

再帰型ニューラルネットワークによるアプリケーションの利用履歴を用いたユーザのコンテキスト推定モデルの構築

神谷 俊充¹ 中村 達哉¹ 前川 卓也¹ 天方 大地¹ 原 隆浩¹

概要: スマートフォン向けのアプリケーションの増加に伴い, ユーザは様々なアプリケーションを自身の状況に応じて使い分けている. 最近では, ユーザの状況として現在の活動内容や同伴者の数, 気分, 疲労度といったより具体的な状態を表すコンテキストが研究の対象として注目を集めている. しかし, コンテキストはユーザの主観的な情報を多く含んでおり, スマートフォンに内蔵されたセンサでは推定が困難なため収集コストが高い. そこで本稿では, アプリケーション利用履歴を用いたユーザのコンテキスト推定モデルを構築する手法を提案する. ユーザのアプリケーション利用履歴とコンテキストの関係は時間の経過とともに変化する時系列データであると考えられるため, 提案モデルでは時系列データの学習が可能な再帰型ニューラルネットワークを用いた. 評価実験の結果, 提案モデルがアプリケーション利用履歴からユーザのコンテキストを推定できることを確認した.

1. はじめに

近年, Android や iOS などの高性能な OS を搭載した携帯端末であるスマートフォンが広く普及している. そして従来の携帯端末であるフィーチャーフォンと比較して, スマートフォンは高性能な CPU や, 加速度, GPS 等の様々なセンサを内蔵しており, これら計算リソースやセンサを活用した様々なアプリケーションが開発されている.

スマートフォンユーザは様々なアプリケーションの中から自身の状況に応じてアプリケーションを使いこなしている. 例えば, 外出先で食事をするレストランを探すためにレストラン検索アプリケーションを利用するといった場面が挙げられる. また, アプリケーションもユーザの状況に応じて様々なサービスを提供している. 例えば, レストラン検索アプリケーションではスマートフォンの GPS センサの情報を用いてユーザの現在地を把握し, 周辺のレストランを自動的に検索する機能がある. このようにユーザの状況と利用するアプリケーションには関連性があると考えられ, これまでにユーザの状況と利用アプリケーションの関係性の分析 [2], [3], [4], [7], [14] や, この関係性を用いたユーザのアプリケーション利用予測の研究 [9], [16], [21] が行われている.

最近では, ユーザの現在の活動内容や同伴者の数, 気分, 疲労度といった, スマートフォンに内蔵されたセンサでは推定が困難な具体的なユーザの状況 (以下, コンテキ

ストと呼ぶ) に注目が集まっている. そこで筆者らの研究グループでは, ユーザのコンテキストとアプリケーションの利用履歴を収集するスマートフォン用アプリケーション「こんてきすとモンスター」[13] を運用してきた. これまでに, こんてきすとモンスターによって得られたコンテキストとアプリケーションの利用履歴を調査した結果, 特定のコンテキスト下でのみ頻繁に利用されるアプリケーションの存在を確認した [8]. さらに, アプリケーションの利用予測において, コンテキストを用いることにより, アプリケーションの利用履歴のみを用いる場合よりも予測精度が向上することを確認した.

しかし, 前述の通りユーザのコンテキストはスマートフォンに内蔵されたセンサでは推定が困難なため, ユーザに直接入力してもらう必要があり収集コストが高い. ユーザのコンテキストを推定できれば, コンテキストの入力に関わるユーザの負担を軽減できるだけでなく, コンテキストが入力されていない過去の期間の情報を補完できると考えられる.

そこで本研究では, こんてきすとモンスターによって収集したデータを用いて, ユーザのアプリケーションの利用履歴からコンテキストを推定するモデルを構築する手法を提案する. ここで, ユーザが利用したアプリケーションとその時のコンテキストの関係は時間の経過とともに変化する時系列データであると考えられる. 時間的に変化する関係性を持つデータに対して推定を行う場合, 短期的な関係性だけでなく長期的な関係性も考慮する必要があ

¹ 大阪大学大学院情報科学研究科マルチメディア工学専攻

る。この問題に対して提案手法では、時系列データの推定に用いられる再帰型ニューラルネットワーク (Recurrent Neural Network, RNN) の一種であり、より長期的なデータ間の関連性の学習が可能である LSTM (Long Short-Term Memory) [6] を推定モデルに用いる。

アプリケーションの利用履歴を用いてコンテキストを推定する際に考慮すべき点として、ユーザが新しくリリースされたアプリケーションをインストールし利用するといった、ユーザの利用アプリケーションが時間の経過によって変化することが挙げられる。この場合、新たにインストールされたアプリケーションに対応するコンテキストのデータが存在しないため、アプリケーション利用時のコンテキストの推定は困難である。しかし、新たにインストールされたアプリケーションと類似した機能を持つアプリケーションが過去に利用されていた場合、新しいアプリケーションは機能が類似したアプリケーションと同じコンテキストで利用されやすいと考えられる。そこで提案手法では、アプリケーションマーケットに存在するアプリケーションの説明文を用いて、アプリケーションを自身の説明文に基づく意味ベクトルによって表現する。これにより、類似した説明文を持つアプリケーションは互いに類似した意味ベクトルとして表現されるため、提案モデルは新たにインストールされたアプリケーションであってもコンテキストを推定できる。

