

# 携帯電話の利用傾向に基づくアプリケーション推薦法の提案

Recommendation Based on Users' Similar Habits of Mobile Applications

巖康平 福井健一 森山甲一 沼尾正行 栗原聡  
Kohei IWAO Kenichi FUKUI Koichi MORIYAMA Masayuki NUMAO Satoshi KURIHARA

大阪大学 産業科学研究所  
Osaka University ISIR

Mobile phones are enhanced recently. And so we can use various applications, for example, time table, weather information, photo-editing tool, etc. Applications relate to people's lifestyles, because they download and use applications they like or need in their lives. On the other hand, people must select the application they want to use from a complex menu structures, as cell phones are downsized. For this problem, they research means to estimate and recommend applications persons want to use. In this paper, we introduce a method to recommend applications from similarity of use logs among hours and spots.

## 1. はじめに

昨今の携帯電話は高機能化が進み、様々なアプリケーションを利用できるようになった。その一方で、小型故の操作の煩雑さを解消するため、ユーザーの所望するアプリケーションを予測し推薦する研究が行われている。

使用するアプリケーションは時間帯や場所に依存すると考えられる。一方、場所などが異なってもアプリケーションを利用する目的が同じであればこの操作も類似する可能性が高いが、従来の研究ではこの類似性は考慮されていない。本研究は個人の行動により時間帯や場所をグルーピングする方法を提案、評価するものである。本手法の応用により、ユーザーが初めて訪れるような場所であっても適切なアプリケーションの推薦が可能になると期待できる。

## 2. 本研究の位置付け

### 2.1 関連研究

上坂らの研究 [Uesaka 09] では、様々なコンテキストを利用し、ナイーブベイズを用いてユーザーの利用するアプリケーションの予測を行なっている。この手法は、比較として用いられた頻度ベースの手法に比べてある程度良い精度を示していたが、更なる精度向上が課題として述べられている。また、彼らの実験においては、時刻と直近操作属性が推薦に有用であり、反対に位置情報はあまり有用でなかったと、興味深い結論を出している。同じくナイーブベイズをアプリケーション予測に応用した研究には Sangoh らの研究 [Sangoh 09] が挙げられる。この研究ではナイーブベイズを用いた手法と共クラスタリングを用いた手法の二つのアプローチが提案されており、閾値によって推薦精度の優劣が入れ替わる結果が報告されている。また、Driss らの研究 [Driss 10] では、携帯電話の利用時刻に着目し、過去のデータにおいて、あるいはその日の別時刻のデータにおいて、人の行動予測に最も役に立つ時間帯の抽出を行っている。さらにあるユーザーから異なるユーザーの行動の予測も試みており、一定の条件下であれば予測も可能であると報告されている。

### 2.2 アプリケーション利用傾向の類似

先行研究として場所や時間による操作パターンや行動を抽出による適切なアプリケーションの推薦が試みられているが、これらの研究は各場所や各時間をそれぞれ独立に考えているが、個人の操作パターンは携帯端末を利用する目的によって類似することも考えられる。

例えば、あるサラリーマン“Aさん”は出勤時である“早朝”、“X駅”にて“時刻表アプリケーション”と“ゲームアプリケーション”を利用したとする(図1)。また、帰宅時である“夕方”、“Y駅”でも“時刻表アプリケーション”と“ゲームアプリケーション”を利用したとする。

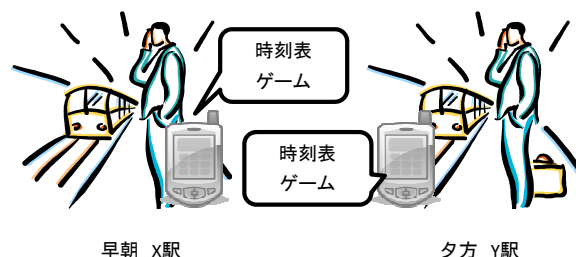


図1: 操作パターンの類似する例

この例の場合 Aさんは“早朝”や“夕方”といった時間帯や、“X駅”や“Y駅”といった場所によらず、出勤時や帰宅時の駅では時刻表とゲームを利用するという共通した目的が存在すると考えられる。上記は例えではあるが、Aさんに限らず、携帯電話の操作パターンや利用傾向には、そのときの目的によるグループのような存在がある可能性が高い。

本研究は、この類似性やグループが存在すると仮定し、携帯電話の利用履歴からこれらの発見を試みるものである。そして、このグループを利用した携帯電話におけるアプリケーションの推薦法を提案する。

