AI システム
- ② 分野適応型機械翻訳システム
- ③ 寄り添い型対話システム

開発するシステムとその応用について、図1に概要を示す。

本プロジェクトの参画機関は豊橋技術科学大学のほか、以下の7機関である。（順不同）

シンフォニアテクノロジー株式会社
 武蔵精密工業株式会社
 新東工業株式会社



図1 システムとその応用

株式会社ウォンツ
 日本マイクロソフト株式会社
 社会福祉法人さわらび会・さわらびグループ
 グローバルデザイン株式会社

以下、各テーマについて概説する。

4 知識抽出 AI システム

専門用語を自動的に抽出し、対訳辞書や用語辞書を作成する機能を開発する。これにより、用語集やオントロジーといった人工知能を用いた自然言語処理システムに必須の要素を自社では開発できないような企業でも人工知能技術を利用可能にする。

自然言語処理などの人工知能システムの性能は大きく向上したが、ウェブ文書の処理や簡単な対話では実用レベルに達していても、ビジネスでの利用には課題が残る。個別の対象にチューニングすることにより、対象分野での精度を向上する必要がある。このためには分野知識としての辞書の作成が必要となり、専門用語の抽出や意味のベクトル表現化の研究がおこなわれている。本テーマでは単語の抽出や、単語間の意味的距離の判定だけでなく、分野の知識を表現するような典型的な語句の抽出（用語抽出）や、語の意味の分類（意味抽出）に関する研究開発を行う。得られた成果を本プロジェクトの出口である他の二つのテーマに適用する。

人工知能システムの対象別チューニングに必要な用語抽出と意味抽出の2つの課題に対し、長い語句の抽出技術と大規模データを用いた意味分類技術の2つの既存技術を発展させて、参画企業の実文書に対して適応可

能なシステムを実現する。

意味抽出システムについては、マニュアルや学術論文を対象に、アルゴリズムの改良（機械学習の活用）、ヒューリスティックスの利用により、性能向上を図っている。

知識抽出システムはウェブ上のテキストや小説を学習データとし、得られたモデルを機械翻訳システムの評価に適用し、モデルの性能を評価している。

5 分野適応型機械翻訳システム

近年、深層学習によって自然言語処理技術は急速に発展したが、機械翻訳システムにおいては、流暢度は高いが、ビジネスで使うには専門用語の誤翻訳など、忠実度に難がある。文の前処理、専門用語辞書の利用、分野のテキストを用いた学習など、個別企業のニーズへのチューニング技術を開発することにより、個々の業界において実ビジネスで利用可能な高性能機械翻訳システムを容易に実現する手法を確立する。

ここでは、既存の機械翻訳エンジンをベースに、分野にチューニングする技術を確立し、各企業の実ニーズに適合した個別の機械翻訳システムを作成する。社内文書や用語集、関連分野の特許対訳を利用し、翻訳システムの学習用データとして利用可能な形式としたデータベースを作成する。構築したデータベースを活用したシステムを試用し、翻訳精度や翻訳プロセスの効率を評価する。個々の企業の文書で出現した課題を抽出し、分野適応技術およびデータベースの改良を行う。これらの成果の他の企業への導入を容易にするため、対応法を定式化する。個別企業へのチューニング技術を実用化するとともに、他企業への提供（販売）の可能性を検討するため、必要

なコストを明確化する。

研究開発に利用する言語資源としては、学習用の対訳データは対象分野の特許文書 10～20 万文対程度を利用する。対象分野の用語集として事前に参画企業で作成されていたものと、翻訳評価実験の中で指摘され、収集された用語セットを利用する。

機械翻訳システムのビジネス利用での大きな課題は、各分野に固有の専門用語等の低頻度語の処理であるが、本プロジェクトにおいては、平成 30 年度に東京都港区と本プロジェクトの参画機関である豊橋技術科学大学、日本マイクロソフト株式会社、グローバルデザイン株式会社 が共同で行った「AI を活用した行政向け自動翻訳サービスの実証実験」において、検証した手法をさらに改良する。

<https://www.city.minato.tokyo.jp/kouhou/aihonyaku.html>

6 寄り添い型対話システム

先に述べたように、人とコンピュータ (AI) の対話、あるいは人と人とのコミュニケーションを支援するコンピュータ (AI) の必要性が高まっている。ここでは対話コミュニケーションを情報の要求と提供の過程ととらえ、様々な応用に対して、同一の処理を適用することを考える。

近未来社会では様々な場面でコンピュータ (AI) と人間とのコミュニケーションが起こる。人と対話しているかのように感じるコミュニケーションが求められるが、目標に向かって、正確に効率よく対話するだけでは敷衍無礼で「人間的でない」と感じる。

このテーマでは、人間と機械の対話モデルを構築し、その実装先として2つのターゲットを定める。一つは介護場面での介護ロボットと被介護者との対話である。知の拠点あいち重点研究プロジェクト第2期で開発された介護ロボットに対話機能を持たせる。また、他の形態のロボットへの適用も目指す。もう一つは機器等の使用法を容易に獲得するなど、ユーザが必要とする情報を対話によって入手するシステムの開発である。サービスの形態としては、チャットボットや FAQ 検索などが考えられる。