2. 関連研究

本章ではまず、センサ情報やスマートフォンの利用情報を用いたユーザのコンテキスト推定に関する研究について述べる。その後、再帰型ニューラルネットワークを用いたコンテキスト推定に関する研究を述べる。

2.1 スマートフォンユーザのコンテキスト推定

スマートフォンの利用状況や内蔵のマイク、センサから収集したデータを用いてユーザのコンテキストを推定する研究が行われている。Sano ら [15] は、通話や SMS の数や長さの平均や合計といったスマートフォンの利用状況とユーザの手首に装着したセンサより収集した加速度や皮膚コンダクタンスを用いて、ユーザがストレスを感じているか否かを推定する手法を提案した。Lu ら [11] は、スマートフォンのマイクを用いて収集したユーザの話し声のピッチ等を特徴量としてユーザのストレスの度合を推定するモデルを構築した。評価の結果、収集データが多いほど高い精度を達成できる一方、データ収集に関するユーザの心理的負担が増大することを確認した。Bogomolov ら [1] は、通話や SMS などの数や長さといったスマートフォンの利用状況やユーザの性格、気象情報から、スマートフォンのセンサを用いることなく、ユーザのストレスの度合を推定する手法を提案した。評価の結果、ユーザのストレス度合を推

定する上でスマートフォンの利用状況、ユーザの性格、気象情報がいずれも重要であることを確認した。LiKamWa ら [10] は、通話履歴、SMS 履歴、アプリケーションの起動回数や利用時間を利用し、Russell の円環モデルに基づくユーザの気分を推定する手法を構築した。またデータが少ないユーザに対する推定精度の向上を目的として、他ユーザのデータを用いて推定を行う手法も提案している。

これらの研究ではスマートフォンに内蔵されたセンサから得られるデータやユーザのスマートフォンの利用状況を用いてユーザのコンテキストを推定している。しかし、マイクや GPS などのセンサ情報にはユーザのプライバシーに関わる情報が多く含まれており、これらの情報を継続的に収集することはユーザにとって心理的な負担が大きい。そこで本研究ではスマートフォンの利用状況のうち、容易に収集可能であり、通話やメッセージの持つ誰に送信するかという情報が含まれず、心理的負担が小さいと考えられるアプリケーションの利用履歴を推定に用いた。またこれらの研究では通話履歴やアプリケーションの利用履歴といったスマートフォンの利用状況とコンテキストの静的な関係のみを用いて推定を行っている。しかし、スマートフォンの利用状況とコンテキストは時間の経過とともに変化する時系列データの一つであると考えられるため、本研究では時系列データの取扱いが可能な再帰型ニューラルネットワークを用いて推定モデルを構築した。

2.2 再帰型ニューラルネットワークを用いたコンテキスト推定

近年、機械学習の手法の 1 つである深層ニューラルネットワークが注目を集めている。その中でも再帰型ニューラルネットワーク (Recurrent Neural Network, RNN) は時系列データを対象とした様々なタスクにおいて優れた成果を挙げている。Vu ら [20] は RNN を拡張した独自の構造を持つニューラルネットワークを提案し、ユーザのスマートフォンに内蔵されている加速度、磁気およびジャイロの 3 つのセンサを用いて、歩いている、走っている、車に乗っているといったユーザの移動状況を推定する手法を提案した。Suhara ら [17] は、専用のスマートフォン用アプリケーションを用いて睡眠時間や日中の気分、服薬の有無など、うつ病患者の詳細な状態に関する情報を収集し、RNN の一種である LSTM (Long Short-Term Memory) [6] とサポートベクトルマシン [19] を用いて患者の現在の症状を推定するモデルを提案した。評価の結果、LSTM を用いたモデルの方が患者の長期的な状態の変化を捉えた推定が可能であることを確認した。

これらの研究により、スマートフォンのセンサや利用状況などの時系列データからユーザの状態やコンテキストを推定する際に RNN が有効であることがわかった。そこで本研究でも同様に、RNN を推定モデルに採用した。また、

表 1 収集対象のコンテキストと分類項目群, 有効時間

コンテキスト	分類項目の例	有効時間
今日	仕事, 休み, 学校, バイトなど	24 時間
活動	仕事, 勉強, 食事, 暇, 音楽など	2 時間
同伴者の人数	誰もいない, 1 人, 2 人, . . . , たくさん	2 時間
同伴者の分類	家族, 恋人, 親戚など	2 時間
場所	学校, 仕事先, 駅・電車など	3 時間
気分	4 段階選択	4 時間
体調	4 段階選択	4 時間
忙しさ	4 段階選択	4 時間

