



自然言語処理のビジネス活用を革新する エンタープライズ向け NLP

「ドメイン特化 NTT-BERT と
構築フレームワーク」

C O N T E N T S

01 CHAPTER.1 はじめに

02 CHAPTER.2 NTTグループの 自然言語処理ノウハウ

03 CHAPTER.3 BERTの登場

04 CHAPTER.4 ドメイン特化 NTT-BERT の実力

05 CHAPTER.5 自然言語処理を 手軽にビジネスで使える時代へ

CHAPTER.1 はじめに

自然言語処理技術の現在の技術水準は汎用精度80%程度が限界と言われていています。この精度において、アンケートの解析/議事録の自動生成/多言語の機械翻訳、といったユースケースを想定した場合、完璧な結果は得られないため、結局人間によるチェックが発生してしまう状況です。現時点で役立つ用途としては、以下のような場面が挙げられます。

①膨大な文章量に対する検索・検知などの人間で

は処理できないタスクを実施

②ある程度間違っても機能上大きな問題にならないことを実施

③分野と用途を限定し個別アレンジすることで高い精度を実現

しかし、決定的な効果を出しにくい点や、高精度を維持するための構築運用コストが非常に高くなってしまふ点が課題となっています。

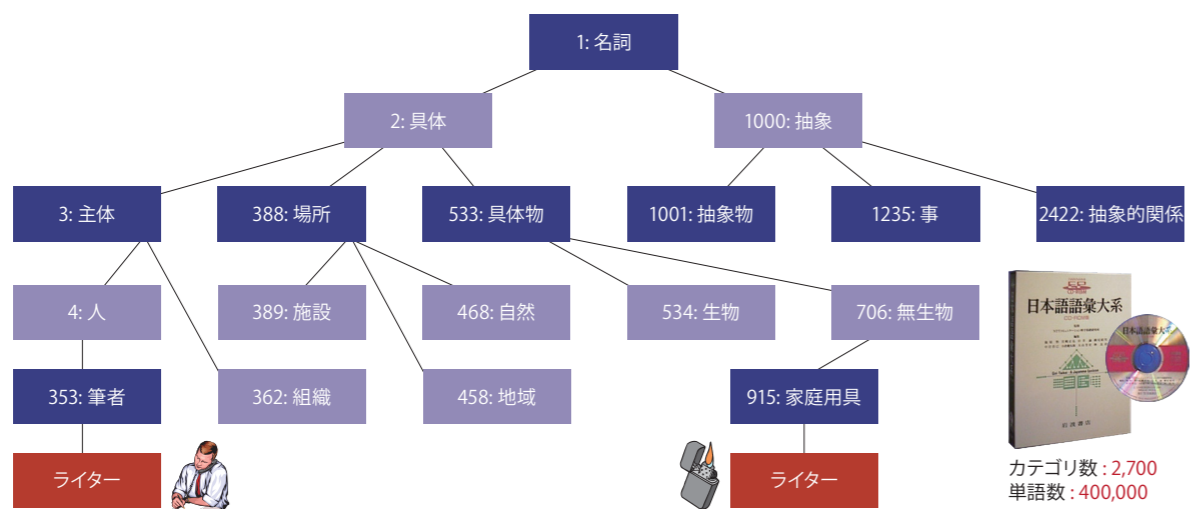
CHAPTER.2 NTTグループの 自然言語処理ノウハウ

NTTグループでは、日本語を対象とした自然言語処理技術の研究を40年以上も続けてきました。世界最大級の日本語辞書を保有し、世界最高水準の意味解析技術や文法に基づく語順変換技術を有しています。また、音声認識・音声感情認識・応答対話の技術も幅広くカバーしています。これらの技術/ノウハウは、技術文書等の長文に対して

も高品質な翻訳を可能とする統計的機械翻訳プラットフォームや、テキスト解析/対話エンジンに利用可能な音声解析処理APIとして、世の中に還元しています。NTTデータでは、NTT研究所の持つ自然言語処理技術/ノウハウをいち早く取り入れ、お客様業務を対象としたPoC実施や実プロジェクトへの適用を進めてきました。



