

顧客ニーズに対応した音声認識用辞書の カスタマイズを自動化する取り組み

Efforts for Development of Technology to Automatically Customize Voice Recognition Dictionaries to Meet Customer Requirements

長 健太 CHO Kenta 阿部 一彦 ABE Kazuhiko

B2B (Business to Business) 向けの音声認識サービスでは、顧客ごとに異なる話題の音声を適切に認識できるように、音声認識用辞書をカスタマイズする必要がある。しかし従来の技術では、カスタマイズには音声認識の専門家による作業が必要であり、数多くの作業を短時間で行うことが困難であった。

東芝デジタルソリューションズ(株)は、知識メディア処理技術を融合し体系化したサービスである東芝コミュニケーションAI “RECAIUS”を提供している。顧客に合わせた音声認識サービスを提供するため、RECAIUS向けに、音声認識用辞書をカスタマイズする自動化技術の開発に取り組んでいる。カスタマイズ作業の効率化を支援するプラットフォームを新規に開発することで、人手による作業時間を従来に比べ約40%削減することができ、多数のカスタマイズ作業を短時間で実施できるようになった。

In the business-to-business (B2B) market, attention has become increasingly focused in recent years on voice recognition. In response to the market demand for various types of voice data meeting the specific requirements of individual customers, the customization of voice recognition dictionaries is necessary. However, the need for experts on voice recognition and reduction of the time required for such complicated customization work are significant issues in this field.

Toshiba Digital Solutions Corporation has been supplying “Toshiba Communication AI RECAIUS” services, which combine various media intelligence technologies including voice recognition, text-to-speech, and intelligent dialogue and image recognition. In these services, we are making efforts to develop a new technology to automatically customize voice recognition dictionaries optimized for each customer. By applying a newly developed platform to efficiently implement the customization processes for voice recognition dictionaries, this technology makes it possible to reduce the time required for customization by approximately 40%, significantly shortening the period required for customizing multiple voice recognition dictionaries.

1. まえがき

東芝デジタルソリューションズ(株)は、様々な知識メディア処理技術を適用したサービスとして、東芝コミュニケーションAI RECAIUSを提供している。人の話した内容をテキスト化する音声認識もそれらの技術の一つであり、コールセンターにおける発話内容のテキスト化を通じたレポート作成や、スマートテレビやカーナビなどの電子機器の自由発話による操作などに利用されている。

音声認識エンジンは、音声認識用辞書を参照して動作する(図1)。音声認識用辞書は、音声信号から音素情報に変換するための音響モデルや、それら音素の列から言語的な特徴を加味してテキストを生成するための言語的なルールや統計情報から成る言語モデル⁽¹⁾などを、音声認識エンジンで使えるデータ形式にしてまとめたものである。

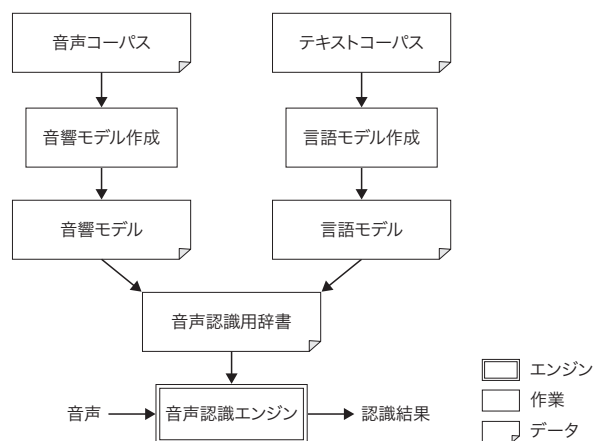


図1. 音声認識の基本動作

音声認識エンジンは、音響モデルや言語モデルなどから作成した音声認識用辞書と音声特徴を照合し、認識結果を出力する。

Basic mechanism of voice recognition

これらのモデルは、音声コーパスやテキストコーパスから作成される。音声は話者や収録環境、テキストは話題などによって変化することから、音声認識の精度を向上させるためには、これらのコーパスを音声認識対象に応じて収集し、適切なモデルを作成して音声認識用辞書に反映する必要がある。