本研究はアプリケーションの利用履歴のみを用いて様々な種類のコンテキストの推定を実現しているという点でこれらの研究より難しい問題を扱っていると言える。

3. こんてきすとモンスターと収集データの概要

3.1 こんてきすとモンスター

本研究では, こんてきすとモンスター [13] により収集されたデータを使用した。こんてきすとモンスターは, ユーザのコンテキストとアプリケーションの利用履歴の収集を目的とした Android 用アプリケーションである。以降では, こんてきすとモンスターにおけるコンテキストとアプリケーションの利用履歴の収集方法について説明する。

3.1.1 コンテキストの収集

こんてきすとモンスターでは, ユーザに対して身の回りの状況に関する質問を定期的に行い, 表 1 に示すコンテキストを収集している。それぞれのコンテキストには有効時間が設定されており, 有効時間が切れた場合, ユーザに通知を行いコンテキストの入力を促すよう設計されている。また, ユーザの長期的な利用を促すため, コンテキストの入力ごとにモンスターが成長するモンスター育成ゲーム要素 [13] と過去のコンテキストのタイムラインや統計情報を表示するライフログ機能 [8] を持つ。

3.1.2 アプリケーションの利用ログの収集

Android では, 端末の画面に表示中のアプリケーションのアプリケーション名とパッケージ名を, 別のアプリケーションのプロセスが取得できる。この機能を用いることによりこんてきすとモンスターはユーザが利用したアプリケーションを常に記録する。具体的には, 画面に表示中のアプリケーションのパッケージ名を取得するバックグラウンドプロセスを常に起動させ, 取得パッケージ名が変化したタイミングをそのパッケージ名を持つアプリケーションの起動とみなし, 起動時刻, ユーザの識別 ID, アプリケーション名 (パッケージ名) を収集する。

3.2 使用したデータの概要

本研究では, 2015 年 2 月 9 日から 2015 年 12 月 10 日までの期間において, こんてきすとモンスターによって収集

表 2 使用データの統計情報

	件数
アプリケーション数	259
利用アプリケーションの履歴数	211,931
コンテキストと紐付いたログ数	14,043
ユーザのログ数の平均値	520.11
ユーザのログ数の中央値	50
全ユーザ中の最大ログ数	4,017
全ユーザ中の最小ログ数	1

された 27 名分のデータを使用した。表 2 に使用したデータの統計情報を示す。当該期間中に収集されたアプリケーションの利用ログ数は 211,931 件であり, ユーザあたりの平均ログ数は 520.1 件, 中央値は 50 件であった。また全アプリケーション利用ログのうち, 「今日」のコンテキストを除いて有効時間内のコンテキストが一つでも紐付けられたログ数は 14,043 件 (6.6%) であった。

4. 提案手法

4.1 提案手法の概要

本研究では, ユーザのコンテキスト推定という問題を, ユーザのコンテキストが推定対象のコンテキストであるか否かを分類する二値分類問題として定義する。提案手法では, 推定対象のコンテキストそれぞれについて再帰型ニューラルネットワークを用いた推定モデルを構築する。

図 1 に提案手法の処理の流れを示す。まず, ユーザが利用したアプリケーションの情報を推定モデルに入力可能な形式に変換するため, ベクトルとして表現する。そして, 事前にユーザ自身のアプリケーション利用履歴を用いて学習した推定モデルを用いて, 入力アプリケーション利用時のユーザのコンテキストが推定対象であるか否かを推定する。以下では, 提案手法の各処理について詳しく説明する。

4.2 利用アプリケーションのベクトル表現

提案手法では, アプリケーションの情報をコンテキスト推定モデル (ニューラルネットワーク) に入力可能なベクトル形式に変換する必要がある。本研究では, ベクトル表現形式として, One-hot ベクトルとアプリケーションの説明文から生成した意味ベクトルの 2 種類のベクトル表現形式を用いた。以下では, それぞれのベクトル表現形式について説明する。

4.2.1 One-hot ベクトル

利用アプリケーションの情報をベクトルで表現する簡便な方法として one-hot ベクトルがある。One-hot ベクトルは各アプリケーションに対応する次元を基底として持つベクトルであり, あるアプリケーションはそのアプリケーションに対応する次元の重みを 1, それ以外の次元の重みを 0 としたベクトルとして表現される。本研究で用いたこんてきすとモンスターのデータには 259 種類のアプリケー

