

高度なコンテキスト情報を活用した ユーザの利用アプリケーションの推定

川上 慶士[†] 原 隆浩[†] 前川 卓也[†] 石塚 宏紀^{††} Xing Xie^{†††}

[†] 大阪大学大学院情報科学研究科 〒565-0871 大阪府吹田市山田丘 1-5

^{††} KDDI 総合研究所 〒356-0003 埼玉県ふじみ野市大原 2-1-15

^{†††} Microsoft Research Asia

E-mail: †{kawakami.keiji,hara,maekawa}@ist.osaka-u.ac.jp, ††hk-ishizuka@kddi-research.jp,

†††xingx@microsoft.com

あらまし スマートフォン向けアプリケーションの爆発的増加に対し、ユーザの状況を考慮した利用アプリケーションの推定は、ユーザビリティ向上の一因になると考える。筆者らの研究グループでは、スマートフォンの利用傾向を収集することを目的に、「こんてきすとモンスター」というアプリケーションを運用し、ユーザの身の周りの状況を表す高度なコンテキスト情報を収集している。そして、得られたユーザのアプリケーション履歴とコンテキスト情報を活用し、再帰的ニューラルネットワークを用いてユーザの利用アプリケーションの推定を行った。その結果、コンテキスト情報を用いた場合の推定精度の向上を確認した。また、学習用のデータが不足したユーザを想定し、他のユーザデータを用いたアプリケーションの利用カテゴリの予測を同時に行い、複数人のデータを用いた予測が有効であることを確認した。

キーワード モバイル、コンテキスト、アプリケーション、スマートフォン、予測

1. はじめに

近年スマートフォン向けのアプリケーションは爆発的に増加している。このような環境下、ユーザは様々なアプリケーションを利用できる一方、自身の端末から目的のアプリケーションを探し出すことが煩雑になっている。

そこで近年、スマートフォン使用履歴を用いたユーザの利用アプリケーションの推定を行う研究が盛んに行われている [2], [9]。ユーザの利用するアプリケーションを推定することで、利用アプリケーションの候補を表示したり、アプリケーションの利用傾向によって画面をカスタマイズすることができ、ユーザビリティ向上に貢献できる。

アプリケーションの利用予測を行う既存研究では、アプリケーションの利用履歴とともに、ユーザの身の回りの状況（以下、コンテキストと呼ぶ）として GPS による位置情報、加速度センサなど、端末から自動的に取得できる情報を利用している。しかし、センサ情報では「何をしているか」、「誰といるか」、「気分が良いか」といったユーザの状況を具体的に表す高度なコンテキストを推測することが困難である。そして、ユーザのアプリケーションの利用傾向は、このような高度なコンテキストを含む様々なコンテキストが影響していると考えられるため、より効果的な利用アプリケーションの予測・推薦を行うためには、これらのコンテキストを考慮する必要がある。

筆者らの研究グループでは、ユーザのスマートフォンの利用傾向を調査することを目的に、アプリケーションの利用履歴とユーザのコンテキスト情報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」[7] を運用している。こ

のアプリケーションでは、ユーザに一定時間毎にコンテキストに関する質問に直接回答してもらうことで現在の活動、同伴者、気分といったユーザの状態を直接表す高度なコンテキスト情報を収集している。このようなスマートフォン利用時の具体的なコンテキスト情報は、ユーザのアプリケーションの利用予測を行う際に有用な情報になると考えられる。

そこで本研究では、こんてきすとモンスターで得られたユーザのアプリケーション履歴とコンテキスト情報を活用し、ユーザのコンテキスト状況を考慮した利用アプリケーションの推定を行う。この際、コンテキスト情報とアプリケーション履歴の時系列性を考慮して、時系列データの予測に用いられる深層学習の 1 種である再帰的ニューラルネットワーク (Recurrent Neural Network, RNN) をアプリケーションの利用予測に用いる。なお、通常の RNN に比べ、時系列データの長期的な記憶が可能である LSTM (Long Short-Term Memory) をネットワーク構造に取り入れる。これは直近の利用パターンだけではなく、より過去の利用パターンも考慮して予測を行うことを目的としている。そして LSTM を取り入れた多層ニューラルネットワーク構造の予測モデルを構築し、予測精度に関して検証を行う。

深層学習を用いてアプリケーションの利用予測を行う場合、ユーザのログが少ない状態だと、学習のためのデータが少なく十分な予測精度を得ることができないことが問題点として挙げられる。そこで、ログが十分にあるユーザのデータから作られた予測モデルを用いて、ターゲットとなるユーザの利用アプリケーションを予測することが解決策として考えられる。

異なるユーザ同士が同じアプリケーションをインストール・

利用しているとは限らないため、本研究ではアプリケーションのカテゴリ分けを行い、ユーザが次回利用するアプリケーションのカテゴリを予測するモデルを作成する。そして、他のユーザのログを用いて作成したモデルを、ターゲットユーザに用いて精度の検証を行う。

以下、第2章では関連研究を紹介し、本研究との関連を述べる。第3章では収集データと利用アプリケーションの予測モデル、利用アプリケーションのカテゴリ予測の概要に関して述べ、第4章で評価実験について述べる。そして、第5章で評価実験での結果を基に考察を行い、最後に本研究のまとめと今後の課題について述べる。

2. 関連研究

本章ではまず、ユーザの利用アプリケーションの予測・推薦に関する研究を紹介し、次に再帰的ニューラルネットワークを用いた予測・推定に関する研究について紹介する。

2.1 利用アプリケーションの予測・推薦に関する研究

はじめに述べたように、スマートフォン向けに提供されるアプリケーションが爆発的に増加したことにより、ユーザが自身の目的に沿ったアプリケーションを見つけることが困難になりつつある。そこで、ユーザのアプリケーションの利用予測に関する研究が数多く行われている。[2], [9].