RECAIUSでは、音声認識用辞書のカスタマイズサービスを行っている。顧客から提供された音声データを用いてカスタマイズを行い、クラウドシステム上などで利用可能な音声認識サービスに組み込んで提供される。

この論文では、当社のRECAIUSによる音声認識サービスを一層有用なものにするために、これまで音声認識の専門家が行っていた音声認識用辞書のカスタマイズ作業を、自動化する取り組みについて述べる。

2. 音声認識用辞書のカスタマイズ作業の現状

音声認識用辞書のカスタマイズ作業の一例として、音声データから言語モデルを作成して音声認識用辞書に登録するまでの作業フローを、図2に示す。

- (1) 音声データの受領・確認 顧客から提供された音声データを確認する。
- (2) 書き起こし・確認 音声からテキストを書き起こし、確認してテキストコーパスとする。
- (3) 言語モデル作成 テキストコーパスの一部、及びあらかじめ作成した汎用的な認識が可能なベース言語モデルを用いて、顧客ごとに言語モデルを作成する。
- (4) 音声認識用辞書作成 言語モデルを音声認識エンジンで使用可能なデータ形式に変換し、音声認識用辞書に登録する。
- (5) 評価 音声及びテキストコーパスの一部を評価データとして用いることで、作成した音声認識用辞書の評価を実施し、カスタマイズ前の辞書との精度比較などを行う。

この作業フローの中では、様々なツールを組み合わせる多様なデータを生成する必要がある。また、それらのデータの管理やツールへの適切なデータの設定などを行うために、音声認識エンジン特有のデータやツールに精通した専門家による作業が必要になる。専門家の数は限られているため、全ての作業で専門家による作業が必要であると、多数のカスタマイズ作業を行うことは困難である。

カスタマイズの作業フローで専門家が必要とされる要因は、主に以下の三つである。

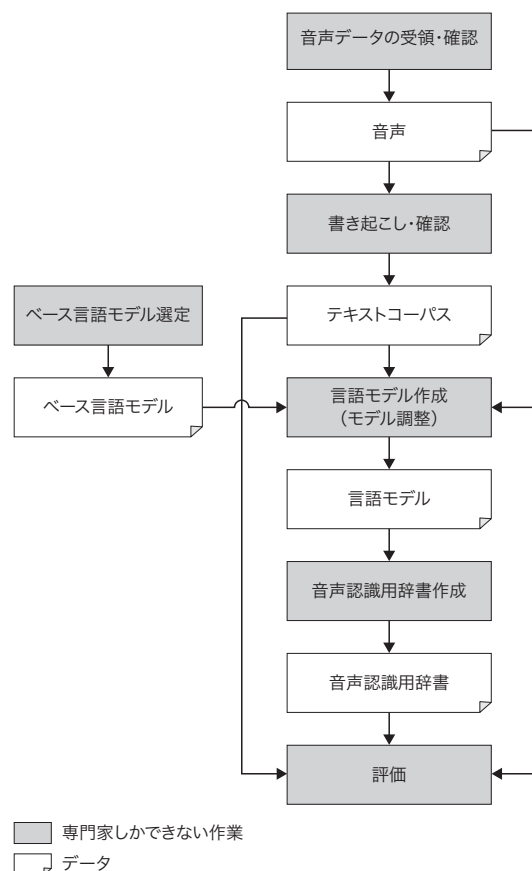


図2. 言語モデルを作成し音声認識用辞書に登録する作業フロー

音声から書き起こしたテキストコーパスを用いて、言語モデル、音声認識用辞書を作成し、その評価を実施する。

Flow of registration of generated language model in voice recognition dictionaries

- (1) データ管理 顧客から受領した音声データや、ツールから生成される中間データ、成果物などを統合的に管理する手段がなく、専門家のノウハウに頼った管理が必要になる。
- (2) ツールハンドリング ツールに入力すべきデータやパラメーターの設定を、作業手順書に沿って実施する必要があり、習熟していないと誤ったツール操作を行う可能性がある。また、ツールの実行結果の確認では、特定のログファイルの特定箇所などを探してチェックする必要がある。
- (3) モデル調整 言語モデル作成では、適切な音声認識が行えるように、言語モデルを調整する作業が必要である。これには、作業による繰り返し調整実験の試行や、調整結果の確認、次の調整実験の設定など、非専門家には難しい作業が含まれている。