6.1 介護ロボット

実在する介護ロボットへの対話システムの実装と評価を通して、質問応答や情報検索といったコンピュータに向けての一方のコミュニケーションから、人間とコンピュータが相互の関係性の中で適切に判断し、行動する双方向のコミュニケーションへの展開を目指す。

対話システムは、音声認識⇒対話モデルの構築⇒応答テキストの生成⇒音声合成といった形態を取る。介護ロボットの実機上に対話システムを実装する。ユーザが機械に対して感じる好感度 (好判定の割合) を人間的だと感じるコミュニケーションの尺度とし、好感度向上を実現する上で重要となる対話モデルの構築を行い、実機での実用化を行う。

実際に介護施設での介護者と被介護者との対話データを収集する。音声データを書き起こしし、介護場面で必要な対話セットを集めた対話用データセットを作成する。また介護施設での実際のニーズを聞き取り調査し、システム開発への知見を得る。介護ロボットがどのような対話を行えば、被介護者の満足度が高まるかを検討する。

6.2 情報提供支援

近年、企業や自治体のウェブページには質問に答えるチャットボットが付いているものが増えてきた。ここで開発するサービスは基本的にはそのようなチャットボットである。柔軟な日本語文での質問を可能とし、ユーザの視点からの正解率の向上を目指す。そのためにはあいまいな入力に対して、適切な情報 (適切な FAQ 項目など) を得ることが必要となる。現在、この精度向上を進めている。

この手法は FAQ をデータとするチャットボットに利用できるだけでなく、入力文に対して適切な情報を検索する「あいまい検索」にも有効となる。応用の観点から見れば、同じ手法を同じデータに対して適用し、結果の提示法を必要に応じて変更することも考えられる。たとえば、利用者とヘルプデスクの担当者が同じデータと検索エンジンを用いるが、利用者はチャットボットとして対話的に利用し、担当者には一度に多くの情報を提示する、といったことが可能となる。

7 社会実装の貢献

自然言語処理をはじめとする人工知能技術は様々な産業の基盤となる技術であり、参画企業はこの技術のユーザとなるとともに、この技術を組み込んだサービスの提供者となりうる。具体化できる産業貢献は以下のようなものがあげられる。

7.1 産業の国際競争力の強化への貢献

国際競争力強化には産業文書（ノウハウ文書、取扱説明書など）の多言語化が必須である。産業に関する情報を表現する日本語をこれまで以上に、迅速に、正確に、低コストで翻訳することが可能になる。

多言語化を効率的に行うことが可能になれば、企業の海外展開（海外拠点）の支援、製品の輸出、知財貿易などに関する競争力が強化され、波及効果は膨大である。

低コストで関連文書を翻訳できれば、より多くの製品（サービス）の情報が多言語化され、より多くのビジネスチャンス（時間・市場）を得ることが可能になる。

7.2 付加価値の追加による競争力の強化

介護ロボットを例にとると以下ようになる。

介護者の指示を的確に把握し、被介護者の心に寄り添った対話を行うことができる介護ロボットを実現する。この技術を得てロボットに実装した参画企業の販売競争力が強化される。

近隣の病院や施設では、このロボットをいち早く導入することが可能となり、それら施設の対外的競争力を強化する。

いろいろな機器に音声での対話能力（音声チャットボット）が付加されることにより、個人レベルのユーザも利用可能になる。また、ハンズフリーが要求される分野にも適用可能な機器となる。

8 人材育成

AIを用いたサービスの社会実装が広まることが求められている。しかしながら、大学等のAI研究者だけではなく、モノづくりに関わる企業人が自分自身でAI技術を使いこなすことが出来なくては、真の活用は行えない。特にAI技術の基盤をなすデータサイエンス技術は

様々な分野のモノづくりにおいて、活用が期待される。豊橋技術科学大学は、文部科学省の「数理及びデータサイエンスに係る教育強化」の協力校に選定されており、「データ科学のものづくり技術への定着を促進する包括的教育環境の整備」事業として、データサイエンスを学ぶ自習・演習型のe-Learning教材を開発している。知の拠点プロジェクトの参画企業や近隣大学には既に教材の提供を行っており、人材育成のための活用が期待される。

また、豊橋技術科学大学の卒業生がプロジェクトの参画企業に就職し、このプロジェクトに参加するという例もあり、大学と企業が連携したOJT的な人材育成も行っている。

9 おわりに

情報技術は実際に社会で利用されなければ、その価値や効果は限定される。社会実装を目標としたプロジェクトの例として、筆者の行っている「知の拠点あいち重点研究プロジェクト」のプロジェクトについて概説した。

このプロジェクトはスタートから1年が経過し、2021年度末の終了に向けて、開発と実証を進めている。その進捗について、改めて報告する機会を持てればと思う。