連絡先: 栗原聡, 大阪大学産業科学研究所, 〒567-0047 大阪府茨木市美穂ヶ丘8-1, TEL:06-6879-8427, Fax:06-6879-8428, kurihara@sanken.osaka-u.ac.jp

### 3. 利用傾向に基づくアプリケーションの推薦

ここでは、先に説明した操作パターンの類似性を定義し、それを基にした共通した目的によるグループの発見方法について述べる。そして最後に、これを利用したアプリケーションの推薦方法を提案する。

#### 3.1 EF-ICF

初めに、類似性を考慮する際に用いた EF-ICF(Event Frequency - Inverse Context Frequency) について説明する。これは、文書中の特徴ご抽出に用いられる TF-IDF(Term Frequency - Inverse Document Frequency)[Peter 07] を基に松本らが提案した [Matsumoto 10] アルゴリズムである。具体的には、TF-IDF における文書をコンテキストに、ある単語をあるアプリケーションに置き換えて考えたものである。すなわち、あるコンテキスト内での利用回数が多く、かつ他のコンテキストではあまり利用されないアプリケーションを抽出するためのアルゴリズムである。すなわち、ef-icf 値が高いアプリケーションは、そのコンテキストでこそ使われるアプリケーションであり、そのコンテキストを表すようなアプリケーションとも考えられる。以下の式に EF-ICF の定義を記す。j はあるコンテキストを表し、i はあるアプリケーションを表す。また、 $freq(i, j)$  は j における i の利用回数、 $NoE$  は j 中のアプリケーションの種類数、 $NC$  はコンテキストの総数、 $Dreq(i)$  は i が登場するコンテキスト数を表す。

$$ef(i, j) = \frac{\log_2(freq(i, j) + 1)}{\log_2(NoE)} \quad (1)$$

$$icf(i) = 1 + \log_2\left(\frac{NC}{Dreq(i)}\right) \quad (2)$$

$$eficf(i, j) = ef(i, j) \times icf(i) \quad (3)$$

#### 3.2 操作パターンの類似性の定義

アプリケーションの利用傾向における、共通した目的によるグループを導くため、本研究では操作パターンの非類似度を定義する。

この非類似度には、先に紹介した“ef-icf 値”を用いる。ef-icf 値は、あるアプリケーションがその場所、その時間でこそ利用されると考えられる度合いを表す値である。反対に言えば、ある場所、ある時間において高い ef-icf 値をもつアプリケーションは、その場所やその時間を表すアプリケーションとも考えられ、その表し具合が ef-icf 値であると捉えることができる。この特徴を利用し、以下のように非類似度  $s$  を定める。また、以後ある時間帯のある場所をエリアと呼び、このエリアは前述のコンテキストと同義である。

$$s(a, b) = \frac{\sum_k |a_k - b_k|}{\sum_k a_k} \quad (4)$$

$s(a, b)$  は、エリア a でのアプリケーションの操作パターンが、エリア b でのアプリケーションの操作パターンにどの程度類似しているか、を表す値である。そして、 $a_k$  はエリア a におけるアプリケーション k の ef-icf 値であり、 $b_k$  はエリア b におけるアプリケーション k の ef-icf 値である。また、式の分母は正規化を行うためのものである。この非類似度  $s$  の値が小さければ小さいほど、エリア a でのアプリケーションの操作パターンが、エリア b での操作パターンに類似している可能性が高い。

#### 3.3 共通した目的によるグループの発見

先に挙げた非類似度  $s$  を基に、共通した目的によるグループを導く。共通した目的によるグループとは、先に述べた非類似度  $s$ (式 (4)) を用いて近傍グラフを作成することで得ることができるあるノードに対する集まりのことである。この場合のあるノードとは推薦の対象となるエリアである。あるエリアについて非類似度  $s$  の値が小さい、すなわち操作パターンに似た特徴が見られるエリアを集めることで、そのエリアに対し、携帯電話を使う目的が似通ったエリアを参照できる環境を構築する。このグループの形成、すなわち近傍グラフの作成については 3.4 節に後述する。