Shin ら [9] は、アプリケーションの利用履歴と、スマートフォンに搭載されたセンサから得られる場所、加速度、照度といった様々なコンテキスト情報を用いて、ベイジアンモデルを作成し、利用アプリケーションの予測を行った。

また利用アプリケーションの予測と関連した研究として、利用アプリケーションの推薦に関する研究も盛んに行なわれている。Girardello ら [3] はアプリケーション推薦システムとして「AppAware」を提案した。AppAware はアプリケーションの追加、削除、更新を監視することで、一般的にインストールされている数が多いアプリケーションの推薦をリアルタイムに行うシステムである。GPS による位置情報を用いることで、ユーザの現在地付近で、その時点で多くインストールされている人気のアプリケーションを推薦することが可能となる。

ここまで利用アプリケーションの予測・推薦に関する研究を紹介したが、これらの研究は、時刻や GPS による位置情報など、センサ等を用いて自動的に取得できる情報のみをコンテキスト情報として用いている。しかし、現在の活動や同伴者、気分といったセンサから自動的に推測することが困難なコンテキストも、アプリケーションの利用傾向に大きく影響を与えるものと考えられる。そこで本研究では、このようなコンテキスト情報を用いてアプリケーションの利用予測に関する取り組みを行った。

2.2 再帰的ニューラルネットワークを用いた予測・推定に関する研究

近年、機械学習の手法の1つである深層学習が注目を集めている [6]。深層学習は画像認識や言語処理、音声認識といった分野で特に成功を収めているが、それ以外の分野に関しても従来の機械学習で行っていたタスクに関して幅広く適用が可能であ

る。そして、時系列データに対しては、再帰的ニューラルネットワーク (Recurrent Neural Network, RNN) を用いる研究が優れた成果をあげている。ここから実際に RNN を予測や推定に関するタスクに適用した研究について紹介する。

近年、RNN を用いて市場の動向を予測する研究が幾つか行なわれている [1], [8]。Ayankoya ら [1] は、市場の穀物価格の時系列データに着目し、RNN をネットワーク構造に組み込んだ穀物の価格を推定するモデルを提案した。実際のリアルタイムの市場データを用いて穀物の価格を予測し、実験に協力した市場関係者の予測した価格よりも誤差が少ない予測に成功した。

RNN を用いたユーザの行動推定に関しても研究が行なわれている。Zhang ら [10] は、ユーザの属性や、広告の情報、広告が表示されている位置、ユーザの一連の操作の履歴から RNN を基にしたモデルを用いて、ページ内の広告がユーザによってクリックが実行されるかどうか推定を行った。

ここまで RNN を用いた予測や推定に関する研究を紹介し、実際に時系列データを用いた予測・推定に関して RNN が有用であることを確認した。そして、こんてきすとモンスターで収集したユーザのコンテキスト情報やアプリケーションの利用履歴の時系列性を考慮すると、アプリケーションの利用予測というタスクに対しても RNN を用いることは有効な方法であると考えられる。アプリケーション予測に関して様々な研究が行なわれているが、RNN を基にユーザの具体的なコンテキスト情報とアプリケーション履歴といった時系列情報を用いてアプリケーションの利用予測を行う研究はこれまでに行なわれていない。そこで本研究では、コンテキスト情報を用いたユーザの利用アプリケーションの予測というタスクに対して再帰的ニューラルネットワークを適用し、予測精度の確認を行った。

3. データ収集と利用アプリケーションの推定手法

本章ではまずログ収集システムである「こんてきすとモンスター」を用いたデータ収集に関して述べる。次に、ユーザの利用アプリケーションの推定モデルについて述べ、最後に情報の不足したユーザを対象としたユーザの利用アプリケーションのカテゴリ予測に関して述べる。

3.1 ログ収集システムを用いたデータ収集

筆者らの研究グループでは、アプリケーション利用ログ収集システム「こんてきすとモンスター」[7] の運用によって、ユーザのアプリケーション利用履歴とコンテキスト情報を収集している。アプリケーションの利用履歴については、画面の最前面にあるアプリケーション ID が変化することをアプリケーションの起動とみなして、そのときのアプリケーション ID と利用時刻をアプリケーション履歴として収集する。ユーザのコンテキスト情報については、ユーザに対して活動や同伴者、気分といった身の回りの状況に関する質問を、有効時間を設けて定期的に行うことで取得する。表1に各コンテキスト項目についての詳細を示す。これらの取得情報をアプリケーションの利用予測に用いることにした。

表 1 コンテキスト項目と選択群, 種類数, 有効時間

項目 (質問文)	選択項目群	種類数	有効時間
今日 (今日は何をするか)	仕事, 休み, 旅行など	100 種類	24 時間
活動 (何をしているか)	仕事, 勉強, 食事, 音楽など	189 種類	2 時間
同伴者の人数 (何人といるか)	1 人, 2 人, たくさんなど	6 種類	2 時間
同伴者の分類 (誰といるか)	家族, 恋人, 同僚など	57 種類	2 時間
場所 (どこにいるか)	学校, 仕事先, 観光地など	150 種類	3 時間
気分 (気分はどうか)	良い, 悪いなど	4 種類	4 時間
体調 (体調はどうか)	良い, 悪いなど	4 種類	4 時間
疲労度 (疲れているか)	元気, 疲れているなど	4 種類	4 時間
忙しさ (忙しいか)	ゆったり, 忙しいなど	4 種類	4 時間

3.2 利用アプリケーションの推定モデルの設計

3.2.1 RNN の概要と LSTM の活用

ユーザがスマートフォンを利用していた時のコンテキスト情報は、アプリケーションの利用予測を行う上で有用な情報となり得る。しかし、3.1 節で紹介した今回のアプリケーション予測の際に用いるユーザのコンテキスト情報は全部で 9 項目 518 種類あり、これら多数のコンテキスト情報とアプリケーションの利用パターンとの複雑な関係性を人手で抽出して予測の際の特徴として用いることが難しい。そこで今回のアプリケーションの利用予測に関して、人手を介さずに自動で特徴抽出が可能な機械学習の方法の 1 つである深層学習を用いることにした。

予測を行う上で、ユーザのアプリケーションの利用傾向に関しても考慮すべき点がある。短期的にはユーザの状況によってアプリケーションの利用傾向が変化するが、長期的に見ると、習慣的なアプリケーションの利用パターンが存在する [5]。アプリケーションの利用予測を行う際に、このような短期的な利用傾向と長期的な利用傾向に対処しなければならない。このような時系列データの短期的なパターンと長期的なパターンに対応できる深層学習の手法の 1 つとして再帰的ニューラルネットワーク (Recurrent Neural Network, RNN) が挙げられる。