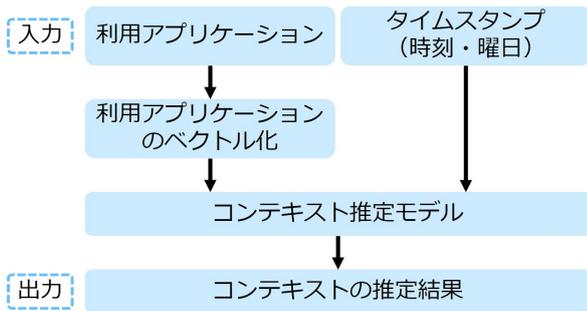


図 1 提案手法の処理の流れ

ションが存在するため、One-hot ベクトルは 259 次元のベクトルとした。

One-hot ベクトルはユーザが利用したアプリケーションの情報を明示的に表現できるが、ユーザのコンテキスト推定においてはいくつか問題がある。まず、ユーザがこれまでに保有していないアプリケーションをインストールして利用した場合、過去にそのアプリケーションを利用したときのコンテキストの情報がないため、コンテキストを推定することが難しい。そして、新しく公開されたアプリケーションなど、ベクトルの次元に含まれないアプリケーションは One-hot ベクトルとして表現できない。そこで本研究ではこれらの問題に対して、次項で述べる意味ベクトル表現を提案する。

4.2.2 アプリケーションの説明文を用いた意味ベクトル

意味ベクトル表現では、アプリケーションの説明文を用いて、任意のアプリケーションをある共通した意味空間のベクトルとして表現する。これにより類似した説明文を持つアプリケーションは互いに類似したベクトルとして表現されるため、ユーザが初めて利用するアプリケーションであっても、過去に利用した類似するアプリケーションのコンテキストの情報を利用してコンテキストを推定できる。

提案手法では意味ベクトルを次の手順で作成する。

(1) Android 向けのアプリケーションマーケット Google Play^{*1}に存在する各アプリケーションの説明文がもつ文書の集合から各単語の TF-IDF 値を計算し、TF-IDF 値の高い上位 10 個の単語をキーワードとして選択する。(2) 各キーワードを word2vec[12] によりベクトル化し、それぞれのベクトルの和をとったものをアプリケーションの意味ベクトルとする。なお、word2vec について提案手法では日本語版 Wikipedia のすべての記事の文章を用いて学習された既存の skip-gram モデル^{*2}[18] を用いた。このモデルでは入力単語は 200 次元のベクトルとして表現される。図 2 にアプリケーションの意味ベクトルの作成例を示す。

*1 <https://play.google.com/store>

*2 http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/

アプリケーションの説明文

LINEは、24時間、いつでも、どこでも、無料で好きなだけ通話やメールが楽しめる新しいコミュニケーションアプリです。...

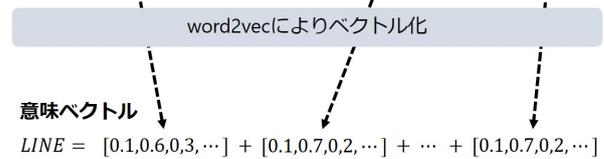


図 2 アプリケーションの説明文を用いた意味ベクトルの作成例 (LINE)

4.3 再帰型ニューラルネットワークを用いたコンテキスト推定モデル

本節ではまず、コンテキスト推定モデルに用いた再帰型ニューラルネットワーク (Recurrent Neural Network, RNN) の一種である LSTM (Long Short-Term Memory) の概要を述べる。その後、コンテキスト推定モデルのネットワークの構成について述べる。

4.3.1 再帰型ニューラルネットワークと LSTM

再帰型ニューラルネットワーク (RNN) は内部に閉路を持つニューラルネットワークの総称である。この構造により過去の情報が内部に保存され、過去と現在の情報から時系列データ内の特徴を捉えた推定ができる。このような特徴から RNN は時系列データを扱うことに適しており、信号処理や自然言語処理などの様々な分野で利用されている。

RNN は過去の情報の特徴を捉えた推定が可能である一方で、長期間のデータの学習では勾配消失問題により学習可能な過去の情報には限りがある。長期的なデータの学習を実現したものとして LSTM と呼ばれるモデルがある。LSTM は RNN の中間層のユニットをメモリユニットで置き換えた構造を持つ。メモリユニットは入力ゲート、出力ゲート、忘却ゲートの 3 つのゲートを用いて効率的に情報を制御することができる。これにより LSTM は通常の RNN と比べて長期的な時系列データを扱うことが可能であり、多くのタスクで一般的に用いられるモデルである。本研究においても、ユーザのアプリケーションの利用履歴とコンテキストの長期的な関係を考慮するため、提案モデルに LSTM を用いた。

4.3.2 コンテキスト推定モデルのネットワーク構成

4.2 節で説明したアプリケーションベクトルを入力としてユーザのコンテキストを推定するモデルを LSTM を用いたニューラルネットワークによって構築する。図 3 に one-hot ベクトルを用いた場合の提案モデルのネットワーク構成を示す。図中の数字は各層の出力の次元数、各層間の文字は直前の層で使用した活性化関数を表す。提案モデルではアプリケーションベクトルとアプリケーションを利用した時刻と曜日を入力として用いた。これらの入力データを全結合ニューラルネットワーク層に入力しそれぞれ 20