このグループを表す近傍グラフの例を図 2 に挙げる。

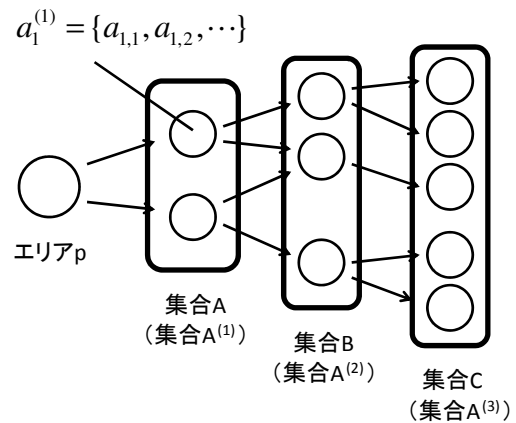


図 2: 近傍グラフの例

このグラフにおいて、エリア p から出る矢印によって直接結びつくノード a の集合を A とする。同様に、集合 A に含まれる各ノード a から出る矢印によって直接結びつくノード b の集合を B、集合 B に含まれる各ノード b から出る矢印によって直接結びつくノード c の集合を C とする。このとき“エリア p は集合 A に一次的に接続する”と呼び、同様に“集合 B に二次的に接続する”、“集合 C に三次的に接続する”と呼ぶ。また、例えばエリア p がエリア q に二次的にも三次的に接続している場合、これは二次的にのみ接続していると考えられる。

本論文では、この集合 A、集合 B、集合 C をエリア p の目的によるグループとして提案する。そして、このグループと定義するグループに属するノードの情報を用い、アプリケーションの推薦を行う。

#### 3.4 グループを用いたアプリケーションの推薦法

アプリケーションの推薦には、先のグループと名付けたグループと、ef-icf 値を用いる。これは前述のように、操作パターンの類似性を ef-icf 値を基に考えているためである。すなわち、各アプリケーションの ef-icf 値の差が小さい組み合わせを、操作パターンが類似していると考えているため、一方のエリアに対する推薦に、もう一方のエリアの ef-icf 値の高いアプリケーションを推薦すればよいと考えたためである。推薦の対象となるエリアを p とし、推薦の手順を以下に記す。

ステップ 1 先に述べた近傍グラフの作成を行う。初めに、非類似度  $s$  に関する閾値を以下のように定義する。

$$s(\mathbf{a}, \mathbf{b}) \leq \delta \tag{5}$$

次に、ノード  $p$  から出発し、最大ホップ数  $N$  に達するまで、式 (5) を満たす場合にノード間を連結していく。これによってエリア  $p$  についてのグループを表す近傍グラフが完成する。

ステップ 2 作成した近傍グラフと以下の式を用いて、エリア  $p$  に対するアプリケーション  $k$  の推薦度  $r_b(k)$  を求める。

$$r_p(k) = (1 - \alpha)p_k + \alpha \sum_{n=1}^N \frac{1}{n} \sum_{a_i^{(n)} \subset A^n} \sum_{a_k \in a_i^{(n)}} a_{i,k}^n \tag{6}$$

式 (6) における  $A^n$  はエリア  $p$  から  $n$  ホップ先のエリアの集合であり、 $a_i^{(n)}$  はその一つである。また、 $p_k$  はエリア  $p$  におけるアプリケーション  $k$  の ef-icf 値を、 $a_{i,k}^n$  はエリア  $a_i^{(n)}$  におけるアプリケーション  $k$  の ef-icf 値を表す。式中の  $\alpha$  はエリア  $p$  と近傍エリアに対する重み付けを、 $\frac{1}{n}$  は近傍内での重み付けを目的とするものである。

ステップ 3 この推薦度  $r_b(k)$  を全てのアプリケーションについて算出し、エリア  $p$  に対して  $r_b(k)$  の高いアプリケーション  $k$  から順に推薦を行う。

## 4. 実験と結果

### 4.1 実験に用いたデータ

本論文では、携帯電話の代わりに小型 PC を一人の被験者に約三ヶ月間持ち歩いてもらい、収集した行動履歴の解析を行った。エリアは計 80 箇所抽出された。アプリケーションのノイズ除去を手作業で行った後、各エリアに対するアプリケーションの推薦結果を用いて、操作の似たグループが作成できているかを評価した。

### 4.2 評価方法

推薦対象エリア  $p$  に対する近傍グラフの情報が、エリア  $p$  に対するアプリケーションの推薦に役立つことを検証するために、下記の二つの推薦順位 “Ranking  $p$ ”, “Ranking  $Np$ ” のアプリケーションの一致率  $d(p)$  を考え、これにより評価する。

- “Ranking  $p$ ”: 推薦対象エリア  $p$  のみの ef-icf 値を用いた推薦順位 (上位 5 位以内)、 $\alpha = 1$  のときの  $r_p(k)$  (式 (6)) を用いる。
- “Ranking  $Np$ ”: 推薦対象エリア  $p$  の近傍エリアの集合  $A^n$  のみの ef-icf 値から、重み付けして得られる推薦順位 (上位 5 位以内)、 $\alpha = 0$  のときの  $r_p(k)$  (式 (6)) を用いる。

$$d(p) = \frac{n_{r=p}}{n_p} \tag{7}$$

$n_{r=p}$  は、推薦対象エリア  $p$  に対する “Ranking  $p$ ” と “Ranking  $Np$ ” のうち、一致するアプリケーションの数であり、 $n_p$  はエリア  $p$  に対する “Ranking  $p$ ” に含まれるアプリケーションの数である。