RNN の入力層では、動画像や音声などの時系列データが、時刻 $t=1$ から時系列に従って入力される。時刻 t の中間層には、時刻 t の入力層の出力と時刻 $t-1$ の中間層の出力が入力される。このように、連鎖的にデータを入力し、時刻 $t-1$ の中間層の出力を時刻 t の中間層に入力することで、時系列に対応した認識が可能となる。

通常の RNN ではネットワークパラメータの更新に出力層から入力層に遡って誤差を伝播させるアルゴリズムを用いるが、中間ユニットが多いと入力層まで誤差を伝播させるまでに多くの勾配計算を行う必要があり誤差の計算が不安定となる。これは勾配消失問題と呼ばれる。これにより、長期間の学習が通常の RNN では不可能であったが、その問題を解決する LSTM (Long Short-Term Memory) [4] という RNN を拡張したニューラルネットワークが提案された。LSTM では必要な誤差だけを伝播させるように入力と出力を制限する構造を持っている。今回はこの時系列データの長期的な学習が可能である LSTM を利用アプリケーション推定に用いることにした。

3.2.2 アプリケーションの利用予測モデルの概要

本項ではアプリケーションの利用予測モデルについて説明する。なお図 1 はアプリケーションの利用予測モデルの構造で

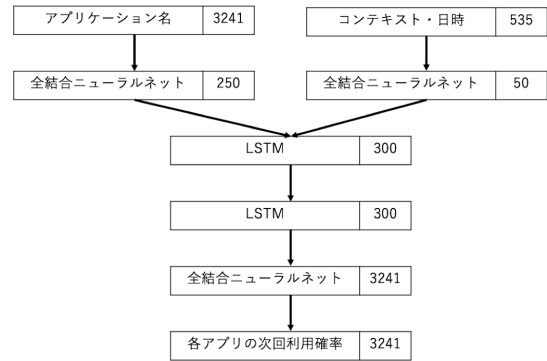


図 1 アプリケーションの利用予測モデルの構造

あり、ある 1 時刻分のアプリケーション利用履歴とコンテキスト・日時に関する情報を予測モデルに入力し、最終的に各アプリケーションの次回利用確率を得る。なお図中に現れる数字は各層の出力次元を表す。以下アプリケーション利用予測モデルの詳細を説明する。

まず利用アプリケーションを表す入力について説明する。これまでと異なるユーザが使用したアプリケーションは全 3241 種類であった。そこで 3241 種類のアプリケーションに関して、それぞれ 1 から 3241 までのアプリケーション ID をつけた。そして、アプリケーションを表す入力ベクトルとして、アプリケーションが該当する要素が 1、残りの要素が 0 となる 3241 次元のベクトルを用いることにした。

次にアプリケーション利用時のコンテキスト情報・日時を表す入力について説明する。複数のコンテキスト情報を単一のベクトルで表現するためにコンテキスト ID を導入した。コンテキスト ID について 1 から 518 まで用意し、「今日」の項目のコンテキストは 1 から 100、「活動」のコンテキストは 101 から 289 といったように、9 項目 518 種類のコンテキストに対して ID を付与する。そしてコンテキスト情報に関して、ある項目 (ここでは c とする) のコンテキスト情報入力から経過時間を表現する値 t_c を入力に導入する。項目 c のコンテキストについて最新のコンテキスト情報の入力からの経過時間を $t_{c_{in}}$ 、そのコンテキストの項目を有効時間を $t_{c_{val}}$ とすると、 t_c は以下のように定義される。

$$t_c = 1 - \min \left\{ \frac{t_{c_{in}}}{t_{c_{val}}}, 1 \right\} \quad (1)$$

この t_c を各コンテキスト 9 項目に対して用意する。

日時に関する情報については、時間帯と平日休日の区別に関する情報を入力に用いる。時間帯に関しては 1 日を 4 時間毎に分割した 6 つの時間帯で表現した。そして平日休日の区別に関する情報は、現在の曜日に関して、月曜日～金曜日であれば平日、土曜日、日曜日であれば休日として表現した。以上からコンテキスト情報に関してはコンテキスト入力経過時間の情報も含めて 527 次元、利用時間帯は 6 次元、平日休日の区別に関する情報は 2 次元の計 535 次元のベクトルをアプリケーション利用時のコンテキスト・日時に関する情報とした。

ここまで入力情報について説明してきたが、この入力をもとに LSTM を用いたニューラルネットワークで学習を行い、アプリケーションの利用予測モデルを生成する。

本項のはじめにも触れた、アプリケーション利用予測モデルのネットワーク構造は図 1 のようになる。ネットワークは LSTM 層を含む多層ニューラルネットワークとなっている。入力層に関しては、利用アプリケーション名を入れる全結合層と、コンテキストと日時の情報を入れる全結合層がある。利用アプリケーションの情報を入力する層は 250 次元の情報を出力する。またコンテキスト・日時の情報を入力する層は 50 次元の情報を出力する。2 つの出力情報に関して \tanh 関数を活性化関数として通し、この出力を 300 次元の情報として結合する。そして、この 300 次元の情報を、次の LSTM 層の入力として使用する。なお LSTM 層は 2 層重ねた構造となっており 1 層目の出力が 2 層目の入力となる。また、1 つの LSTM 層のユニット数は 300 である。LSTM 層から出力された 300 次元の情報を全結合層に入力し、アプリケーション種類数である 3241 次元の情報を出力する。最後にこの出力を活性化関数である softmax 関数に通すことで、各アプリケーションが次に利用される確率が出力される。出力されたベクトルの各要素はその要素が該当するアプリケーションが次に利用される確率を表している。