3. 専門家に頼らないカスタマイズを実現する要件

2章で述べた現状の問題を解決するには、以下の要件を満たす作業フローを実現する必要がある。

- (1) 作業者間での作業分担とデータの共有 作業フローを幾つかのまとまりに分けて、各作業者間でデータを共有できるようにする。これにより、専門家を必要としない作業を非専門家が分担できる。
- (2) ツールハンドリングの容易化 ツールの実行に必要な入力設定や実行結果の妥当性の判断を、専門知識がなくてもできるようにすることで、非専門家でも作業可能にする。
- (3) モデル調整の自動化 モデル調整は、調整用評価データに基づいて自動的に行われるようにする。

これらの要件を実現するために、以下の3種類のアプローチが考えられる。

- (1) 作業手順書の詳細化 データのファイル管理方式などは作業手順書を詳細化し、作業者が変わっても作業内容に差異が出ないようにする。この方法ではシステム開発などを行う必要はないが、詳細化を行えば行うほど実施工数が増えて作業ミスが入り込む可能性が増大する。
- (2) 既存CIツールなどの活用 ソフトウェア開発でプログラムの自動ビルドやテストなどを行うためのCI (Continuous Integration: 継続的インテグレーション) ツールを用いて、自動化する方法である。これによって、システムの開発工数を抑えられるが、CIツールはコードのビルドなどの定型作業を対象としたツールであり、複数作業者による作業の統合や、作業途中での人手による確認作業などを行うことは難しい。
- (3) 新規プラットフォーム開発 カスタマイズ作業の効率化を支援するための専用の新規プラットフォームを開発する。システム開発の工数は掛かるが、実際のワークフローに合わせたシステムのカスタマイズなどが容易にできる。

今回は、要件の実現に最も適した(3)の新規プラットフォーム開発というアプローチを選択した。

4. プラットフォーム開発

今回開発したプラットフォームは、以下の特長を備えている。

- (1) ジョブセット 一連の作業を実施する上でのデータ管理及びツールハンドリングを、ジョブセットと呼ば

れる単位で定型化可能である。また、ジョブセットは設定された任意のリモートマシンで実行可能である。実行した結果は、プラットフォームが備えるDB (データベース) 内に一元管理され、他のジョブセットからその結果を参照できる。これによって、あるジョブセットの実行結果を入力として他のジョブセットを実行し、複数のジョブセットを連携動作させることができる。

- (2) ジョブセット実行・結果確認UI ジョブセットの実行は、Webブラウザ上から行うことができる。実行済みジョブセットのリストは、ブラウザ上から閲覧でき、各ジョブセットをクリックすればその実行結果の詳細を確認できる。他のジョブセットの結果を入力とする場合は、実行UI (ユーザーインターフェース) 上から対象とするジョブセットをクリックすることで指定でき、ジョブセット間のデータ流通が容易になる。

- (3) モデル自動調整 モデル調整する作業は、複数のモデル調整パターンに対して評価実験を繰り返し、最適な評価結果が得られるモデルを提示することで、自動化を実現した。

開発したプラットフォームの利用フローを、図3に示す。

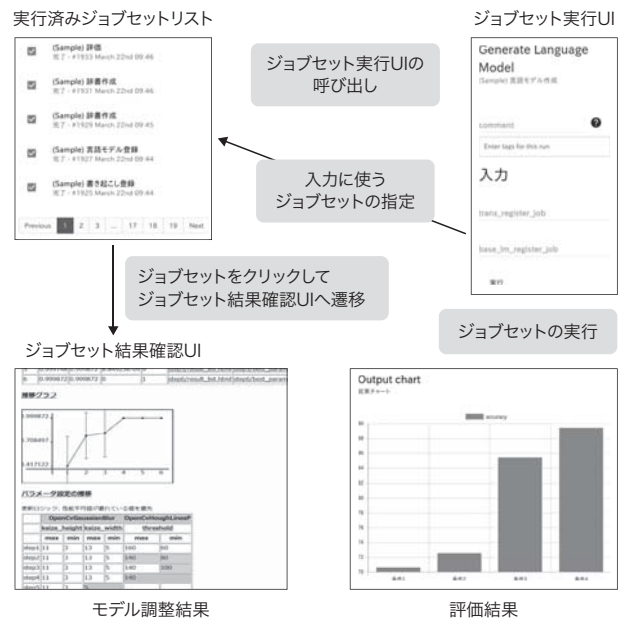


図3. 音声認識用辞書のカスタマイズ作業の効率化を支援するプラットフォームの利用例

プラットフォーム利用者は、ジョブセットの実行や、その結果確認などをWebブラウザ上から実施できる。

Example of usage of newly developed platform to efficiently customize voice recognition dictionaries