以上の評価を近傍グラフの得られた各エリアに対して行った。

## 4.3 実験結果

本論文では、“ $\delta = 0.5$ ”, “ $N = 3$ ” の場合 (実験 1) と、“ $\delta = 0.25$ ”, “ $N = 3$ ” の場合 (実験 2) の二通りの条件で実験を行った。本論文には結果の一部を記す。図 3 は実験 1 での近傍グラフの一つであり、表 1 は実験 1 での推薦順位の一つである。

実験 1 における全体の平均一致率は  $d_{average} = 0.74$  (小数点第 2 位以下切り捨て)、実験結果 2 における全体の平均一致率は  $d_{average} = 0.75$  といずれも高い数値となり、類似するエリアを上手くグルーピングできたとと言える。

一方、閾値  $\delta$  が大きければ、より多くのエリアについて近傍エリアが得られる、すなわち推薦を行うことができ、また得られるグラフの大きくなることが分かった。より細かな実験を行うことで、最適な  $\delta$  を導く必要があり、最大ホップ数  $N$  についても同様である。

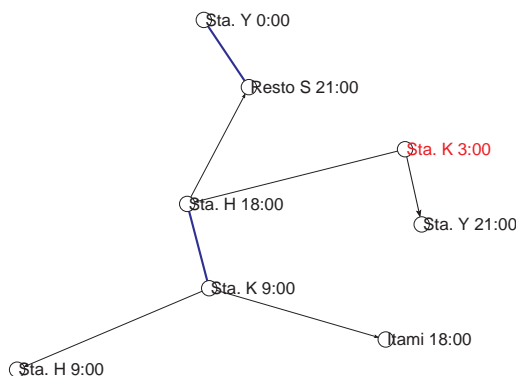


図 3: 実験 1 での駅 K 3:00-6:00 に対する近傍グラフ

推薦順位 “Ranking $Np$ ”	推薦順位 “Ranking $p$ ”
航空会社	航空会社
無線通信の設定	ウェブブラウザ
ウェブブラウザ	
無線接続関係	
ウェブメール	

表 1: 実験 1 での駅 K 3:00-6:00 に対する推薦順位

## 5. まとめ

本論文では、高機能化し複雑化する携帯電話の操作に対し、特に使用されたアプリケーションの類似性から推薦を行う手法を提案し、実験によってこのグルーピングの評価を行った。これにより時間帯や場所が異なる場合でも、携帯電話に求める事柄が同じエリアの情報を基に適切なアプリケーションの推薦が可能になると期待できる。今後の課題として、時系列に沿ってこのグルーピングを行っていき、ある場所や時間での情報が少ないときにも上手くグルーピングできるか評価する必要がある。また、今回実験に用いたデータにはあまり見られなかったが、出張先や旅行先がいくつか現れるようなデータに対しても評価を行う必要があると考えられる。

## 参考文献

- [Uesaka 09] 上坂大輔, 岩本健嗣, 村松茂樹, 横山浩之: 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. FIT2009(第8回 情報科学技術フォーラム), Vol. 4, pp. 23 - 29, 2009.
- [Sangoh 09] Sangoh Jeong, S Kalasapur, D Cheng, H Song, and Hyuk Cho: Clustering and naive bayesian approaches for situation-aware recommendation on mobile devices. Machine Learning and Applications, 2009. ICMLA 09. International Conference on, pp. 353 - 358, 2009.
- [Driss 10] Driss Choujaa and Naranker Dulay: Predicting human behaviour from selected mobile phone data points. UbiComp' 10 Proceedings of 12th ACM international conference on Ubiquitous computing, pp. 105 - 108, 2010.
- [Matsumoto 10] 松本光弘, 清原良三, 沼尾正行, 栗原聡: 位置情報を含む携帯端末利用履歴からのコンテキストに基づく最適アプリケーション推定法の提案. 情報処理学会数理モデル化と問題解決研究会 (MPS) 研究報告, Vol. 2010-MPS-80, No. 25, pp. 1 - 8, 2010.
- [Peter 07] Peter Sunehag: Using two-stage conditional word frequency models to model word burstiness and motivating tf-idf. Artificial Intelligence and Statistics (AISTATS 2007), pp. 1 - 8, 2007.