予測モデルの学習に関して、各層の重み W とバイアス b の値を訓練データの学習によって調整する。重みやバイアスの更新のために用いる損失関数として、ターゲットのアプリケーションが次回利用される確率 P_{target} の対数の負の平均をとったものを使用する。そして、損失関数 $loss$ は以下のように定義される。なお N はアプリケーションの全履歴数を表す。

$$loss = -\frac{1}{N} \sum_{i=1}^N \ln P_{target_i} \quad (2)$$

この損失関数が最小となるように学習を行う。なお学習には、関数の傾きから関数の最小値を探索する最適化アルゴリズムの 1 つである最急降下法を用いる。また学習パラメータの更新量を定める学習係数に関しては、学習の初期は大きな値を選び、学習の進捗によって小さくする方法が一般的に用いられる。そこで訓練データを用いた予測モデルの学習に関して全 26 回行い、初めの 6 回の学習では学習係数を 1.0 とし、その後の学習では減衰率を 0.8 として徐々に学習係数が小さくなるように設定した。

3.3 利用アプリケーションのカテゴリ予測の概要

本節では、ログが不足したユーザを想定した他のユーザのデータを用いた利用アプリケーションのカテゴリの予測の概要に関して述べる。

3.3.1 アプリケーションのカテゴリ予測の意義

ログが不十分なユーザに関しては、予測モデルの学習をうまく行うことができず予測精度が低下する。情報が不足している状態で、ユーザのアプリケーションの利用予測を行う方法としては、他のユーザのデータから作成した予測モデルを利用して予測を行うことが方法として考えられる。しかし、ターゲットとなる情報が不足しているユーザと、予測モデルを生成した

表 2 アプリケーションのカテゴリ分けと利用率

大カテゴリ	小カテゴリ	比率	
通信	メッセージ	0.1956	
	通話	0.0237	
	メール	0.1298	
SNS	SNS	0.2071	
	出会い	0.0000	
ブラウザ	ブラウザ	0.1072	
ツール	ツール	0.0475	
	アート&デザイン	0.0000	
	ライブ&ライブ	0.0001	
	カスタマイズ	0.0048	
システム	システム	0.0581	
ゲーム	ゲーム	0.0424	
情報	地図&ナビ	0.0110	
	旅行&地域	0.0002	
	自動車	0.0000	
	天気	0.0013	
	ライフスタイル	0.0006	
	健康&フィットネス	0.0024	
	住まい&インテリア	0.0000	
	美容	0.0000	
	医療	0.0000	
	出産&育児	0.0000	
	イベント	0.0001	
	フード&ドリンク	0.0011	
	スポーツ	0.0007	
	ショッピング	0.0008	
	ニュース&雑誌	0.0171	
	仕事&勉強	ビジネス	0.0015
		ファイナンス	0.0023
		仕事効率化	0.0461
		書籍&参考書	0.0019
		教育	0.0016
写真	写真	0.0381	
	音楽&オーディオ	0.0123	
マルチメディア	動画&プレイヤー&エディタ	0.0205	
	エンタメ	0.0205	
娯楽	コミック	0.0036	

ユーザの利用しているアプリケーションは異なっているものも多く、そのまま利用アプリケーションを予測する場合は、予測精度が悪くなることが考えられる。そこで、アプリケーションに関して「ゲーム」、「SNS」といったカテゴリ分けを行い、カテゴリを予測するモデルをログの不足したユーザに適用することが解決策として考えられる。

適切なアプリケーションのカテゴリ分けを行い、一定の精度が保たれた利用アプリケーションのカテゴリ予測を行うことができれば、ログの不足したユーザに対しての予測の補助として、利用アプリケーションのカテゴリ予測は意義があると考えられる。

3.3.2 アプリケーションのカテゴリ分け

まず Android OS のアプリケーションマーケットである Google Play に記載されているアプリケーションカテゴリを基にして 37 種類のアプリケーションのカテゴリを定義し、これを基準のカテゴリ（小カテゴリ）とした。またカテゴリの粒度を変えて予測精度を調査するために、基準のカテゴリをさらに統合し 11 種類のカテゴリをもつ大カテゴリを定義した。表 2 はカテゴリ分けの結果とそれぞれのカテゴリの利用率を示したものである。この 2 つのカテゴリ分けを基に、他のユーザデータを用いた利用アプリケーションのカテゴリ予測を行う。

3.3.3 他のユーザのデータを用いたカテゴリ予測の概要

他のユーザのデータを用いたアプリケーションのカテゴリ予測について述べる。予測モデルを生成するのに十分なデータがあるユーザが N 人いるとする。まず 3.3.2 項で定義した 2 種類のカテゴリ分けから、 N 人のそれぞれのデータを基にカテゴリ予測を行うモデルを N 個用意する。なおモデルを作成する際に入力用いるコンテキスト情報について、選択肢の多い項目のコンテキストについては、ユーザによって入力しているコンテキストが、かなり異なることが考えられるため、「今日」、「活動」、「場所」、「同伴者の種類」の項目のコンテキスト情報に関して表 3 のようにカテゴリ分けを行い、コンテキストがより

表3 カテゴリ分けしたコンテキスト情報

項目	内容
今日 (今日は何をするか)	仕事・勉強, 休暇
活動 (何をしているか)	遊び・娯楽, 日常生活, 勉強・仕事
場所 (どこにいるか)	自宅, 日常, 外出, 移動
同伴者の種類 (誰といるか)	親しい人, 仕事

広義なものになるようにした。

そして、アプリケーション履歴とコンテキスト情報が不足しているユーザがいると仮定する。その情報の不足したユーザのデータを予め生成された N 人の予測モデルに適用し、予測精度を調査する。そして、予測精度の良かった上位 k 人のデータを用いて新たに予測モデルを生成し、情報の不足したユーザのカテゴリ予測を行う。

この概要を基に、4章で他のユーザのデータを用いた場合の利用カテゴリの予測精度の調査を行う。

4. 評価実験

本章では、3.2.2項で述べた予測モデルを基に、ユーザのアプリケーションの利用予測を行った結果について述べる。また他のユーザのデータを用いた場合のカテゴリ予測に関する実験結果についても述べる。