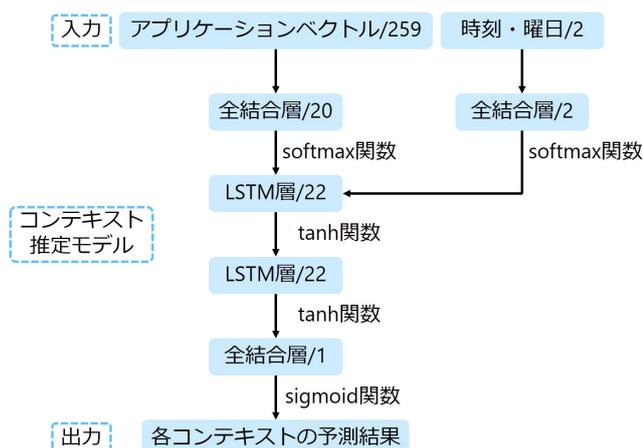


図 3 コンテキスト推定モデルのネットワーク構成 (one-hot ベクトル)

次元と 2 次元の情報に圧縮する。圧縮後、各層の出力を 22 次元の情報として結合し、1 つ目の LSTM 層に入力する。最後に、2 つ目の LSTM 層の出力を全結合ニューラルネットワーク層に入力し、ユーザのコンテキストが推定対象のコンテキストであるかどうかを表す確率を出力する。

推定モデルの学習に関しては各層の重みとバイアスの値を訓練データによって調節することによって行う。重みやバイアスの更新のために用いる損失関数として、出力層の出力 P の対数の負の平均をとったものを使用する。損失関数 $loss$ は以下の式で定義される。なお、 N は訓練データ数を表す。

$$loss = -\frac{1}{N} \sum_{i=1}^N \ln P_i \quad (1)$$

学習にはミニバッチ勾配降下法を用いた。学習パラメータの更新量を決定する学習率は 0.01 とし、バッチ数 50 での学習を 100 回繰り返すように設定した。

5. 評価実験

5.1 評価環境

5.1.1 データセット

本実験では、データセットとして 3 章で述べたこんてきすととモンスターにより収集されたアプリケーションの利用ログとコンテキストのデータを用いる。推定対象のコンテキストは「場所」と項目の重複が多い「活動」、また、入力頻度の少ない「同伴者の分類」を除いたコンテキストとした。表 3 に推定対象のコンテキストの一覧を示す。提案モデルを学習するために、一定量のデータが必要であることを考慮し、データセット中から表 3 に示す「場所」、「気分」といったコンテキスト項目が紐付けられたアプリケーションの利用ログがそれぞれ 400 件以上存在するユーザ 5 名に対して実験を行った。これらユーザのデータはアプリケーションの利用時刻によって整列された時系列データで

表 3 推定対象のコンテキスト一覧

コンテキスト	正例のラベル
場所	自宅
	学校・職場
	移動中
	外出中
同伴者の人数	誰かという
気分	明るい
体調	元気
忙しさ	忙しい

あり、本実験では各ユーザのデータを期間前半の 8 割と期間後半の 2 割に分割した。そして期間前半を訓練データとして各ユーザのコンテキスト推定モデルの学習に使用し、期間後半をテストデータとして各ユーザの推定モデルの性能評価に利用した。表 4 に本実験で使用するデータセットの統計情報を示す。

5.1.2 評価指標

表 4 より本実験で用いたデータセットはユーザおよびコンテキストの種類により正例か負例のどちらかにデータ数が偏った不均衡データであることがわかる。不均衡データに対する二値分類において単純な正解率を評価指標として用いると、正例および負例のうちデータ数が多い方のラベルを常に出力し続けることで容易に高い精度を達成できるため、モデルの精度を適切に評価できないことが知られている [5]。

そこで本実験では、マクロ平均 F 値を評価指標として用いる。マクロ平均 F 値は 0 から 1 の値を取るスコアであり、値が 1 に近いほど推定モデルの性能が良いことを表す。本実験におけるマクロ平均 F 値 ($F1$ -score) は以下の式で定義される。

$$F1\text{-score} = \frac{1}{2}(F1_p\text{-score} + F1_n\text{-score}) \quad (2)$$

ここで、 $F1_p\text{-score}$ と $F1_n\text{-score}$ はそれぞれ正例、負例に対する F 値であり以下の式で定義される。

$$F1_p\text{-score} = \frac{2recall_p \cdot precision_p}{recall_p + precision_p} \quad (3)$$

$$F1_n\text{-score} = \frac{2recall_n \cdot precision_n}{recall_n + precision_n} \quad (4)$$

