

携帯電話における長期的操作履歴の解析と状況適応型操作予測技術の検討 Long-term Data Analysis and Operation Prediction for Context-Aware UI of Mobile Phones

上坂 大輔[†] 岩本 健嗣[‡] 村松 茂樹[†] 横山 浩之[†]
Daisuke Kamisaka[†] Takeshi Iwamoto[‡] Shigeki Muramatsu[†] Hiroyuki Yokoyama[†]

1. はじめに

通話やメールなどの基本的なコミュニケーション手段だけに留まらず、今やカメラや音楽プレーヤー、ナビゲーションシステムなど、多様なアプリケーションやサービスを提供し、人々の日常に欠かせないツールとなっている携帯電話は、様々な属性（年齢、性別、職業など）や目的、ライフスタイルを持ったユーザに、様々な状況下で利用される。その一方で、既存の携帯電話のユーザインタフェース（UI）の多くは、階層の深いメニューのような煩雑なものになりがちであり、ユーザはしばしば、使いたいアプリケーションに使いたいタイミングで素早くアクセスすることができない状況に陥ってしまう。

これまでも、ユーザが目的とするアプリケーションへ素早くアクセスすることを支援するための技術は、多数提案されている。例えば、ユーザが並び順をカスタマイズすることのできるメニューやアプリケーションランチャを搭載した携帯電話はよく知られている。また、最もよく使われる項目をメニューの最上位に置く MFU (Most Frequently Used) 法や、最近使われた項目を最上位に置く MRU (Most Recently Used) 法なども、しばしば用いられる方法である。これらの手法は、平均的には効果があると言えるが、その時々によってダイナミックに変化するユーザの状況（時間や場所など）には追従することができない。

我々が最終的に目指すところは、ユーザのその時々状況に応じて、適切なアプリケーション（あまり頻繁には利用されないものも含めて）に素早くアクセス可能な、「状況適応型」の UI を携帯電話上に実現することである。

携帯電話の UI には、単純なリスト形式のものや、タイル状のもの、アプリケーションランチャ、ショートカット、階層構造を持ったメニューなど、様々なものが存在する。ここに新たな使いやすい UI を提案することは一つの有効な方法であると考えられる。しかし、どのような UI にも、アクセスのしやすい（端的には必要なキー操作回数の少ない）場所と、そうでない場所が存在し、「どの場所にどの項目を配置するか」が問題となってくる。

我々は、多種多様な既存の UI の多くは、それが 1 次元的なものであれ階層構造を持ったものであれ、前記の「アクセスのしやすさ」に基づく単純な「並びモデル」に抽象化する事が可能であると考えている（図 1）。このモデルに基づくと、状況適応型の UI を実現する事とは、状況に応じて並びを最適化する問題に抽象化することができる。最適な並びを推定する事とは、個々の項目について、現在の状況下で使われそうな度合いを算出し、それに基づいて項目を並び替えることである。これによって、正解の操作を並びの最上位に推定することに失敗したとしても、並びの上位にあるほど、操作ステップ数を削減することができる。並びを最適化する事ができれば、様々な UI に「状況

適応性」を付加する事が可能である。

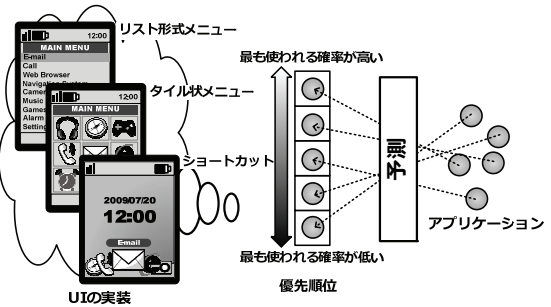


図 1 並びモデルと UI

携帯電話の操作予測に関しては、いくつかの先行研究がある。しかしその多くは、MFU 法でも比較的容易である、頻繁に使用されるアプリのみからただ 1 つの候補を推定するだけのものなどである。これらは、全ての項目を対象とした、並びの最適化問題には適用できない。

本論では、実際の商用の携帯電話において長期間（1～8ヶ月間）収集した約 50 名分のユーザの操作履歴（ログ）を解析し、機械学習によって、位置や時間といった観測可能な属性情報から、携帯電話のアプリケーションの並びを推定する事の実現可能性について論ずる。最終的に、あまり頻繁には使われないアプリケーションであっても、並びの上位に推定することができた。加えて、推定に必要な属性やデータ量、候補数について明らかにすることができた。

本論の構成は次の通りである。まず 2 章では、携帯電話のユーザビリティ向上についての関連研究について述べる。3 章では、我々のアプローチについて述べる。4 章では、実際の携帯電話の操作履歴を収集するロガーの設計、実装、およびデータ収集実験について説明する。5 章で、収集済みデータの解析と操作予測の実現可能性について議論し、6 章で結論を述べる。

2. 関連研究

主として物理的な制約から生じる携帯電話の使いにくさの問題については従来から認識されており、UI の設計、テスト、評価や、その方法論については、広く研究されている[1-3]。携帯電話は文字通り、ユーザによって様々な環境に携行され利用される[4]。そのため、携帯電話あるいは付随するセンサによって観測可能な情報から、ユーザ自身や周辺環境に関する、その時々状況（コンテキスト情報）を抽出することは、モバイルコンピューティング分野において共通の大きなテーマとなっている。関連する研究としては、例えば、コンテキストに応じた情報配信[5-7]、移動手段の予測[8]、健康管理[9]、電源管理[10]などが挙げられる。これらにはしばしば機械学習（例えば決定木やページアンネットワークなど）が利用される。しかしながら、機

[†](株) KDDI 研究所 KDDI R&D Laboratories Inc.