4.1 アプリケーションの利用予測

本節ではユーザのアプリケーションの利用予測に関して評価手法と実験結果を述べる。

4.1.1 評価手法

アプリケーションの履歴数が5,000件以上のユーザ25名のデータを今回の評価実験で用いることにした。なお選定ユーザ25名の履歴数の平均値は8189.9回、中央値は7145.0回、コンテキストの入力回数の平均値は1511.2回、中央値は1525.0回、こんてきすとモンスターの継続日数は、平均値が57.9日、中央値は53.3日であった。そして各ユーザの履歴に関して、前半期間と後半期間でデータを均等に分割した。そして前半期間のデータを訓練データとし、後半期間のデータをテストデータとして用いて予測精度の評価を行った。

なお評価実験では、各ユーザの推定モデルが出力した次回利用される確率の高い k 件のアプリケーション候補の中に、実際に利用されたアプリケーションが含まれていた場合に、利用アプリケーションの予測が成功したとみなす。そして評価指標として、正解率を以下のように定義する。

$$\text{正解率} = \frac{\text{予測が成功した回数}}{\text{テストデータのアプリケーション利用回数}}$$

訓練データとテストデータに関して、コンテキスト情報に関する要素は全て0に変更したコンテキスト情報を用いない場合と利用率上位 k 件並べる Most Frequently Used (MUF) 法による予測を比較対象とする。

4.1.2 利用アプリケーション予測の正解率

本項では利用アプリケーションの予測結果について述べる。図2は予測されるアプリケーション利用候補の数を変えた場合の、それぞれのユーザ25名に関して、コンテキスト情報を使用した予測モデルとコンテキスト情報を使用しなかった予測モ

デルの正解率の平均と、MFU法による正解率の平均をまとめたものである。

図2から、コンテキスト情報の有無に関わらず予測モデルの正解率は、どの予測候補数においても、MFU法より高い値を示しており、単純に利用頻度の高いアプリケーションを予測する場合と比較すると高い精度が出ていることが分かる。そして、予測モデルの正解率に関して、コンテキスト情報を用いて学習した予測モデルの方が、コンテキスト情報を用いない場合に比べて、どの予測候補数でも正解率が高いことが分かる。またコンテキスト情報を使用したモデルの正解率とコンテキスト情報を使用しないモデルの正解率の差は、予測候補数を増やすに連れて小さくなることも分かる。

4.1.3 カテゴリ別の正解率

次にアプリケーションをカテゴリ別に分類し、上位1件を予測する場合のカテゴリ別の正解率と利用率を比較した。図3の各カテゴリの左の棒グラフは上位1件を予測する場合のコンテキスト情報を使用しない予測モデルの正解率とコンテキスト情報を使用した場合の正解率との差を積み上げた棒グラフとなる。ただし、コンテキスト情報を用いることで正解率が低下した場合は、コンテキスト情報を用いない場合の正解率を表す棒グラフの下部にコンテキスト情報を用いた場合との正解率の差を表示している。また各カテゴリでの右の棒グラフは各カテゴリの利用率を示したものである。

図3より、「SNS」以外のカテゴリでコンテキスト情報を用いた予測モデルの方が、コンテキスト情報を用いない予測モデルに比べて正解率が高いことが分かる。「仕事効率化」、「写真」、「動画プレイヤー&エディタ」、「音楽&オーディオ」といったカテゴリはコンテキスト情報がないモデルの正解率に比べて、コンテキスト情報を用いた場合の正解率の向上度合いが大きく、コンテキスト情報とアプリケーションの利用傾向に関するアプリケーションが多いカテゴリ群であることが伺える。逆に「ニュース&雑誌」は、利用率に比べて、コンテキスト情報がないモデルの正解率がかなり高く、コンテキスト情報を用いた場合の正解率の向上度合いが小さいことが分かる。このことから「ニュース&雑誌」のカテゴリのアプリケーションは時系列のアプリケーション利用履歴だけでも予測を行いやすいことが伺える。

次に図3と同様な情報を上位5件を予測する場合で算出したグラフを図4で示す。図4から、上位5件を予測する場合に関しては全てのカテゴリでコンテキスト情報を用いた場合の正解率の方が高いことが分かる。また、「通信」、「SNS」、「ブラウザ」といった比較的利用率の高いカテゴリについては、上位5件の時点でコンテキスト情報がないモデルの正解率がかかなり高く、コンテキスト情報を付与した場合の正解率の向上度合いが他のカテゴリに比べると低いことが分かる。

利用率の低いカテゴリのアプリケーションに関しては、上位1件を予測する場合と比較すると度合いは低くなるものの、コンテキストを付与した場合の一定の正解率向上の効果が見られ、特に「移動」や「勉強」といったカテゴリでコンテキスト情報による正解率向上の効果がみられる。ただし、「カスタマイ

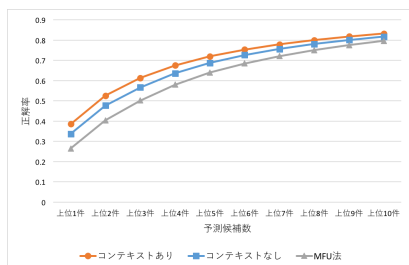


図 2 予測アプリケーション候補数を変化させた場合の正解率

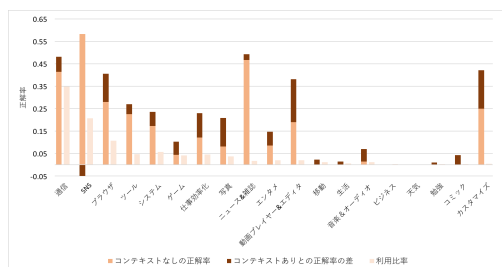


図 3 カテゴリ別のコンテキスト情報の有無による正解率と利用率 (上位 1 件)

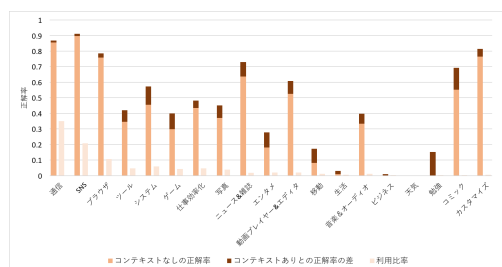


図 4 カテゴリ別のコンテキスト情報の有無による正解率と利用率 (上位 5 件)