プラットフォームのユーザーには、まず実行済みジョブセットリストUIが提示される。ユーザーはこのUIからジョブセット実行UIを呼び出し、ジョブセット実行に必要な入力引数を指定して実行できる。また、ジョブセットの実行結果を次のジョブセットの入力としたいときは、入力引数として実行済みジョブセットをリストから選択する。

ジョブセットリスト内のジョブセットをクリックすると、結果確認UIへ遷移する。結果確認UIでは、モデルを自動調整した結果や、既存辞書とカスタマイズ済み音声認識用辞書の精度評価結果などを確認できる。

5. プラットフォーム導入効果の検証

音声認識サービスに、4章で述べたプラットフォームを用いた音声認識用辞書のカスタマイズを導入した。実際の運用を通して、プラットフォームの導入後に、どのような効果があったかを検証する。

図2と同じ作業フローに、プラットフォームを導入したものを、図4に示す。ジョブセットによる作業とジョブセットを使わない作業が示されている。更に、それぞれの作業は、専門家による実施が必要な作業と、非専門家でも実施可能な作業に分かれる。すなわち、言語モデル作成や、音声認識用辞書の作成、評価など主なカスタマイズ作業のジョブセットについては、非専門家が実施する。これに対して、音声データや、書き起こしテキスト、ベース言語モデルなどの、登録前にデータの確認や選定が必要な作業やデータ登録ジョブセットは、専門家が実施する。

- (1) データ管理 データ登録のジョブセットを用いて、音声や書き起こしテキストなどの登録を行う。登録されたデータは、作業フローごとにプロジェクトと呼ばれる単位に分けて管理される。また、登録時にタグを付加し、タグに基づいて対象データの検索を行うことができる。

プロジェクト内のデータは、プロジェクトに参加している作業員間で共有できるため、データを共有した作業員間で作業分担が容易になる。これによって、ファイル置き場の管理や、ファイル名などに基づいた対象データの検索などを行う必要がなくなる。

- (2) ツールハンドリング ジョブセット実行UIから実行済みのジョブセットの結果を入力引数として選択して実行することで、ジョブセットに設定されたツールが実行される。ツールの実行結果は、ジョブセット結果確認UIから確認できる。ジョブセットの実行結果を作業員間で検証したい場合には、対応するジョブセットID（識