Figure.1 NTT 研究所が持つ世界最大級の日本語辞書と、世界最高水準の言語解析技術



文法に基づく構文構造解析と、述語の係り受け関係に基づく意味構造解析を併用し、日本語文章の意味を正しく理解。

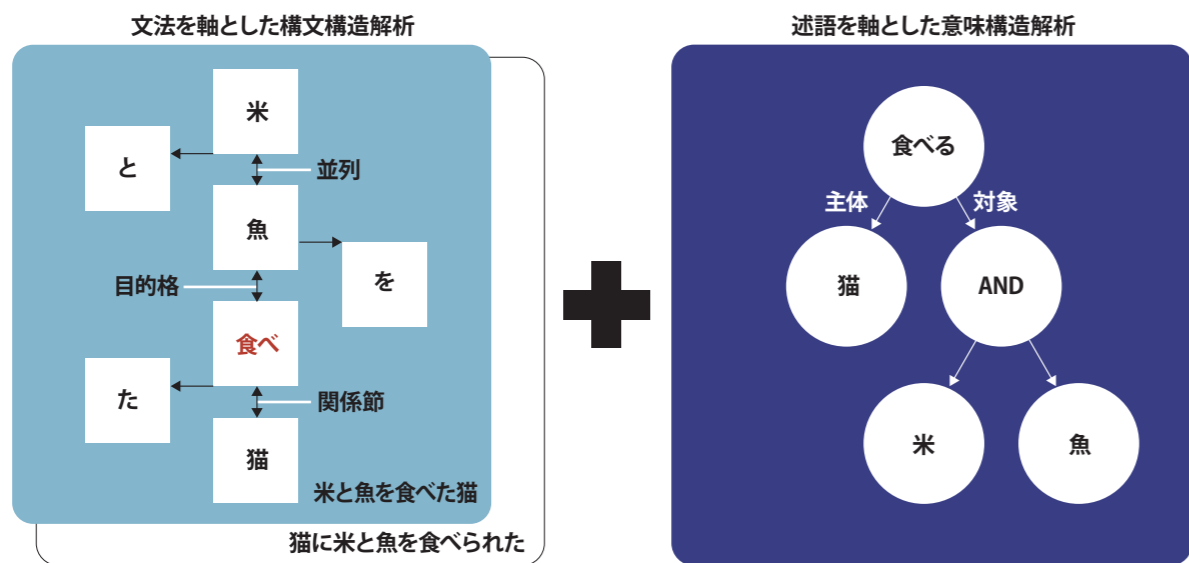
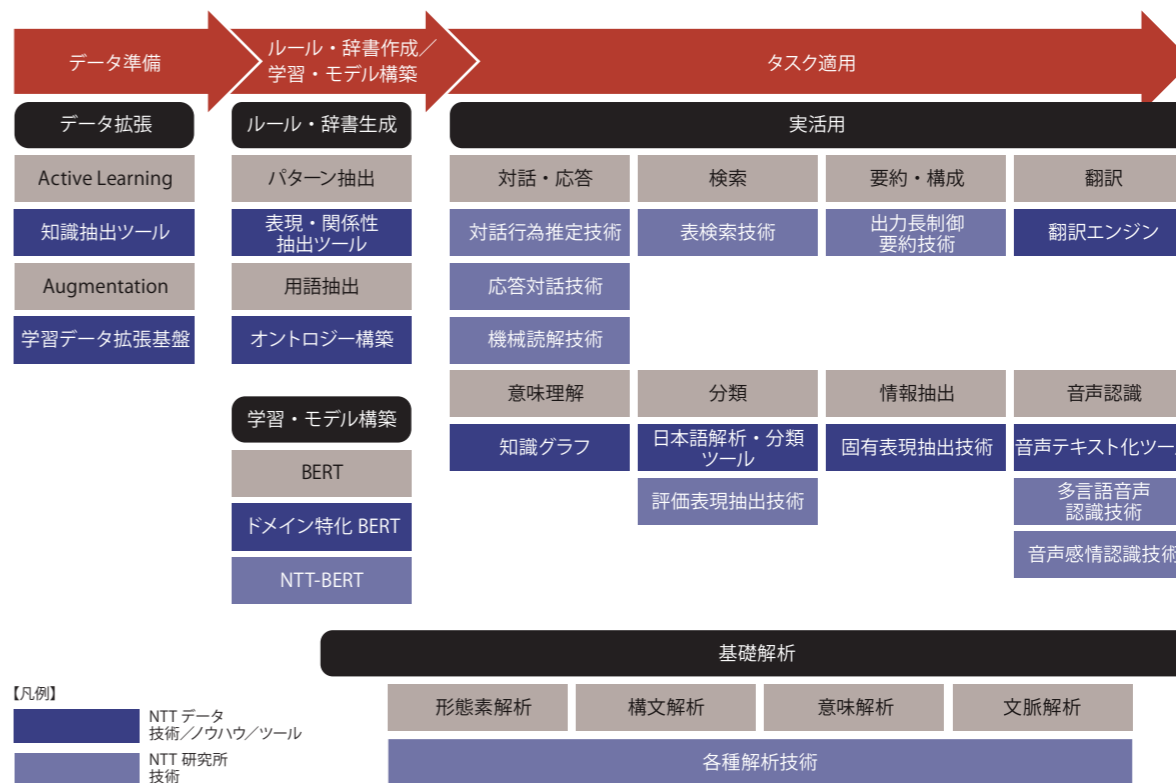