$precision_{p,n}$ および $recall_{p,n}$ はそれぞれ正例、負例に対する適合率と再現率であり、

$$precision_p = \frac{TP}{TP + FP} \quad (5)$$

$$precision_n = \frac{TN}{TN + FN} \quad (6)$$

$$recall_p = \frac{TP}{TP + FN} \quad (7)$$

$$recall_n = \frac{TN}{TN + FP} \quad (8)$$

として定義される。 TP , FP , TN , FN はそれぞれ表 5

表 4 評価に用いたデータセットの統計情報

コンテキスト	ユーザ	訓練データ		テストデータ	
		正例	負例	正例	負例
自宅	A	355	2,196	82	556
	B	616	336	193	46
	C	1,186	938	323	209
	D	206	195	25	76
	E	1,510	1,350	581	135
学校・職場	A	1,742	809	342	296
	B	198	754	36	203
	C	351	1,773	146	386
	D	99	302	34	67
	E	323	2,537	0	716
移動中	A	297	2,254	99	539
	B	138	814	10	229
	C	312	1,812	63	469
	D	39	362	0	101
	E	1,027	1,833	135	581
外出先	A	157	2,394	76	562
	B	0	952	0	239
	C	275	1,849	0	532
	D	57	344	42	59
	E	0	2,860	0	716
同伴者がいる	A	630	1,338	190	303
	B	575	165	185	0
	C	349	1,579	38	445
	D	53	278	6	77
	E	1,253	1,054	442	135
気分が明るい	A	2,153	503	554	110
	B	257	808	0	267
	C	2,236	0	560	0
	D	444	20	117	0
	E	1,832	1,068	609	116
体調がよい	A	2,019	737	664	25
	B	217	840	0	265
	C	2,190	51	561	0
	D	446	27	119	0
	E	1,078	1,824	154	572
忙しい	A	2,255	318	606	38
	B	724	346	202	66
	C	2,097	132	558	0
	D	356	108	117	0
	E	2,590	314	726	0

表 5 クロス表

		データセットのラベル	
		正例	負例
推定モデルの出力	正例	True Positive (TP)	False Positive (FP)
	負例	False Negative (FN)	True Negative (TN)

で表されるデータのラベルと推定モデルの出力のクロス表に対応するデータ数である。以上の定義からマクロ平均 F 値は、正例と負例の両方を適切に推定できた場合のみ最大値 1 をとるため、不均衡データに対する分類問題に関する研究において用いられている指標である。

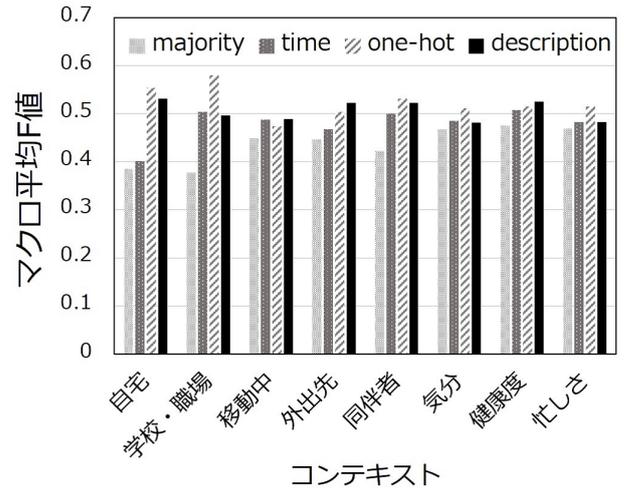


図 4 ユーザ全体の推定精度の平均

5.1.3 比較手法

本実験では以下のコンテキスト推定モデルについて比較を行った。

- Majority : 訓練データの正例と負例のうちデータ数が多い方のラベルを常に出力するモデル。
- Time : アプリケーション利用ログの時刻情報 (時間帯・曜日) のみを用いてコンテキストの推定を行うモデル。モデルのネットワーク構成は 4.3.2 項で説明した構成からアプリケーションの利用履歴の情報を除いたものとした。
- One-hot : 4.3.2 項のコンテキスト推定モデルにおいて、アプリケーションベクトルとして 4.1.2 項で説明した one-hot ベクトルを用いたモデル。
- Description : 4.3.2 項のコンテキスト推定モデルにおいて、アプリケーションベクトルとして 4.2.2 項で説明したアプリケーションの説明文から生成した意味ベクトルを用いたモデル。

5.2 実験結果

実験結果を図 4 に示す。図 4 は推定対象のコンテキストごとにユーザ 5 名のマクロ平均 F 値の平均値を表している。図 4 より、比較手法 (Majority および Time) と提案手法 (One-hot および Description) を比較すると、いずれのコンテキストにおいても、アプリケーションの利用履歴を用いた提案手法のどちらかが最も高い推定精度を達成している。このことからユーザのコンテキスト推定において、アプリケーションの利用履歴を用いることは有用であると考えられる。

One-hot と Description を比較すると、「自宅」、「学校・職場」、「同伴者」、「気分」、「忙しさ」で One-hot が Description より高い精度を達成している。モデル間で推定精度が異なる理由として、コンテキストごとにアプリ