[‡] 富山県立大学 Toyama Prefectural University

機械学習によって携帯電話のユーザビリティを向上しようという試みについては、あまり見られない。

松本らは、位置や時刻に基づく、携帯電話のアプリケーションの並べ替え手法について提案している[11]。15種類のアプリケーションを位置・時刻情報とともにクラスタリングし、各クラスタ内でアプリケーションを利用頻度に基づいて優先順位付けしている。実験によれば、MFU法に比べて、正解のアプリケーションをメニューの高い順位に推定することができるが、携帯電話ではなく、PCにおける、1人の被験者の、短期間のログデータのみによる評価結果である。複数の被験者による、長期的ログデータを使用し、時間や位置以外の属性（バッテリー残量など）加えた検討が必要であると考えられる。

Bridleらは、携帯電話のテキストメッセージと音声通話と呼び出すのに要するキー操作回数を削減するための、ショートカット手法を提案している[12]。この中では、ContextPhone[13,14]プロジェクトによって公開されている、実際の携帯電話(Nokia Series 60)のログデータが使用されている。MRU法、MFU法、C4.5(決定木)、単純ベイズ法、フィッシャーの正確確率検定、および、MFU法と単純ベイズ法のハイブリッド手法が比較評価されている。しかし、予測対象が2種類だけであるため、全ての項目を対象とした状況適応型UIを実現するには不十分である。対象を頻繁に使われる特定のアプリケーションのみに限るのであれば、MFU法のような従来手法であっても、実用的な精度が得やすいと考えられる。あまり頻繁に使われないアプリケーションの推定は重要な課題であり、並びを予測するために不可欠である。さらに、彼らは1回の推定で1つの操作を予測することとどまっている。単一操作の予測は、予測失敗のリスクをはらんでいるが、多くの操作候補は逆にユーザを混乱させ、操作性を低下させる原因となる。そのため、安定的な精度を得るための最小限の候補数についての検討が必要である。

3. アプローチ

3.1 要件定義

機械学習によって、観測可能な属性値を用いて、MFU法などの従来法よりも精度の高い、状況適応性を持った優先順位付けを行うためには、以下の項目を明らかにする必要がある。

- 必要な属性：位置や時間などの観測可能な属性のうち、操作を精度良く予測するのに必要な属性の種類。
- 必要なデータ量：操作を精度良く予測するのに必要な操作履歴の学習データ量（学習期間）。
- 必要な候補数：UIが持つことのできる候補数が限られている場合（ショートカットなど）の、十分な予測精度を保つのに最小限必要な候補数。
- 個人差：利用スタイルの異なる複数の被験者による評価結果の違い。

3.2 アプローチ

前述の要件を明らかにするため、以下の手法を採用した。

- 数ヶ月に渡る、実際の携帯電話の操作履歴の収集を行う。操作履歴には、アプリケーションの利用（音声通話含む）、位置、時刻のほか、電波強度などの観測可能な情報群が含まれる。携帯電話実機上で操作履歴を収集するためのロガーが必要である。詳細は4章で述べる。
- 収集された操作履歴をオフラインで解析する。並びが機械学習によって推定され、精度や必要な属性などが評価される。機械学習に使われる属性情報には、前処理（離散化など）が必要であるが、高レベルなセマンティクスの抽出は行わない。一般に携帯電話は操作時間よりも無操作時間の方が支配的であるため、その間推定処理を行い続けるのは消費電力の観点からも無駄が多い。端末を開く動作は、ユーザが何か操作を行う際の明示的な合図であると捉えることができるため、推定トリガとして有用である。そこで、機械学習は、携帯電話が開かれた後の最初の操作を推定対象とする。単純化のため、ユーザが何もしなかった場合を含め、推定の失敗はユーザの操作を妨げる要因にはならないと仮定する。詳細は5章で述べる。
- 推定結果を利用したUIとしては、既存のものを想定する（典型的なリスト形式のメニューなど）。本論では、UI自体の認知性や快適性といった要素[20]を含めたユーザビリティ評価は対象としていない。

4. ログ収集

4.1 設計

携帯電話の操作履歴を収集するためのロガーを開発した。ロガーは、携帯電話の日常的な操作履歴に関するデータを、可能な限り歪めずに観測・収集する必要がある。ロガー設計にあたっては、以下を要件として定義した。

- 商用の携帯電話上で動作すること（実験用の特殊な携帯電話ではないこと）。
- 実際のサービス、アプリケーション、機能がそのまま通常通り利用可能であること（実験用のアプリケーションではないこと）。
- ログ収集によって携帯電話の機能性を阻害しないこと（機能の競合やパフォーマンスの著しい低下を起こさないこと）。
- アドレス帳の個人名やアクセスしたWebサイト名などのプライベートな情報は記録しないこと。