Figure.2 NTT グループの自然言語処理技術マップ



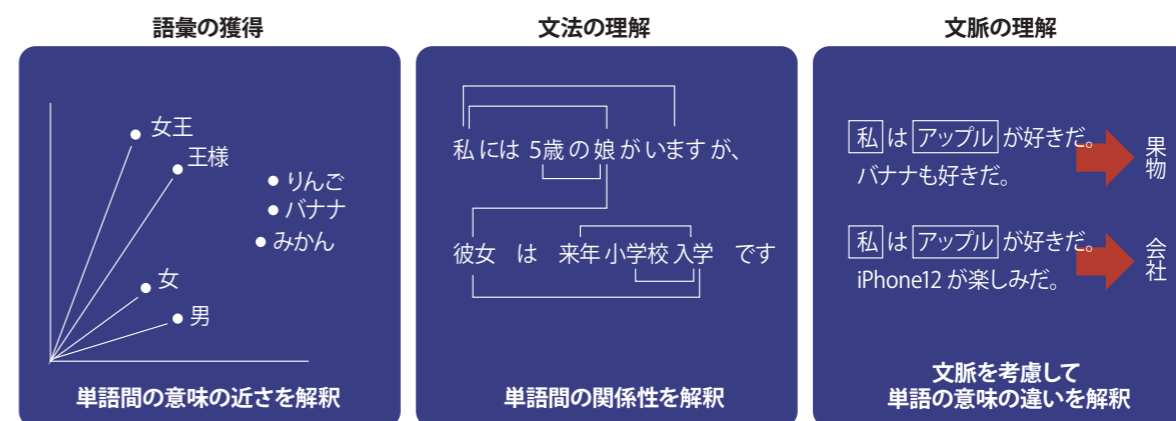
CHAPTER.3 BERT の登場

BERT (Bidirectional Encoder Representations from Transformers) は、2018 年 10 月に Google が発表した「自然言語処理向けの事前学習手法&汎用言語モデル」です。大規模データの教師なし

学習により高性能を実現する方法で、自然言語処理分野の様々なベンチマークにおいて従来モデルの精度を上回る結果を出しています。NTT でもこの技術を取り入れ、これまで蓄積され

