## 音声操作で見たい番組をすぐに 見つける対話型意図理解技術

# 音声による対話から意図を理解し、膨大なコンテンツをすばやく検索

東芝はテレビ (TV) 視聴の新たなスタイルとして、ユーザーが設定した録画チャンネルと時間帯の番組を常時録画する、タイムシフトマシンを提案しています。これによって、いつでも時間を遡ってTV番組を見ることができるようになりました。

これまで見たい番組は、キーワードやジャンルなどを 指定して検索していましたが、今回、自然言語を用いた音 声による対話と番組情報のクラウドコンピューティングを 組み合わせることで、簡単に、かつスピーディに見たい番 組やシーンを見つけ出すことができる"対話型意図理解 技術"を開発しました。

# ●曖昧性のない表現 「2014年10月1日の東芝太郎が出演している BS東芝テレビジョンの音楽番組が見たい」 ●曖昧な人間の発話例 「昨日やった東芝太郎の音楽番組ない?」 ・ 音声認識の失敗例 「昨日あった東芝の音楽番組な」 ・ 意図理解の結果 昨日(2014年10月1日)の"東芝"をキーワードとした、放送ジャンルが"音楽"の番組について 録画済みの番組一覧から検索して、検索にヒットした番組を全て表示し、ユーザーに選択を委ねる

**図1. 対話型意図理解技術の利点** 一 曖昧な表現を許容した対話をサポート することで、ユーザーに負担の少ない番組検索を提供できます。

# クラウドシステム 「TimeOn」 メタ情報 エンジン 音声認識部 意図理解部 意図理解 エンジン 読み変換 対話制御 アプリケーション 音声合成 エンジン 音声の理 スピーカ マイク付きリモコン マイク付きリモコン

図2. 対話型意図理解技術の構成 — 製品寿命が長いTV製品をサポートするために、主要な処理はクラウドシステム上に配置しています。

### タイムシフトマシンをもっと便利に

液晶TV レグザシリーズのタイムシフトマシン機能は、ユーザーが指定したチャンネル数と録画時間に応じて複数チャンネルの番組をまるごと録画するため、これらの番組は過去番組表などを使って放送時間に関係なく視聴できます。

またレグザクラウドサービス「Time-On」では、検索キーワードを指定してシーンを検索することで、見たいポイントから視聴することができます。しかし、録画された番組数が膨大になると、見たい番組やシーンを見落としたり、検索の指示に手間と時間が掛かったりするようになってしまいます。

そこで東芝は、番組を見つけたい、 番組を絞り込みたい、シーンを検索した いといったユーザーの要求に対して、自 然言語による音声対話から的確に判断 して結果を提示することができる"対 話型意図理解技術"を開発しました。

### 対話型意図理解技術の特長

一般に、ユーザーが何か物事を指示する場合、唯一の表現というものはなく、多種多様な表現があり、また曖昧な表現もあります。特に自由発話を許容するシステムの場合は、これらが顕著になります。そのため表現それぞれについて、処理方法を全て規定することは極めて難しくなります。

そこで今回、ユーザーの発話を音声 認識した表現からその意図を推定する ために、事前に様々な表現を大量に学 習させた意図理解エンジンを用いるこ とにしました。このとき、現在のTV側 の状態を適宜意図理解エンジンに提供 することで、正確な意図推定を実現して います。また曖昧な表現については、クラウドサービス側のメタ情報エンジンと連携することで、ユーザーが検索したいキーワードの検出精度を高める仕組みを採用しています。最終的に検索処理などで得られた結果は、GUI (グラフィカルユーザーインタフェース) だけでなく合成音声でも知らせることで、ユーザーはTVとの対話を楽しむことができます。

このように対話型意図理解技術は、ユーザーとの間で音声発話と音声合成、及びGUIによって対話しながら、TV機能やクラウドサービスとも連携してユーザーの意図を的確に把握し、**図1**に示すように意図どおりの結果を返すことができる点が、大きな特長になっています。

### 対話型意図理解技術の構成

対話型意図理解技術は、図2に示す

ように主に音声認識部、意図理解部、 及び音声合成部の三つの要素から構成 されています。これらは、それぞれクラ ウドサービスと連携する仕組みになっ ています。

最初にユーザーの肉声を受け付ける 音声認識部は、高音質を保つために マイク付きリモコンで集音した音声を、 2.4 GHz帯の無線通信でTV本体の音 声処理に送るとともに、音声処理から インターネット回線を通してクラウドシ ステム上に配置されている音声認識エ ンジンにも送信します。

意図理解部は、音声処理から送られてきた音声認識後のユーザー発話文を元に、次に実行する処理、画面に表示する内容、及び合成音声で音声発話する内容を決定する意図理解エンジンと、意図理解エンジンからの処理依頼をTV本体側の動作にえり分ける対話制

御アプリケーションから成ります。意図 理解エンジンで、意図を推定するうえ で不足している情報があれば、対話制 御アプリケーションは相互に連携し、 適宜TV側の状態を送信しつつ、次に 実行する処理や、画面に表示する内容、 合成音声で音声発話する内容などを的 確に処理します。TV内の録画番組を検 索する場合、もし検索キーワードが省 略語や別名であれば、検索キーワード をメタ情報エンジンに送り、正しい検索 キーワードに変換してから検索を実行 します。

音声合成部は、誤った発話を防ぐためにキーワードの読みをクラウドシステム上のサーバから取得し、TV内蔵の音声合成エンジンで音声合成します。

対話型意図理解技術には、音声認識、意図理解、及び音声合成のための 最新の辞書が必要になります。そのた めクラウドシステム上に、一貫性を持た せた常に最新の辞書を配置する構成に なっています。

このように、それぞれの処理部が密接に連携することで、対話型意図理解技術を実現しています。

### 今後の展望

対話型意図理解技術は、当社製UHDTV (超高精細度テレビ: 3,840×2,160 画素) のレグザ Z10X に、初めて導入されました。ユーザーは、音声入力が可能なマイク付きリモコンを用いて自由に発話し、発話意図に対応した合成音声やGUIを通して、TVと対話することが可能になりました。これによって、TV 視聴をより簡単な音声操作で実現できるようになりました。

対話型意図理解技術は、これからも 進化していきます。例えば、対話の履歴 を活用した推薦番組や操作の提案、よ り的確な番組の絞込みなどが考えられ ます。

今後、この対話型意図理解技術の改良を進め、TVを更に便利に使えるように努めるとともに、TV以外の製品への応用にも取り組んでいきます。

### 伊藤 能康

研究開発センター ライフスタイルソリューション開発センター エンベデッドソフトウェア技術開発部主務