4.2 実装

開発したロガーについて説明する。ロガーは BREW 3.1 [15] 準拠のアプリケーションソフトウェアであり、auの携帯電話上で動作する。図2にロガーの構成を示す。ロガー自身も一般の BREW アプリケーションと同様の形態をとる。ロガーはバックグラウンドに常駐し、以下の情報を収集する。

- 操作：ユーザによって行われた操作（音声通話、Eメールの送信、Webブラウジング、その他一般のアプリケーション利用など）。キー操作のログは含まない。
- イベント：Eメール受信、音声着信、アラームなどのイベント。

- ・位置：GPSによって得られた、緯度経度で示される、ユーザの位置。被験者のプライバシーに考慮し、地図上にプロットしたり、意味のある地理的名称へ変換したりすることは行わないこととした。
- ・フリップの状態：折りたたみ式の携帯電話が開いているかどうかを示す値。
- ・電波強度：無線電波強度の値。5段階で示される。
- ・バッテリー残量：充電状態（充電しているかどうか）を含む、バッテリー残量を示す値。
- ・マナーモード状態：マナーモードになっているかどうかを示す2値情報。

全ての情報は携帯電話内部の不揮発性メモリに記録され、データ収集実験後、外部ストレージにコピーされる。

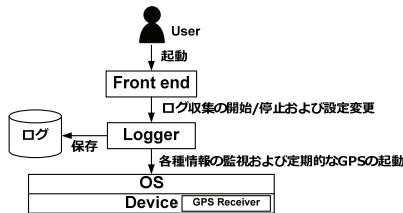


図2 ロガーの構成

4.3 実験

携帯電話の操作履歴を収集するため、データ収集実験を実施した。推定に必要なデータ量を評価するため、長期間の利用データが必要である。被験者数は20代後半～50代前半の会社員の男女48名で、実験期間は1～8ヶ月（被験者によって異なる）である。各被験者に、ロガーをインストールした携帯電話を普段通り使用してもらい、実験期間終了後、データを回収し、解析を行った。

5. 解析と考察

機械学習による操作予測の実現可能性を検証するため、収集した操作履歴情報の解析・評価を行った。

はじめに、1名の被験者(S-0)の長期間(約8ヶ月)のデータを解析し、その結果を、他の47名の被験者の短期間(1～3ヶ月)のデータを用いて検証した。解析は3.1章で述べた評価(必要な属性、データ量、候補数、および個人差)と、前処理から構成される。なお、解析中で使用する種々のアルゴリズムについては、Weka [16]のものを利用した。

5.1 前処理

事前に、収集された各種属性情報は、被験者別に、時間をキーとして統合する。各属性の値の1組が1つのインスタンスである。独立同分布(i.i.d.)を仮定し、インスタンス相互の時間順序などは考慮しないものとする。

まず、結合したデータに対し、次に説明する各種の前処理を実施した。解析に必要なないデータを、前フィルタにより除去した。解析に必要なないデータとは、システムプロセス(ロガー上はユーザプロセスと同様に観測される)、アプリケーション内部の操作に関わるもの(コピー&ペースト操作など)、欠損値(位置取得失敗など)を含むデー

タ、極端に利用回数が少ない操作に関するもの(全実験期間中、数回しか使用されなかったもの)、および、無操作時間中のデータである。前フィルタにより、以後の学習、推定、および評価は、ユーザが何らかの操作を行った場合についてのみに、行なわれることになる。

前フィルタに続き、機械学習に効果的であると考えられるいくつかの属性を追加した(表1, 属性追加)。オフィスからの距離のみ、事前知識を用いて計算した。前述の通り、高次のセマンティクスを表現する属性の抽出は行っていない。

表1 属性と前処理

属性	前処理
時刻	連続値化, 離散化
曜日	
緯度	離散化
経度	離散化
位置クラスタ	属性追加
オフィスからの距離	属性追加, 離散化
バッテリー残量	
電波強度	
マナーモード状態	
直近操作	属性追加
直近操作からの経過時間	属性追加, 離散化
直近イベント	属性追加
直近イベントからの経過時間	属性追加, 離散化
携帯電話を開いてからの経過時間	属性追加, 離散化

位置クラスタは緯度経度から k-means を用いて抽出した。各被験者のオフィスの所在地からの距離を用いるのは、オフィスは平日における活動の中心であると考えられるためである。24時間の時刻情報は0～1の連続値に変換した。直近の操作とイベント、およびそれらからの経過時間は、ユーザの操作に影響を与えると考えられる(例えば不在着信の後の発信など)。ほか、携帯電話のフリップを開いた後の経過時間も追加した。

機械学習アルゴリズムが扱えるように、全ての連続値属性は、KononenkoのMDL基準[17]に基づき離散化される。バッテリー残量、電波強度、マナーモード状態に関する情報ははじめから離散値として観測されるため離散化の対象とはならない。なお、フリップの開閉状態の情報は、直接属性としては使用せず、経過時間などを判断するフラグとして使用した。

