

Raspberry Pi と Arduino を利用した音声入出力が可能な 性格を踏まえた感情表現のできる対話システム

呉敏 長名優子

東京工科大学 コンピュータサイエンス学部

1 はじめに

テキストや音声を入力とした対話システムが普及してきており、有名なものとして、Apple の Siri や Microsoft の りんななどが挙げられる。対話システムの実用化が進むのに伴い、様々な対話システムに関する研究も行われている。

そのひとつとして性格を踏まえた感情表現のできる対話システム [1] が提案されている。このシステムでは、ユーザ主導とシステム主導による対話とが実現されている。ユーザ主導による対話は、ユーザの発話により対話を開始され、ユーザの発話に対する応答を行う処理である。システム主導による対話は、ユーザの発話間隔を取得し、一定時間沈黙が続いた場合にシステムからユーザに対して質問を行う処理である。システムの応答は、性格により感情表現に違いが生じることとなる。ユーザとの対話による東京工科大学の情報案内が実現されているが、システムとの対話はテキストベースのものとなっている。

本研究では、Raspberry Pi と Arduino を利用した音声入出力が可能な性格を踏まえた感情表現のできる対話システムを提案する。提案システムは、性格を踏まえた感情表現のできる対話システムに音声機能を加えたものである。

2 性格を踏まえた感情表現のできる対話システム

ここでは、提案システムで用いる性格を踏まえた感情表現のできる対話システムについて説明する。図 1 に性格を踏まえた感情表現のできる対話システムの構造を示す。このシステムは、(1) 発話理解処理、(2) 対話管理処理、(3) 問いかけ処理の 3 つの部分で構成されている。

発話理解処理では、タイプ判定および知識ベース検

索のためにユーザ発話に含まれるキーワードを取得する。続いて、取得したキーワードに基づいてユーザ発話のタイプを判定する。タイプ判定では、ユーザの発話を (1) 検索要求、(2) 質問応答、(3) その他の 3 種類のタイプに分ける。また、応答された話題のデータを保持して利用できるようにトピック同定を行う。

対話管理処理では、発話のタイプによって知識ベース検索または応答方針決定を行う。ユーザ発話のタイプが検索要求と質問応答と判定された場合には、取得されたキーワードと同定されたトピック名を用い、知識ベースに蓄積した応答文を検索する。取得した応答文はそのまま利用して応答生成を行う。求める情報が知識ベースにない場合は、テンプレートをを用いて応答文を生成する。知識ベースは、東京工科大学公式ホームページに含まれるデータから構成されており、トピックが階層構造で保持されている。ユーザ発話のタイプがその他と判定された場合には、発話の内容に応じて応答方針が決定される。例えば、発話の内容が挨拶であれば挨拶を、賛辞であればお礼、肯定、否定の中から選ばれて返すことになる。続いて、対話システムの感情の更新を行う。感情は、対話の流れの中で抱いている感情とユーザの発話に対して抱く感情とから更新される。ユーザ発話に対して抱く感情は設定されている性格によって異なる。感情表現では、更新後の感情を感情表現シソーラスとテンプレートをを用いて応答文を生成する。

問いかけ処理は、沈黙が一定時間続いた場合にシステムからユーザに対して対話を開始するものである。

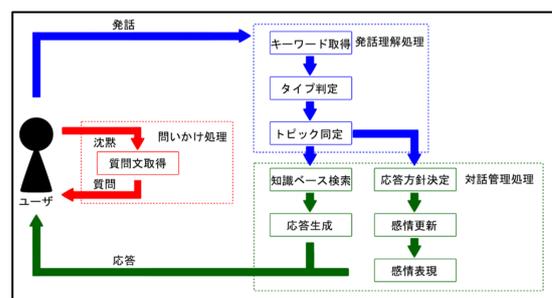


図 1: 性格を踏まえた感情表現のできる対話システムの構造

Dialogue System Capable of Expressing Emotions based on Personality with Audio Input/Output using Raspberry Pi and Arduino
Min Wu and Yuko Osana (Tokyo University of Technology, osana@stf.teu.ac.jp)

同定された現在のトピックとトピックの履歴を用い、まだ話題になっていない内容を持つトピックを探し、該当するトピックを発見したら「(該当するトピック)について説明しますか」と問う。

3 Raspberry Pi と Arduino を利用した音声入出力が可能な性格を踏まえた感情表現のできる対話システム

ここでは、提案する Raspberry Pi と Arduino を利用した音声入出力が可能な性格を踏まえた感情表現のできる対話システムについて説明する。提案手法では、対話システムは 2 で述べた性格を踏まえた感情表現のできる対話システムと同様の方法で実現する。ここでは、音声による入出力を実現する部分について説明する。

3.1 ハードウェア

Raspberry Pi 3 Model B+

本研究で用いる Raspberry Pi 3 Model B+ は、小型のシングルボードコンピュータと呼ばれ、Raspberry Pi の一種である。microSD カードから Linux を起動し、Python 言語を用いてロボットの音声機能を制御する。また、シリアル通信で Raspberry Pi から Arduino にロボットの制御命令を送信する。

Arduino Uno

Arduino Uno は機械を制御するためのマイコンボードであるが、Linux のような OS は搭載していない。これを用いて、LED でロボットの顔の表情を、サーボモータでロボットの首・腕・足の動きを制御する。ロボット本体

ロボットとしては、二足人型ロボット Rapiro(図 2) を使用する。Rapiro では、12 個のサーボモータを使って関節を動かし、Arduino と交換性がある制御基板でロボットの動作を制御する。頭部には、Raspberry Pi を搭載することができる。Rapiro を介して会話するときに、腕と頭を振ったり、目も色を変えたり、点滅させたりすることもできる。

3.2 ソフトウェア

SoX (Sound eXchange)

SoX は、Linux の標準的な音声ファイルを編集する音声処理ライブラリである。基本的にコマンドラインで、音声の録音や再生、様々な形式の音声ファイルの変換などの操作ができる。本研究では、SoX ライブラ



図 2: Rapiro

リの rec コマンドを用い、Google の音声認識サービスが処理できる FLAC フォーマットの音声ファイルで録音することにする。

Google Speech API

本研究では、Google Speech API のオンライン型の音声認識を使用する。Google Developer Console サイトから取得した API を利用するには、requests ライブラリの post() メソッドの引数として、認識用 URL と FLAC 形式の音声データ、サンプリングレートなどを指定する。音声認識の処理結果は、JSON (JavaScript Object Notation) データフォーマットで返される。

Open JTalk

Open JTalk は、テキストを日本語音声に変換することができるソフトウェアである。本研究で用いる JTalk は、日本語音声合成ソフトウェア本体、日本語解析のための辞書ファイル、音声データの 3 つの部分から構成される。Open Jtalk の音声データの声は少し硬いため、オープンソースソフトウェア MMDAgent の女性の音声データ「Mei」を使用する。

4 実験

従来のテキストベース [1] のシステムとロボットを用いた提案システムとを用いて実験を行い、ロボットを用いることで、よりユーザが親しみを感じることができることなどを確認した。

参考文献

- [1] 渡部哲也, 長名優子: “性格を踏まえた感情表現のできる対話システム,” 電子情報通信学会総合大会, 2019.
- [2] 鄭立: 自然会話ロボットを作ろう, 秀和システム, 2016.