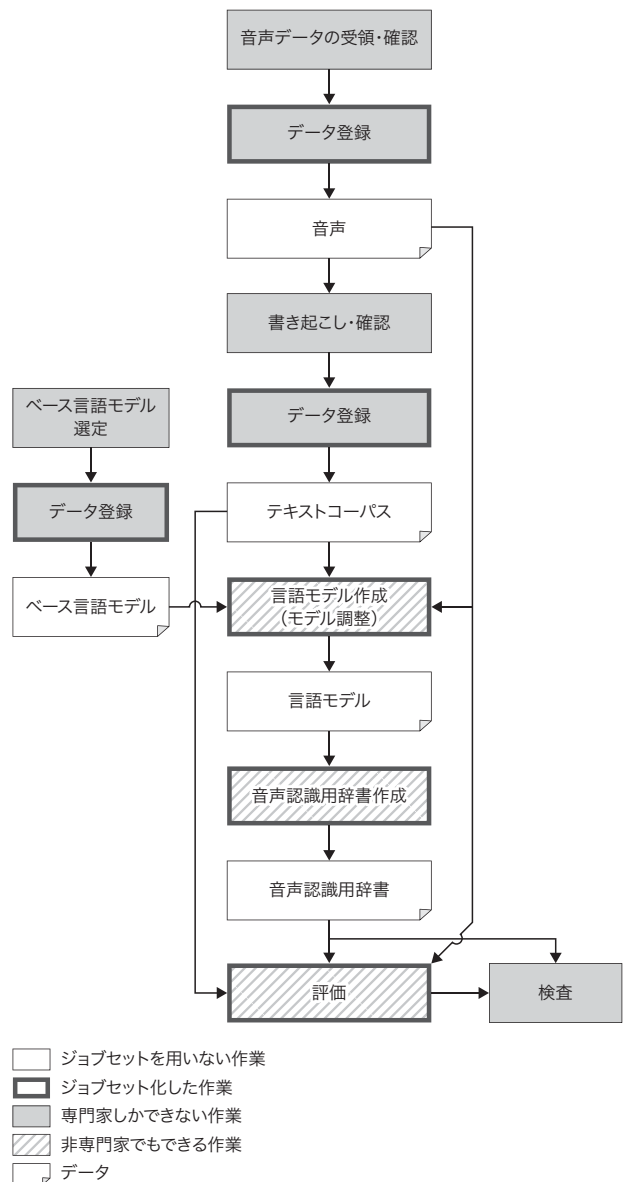


図4. プラットフォームを活用して言語モデルを作成し音声認識用辞書に登録する作業フロー

作業員間でデータの共有ができるようになり、ジョブセットを用いることで、カスタマイズの主要な作業である言語モデル作成などを非専門家でも実施できる。

Flow of registration of language model in voice recognition dictionaries applying newly developed platform

別情報)を通知することで実行内容の確認や再現ができ、作業員間での検証作業が容易になる。

- (3) モデル調整 モデルの調整作業は、言語モデル作成ジョブセットにおいて音声及びテキストコーパスの書き起こしテキストデータを調整対象の評価データとして、自動的に行われる。結果は、ジョブセット結果確

認UIから確認できる。また、評価ジョブセットにおいて調整されたモデルを用いた評価結果を確認できる。

このように、ジョブセットを用いて作業を行うことで、カスタマイズ作業の大部分を占める言語モデル作成などの作業を、非専門家でも実施できるようになった。また、データはデータ登録ジョブセットを用いて管理されるようになり、作業間でのデータ共有が容易になった。更に、ジョブセットにツールを設定しての実行作業やモデル調整作業の自動化を行うことで、人手による作業時間の約40%を削減することができた。

6. あとがき

音声認識サービスにおける音声認識用辞書のカスタマイズ作業にプラットフォームを導入することで、作業の自動化、定型化を進め、主要な作業を非専門家でも実施可能な作業フローを確立した。また、ツールハンドリングやモデル調整作業の自動化により、人手による作業時間を従来に比べて約40%削減した。

今後の課題として、以下が挙げられる。

(1) ジョブセット構築の容易化 今後、カスタマイズサービスの充実に伴って、ジョブセットの拡充を行う必要がある。現在、ジョブセットの構築そのものに専門家による作業が必要であるが、ジョブセットの構築を容易にし、非専門家でもジョブセット構築を行えるようにすることで、作業の自動化を専門家に頼らず実施可能にする。

(2) データ管理方式の更なる洗練化 現在、各種データ登録ジョブセットで登録時に付加したタグによってデータを管理しているが、多数のデータやプロジェクト横断で利用されるベースモデルなどの検索を行うためには、カスタマイズ対象の音声に合わせた適切なデータを検索可能なデータベース設計が必要である。

これらの開発を進め、音声認識用辞書のカスタマイズ作業のフロー全体を自動化し、作業者の介入を不要にすることを目指す。

文献

- (1) 益子貴史, 芦川将之. 同時通訳や音声対話の実用化に向けた大語彙音声認識技術, 東芝レビュー, 2013, 68, 9, p.6-9.



長 健太 CHO Kenta, D.Eng.

東芝デジタルソリューションズ(株)
ソフトウェア&AIテクノロジーセンター システム&サービス技術部
博士(工学) 情報処理学会・日本ソフトウェア科学会・
電気情報通信学会会員
Toshiba Digital Solutions Corp.



阿部 一彦 ABE Kazuhiko

東芝デジタルソリューションズ(株)
RECAIUS 事業推進室 事業開発部
Toshiba Digital Solutions Corp.