次に、問題を単純化するため、2つの属性の値に基づく後フィルタによって、周期的・日常的であると考えられるデータを抽出した。後フィルタに用いる属性の1つは曜日であり、より周期性・規則性があると考えられる、平日(月～金曜)となる期間のデータのみを取り出した。2つ目として、オフィス所在地からの距離が極端に遠い(50km以上とした)場合のデータは除外した(例えば、旅行や出張などは非日常である事が多いため)。

最後に、各インスタンスに重み付けを行ない、各アプリケーションのデータ量を、それが使用された時間の長さに関わらず、使用回数に基づき等しくなるよう、正規化した。データサイズ削減のため、ロガーは値に変化があった場合のみ記録を行う(例えば電波強度ログの場合、その値が変化した場合のみ記録する)。そのため、結合後のログデータは、そのままでは記録密度の粗な部分と密な部分がある。

これを均一化する方法としては、時間を基準とする方法があるが、その場合、長時間利用されたアプリケーションほど、多くの学習データを持ってしまふ(=推定されやすくなる)ことになる。利用時間は、ユーザ個人の利用スタイルの他、アプリケーションの性質によっても大きく異なる(例えばブラウザは何十分も使う場合があるが、音量設定変更操作にはそれほどの時間はかからない)。推定は、その使用時間の長さに関わらず、1回の利用につき1回であるため、1回の利用に関するデータが利用時間に応じた個数、学習に使用されることは不相当である。そのため、均一化は回数ベースで行う。具体的には、ある1回のアプリの使用(起動から終了まで)に関する重み付きインスタンス数の総和が1.0となるようにした。

被験者 S-0 についての、各アプリケーションの、前処理完了後の重み付き利用頻度を図3に示す。アプリケーションの種類は15種類であった。この中で、特にアプリケーションCの利用頻度が支配的であるということがわかる。

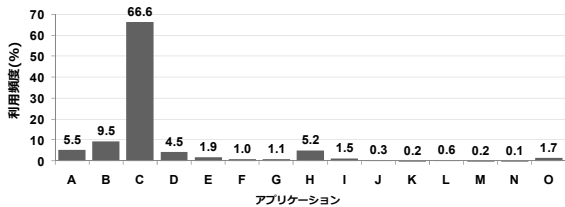


図3 アプリケーションの利用頻度分布

機械学習と従来の方式の比較評価のために、それぞれ代表的なものとして、単純ベイズ法(Naive Bayes, 以下 NB)と MFU 法(ZeroR)を採用した。単純ベイズ法とは、各属性が条件付き独立であるという前提に基づく分類器である。

5.2 属性選択

操作履歴収集実験においては、可能な限り多くの種類の情報を収集したが、実際に推定に有用な属性は限られていると考えられる。携帯電話上への実装を考えた場合、収集・保存される属性の種類は、データサイズの観点から、最小限にするべきである。属性選択は機械学習にとって共通の問題であると言える。

Wrapper アルゴリズム[19]を用い、属性の操作予測推定精度に対する寄与度を評価した。操作予測精度としては、ユーザが行いそうな操作ただ1つを推定してそれが正しかったかどうかを評価するのではなく、操作の並びを予測し、正解の操作がその並びの何番目に存在したかを評価した。ここで、「Rank」という指標を導入する。これは、事後確率分布に基づく順位であり、Rankの数値は、事後確率が大きいものほど小さくなり、「Rankが高い」と表す。実際の操作のRankが1であると推定されたということは、当該項目はリスト形式のメニューで言えば1番上(最もアクセスしやすい場所)に配置されたということの意味する。前述の通り、たとえ正解の操作のRankを1に推定できなくても、頻度学習でのRankが10のものをNBで3に推定することができれば、それには大きな意味がある。

評価は10分割交差検定によって行なった。図4に結果を示す。選択率とは、10回の検定中、その属性が推定に寄与した検定数の割合である。

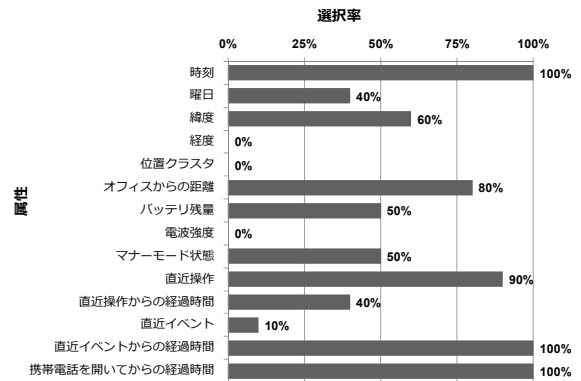


図4 操作推定精度に対する属性の寄与度

図4によれば、時刻は全ての検定で選択されたが、曜日はあまり選択されなかった。前フィルタで休日のデータは除外しているため、平日内の曜日間ではあまり操作に差がないためと考えられる。

位置に関する属性としては、位置クラスタは全く選択されなかったが、オフィスからの距離は高い割合で選択されていた。緯度と経度の寄与度にも大きな差があることから、おそらく、ユーザの位置が携帯電話の操作に何らかの影響を与えている事は間違いないが、その特徴量抽出手法に検討の余地があると考えられる。k-means法のパラメータか、代替するクラスタリング手法を用いることで寄与度が向上する可能性がある。

