7U-03

LSTM と音声対話によるコンテキストアウェアシステムの提案

直井　波輝 †　　顧　優輝 †　　真部　雄介 †　　菅原　研次 †
† 千葉工業大学大学院情報科学研究科

1　はじめに

近年，ユーザが次にとる行動を予測し，情報提供を行うコンテキストアウェアサービスが普及し始めている．コンテキストアウェアサービスとは，M2M 技術やクラウドサービスの基で，多種多様なデータを利用し，これらのデータから現実世界の状況（コンテキスト）を判断し，自律的な制御を行うサービスと定義されている [1]．

従来までのコンテキストアウェアサービスの多くは，ロケーションベーステクノロジーと呼ばれる位置情報に基づくものが主流だった [3]．例えば，グーグル株式会社はユーザのコンテキストに合わせた情報カードを表示する機能を搭載したスマートフォンアプリ「Google Now」[2] を開発した．「Google Now」は，ユーザの位置情報と位置履歴を利用することにより，利用する電車の時刻表，現在地と勤務地の時刻表，終電情報などを表示させる機能がある．さらに，末田ら [4] は，位置情報に基づくコンテキスト生成方式の提案をした．ユーザやその周囲のコンテキストをより詳しく把握するため，ユーザの位置情報を用いている．ユーザの位置情報により，訪問履歴や同伴者といった情報を新たなコンテキストを生成する方式により，コンテキスト数を効率的に拡張できることを示している．

しかしながら，コンテキストの定義は環境や状況といったユーザを取り巻く情報のことであり [5]，位置情報や情報機器の操作履歴といった間接的な計測ではなく，直接人間の行動を認識すれば，きめ細かいサービスの提供が可能になると考える．さらに，一方的な情報提供ではなく，インタラクティブな情報提供により細かいサービスの提供が可能になると考える．

そこで本研究では，設置型センサを用いて人間の行動を直接計測し，粒度の細かい行動認識を提案する．さらに，音声対話を用いることにより，インタラクティブな情報提供を可能とする．なお，本稿では行動認識のために必要となる動作認識とユーザに対する情報提供に関して行う．

2　提案方式

図 1に提案手法の処理手順を示す．提案手法は (1) データ登録部，(2) 動作認識部，(3) サービス提供部の 3 つの部分で構成される

まず，登録の流れは Kinect の RGB-D カメラにより行動をセンシングし，被験者の関節データを取得する．このデータに対して特徴抽出を行い認識器の学習を経て，この認識器を動作認識器に登録する．学習には，ニューラルネットワークの一種である Long short-term memory（LSTM）を用いた．LSTM は, 時系列データを扱う Recurrent Neural Network(RNN) の一種である．RNN は学習時の勾配消失によって長い時系列データは扱うことが難しかったが，LSTM は長い系列データも扱うことができる．動作を行う時間は個人によって異なることから，長い系列データでも扱える LSTM を用いる．

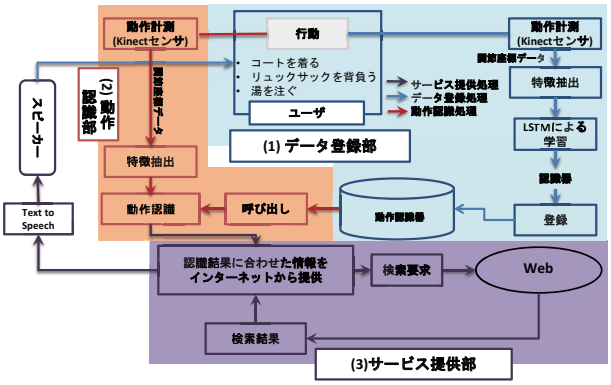

図1　提案手法の流れ

次に動作認識の流れは，登録の流れと同様に計測，特徴抽出を行い，事前に作成した動作認識器を用いて動作認識を行う．本研究で用いる動作認識の対象を「湯を注ぐ」（以下，動作 A）と「コートを着る」（以下，動作 B），「リュックサックを背負う」（以下，動作 C）とする．

そしてサービス提供の流れは，動作認識の結果に合わせて Web から情報を取得する．本研究で用いるサービス提供は，動作 A に対して，即席カップめん調理行動とみなし，3 分測り通知する．動作 B と C に対しては，帰宅行動とみなし直近の電車のダイヤを提供することを想定する．

最後に，取得した情報を音声対話エージェントを用いてスピーカーよりユーザに提供する．対話エージェントは，MMDAgent[6] を用いる．

2.1　特徴抽出・認識器

登録の流れにある特徴抽出，認識器について述べる．特徴抽出では，人物の手やうで，頭を中心とする上半身の関節座標から角度および，相対的な位置関係を算出する．

まず，角度の算出は式 (1)-(3) により，xy, yz, zx 平面上の値をそれぞれ算出する．

$$\vec{a}_{PQ} = (P_k(t) - P_o(t), Q_k(t) - Q_o(t)) \quad (1)$$

$$\vec{b}_{PQ} = (P_l(t) - P_o(t), Q_l(t) - Q_o(t)) \quad (2)$$

$$\theta_{PQ}(t) = \frac{180}{\pi} \times arccos\left(\frac{\vec{a}_{PQ} \cdot \vec{b}_{PQ}}{|\vec{a}_{PQ}||\vec{b}_{PQ}|}\right) \quad (3)$$

ただし，$(P, Q) = \{(x, y), (y, z), (z, x)\}$とする．

A context-aware system using spoken dialogue and Long short-term memory

†Namiki NAOI　†Yuki KAERI　†Yusuke MANABE　†Kenji SUGAWARA
†Graduate School of Information and Computer Science, Chiba Institute of Technology