ケーションの利用傾向が学習データ（期間前半）とテストデータ（期間後半）の間で変化する場合が挙げられる。そこで、訓練データとテストデータの間でユーザが利用するアプリケーションの種類がどれくらい変化したかについて調査した。変化の度合を測るための指標として、Jaccard 係数を用いた。Jaccard 係数は、2つの集合 X , Y の類似度を表現する指標であり以下のように定義される。

$$Jaccard_{X,Y} = \frac{|X \cap Y|}{|X \cup Y|} \quad (9)$$

ここで、 X は訓練データに含まれるアプリケーションの集合、 Y はテストデータに含まれるアプリケーションの集合である。

表 6 に訓練データとテストデータに含まれるアプリケーションの種類 Jaccard 係数を正例、負例のそれぞれについて算出した結果を示す。表 6 より、One-hot が Description より高い精度を達成しているコンテキストは、正例と負例の Jaccard 係数がどちらも大きい傾向にあることがわかる。つまり、これらのコンテキストはユーザの利用アプリケーションが時間の経過によって変化しづらいコンテキストであり、そのコンテキストで利用されやすいアプリケーションを学習することができれば高い精度を達成できると考えられる。One-hot は利用されたアプリケーションの情報を明示的に表現できるためこのような特徴を持つコンテキストに対しては One-hot が Description よりも高い精度を達成できたとと言える。

Description で高い精度を達成したコンテキストは正例の Jaccard 係数が小さい傾向にあり、これらのコンテキスト下ではユーザが利用するアプリケーションが時間の経過によって変化しやすいと言える。また表 4 より、これらのコンテキストでは正例の訓練データ数も少ない。そのため、これらのコンテキストに対して One-hot は利用アプリケーションとコンテキストの関係を十分に学習できなかったと考えられる。一方、Description は利用アプリケーションを意味ベクトルとして表現することにより、利用アプリケーションが変化しても説明文が類似した過去の利用アプリケーションの情報からアプリケーションとコンテキストの関係を学習できた結果、これらのコンテキストに対して高い精度を達成したと考えられる。

忙しさは正例の Jaccard 係数が最大であり、かつ、負例の Jaccard 係数が最小のコンテキストであるが、One-hot の方が高い精度を達成している。これは他のコンテキストと比較して、ユーザが忙しいときに利用するアプリケーションがほとんど変化しないため、忙しいときに利用するアプリケーションの情報を十分に学習できた結果、One-hot の方が高い精度を達成したと考えられる。

以上の結果から、アプリケーションの利用履歴を用いた提案モデルはユーザのコンテキスト推定に有効であることがわかった。また、ユーザが利用するアプリケーションの

表 6 各コンテキストにおける Jaccard 係数のユーザ平均

コンテキスト	Jaccard 係数	
	正例	負例
自宅	0.440	0.426
学校・職場	0.414	0.417
移動中	0.289	0.465
外出先	0.149	0.446
同伴者（人数）	0.492	0.349
気分	0.370	0.257
体調	0.349	0.222
忙しさ	0.495	0.141

種類が時間の経過によって変化しづらいコンテキストの推定では One-hot が有効であり、利用アプリケーションの種類が変化しやすいコンテキストでは Description が有効であることがわかった。また、One-hot はユーザの保有アプリケーションがすべて既知であることを想定しているが、日々新しいアプリケーションが大量に公開されており、常にユーザが利用するアプリケーションの種類を把握するのは困難である。One-hot ではベクトルの次元に含まれないアプリケーションが利用される度に新しいベクトルを作成する必要があり、同時に新しいベクトルを用いて推定モデルを再学習させなければならない。一方、Description はアプリケーションの説明文を用いることで、任意のアプリケーションを共通の意味空間内のベクトルとして表現するため、新しいアプリケーションであっても推定モデルを再学習させる必要がないという利点がある。

6. おわりに

本研究では、ユーザが利用したアプリケーションを入力として、そのアプリケーションが利用された時のユーザのコンテキストが対象のコンテキストであるかどうかの二値分類を行うモデルを再帰型ニューラルネットワークを用いて構築する手法を提案した。提案手法では、入力の利用アプリケーション情報の表現方法として、アプリケーションの ID 情報に基づく One-hot ベクトル表現とアプリケーションの説明文に基づく意味ベクトル表現の 2 種類を用いた。また、利用アプリケーションとコンテキストの時間的な関係性を考慮し、長期的な関係を考慮可能な再帰型ニューラルネットワークの 1 種である LSTM を推定モデルに使用した。

提案手法の有効性を検証するために、こんてきすとモニターにより収集したデータを用いた実験を行った。その結果、アプリケーションの利用履歴を使わない比較手法より提案手法が高い精度でユーザのコンテキストを推定できることを確認した。また、提案手法間で性能を比較したところ、ユーザが利用するアプリケーションが時間の経過によって変化しづらいコンテキストの推定には one-hot ベクトル表現が、ユーザが利用するアプリケーションが変化し

やすいコンテキストの推定には意味ベクトル表現が有効であることがわかった。

今後の課題として、スマートフォンに内蔵されたセンサの情報や容易に収集可能なコンテキストを用いた推定精度の向上と、アプリケーションの利用履歴数が少ないユーザに対して履歴数が多いユーザの情報を用いてコンテキストを推定する手法について検討する予定である。