直近情報関連の属性については、特徴的な違いが見られた。直近操作の寄与度は高いが、その経過時間の寄与度は低い。また、直近イベントの寄与度は低い、直近イベントからの経過時間の寄与度は高い。言い換えると、経過時間に依らない「前回の操作内容」と、内容に依らない「直前にイベントが発生したかどうか」が、現在の操作に影響を与えているということがわかった。

以後の評価において、NBで使用する属性は、個々の学習データセットから、学習の度にWrapperアルゴリズムによって選択されるものとする。

5.3 精度評価

アプリケーション別の平均Rankを10分割交差検定により評価した。学習データには全てのデータが使用されるが、推定に際しては、携帯電話を開いた後最初に行った操作のみが対象となる事に注意されたい。

図5に結果を示す。横軸が平均Rankであり、左側に寄るほど(棒グラフが短い程)Rankが高く、良い結果である事を意味する。これによると、全平均Rank(図中、Avg.)は、ZeroR, NBともにほぼ同じ結果であった。これは、全平均は、データ中で大半の割合を占めるアプリケーションCの精度に引張られるためであり、単純にこれをZeroRとNBの性能と見なすことはできない(当然のことながら、ZeroRは頻度が高いアプリケーションに対しては高い精度で推定が可能である)。個々のアプリケーションのRankを見てみると、NBは、ZeroRではあまり高いRankには推定できなかったアプリケーションを、高いRankに推定できていることがわかった。例えば、アプリケーションF, G, I, L, Oは、図3(5.1章)によれば利用頻度は高く

ないが、その Rank は ZeroR と比べ 3 以上上昇していることがわかる。

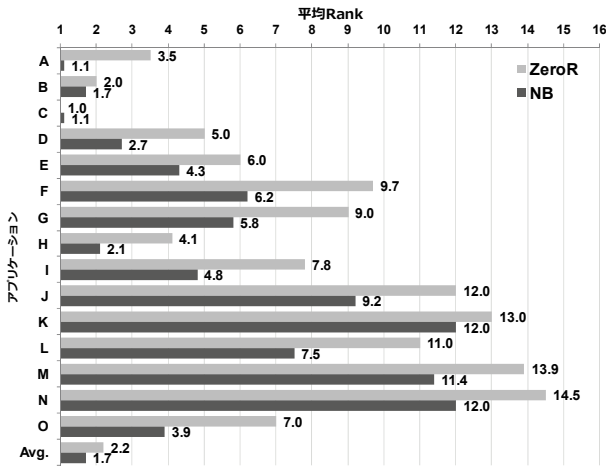


図 5 推定精度の評価結果

NBによる ZeroR からの Rank の改善度の違いは、アプリケーションの状況依存性の違いから生じていると考えられる。一般に携帯電話は常に携帯され、様々な状況下で利用されるため、利用頻度が高いということは、状況にあまり依存していないことを暗に意味していると考えられる。逆に、あまり頻繁に使われないアプリケーションは、特定の状況に強い関連性がある場合があると考えられる。

5.4 必要なデータ量

携帯電話の使い方が周期的で変化しない(例えば、ある1ヶ月間と次の1ヶ月間で、概ね同じような使い方である)という前提では、ZeroRの精度はデータ量(学習期間と言うこともできる)に依存せず一定である。これに対してNBはデータ量が不足していると十分な精度で推定を行なうことができない。しかし一方で、実際の実装を想定すると、学習に使用するデータ量は最小限にすべきである。そこで、学習に使用するデータ量を増やしていきながら、推定精度がどのように変化していくかを評価した。

テストデータセットは全データ中の最新の1ヶ月分に固定し、学習データセットのサイズを、テストデータの直前1ヶ月間から過去方向に段階的に拡張した。図6に学習曲線で結果を示す(アプリケーションK, M, Nはテストデータ中では1回も使用されなかったため省略している)。横軸が学習期間、縦軸がRankであり、図中、上側に寄るほどRankが高く良い結果であるという事を意味する。

これによると、学習データサイズを増やすと精度が上昇するアプリケーション群(学習可能群)と、逆に学習データサイズを増やしても精度がほとんど上昇しないアプリケーション群(学習困難群)の、2群に分類できることがわかった。学習可能群では、特に最初の1ヶ月で顕著な精度の上昇が確認されたが、アプリケーションAおよびCに関しては最初から比較的高い精度であることがわかる。学習困難群としては、アプリケーションB, E, F, J, Lについて、精度の上昇がほとんど確認できなかった。

これにはいくつかの原因が考えられる。アプリケーションBは、精度にZeroRとあまり差が無く(5.3章, 図5)

かつ横ばいであるため、頻度が高いことによりRank2を保っているものの、状況依存性は低いと考えられる。一方、アプリケーションE, F, J, Lについては、精度の上昇はほとんど見られないが、ZeroRと比較するとRankは高く、ある程度属性との因果関係はあると考えられる。これを改善するためには、前処理における特徴量の抽出手法、あるいはNB自体を別の手法に変える必要があると考えられる。