Figure.3 BERT ができること

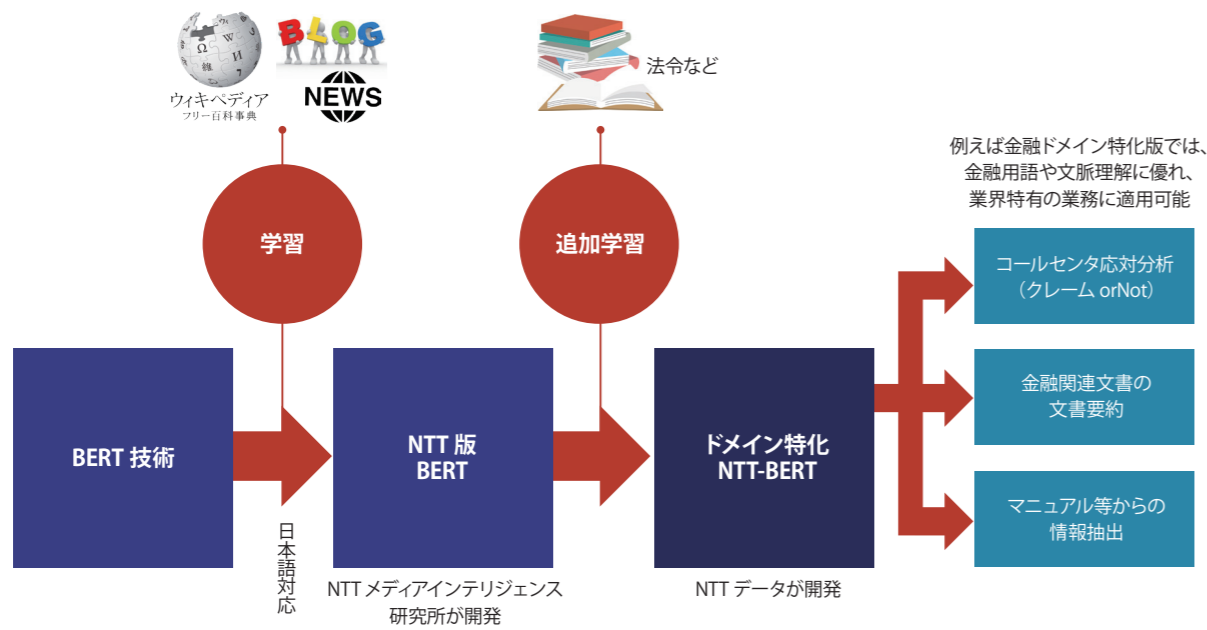
インプットとなるテキストを与えると、機械が自動的に 語彙、文法、文脈理解に相当する知識を学習して、モデルを生成



たノウハウを活用し日本語に対応した NTT 版 BERT を用意しました。さらに NTT データでは、NTT 版 BERT を元に「ドメイン特化 NTT-BERT と

構築フレームワーク」を開発することで、価値ある精度を達成しながら構築運用のコストを抑えられるエンタープライズ向け NLP を実現しました。

Figure.4 ドメイン特化 NTT-BERT



CHAPTER.4 ドメイン特化 NTT-BERT の実力

①価値ある精度

金融業界向けに開発した「金融ドメイン特化 NTT-BERT」は、融資関連文書や銀行業務に関する文書を追加学習のインプットに使用しました。精度を評価するにあたり、実際の金融文書に登場する文章の穴埋め問題や、業界で用いられている資格試験の問題を解かせたところ、従来の BERT よりも高い精度を確認できています。金融文書内での専門用語や言い回しを正しく解釈することで、金融機関の業務担当者のように文脈に沿った処理が可能になります。具体的なユースケースとしては、チャットボットによる問い合わせ対応／FAQ の回答自動引き当て／財務情報からのリスク抽出／稟議書の記載内容チェック／日報からの情報抽出等