ここで，P および Q は，用いる座標の軸を示しており，t はタイムスタンプを示している．o は角度を算出する関節部位を示しており，k と l はそれぞれ o に隣接する関節部位を表しており，用いた 5 つの組み合わせを表 1に示す．

表1　角度算出に用いた $k-o-l$

k	o	l
Head	Neck	SpineShoulder
SpineShoulder	ShoulderLeft	ElbowLeft
ShoulderLeft	ElbowLeft	WristLeft
SpineShoulder	ShoulderRight	ElbowRight
ShoulderRight	ElbowRight	WristRight

次に，相対的な位置関係の算出は，式 (4),(5) により，x, y, z 軸上の頭と右手，頭と左手の値をそれぞれ算出．

$$R_{HR} = (x_{HR}(t) - x_H(t), y_{HR}(t) - y_H(t), z_{HR}(t) - z_H(t)) \quad (4)$$
$$R_{HL} = (x_{HL}(t) - x_H(t), y_{HL}(t) - y_H(t), z_{HL}(t) - z_H(t)) \quad (5)$$

ここで，H は頭の座標を示しており，HR, HL はそれぞれ右手，左手の座標を示している．また t はタイムスタンプを示している．

最後に，LSTM を用いて認識器は，角度の算出による 15 次元の特徴量と相対的な位置関係の算出による 6 次元の 21 次元の特徴量を入力とし，動作 A,B,C に対する帰属確率の 3 次元を出力として作成する．

3　評価実験

実験概要と結果を述べる．

3.1　実験概要

千葉工業大学津田沼キャンパス 7 号館 5 階第 6 研究室の室内を実験環境とし，Microsoft 社の Kinect for Windows v2(Kinect) を 2 台使用し実験を行った．それぞれ K1，K2 とし K1 で動作 A を計測し，K2 で動作 B と C を計測する．それぞれ 21 回，17 回，26 回計測し，うちそれぞれ 3 回をテストに使い，それ以外を訓練データとして使用した．また，コートは P コートを用いる．被験者は 1 名である．認識器の作成時のハイパーパラメータは，100 エポック，中間層を 1 層，中間層のそれぞれのユニット数 300，バッチサイズ 20，ドロップアウトレート 0.5 とした．この時，エポック間の平均認識率を算出したところ，テストデータに対する認識率が高かった．この値で学習した時の認識率を図 2に示す．

3.2　認識器の評価

図 2より検証データに対しての認識率が大きいエポックは 99 で 91.9% であった．また，100 エポック中の平均を出したところ，訓練データに対して認識率 93.8%，検証データに対しての認識率 87.4% となった．

3.3　実験結果・評価

図 3に 9 つのテストデータの認識結果を示す．全体の認識率は 91.9% となった．そして，動作 A は 100% という良好な結果となった．一方，動作 B は動作 C に認識され，動作 C は動作 B に認識されることがあったが，

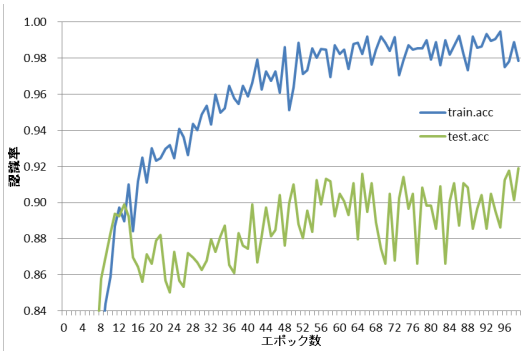

図2　エポック毎の認識率

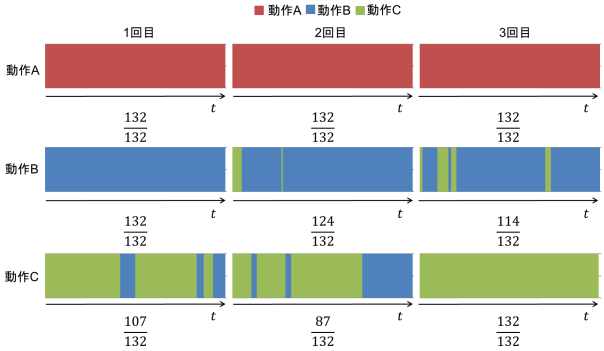

図3　各テストデータに対する認識結果

動作 B は 93.4%，動作 C は 82.3% と概ね良好な結果となった．

4　おわりに

動作認識の結果，設置型センサを用いた LSTM による動作認識は有意であることを示した．しかし，これは動作ごとに個別にデータを計測し，学習，認識をした場合である．実世界は，動作が連続して行われるので，このような場合においても有意であることを示す必要がある．

また，認識結果に基づいた情報提供は可能になったが，インタラクティブな情報提供を実装していない．このため，インタラクティブな情報提供の場合での有用性を示す必要がある．

参考文献

[1] 髙塚他: 異種分散 Web サービスに基づくコンテキストアウェアサービスの管理フレームワークの提案, 信学技報, Vol. 113, No. 326, pp. 71–76, 2013.
[2] Google Now で欲しい情報をちょうどいいタイミングで, https://www.google.com/intl/ja/landing/now/(最終閲覧日: 2018.01.08)
[3] 田崎, “空気を読む”システムを実現する Context-Aware Computing（第 6 回,https://it.impressbm.co.jp/articles/-/7777(最終閲覧日：2018.01.08).
[4] 末田他: 位置情報に基づくコンテキスト生成方式, 情処研報（UBI）, Vol. 2004, No. 39, pp. 29–34, 2004.
[5] 中村他: コンテクストアウェアコンピューティングとコンテクストの定式化, 人工知能学会研究会 (SIG-SWO-7), Vol. SIG-SWOA402-03,2004.
[6] MMDAgent, http://www.mmdagent.jp/(最終閲覧日: 2018.01.11)