ズ」に関しては利用率は低いものの、上位 5 件を予測する場合にはコンテキスト情報を用いない場合でも正解率がかなり高くコンテキスト情報による正解率の向上度合いが低い。

利用頻度が高いカテゴリのアプリケーションに比べて、利用頻度が低いカテゴリのアプリケーションの方がコンテキスト情報による正解率の向上度合いが大きい傾向がある理由として、「通信」、「SNS」、「ブラウザ」といったカテゴリのアプリケーションはユーザのコンテキスト状況に関係なく常に使われる傾向があるが、より利用頻度の少ないアプリケーションに関しては利用傾向がユーザのコンテキストによって影響を受けやすいことが原因として考えられる。

4.1.4 コンテキスト項目別の正解率

次に、各コンテキスト情報の予測精度への影響を調査した。今回対象としたコンテキストは、こんてきすとモンスターの取得方法でしか得ることが困難であると想定される「活動」、「同伴者（種類・人数）」、「主観情報（気分・健康・疲労度・忙しさ）」とした。訓練データとテストデータに関して該当のコンテキスト情報以外の要素を 0 とし、該当コンテキスト以外はコンテキスト情報がないものとした。そして、「活動」、「同伴者」、「主観情報」の 3 つの項目に関して、予測モデルを作成し、正

表 4 予測候補数を変化させた場合のコンテキスト項目別の平均正解率

コンテキスト項目	上位 1 件	上位 5 件	上位 10 件	上位 15 件
全て	0.3858	0.7205	0.8329	0.8859
活動	0.3685	0.7092	0.8290	0.8853
同伴者	0.3614	0.7027	0.8235	0.8817
主観情報	0.3548	0.6990	0.8225	0.8800
なし	0.3372	0.6880	0.8169	0.8766

解率を算出した。表 4 は「活動」、「同伴者」、「主観情報」のコンテキストのみを用いた正解率と全てのコンテキストを使用した場合、全くコンテキストを使用しなかった場合の正解率をまとめたものである。どの予測候補数でも、全てのコンテキストを用いた場合が最も正解率が高いことが分かる。「活動」、「同伴者」、「主観情報」のみをそれぞれ用いた場合に関して、全くコンテキストを使用していない場合に比べると、どの予測候補数でも正解率が向上していることが分かる。またコンテキスト項目間の正解率を比較した場合、どの予測候補数に関しても、「活動」の項目が最も正解率が高く、次に高い項目は「同伴者」の項目であり、最も正解率が低かったのは「主観情報」の項目であった。

4.2 他のユーザのデータを用いた利用カテゴリ予測

本節では、3.3 節で述べた他のユーザのデータを用いたアプリケーションのカテゴリ予測に関する評価実験とその結果について述べる。

4.2.1 評価手法

評価実験に用いるデータは 4.1.1 項で選定した 25 名のユーザのデータとする。あるユーザに関してログが不足しており、残り 24 名のユーザのデータから利用カテゴリの予測を行う状況を仮定する。まず、自分以外のユーザ 24 名のカテゴリ予測モデルに対して、テストデータの中で、期間初期のアプリケーション利用履歴とコンテキスト情報・日時がセットになった 500 件のデータを入力しそれぞれ正解率を得る。

予測候補の数を上位 1 件から上位 3 件までで変化させた場合の正解率の平均が高い上位 k 人のユーザを選定する。次にその k 人のユーザの訓練データを統合して 1 つの訓練データとする。これを基に学習を行い予測モデルを作成する。複数のユーザデータを用いて作成したモデルに対し、対象ユーザのテストデータの 501 件目以降を用いて正解率を算出する。

4.2.2 実験結果

他のユーザのデータを用いた利用カテゴリの予測結果について述べる。訓練データに用いるユーザの数を 1 名、3 名、5 名、7 名と変化させた場合の予測候補数 5 件までの正解率の平均と自身のデータを用いて作成した予測モデルの正解率の平均を大カテゴリについては表 5、小カテゴリについては表 6 に示す。2 つのカテゴリ分けに共通することとして、1 名のユーザのデータを用いた予測モデルが最も正解率が低く、使用するユーザの数を増加させる毎に正解率が向上し、自身のモデルの正解率に近づくことが分かる。しかし、用いるユーザを 5 名から 7 名に増加させた場合は正解率が低下、または、ほぼ変化しないことが分かる。このことから複数のユーザのデータを用いることで正解率は向上するが、一定の数を超えると、正解率は向上しなくなる事が分かる。

表5 訓練データを変えた場合の利用カテゴリの予測結果（大カテゴリ）

用いるユーザ	上位 1 件	上位 2 件	上位 3 件	上位 4 件	上位 5 件
1 名	0.4010	0.5886	0.6833	0.7454	0.7886
3 名	0.4267	0.6241	0.7271	0.7870	0.8299
5 名	0.4399	0.6374	0.7404	0.8017	0.8445
7 名	0.4316	0.6228	0.7223	0.7858	0.8314
ユーザ自身	0.4729	0.6657	0.7631	0.8292	0.8756

表6 訓練データを変えた場合の利用カテゴリの予測結果（小カテゴリ）

用いるユーザ	上位 1 件	上位 2 件	上位 3 件	上位 4 件	上位 5 件
1 名	0.3534	0.5174	0.6101	0.6708	0.7157
3 名	0.3777	0.5534	0.6530	0.7201	0.7646
5 名	0.3897	0.5651	0.6601	0.7284	0.7760
7 名	0.3849	0.5647	0.6629	0.7296	0.7758
自身のモデル	0.4382	0.6101	0.7127	0.7830	0.8304

小カテゴリに比べて大カテゴリの方が、5名のデータを使用した時の正解率と自身のモデルを用いた時の正解率の差が小さく、他のユーザのデータを用いた場合でも一定の精度を保って予測を行えることが分かる。