謝辞 本研究の一部は、文部科学省科学研究費補助金・基盤研究(A)(26240013)によるものである。ここに記して謝意を表す。

参考文献

- [1] A. Bogomolov, B. Lepri, M. Ferron, F. Pianesi, and A.S. Pentland, "Daily stress recognition from mobile phone data, weather conditions and individual traits," Proc. Int'l Conf. on Multimedia, pp.477-486, 2014.
- [2] K. Church, D. Ferreira, N. Banovic, and K. Lyons, "Understanding the challenges of mobile phone usage data," Proc. Int'l Conf. on Human-Computer Interaction with Mobile Devices and Services, pp.504-514, 2015.
- [3] T.M.T. Do, J. Blom, and D. Gatica-Perez, "Smartphone usage in the wild: A large-scale analysis of applications and context," Proc. Int'l Conf. on Multimodal Interfaces, pp.353-360, 2011.
- [4] D. Ferreira, J. Goncalves, V. Kostakos, L. Barkhuus, and A.K. Dey, "Contextual experience sampling of mobile application micro-usage," Proc. Int'l Conf. on Human-Computer Interaction with Mobile Devices and Services, pp.91-100, 2014.
- [5] H. He and E.A. Garcia, "Learning from imbalanced data," IEEE Trans. Knowledge and Data Engineering, vol.21, no.9, pp.1263-1284, 2009.
- [6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol.9, no.8, pp.1735-1780, 1997.
- [7] S.L. Jones, D. Ferreira, S. Hosio, J. Goncalves, and V. Kostakos, "Revisitation analysis of smartphone app use," Proc. Int'l Conf. on Pervasive and Ubiquitous Computing, pp.1197-1208, 2015.
- [8] 川上慶士, 大澤 純, 原 隆浩, 西尾章治郎, "ゲーミフィケーションに基づくコンテキスト収集システムにおけるライフログ機能の導入とその効果の検証," 情報処理学会マルチメディア, 分散, 協調とモバイルシンポジウム論文集, pp.714-721, 2015.
- [9] Z.X. Liao, Y.C. Pan, W.C. Peng, and P.R. Lei, "On mining mobile apps usage behavior for predicting apps usage in smartphones," Proc. Int'l Conf. on Information & Knowledge Management, pp.609-618, 2013.
- [10] R. LiKamWa, Y. Liu, N.D. Lane, and L. Zhong, "Mood-scope: Building a mood sensor from smartphone usage patterns," Proc. Int'l Conf. on Mobile Systems, Applications, and Services, pp.389-402, 2013.
- [11] H. Lu, D. Frauendorfer, M. Rabbi, M.S. Mast, G.T. Chittaranjan, A.T. Campbell, D. Gatica-Perez, and T. Choudhury, "Stressense: Detecting stress in unconstrained acoustic environments using smartphones," Proc. Conf. on Ubiquitous Computing, pp.351-360, 2012.
- [12] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," Advances in Neural Information Processing Systems, pp.3111-3119, 2013.
- [13] 大澤 純, 岩田麻祐, 原 隆浩, 西尾章治郎, "ユーザコンテキストと利用アプリケーションの関連性調査のためのゲームを用いたアプリケーション利用ログ収集システム," 情報処理学会マルチメディア, 分散, 協調とモバイルシンポジウム論文集, pp.1855-1862, 2012.
- [14] A. Oulasvirta, T. Rattenbury, L. Ma, and E. Raita, "Habits make smartphone use more pervasive," Personal and Ubiquitous Computing, vol.16, no.1, pp.105-114, 2012.
- [15] A. Sano and R.W. Picard, "Stress recognition using wearable sensors and mobile phones," Humaine Association Conf. on Affective Computing and Intelligent Interaction, pp.671-676, 2013.
- [16] C. Shin, J.H. Hong, and A.K. Dey, "Understanding and prediction of mobile application usage for smart phones," Proc. Int'l Conf. on Ubiquitous Computing, pp.173-182, 2012.
- [17] Y. Suhara, Y. Xu, and A. Pentland, "Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks," Proc. Int'l Conf. on World Wide Web, pp.715-724, 2017.
- [18] 鈴木正敏, 松田耕史, 関根 聡, 岡崎直観, 乾健太郎, "Wikipedia 記事に対する拡張固有表現ラベルの多重付与," 言語処理学会第 22 回年次大会, 2016.
- [19] I. Tschantzaris, T. Joachims, T. Hofmann, and Y. Altun, "Large margin methods for structured and interdependent output variables," Journal of machine learning research, vol.6, pp.1453-1484, 2005.
- [20] T.H. Vu, L. Dung, and J.C. Wang, "Transportation mode detection on mobile devices using recurrent nets," Proc. Int'l Conf. on Multimedia Conference, pp.392-396, 2016.
- [21] X. Zou, W. Zhang, S. Li, and G. Pan, "Prophet: What app you wish to use next," Proc. Int'l Conf. on Pervasive and Ubiquitous Computing, pp.167-170, 2013.