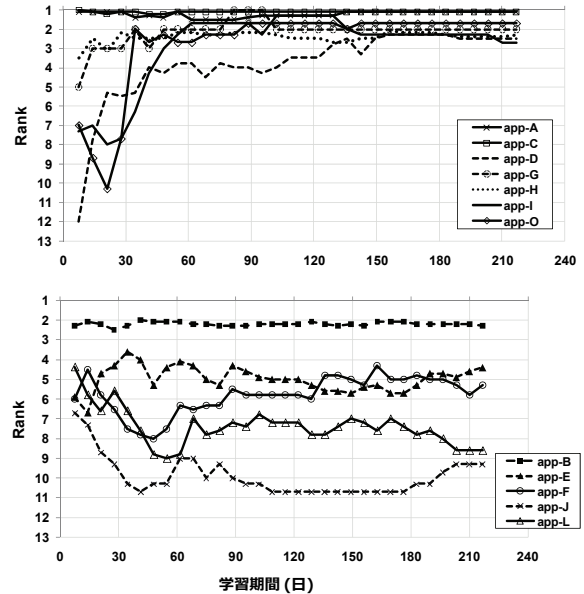


図 6 学習曲線(上:学習可能群, 下:学習困難群)

5.5 必要な候補数

UIの実装によっては、全ての候補を含む並びを許容していない場合がある。例えば、待受画面のショートカットアイコンなどの場合、全てのアプリケーションを並べるのは非現実的である。

そこで、候補数と精度の関係を評価した。正解の操作についてのRankの推定結果が、指定候補数内に収まっていれば正解、そうでなければ不正解とし、精度は、推定回数に対する正解回数の割合で示す。候補数を1から15まで変化させた場合の精度の評価結果を、図7に示す。

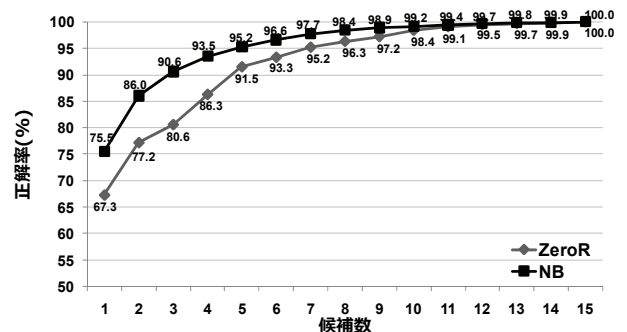


図 7 候補数と正解率

結果から、候補数が少ないほど、NBのZeroRに対する優位性(精度の差)が高いということがわかった。具体的

には、75%の精度を実現するために、ZeroR では 2 つの候補が必要であるのに対し、NB は 1 つの候補で十分であった。また、90%の精度を実現する場合でも、NB は 3 候補で十分であった(ZeroR は 5 つを要した)。ZeroR では、推定結果の並び順は概ね固定的であるため、頻度の順位が指定候補数以下のアプリケーションの推定は全て不正解となるが、NB の場合は与えられる属性に応じて変化させることができるため、候補数が制限されている場合であっても、ZeroR に対して高い精度で推定が可能であると考えられる。

5.6 個人差の検証

長期間データを用いた評価で得られたこれまでの結果を、他の 47 名分の短期間(1~3ヶ月)のデータを用いて検証した。

最初に、属性選択について、長期間データと同様に、被験者別に推定精度に対する属性の寄与度を評価した。図 8 に全被験者で平均した、各属性の選択率を示す。

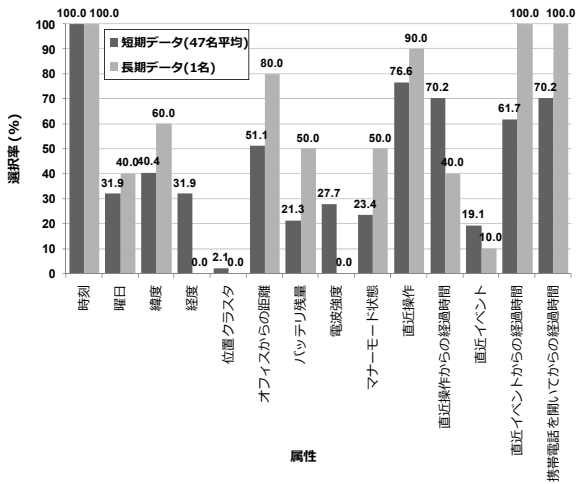


図 8 操作推定精度に対する属性の寄与度(短期と長期)

時刻や位置クラスについては、長期間データの結果と類似した傾向を示したが、その他に関しては違いも見られた。オフィスからの距離の選択率は減少した。これは被験者個々の環境の差から生じているものと考えられる。直近イベントからの経過時間についての選択率は下がる結果となったが、直近操作については長期間データ同様に高い結果となった。時刻と直近操作は、NB による操作予測に有効であると判断できる。47 名のデータで高い割合で選択され、かつ長期間データでは選択されなかった属性は、確認できなかった。従って、長期間データの被験者は特殊なケースではなかったと判断できる。

続いて、NB の精度を、ZeroR に対する Rank の改善度で評価した。前述の通り ZeroR で低い Rank となったアプリケーションが NB により大幅に改善する場合があったため、最も改善された/されなかったアプリケーションを、平均とともに図 9 に示した。図中、左側に寄るほど精度が改善されたことを意味する。平均的には改善度はほぼ 0 であったが、全体的には、最も改善されたアプリケーションの Rank 上昇は高く、かつ、最も改善されなかったアプリケーションの Rank 下降も小さな幅に抑えられていることがわ