の使い道が考えられます。

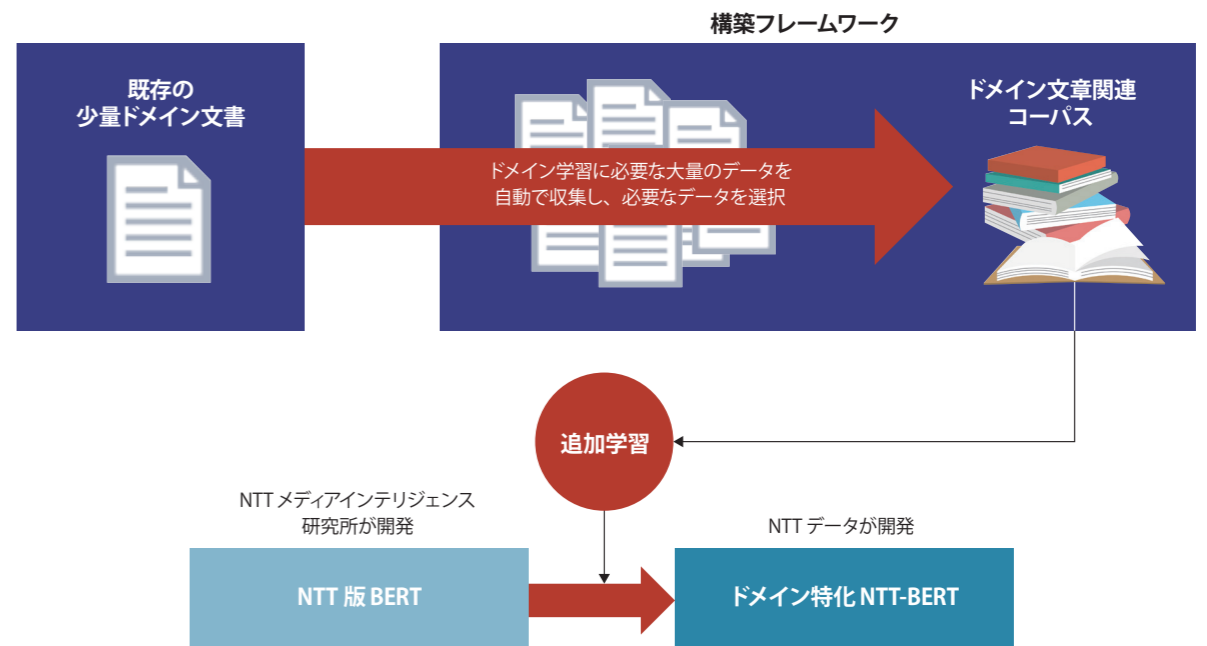
②重荷にならない構築運用

従来の自然言語処理をエンタープライズに活用する場合、運用コストが非常に高くなってしまいう課題がありました。モデル構築時に用途に特化した文書を大量に準備し学習させる必要があり、その後も定期的に変更やアップデートの際に同様の対応が必要になるからです。ドメイン特化 NTT-BERT では、構築フレームワークを用意することでこの課題に対処します。既存の少量ドキュメントを下準備なく読み込ませるだけで、ドメイン学習に必要な大量のデータを自動で収集し、必要なデータを選択して学習できるため、効率的にモデルを構築・アップデートしていくことが可能です。

Figure.5 ドメイン特化 NTT-BERT の性能検証結果



Figure.6 構築フレームワークが追加学習を自動化



CHAPTER.5 自然言語処理を 手軽にビジネスで使える時代へ

NTT データでは、ドメイン特化 NTT-BERT と構築フレームワークを活用し、様々なドメインでお客様務

への NLP 適用を推進しています。具体的な適用先ドメインとユースケースの実績例を図7に示します。

Figure.7 エンタープライズ向け NLP の適用実績例



各業界における書類審査／日報や記録からの情報抽出／問合せ応答など、意味を理解した AI による適切な自動処理の幅が広がることで、社会全体

の高度化に寄与できると信じています。自然言語処理を気軽にビジネスで導入できる時代はすぐそこまで来ています。