次に、上位1件を予測する場合について訓練データに用いるユーザの数を変化させた時のカテゴリ別の正解率を算出した。まず大カテゴリについて、上位1件を予測する場合の結果を表7に示す。自身のデータを用いたモデルの正解率が最も高いカテゴリは6個であった。自身のデータを用いたモデル以外の正解率の方が高いカテゴリも幾つかあり、「通信」のカテゴリは1名のモデル、「情報」、「写真」のカテゴリは5名のモデル、「SNS」、「システム」のカテゴリは7名のモデルの正解率が最も高かった。カテゴリ別の正解率の平均を見ると、自身のモデルを除いた場合は、5名の場合の正解率が最も高い。

次に小カテゴリについて、上位1件を予測する場合の結果を表8に示す。なお、どのモデルにおいても正解率が0%のカテゴリは表8から除外している。自身のデータを用いたモデルの正解率が最も高いカテゴリは15個あった。5名のユーザを用いた場合の正解率に比べ、自身のモデルを用いた場合の正解率が10%以上高いカテゴリは、「ニュース&雑誌」、「エンタメ」、「動画プレイヤー&エディタ」、「コミック」、「カスタマイズ」の5つのカテゴリであった。また大カテゴリと同様に、自身のモデルを用いるよりも他のユーザのモデルを用いた方が正解率の高いカテゴリが幾つか見られ、「メッセージ」は1名のモデル、「SNS」、「システム」は7名のモデル、「地図&ナビ」は5名のモデルの正解率が最も高かった。表中のカテゴリに限定して、カテゴリ別の正解率の平均を見ると、大カテゴリと同じく、自身のモデルを除いた場合は、5名の場合の正解率が最も高い。

2つのカテゴリ分けに関して、ユーザを1名から複数人に増やすことで全体の正解率と同様に、多くのカテゴリの正解率が向上していることが分かる。また自身のモデルを用いる場合と他のユーザを用いる場合（5名）のカテゴリ別の正解率の差の平均は、小カテゴリは8.79%であったのに対し、大カテゴリの正解率の差の平均は2.88%であった。このことから、大カテゴリの方が自身のモデルとのカテゴリ別の正解率の差が平均して小さいことが分かる。

小カテゴリに関してカテゴリの利用比率と、自身のモデルと

表7 他のユーザデータを用いたモデルのカテゴリ別の上位1件の正解率（大カテゴリ）

カテゴリ	1名のモデル	3名のモデル	5名のモデル	7名のモデル	自身のモデル
通信	0.7484	0.7115	0.7016	0.6875	0.7261
SNS	0.4239	0.4962	0.5298	0.5482	0.5137
ブラウザ	0.3153	0.3744	0.3874	0.4052	0.4476
ツール	0.1203	0.1857	0.2837	0.1663	0.3510
システム	0.1113	0.1484	0.1506	0.1946	0.1853
ゲーム	0.1457	0.2698	0.2743	0.2698	0.3150
情報	0.2424	0.2405	0.3125	0.3072	0.2668
仕事・勉強	0.1619	0.1883	0.2124	0.1854	0.2280
写真	0.0793	0.2956	0.3104	0.2662	0.2367
マルチメディア	0.0679	0.1420	0.1431	0.1431	0.2384
娯楽	0.0072	0.0370	0.0578	0.0253	0.1722
平均	0.2203	0.2809	0.3058	0.2908	0.3346

表8 他のユーザデータを用いたモデルのカテゴリ別の上位1件の正解率（小カテゴリ）

カテゴリ	1名のモデル	3名のモデル	5名のモデル	7名のモデル	自身のモデル
メッセージ	0.6021	0.5700	0.5644	0.5669	0.5983
通話	0.1131	0.1644	0.1568	0.1497	0.1973
メール	0.3251	0.3251	0.3489	0.3391	0.3844
SNS	0.6400	0.6846	0.6819	0.6963	0.6810
ブラウザ	0.3066	0.3821	0.3992	0.3508	0.4762
ツール	0.1746	0.2647	0.2701	0.2694	0.3498
システム	0.2090	0.1668	0.2251	0.2350	0.2108
ゲーム	0.2345	0.3452	0.3492	0.3475	0.3539
仕事効率化	0.0586	0.2028	0.2151	0.2393	0.2772
写真	0.2433	0.3049	0.3212	0.3040	0.3277
ニュース&雑誌	0.2941	0.3018	0.3466	0.3515	0.4677
エンタメ	0.0225	0.0332	0.0332	0.0421	0.1336
動画&エディタ	0.0086	0.0838	0.1217	0.1332	0.3501
地図&ナビ	0.0254	0.0342	0.0949	0.0795	0.0734
音楽&オーディオ	0.0261	0.0401	0.0321	0.0331	0.1110
コミック	0.0000	0.0000	0.0000	0.0000	0.3052
カスタマイズ	0.0000	0.0021	0.0021	0.0000	0.4444
健康&フィットネス	0.0000	0.0000	0.0000	0.0000	0.0359
書籍&参考書	0.0000	0.0000	0.0000	0.0000	0.0552
平均	0.1760	0.2056	0.2191	0.2177	0.3070

他のユーザデータを用いたモデル（5名）との正解率の差の相関係数を見たところ、 -0.4070 という値を得た。これは、カテゴリの利用比率が低いほど正解率の差が大きくなる傾向、すなわち、カテゴリに属するアプリケーションの利用比率が低いほど自身のモデルと比較して正解率が低下する傾向があることを示している。同様に大カテゴリに関しても、同じように相関係数を算出したところ、 -0.1319 という値を得た。小カテゴリと同様の相関が見られるが程度としてはかなり低いことが分かる。

以上から他のユーザデータを使用して予測を行う場合、カテゴリ別での結果に着目した場合でも、大カテゴリの方が精度を保って予測を行うことができることが判明した。

5. 考察

本章では、評価実験の結果を基に考察を行い、今後の課題を述べる。

5.1 コンテキスト情報の選択

4.1.4項でコンテキスト情報を組み合わせることで、予測精度が向上することが確認できたが、コンテキスト項目別で見た場合、主観情報に関するコンテキストに比べて同伴者や活動のコンテキスト情報を用いた方が予測精度が向上した。この結果から、コンテキストの項目間でアプリケーションの利用予測に対する効果の違いが確認できた。予測により効果のあるコンテキスト項目を発見し、コンテキストの入力項目を減らした場合でも予測精度を維持することができれば、ユーザの入力負担を軽減することができる。