かる。被験者 S-5 については、ほぼ 1 種類のアプリケーションしか利用していなかった(約 85%を占めた)ため、Wrapper アルゴリズムによって全ての属性が推定に寄与しないものと判断されており、結果として NB は ZeroR と等価となった。

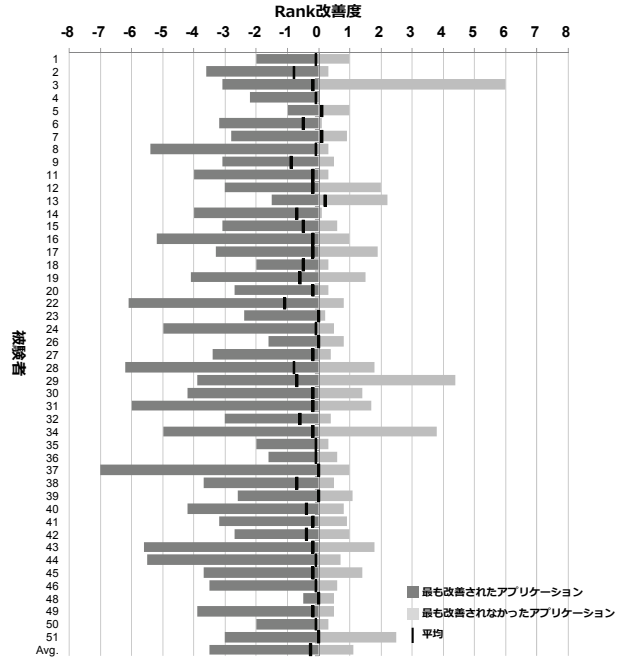


図 9 NB の ZeroR に対する推定精度の改善度

次に、学習に必要なデータ量を評価した。47 名の中から比較的長期間のデータが存在する 2 名を抽出し、評価した。被験者 S-1 と S-22 の学習曲線を、図 10 と図 11 にそれぞれ示す。

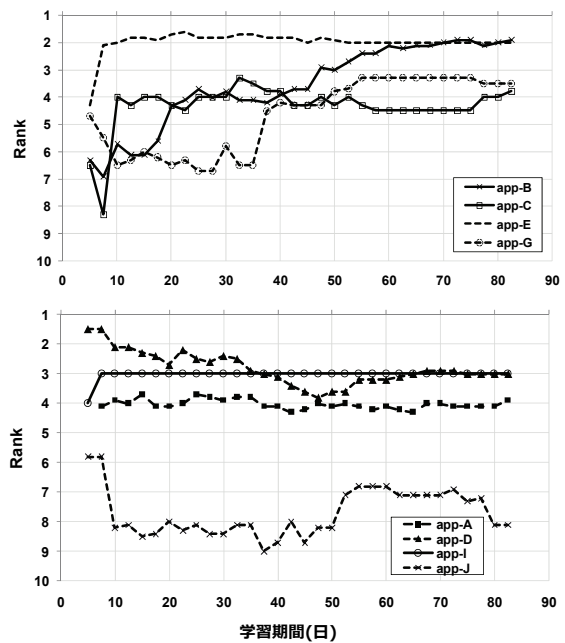


図 10 S-1 の学習曲線(上:学習可能群, 下:学習困難群)

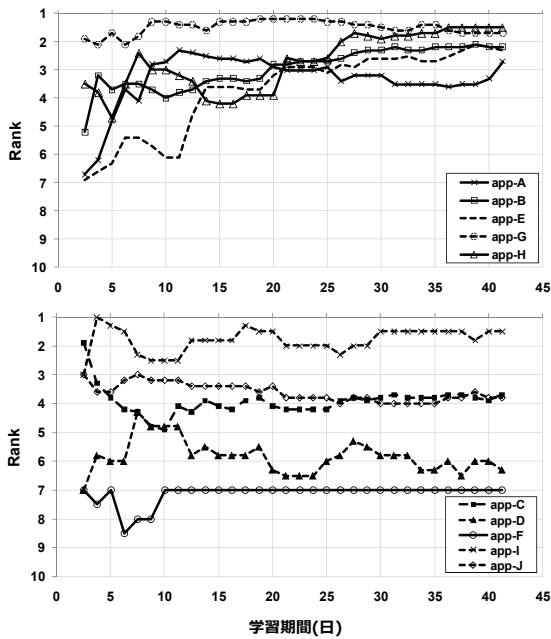


図11 S-22の学習曲線(上:学習可能群, 下:学習困難群)

両被験者とも、長期間データ (5.4章, 図6) 同様、学習データを増加させると精度が上昇する学習可能群と、そうでない学習困難群に分けることができた。学習困難群については、やはり ZeroR に対してはある程度良い精度であったが、さらなる精度向上は今後の課題である。

最後に、候補数と精度の関係について評価した。図12に、候補数を変化させた場合の平均精度の変化を示す(全被験者平均)。候補数が少ないほどNBのZeroRに対する優位性は高かったが、長期間データ (5.5章, 図7) ほど顕著な差は認められなかった。アプリケーション間の利用頻度の差があまり無かったか、あるいは、期間が短いため、NBが得意とする「頻度は低いが高精度」アプリケーションのデータの割合が少なかったためと考えられる。