また現在取得しているコンテキストに関してもセンサ情報で

代替できそうなコンテキストが存在する。例えば場所のコンテキストは GPS データ、健康のコンテキストに関しては歩数計や心拍数計のデータを用いるといったことが考えられる。そのようなセンサデータを代わりに使用した場合に精度がどの程度出るのかを同時に検証する必要があると考える。

5.2 他のユーザデータを用いた場合のカテゴリ予測の精度

4.2.2 項で他のユーザデータを用いてカテゴリ予測を行うときに、自身のモデルに対しほぼ精度を下げずに予測できるカテゴリと、予測精度が下がるカテゴリが存在した。

小カテゴリに関して、5名のデータを使用した予測モデルの正解率に着目した場合、「メッセージ」、「メール」といったカテゴリの正解率は自身のモデルの正解率と比較して90%以上の正解率であったが、「エンタメ」、「音楽&オーディオ」といったカテゴリは自身のモデルの正解率と比較して50%以下の正解率であった。

小カテゴリに関して、各カテゴリに平均して51.85種類のアプリケーションが属するが、利用率の比較的高い「メッセージ」や「メール」に属するアプリケーションがそれぞれ11種類と17種類であり利用率に対してアプリケーションの種類が少ないのに対し、利用率が比較的低い「エンタメ」や「音楽&オーディオ」にはそれぞれ68種類と63種類のアプリケーションが属しており、利用率に対してアプリケーションの種類が多い。「エンタメ」や「音楽&オーディオ」としてカテゴリにまとめたときに、「メッセージ」や「メール」のカテゴリに比べてユーザの利用傾向にばらつきがあることが伺える。このような、利用頻度が低くユーザの利用傾向が異なるカテゴリのアプリケーションに関しては、自身のモデルと同等の精度を保つて予測を行うことが難しいと考えられる。

カテゴリ数を少なくし、1つのカテゴリの利用率を高めることが対策として考えられるが、ユーザビリティの観点からは好ましいとは言えない。そこで、ユーザの利用傾向が類似しているアプリケーションを調査し、利用傾向が類似しているアプリケーション群を新たなカテゴリとして定義し直し、他のユーザデータを用いたカテゴリ予測に適用することが対策として考えられる。

6. おわりに

本研究では、ユーザのスマートフォン使用履歴を収集するシステムである「こんてきすとモンスター」で収集したアプリケーション利用履歴とコンテキスト情報を活用したユーザの利用アプリケーションの予測を行った。

アプリケーション利用履歴の時系列性を考慮して、時系列データの長期的な記憶が可能である再帰的ニューラルネットワークの1種であるLSTM (Long Short-Term Memory) を用いてネットワークを構築し、ユーザのアプリケーション履歴とコンテキスト情報を入力として利用アプリケーションの予測を行った。その結果、コンテキスト情報を用いた予測モデルの方が、コンテキスト情報がない予測モデルに比べて正解率が高いことが分かり、特に利用頻度が低いアプリケーションで正解率が向上することが判明した。

また本研究では、履歴の不足したユーザのアプリケーションの利用予測を目的として、他のユーザのデータを用いた利用アプリケーションのカテゴリ予測を同時に行った。その結果、単独のユーザのデータを用いるよりは複数人のデータを用いることで予測精度が向上し、自身のデータを用いたモデルを使用する場合の予測精度に近づくことが判明した。また、カテゴリ数の少ないカテゴリ分けによる予測の方がより自身のモデルの正解率に近づくことも判明した。

今後の課題としては、センサ情報と高度なコンテキスト情報を組み合わせたアプリケーションの利用予測の精度の調査や他のユーザデータを用いた予測により適したカテゴリ分けの検討などが挙げられる。

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)、文部科学省国家課題対応型研究開発推進事業 - 次世代 IT 基盤構築のための研究開発 - 「社会システム・サービスの最適化のための IT 統合システムの構築」、マイクロソフト CORE プロジェクト (CORE 11) の研究助成によるものである。ここに記して謝意を表す。

文 献

- [1] K. Ayankoya, A.P. Calitz, and J.H. Greyling, "Using Neural Networks for Predicting Futures Contract Prices of White Maize in South Africa," Proc. the Annual Conf. of the South African Institute of Computer Scientists and Information Technologists, vol.3, pp.1-10, 2016.
- [2] R. Baeza-yates, D. Jiang, and B. Harrison, "Predicting The Next App That You Are Going To Use," Proc. ACM Int'l Conf. on Web Search and Data Mining, pp.285-294, 2015.
- [3] A. Girardello and F. Michahelles, "AppAware: Which Mobile Applications Are Hot?," Proc. Int'l Conf. on Human Computer Interaction with Mobile Devices and Services, pp.431-434, 2010.
- [4] S. Hochreiter and J.J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol.9, no.8, pp.1-32, 1997.
- [5] S.L. Jones, D. Ferreira, S. Hosio, J. Goncalves, and V. Kostakos, "Revisitation Analysis of Smartphone App Use," Proc. Int'l Conf. on Ubiquitous Computing, pp.1197-1208, 2015.
- [6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol.521, no.7553, pp.43-444, 2015.
- [7] 大澤 純, 岩田麻祐, 原 隆浩, 西尾章治郎, "ユーザの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," データ工学と情報マネジメントに関するフォーラム (DEIM), 2013.
- [8] A.M. Rather, A. Agarwal, and S. V.N, "Recurrent neural network and a hybrid model for prediction of stock returns," Expert System with Application, vol.42, pp.3234-3241, 2015.
- [9] C. Shin, J.H. Hong, and A.K. Dey, "Understanding and Prediction of Mobile Application Usage for Smart Phones," Proc. Int'l Conf. on Pervasive and Ubiquitous Computing, pp.173-182, 2012.
- [10] Y. Zhang, H. Dai, C. Xu, J. Feng, T. Wang, J. Bian, B. Wang, and T.Y. Liu, "Sequential Click Prediction for Sponsored Search with Recurrent Neural Networks," Association for the Advancement of Artificial Intelligence, pp.1369-1375, 2014.