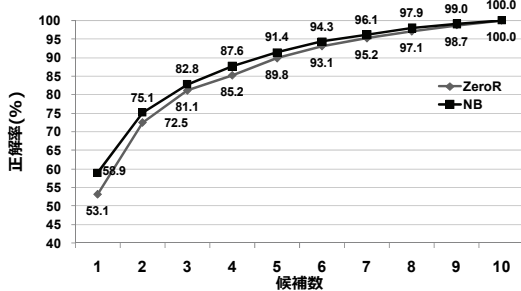


図12 候補数と正解率 (短期)

6. おわりに

本論では、機械学習を操作の並びを予測する方法として、長期的に収集した携帯電話の操作履歴を用いて解析・評価した。評価結果によれば、機械学習は、従来の方法では精度良く推定することが困難な、頻度の低いアプリケーションについても、比較的高い精度で推定することができた。特に時刻と直近操作属性が推定に有用であり、逆に位置情報は、現状の特徴量抽出方法では、あまり有用ではなかつ

た。十分なデータがあれば、90%の精度を得るためには3つの候補があれば十分であるということがわかった。しかしながら、長期的なデータを用いても、ある一定以上精度が上昇しないアプリケーションも存在することがわかった。

現時点では、最適なパラメータの探索などは行っていない。また、問題を単純化するため、周期的で推定が容易と考えられるデータのみを対象を限定している。最適なパラメータの決定や、ヒューリスティックの導入、今回対象としなかったデータ (休日など) への適用、認知性や快適性に関する検討などは、今後の課題である。

参考文献

- [1]Tomas Klockar et al., "Usability of mobile phones", the 19th International Symposium on Human Factors in Telecommunication (2003).
- [2] Harold Thimbleby, "Analysis and Simulation of User Interfaces", BCS Human Computer Interaction 2000, McDonald, S, Waern Y & Cockton G (eds), XIV: 221-237.
- [3]Kristian Kiili, "Evaluating WAP Usability: "What Usability?," Wireless and Mobile Technologies in Education, IEEE International Workshop on, vol. 0, no. 0, pp. 169, IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'02) (2002).
- [4]Ja Blom et al., "Contextual and Cultural Challenges for User Mobility Research", CACM, Vol.48, No.7, pp.37-41 (2005).
- [5]K. Hanamura, K. Kawabata et al., "Intention-Aware Information-Delivery System For Mobile Device", Fuzzy Systems (FUZZ) 2003, Vol. 2, pp.1110-1115 (2003).
- [6]Takeshi Nakatsuru, Koji Murakami et al., "Context-Aware Information Provision to the Mobile Phone Standby Screen," 7th International Conference on Mobile Data Management (MDM'06), vol. 0, no. 0, pp. 43 (2006).
- [7]D. Morikawa, M. Honjo, A. Yamaguchi et al., "A Design and Implementation of a Dynamic Mobile Portal Adaptation System", WPMC2005, Aalborg, Denmark (2005).
- [8]S. Reddy, et al. "Determining Transportation Mode On Mobile Phones", IEEE International Symposium on Wearable Computers (ISWC) (2008).
- [9]Maitland, J. et al., "Increasing the Awareness of Daily Activity Levels with Pervasive Computing", Pervasive Health Conference and Workshops (2006).
- [10] Ravi, N et al., "Context-aware Battery Management for Mobile Phones", Pervasive Computing and Communications (PerCom) 2008, pp 224-233 (2008).
- [11] Matsumoto, M et al., "Proposition of the context-aware interface for cellular phone operations", Networked Sensing Systems (INSS) 2008, 17-19 June 2008, pp. 233-233 (2008).
- [12] Bridle, R. and McCreath, E., "Inducing shortcuts on a mobile phone interface", the 11th international Conference on intelligent User interfaces (IUI '06), 2006, pp.327-329 (2006).
- [13] Context Project Dataset. Retrieved July 18, 2005, from Department of Computer Science and HIIT Basic Research Unit, University of Helsinki Web site: <http://www.cs.helsinki.fi/group/context>.
- [14] M.Raento et al., "ContextPhone: A Prototyping Platform for Context-Aware Mobile Applications", IEEE Pervasive Computing, 4 (2), 51-59 (2005).
- [15] BREW, Binary Runtime Environment for Wireless, <http://brew.qualcomm.com/brew/en>, Qualcomm.
- [16] Ian H. Witten and Eibe Frank, "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations", Morgan Kaufmann (1999).
- [17] I. Kononenko, "On biases in estimating multi-valued attributes", Proc. of the Fourteenth Int'l Joint Conf. on Artificial Intelligence (IJCAI'95), pp. 1034-1040 (1995).
- [18] T.M. Mitchell, "Machine Learning," McGraw-Hill (1997).
- [19] Ron Kohavi and George John, "Wrappers for Feature Subset Selection (late draft)", Artificial Intelligence journal, special issue on relevance, Vol. 97, Nos 1-2, pp. 273-324.
- [20] 黒須 正明, "ユーザビリティ概念の構造", ヒューマンインタフェースシンポジウム 11, pp. 351-356 (1996).